Influence of Calibration Method on Distortion Product Otoacoustic Emission Measurements. II. Threshold Prediction

Abigail R. Rogers,[1,2] Sienna R. Burke,[2,3] Judy G. Kopun,[2] Hongyang Tan,[2] Stephen T. Neely,[2] and Michael P. Gorga[2]

Objectives: Distortion product otoacoustic emission (DPOAE) stimulus calibrations are typically performed in sound pressure level (SPL) before DPOAE measurements. These calibrations may yield unpredictable DPOAE response levels, presumably because of the presence of standing waves in the ear canal. Forward pressure level (FPL) has been proposed as an alternative method for stimulus calibration because it avoids complications due to standing waves. DPOAE thresholds after four FPL calibrations and one SPL calibration were compared with behavioral thresholds to determine which calibration results in data that yield the highest correlations between the two threshold estimates.

Design: Fifty-two subjects with normal hearing and 103 subjects with hearing loss participated in this study, with ages ranging from 11 to 75 yr. These were the same individuals whose data were used to address the influence of calibration method on test performance in an accompanying article. DPOAE input/output (I/O) functions were obtained at f_2 frequencies of 2, 3, 4, 6, and 8 kHz with the primary frequency ratio fixed at $f_2/f_1 \approx 1.22$. L_1 was set according to the equation $L_1 = 0.4L_2 + 39$ with L_2 levels ranging from -20 to 70 dB SPL and FPL in 5-dB steps. I/O functions were obtained at each frequency for each of the five stimulus calibrations: SPL, daily FPL at room temperature, daily FPL at body temperature, reference FPL at room temperature, and reference FPL at body temperature. DPOAE thresholds were estimated using two methods. In the first method, DPOAE threshold was taken as the lowest L_2 for which DPOAE level is 3 dB or greater than the noise floor (signal to noise ratio $\geq$3 dB). In a second method, a linear regression method first described by Boege & Janssen (2002) and later adapted by Gorga et al. (2003), all DPOAE levels in each I/O function are converted to linear pressure and extrapolated to 0 μPa, at which the L_2 is taken as threshold. Correlations of DPOAE thresholds with behavioral thresholds were obtained for each frequency, calibration method, and threshold-prediction method.

Results: Correlations were greatest for frequencies of 3 to 6 kHz and lowest for 8 kHz, consistent with previous frequency effects. Calibration method made little difference in correlations between DPOAE and behavioral thresholds at any frequency. A small difference was noted in correlations for the two threshold prediction methods, with the linear regression method yielding slightly higher correlations at all frequencies.

Conclusions: Little difference in threshold correlations was observed among the five calibration methods used to calibrate the stimuli before DPOAE measurements. These results were not anticipated, given the known effects of standing waves on ear-canal estimates of SPL at the plane of the probe. In addition, there was no effect of temperature (body versus room) or timing (daily versus reference) for FPL calibrations. It may be important to note that differences between SPL and FPL calibrations should not be seen if a standing wave does not occur at the plane of the probe at or near the frequency being tested. The frequencies (2 to 8 kHz) were chosen because it was expected that effects from standing waves would occur between these frequencies because of the typical lengths of ear canals for the age group tested. Because measurements were taken at only five discrete frequencies in the interval, it is possible that standing waves were present but did not affect the specific test frequencies. In total, these results suggest that SPL calibrations may be adequate when attempting to predict pure-tone thresholds from DPOAEs, despite the fact that they are known to be susceptible to errors associated with standing waves.

(Ear & Hearing 2010;XX;1–●)

[1]Department of Speech and Hearing Sciences, Indiana University, Bloomington, Indiana; [2]Boys Town National Research Hospital, Omaha, Nebraska; and [3]Department of Speech and Hearing Sciences, University of Maryland, College Park, Maryland.

INTRODUCTION

The measurement of distortion product otoacoustic emissions (DPOAEs) relies on specific stimulus conditions, one of which is the absolute and relative levels of the two primary frequencies that elicit the response. One factor that can influence stimulus level is the variance in individual ear-canal acoustics. To ensure that intended stimulus levels are presented to the ear, in situ calibrations typically are performed in sound pressure level (SPL) before measurements. SPL is estimated at the plane of the probe, and it is assumed that this value represents the level at the plane of the tympanic membrane. However, SPL calibrations are susceptible to influence from standing waves at specific frequencies depending on the dimensions of the ear canal (Siegel 1994, 2007; Siegel & Hirohata 1994). This situation can result in over- or underestimation of stimulus level at the eardrum, and thus a disparity between actual and intended stimulus levels. That is, the level measured at the plane of the probe would not accurately reflect the input level to the middle ear. Recently, Scheperle et al. (2008) showed that stimuli calibrated in forward pressure level (FPL) provide more consistent DPOAE measurements than the stimuli calibrated in SPL when probe-insertion depth was varied. Their results were obtained on a group of normal-hearing subjects. The observation that the calibration method can affect DPOAE level because of errors in stimulus-level calibrations suggests that these errors might also influence the clinical accuracy of DPOAE measurements. The purpose of this study is to extend the work of Scheperle et al. by comparing SPL and FPL calibrations using correlations between predicted DPOAE thresholds and pure-tone behavioral thresholds for a group of normal-hearing and hearing-impaired subjects. In a companion article, Burke et al. (2010) tested the extent to which calibration method affects test performance, defined as the ability of DPOAEs to accurately classify ears as either normal hearing or hearing impaired.

Correlations between DPOAEs and audiometric thresholds have been reported previously (Martin et al. 1990; Gorga et al.

0196/0202/10/XX0X-0001/0 • Ear & Hearing • Copyright © 2010 by Lippincott Williams & Wilkins • Printed in the U.S.A.

1993, 1997). These correlations were limited to hearing losses no greater than 50 to 60 dB HL (Gorga et al. 1997), which was expected, given the relationship between outer hair-cell damage (the presumed generators of OAEs) and sensitivity loss. Similar to the estimates of test performance (Gorga et al. 1993, 1997, 2000), correlations between DPOAE data and audiometric thresholds are highest for mid-to-high frequencies (Gorga et al. 2003). In some cases, DPOAE levels for fixed-level stimuli were correlated with behavioral thresholds, whereas in other cases, DPOAE threshold (defined as the stimulus level producing a DPOAE some criterion number of decibel above the noise) was correlated with audiometric threshold (Martin et al. 1990; Gorga et al. 1997). As an alternative to these measurements, Boege and Janssen (2002) developed a method in which DPOAE levels obtained from an input/output (I/O) function were converted to pressure and fit with a linear equation. The L_2 (the level of the higher frequency in the primary frequency pair) at which DPOAE pressure equals 0 μPa was defined as threshold in this linear regression method. Boege and Janssen (2002) demonstrated that, after the application of certain inclusion criteria, their method provided DPOAE threshold estimates that correlated with behavioral thresholds. Using fits to the DPOAE I/O function to provide an estimate of DPOAE threshold has since been replicated and extended by Gorga et al. (2003) and Johnson et al. (2007). Although both threshold estimation methods revealed correlations between DPOAE thresholds and behavioral thresholds, variability and prediction errors were evident. Notably, both approaches used in situ SPL calibrations at the plane of the probe to set the stimulus level during DPOAE measurements.

The effect of standing waves on the calibration of stimulus level could potentially affect behavioral threshold estimations based on DPOAEs. The presence of incident and reflected waves in the ear canal creates pressure nulls at specific locations determined by an interaction between stimulus frequency and the dimensions of the canal. The level of sound in the ear canal will be underestimated when incident and reflected waves are out of phase at the point of measurement (the plane of the probe), resulting in a higher stimulus level than intended. This overcorrection could affect threshold prediction in which it could be (1) underestimated when incident and reflected waves are out of phase or (2) to a lesser extent, overestimated when incident and reflected waves are in phase at the probe. The influence of standing waves on DPOAE threshold estimates based on fits to the I/O function may be less, although this is unknown at the present time. Typically, standing waves are not a problem for frequencies below ~2 kHz, because these lower frequencies have wavelengths that are large relative to the length of the typical ear canal. However, information about auditory function above 2 kHz may be affected by standing waves, given the shorter wavelengths. This may be of a particular clinical concern, especially for hearing loss caused by noise exposure and ototoxicity at which higher frequencies are first affected by the cochlear insult.

As alternatives to SPL, calibrations obtained in both sound intensity level (SIL) and FPL avoid the problem of standing waves (Neely & Gorga 1998; Scheperle et al. 2008). SIL calibrations do not contain the reactive components of impedance, which are the cause of standing waves, and FPL calibrations quantify the forward pressure component of a sound wave only. Scheperle et al. noted effects of stimulus frequency on DPOAE level when comparing SPL, SIL, and FPL calibrations, with SPL demonstrating the greatest variability. However, it has been observed that stimuli calibrated in SIL may become uncomfortably loud, especially at high frequencies (Scheperle et al. 2008). It is mainly for this reason that FPL calibration was used in this study. An additional advantage of FPL over SIL is that it is more easily comparable with SPL, as it shares the same reference. The measurement of an incident pressure wave independent of its reflected component requires information about the source and load characteristics of the transmission line. This requires measurements obtained in known acoustic loads to calculate the Thévenin equivalent characteristics of the source. The source characteristics are then used to determine the ear-canal load impedance during in situ calibration. However, estimates of source characteristics may be sensitive to temperature and (somewhat surprisingly) variable from day to day (Scheperle et al. 2008). It remains to be determined whether these measurements need to be performed daily or if an average or some other ideal calibration from a set of measurements is preferable to obtain the most reliable estimates of level, and, therefore, the most accurate estimates of cochlear function from DPOAE measurements.

This study examines the effects of calibration method on estimates of DPOAE thresholds, defined either as the lowest L_2 at which the signal to noise ratio (SNR) is $\geq$3 dB or by the linear-regression (LR) method first described by Boege and Janssen (2002). The outcome measure in this study was the accuracy with which DPOAE thresholds, estimated from both techniques, predicted behavioral thresholds in subjects with normal hearing and hearing loss. Four different FPL calibration conditions were used to examine the effects of temperature (body versus room) and the timing of the calibration (daily versus a reference taken from repeated measurements before all data collection). For comparison purposes, data were collected using an SPL calibration as well. DPOAE I/O functions were measured at 2, 3, 4, 6, and 8 kHz using each of the five calibration procedures. Effects of stimulus frequency, calibration method, and method of threshold estimation were examined to determine which calibration method and which DPOAE threshold estimate results in the highest correlation with behavioral threshold. If one calibration method is shown to be superior, it would be reasonable to implement it in clinical applications of DPOAE measurements. If there is no difference in correlations then the simpler SPL calibration that is in current use may be preferred.

METHODS

Subjects

Fifty-two subjects with normal hearing and 103 subjects with hearing loss participated in this study, with ages ranging from 11 to 75 yrs. These subjects were the same individuals whose data were used to address the influence of calibration method on test performance in an accompanying article (Burke et al. 2010). Table 1 shows the number of ears at each frequency for each behavioral threshold, expressed in dB HL. Note that the largest cell is the one that includes thresholds exceeding 65 dB HL. Because we were interested in evaluating the relation between behavioral and DPOAE thresholds for a wide range of hearing losses, no restrictions were made

TABLE 1. Number of ears at each test frequency per threshold

Threshold (dB HL)	Frequency (Hz)				
	2000	3000	4000	6000	8000
−10	1	1	0	0	0
−5	3	4	2	6	6
0	10	14	17	9	14
5	15	14	17	19	20
10	22	15	12	12	10
15	17	10	6	9	13
20	11	11	9	5	5
25	8	11	6	5	4
30	8	8	7	6	4
35	7	5	9	7	4
40	5	5	6	7	4
45	7	11	5	7	4
50	7	5	10	5	5
55	7	10	10	9	6
60	6	5	8	10	8
65	5	6	10	7	9
>65	16	20	21	32	39

regarding degree and configuration of hearing loss during subject recruitment. However, extra effort was made to include subjects whose behavioral thresholds ranged from 0 to 65 dB HL. Our interest in focusing on this range relates to the need to have measurable DPOAEs from which predictions of behavioral thresholds can be made. Individuals with hearing losses outside this range were included, but they represented a minority of the subjects. Ideally, analyses would be simplified if the number of observations were equal in all cells in Table 1. Given the source of subjects for this study, it was not possible to achieve this goal. However, every cell between 0 and 60 dB HL had at least four subjects at all frequencies, which should be sufficient for the correlation analyses that will be described subsequently. Notably, only 5 of 70 cells between 0 and 65 dB HL had only four subjects, and all of these occurred at 8 kHz. All other cells include data from a larger number of subjects.

For the purpose of global characterizations, an individual was considered to have normal hearing if thresholds at all five test frequencies (in addition to other typical audiometric frequencies) were ≤20 dB HL. Hearing impairment for any subject was defined as behavioral thresholds exceeding 20 dB HL at any octave or interoctave frequency from 2 to 8 kHz. For the purpose of data analysis, thresholds were considered at each individual frequency for each subject. Hearing losses were generally determined to be of cochlear origin based on clinical history and other common clinical tests. However, the specific etiology of hearing loss was not considered during subject recruitment.

Equipment

Data collection, source calibrations, calculations of Thévenin equivalents, and FPL conversions were performed using custom-designed software (EMAV version 3.1; Neely & Liu 1994). DPOAE stimuli were produced and responses were recorded using a 24-bit soundcard (CardDeluxe; Digital Audio Labs, Chanhassen, MN) housed in a PC. The two primary tones (f_1 and f_2) were produced by separate channels of the soundcard and sent to two loudspeakers in a probe-microphone system (ER-10C, Etymotic Research, Elk Grove Village, IL), which was coupled to the ear using a foam tip. Pure-tone air and bone conduction behavioral thresholds were measured using a Grason-Stadler GSI-61 audiometer with EAR-Tone ER-3A earphones.

Stimuli

DPOAE I/O functions were obtained at 2, 3, 4, 6, and 8 kHz. These frequencies were chosen because they are clinical test frequencies whose levels may be affected by standing waves in the ear canal. The $2f_1 - f_2$ DPOAE was measured with the primary–frequency ratio (f_2/f_1) ≈1.22. These stimulus-frequency parameters are in common clinical use and similar to those used in previous efforts to predict behavioral thresholds from DPOAEs (Martin et al. 1990; Boege & Janssen 2002; Gorga et al. 2003). The L_2 starting level was set to 70 dB SPL or FPL and decreased in 5-dB steps. L_1 was set according to the equation $L_1 = 0.4L_2 + 39$, a commonly used primary-level paradigm described by Kummer et al. (1998, 2000). This relationship between primary levels was also used in the two studies that correlated behavioral thresholds with DPOAE thresholds that were predicted from fits with the DPOAE I/O function (Boege & Janssen 2002; Gorga et al. 2003). There is evidence that other relationships between L_1 and L_2 might result in larger DPOAEs, at least among subjects with normal hearing (Neely et al. 2005; Johnson et al. 2007). However, this study was designed to determine the influence of calibration procedure on threshold predictions from DPOAE data, not the algorithm used to select stimulus levels. Given this goal, it seemed appropriate to select stimulus levels the same way as in previous studies in which threshold predictions were made.

Calibration Methods

Calibration procedures for this study have been described in detail previously (Burke et al. 2010) and will only be briefly summarized here. FPL calibration procedures were performed using a set of five brass cavities with known load impedances to determine the Thévenin equivalent characteristics of the stimulus source. The foam ear tip of the ER-10C probe-microphone system was coupled to each cavity, and a wide-band chirp stimulus (sampling rate of 32 kHz) was presented. The pressure response of each loudspeaker was measured with the probe microphone, and the software used this measurement along with the known load impedance to estimate the source impedance and pressure. A comparison between the estimated source pressure and measured pressure was expressed as an error value. An in situ calibration was performed before the measurement of each I/O function to measure the pressure response in the ear, and the source impedance and pressure were used along with load pressure to determine the ear-canal load impedance. A transform was then derived to convert SPL to FPL for each of the four calibration conditions.

FPL calibration conditions included (1) a daily calibration at room temperature, (2) a daily calibration at body temperature, (3) a reference calibration at room temperature, and (4) a reference calibration at body temperature. For comparison purposes, SPL calibration, which is the approach to calibration that is currently in widespread use, was obtained in situ before DPOAE measurement. Body temperature calibrations were obtained by heating the cavity set with a heating pad and measuring the temperature inside each cavity with a digital thermometer. Temperatures varied from 96 to 102°F. Calibrations were performed at both room and body temperatures in view of the results described by Scheperle et al. (2008), which showed effects of temperature on FPL calibrations. In this study, another goal (in addition to the one in which the effects of calibration method were assessed) was to determine whether these temperature effects were significant in the relationship between behavioral and DPOAE thresholds.

Reference calibrations at both room and body temperatures were obtained before the initiation of any data collection. Twenty-five calibrations were obtained for both temperatures, and the calibrations with the smallest errors associated with source estimates were taken as the reference calibrations for each temperature. For each day of data collection, a daily calibration was also obtained for both temperatures. As in the study by Scheperle et al. (2008), calibrations were variable in terms of the error values obtained. It seems that the variability in error values may be due to differences in the insertion or compression of the foam tip in the coupler (Burke et al. 2010).

Procedures

A routine hearing evaluation was completed to provide pure-tone thresholds and to determine whether subjects met inclusion criteria. Just before DPOAE measurement, an otoscopic examination was performed, and a 226-Hz tympanogram was obtained to evaluate middle-ear status. Pure-tone thresholds were measured at octave frequencies of 0.25 to 8 kHz and interoctave frequencies of 3 and 6 kHz using clinical procedures (ASHA 2005). DPOAE I/O functions (DPOAE level as a function of L_2) were measured at f_2 frequencies of 2, 3, 4, 6, and 8 kHz for each of the four FPL calibration conditions as well as for the SPL calibration, for a total of 25 I/O functions per subject. The orders of the conditions were counterbalanced across subjects. To increase data-collection efficiency, I/O function measurements started with an $L_2 = 70$ dB SPL or FPL, decreased in level in 5-dB steps, and were terminated manually once the DPOAE level was <3 dB greater than the noise floor (i.e., an SNR <3 dB) for each individual subject.

Measurements for each test condition were terminated automatically when one of the following measurement-based stopping rules was met: the noise floor was ≤−25 dB SPL, 32 sec of artifact-free averaging time had elapsed, or the SNR was >60 dB. These rules were selected so that measurement would ideally end when the noise-floor criterion was met and would never stop on the SNR criterion. To obtain reliable estimates of DPOAE thresholds, it was important to ensure that the noise floor was reduced as much as possible so that DPOAE responses were measured at low levels and over a wide dynamic range. The noise-floor criterion was selected in relation to a conservative estimate of the level at which system distortion occurs. If averaging was continued with a noise floor at or below −25 dB SPL (FPL), it would be unknown whether responses measured were biological in nature. For several subjects, however, the noise-floor criterion could not be met before 32 sec of artifact-free averaging had elapsed. For these subjects, the DPOAE and noise levels were taken as an average across the 32 sec, which may have resulted in higher threshold estimates. However, care was taken to ensure that noise levels were kept relatively low (i.e., data collection was paused or restarted) even if they did not meet the −25 dB SPL criterion.

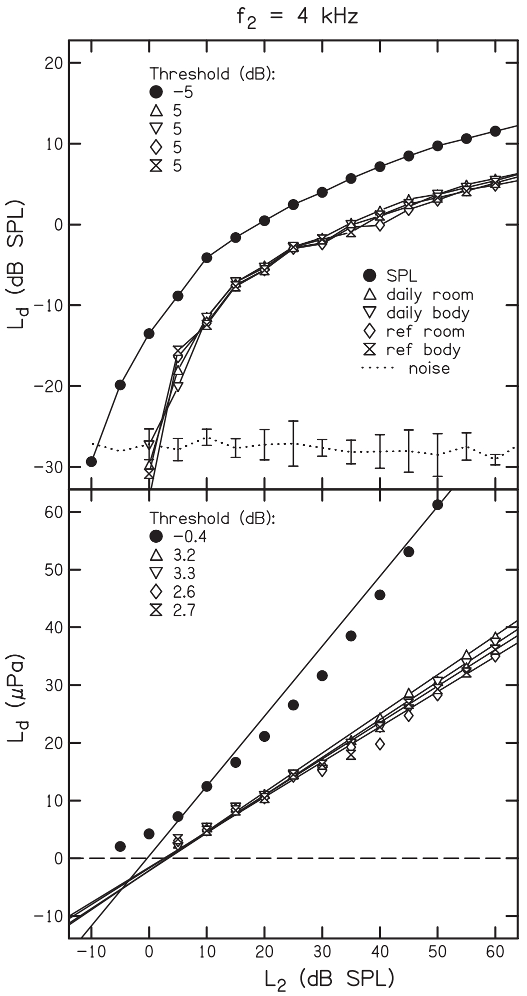

Fig. 1. An example of both distortion product otoacoustic emission (DPOAE) threshold-estimation methods for all five calibration methods based on data from one normal-hearing subject with a behavioral threshold of 5 dB HL at 4 kHz. In the top panel, DPOAE input/output (I/O) functions are shown with DPOAE level (dB SPL) as a function of L_2 (dB SPL and FPL). Thresholds were defined as the lowest L_2 for which the SNR ≥3 dB. For the linear-regression method, demonstrated in the bottom panel, DPOAE I/O functions are shown after transformation of the data into pressure (μPa). Linear functions were fit to the data and solved for the values of L_2 resulting in pressures of 0 μPa, which were defined as threshold.

A longer averaging time was not used because of the need to collect data on a large sample of normal-hearing and hearing-impaired subjects for this study. The calculation of noise levels in the EMAV software was the same for both parts I and II and is described in detail by Burke et al. (2010). Briefly, data were stored in 11 frequency bins, one at $2f_1 - f_2$ and five bins on each side of $2f_1 - f_2$, in 2-sec alternating subaverages. The contents of the subaverages were summed to estimate signal level and the level in the $2f_1 - f_2$ frequency bin and subtracted to estimate the noise level in all 11 bins.

Threshold Prediction

Two methods were used to estimate DPOAE thresholds. Figure 1 illustrates these methods and the estimated DPOAE thresholds for each calibration method at 4 kHz for a single normal-hearing subject, whose audiometric threshold was 5 dB

HL at this frequency. In the simplest method, which is shown in the upper panel, threshold was estimated at the lowest L_2 for which DPOAE level was ≥3 dB above the noise floor (i.e., SNR ≥3 dB). This threshold definition depends on the level of the noise floor, which is why the primary measurement stopping rule was a noise level of ≤−25 dB SPL. In those conditions in which measurement did not stop based on the noise level, it is likely that threshold was overestimated. However, as stated earlier, care was taken to ensure that noise levels were kept relatively low. Thresholds estimated in this fashion are provided as insets in the top panel of Figure 1 for each of five calibration methods. Note that the DPOAE threshold for SPL in this case is 10 dB lower than for any of the FPL calibrations. A possible explanation for this finding would be that, during SPL calibration, a standing wave occurred resulting in destructive interaction between incident and reflected waves. As a consequence, the level at the plane of the probe was less than the level at the eardrum, producing greater stimulus levels than specified, which resulted in larger DPOAEs than expected, leading to an underestimation of threshold.

The second method, shown in the lower panel of Figure 1, followed procedures that were described by Boege and Janssen (2002) and later adapted by Gorga et al. (2003). For each DPOAE I/O function, DPOAE levels were converted to pressure and fit with a linear equation. Linear-regression analysis was performed to extrapolate DPOAE amplitude to 0 μPa. The L_2 at which this DPOAE pressure occurred was taken as DPOAE threshold, which subsequently was compared with the behavioral threshold. To fit the I/O function with a linear equation several different criteria had to be met. At least three data points on the I/O function had to have SNRs of ≥6 dB. In addition, there were inclusion criteria associated with the linear regressions which are as follows: (1) the slope of the individual linear regressions had to be ≥0.2 μPa/dB, (2) the variance accounted for (r^2) had to be ≥0.8, and (3) the standard error had to be ≤10 dB. These criteria were similar to those specified by Boege and Janssen (2002), who demonstrated that their use tended to reduce the difference between DPOAE and behavioral thresholds. Estimated DPOAE thresholds from linear regressions were included in comparison with behavioral thresholds only if they met the ≥6 dB SNR inclusion criteria and the three inclusion criteria associated with the linear-regressions. Referring to Figure 1, although the differences between SPL and FPL calibrations are less using the linear-regression method of threshold estimation as opposed to the simpler SNR-based method, it is still the case that the DPOAE threshold based on SPL calibration was lower than any of the thresholds based on FPL calibration. The standing-wave interpretation that was used to explain the SNR-based threshold discrepancies could explain these findings as well.

RESULTS

Individual Examples of DPOAE I/O Functions

For examples of individual I/O functions, see Figure 3 in the article by Burke et al. (2010). These I/O functions are representative of the functions measured in other subjects and are consistent with expectations for both normal-hearing and hearing-impaired subjects.

Effects of Calibration Method

Figure 2 displays behavioral thresholds (dB HL) as a function of DPOAE thresholds (dB SPL and FPL) for each calibration method collapsed across frequency. In the left column, the SNR-based method is used to estimate DPOAE thresholds. In the right column, the linear-regression method is used to estimate DPOAE thresholds. The solid line in each panel represents the best-fit line to the data. Also shown as insets in the upper left corner of each panel are correlation coefficients, the number of threshold comparisons, and standard errors.

In an effort to improve the accuracy with which the data from DPOAE I/O functions predicted behavioral thresholds, all behavioral thresholds <0 dB HL were set to 0 dB HL and all estimated DPOAE thresholds <0 dB SPL (FPL) were set to 0 dB SPL (FPL). This truncation was based on the assumptions that any DPOAE threshold below 0 dB would be particularly affected by the noise floor and common clinical procedures may not warrant presentation levels <0 dB HL for the purpose of behavioral-threshold measurements, because maximum permissible noise levels in audiometric test rooms are based on ambient noise masking at 0 dB HL (Frank et al. 1993). Another truncation occurred at the upper limits of behavioral and estimated DPOAE thresholds is that any threshold >60 dB SPL (FPL) was set to 60 dB SPL (FPL), which was based on the assumption that a complete loss of outer hair cells (the presumed generators of OAEs) would produce a hearing loss no greater than 60 dB in the absence of inner hair-cell damage. The differences between correlation coefficients for both threshold-prediction methods are <0.05 for each of the five calibration methods when data are collapsed across frequency. The largest correlation coefficient was for the daily body temperature calibrations (0.82 and 0.84 for SNR-based and linear-regression methods, respectively) and the smallest correlation was observed for the reference room temperature calibrations (0.78 and 0.80 for SNR-based and linear-regression methods, respectively). As before, these differences were minimal. In an upcoming section, we introduce Figure 5, which shows these data separately for each frequency.

Effects of Threshold Prediction Method

Referring again to Figure 2, the differences between correlation coefficients across calibration methods are <0.03 between the two threshold-prediction methods, although the linear-regression method consistently yielded slightly higher correlations. This may have been partially because of the inclusion criteria for this method. Specifically, by eliminating cases in which <3 points on the I/O function had an SNR ≥6 dB, we increase the chance that thresholds will not be predicted for subjects whose hearing thresholds exceed 60 dB HL.

Figure 3 displays the cumulative proportion of subjects as a function of behavioral threshold for each calibration method that failed to meet the inclusion criteria associated with the linear-regression method of estimating DPOAE threshold. Failed SNR refers to the proportion of subjects who failed to meet the criterion that the I/O function must have three points at which SNR is ≥6 dB for the linear regression to be applied. Failed linear regression refers to the proportion of subjects failing to meet the additional criteria associated with regression analysis: (1) the slope of the individual linear regressions had to be ≥0.2 μPa/dB, (2) the variance accounted for (r^2) had to

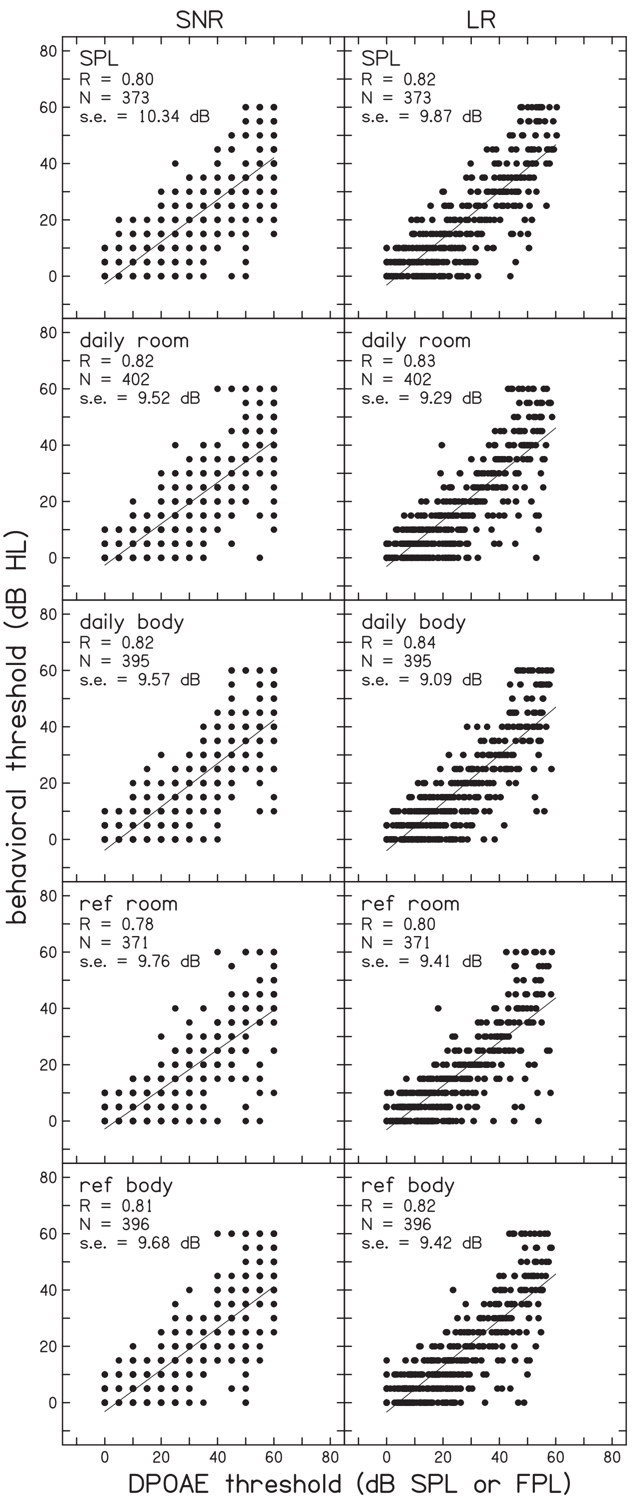

Fig. 2. Behavioral thresholds (dB HL) as a function of distortion product otoacoustic emission (DPOAE) thresholds (dB SPL and FPL) for each calibration method collapsed across frequency. In the left column, the SNR-based method is used to estimate DPOAE thresholds. In the right column, the linear-regression method is used to estimate DPOAE thresholds. The solid line in each panel represents the best-fit line to the data. Also shown as insets in the upper left corner of each panel are correlation coefficients, the number of threshold comparisons, and standard errors. In an effort to improve the accuracy with which the data from DPOAE input/output (I/O) functions predicted behavioral thresholds, all behavioral thresholds <0 dB HL were set to 0 dB HL and all estimated DPOAE thresholds <0 dB SPL and FPL were set to 0 dB SPL & FPL. Additionally, any threshold >60 dB was set to 60 dB as explained in the text.

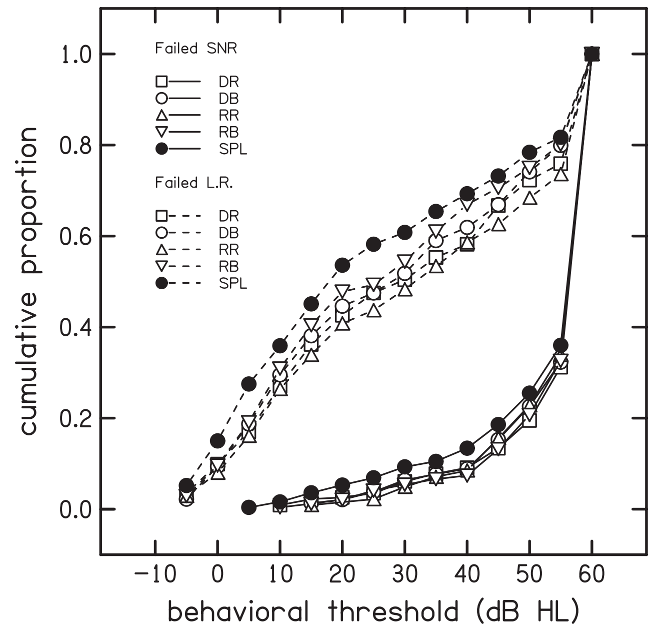

Fig. 3. Cumulative proportions of subjects who failed the inclusion criteria for the linear regressions in the linear-regression threshold estimation method as a function of behavioral threshold (in dB HL). Failed SNR refers to proportion of subjects who failed the SNR criterion in which the input/output (I/O) function must have at least three points with an SNR of 6 dB or greater. This criterion is applied first. Failed linear regression refers to those subjects who then failed the linear-regression criteria, which are used to determine how well a linear equation fits the transformed distortion product otoacoustic emission (DPOAE) I/O functions. These criteria included: (1) the slope of the individual linear regressions, which had to be ≥0.2 μPa/dB, (2) the variance accounted for (r^2), which needed to be ≥0.8, and (3) the standard error, which had to be ≤10 dB). The SNR criterion associated with this figure should not to be confused with the SNR-based threshold-prediction method described throughout this article. This figure applies to the linear-regression method of threshold estimation only. DR, daily room; DB, daily body; RR, reference room; RB, reference body.

be ≥0.8, and (3) the standard error had to be ≤10 dB. The upper cluster of functions represents the proportion of subjects failing to meet the criteria associated with the regression analysis, whereas the lower cluster of functions describe the proportions failing to meet the SNR criterion. Approximately 95% of the DPOAE I/O functions that were excluded based on the SNR criterion had behavioral thresholds >30 dB HL. Thus, it is highly probable that a subject whose I/O function could not be used during the linear-regression analysis because they did not produce responses with positive SNRs had hearing loss. Approximately 70% of the I/O functions that were excluded based on the criteria associated with the linear fit had behavioral thresholds of 40 dB HL or less. Thus, it is likely that subjects who failed the linear-regression criteria have normal to near-normal hearing. This has implications regarding the extent to which the linear-regression method and its associated criteria were applicable. Since the SNR criterion was applied first, every subject failing to meet the criteria associated with the linear fits produced at least three responses with SNRs ≥6 dB. Among all five of the calibration methods, I/O functions obtained using SPL calibrations had the highest cumulative percentage of those failing both sets of inclusion criteria.

Effects of Degree of Hearing Loss

For each calibration method, the difference between DPOAE and behavioral thresholds as a function of behavioral thresholds collapsed across frequency is shown in Figure 4. This figure is used to illustrate differences between calibration methods and DPOAE threshold-estimation methods in relation to degree of hearing loss. The open symbols represent the differences when DPOAE thresholds were estimated using the SNR-based method, and the closed symbols represent the differences when DPOAE thresholds were estimated using the linear-regression method.

Frequency Effects

In Figure 5, behavioral thresholds (dB HL) as a function of DPOAE thresholds (dB SPL and FPL) are shown for each frequency (shown separately in each row) and for each calibration method (shown separately in each column) when the linear-regression threshold-estimation method is used. Frequency effects for the SNR-based threshold-estimation method are essentially the same and thus not shown. Correlations were greatest at 3, 4, and 6 kHz and lowest at 8 kHz, with intermediate correlations at 2 kHz. More importantly, there does not seem to be an effect of calibration method at any frequency. There are conditions in which one of the FPL calibration methods resulted in the highest correlation, but there are also conditions in which the correlations between behavioral and DPOAE thresholds were highest after the SPL calibration. In any case, the differences in correlations among calibration methods were small.

CONCLUSIONS

The purpose of this study was to determine the extent to which stimulus-level calibration method affects the prediction of behavioral thresholds from DPOAE thresholds. Although differences were observed among calibration procedures, the differences were always small, and there were no consistent trends among calibration method relative to threshold predictions. Recall that the calibrations were taken at both room and body temperatures. A lack of difference among calibration methods indicates a lack of temperature effects. Results also indicate that there was no effect of time at which the calibration was taken (daily versus a reference taken on a single day before any data collection). Effects of frequency were similar to those noted by Gorga et al. (2003). Regardless of the way stimulus levels were calibrated, correlations between DPOAE and behavioral thresholds were largest for frequencies between 3 and 6 kHz and lowest at 8 kHz. Correlations were slightly higher when DPOAE thresholds were estimated using the linear-regression technique, compared with the SNR-based technique, but again these differences were small and inconsistent.

Referring to Figure 4, the differences between the two threshold-estimation methods as a function of behavioral thresholds are small. However, almost without exception, there is a better agreement between behavioral thresholds and those estimated using the linear-regression method, compared with the SNR-based method. Note that differences between DPOAE and behavioral thresholds are relatively constant for behavioral thresholds up to about 45 dB HL, with DPOAE threshold 5 to 15 dB higher than behavioral thresholds. At and above behavioral thresholds of 45 dB HL, there is a trend toward decreasing

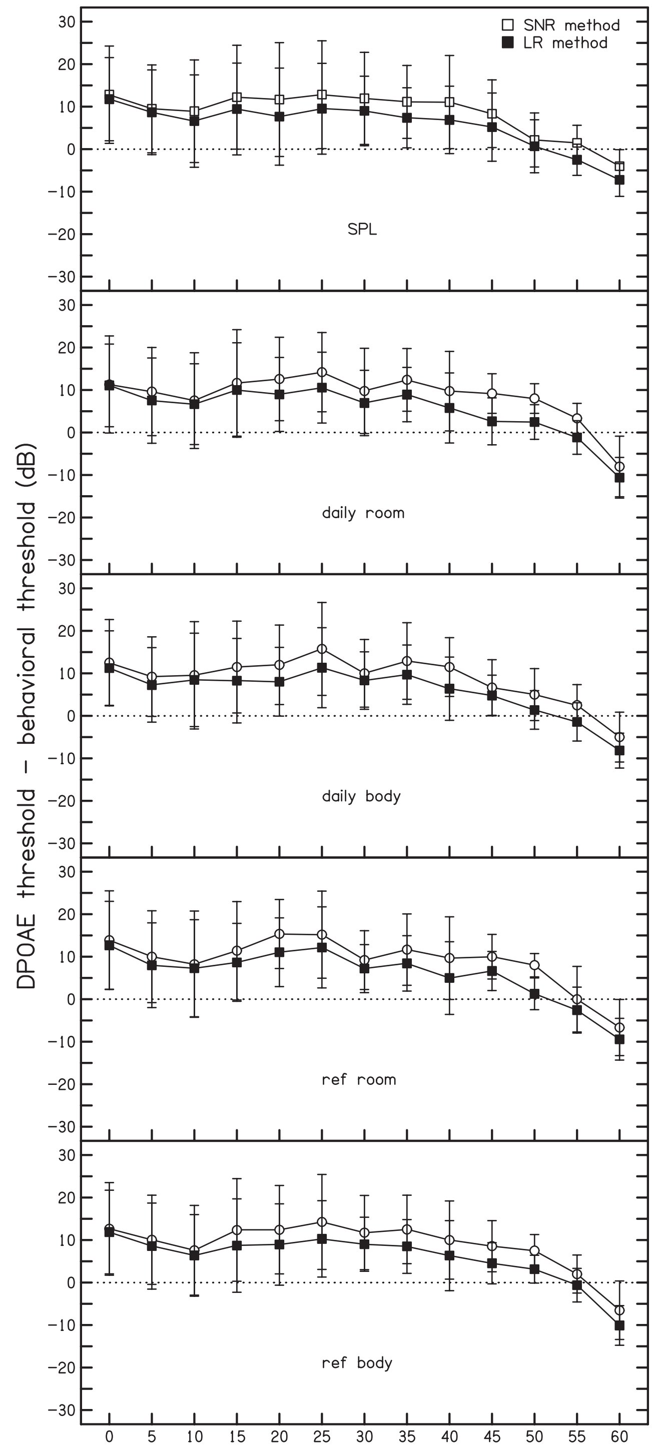

Fig. 4. For each calibration method, the difference between distortion product otoacoustic emission (DPOAE) and behavioral thresholds as a function of behavioral thresholds is collapsed across frequency. Open symbols represent the differences when the SNR-based method was used to estimate DPOAE thresholds, whereas closed symbols represent the case when DPOAE thresholds were estimated by the linear-regression method.

differences as behavioral threshold increases, with some differences below zero. This may be due to a smaller number of conditions at which thresholds could be estimated from DPOAEs, the truncation of thresholds as described in relation

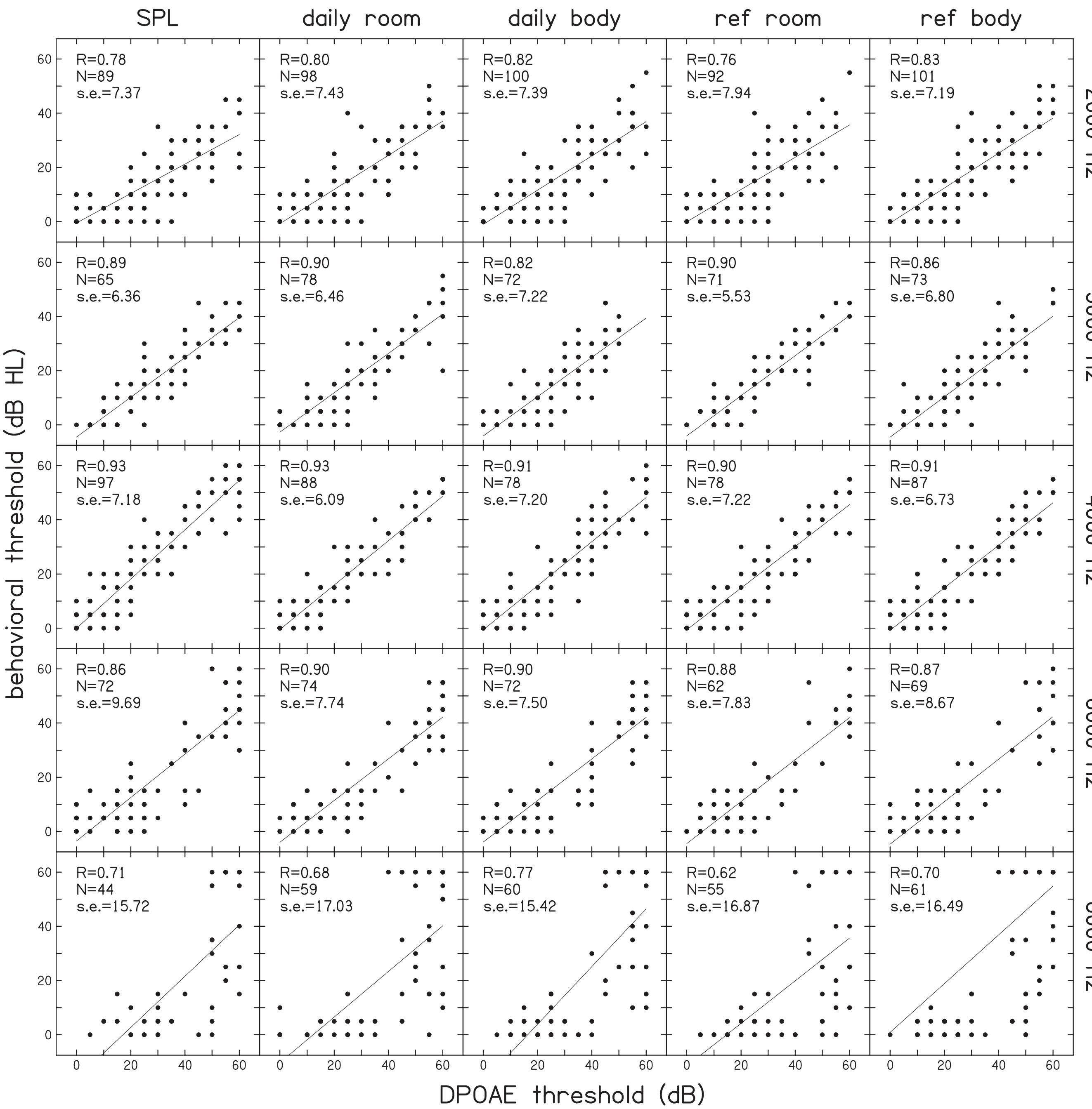

Fig. 5. Behavioral thresholds (dB HL) as a function of distortion product otoacoustic emission (DPOAE) thresholds (dB SPL and FPL) at each frequency (shown separately in each row) and for each calibration method (shown separately in each column) when the linear regression threshold estimation method is used. Insets in each panel provide correlations, number of subjects, and standard errors.

to Figure 2, or more accurate predictions of behavioral thresholds at these levels. Importantly, there is no apparent effect of calibration method. Although it seems that DPOAE thresholds for both threshold-estimation methods overestimate behavioral thresholds, the size of this effect seems to be independent of calibration procedure, further demonstrating that there was little or no effect of calibration method on the extent to which DPOAEs accurately predict behavioral threshold. In total, these results suggest that SPL calibrations may be adequate when attempting to predict pure-tone thresholds from DPOAEs, at least for the five frequencies studied in this experiment.

These findings are in agreement with those reported by Burke et al. (2010) in a companion article in which the effects of stimulus calibration on DPOAE test performance were assessed. Burke et al. found a calibration effect only at 8 kHz with essentially no differences in performance related to calibration method at other frequencies. The results from this study and from that by Burke et al. were not anticipated, given the known effects of standing waves on estimates of ear-canal SPL at the plane of the probe (Siegel 1994, 2007; Siegel & Hirohata 1994; Dreisbach & Siegel 2001). The results are particularly surprising because frequencies were studied at which standing-wave effects are expected to occur. However, it is important to note that differences between calibrations will only occur at or near frequencies at which standing waves are present. In this study and in its current clinical form, the measurement of DPOAEs is performed at discrete frequency intervals. This approach samples frequency broadly, and it is possible that the dimensions of the ear canals of this sample of subjects had few occurrences of standing waves at specific test frequencies, even though we evaluated DPOAE threshold estimations for octave and interoctave frequencies in a range in which standing-wave problems are expected.

Although the results of this study do not suggest a need to change the current calibration methods for DPOAEs, it is possible that future changes in DPOAE test protocol (i.e., the inclusion of more test frequencies or swept frequencies) may warrant further investigation of calibration effects. Despite our inability to detect a calibration-method effect, it remains the case that SPL is susceptible to standing-wave problems and FPL is not. Furthermore, Scheperle et al. (2008) reported that FPL calibrations provide more consistent DPOAE measurements when probe-insertion depth is varied, suggesting that FPL calibrations may yield less variable test results if the probe is removed and reinserted during testing, a fairly common clinical scenario, or if DPOAEs are being monitored over time. Thus, the increase in calibration time required for FPL calibrations may still be warranted to provide a consistently reliable estimate of level in the ear canal.

ACKNOWLEDGMENTS

The authors thank Sandy Estee for her assistance in subject recruitment and scheduling of subjects.

This work was supported by NIH grants DC T35 8757, DC R01 2251, DC P30 4662, and DC R01 8318.

These data were presented at the 2009 meeting of the American Auditory Society.

Address for correspondence: Michael P. Gorga, PhD, Boys Town National Research Hospital, 555 North 30th Street, Omaha, NE 68131. E-mail: Michael.gorga@boystown.org.

Received July 8, 2009; accepted February 6, 2010.

REFERENCES

American Speech-Language-Hearing Association. (2005). *Guidelines for Manual Pure-Tone Threshold Audiometry* [Guidelines]. Retrieved from www.asha.org/policy, December 22, 2009.

Boege, P., & Janssen, T. (2002). Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears. *J Acoust Soc Am*, *111*, 1810–1818.

Burke, S. R., Rogers, A. R., Neely, S. T., et al. (2010). Influence of calibration method on DPOAE measurements. I. Test performance. *Ear Hear. 31,* xxx–xxx.

Dreisbach, L. E., & Siegel, J. H. (2001). Distortion-product otoacoustic emissions measured at high frequencies in humans. *J Acoust Soc Am, 110(Pt 1)*, 2456–2467.

Frank, T., Durrant, J. D., & Lovrinic, J. M. (1993). Maximum permissible ambient noise levels for audiometric test rooms. *Am J Audiol*, *2*, 33–37.

Gorga, M. P., Neely, S. T., Bergman, B. M., et al. (1993). Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses. *J Acoust Soc Am, 93,* 2050–2060.

Gorga, M. P., Neely, S. T., Ohlrich, B., et al. (1997). From laboratory to clinic: A large scale study of distortion product otoacoustic emissions in ears with normal hearing and ears with hearing loss. *Ear Hear, 18,* 440–455.

Gorga, M. P., Neely, S. T., Dorn, P. A., et al. (2003). Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions. *J Acoust Soc Am*, *113*, 3275–3283.

Gorga, M. P., Nelson, K., Davis, T., et al. (2000). DPOAE test performance when both $2f_1$-f_2 and $2f_2$-f_1 are used to predict auditory status. *J Acoust Soc Am*, *107,* 2128–2135.

Johnson, T. A., Neely, S. T., Kopun, J. G., et al. (2007). Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance. *J Acoust Soc Am, 122,* 3539–3553.

Kummer, P., Janssen, T., & Arnold, W. (1998). The level and growth behavior of the 2f1 − 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.

Kummer, P., Janssen, T., Hulin, P., et al. (2000). Optimal L_1-L_2 primary tone level separation remains independent of test frequency in humans. *Hear Res*, *146,* 47–56.

Martin, G. K., Ohms, L. A., Franklin, D. J., et al. (1990). Distortion product emission in humans. III. Influence of sensorineural hearing loss. *Ann Otol Rhinol Laryngol Suppl*, *147*, 30–42.

Neely, S. T., Johnson, T. A., & Gorga, M. P. (2005). Distortion-product otoacoustic emission measured with continuously varying stimulus level. *J Acoust Soc Am*, *117*, 1248–1259.

Neely, S. T., & Gorga, M. P. (1998). Comparison between intensity and pressure as measures of stimulus level in the ear canal. *J Acoust Soc Am*, *104*, 2925–2934.

Neely, S. T., & Liu, Z. (1994). EMAV: Otoacoustic Emission Averager. *Technical Memo No. 17.* Boys Town National Research Hospital, Omaha, NE.

Scheperle, R. A., Neely, S. T., Kopun, J. G., et al. (2008). Influence of *in situ*, sound-level calibration on distortion-product otoacoustic emission variability. *J Acoust Soc Am*, *124*, 288–300.

Siegel, J. H. (2007). Calibrating otoacoustic emission probes. *Otoacoustic Emissions: Clinical Applications* (3rd ed.). New York: Thieme.

Siegel, J. H. (1994). Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. *J Acoust Soc Am, 95(Pt 1)*, 2589–2597.

Siegel, J. H., & Hirohata, E. T. (1994). Sound calibration and distortion product otoacoustic emissions at high frequencies. *Hear Res*, *80*, 146–152.