Published in final edited form as:
Ear Hear. 2019 ; 40(6): 1345–1358. doi:10.1097/AUD.0000000000000714.

Effects of Forward- and Emitted-Pressure Calibrations on the Variability of Otoacoustic Emission Measurements Across Repeated Probe Fits

Tom Maxim, **Christopher A. Shera**, **Karolina K. Charaziak**, **Carolina Abdala**
Department of Otolaryngology, University of Southern California, Auditory Research Center, Keck School of Medicine, Los Angeles, California, USA.

Abstract

Objective: The stimuli used to evoke otoacoustic emissions (OAEs) are typically calibrated based on the total sound-pressure level (SPL) measured at the probe microphone. However, due to the acoustics of the ear-canal space (i.e., standing-wave interference), this method can underestimate the stimulus pressure reaching the tympanic membrane at certain frequencies. To mitigate this effect, stimulus calibrations based on forward-pressure level (FPL) can be applied. Furthermore, the influence of ear-canal acoustics on measured OAE levels can be compensated by expressing them in emitted-pressure level (EPL). To date, studies have used artificial shallow vs deep probe fits to assess the effects of calibration method on changes in probe insertion. In an attempt to better simulate a clinical setting, the combined effects of FPL calibration of stimulus level and EPL compensation of OAE level on response variability during routine (non-contrived) probe fittings were examined.

Design: The distortion component of the distortion-product OAE (DPOAE) and the stimulus-frequency OAE (SFOAE) were recorded at low and moderate stimulus levels in 20 normal-hearing young-adult subjects across a five-octave range. In each subject, three different calibration approaches were compared: (1) the conventional SPL-based stimulus calibration with OAE levels expressed in SPL; (2) FPL stimulus calibration with OAEs expressed in SPL; and (3) FPL stimulus calibration with OAEs expressed in EPL. Test and re-test measurements were obtained during the same session and, in a subset of subjects, several months after the initial test. The effects of these different procedures on the inter- and intra-subject variability of OAE levels were assessed across frequency and level.

Results: There were no significant differences in the inter-subject variability of OAE levels across the three calibration approaches. However, there was a significant effect on OAE intra-subject variability. The FPL/EPL approach resulted in the overall lowest test-rest differences in DPOAE level for frequencies above 4 kHz, where standing-wave interference is strongest. The benefit was modest, ranging on average from 0.5 to 2 dB, and was strongest at the lower stimulus level. SFOAE level variability did not show significant differences among the three procedures, perhaps due to insufficient signal-to-noise ratio and non-optimized stimulus levels. Correlations

Address correspondence to: Carolina Abdala, University of Southern California, Keck School of Medicine, Department of Otolaryngology, Auditory Research Center, 1640 Marengo St., Ste. 326, Los Angeles, CA 90033, USA. carolina.abdala@usc.edu.

Conflicts of Interest

were found between the short-term replicability of DPOAEs and the benefit derived from the FPL/EPL procedure: the more variable the DPOAE, the stronger the benefit conferred by the advanced calibration methods.

Conclusions: Stimulus and response calibration procedures designed to mitigate the effects of standing-wave interference on both the stimulus and the OAE enhance the repeatability of OAE measurements and reduce their dependence on probe position, even when probe shifts are small. Modest but significant improvements in short-term test-retest repeatability were observed in the mid- to high-frequency region when using combined FPL/EPL procedures. Authors posit that the benefit will be greater in a more heterogeneous group of subjects and when different testers participate in the fitting and refitting of subjects, which is common practice in an audiology clinic. The impact of calibration approach on OAE inter-subject variability was not significant, possibly due to a homogeneous subject population and because factors other than probe position are at play.

INTRODUCTION

Otoacoustic emissions (OAEs), low-level sounds produced by the healthy inner ear, enjoy widespread use in clinics and research laboratories, including applications such as neonatal hearing screening, pediatric hearing evaluations, monitoring protocols for ototoxicity and chronic noise exposure, defining development and aging of the inner ear, and probing cochlear mechanics (see Shera and Abdala, 2012 and Abdala et al. 2016 for a review). While OAEs are useful probes of cochlear function, their utility is strongly impacted by variability among normal-hearing subjects (*inter*-subject variability) and within a single subject tested repeatedly over time (*intra*-subject variability). For instance, standard deviations of distortion-product (DP) OAE levels among a group of normal-hearing individuals range from 6 to 12 dB across frequency (f_2 = 0.75 to 8 kHz) at moderate stimulus levels (Poling et al. 2014; Abdala et al. 2018a). A large range of normative OAE levels can hamper classification of an ear as hearing or hearing-impaired during clinical evaluation. Therefore, methods that can control for sources of variability among individuals may improve clinical decision-making.

Intra-subject variability, on the other hand, affects the sensitivity of an OAE test in detecting changes in cochlear function within an individual over time. Test-retest repeatability of OAE results have been measured in various ways (e.g., Roede et al. 1993; Dreisbach et al. 2006; Lapsley Miller et al. 2006; Marshall et al. 2009; Reavis et al. 2015; Dreisbach et al. 2018). Reavis et al. (2015) performed a meta-analysis of 10 published studies examining DPOAE test-retest repeatability based on the standard error of measurement at 1, 2, 4, and 6 kHz. The normative 90% test-retest range, depending on the f_2 frequency and time between tests, was ±3.8 dB at day 1 and ±5.6 dB at day 20. Overall, the estimates suggest a normative range of DPOAE test-retest values exceeding 10 dB for higher frequencies and hovering around 8 dB for low-to-mid frequencies. Improving the test-retest repeatability of OAE measurements would increase their utility in detecting true changes in cochlear health over time.

Calibration Inaccuracies

The sources of OAE measurement variability are multifactorial and difficult to disentangle. The present study considers one possible contribution to this variability: the effects of ear-canal acoustics on both the stimulus and OAE pressures. Typically, the stimulus levels used to evoke OAEs are calibrated using in-the-ear measurements of total pressure at the probe microphone. This method—referred to here as the SPL calibration method—attempts to compensate for disparate ear-canal volumes among subjects by adjusting the earphone voltage to achieve target stimulus levels at the probe microphone (e.g., Siegel 1994). However, the total pressure at the probe microphone is not always a good predictor of the stimulus pressure reaching the eardrum. For example, the pressure at the probe microphone is affected by the presence of ear-canal standing waves that arise through the interaction between forward (i.e., traveling toward the eardrum) and reverse waves (i.e., stimulus waves partially reflected at the eardrum and traveling toward the OAE probe). Specifically, when the effective ear-canal length (i.e., the distance between the probe microphone and the eardrum) is equal to an odd multiple of a quarter of the sound wavelength, the forward and reverse stimulus waves at the probe microphone are out of phase and largely cancel one another. As a result, the family of so-called "quarter-wave pressure nulls" appears at the microphone but not at the tympanic membrane. These nulls cause problems for SPL-based calibration methods where the total stimulus pressure is controlled at the microphone location rather than at the eardrum. In particular, at frequencies near the quarter-wave null (and its odd multiples), the SPL-based calibration curve compensates for the null by dictating that the earphones be driven with inappropriately high voltages. Consequently, although the pressure at the microphone is well controlled, the pressure at the eardrum can exceed the target level by as much as 20 dB, leading to considerable imprecision in OAE estimates (Siegel 1994; Siegel & Hirohata 1994). The magnitude and frequency of these errors vary both from person to person—just as ear-canal lengths and shapes vary—and within an individual upon repeated testing due to changes in probe placement. In adult subjects, the quarter-wave null frequency ranges from 2.1 to 6 kHz (mean ~4 kHz), implying that the largest standing-wave-related errors occur in the high frequencies (Richmond et al. 2011; Reuven et al. 2013).

Alternative Calibration Procedures

Forward-pressure level (FPL) earphone calibration was developed to correct for standing-wave interference in estimating stimulus levels presented to the ear. In the FPL-based calibration procedure, forward-going stimulus waves in the ear canal are separated from reverse waves reflected off the eardrum so that the *forward-* rather than total sound pressure can be controlled at the microphone (Farmer-Fedor and Rabbit 2002; Scheperle et al. 2008). Thus, unlike SPL-based calibration, FPL calibration is not susceptible to errors arising from quarter-wave nulls. Multiple studies have shown that the dependence of both OAE amplitudes and audiometric thresholds on probe position in the ear canal is reduced when using FPL-based compared to conventional SPL-based earphone calibrations (Scheperle et al. 2008; Souza et al. 2014; Charaziak & Shera 2017). However, others have shown slight and somewhat inconsistent FPL-based improvements in DPOAE performance for the detection of hearing loss and no improvement when using DPOAEs in the prediction of audiometric thresholds (Burke et al. 2010; Rogers et al. 2010; Reuven et al. 2013).

Although FPL-based stimulus calibration methods help to minimize the undesirable effects of ear-canal acoustics on stimulus level, standing-wave interference also contaminates the evoked OAE. Multiple reflections within the ear canal artificially boost the measured OAE pressure (by ~10 dB) at frequencies near even multiples (including 0) of the half-wave resonant peak; the frequency of this peak depends on individual ear-canal length and the precise probe position within the ear-canal but is typically around 8 kHz in adult ears (Charaziak & Shera 2017). To help mitigate the effects on OAE levels, one can convert the measured OAE amplitudes to emitted pressure level (EPL). The conversion to EPL is equivalent to isolating the initial outgoing OAE wave at the eardrum from all its subsequent reflections within the enclosed ear-canal space (Keefe 1997; Charaziak & Shera 2017); the result is an estimate of the OAE as it would be measured in an anechoic ear canal of the same cross-sectional area. Just as the use of forward pressure calibrates the stimulus generation to reduce the effects of standing-wave interference, so the conversion to emitted pressure “calibrates” the recording of the OAE response. When combined with FPL-based stimulus calibration, the use of EPL-based response calibration eliminates the dependence of OAE levels on deep-vs-shallow ear-canal probe placements (Charaziak & Shera 2017).

In the present experiments, we test the effectiveness of three different combinations of stimulus/response calibration procedures for reducing the inter- and intra-subject variability of measured OAE levels: (1) using conventional SPL-based calibrations for both the stimulus and the OAE response (SPL/SPL); (2) using FPL-based stimulus calibration while employing conventional SPL-based OAE calibration (FPL/SPL); and (3) combining FPL-based stimulus calibration with EPL-based response calibration (FPL/EPL). We study how these three combinations (SPL/SPL, FPL/SPL, FPL/EPL) impact the variability of OAE levels measured across normal-hearing adult subjects, as well as within subjects upon repeated testing. Unlike previous reports that intentionally changed the probe position between measurements (e.g., shallow vs. deep), we asked our trained tester to fit the probe with the goal of achieving a best fit each time. This condition may better approximate differences produced by routine clinical fits across patients or in the same patient over time; consequently, our protocol may be more ecologically valid in assessing the translational relevance of FPL and EPL for OAE measurements.

Four factors distinguish the present study from previous evaluations of stimulus and response calibration procedures: (1) We examined calibration effects on OAEs generated via two distinct intra-cochlear mechanisms: Stimulus-frequency OAEs (SFOAE), a reflection-type emission, and DPOAEs, a nonlinear distortion emission; (2) We used the unmixed distortion component of the DPOAE rather than the total, mixed DPOAE; (3) We measured OAE variability both within and across subjects; and (4) We attempted to simulate pseudo-clinical conditions by having the tester focus on obtaining a good probe fit rather than contriving to produce a deep or shallow fit.

MATERIALS AND METHODS

Subjects

Twenty young adult subjects (11 F; 9 M) ranging from 22 to 28 years old (mean 25.3 years) participated in this study. Twelve right ears and 8 left ears were tested. All subjects denied any history of otologic disease, hearing loss, or chronic noise exposure. All had normal otoscopic exams and Type-A tympanograms (226-Hz probe tone) with peaks between ±50 daPa. Audiometric thresholds were ≤15 dB HL for frequencies tested at 1-octave intervals between 500 and 8000 Hz. Informed consent was obtained prior to participation in accordance with the Institutional Review Board (IRB) of the University of Southern California.

Instrumentation

Békésy audiometric threshold tracking and OAE testing were completed in a double-walled sound-attenuating IAC booth. A BabyFace Pro USB High Speed Audio Interface (RME Audio, Germany) and ER-10X probe system (Etymōtic Research, Elk Grove Village, IL) controlled by custom software written in MATLAB (Mathworks, Natick, MA) were used to generate stimulus waveforms and record the ear-canal pressures. Microphone voltages were amplified (+20 dB) and high-pass-filtered (300-Hz cutoff frequency) before A/D conversion. OAE testing was performed with the subject reclined in an ergonomic chair. The probe cable was suspended from the ceiling and the probe tip was carefully positioned into the ear canal with the goal of achieving a deep and stable fit, whereupon the cable was secured using a nylon headband. Subjects rested quietly or watched a subtitled video during testing.

Calibration Methods

This study compares OAE measurements obtained using three different methods of stimulus and response calibration. By “SPL-based stimulus calibration” we mean the conventional in-the-ear calibration technique that controls the *total* stimulus pressure at the probe microphone across subjects and frequencies. By contrast, FPL-based stimulus calibration corrects for the effects of ear-canal standing waves on stimulus level by controlling only the amplitude of the forward-traveling stimulus wave, separating it from any energy reflected from the eardrum (Scheperle et al. 2008). After recording the emissions, we employed two different “response calibrations”, expressing emission levels in either the conventional way using the *total* measured OAE pressure (SPL) or after compensating for standing-wave effects by extracting the emitted pressure (EPL).

SPL-Based Stimulus Calibration •—To implement the conventional, in-the-ear method of stimulus calibration based on the total ear-canal pressure measured by the probe, we fit the ER10X probe into the ear canal and presented moderate-level chirps to measure a stimulus calibration function. At any given frequency, the stimulus calibration function specifies the complex-valued pressure (in this case, the total pressure) produced at the probe microphone when a sinusoid of amplitude 1 volt is presented to the earphone. The conventional stimulus calibration function based on total pressure, denotedC_{stim}, thus has units of total complex pressure (in Pascals) per volt. During subsequent OAE measurements, the calibration function is used to determine the driving voltage needed to achieve target stimulus levels across frequency. The chirp calibration measurement and computation of the stimulus calibration function was repeated every three to four minutes throughout the OAE test session. Note that the stimulus calibration function also provides valuable information about the quality of the seal (e.g., the presence of low-frequency leaks), the location of the half-wave resonance peak, and the probe-insertion depth.

FPL-Based Stimulus Calibration •—The FPL-based stimulus calibration method controls the value of the forward ear-canal pressure rather than the total pressure at the microphone. We implemented the method by replacing the conventional stimulus calibration function,C_{stim}, with the forward-pressure calibration function,$C_{\text{stim}-\text{FPL}}$, defined by the equation

$$C_{\text{stim}-\text{FPL}} = \frac{C_{\text{stim}}}{1+R}, \tag{1}$$

where R represents the ear-canal pressure reflectance measured at the OAE probe microphone (Scheperle et al. 2008), as obtained from the chirp calibration measurements described above. To derive R, the Thévenin-equivalent OAE-probe parameters (source pressure and source impedance vs frequency) need to be known. Details of the Thévenin calibration procedures are described elsewhere (Scheperle et al. 2008; Charaziak & Shera 2017). In short, Thévenin-equivalent probe parameters were obtained daily in the ER-10X calibrator (brass tube, inner diameter 7.9 mm) at room temperature using five settings of the calibrator length (78.4, 64.8, 35.8. 29.7, 24.6 mm) with the goal of achieving total “source-calibration errors” less than 1 (see Scheperle et al. 2011). With the known probe parameters, the ear-canal acoustic impedance and corresponding characteristic impedance can be derived from the chirp calibration measurement and R can be calculated (Scheperle et al. 2008; Charaziak & Shera 2017).

EPL-Based Response Calibration •—Emission levels are conventionally reported by giving the total emission pressure measured by the microphone. Total pressures are obtained from the microphone output voltage using the microphone response calibration function, C_{resp}, which is also known as the microphone sensitivity curve and has units of volts per Pascal. In this paper, we use the term “SPL-based response calibration” to refer to the use of this conventional response calibration function,C_{resp}, to obtain the total OAE pressure. To obtain the *emitted* rather than the total OAE pressure from the measured microphone voltage, we simply replace the functionC_{resp} with the emitted-pressure response calibration function,$C_{\text{resp}-\text{EPL}}$, approximated by the equation,

$$C_{\text{resp}-\text{EPL}} \cong C_{\text{resp}} \frac{T\left(1 + R_s\right)}{1 - RR_s}, \tag{2}$$

where R is ear-canal reflectance,R_s is the probe reflectance, and T is the ear-canal transmission coefficient [see Charaziak and Shera (2017) for details]. Calculation of the emitted OAE pressure is readily performed *post hoc* using information obtained from the chirp calibration function and the Thévenin-equivalent probe parameters.

Protocol

After completing screening audiometry and tympanometry, subjects performed the Békésy threshold-tracking test, which took roughly 10 to 15 minutes (see Lee et al. 2012). Pulsed-tone stimuli for the Békésy audiometry were delivered through the ER-10X probe and presented at sequentially varying voltage levels. Subjects used an Aerb Mini Keyboard (Shiller Park, IL) to press and hold down a designated key whenever a pulsed tone was heard and release the key when the sound was no longer audible. The protocol required six ascending runs and the midpoints between reversals were computed for each. Threshold was reached when the standard error of the mean was less than 1 dB. Békésy thresholds were obtained at 0.5, 1, 1.5, 2, 3, 4, 6, 8, and 16 kHz. Note that the earphone voltage at threshold can be expressed using either the corresponding total pressure (SPL) or forward-pressure level (FPL). Although the primary objective of this study was to examine the impact stimulus/response calibration methods on OAE variability, we also provide a brief report on the impact of calibration on the stability of Békésy thresholds for the limited subset of subjects who returned for a second visit.

The OAE protocol was repeated in two consecutive test blocks (A and B). Each block contained the same eight conditions (DPOAE or SFOAE, with SPL- or FPL-based stimulus calibration, at low or moderate stimulus levels) and differed only in the order of presentation, which was determined randomly. The probe with its standard ear-tip was fit carefully at the start of each block but was not adjusted during data collection. At the conclusion of block A, the probe was removed and the subject was allowed a short break (about five to ten minutes) before refitting the probe and continuing with block B. Altogether, each session lasted approximately two hours, including roughly 40 minutes of actual OAE recording time. Five subjects were tested in a third block (block C) at a time one-to-three months after the initial test. The five retested subjects were selected based on their ability to sit quietly and their favorable OAE signal-to-noise ratios (SNR). A repeat of the Békésy threshold tracking was also performed during block C.

Recording and Stimulus Parameters

Both DPOAEs and SFOAEs were evoked with sweeping tones, which improves testing efficiency when compared to emissions recorded at discrete frequencies (Kalluri & Shera

2013; Abdala et al. 2015). Presenting the tones in concurrent “stacked” frequency segments further reduces data-acquisition time while providing close reproduction of data collected by means of a single-sweep (Abdala et al. 2018b). Consistent with pilot testing in normal-hearing young-adults, the number of sweeps required to obtain sufficient SNR ranged from 24 to 64.

DPOAEs at frequency $2f_1–f_2$ were evoked using a pair of tones, f_1 and $f_{2,}$ with the ratio f_2/f_1 fixed at 1.22. The stimuli were swept logarithmically upward from f_2 = 0.626 to 16 kHz at a rate of 1 octave per second. Three concurrent segments spanning 1.55 octaves each were presented simultaneously, with an overlap of 0.1 octaves between stacked segments. We applied phase-rotation averaging to cancel the primary tones before analysis (Whitehead et al. 1996). Three stimulus segments with different primary-tone starting phases (ϕ) are interleaved such that the primary tones f_1 and f_2 cancel when the responses are averaged and only the DPOAE at $2f_1–f_2$ remains: $p_{\mathrm{DPOAE}} = (p_{\phi \mathrm{I}} + p_{\phi 2} + p_{\phi 3})/3$. Tones were presented at two stimulus levels (“moderate” and “low”). When recorded using SPL-based calibration, the moderate stimulus level was defined as L_1, L_2 = 65,65 dB SPL and the low stimulus level condition was L_1,L_2 = 55,40 dB SPL. Primary-tone level separations were set using the so-called “scissors” method (Kummer et al. 1998). In the human ear canal, the total sound pressure expressed in dB SPL is roughly 3 dB greater than the forward pressure (FPL) at low-to-mid frequencies. Therefore, when recorded using FPL-based calibration, the moderate primary tones were L_1,L_2 = 62,62 dB FPL and the low level primary tones were L_1,L_2 = 52,37 dB FPL. Data collection stopped after 24 to 48 artifact-free sweeps had been obtained across frequency.

SFOAEs were measured with a probe tone (frequency f_p) presented from 0.5 to 16 kHz using a modified interleaved suppression paradigm (Shera & Guinan 1999; Abdala et al. 2018b). Responses to four stimulus combinations were measured: p_1 = probe tone alone, p_2 = probe and suppressor tone (+polarity), p_3 = probe tone alone, and p_4 = probe and suppressor tone (–polarity). The SFOAE time waveform was extracted from the 4 response waveforms using the formula: $p_{\mathrm{SFOAE}} = (p_1 + p_3 – p_2 – p_4)/2$. The probe and suppressor tones were swept downwards logarithmically at a rate of one octave per second. Five one-octave-wide frequency segments were presented concurrently in a stacked fashion to expedite data collection. As with the DPOAEs, SFOAEs were recorded at “moderate” and “low” probe levels (L_p = 40 and 20 dB SPL, respectively, or L_p = 37 and 17 dB FPL, respectively). The suppressor tone (frequency f_s) was presented at L_s = 60 dB SPL (or 57 dB FPL) and at a frequency slightly below f_p (f_s/f_p = 0.95). SFOAE data collection stopped after 64 artifact-free sweeps had been obtained across frequency.

Real-Time Glitch Detection and Offline Artifact Rejection

During data collection, each OAE time waveform was analyzed in the frequency domain using least-squares fitting (see below) and the median magnitude was calculated and updated with each new sweep. To ensure sufficient data for analysis after artifact rejection, any data point differing by 2 or more standard deviations from the current median OAE level was termed a “glitch” and triggered an additional sweep. Data collection stopped when all frequency points had reached the target minimum number of glitch-free sweeps.

Identification and rejection of true artifacts was performed offline during data analysis. OAE artifacts were identified as frequency points whose magnitude exceeded four (DPOAE) or five (SFOAE) standard deviations from the final median amplitude. These boundaries were determined from pilot work assessing the effect of point rejection on OAE SNR. Once identified, artifactual points were linked to the corresponding response waveform and a time segment centered around the artifact frequency and equal in duration to 10% of the analysis window was eliminated.

OAE Estimates

SFOAEs and DPOAEs were extracted from the recorded ear-canal signals using a least-squares fitting (LSF) technique applied to the recorded time waveform (Long et al. 2008; Kalluri & Shera 2013). Methods and rationale for the LSF procedures are detailed elsewhere (Abdala et al 2015; Abdala et al 2018b). Briefly, the OAE time waveform (i.e., p_{DPOAE} or p_{SFOAE}) was segmented into moving analysis windows that shifted in 0.01 octave steps. Models for the stimuli, suppressors, and OAEs were created. The amplitude and phase of the signals of interest within each analysis window were then estimated by minimizing the sum of the squared residuals between the model and the data to achieve the best fit.

The LSF model was applied at 100 points per octave resulting in OAE spectra consisting of ~500 points across the 5-octave test range. The noise floor at each frequency was estimated by averaging four LSF spectral levels computed at frequencies close to the OAE. For DPOAEs, the four noise frequencies were {0.90, 0.88, 0.86, 0.84} times the DPOAE frequency; for SFOAEs, they were {1.10, 1.12, 1.14, 1.16} times the SFOAE frequency. The LSF procedure employs analysis bandwidths (i.e., window durations) that vary continuously as a function of frequency with the goal of keeping constant the number of spectral fine-structure periods (for DPOAEs) or cycles of phase rotation (for SFOAEs) in each analysis window. For SFOAEs, the bandwidth of the LSF analysis window shifted from 0.16 (at 500 Hz) to 0.038 octaves (at 16 kHz). To improve the SFOAE estimation, a delay term was implemented in the analysis based on normative SFOAE delays reported in Shera et al. (2002). In this study, we separated the distortion component of the DPOAE from the reflection component. We removed the reflection component from the total DPOAE by employing a larger LSF bandwidth (~1.75 periods of DPOAE fine structure in each window). The larger bandwidth smooths the response, effectively eliminating long-latency energy associated with the reflection component. The DPOAE noise was passed through the same analysis window that extracted the distortion component and served as a reference for SNR measures. Henceforth, we use the term DPOAE to mean the separated distortion component of the total DPOAE.

With the goal of reducing the noise floor and eliminating uninformative sources of variability, we filtered the SFOAE measurements to focus on the primary reflection at the probe frequency (Shera and Bergevin 2012; Moleti et al. 2012; Abdala et al. 2018b). Analyses of SFOAEs have implemented various signal processing methods to eliminate longer-latency contributions (e.g., Konrad-Martin & Keefe, 2005; Moleti et al., 2012; Shera & Bergevin, 2012; Biswal & Mishra, 2018; Abdala et al., 2018b). We applied a time-domain filtering technique (inverse FFT) to extract the principal SFOAE while eliminating both probe/suppressor contamination (at shorter latencies) and multiple cochlear reflections (at longer latencies). SFOAE spectral data were resampled with 10-Hz frequency resolution and overlapping (50 Hz overlap) Hann-windowed segments of the SFOAE were transformed into the time domain for windowing. The time-domain windows were centered at times given by published SFOAE delay curves (Shera et al. 2002), $\tau(f)$, and varied with frequency according to a power-law function. The time windows used to extract the principal SFOAE spanned the region between the curves $\tau_{short} = 0.5\tau(f)$ on the short-latency side and $\tau_{long} = 1.5\tau(f)$ on the long-latency side, consistent with the work of Moleti et al. (2012). The windowed data were then transformed back into the frequency domain using the FFT. The SFOAE noise was passed through the same time-domain filter as the emissions and served as a reference for SNR measures. Although it lowered the noise floor and reduced the variability of the data, IFFT filtering had no effect on our findings or conclusions.

Statistical Analysis

We used three different measures to assess the effect of calibration method on the inter- and intra-subject variability of OAEs: (1) the variability of OAE level across subjects (standard deviations); (2) the short-term test-retest repeatability of OAE levels within an ear, quantified by comparing results from blocks A and B, which were recorded within the same session; and (3) the longer-term test-retest repeatability of OAE level within an ear, quantified by comparing Blocks A and C.

OAE levels were binned into one-third-octave frequency bands for analysis with center frequencies (CtrFrqs) ranging from 0.8 to 12.6 kHz. Despite the uncertainties associated with high-frequency acoustic Thévenin calibrations, Charaziak and Shera (2017) reported benefits consistent with theory when using FPL/EPL out to stimulus frequencies as high as 16 kHz; we employ a similar frequency range here. Inter-subject variability was evaluated by calculating the standard deviations (SD) of OAE level within each bin; 95% confidence intervals for the SDs were generated by resampling with replacement. Short-term test-retest repeatability was assessed by computing $\Delta AB\{OAE\}$, defined as the OAE level difference between blocks A and B ($OAE_A - OAE_B$). The mean absolute difference in level $|\Delta AB\{OAE\}|$ was binned into one-third-octave frequency bands for statistical analysis using repeated measures ANOVAs. Corrections to the criterion p value were applied to account for multiple comparisons across frequency.

Longer-term (1 to 3 month) test-retest repeatability was assessed in five of the original 20 subjects by computing $\Delta AC\{OAE\}$, defined as the OAE level difference between Blocks A and C ($OAE_A - OAE_C$). Although we do not report group statistics on this small subset, we examine and display individual data and trends from these five subjects. We hypothesized that the intra-subject variability of OAE amplitude would be larger for long- vs short-term replications hence any benefits conferred by the FPL/EPL calibration method would be greater for long- vs short-term repeated OAE tests.

Data Cleaning

Because artifactual data with poor SNR artificially inflate the variance, possibly obscuring the true effects of calibration method on test-retest repeatability, it was crucial to begin with a clean data set. Before analysis, OAE data were therefore cleaned by enforcing a minimum SNR criterion of 6 dB. In the low stimulus level condition, 7% of the available points were eliminated; in the moderate stimulus level, 4% were eliminated. Later, when these data were binned into one-third-octave frequency bands, any bin retaining less than 50% of its possible number of data points was eliminated altogether. Once the OAE data had been binned, any values further than 2 SDs from the mean were considered outliers and removed.

The cleaned OAE data set with outliers eliminated produced OAEs with mean signal-to-noise ratios as follows: DPOAEs evoked at low stimulus levels (the worst-case scenario for SNR) ranged from 15 dB SNR (at 12.6 kHz) to 33 dB SNR (at 6.1 kHz); the mean SNR of the SFOAE evoked at moderate stimulus levels ranged from 17 dB (at 12.6 kHz) to 33 dB (at 1.24 kHz). The standard deviations of the SNR ranged from 6 to 11 dB for both OAEs. Note that the low-level condition for the SFOAEs could not be analyzed for calibration effects due to the relative paucity of data at frequencies > 3 kHz that survived the cleaning process.

RESULTS

Inter-Subject Variability

Figure 1 shows mean DPOAE (left panel) and SFOAE (right panel) levels ±1 standard deviation (shaded regions) obtained in the moderate-level stimulus condition. The insets in each column display superimposed mean OAE levels for each of the 3 calibration conditions. The effects of FPL-based calibration on OAE level are evident in these insets. The results are consistent with SPL-based calibrations producing incorrect (and elevated) stimulus levels at the eardrum for stimulus frequencies near the quarter-wave nulls. (The reductions in OAE level caused by FPL-based calibrations led to insufficient SNR and the elimination of some CtrFrq conditions, primarily for the SFOAE.) Extracting the emitted OAE pressure decreases OAE levels further, although measurement SNRs are not affected. For both OAE types, the largest EPL-related reductions are observed at low frequencies (<4 and <2 kHz for DPOAEs and SFOAEs, respectively) and high frequencies (>10 kHz and >6 kHz, respectively). These findings are consistent with EPL corrections removing artificial boosts in OAE levels at low frequencies (due to closing the ear canal) and at high frequencies (due to the half-wave resonances; note that at f_2 = 10 kHz, the $2f_1$–f_2 = 6.4 kHz).

Standard deviations (SDs) of OAE level for the group were binned into one-third-octave frequency bands as a gauge of inter-subject variability. Figure 2 shows these SDs for DPOAEs and SFOAEs in the moderate-level stimulus condition. Although SDs increase at higher center frequencies, where the largest errors due standing-wave interference are expected, we found no significant differences across the three calibration methods, as indicated by the overlapping 95% CIs. Similar overlap was observed in the low-level stimulus condition (not shown). This finding suggests that intra-subject differences in ear-canal acoustics (or at least those compensated using the combined FPL/EPL calibration procedure) are not the main sources of variance in OAE level across these subjects. Note that the SFOAE data (bottom panel) recorded with moderate-level stimuli are only shown through 10 kHz because of poor SNR at the highest center frequency (12.6 kHz). For the low-level condition, SFOAE levels with sufficient SNR were only measurable up to 4 kHz;

unfortunately, this largely eliminates the frequency range where potential benefits of the advanced calibration techniques explored here might best be assessed.

Unlike the smoother spectra of the DPOAE distortion component, SFOAE spectra display a pronounced pattern of widely spaced peaks and notches (i.e., macrostructure). Minima in SFOAE spectra generally provide unstable estimates of level, shifting abruptly in response to small changes in stimulus level (Abdala et al. 2018b) or temporal variations in cochlear or middle-ear physiology. Because inclusion of these unstable data reduces mean SFOAE levels and SNR while potentially increasing variability across subjects, we undertook an alternative assessment in which only SFOAE levels near macrostructure peaks were included in the one-third-octave averages. For our purposes, SFOAE peaks were required to have at least 6 dB SNR and rise at least 2 dB above the adjacent troughs on either side. Although mean peak-picked SFOAE levels increased by a few dB compared to means calculated using the full spectra, the variability across subjects remained insensitive to calibration method.

One reason that OAE inter-subject variability appears little affected by calibration method may stem from our binning of the data into one-third-octave intervals with pre-determined center frequencies (CtrFrqs). As noted above, we expect the effects of stimulus calibration to be largest near frequencies corresponding to quarter-wave nulls and those of response calibration to be largest near half-wave resonance peaks. But because these frequencies vary across subjects, the use of fixed center frequencies may smear out and dilute the observable effects across bins. To examine that possibility, we identified the frequency of the first half-wave resonance peak in each individual ear canal. The half-wave peak frequencies ranged from 7 to 10 kHz across subjects, with a mean of 8.4 kHz. We estimated the first quarter-wave null frequency by dividing the half-wave frequency by 2. Mean OAE levels and their SDs across subjects were then re-calculated in half-octave frequency bands centered on each individual's measured quarter- and half-wave frequencies. Despite this more focused analysis, the effect of calibration method on OAE inter-subject variability remains insignificant in our data.

Although using the EPL-based response calibration method [Eq. (2)] to extract the emitted OAE pressure reduces the impact of standing waves on OAE levels, it does not compensate for variability in ear-canal cross-sectional area, another geometric property of the ear canal that influences measured OAE levels (e.g., by affecting the total volume of the residual ear-canal space). To explore whether inter-subject variability in ear-canal area contributes to the observed variability of OAE levels, we normalized the emitted pressures by the corresponding characteristic (or surge) impedance to obtain the emitted pressure that would have been measured in an anechoic ear canal of standard cross-sectional area (characteristic impedance of 84.5 CGS Ohms, corresponding to an 8-mm diameter tube). Although the use of this "normalized EPL" (nEPL) slightly reduces the scatter in OAE levels across subjects, the inter-subject variability (SDs) for three calibration methods (FPL/SPL, FPL/EPL, and FPL/nEPL) remains statistically indistinguishable. We explore additional reasons for this finding in the Discussion.

Short-Term Intra-Subject Variability

The OAE level difference ΔAB{OAE} between the recordings of Blocks A and B completed within the same session provides a measure of short-term intra-subject variability. We calculated this difference metric for each of the three calibration methods to assess whether improvements in intra-subject variability were observed when using FPL and EPL techniques. Overall, test-retest repeatability was outstanding, and ΔAB{OAE} values were small. Figure 3 shows mean values of |ΔAB{OAE}| for the three calibration methods in the moderate level stimulus condition for the DPOAE and SFOAE. For DPOAEs and SFOAEs, the mean short-term |ΔAB{OAE}| typically ranges from 1 to 3 dB across frequency for both low and moderate levels and for all calibration methods. There is a trend for slightly poorer replicability (i.e., increased |ΔAB{OAE}|) at higher frequencies and lower stimulus levels.

We conducted a two-factor repeated-measures ANOVA of calibration method (3) x frequency (13) on |ΔAB{OAE}|. Results for the DPOAEs recorded with low and moderate stimulus levels show an effect of calibration method (low: $F = 10.34$; $p = 0.004$; moderate: $F = 4.45$; $p = 0.0445$), and a marginal effect of frequency for low levels only (low: $F = 3.9$; $p = 0.062$; moderate: $F = 5.01$; $p = 0.036$) on |ΔAB{OAE}| with no interaction. We subsequently conducted one-factor ANOVAs at each center frequency separately, implementing a Bonferroni correction to accommodate multiple analyses. Statistically significant effects of calibration method on |ΔAB{DPOAE}| at each center frequency are marked with an asterisk on Fig. 4 for the low- (top panel) and moderate-level stimulus (middle panel) conditions. At the higher frequencies (≥ 4 kHz), the SPL/SPL method produced the poorest test-retest DPOAE repeatability among the three calibration approaches. The benefits of using either FPL/SPL or FPL/EPL calibrations produced an average improvement in DPOAE test-retest repeatability on the order of 0.5 to 2 dB relative to the SPL/SPL approach. We suspect that the modest size of these benefits can be explained by the consistency of probe fit achieved by our single tester. By striving to produce a stable and consistent fit for both Block A and B measurements, the tester may have minimized any changes in ear-canal acoustics, limiting the benefit obtainable from these calibration techniques. The small mean |ΔAB{DPOAE}| values achieved using conventional SPL/SPL calibrations support this notion. The fact that the fit and refit were conducted in the same session only ~1 hour apart likely contributed as well.

To further specify the calibration effects on intra-subject variability of the DPOAE, we conducted pairwise comparisons (with Bonferoni correction) at each of the center frequencies that showed an effect of calibration method. For the moderate-level condition, both FPL/SPL and FPL/EPL showed significantly less test-retest variability when compared to the SPL/SPL method at two center frequencies: 5 and 10 kHz; however, only the FPL/EPL showed significant improvement (vs SPL/SPL) in two additional frequencies: 6.3 and 12.6 kHz. For the low-level condition, at a center frequency of 4 kHz, the FPL/EPL condition showed significantly less intra-subject variability compared to SPL/SPL whereas the FPL/SPL condition did not. The |ΔAB{DPOAE}| never differed when the two advanced calibration methods were compared. These findings show that at some frequencies, the addition of the EPL correction to OAE level reduces DPOAE variability beyond the benefit conferred by simply calibrating the stimulus in FPL.

Values of |ΔAB{SFOAE}| for the moderate-level stimulus are shown in the bottom panel of Fig. 4. To ensure adequate SNR, we only analyzed the SFOAE data in the moderate stimulus-level condition. There was no significant effect of calibration method on the SFOAE level but there was a frequency effect ($F = 17.4$; $p < 0.001$); there was no interaction. Despite the absence of statistical significance, the overall pattern of results for |ΔAB{SFOAE}| at frequencies where the largest calibration effects are expected (CtrFrq ≥ 4 kHz) show noteworthy trends that match those predicted by theory. In particular, comparison of the three curves indicates that the benefits conferred by the FPL/EPL combination, i.e., where mean values of $|\Delta AB\{SFOAE_{FPL/EPL}\}|$ are the smallest, appear attributable primarily to the use of FPL in the frequency range associated with the first quarter-wave null (4–5 kHz) and primarily to EPL at higher frequencies associated with the half-wave resonance (> 6 kHz). Because of the multiple frequencies involved in the production of DPOAEs (f_1, f_2, $2f_1–f_2$), clear patterns like this, especially those associated with quarter-wave nulls, may be obscured in DPOAE data (Charaziak & Shera 2017). In the Discussion section, we consider possible reasons why there was not a statistically significant effect of calibration method on SFOAE level.

Correlations: A Look at Individual Subjects—Although mean |ΔAB{OAE}|s are small, some subjects within the group have relatively large differences, indicating less repeatable OAE recordings. To further explore the relationship between calibration method and OAE repeatability, we correlated DPOAE $|\Delta AB\{DPOAE_{SPL/SPL}\}|$, an indicator of the within-subject repeatability obtained using conventional calibration, with the benefit provided by FPL/EPL calibration in individual ears. Benefit was defined as $|\Delta AB\{DPOAE_{SPL/SPL}\}| – |\Delta AB\{DPOAE_{FPL/EPL}\}|$. We predicted that ears producing less repeatable DPOAEs (i.e. those with larger $|\Delta AB\{DPOAE_{SPL/SPL}\}|$) would show greater benefits from application of the advanced calibration methods. We computed correlations at the six center frequencies ≥ 4 kHz and found significant positive correlations for the low stimulus-level condition at four center frequencies: 4 kHz, 5 kHz, 6.2 kHz and 12.6 kHz. The correlations, shown in Fig. 5, explain between 32 and 89% of the variance. In the moderate-level condition (not shown), all 6 center frequencies show significant positive correlations and explain between 32 and 64% of the variance. Therefore, the more intra-subject variability there was to explain, the more effective the FPL/EPL calibration methods in improving OAE repeatability. It appears that mean FPL/EPL benefits observed in this study (ranging on average from 0.5 to 2 dB) were modest due to a ceiling effect; because most ears had small test-retest variability regardless of calibration method, the maximum achievable benefit was generally small. However, the correlations indicate that in ears with more variability, the benefit was substantially larger. The FPL/EPL calibration techniques worked as intended, conferring maximal benefit when there is variability in the OAE data due to differences in ear-canal acoustics between repeated measurements.

To assess the consistency of the probe fit between Blocks A and B, we considered another metric derived directly from the chirp calibration measurements. If the probe positions for Blocks A and B in a given ear are not identical, the calibration curves will differ. We calculated the frequency-dependent difference, ΔAB{CAL}, between the calibration curves taken at the beginning of Blocks A and B, as shown in the top panel of Fig. 6 (A and B are shown in blue and red, respectively; ΔAB{CAL} is shown in gray and plotted re: the right-hand ordinate). For each subject, we then correlated mean ΔAB{CAL} averaged over two different one-third-octave frequency bands centered on the half-wave peak and quarter-wave null frequencies, with the corresponding values of ΔAB{$\text{DPOAE}_{\text{SPL/SPL}}$} and ΔAB{$\text{DPOAE}_{\text{FPL/EPL}}$} averaged over the same frequency bands. The analysis was conducted in the moderate stimulus level condition only. We predicted that ΔAB{$\text{DPOAE}_{\text{FPL/EPL}}$} would show little dependence on ΔAB{CAL}, since changes in probe position from Block A to B should be appropriately corrected by the FPL/EPL calibrations. In contrast, we predicted that ΔAB{$\text{DPOAE}_{\text{SPL/SPL}}$} would correlate more strongly with ΔAB{CAL} and that, furthermore, the correlation should generally be *negative*—spurious (i.e., standing-wave related) increases in the calibration function result in smaller voltages being applied to the earphone, lower stimulus levels at the eardrum, and thus lower DPOAE levels.

Consistent with these predictions, we found significant negative correlations between ΔAB{CAL} and ΔAB{$\text{DPOAE}_{\text{SPL/SPL}}$} (Fig. 6, black fit; 27% variance explained) but not between ΔAB{CAL} and ΔAB{$\text{DPOAE}_{\text{FPL/EPL}}$} (Fig. 6, gray fit; 1% variance explained). This result suggests that FPL/EPL calibrations are better at correcting for changes in ear-canal acoustics after probe refitting than the conventional calibration approach. We should note that although the ΔAB{$\text{DPOAE}_{\text{SPL/SPL}}$} vs ΔAB{CAL} correlation followed the predicted pattern, the correlations were not especially robust (e.g., to the choice of frequency band). This is likely because the spread in ΔAB{CAL} was relatively small across subjects, consistent with our observation that the tester produced very uniform probe fits across measurement blocks.

Long-Term Intra-Subject Variability

Five subjects were retested with the same protocol (block C) 1 to 3 months after the first recording session (block A) to assess the long-term repeatability of OAEs recorded with the different calibration approaches. To determine whether overall test-retest stability of the OAE differs for short- and long-term repeats, we compared ΔAB{OAE} to ΔAC{OAE} at the 6 highest center frequencies. In general, the longer time interval between the probe re-fitting resulted in poorer overall OAE repeatability (i.e., |ΔAC{OAE}| > |ΔAB{OAE}|). Among the five subjects, |ΔAB{DPOAE}| ranged from 0.5 to 1.5 dB across CtrFrq whereas |ΔAC{DPOAE}| ranged from 1 to 3.6 dB for both stimulus level conditions. This increase in DPOAE intra-subject variability with time between tests is evident in Fig. 7 for 1 subject in the SPL/SPL (top panel) and FPL/EPL (bottom panel) conditions.

As with |ΔAB{DPOAE}|, the values of |ΔAC{DPOAE}| were typically smaller for FPL/EPL calibrations than for SPL/SPL calibrations. Overall, the mean benefit of using FPL/EPL over conventional SPL/SPL was just over 0.5 dB, but it was consistently present at each CtrFrq. Therefore, whether the retest is done within a session or after 1 to 3 months, the FPL/EPL approach improves the repeatability of OAE levels. Although we predicted that the FPL/EPL benefit would be greater for the long-term (|ΔAC{$\text{DPOAE}_{\text{SPL/SPL}}$}| − |ΔAC{$\text{DPOAE}_{\text{FPL/EPL}}$}|) compared to the short-term (|ΔAB{$\text{DPOAE}_{\text{SPL/SPL}}$}| − |ΔAB{$\text{DPOAE}_{\text{FPL/EPL}}$}|) repeats, this enhanced benefit is difficult to quantify because the benefits are small to begin with. The trend, however, is in the correct direction in 9 out of 10 measures: Slightly greater FPL/EPL benefits are seen in long-term vs short-term analyses of DPOAE and SFOAE repeatability, as shown in Fig. 8 for the five subjects.

Intra-subject Variability of the Audiogram

Because we had Békésy threshold tracking data available for blocks A and C in five subjects, we examined the repeatability of hearing thresholds expressed in SPL vs FPL. To this end, we calculated the differences, $\Delta AC\{BEK_{SPL}\}$ and $\Delta AC\{BEK_{FPL}\}$, between Békésy thresholds obtained during blocks A and C and expressed in either SPL or FPL. For frequencies above ~2 to 3 kHz, $|\Delta AC\{BEK\}|$ was often larger for thresholds expressed in SPL than in FPL, most notably in subject 003 shown in the upper panel of Fig. 9. The FPL benefit, defined as the difference between $\Delta AC\{BEK_{SPL}\}|$ and $\Delta AC\{BEK_{FPL}\}|$, ranged from a few dB to nearly 15 dB in this subject. Two out of five subjects did not show strong improvements in the repeatability of their Békésy thresholds when stimulus was expressed in FPL, as noted for subject 012 shown in the lower panel of Fig. 9. However, whenever there was a difference of at least 5 dB between thresholds expressed in FPL and SPL, the FPL-calibrated Békésy threshold was more repeatable than the SPL threshold.

DISCUSSION

Intra-Subject Variability of DPOAEs

Past work has shown improvements in DPOAE test-retest repeatability with FPL stimulus calibration when the probe position in the ear canal was intentionally varied between deep and shallow. Here, we tested whether improvements were also observed when the probe refitting was done simply with the goal of achieving a “good fit,” rather than producing a clear change in fit between two block measurements. Our results indicate that the use of forward- and emitted-pressure calibrations significantly improves the repeatability of OAE measurements, although the improvements are necessarily modest when the test-retest variations in probe fit are small (Fig. 4). We also observed improvements in the long-term repeatability of OAEs (Figs. 7 and 8).

One reason for the modest improvement in OAE repeatability may be that the probe fits changed little between measurements so there was little potential benefit to be had. As illustrated in Fig. 6, the ear-canal acoustics in the majority of our subjects remained fairly stable between blocks A and B (see tight distribution of $\Delta AB\{CAL\}$ around zero in the lower panel). As further support for this notion, DPOAE level differences from block A to B were smaller than test-retest figures reported in the OAE literature. Our unusually low intra-subject variability limited the potential benefits of using the FPL/EPL calibrations. For DPOAEs recorded using SPL/SPL calibrations, we recorded mean absolute test-retest differences within the same ear of 0.8 to 3 dB for center frequencies ranging from 0.79 to 12.6 kHz. To compare our measures of OAE intra-subject variability to those in the literature, we multiplied the group standard deviation of $\Delta AB\{DPOAE_{SPL/SPL}\}$ by 1.64 to generate a normative 90% range of accepted test-retest differences, much like those derived by Reavis and colleagues (2015), who conducted a meta-analysis of ten different studies. In our data, 90% of the DPOAE test-retest levels lie within ±2.1 dB at 2 kHz, ±1.3 dB at 4 kHz,

±2.5 dB at 8 kHz, and ±3 dB at 12.6 kHz. These 90% ranges are significantly smaller than those reported in previous literature (see Reavis et al. 2015), even for very high-frequencies (e.g., Dreisbach et al. 2006, 2018).

Intra-Subject Variability of SFOAEs

In our measurements, SFOAE repeatability was not significantly affected by calibration method. At least in part, however, this is due to the unreliability (poor SNR) of SFOAE measurements in the low-level condition above 4 kHz. One consequence of using FPL rather than SPL stimulus calibration is that relative OAE levels (and therefore SNRs) are reduced at frequencies where SPL calibration produces larger-than-intended stimulus levels at the eardrum (see Fig. 1 insets). Even in the moderate-level condition, however, where SNR was adequate at all but the 12.6 kHz CtrFrq, SFOAE test-rest repeatability did not improve in the FPL/EPL calibration condition. We suspect this is a consequence of the saturation of SFOAE growth curves. For the stimulus parameters used here, SFOAE levels begin to grow compressively at sound levels that are 8 to 10 dB lower than those producing compressive growth in DPOAE levels (Abdala & Kalluri 2017; Abdala et al. 2018b). Average SFOAE “compression thresholds” are approximately 33 to 35 dB SPL, depending somewhat on frequency. At our “moderate” stimulus level (40 dB SPL), which falls in the compressive region of the growth curve, SFOAE levels are relatively insensitive to changes in stimulus level and are therefore also insensitive to small changes in probe fit and/or calibration method. (In our low-level condition, where SFOAE levels would presumably have been more responsive to calibration effects, the SNRs became problematic.) In retrospect, a better choice of probe level for the SFOAE measurements in this study might have been something like 30 dB SPL (27 dB FPL), a level below the compression threshold but with sufficient SNR for reliable measurement.

Inter-Subject Variability of OAEs

We found no significant reductions in the variability of either DPOAE (Fig. 2) or SFOAE levels across subjects when using either the FPL/SPL or the FPL/EPL calibration methods. Differences across subjects in ear-canal anatomy and probe insertion should produce varied standing-wave patterns. One would expect these variations to affect both stimulus calibration and OAEs, producing differences in measured OAE levels. Theoretically, these inter-subject differences due to variation in probe insertion can be reduced although perhaps not entirely eliminated by using FPL/EPL calibration methods. Failing to observe significant reduction of OAE variability across subjects with FPL/EPL is therefore perplexing. It suggests that variations in ear-canal acoustics were not the main source of variability in OAE levels, at least within our cohort of subjects. Other factors left uncontrolled by our calibration methods but potentially contributing to inter-subject variability include: differences in the integrity of outer hair cells among normal-hearers, sub-clinical differences in hearing thresholds, differences in cochlear nonlinearities and OAE growth curves, differences in forward and reverse middle-ear transmission, subject sex and age, among others. Individually or in combination, these factors may well have provided the dominant sources of variation among OAE level estimates.

The lack of calibration-method effect on OAE inter-subject variability may also stem from the homogeneity of our subject group. Although we were initially surprised by the apparent insensitivity of OAE inter-subject variability to calibration method, in retrospect, the homogeneous nature of our subject group (all normal-hearing, young-adult college students aged 22 to 28 years) likely produced a sample with relatively limited variation in ear canal shape, length, and volume. Limited physical variation would naturally reduce the amount of variability attributable to ear-canal acoustics and decrease the impact of advanced calibration techniques. To explore this idea, Fig. 10 compares SDs from DPOAE levels in the current database levels to SDs calculated for DPOAEs obtained from more diverse subject groups (Poling et al. 2014; Abdala et al. 2018a). The data set from Poling et al. includes unmixed DPOAEs from 350 normal-hearing subjects ranging from 10 to 65 years of age measured at primary levels of 65,55 dB SPL; the second data set, from our own laboratory, included the separated distortion component of the DPOAE measured at primary levels of 65,65 dB SPL in 77 subjects ranging from 18 to 76 years of age. Although the DPOAE measurement paradigms differ somewhat between studies, matching recording parameters as closely as possible reveals that the current DPOAE data (using conventional calibration) has SDs 3 to 4 dB smaller than those of the other two studies. Whereas our current SDs range from 4 to 8 dB (mean = 5.3 dB), those of Poling et al. range from 6 to 12 dB (mean = 8.8 dB), and those of Abdala et al. from 7 to 11 dB (mean = 8.5 dB) over a comparable frequency span. (The slightly more variable data of the Poling et al. study may reflect their use of the total DPOAE rather than its unmixed distortion component.) Evidently, the variability of OAE level in our cohort was atypically small at mid-to-high frequencies, where the largest benefits from advanced calibration techniques are expected. Thus, by limiting our sample to young-adult students we may have encountered a ceiling effect. Of course, many factors other than ear-canal acoustics presumably contribute to the scatter of OAE amplitude across subjects, further complicating the issue.

Clinical Simulation

Even though we did see significant improvements in the repeatability of the DPOAE with the implementation of FPL/EPL methods, we expect the benefits would have been larger with a diverse subject group and with different testers doing the initial and repeated probe fittings. Correlations support this idea since the larger the intra-subject variability, the more benefit observed from the use of FPL/EPL calibration methods (see Fig. 5). One might therefore wonder whether our experimental paradigm effectively simulated clinical conditions.

In one respect, our data more closely simulated clinical audiology than many of the initial studies of FPL calibration. We opted for a more “natural” variation of probe fit and did not attempt to artificially induce varied patterns of standing-wave interference by manipulating the depth of probe insertion. Instead, we trained a highly motivated tester to achieve the best fit he could each time he tested. In practice, this method produced relatively small differences in probe position across and within subjects upon repeated testing. The values of ΔAB{OAE} we observed were therefore presumably more representative of clinical practice than those observed when employing intentionally deep and shallow fits likely to produce large FPL-related benefits. However, our study failed to simulate a clinical setting in other important ways: For example, patients in a typical audiology clinic generally span a diverse age range (infants, children, young, and elderly adults). This heterogeneous group would produce higher variability in OAE levels across ears, some of which, no doubt, would be due to differences in ear-canal length and probe insertion. Our group did not have such diversity.

Additionally, one might also expect that patients undergoing serial OAE monitoring (e.g., for chemotherapy or noise exposure) would, over time, be tested by multiple audiologists with varying levels of experience and/or different approaches to probe fitting. Typically, the time between such serial tests would also be on the order of months or years and not the minutes assessed here by our measure of short-term intra-subject variability. Indeed, both our analyses and those of others (Reavis et al. 2015) suggest that test-retest variability increases with time between tests. Although our study offers a more realistic test of calibration effects than those imposing artificially deep and shallow fits (e.g., Scheperle et al. 2008; Charaziak & Shera 2017), neither the homogeneous subject population, the use of a single tester for all fitting and refitting, nor the short time interval between tests is especially representative of OAE assessments conducted in a clinical setting.

For these reasons, we conjecture that larger benefits of FPL/EPL calibration would be observed in more realistic clinical conditions. Even with the atypically small variability we saw across and within subjects, significant improvements were achieved when using FPL stimulus calibration. Consistent with past work (Charaziak & Shera, 2017), additional benefits were achieved when using the FPL stimulus calibration combined with the EPL correction to DPOAE level. This suggests that the optimal calibration strategy is a combination of both FPL and EPL methods.

Acoustic Calibrations in the Human Ear Canal

In many respects, the development and validation of reliable procedures for determining the Thevenin-equivalent source parameters and ear-canal reflectances required to implement FPL- and EPL-based stimulus and response calibrations remain a work in progress. Even during the relatively short period since this study was initiated, new methods have appeared that promise to help circumvent some of the difficulties—including contamination by non-propagating, higher-order acoustic modes (evanescent waves), measured ear-canal reflectance magnitudes exceeding unity at some frequencies, and other troubles—that confound accurate measurement of acoustic impedance and reflectance in the human ear canal, especially at high frequencies (Norgaard et al. 2017; Norgaard et al. 2018; Siegel et al. 2018). Nevertheless, as demonstrated here and elsewhere (e.g., Scheperle et al. 2008, 2011; Charaziak and Shera 2017), the benefits conferred by the use of FPL and/or EPL-based procedures appear robust to the as-yet imperfect nature of the acoustic calibrations. We expect these benefits will only increase as methods of calibration improve over time.

Previous Findings

Our results are consistent with previous work investigating the effects of FPL stimulus calibration on DPOAEs. Scheperle and colleagues (2008) found that FPL was a more reliable stimulus calibration method than SPL, reducing the variability of changes induced by contrived differences in probe-insertion depth. Subsequent work attempted to translate this effect into enhanced clinical utility, which proved more elusive. Reuven et al. (2013) found that when OAE measurements were taken at standing-wave frequencies (rather than the arbitrary center frequencies of averaged frequency bands), FPL calibration provided a significant improvement in the ability of DPOAEs to detect hearing loss. Another group (Rogers et al. 2010) explored whether FPL calibration affected the ability of DPOAEs to predict hearing thresholds but found no benefit. To date, documented improvements in OAE test-retest variability using FPL calibration have not been easily translated into tangible improvements in clinical practice and diagnostics. These are more difficult to detect and appear to depend strongly on methodological choices.

A series of studies have also tested the effects of FPL-based and other similar calibration methods on the stability of behavioral thresholds (e.g. Souza et al. 2014). Most recently, Lapsley Miller and colleagues (2018) studied the reliability of pure-tone audiometry in young-adult subjects using in-the-ear SPL calibrations, SPL calibrations in a coupler, and the FPL calibration method. They found that whereas intra-subject variability of audiometric thresholds with FPL averaged only 2 dB across frequency, SPL-calibrated thresholds showed significantly increased variability at frequencies between 4 and 8 kHz. Others have reported a similar result, though smaller improvements were noted (Withnell et al. 2014). In the present study, three out of five of the subjects tested two times with Békésy tracking showed reduced (i.e. improved) test-retest variability when thresholds were expressed in FPL. It is not yet clear, however, whether these FPL improvements in the repeatability of audiometric thresholds will translate into earlier or more sensitive detection of hearing loss induced by noxious factors such as noise exposure or ototoxins.

ACKNOWLEDGMENTS

We thank Yeini Guardia and Ping Luo for assistance in tester training and data analysis and Sumit Dhar for generously sharing his data and for helpful discussions. This work was supported by NIH/NIDCD Grants R01 DC003552 (CA) and R01 DC003687 (CAS) and by a USC Dean's Research Scholarship (TM).

Source of Funding

This research was funded by NIH/NIDCD Grants R01 DC003552 (CA), R01 DC003687 (CAS), and by a USC Dean's Research Scholarship (TM).

REFERENCES

Abdala C, & Kalluri R (2017). Towards a joint reflection-distortion otoacoustic emission profile: Results in normal and impaired ears. J Acoust Soc Am, 142, 812. [PubMed: 28863614]

Abdala C, Luo P, Shera CA (2015). Optimizing swept-tone protocols for recording distortion-product otoacoustic emissions in adults and newborns. J Acoust Soc Am, 138, 3785–3799. [PubMed: 26723333]

Abdala C, Winter M, Shera CA (2016). Otoacoustic Emissions in Infants and Children: An Updated Approach In Tharpe AM & Seewald R (Eds.), Comprehensive Handbook of Pediatric Audiology (2nd ed) (pp. 475–504). San Diego, CA: Plural Publishing, Inc.

Abdala C, Ortmann AJ, Shera CA (2018a). Reflection- and Distortion-Source Otoacoustic Emissions: Evidence for Increased Irregularity in the Human Cochlea During Aging. J Assoc Res Otolaryngol. 19, 493–510. [PubMed: 29968098]

Abdala C, Guardia YC, Shera CA (2018b). Swept-tone stimulus-frequency otoacoustic emissions: Normative data and methodological considerations. J Acoust Soc Am, 143, 181. [PubMed: 29390734]

Biswal M and Mishra S (2018). Comparison of time-frequency methods for analyzing stimulus frequency otoacoustic emissions. 143: 626.

Burke SR, Rogers AR, Neely ST, et al. (2010). Influence of calibration method on distortion-product otoacoustic emission measurements: I. test performance. Ear Hear, 31, 533–545. [PubMed: 20458246]

Charaziak KK, & Shera CA (2017). Compensating for ear-canal acoustics when measuring otoacoustic emissions. J Acoust Soc Am, 141, 515–531. [PubMed: 28147590]

Dreisbach L, Zettner E, Chang Liu M, et al. (2018). High-Frequency Distortion-Product Otoacoustic Emission Repeatability in a Patient Population. Ear Hear, 39, 85–100. [PubMed: 28678077]

Dreisbach LE, Long KM, Lees SE (2006). Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults. Ear Hear, 27, 466–479. [PubMed: 16957498]

Farmer-Fedor BL, Rabbitt RD. (2002) Acoustic intensity, impedance and reflection coefficient in the human ear canal. J Acoust Soc Am 112:600–620. [PubMed: 12186041]

Kalluri R, & Shera CA (2013). Measuring stimulus-frequency otoacoustic emissions using swept tones. J Acoust Soc Am, 134, 356–368. [PubMed: 23862813]

Keefe DH (1997). Otoreflectance of the cochlea and middle ear. J Acoust Soc Am, 102, 2849–2859. [PubMed: 9373972]

Konrad-Martin D & Keefe DH (2005).Transient-evoked stimulus-frequency and distortion-product otoacoustic emissios in normal and impaired ears. J Acoust Soc Am, 117, 3799–3815. [PubMed: 16018483]

Kummer P, Janssen T, Arnold W (1998). The level and growth behavior of the 2 f1-f2 distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss. J Acoust Soc Am, 103, 3431–3444. [PubMed: 9637030]

Lapsley Miller JA, Marshall L, Heller LM, et al. (2006). Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss. J Acoust Soc Am, 120, 280–296. [PubMed: 16875225]

Lapsley Miller JA, Reed CM, Robinson SR, et al. (2018). Pure-Tone Audiometry With Forward Pressure Level Calibration Leads to Clinically-Relevant Improvements in Test-Retest Reliability. Ear Hear, 39, 946–957. [PubMed: 29470259]

Lee J, Dhar S, Abel R, et al. (2012) Behavioral hearing thresholds between 0.125 and 20 kHz using depth-compensated ear simulator calibration. Ear Hear, 33, 315–329. [PubMed: 22436407]

Long GR, Talmadge CL, Lee J (2008). Measuring distortion product otoacoustic emissions using continuously sweeping primaries. J Acoust Soc Am, 124(3), 1613–1626. [PubMed: 19045653]

Marshall L, Lapsley Miller JA, Heller LM, et al. (2009). Detecting incipient inner-ear damage from impulse noise with otoacoustic emissions. J Acoust Soc Am, 125, 995–1013. [PubMed: 19206875]

Moleti A, Longo F, Sisto R (2012). Time-frequency domain filtering of evoked otoacoustic emissions. J Acoust Soc Am, 132, 2455–2467. [PubMed: 23039440]

Nørgaard KR, Fernandez-Grande E, Laugesen S (2017) Incorporating evanescent modes and flow losses into reference impedances in acoustic Thévenin calibration. J Acoust Soc Am 142:3013 [PubMed: 29195468]

Nørgaard KR, Neely ST, Rasetshwane DM (2018) Quantifying undesired parallel components in Thévenin-equivalent acoustic source parameters. J Acoust Soc Am 143:1491 [PubMed: 29604709]

Poling GL, Siegel JH, Lee J, et al. (2014). Characteristics of the 2f(1)-f(2) distortion product otoacoustic emission in a normal hearing population. J Acoust Soc Am, 135, 287–299. [PubMed: 24437769]

Reavis KM, McMillan GP, Dille MF, et al. (2015). Meta-Analysis of Distortion Product Otoacoustic Emission Retest Variability for Serial Monitoring of Cochlear Function in Adults. Ear Hearing, 36, e251–260. [PubMed: 25985018]

Reuven ML, Neely ST, Kopun JG, et al. (2013). Effect of calibration method on distortion-product otoacoustic emission measurements at and around 4 kHz. Ear Hearing, 34, 779–788. [PubMed: 24165303]

Richmond SA, Kopun JG, Neely ST, et al. (2011). Distribution of standing-wave errors in real-ear sound-level measurements. J Acoust Soc Am, 129, 3134–3140. [PubMed: 21568416]

Roede J, Harris FP, Probst R, et al. (1993). Repeatability of distortion product otoacoustic emissions in normally hearing humans. Audiology, 32, 273–281. [PubMed: 8216026]

Rogers AR, Burke SR, Kopun JG, et al. (2010). Influence of calibration method on distortion-product otoacoustic emission measurements: II. threshold prediction. Ear Hear, 31, 546–554. [PubMed: 20458245]

Scheperle RA, Neely ST, Kopun JG, et al. (2008). Influence of in situ, sound-level calibration on distortion-product otoacoustic emission variability. J Acoust Soc Am, 124, 288–300. [PubMed: 18646977]

Scheperle RA, Goodman SS Neely ST. (2011). Further assessment of forward pressure level for in situ calibration. J Acoust Soc Am, 130, 3882–3892. [PubMed: 22225044]

Shera CA, & Guinan JJ Jr. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs. J Acoust Soc Am, 105, 782–798. [PubMed: 9972564]

Shera CA, Guinan JJ Jr., Oxenham AJ (2002). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc Natl Acad Sci USA, 99, 3318–3323. [PubMed: 11867706]

Shera CA, & Abdala C (2012). Otoacoustic Emissions - Methods and Applications In Tremblay KL & Burkard RF (Eds.), Translational Perspectives in Auditory Neuroscience: Hearing Across the Lifespan - Assessment and Disorders (pp. 123–159). San Diego, CA: Plural Publishing, Inc.

Shera CA and Bergevin C (2012) Obtaining reliable phase-gradient delays from otoacoustic emission data. J Acoust Soc Am 132:927–943 [PubMed: 22894215]

Siegel JH (1994). Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. J Acoust Soc Am, 95, 2589–2597.

Siegel JH, & Hirohata ET (1994). Sound calibration and distortion product otoacoustic emissions at high frequencies. Hear Res, 80, 146–152. [PubMed: 7896573]

Siegel JH, Nørgaard KR, Neely ST (2018) Evanescent waves in simulated ear canals: Experimental demonstration and method for compensation. J Acoust Soc Am 144:2135 [PubMed: 30404523]

Souza NN, Dhar S, Neely ST, et al. (2014). Comparison of nine methods to estimate ear-canal stimulus levels. J Acoust Soc Am, 136, 1768–1787. [PubMed: 25324079]

Whitehead ML, Stagner BB, Martin GK, et al. (1996). Visualization of the onset of distortion-product otoacoustic emissions, and measurement of their latency. J Acoust Soc Am, 100(3), 1663–1679. [PubMed: 8817893]

Withnell RH, Jeng PS, Parent P, et al. (2014). The clinical utility of expressing hearing thresholds in terms of the forward-going sound pressure wave. International Journal of Audiology, 53, 522–530. [PubMed: 24825368]

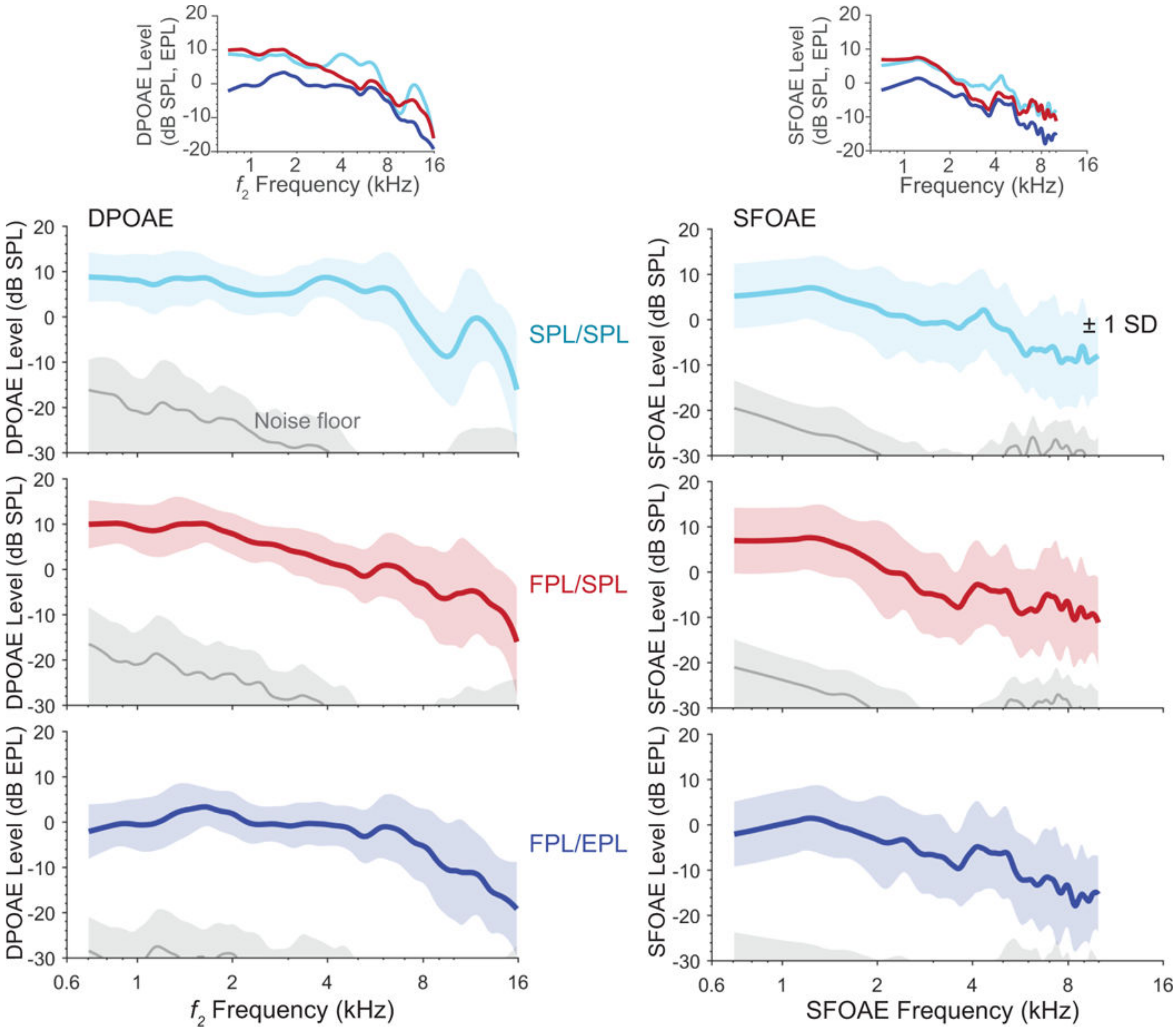

Figure 1.

Mean DPOAE and SFOAE levels (±1 SD – shaded regions) for three stimulus/response calibration methods (SPL/SPL, FPL/SPL, FPL/EPL) measured using moderate-level stimuli. The corresponding mean noise floor is shown in gray. The insets show OAE level superimposed for each condition to illustrate the impact of calibration method on OAE level.

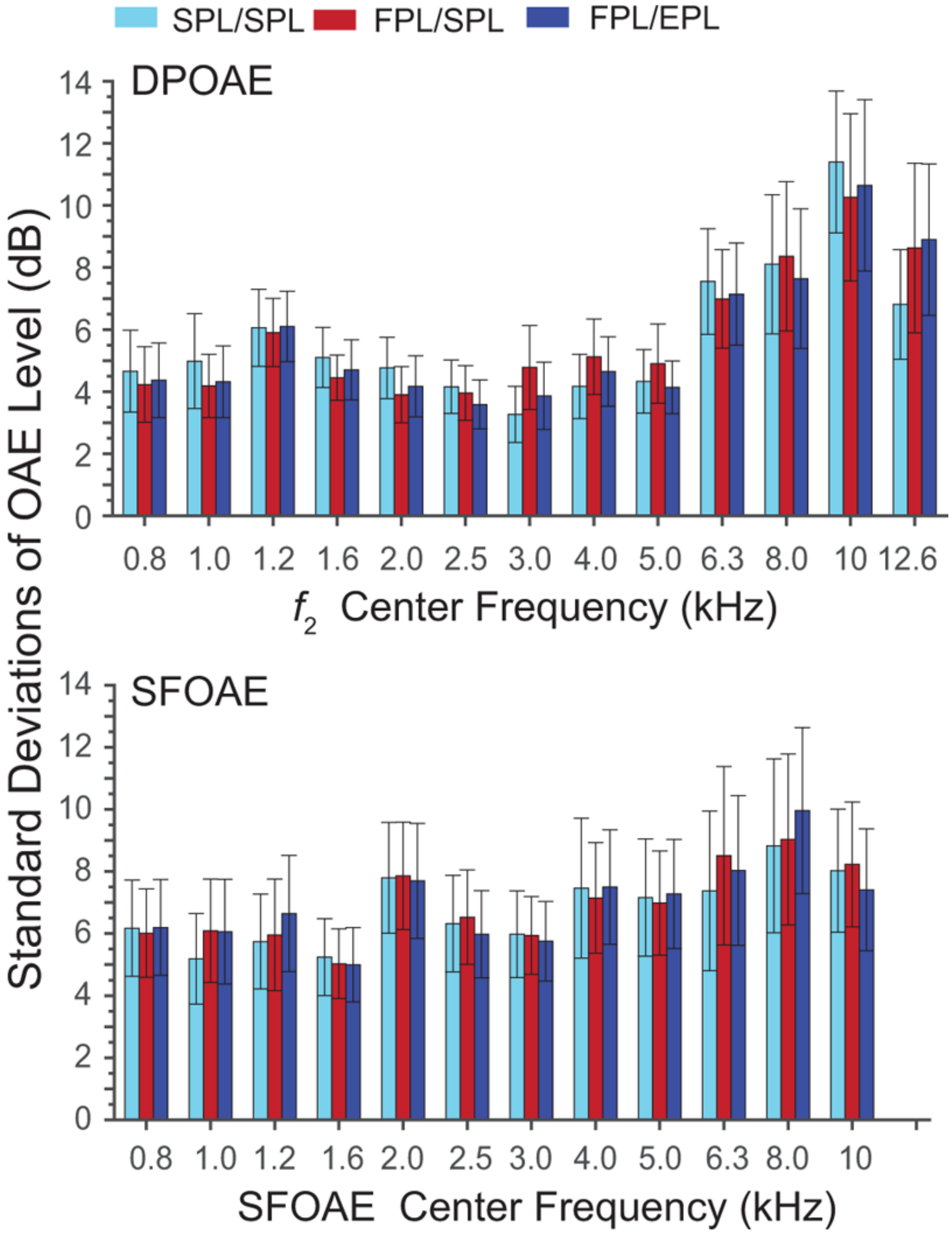

Figure 2.

Standard deviations (SD) of DPOAE and SFOAE level plotted in one-third-octave frequency bands for the three calibration methods (denoted by color). Only data obtained with moderate-stimulus levels are shown. The 95% CIs for SD were generated via resampling.

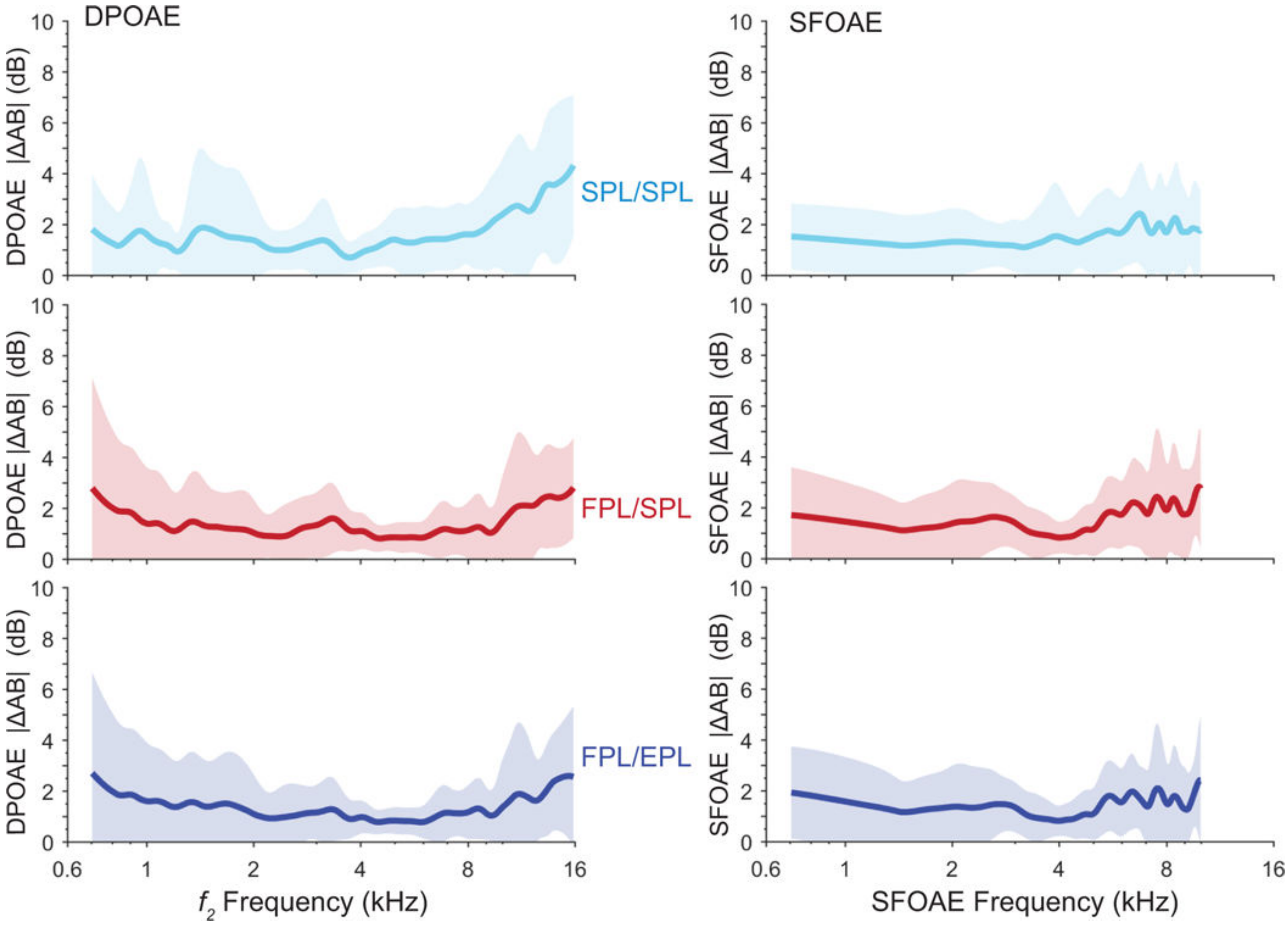

Figure 3.

The mean absolute level differences, |ΔAB{OAE}|, between OAE levels recorded during blocks A and B within the same test session, with a probe refit between the blocks. The shaded region shows ±1 SD. Both |ΔAB{DPOAE}| and |ΔAB{SFOAE}| were measured with moderate-level stimuli.

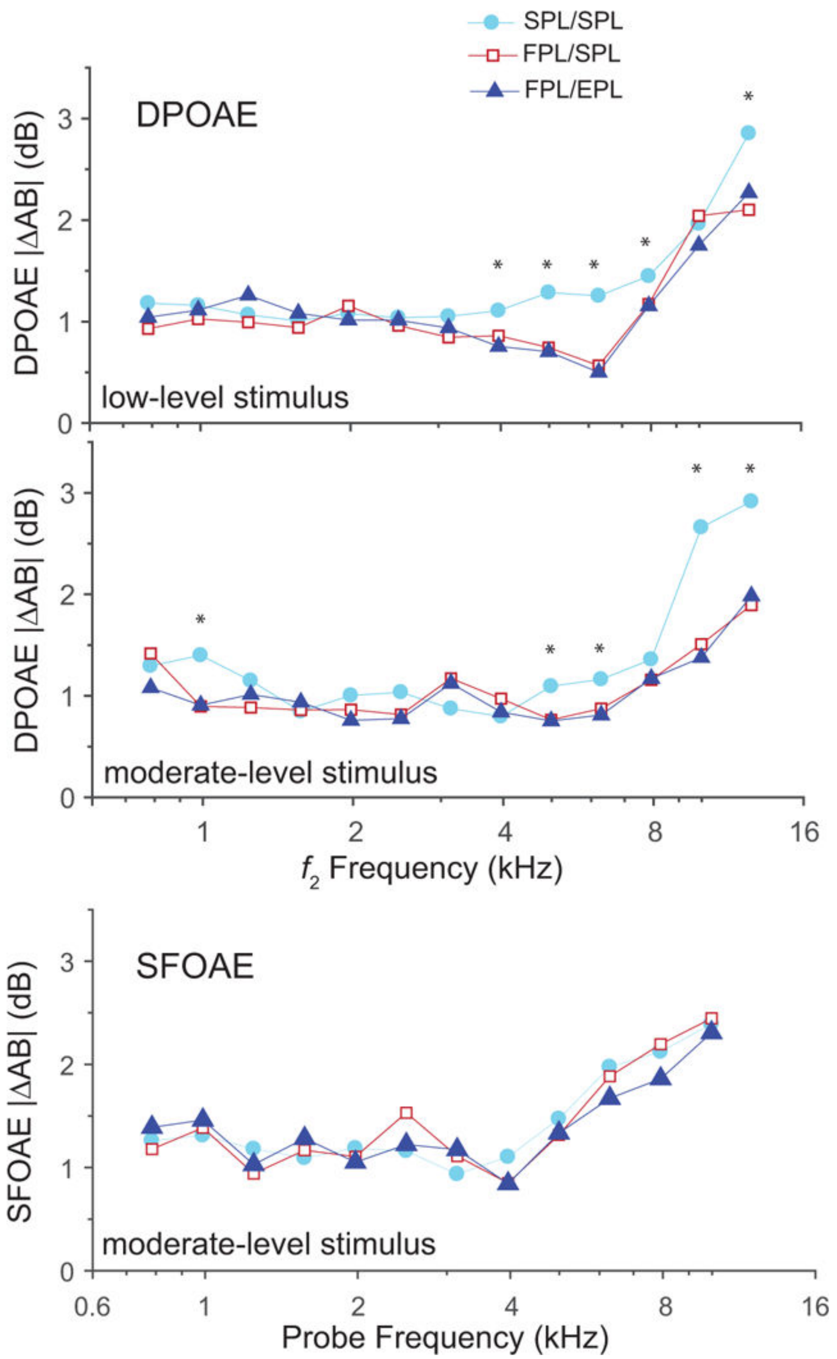

Figure 4.

Mean absolute differences |ΔAB{OAE}| between DPOAE levels (top and middle panels) and SFOAE levels (bottom panel) recorded during blocks A and B, averaged over one-third-octave frequency bands. The top and middle panels show DPOAE results obtained in the low- and the moderate-level stimulus conditions, respectively. Both the FPL/SPL and FPL/EPL calibration methods produced a significant improvement (i.e., smaller |ΔAB{DPOAE}|) over conventional SPL/SPL calibration for 5 frequencies in each panel (see asterisks), mostly for frequencies ≥ 4 kHz, where standing-wave interference is greatest. The bottom panel shows SFOAE results for the moderate-level stimulus. Although the results show no significant effect of calibration method on the SFOAE, the trends at frequencies ≥ 4 kHz are noteworthy and consistent with theory (see text).

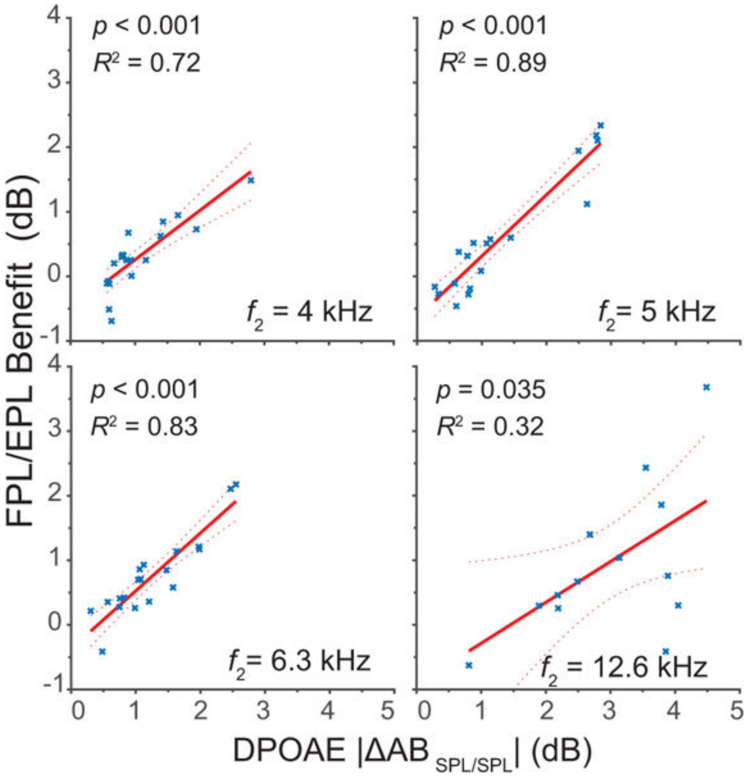

Figure 5.

Correlations between the benefit achieved by FPL/EPL calibration, defined as $|\Delta AB\{DPOAE_{SPL/SPL}\}| – |\Delta AB\{DPOAE_{FPL/EPL}\}|$, and the overall repeatability of the DPOAE data gauged by $|\Delta AB\{DPOAE_{SPL/SPL\}}|$ for frequencies ≥ 4 kHz. Four significant correlations (out of six) are shown here for the low-level condition. DPOAEs with greater intra-subject variability are associated with stronger FPL/EPL benefits.

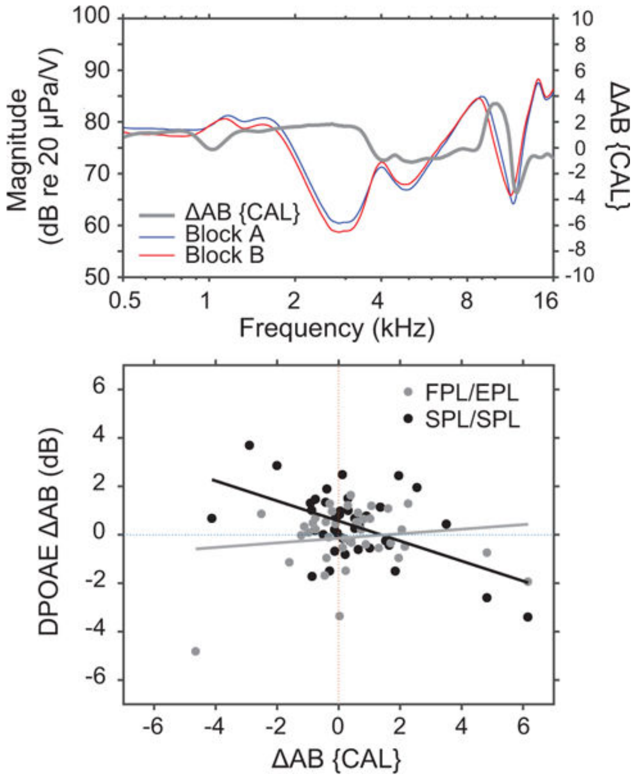

Figure 6.

The upper panel shows examples of chirp calibration curves obtained at the start of blocks A and B (red and blue) and the difference between the two (ΔAB{CAL} (gray line re: right ordinate) for one subject. In the lower panel, correlations are shown between DPOAE test-retest replicability (ΔAB{DPOAE}) in SPL/SPL (black) or FPL/EPL conditions (gray) and stability of the probe fit gauged by ΔAB{CAL}. Both ΔAB{OAE} and ΔAB{CAL} were calculated in one-third-octave frequency bands centered at the half-wave resonant frequency and the quarter-wave null frequency.

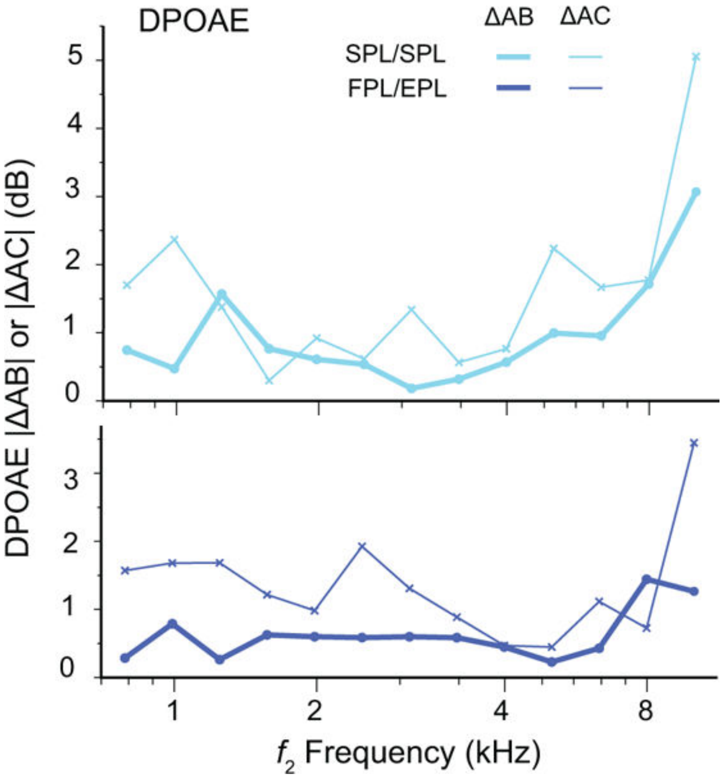

Figure 7.

Mean absolute DPOAE level differences for blocks A and B (|ΔAB{DPOAE}|, thick lines) and blocks A and C (|ΔAC{DPOAE}|, thin lines) averaged in one-third-octave bands at moderate stimulus levels for one representative subject. Data obtained using SPL/SPL (cyan) and FPL/EPL (blue) calibration methods are shown.

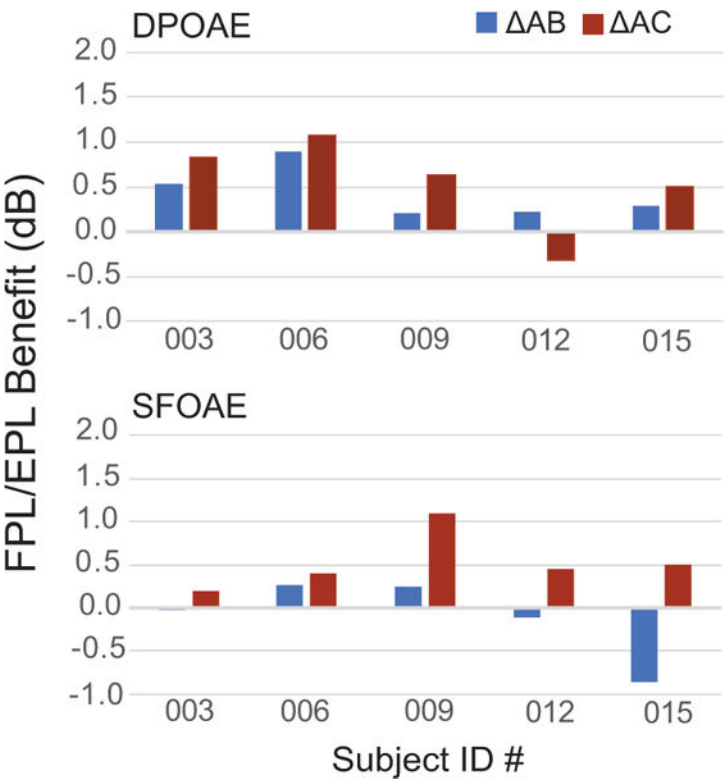

Figure 8.

Benefits achieved by the FPL/EPL calibration method to measures of long-term DPOAE (top panel) and SFOAE (bottom) test-retest repeatability in 5 subjects. Benefit is defined as: $|\Delta AC\{OAE_{SPL/SPL}\}| - |\Delta AC\{OAE_{FPL/EPL}\}|$. Data for the moderate stimulus level condition are shown.

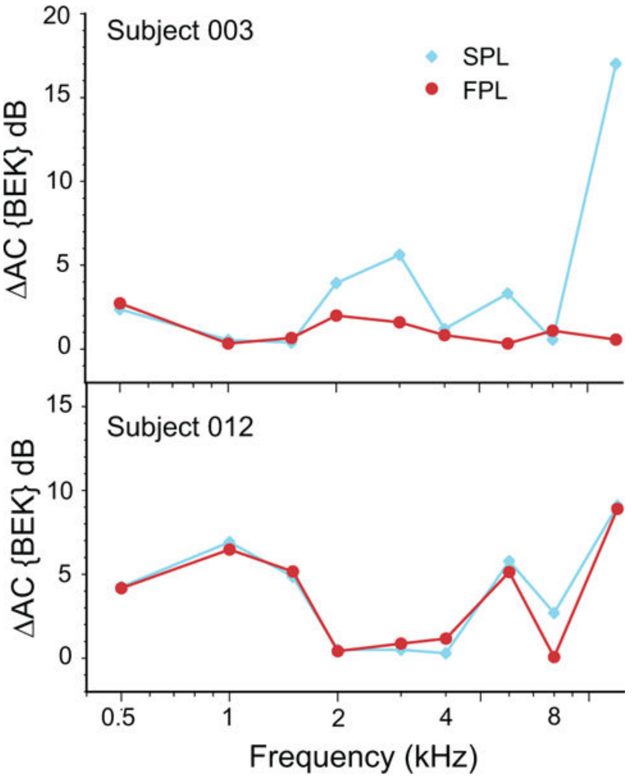

Figure 9.

Mean absolute differences between Békésy-tracked thresholds obtained during blocks A and C for: SPL calibration (|ΔAC{$\mathrm{BEK}_{\mathrm{SPL}}$}|) and FPL calibration (|ΔAC{$\mathrm{BEK}_{\mathrm{FPL}}$}|). Data from two subjects are presented, one showing significant improvement due to FPL calibration (003) and the other with little change (012).

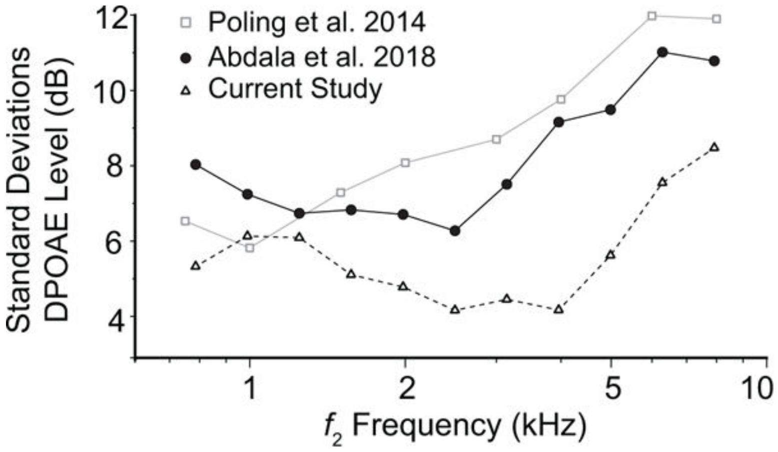

Figure 10.

The standard deviations of mean DPOAE levels obtained from three independent studies. OAEs for all studies were measured using conventional SPL/SPL stimulus calibration methods.