Published in final edited form as:

Ear Hear. 2010 February ; 31(1): 74–83. doi:10.1097/AUD.0b013e3181b71924.

Clinical test performance of distortion product otoacoustic emissions using new stimulus conditions

Tiffany A. Johnson, Ph.D.^a, Stephen T. Neely, D.Sc.^b, Judy G. Kopun, M.A.^b, Darcia M. Dierking, Au.D.^b, Hongyang Tan, M.S.^b, and Michael P. Gorga, Ph.D.^b

^a Speech-Language-Hearing, University of Kansas, Lawrence, KS

^b Boys Town National Research Hospital, Omaha, NE

Abstract

Objectives—To determine if new stimulus parameters, which have been shown to produce large DPOAE levels in a group of normal-hearing listeners (Neely et al., 2005; Johnson et al., 2006), result in more accurate identification of auditory status and more accurate predictions of behavioral threshold than traditional stimulus conditions.

Design—DPOAE I/O functions for 8 f_2 frequencies ranging from 0.7 to 8 kHz were recorded from 96 ears with normal hearing and 226 ears with sensorineural hearing losses ranging from mild to profound. The primary-level differences and primary-frequency ratios were set according to the stimulus relations developed by Johnson et al. (2006). The accuracy of the dichotomous decision task (area under the relative operating characteristic curve, A_{ROC}) for these new stimulus conditions was evaluated as a function of L_2 and was compared to previous reports in the literature where traditional stimuli were used (Stover et al., 1996). Here, traditional stimuli are defined as $L_1=L_2+10$ and $f_2/f_1=1.22$ for all L_2 and f_2 values. In addition to I/O functions, DPgrams with $L_2=55$ dB SPL and f_2 ranging from 0.7 to 8 kHz were recorded for each subject using the traditional stimuli. This provided a direct within-subject comparison of A_{ROC} for moderate-level stimuli when the new and traditional stimuli were used. Finally, the accuracy with which DPOAE thresholds predicted behavioral thresholds was evaluated in relation to previous reports in the literature for two definitions of DPOAE threshold, one where the entire I/O function was used to make the prediction and a second where the lowest L_2 producing an $SNR \geq 3$ dB was used.

Results—There was no evidence that the new stimuli improved the accuracy with which auditory status was identified from DPOAE responses. With both the new and traditional stimuli, moderate stimulus levels ($L_2 = 40$ to 55 dB SPL) resulted in the most accurate identification of auditory status. When $L_2 = 55$ dB SPL, the new stimuli produced A_{ROC} values that were equivalent to those observed with traditional stimuli. The new stimuli resulted in more accurate prediction of behavioral threshold for several f_2 's when using the entire I/O function, although the effect was small. Furthermore, using the entire I/O function to predict behavioral threshold results in more accurate predictions of behavioral threshold than using the SNR definition, although this approach can be applied to a smaller percentage of ears.

Conclusions—The new stimuli that had previously been shown to produce large DPOAE levels in normal-hearing listeners (Neely et al., 2005; Johnson et al., 2006) do not result in more accurate identification of auditory status and have only a small positive effect on the prediction of behavioral threshold.

Keywords

distortion product otoacoustic emissions; test performance; threshold prediction; hearing screening

Introduction

Distortion product otoacoustic emissions (DPOAEs) have been used to identify and diagnose hearing loss, including screening for hearing loss in newborns and infants. A number of studies have investigated the accuracy with which DPOAEs separate normal from impaired ears (e.g., Gorga et al., 1993, 1997; Kim et al., 1996; Stover et al., 1996; Musiek & Baran, 1997; Norton et al., 2000). The results of these studies have suggested that DPOAEs most accurately identify auditory status from 2 to 4 kHz, although errors in identification are observed at all frequencies. Several investigators also have shown a relationship between thresholds derived from DPOAE measures and pure-tone behavioral thresholds (Martin et al., 1990; Nelson and Kimberley, 1992; Dorn et al. 1999; Boege and Janssen, 2002; Gorga et al., 2003; Oswald & Janssen, 2003; Johnson et al., 2007). In all of these studies, the relationship between DPOAE and behavioral thresholds is characterized by variability.

One factor that may influence the accuracy with which DPOAEs identify auditory status or predict behavioral threshold is the choice of stimulus parameters used to elicit these responses, especially as a function of frequency. DPOAEs are evoked with pairs of primary tones with different frequencies, f_1 and f_2 ($f_2 > f_1$), and with levels, L_1 and L_2 , that may be equal or different. Many combinations of primary frequencies and levels are possible. Given the large number of potential combinations, it is not possible to evaluate the influence of all combinations on the accuracy with which DPOAEs identify auditory status in large groups of subjects. Instead, most studies of the clinical accuracy of DPOAEs use stimulus parameters that have been identified previously as producing robust DPOAE levels in normal-hearing ears (e.g., Harris et al., 1989; Brown and Gaskill, 1990; Gaskill and Brown, 1990; Hauser and Probst, 1991; Brown et al., 1994; Whitehead et al., 1995 a, b; Kummer et al., 1998). Based on these studies, commonly used parameters include primary-frequency ratios (f_2/f_1) of approximately 1.2 and L_1 , L_2 relationships that are either fixed across level (e.g., Kim et al., 1996; Stover et al., 1996; Gorga et al., 1997) or vary with L_2 according to the relation $L_1 = 0.4 \cdot L_2 + 39$ dB (e.g., Kummer et al., 1998, 2000; Dorn et al., 2001; Boege & Janssen, 2002; Gorga et al., 2003). No large-scale human studies have evaluated the accuracy of DPOAEs when f_2/f_1 differs from approximately 1.2 or when stimulus parameters vary with f_2 and L_2 . Data reported by Whitehead et al. (1995b) suggest that, at least for rabbits, allowing f_2/f_1 and L_1 , L_2 to vary with frequency and level increased the likelihood that DPOAEs would be reduced following cochlear insult.

Two recent studies have demonstrated that larger DPOAE levels occur in normal-hearing subjects when primary-frequency ratios and/or level differences are allowed to vary with f_2 and L_2 (Neely et al., 2005; Johnson et al., 2006). Neely et al. (2005) demonstrated that the optimal¹ primary-level difference depends on frequency, with the largest effects observed for the two highest frequencies they studied ($f_2 = 4$ and 8 kHz). In contrast, Kummer et al. (2000) reported that the primary-level difference producing the largest DPOAE level was independent of frequency. Johnson et al. (2006) measured larger DPOAE levels when primary-frequency ratio increased and primary-level difference decreased as f_2 decreased and L_2 increased. The data reported by both Neely et al. and Johnson et al. were collected in normal-hearing ears and, therefore, cannot be used to describe the accuracy with which they

¹Here, and throughout the manuscript, we use “optimal” to refer to those stimulus conditions (primary-frequency ratio and primary-level difference) that, on average, result in the largest DPOAE for subjects with normal hearing.

identify auditory status or predict threshold. Test performance can only be determined when data from both normal-hearing and hearing-impaired ears are included. Similarly, understanding the relationship between behavioral threshold and DPOAE threshold requires data from ears whose thresholds range from normal up to at least a moderate degree of loss. While optimal stimulus conditions are expected to produce larger responses, it remains undetermined whether these stimulus conditions influence the accuracy of either the dichotomous-decision task or the threshold-prediction task.

The purpose of the present study was to evaluate DPOAE test performance and the accuracy with which DPOAE thresholds predict behavioral thresholds when using the optimal stimulus parameters derived from our earlier work. It was hypothesized that the optimal parameters would result in a more accurate dichotomous decision and a stronger relationship between behavioral and DPOAE thresholds.

Methods

A. Subjects

Data were collected from 51 subjects (34 females) with normal hearing and 119 subjects (55 females) with sensorineural hearing loss. An additional 5 subjects (5 females) had normal hearing in one ear and sensorineural hearing loss in the opposite ear. For these 5 subjects, results for the normal-hearing ear were included with the normal data and results for the hearing-impaired ear were included with the impaired data. For most subjects, data were collected from both ears; however, in some cases data could not be collected for both ears. Reasons for which this occurred included the presence of a cochlear implant in the non-test ear, middle-ear dysfunction in the non-test ear, asymmetric hearing loss of unknown etiology, and subject time constraints. As a result, data were collected in a total of 96 normal-hearing ears and 226 hearing-impaired ears. Normal hearing was defined as thresholds ≤ 20 dB HL (ANSI, 1996) for the octave and inter-octave frequencies from 0.25 to 8 kHz. For the purposes of counting the number of subjects, subjects were considered hearing impaired if one or more pure-tone behavioral thresholds were > 20 dB HL for the same frequency range.² Only subjects with sensorineural hearing losses (defined as air-bone gaps < 15 dB at the octave frequencies of 0.5 through 4 kHz and normal 226-Hz tympanograms) were included in the study. The subjects with normal hearing ranged in age from 11 to 58 years, while the subjects with impaired hearing were between 11 and 80 years of age.

The degree of hearing loss ranged from mild to profound. Table 1 shows the distribution of thresholds across frequency for those subjects who participated in the study. While subjects presented with audiometric thresholds ranging from normal hearing to profound hearing loss, effort was made to increase the representation of subjects with normal hearing and mild or moderate hearing losses on the assumption that these subjects would be more likely to produce DPOAEs, compared to subjects with greater degrees of hearing loss. Emphasis on subjects with these degrees of hearing loss may have had the inadvertent effect of reducing estimates of test performance relative to the performance one might observe in an unselected sample of subjects. This would occur because the overlap between ears with normal and impaired hearing is likely to be greater for mild and moderate degrees of hearing loss.

²As will be described in more detail below, when evaluating the accuracy with which DPOAEs separate normal from impaired ears (the dichotomous decision) or the accuracy of DPOAE threshold prediction, the determination of normal or impaired was made on a frequency-by-frequency basis.

B. Equipment and Calibration

All data were collected using custom-designed software (EMAV, Neely and Liu, 1993) that controlled a 24-bit soundcard (CardDeluxe, Digital Audio Labs) housed in a PC. An ER-10C (Etymotic Research) probe microphone was used to present stimuli and record responses. The ER-10C had been modified to remove 20 dB of attenuation so that we were able to achieve stimulus levels as high as 80 dB SPL.

Stimulus levels were calibrated *in situ* in sound pressure level (SPL) at the plane of the ER-10C probe. In-the-ear pressure calibration may introduce variability into the measure as a result of acoustic reflection from the eardrum in the ear canal (Siegel, 1994; Siegel and Hirohata, 1994; Neely and Gorga, 1998). Calibration of stimulus levels in intensity or forward pressure may reduce some of the variability associated with pressure calibration (Neely and Gorga, 1998; Scheperle et al., 2008); however, intensity and forward-pressure calibrations have not been widely used and *in situ* SPL calibration remains the standard of practice for DPOAE measures. While eardrum reflection associated with SPL calibration may be a potential source of error, we would expect that any calibration errors would equally influence test performance and threshold predictions for standard and optimal stimulus conditions. Thus, the comparisons being made in this study should be valid.

C. DPOAE Stimuli

DPOAE data were collected in response to pairs of primary tones ($f_1, f_2; f_2 > f_1$) with f_2 ranging from 0.7 to 8 kHz in $\frac{1}{2}$ -octave steps. The level of f_2 (L_2) ranged from -20 to 80 dB SPL in 5-dB steps. The level of f_1 (L_1) and f_2/f_1 were set according to the following equations:

$$L_1 = 80 + 0.137 \cdot \log_2(18/f_2) \cdot (L_2 - 80) \quad (\text{EQ. 1})$$

$$f_2/f_1 = 1.22 + \log_2(9.6/f_2) \cdot (L_2/415)^2. \quad (\text{EQ. 2})$$

The stimulus conditions represented by these equations have been shown to produce, on average, the largest DPOAE levels (L_d) in normal-hearing human ears (Johnson et al., 2006). We will refer to these stimulus settings as the optimal stimulus parameters.

In addition to the DPOAE I/O functions recorded for the optimal stimulus conditions described above, a DPgram was recorded for each subject. For the DPgram, f_2 was set at octave and half-octave frequencies from 0.7 to 8 kHz but L_2 was fixed at 55 dB SPL and traditional primary level and frequency relationships were used where L_1, L_2 was fixed at 65, 55 dB SPL and $f_2/f_1 = 1.22$ for all f_2 .³ These stimuli were chosen for comparison to large-scale studies in which the accuracy of the dichotomous decision was evaluated (e.g., Stover et al., 1996; Gorga et al., 1997).

Recording I/O functions using the stimulus conditions described in Eqs. 1 and 2 not only allowed us to evaluate the test performance of DPOAEs in determining auditory status, but also provided the opportunity to determine the accuracy with which DPOAEs predict behavioral thresholds for stimuli that produce larger DPOAE levels in normal-hearing subjects. Furthermore, the DPgram served as a control condition and allowed us to directly

³Here, and throughout the manuscript, we use “traditional” to refer to those stimulus conditions where $L_1 = L_2 + 10$ and $f_2/f_1 = 1.22$ for all L_2 and f_2 .

compare the accuracy of the dichotomous decision for the optimal (Johnson et al., 2006) and the traditional stimuli, at least for the stimulus level for which test performance is best (Whitehead et al., 1995a; Stover et al., 1996; Johnson et al., 2007).

D. Procedures

All data were collected in a sound-treated room. Subjects were seated in a reclining chair and slept, read quietly, or watched a silent, captioned movie. Data were typically collected in one or two 2-h sessions. Subjects with both normal and impaired hearing were required to have normal 226-Hz tympanograms at each data-collection session.

Pure-tone behavioral thresholds were obtained at the first session using standard clinical procedures with a 5-dB step size. Either TDH-39 supra-aural headphones or ER-3A insert earphones were used for threshold testing. These headphones were calibrated according to their respective standards (ANSI, 1996). While DPOAE data were collected at octave and half-octave frequencies from 0.7 to 8 kHz, behavioral-threshold measurements were obtained at the octave and inter-octave frequencies used in clinical audiometry (e.g., 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, and 8 kHz). As a result, there were slight differences in the frequencies at which DPOAE and behavioral thresholds were compared for the half-octave/inter-octave comparisons. These differences were small, and would not be expected to affect comparisons between audiometric and DPOAE data, and they would be presumed to have the same effect regardless of whether optimal or traditional stimulus conditions were used.

DPOAE data were collected as either DPgrams (L_2 constant and f_2 changing) or I/O functions (L_2 changing and f_2 constant). When recording I/O functions, L_2 was decreased until the response was < 3 dB above the noise floor. L_d was estimated as the level in the $2f_1$ - f_2 frequency bin. The noise floor was estimated from the $2f_1$ - f_2 frequency bin as well as the 5 bins on either side of the $2f_1$ - f_2 frequency. Estimation of noise level in the $2f_1$ - f_2 bin was accomplished by alternately storing 0.25-s samples of the recorded response into one of two buffers. The buffers were summed to provide an estimate of L_d (at $2f_1$ - f_2). They were subtracted to provide a noise estimate. By defining noise from the $2f_1$ - f_2 frequency bin, we estimate noise and response at the same frequency. By including the levels in the adjacent frequency bins, a more stable estimate of noise was obtained than if noise was estimated from the $2f_1$ - f_2 bin alone.

Measurement-based stopping rules were used during data collection. For each condition, averaging continued until the noise floor was ≤ -25 dB SPL or 32 s of artifact-free averaging time had elapsed, whichever occurred first. The use of measurement-based stopping rules results in more consistent noise levels across conditions and subjects because the averaging time was increased whenever the noise level exceeded -25 dB SPL. It should be noted, however, that the extent to which data collection ended on the noise-floor stopping rule was frequency dependent. With few exceptions, data collection stopped on this rule for higher f_2 frequencies and data collection almost always stopped on the test-time limit for lower f_2 frequencies. Longer test times could have been used for low frequencies, which would have resulted in more uniform noise floors across conditions and subjects; however, the additional time would have made it difficult to collect data on the large number of subjects needed to answer our primary questions and complete sets of data would not have been possible from many of the subjects with hearing loss. As a consequence, the test-time compromise was necessary. Furthermore, this compromise should have no effect on the primary questions being addressed in this study, namely to determine the extent to which optimal stimulus conditions affect test performance and threshold predictions.

E. Data Analyses

The influence of stimulus condition on the accuracy with which DPOAEs identified hearing loss was assessed using clinical decision theory (CDT) (Swets and Pickett, 1982; Swets, 1988). CDT is well suited to assessing the accuracy with which diagnostic tests make a dichotomous decision, such as normal versus impaired hearing, and has been used previously to evaluate the test performance of DPOAEs (e.g., Gorga et al., 1993, 1997, 1999, 2000, 2005; Kim et al., 1996; Stover et al., 1996; Dorn et al., 1999; Johnson et al., 2007). In the application of CDT to DPOAE data, pure-tone behavioral thresholds served as the gold standard to which DPOAE data (L_d in the present study) were compared. Ears with behavioral thresholds > 20 dB HL were defined as hearing impaired, while ears with behavioral thresholds ≤ 20 dB HL were defined as normal hearing. The classification of normal versus impaired was made on a frequency-by-frequency basis. For example, an ear with a behavioral threshold of 15 dB HL at 1 kHz and a threshold of 30 dB HL at 2 kHz would be defined as normal at 1 kHz and impaired at 2 kHz for the purposes of data analyses.

A description of DPOAE test performance was obtained by computing hit rates (sensitivity), which is the proportion of ears with hearing loss that were correctly identified, and corresponding false-alarm rates (one minus the specificity), which is the proportion of normal-hearing ears incorrectly identified as hearing impaired. These analyses were performed for all possible DPOAE levels at each f_2 and L_2 . Relative operating characteristic (ROC) curves (plots of hit rate versus false-alarm rate) were constructed for these conditions and the area under each ROC curve (A_{ROC}) was computed. A_{ROC} provides a single estimate of test accuracy and ranges in value from 0.5 where hit and false-alarm rates are equal (chance performance) to 1.0 (perfect test performance) where the hit rate is 100% for all false-alarm rates, including a false-alarm rate of 0%. In the present study, A_{ROC} values were computed at each L_2 and f_2 combination for the optimal stimulus parameters and at each f_2 for the control condition (DPgram) recorded with the traditional stimuli where only $L_2=55$ dB SPL was tested. Like the selection of maximum averaging time, the decision to include measurements limited to this control condition was necessary, given the large number of subjects on whom data were needed in order to assess test performance and threshold predictions. Because the present study limited control-condition measurements to a single level condition, the previously published results from Stover et al. (1996), in which test performance was assessed in relation to L_2 , will be compared to the present findings. Those data also were collected in a relatively large sample of subjects with both normal and impaired hearing under similar measurement conditions. The main difference between those data and the present data is that the primary-level difference in Stover et al. was held constant at 10 dB for all L_2 and the primary-frequency ratio was fixed at 1.2 for all f_2 , whereas, in the present study, the primary-level differences and primary-frequency ratios vary in a way that optimizes response level in normal-hearing subjects.

Results

A. DPOAE Test Performance

In Fig. 1 (taken from Stover et al., 1996), A_{ROC} is plotted as a function of L_2 with data for the octave frequencies from 0.5 to 8 kHz plotted in the left panel and half-octave frequencies plotted in the right panel. A_{ROC} increased as L_2 increased up to an L_2 of approximately 50 to 55 dB SPL, beyond which A_{ROC} remained the same or decreased slightly with further increases in L_2 .⁴ This pattern was observed for all f_2 frequencies. Due to output limitations in the hardware used by Stover et al. (1996), it was not possible to explore test performance for values of L_2 exceeding 65 dB SPL. Data such as these suggest that moderate-level stimuli result in the best DPOAE test performance.

Figure 2 plots A_{ROC} for the optimal stimulus parameters as a function of L_2 based on the present data. For comparison purposes, results are also shown for the DPgram control condition (data points on the right side of each panel), in which $L_1, L_2 = 65, 55$ dB SPL. For ease of comparison between the I/O-function data (optimal stimuli) and the DPgram data (traditional stimuli), each panel plots data for 2 f_2 's as indicated in the legend. Maximum A_{ROC} values (i.e., best test performance) were observed for moderate stimulus levels (L_2 of 40 to 55 dB SPL) for the optimal stimulus conditions. When L_2 either decreased or increased relative to these levels, A_{ROC} decreased, indicating that both higher and lower stimulus levels resulted in more diagnostic errors. The decrease in performance with lower primary levels is attributed to an increase in the false-alarm rate, whereas the decrease in test performance for high primary levels is attributed to an increase in the false-negative rate. These relationships between L_2 and A_{ROC} are consistent with the Stover et al. (1996) data (re-plotted here in Fig. 1) that were obtained with traditional stimulus parameters and suggest that the optimal stimuli did not change the fundamental relationship between L_2 and A_{ROC} . However, because we were able to achieve L_2 levels as high as 80 dB SPL, the present data show the decrease in test performance for high-level stimuli more clearly than has been reported previously. For nearly all f_2 frequencies, the maximum A_{ROC} observed with the optimal stimulus conditions was similar (within ± 1 -3%) to that achieved with the control condition (DPgram), although the A_{ROC} for the control condition was larger for 6 of the 8 f_2 's.

Figure 3 plots A_{ROC} as a function of f_2 when $L_2 = 55$ dB SPL (a level that produced A_{ROC} values at or near maximum for all f_2 's) and L_1 and f_2/f_1 are set either at the level specified by the optimal stimuli (filled circles) or $L_1 = 65$ dB SPL and $f_2/f_1 = 1.22$ (traditional parameters, open circles). These data provide information similar to that shown in Fig. 2 but facilitate the comparison of the results for optimal and traditional stimuli. For this L_2 , A_{ROC} for both optimal and traditional stimuli exceeded 0.80 for all f_2 's and exceeded 0.90 for $f_2 \geq 1.4$ kHz. At each f_2 , the differences in A_{ROC} for the optimal and traditional stimuli were small ($< 5\%$ in most cases) and were not statistically significant ($p > 0.05$; Bamber, 1975). This suggests that the optimal stimuli do not result in more accurate identification of auditory status.

In an effort to gain insight into why maximum A_{ROC} values were not larger for optimal stimulus conditions, we compared the DPOAE levels observed for both optimal and traditional stimuli when recorded in the same ear, which is shown in Fig. 4. Here, the mean difference in DPOAE level between optimal and traditional stimuli is plotted as a function of f_2 for the condition in which $L_2 = 55$ dB SPL (the stimulus level that was tested for both traditional and optimal conditions in the present study). The dotted horizontal line represents the point where the DPOAE levels for optimal and traditional stimulus parameters are equal. Positive values indicate that the new parameters produced larger DPOAE levels, while negative values indicate that DPOAE levels were larger for traditional stimulus parameters. For both normal (filled squares) and impaired (open squares) ears, the optimal stimulus parameters produced similar increases in L_d , compared to the traditional stimulus parameters when $f_2 = 0.7$ -2.8 kHz. Thus, the centers of the normal and impaired response distributions were shifted towards higher L_d 's by equal amounts. Assuming that the variance is similar for traditional and optimal conditions, then the overlap between normal and impaired distributions would not have changed, which would explain why A_{ROC} did not increase with the use of optimal stimuli at these frequencies.

⁴Please note that Stover et al. (1996) also evaluated A_{ROC} when using DPOAE threshold to predict auditory status. A_{ROC} values for DPOAE threshold are shown as nonarticulated symbols at the far right of each panel. While A_{ROC} values for DPOAE threshold compare favorably to A_{ROC} values obtained when using DPOAE level to predict auditory status, DPOAE threshold could not be established in a large proportion of subjects. As a result, A_{ROC} based on DPOAE threshold may be artificially elevated and would not be expected to perform well in clinical applications. Please see Stover et al. (1996) for more details.

A different pattern is observed in Fig. 4 for $f_2 = 4$ and 5.6 kHz. For these two frequencies, L_d was less in the normal-hearing group for the optimal stimuli, compared to the DPOAE level observed for the traditional condition, whereas little or no change in level was observed for the impaired ears. Given this pattern, one would predict that test performance based on L_d as the criterion measure (and estimated by A_{ROC}) would decrease at these two frequencies because the centers of the two distributions of responses (normal and impaired) have moved closer together (by as much as 8 or 9 dB in the case of $f_2 = 5.6$ kHz). Indeed, A_{ROC} at these frequencies was less with the optimal stimuli than with the traditional stimuli, although this difference was not significant and was observed at several other f_2 frequencies.

Finally, at 8 kHz, the optimal stimulus conditions resulted in a 5-dB increase in DPOAE level in normal ears, but no change in level for impaired ears. Given this observation, one would predict greater separation between normal and impaired distributions for the optimal stimulus conditions, translating into a greater A_{ROC} , which was the case. However, only a 3% increase in A_{ROC} was observed, and again the increase was non-significant.

Taken together the results in Figs. 1-4 suggest that the optimal stimuli do not result in more accurate identification of auditory status than the traditional stimuli used in most clinical applications. Furthermore, while the optimal stimuli resulted in larger DPOAE levels for most f_2 frequencies, this increase in level was observed for both normal and impaired ears, and at two frequencies (4 and 5.6 kHz) the optimal stimuli resulted in smaller DPOAEs than the traditional stimuli in normal ears. This latter finding was unexpected and is not easily explained.

B. Behavioral Threshold Predictions from DPOAE Data

In addition to exploring the influence of optimal stimulus conditions on DPOAE test performance, we examined the accuracy with which DPOAE thresholds obtained with these stimuli predicted pure-tone behavioral thresholds. Two definitions of DPOAE threshold were used and both were correlated with pure-tone behavioral thresholds. In the first, DPOAE threshold was defined as the lowest L_2 for which the signal-to-noise ratio (SNR) was ≥ 3 dB and for which all L_2 's above this level produced SNRs > 3 dB. We will refer to this as the SNR approach. In order to reduce the contribution of system distortion to the results, cases where the lowest L_2 producing a 3-dB SNR was > 65 dB SPL (i.e., the DPOAE threshold was > 65 dB SPL) were treated as no response and threshold was not predicted. This decision is justified, if we assume that DPOAEs are a reflection of the status of the outer hair cells (OHC) and that it is unlikely that hearing losses exceeding 60 dB can be attributed solely to OHC dysfunction. Data describing the relationship between behavioral thresholds and DPOAE thresholds using the SNR approach are plotted in the Fig. 5, with each panel representing a different f_2 . The solid lines in each panel are the best-fit lines, with the equations shown as insets in each panel. Table 2 lists the percentage of ears meeting inclusion criteria, correlations, and standard errors for the two threshold-prediction methods used in the present study. For comparison purposes, similar results from Gorga et al. (2003) are also provided. The data plotted in Fig. 5 and Table 2 suggest that there is a relationship between behavioral threshold and DPOAE threshold predicted using the SNR approach, with the strongest correlations observed for f_2 's between 1.4 and 8 kHz. These are also the frequencies for which behavioral threshold could be predicted for the largest percentage of ears, ranging from 78 ($f_2=2.8$ kHz) to 94% ($f_2=4$ kHz).

Results for the second definition of DPOAE threshold are plotted in Fig. 6. As in Fig. 5, behavioral threshold (dB HL) is plotted as a function of DPOAE threshold (dB SPL), with each panel representing data for a different frequency. Here, DPOAE threshold was estimated using the technique described by Boege and Janssen (2002) for predicting a behavioral threshold from the DPOAE I/O function. In this approach, which we will refer to

as the I/O approach, data for L_2 levels from 20 to 65 dB SPL were evaluated to determine if at least 3 points on the I/O function produced an $\text{SNR} \geq 6$ dB. If an I/O function had fewer than 3 points meeting this criterion, it was excluded from further analysis and a DPOAE threshold was not predicted using this approach. For those I/O functions meeting the SNR inclusion criterion, L_d (dB SPL) was converted to pressure (μPa) and plotted as a function of L_2 (dB SPL). The resulting semi-log (μPa vs. dB SPL) I/O functions were fit with linear equations. The slope, correlation coefficient, and standard error resulting from this fit were evaluated to determine if the I/O functions met three additional criteria (described below) and could be included in further analysis (Boege and Janssen, 2002). Specifically, I/O functions were required to have slopes $\geq 0.2 \mu\text{Pa}/\text{dB}$, variances accounted for (r^2) ≥ 0.8 , and standard errors ≤ 10 dB. Those I/O functions not meeting one or more of these regression-based inclusion criteria were excluded from further analysis, meaning that a DPOAE threshold could not be estimated using this approach. For those I/O functions meeting these regression criteria, the DPOAE threshold was estimated by solving the linear equation for the L_2 that produced a DPOAE of $0 \mu\text{Pa}$. The DPOAE thresholds plotted in Fig. 6 represent those L_2 's at which a DPOAE of $0 \mu\text{Pa}$ was estimated when the linear equation was solved. The solid line in each panel is the best-fit line for the data, whose equation is shown as an inset in each panel. Table 2 lists the percentage of ears meeting all of the I/O function inclusion criteria, correlations (r), and the standard errors for the data plotted in Fig. 6. As was the case for the simpler SNR approach (shown in Fig. 5 but also summarized in Table 2), there is a relationship between behavioral thresholds and DPOAE thresholds predicted from the I/O function. The strongest correlations were observed for f_2 's from 1.4 through 8 kHz, frequencies for which the percentage of ears for which a DPOAE threshold could be predicted ranged from 42 ($f_2=1.4$ kHz) to 54% ($f_2=4$ kHz).

Comparison of the correlations between behavioral thresholds and DPOAE thresholds for the SNR and I/O approaches indicates that the correlations were similar for the two approaches but that a DPOAE threshold could be predicted for a larger percentage of ears when using the SNR approach (by as much as 40-49% for some f_2 's) as compared to the I/O approach. When the correlations between SNR-based DPOAE thresholds and behavioral thresholds for the present study were compared to those reported by Dorn et al. (2001) for different stimulus parameters, similar correlations were observed. Therefore, the near equivalence in the correlations between the SNR and the I/O-approach thresholds and behavioral threshold for the present study does not appear to be related to the optimal stimulus parameters. In contrast to the similarities in correlations, comparison of the standard errors for the two approaches (see Table 2) suggests differences in the accuracy with which they predict behavioral threshold. At each f_2 , the standard error for the I/O approach is smaller (by as much as 6 dB) than the standard error for the SNR approach. Smaller standard errors suggest that the predictive accuracy of the I/O approach is better than the predictive accuracy of the SNR approach. The similarity in the strength of the correlations for the I/O approach and the SNR approach suggest that there is a similar linear relationship between the behavioral and DPOAE thresholds estimated using the two methods. However, even though a larger percentage of DPOAE thresholds could be estimated with the SNR approach, the increase in the number of ears is accompanied by a decrease in the accuracy of the prediction as compared to the I/O approach.

Data that describe the influence of stimulus parameters on the accuracy of the I/O function approach are shown in Fig. 7 and Table 2. Figure 7 plots data describing differences in the strength of correlation (upper panel, A) and the percentage of thresholds predicted (lower panel, B) when the optimal stimuli are used, compared to stimuli used by Gorga et al. (2003). Gorga and colleagues used the Boege and Janssen (2002) approach to predict behavioral threshold from the DPOAE I/O function, but explored frequency effects in more detail than reported by Boege and Janssen. The Gorga et al. results are reproduced here

because they provide a basis for evaluating the influence of the optimal stimulus parameters on the accuracy of the behavioral threshold comparison for the I/O approach. Gorga et al. used a fixed f_2/f_1 ratio of 1.22 and set primary levels according to the relation $L_1 = 0.4L_2 + 39$ dB (Kummer et al., 1998; Janssen et al., 1998) for L_2 levels up to 65 dB SPL, beyond which $L_1 = L_2$.

In Fig. 7A, differences in the correlations for the two stimulus approaches are plotted as a function of f_2 . The dotted, horizontal line represents the point where the correlations are equal for the two approaches, while positive values indicate a stronger correlation for the optimal stimuli (used in the present study) and negative values indicate a stronger correlation for the stimuli used by Gorga et al. (2003). In many cases, the correlations were similar for the two stimulus approaches with correlation coefficients within 0.05 ($f_2 = 1.0, 4.0, 5.6,$ and 8 kHz). Changes in correlation exceeded 0.05 and favored the optimal stimuli at 0.7, 1.4, and 2.8 kHz. The only correlation change exceeding 0.05 that favored the stimuli used by Gorga et al. was observed at 2.0 kHz. In Fig. 7B, differences in the percentage of thresholds that could be predicted from the I/O function (those cases meeting all the inclusion criteria) are plotted as a function of f_2 . As in Fig. 7A, the dotted line indicates the point where the two stimulus approaches produced equivalent results. Positive values favor the optimal stimuli and negative values favor the stimuli used by Gorga et al. For nearly every frequency, a larger percentage of ears met all the inclusion criteria (SNR and regression) for optimal stimuli, compared to the results reported by Gorga et al. These differences ranged from 4 ($f_2 = 2$ kHz) to 11% ($f_2 = 1.0$ kHz). At two frequencies ($f_2 = 4.0$ & 5.6 kHz), between 1 and 3% more ears met all the inclusion criteria for the stimuli used by Gorga et al.

A summary of the data plotted in Fig. 7 is shown in Table 2 where the percentage of ears meeting inclusion criteria and the correlations for the Gorga et al. (2003) data are listed. Table 2 also provides the standard errors for the Gorga et al. data, which were not plotted in Fig. 7. At 5 of 8 f_2 's (0.7, 1, 1.4, 2.8, and 5.6 kHz), the standard error for the I/O approach using the optimal stimuli was smaller than the standard error for the stimuli used by Gorga et al., although in all cases except $f_2 = 0.7$ kHz, the differences were ≤ 2 -3 dB. In most cases, the I/O approach, regardless of stimulus paradigm, resulted in smaller standard errors (greater predictive accuracy) than those observed with the SNR approach.

The data shown in Figs. 5-7 and Table 2 are consistent with the idea that there is a relationship between behavioral threshold and DPOAE threshold predicted with either the SNR approach or the I/O approach. This relationship is best for f_2 's between 1.4 and 8 kHz, with $f_2 = 2.8$ and 4 kHz producing the strongest correlations. The use of optimal stimuli did not consistently affect the correlation for either the SNR or the I/O approach as compared to other stimulus approaches. For most f_2 's, the I/O approach using the optimal stimuli resulted in the most accurate prediction of behavioral threshold, as indicated by smaller standard errors, although the effect was small for most frequencies. Additionally, it appears that using the entire I/O function to predict behavioral threshold, regardless of stimulus paradigm, resulted in smaller standard errors than the SNR approach, although the SNR approach can be applied to a larger percentage of ears.

Discussion

In summary, the data reported here suggest that the optimal DPOAE stimulus conditions described by Neely et al. (2005) and Johnson et al. (2006) do not improve the accuracy with which auditory status is predicted from DPOAE measurements over what is achieved with traditional stimulus conditions. Moderate-level primaries ($L_2 = 40$ to 55 dB SPL) with optimal conditions result in the best test performance (as defined by A_{ROC}), with smaller

A_{ROC} values observed for both higher and lower primary levels. These results are similar to what has been described with more traditional stimulus conditions (e.g., Whitehead et al., 1995a; Gorga et al., 1997). Furthermore, the stimulus parameters used in many clinical applications of DPOAEs ($f_2/f_1 = 1.22$; $L_1, L_2 = 65, 55$ dB SPL) resulted in A_{ROC} values that were equivalent to those obtained for the optimal stimulus parameters when $L_2 = 55$ dB SPL. When the relationship between DPOAE thresholds and behavioral thresholds was evaluated, the optimal stimuli did not consistently strengthen the relationship between the DPOAE threshold and behavioral thresholds. Correlations between behavioral and DPOAE thresholds with the optimal stimuli for the SNR approach were similar to those reported by Dorn et al. (2001) using different stimuli. Similarly, the optimal stimuli resulted in both stronger and weaker correlations for the I/O approach when compared to similar correlations reported by Gorga et al. (2003). When using the entire I/O function to predict behavioral threshold, the optimal stimuli resulted in more accurate predictions (smaller standard errors) than the stimuli used by Gorga et al., although the effect was small and was only observed at 5 of 8 f_2 's. The SNR approach can be applied to a larger percentage of ears than the I/O approach; however, the accuracy of the SNR approach is less than the accuracy of the I/O approach, regardless of stimulus paradigm.

DPOAE Test Performance

These results suggest that the optimal stimulus conditions, which previously had been shown to produce larger DPOAE levels in normal-hearing ears (Neely et al., 2005; Johnson et al., 2006), do not result in a decrease in the overlap between the distributions of responses from normal and impaired ears at the moderate stimulus levels that produce the best test performance. Overall error rates are smallest for moderate-level conditions and increase for both higher and lower stimulus levels regardless of stimulus paradigm. Expanding on the previous work demonstrating that optimal stimuli produce larger responses in normal-hearing ears, it appears that at most frequencies these stimuli also produce an equivalent increase in DPOAE levels for impaired ears. Furthermore, for some f_2 's (4 and 5.6 kHz), the optimal stimuli have the undesirable effect of producing smaller DPOAEs in normal ears than the traditional parameters.

One factor that may have contributed to the lack of a stimulus effect may relate to the bias in the distribution of hearing losses. More ears were included with normal hearing and mild or moderate hearing losses, compared to the numbers with severe or profound losses. It is reasonable to assume that the overlap in the distributions of responses from normal and impaired ears will be greater when there are a disproportionate number of ears with mild and moderate losses. Less overlap is expected between normal ears and ears with severe or profound hearing loss. Thus, our emphasis on ears with lesser degrees of hearing loss might have negatively impacted estimates of test performance, which depend on the extent of the overlap between normal and impaired distributions. This argument, however, is countered by the fact that when test performance was compared between optimal stimulus conditions and more traditional conditions for moderate level stimuli, no differences were observed (see Fig. 3).

In the present study, moderate-level stimuli ($L_2 = 40$ -55 dB SPL) produce the best test performance, regardless of whether optimal or traditional relationships between L_1 and L_2 or between f_2 and f_1 are used. This finding is consistent with the results reported previously by Whitehead et al. (1995a) and Stover et al. (1996), although the trend is more clearly shown in the present data (at least in comparison to the results reported by Stover et al.) because we were able to achieve stimulus levels as high as 80 dB SPL (in contrast to maximum levels of 65 dB SPL achieved by Stover and colleagues). The reasons for the poor performance at low levels results from the fact that some subjects with normal hearing do not produce responses for low-level stimuli, thus driving up the false-positive rate. The poorer performance for

high-level stimuli results from the fact that some ears with hearing loss (especially ears with mild losses) produce responses to high-level stimuli, thus driving up the false-negative rate. The present data confirm earlier findings, in which it was shown that the best balance between false-positive and false-negative errors occurs for moderate stimulus levels.

It is not clear why the optimal stimulus parameters had the unexpected and undesirable effect of producing smaller DPOAE levels in normal ears than the traditional parameters for $f_2 = 4$ and 5.6 kHz. The stimulus equations developed by Johnson et al. (2006) were based on data collected with $f_2 = 1, 2, 4,$ and 8 kHz for a wide range of L_1, L_2 and f_2/f_1 combinations, using the same in-the-ear calibration of stimulus levels. While $f_2 = 5.6$ was not directly tested in their study, $f_2 = 4$ kHz was included and the Johnson et al. results suggested that larger DPOAE levels should be expected with the new stimulus conditions at this frequency. This expectation is also consistent with results reported by Neely et al. (2005), in which the largest impact of optimizing stimulus level on L_d was observed for 4 and 8 kHz. Thus, the observation of smaller response levels at both 4 and 5.6 kHz are inconsistent with previous data that were obtained under essentially identical conditions. The present data suggest that the optimal stimulus equations did not generalize to a new group of ears when $f_2 = 4$ and 5.6 kHz, although the reasons why this is the case are not apparent.

Threshold Prediction

While the nearly equivalent increase in L_d for both normal and impaired ears when using the optimal stimulus parameters was an undesirable outcome for the dichotomous-decision task, it might have a positive influence on the accuracy with which behavioral threshold is predicted from the DPOAE I/O function. If more ears with hearing loss produce measurable responses when using the optimal stimuli, then it might be possible to predict a behavioral threshold from the I/O function for a larger proportion of ears with hearing loss and perhaps the correlation would be improved. When compared to results obtained with the more traditional stimuli used by others (Gorga et al., 2003; see Table 2 and Fig. 7), the strength of the correlation showed both increases and decreases for the optimal stimulus conditions, and in many cases the changes were small. The optimal stimuli did result in a larger percentage of ears for which the behavioral threshold could be predicted from the I/O function (see Table 2) for all f_2 's except 4 and 5.6 kHz, although, this increase was typically small ($< 10\%$ in most cases). Even with increases in the number of ears meeting all inclusion criteria for the I/O function approach with the optimal stimuli, the simpler SNR approach could be applied to as many as 40-49% more ears than the I/O approach and produced nearly equivalent correlations. This indicates that a similar linear relationship between behavioral threshold and DPOAE threshold exists for the two approaches to predicting behavioral threshold from DPOAE data. However, accuracy of the prediction favors the I/O approach as indicated by the smaller standard errors for I/O approach regardless of stimulus paradigm. Although this outcome is unrelated to the primary focus of the paper (determining the influence of optimal stimulus parameters on DPOAE accuracy), it suggests that including data from the entire I/O function when predicting a behavioral threshold results in a more accurate prediction, albeit one that can be made for a smaller number of ears. Using the optimal stimuli in conjunction with the I/O approach resulted in the smallest standard errors for the majority of f_2 's, although in many cases the differences were only 1 to 2 dB.

Summary—It appears that the optimal stimuli used in this study do not result in an improvement in test performance, as measured by A_{ROC} . The present results are consistent with previous findings, in that they suggest that the use of moderate-level primaries ($L_2 = 50$ or 55 dB SPL), an L_1 at least slightly larger than L_2 (the optimal difference being less important), and $f_2/f_1 = 1.2$ are the primary determining factors affecting the accuracy with

which DPOAE measurements predict auditory status. Furthermore, predictions of behavioral threshold, using either DPOAE SNR or linear fits to the DPOAE I/O function, were either not improved or showed only small (1-2 dB) improvements when using optimal stimuli as compared to other stimulus parameters.

Acknowledgments

Work supported by the NIH NIDCD F32-DC007536, R01-DC02251, P30-DC04662. The authors thank Sandy Estee for assistance with subject recruitment, Connie Converse and Elizabeth Kennedy for assistance with data collection, and Lauren Baranowski for assistance with figure preparation. We would also like to thank two anonymous reviewers for their suggestions regarding this manuscript. Portions of this work were presented at the 2006 American Auditory Society Meeting in Scottsdale, AZ.

References

- ANSI. ANSI 3.6-1996. American National Standards Institute; New York: 1996. Specifications for audiometers.
- Bamber D. The area above the ordinal dominance graph and the area below the receiver operating characteristic graph. *J Math Psychol* 1975;12:387–415.
- Boege P, Janssen T. Pure-tone threshold estimation from extrapolated distortion product otoacoustic emission I/O-functions in normal and cochlear hearing loss ears. *J Acoust Soc Am* 2002;111:1810–1818. [PubMed: 12002865]
- Brown AM, Sheppard SL, Russell PT. Acoustic distortion products (ADP) from the ears of term infants and young adults using low stimulus levels. *Br J Audiol* 1994;28:273–280. [PubMed: 7735156]
- Brown AM, Gaskill SA. Measurement of acoustic distortion reveals underlying similarities between human and rodent mechanical responses. *J Acoust Soc Am* 1990;88:840–849. [PubMed: 2212309]
- Dorn PA, Konrad-Martin D, Neely ST, et al. Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears. *J Acoust Soc Am* 2001;110:3119–3131. [PubMed: 11785813]
- Dorn PA, Piskorski P, Gorga MP, et al. Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses. *Ear Hear* 1999;20:149–163. [PubMed: 10229516]
- Gaskill SA, Brown AM. The behavior of the acoustic distortion product, $2f_1-f_2$, from the human ear and its relation to auditory sensitivity. *J Acoust Soc Am* 1990;88:821–829. [PubMed: 2212308]
- Gorga MP, Dierking DM, Johnson TA, et al. A validation and potential clinical application of multivariate analyses of distortion-product otoacoustic emission data. *Ear Hear* 2005;26:593–607. [PubMed: 16377995]
- Gorga MP, Neely ST, Bergman B, et al. Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses. *J Acoust Soc Am* 1993;93:2050–2060. [PubMed: 8473617]
- Gorga MP, Neely ST, Dorn PA. Distortion product otoacoustic emission test performance for a priori criteria and for multi-frequency audiometric standards. *Ear Hear* 1999;20:345–362. [PubMed: 10466570]
- Gorga MP, Neely ST, Dorn PA, et al. Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions. *J Acoust Soc Am* 2003;113:3275–3284. [PubMed: 12822800]
- Gorga MP, Neely ST, Ohlrich B, et al. 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* 1997;18:440–455. [PubMed: 9416447]
- Gorga MP, Nelson K, Davis T, et al. Distortion product otoacoustic emission test performance when both $2f_1-f_2$ and $2f_2-f_1$ are used to predict auditory status. *J Acoust Soc Am* 2000;107:2128–2135. [PubMed: 10790038]

- Harris FP, Lonsbury-Martin BL, Stagner BB, et al. Acoustic distortion products in humans: Systematic changes in amplitude as a function of f_2/f_1 ratio. *J Acoust Soc Am* 1989;85:220–229. [PubMed: 2921404]
- Hauser R, Probst R. The influence of systematic primary-tone level variation L_2-L_1 on the acoustic distortion product emission $2f_1-f_2$ in normal human ears. *J Acoust Soc Am* 1991;89:280–286. [PubMed: 2002169]
- Janssen T, Kummer P, Arnold W. Growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission in tinnitus. *J Acoust Soc Am* 1998;94:2659–2669.
- Johnson TA, Neely ST, Garner CA, et al. Influence of primary-level and primary-frequency ratio on human distortion product otoacoustic emissions. *J Acoust Soc Am* 2006;119:418–428. [PubMed: 16454296]
- Johnson TA, Neely ST, Kopun JG, et al. Distortion product otoacoustic emissions: Cochlear-source contributions and clinical test performance. *J Acoust Soc Am* 2007;122:3539–3553. [PubMed: 18247762]
- Kim DO, Paparello J, Jung MD, et al. Distortion product otoacoustic emission test of sensorineural hearing loss: Performance regarding sensitivity, specificity, and receiver operating characteristics. *Acta Oto-Laryngol* 1996;116:3–11.
- Kummer P, Janssen T, Arnold W. The level and growth behavior of the $2f_1-f_2$ distortion product otoacoustic emission and its relationship to auditory sensitivity in normal hearing and cochlear hearing loss. *J Acoust Soc Am* 1998;103:3431–3444. [PubMed: 9637030]
- Kummer P, Janssen T, Hulin P, et al. Optimal L_1-L_2 primary tone level separation remains independent of test frequency in humans. *Hear Res* 2000;146:47–56. [PubMed: 10913883]
- Martin GK, Ohlms LA, Franklin DJ, et al. Distortion product emissions in humans III. Influence of sensorineural hearing loss. *Ann Otol Rhinol Laryngol* 1990;99:30–42.
- Musiek FE, Baran JA. Distortion product otoacoustic emissions: Hit and false-positive rates in normal hearing and hearing-impaired subjects. *Am J Otol* 1997;18:454–461. [PubMed: 9233485]
- Neely ST, Gorga MP. Comparison between intensity and pressure as measures of sound level in the ear canal. *J Acoust Soc Am* 1998;104:2925–2934. [PubMed: 9821338]
- Neely, ST.; Liu, Z. Technical Memorandum 17. Boys Town National Research Hospital; Omaha, NE: 1993. EMAV: Otoacoustic emission averager.
- Neely ST, Johnson TA, Gorga MP. Distortion-product otoacoustic emission measured with continuously varying stimulus level. *J Acoust Soc Am* 2005;117:1248–1259. [PubMed: 15807014]
- Nelson DA, Kimberley BP. Distortion-product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss. *J Sp Hear Res* 1992;35:1142–1159.
- Norton SJ, Gorga MP, Widen JE, et al. Identification of neonatal hearing impairment: Evaluation of transient evoked otoacoustic emission, distortion product otoacoustic emission, and auditory brain stem response test performance. *Ear Hear* 2000;21:508–528. [PubMed: 11059707]
- Oswald JA, Janssen T. Weighted DPOAE input/output functions: A tool for automatic assessment of hearing loss in clinical applications. *Z Med Phys* 2003;13:93–98. [PubMed: 12868334]
- Scheperle RA, Neely ST, Kopun JG, et al. Influence of *in situ*, sound-level calibration on distortion-product otoacoustic emission variability. *J Acoust Soc Am* 2008;124:288–300. [PubMed: 18646977]
- Siegel JH. Ear-canal standing waves and high-frequency sound calibration using otoacoustic emission probes. *J Acoust Soc Am* 1994;95:2589–2597.
- Siegel JH, Hirohata ET. Sound calibration and distortion product otoacoustic emissions at high frequencies. *Hear Res* 1994;80:146–152. [PubMed: 7896573]
- Stover LJ, Gorga MP, Neely ST, et al. Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *J Acoust Soc Am* 1996;100:956–967. [PubMed: 8759949]
- Swets JA. Measuring the accuracy of diagnostic systems. *Science* 1988;240:1285–1293. [PubMed: 3287615]
- Swets, JA.; Pickett, RM. Evaluation of Diagnostic Systems: Methods from Signal Detection. New York: Academic; 1982.

- Whitehead ML, McCoy MJ, Lonsbury-Martin BL, et al. Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. I. Effects of decreasing L_2 below L_1 . J Acoust Soc Am 1995a;97:2346–2358. [PubMed: 7714254]
- Whitehead ML, Stagner BB, McCoy MJ, et al. Dependence of distortion-product otoacoustic emissions on primary levels in normal and impaired ears. II. Asymmetry in L_1, L_2 space. J Acoust Soc Am 1995b;97:2359–2377. [PubMed: 7714255]

Figure 1.

ROC curve area (A_{ROC}) as a function of L_2 (dB SPL) for the nine frequencies evaluated by Stover et al. (1996) using traditional stimulus parameters. Octave frequencies from 0.5 to 8 kHz are plotted in the left panel and interoctave frequencies from 0.7 to 5.6 kHz are plotted in the right panel. [Reprinted with permission from Stover, L., Gorga, M.P., Neely, S.T., & Montoya, D. (1996). Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *Journal of the Acoustical Society of America*, 100, 956-967. Copyright 1996, American Institute of Physics.]

Figure 2.

Area under the ROC curve (A_{ROC}) as a function of L_2 for the eight f_2 frequencies tested using the optimal stimulus parameters. The parameter within each panel is f_2 . The symbol plotted at the far right of each panel (and not connected to the other symbols) represents A_{ROC} for the control condition ($f_2/f_1 = 1.22$; $L_1, L_2 = 65, 55$ dB SPL).

Figure 3.

Area under the ROC curve (A_{ROC}) as a function of f_2 when $L_2 = 55$ dB SPL. The parameter is stimulus condition. The open and filled circles represent A_{ROC} values for the present study when L_1 and f_2/f_1 were specified by the optimal stimulus parameters (filled circles) or were set to the traditional settings of $L_1 = 65$ dB SPL and $f_2/f_1 = 1.22$ (open circles).

Figure 4.

The difference in DPOAE level (dB) for the optimal stimulus parameters as compared the traditional stimulus parameters when $L_2 = 55$ dB SPL plotted as a function of f_2 . The parameter is the ear category (normal vs. impaired). The dotted line represents the point where both types of stimuli produce equal DPOAE levels.

Figure 5.

Behavioral threshold (dB HL) as a function of DPOAE threshold (dB SPL) for the SNR approach to estimating DPOAE threshold. Each panel represents a different f_2 , as indicated on the figure. The solid line in each panel is the best-fit line for the data, with the equation for the best-fit line shown in the bottom right corner of each panel.

Figure 6.

Behavioral threshold (dB HL) as a function of DPOAE threshold (dB SPL) for the I/O approach to estimating DPOAE threshold. As in Fig. 5, each panel represents a different f_2 . Shown within each panel is the best-fit line (solid line) along with its equation (bottom right corner).

Figure 7.

Upper panel (A): Difference in the strength of the correlation between DPOAE and behavioral thresholds for the I/O approach to estimating DPOAE threshold using optimal stimuli as compared to the stimuli used by Gorga et al. (2003). The dashed line indicates the point at which the difference in the correlation for the two stimulus approaches is 0. Lower panel (B): Differences in the percentage of ears for which a behavioral threshold could be predicted from the DPOAE I/O function when using either the optimal or the Gorga et al. stimuli. As in A, the dashed line indicates the point where the percentage difference is 0.

TABLE 1

Number of ears tested in each hearing-loss category.

Category	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	8000 Hz
Normal (≤ 20 dB HL)	248	215	187	140	115	116
Mild (25-40 dB HL)	43	66	68	69	50	30
Moderate (45-60 dB HL)	19	23	36	64	87	79
Severe (65-85 dB HL)	10	10	21	33	45	65
Profound (≥ 90 dB HL)	2	8	10	16	25	32

TABLE 2

Summary of the results for the two approaches to threshold prediction (SNR approach and I/O approach). Results reported by Gorga et al. (2003) are also shown for comparison purposes. In each case, the percentage of ears meeting the inclusion criteria is shown along with the correlation coefficients (r) and associated standard errors for each of the eight frequencies (see Figs 5 and 6).

f_2 (Hz)	% of ears meeting inclusion criteria				Correlation				Standard error (dB)			
	SNR Approach	I/O Approach	Gorga et al. (2003)		SNR Approach	I/O Approach	Gorga et al. (2003)		SNR Approach	I/O Approach	Gorga et al. (2003)	
707	57	36	30		0.65	0.67	0.49		11.64	7.29		13.8
1000	72	44	33		0.59	0.64	0.66		13.89	9.09		11.6
1414	91	42	36		0.74	0.74	0.68		13.70	9.15		11.2
2000	81	44	40		0.79	0.67	0.74		14.02	11.61		10.6
2828	78	47	38		0.85	0.86	0.74		13.03	9.26		12.5
4000	94	54	55		0.82	0.87	0.85		18.15	12.06		11.2
5656	79	40	44		0.78	0.71	0.74		19.32	15.15		16.3
8000	85	39	30		0.73	0.72	0.76		21.59	19.61		19.2