Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions

Michael P. Gorga,[a] Stephen T. Neely, Patricia A. Dorn, and Brenda M. Hoover
Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

(Received 28 October 2002; revised 5 March 2003; accepted 6 March 2003)

Recently, Boege and Janssen [J. Acoust. Soc. Am. **111**, 1810–1818 (2002)] fit linear equations to distortion product otoacoustic emission (DPOAE) input/output (I/O) functions after the DPOAE level (in dB SPL) was converted into pressure (in μPa). Significant correlations were observed between these DPOAE thresholds and audiometric thresholds. The present study extends their work by (1) evaluating the effect of frequency, (2) determining the behavioral thresholds in those conditions that did not meet inclusion criteria, and (3) including a wider range of stimulus levels. DPOAE I/O functions were measured in as many as 278 ears of subjects with normal and impaired hearing. Nine f_2 frequencies (500 to 8000 Hz in $\frac{1}{2}$-octave steps) were used, L_2 ranged from 10 to 85 dB SPL (5-dB steps), and L_1 was set according to the equation $L_1 = 0.4L_2 + 39$ dB [Kummer *et al.*, J. Acoust. Soc. Am. **103**, 3431–3444 (1998)] for L_2 levels up to 65 dB SPL, beyond which $L_1 = L_2$. For the same conditions as those used by Boege and Janssen, we observed a frequency effect such that correlations were higher for mid-frequency threshold comparisons. In addition, a larger proportion of conditions not meeting inclusion criteria at mid and high frequencies had hearing losses exceeding 30 dB HL, compared to lower frequencies. These results suggest that DPOAE I/O functions can be used to predict audiometric thresholds with greater accuracy at mid and high frequencies, but only when certain inclusion criteria are met. When the SNR inclusion criterion is not met, the expected amount of hearing loss increases. Increasing the range of input levels from 20–65 dB SPL to 10–85 dB SPL increased the number of functions meeting inclusion criteria and increased the overall correlation between DPOAE and behavioral thresholds. [DOI: 10.1121/1.1570433]

PACS numbers: 43.64.Ha, 43.64.Jb [BLM]

I. INTRODUCTION

The auditory system behaves nonlinearly under normal conditions. Evidence of this nonlinearity can be found in a number of different phenomena, including the production of distortion product otoacoustic emissions (DPOAEs). DPOAEs are observed when two sounds (f_1 and f_2, f_2 slightly higher in frequency than f_1) are presented to the normal ear. These sounds interact in the cochlea at a place close to the best place for the higher of the two frequencies (f_2), producing intermodulation distortion, the largest component of which occurs at a frequency equal to $2f_1$-f_2. It is thought that these nonlinear phenomena are produced by forces exerted by the outer hair cells (OHCs) on basilar membrane mechanical responses (e.g., Brownell, 1990). When cochlear damage exists that affects the OHCs, thresholds are elevated and nonlinear behaviors are reduced or eliminated (e.g., Dallos *et al.*, 1980). Therefore, it is not surprising that DPOAEs are reduced or eliminated by OHC damage as well. These observations have led to the application of DPOAE measurements in efforts to determine auditory status (e.g., Martin *et al.*, 1990; Gorga *et al.*, 1993, 1997, 2000; Kim *et al.*, 1996). With few exceptions, DPOAE measurements have been used to make dichotomous decisions, in which an ear is classified as having either normal hearing or hearing loss. The results from these studies are similar in that DPOAE measurements classify ears dichotomously with greater accuracy for mid and high frequencies, compared to the accuracy that is achieved at lower frequencies.

Several studies have attempted to go beyond this simple, two-state classification scheme and predict auditory thresholds from DPOAE measurements. For example, Martin *et al.* (1990) and Gorga *et al.* (1996) related DPOAE threshold to audiometric threshold. While both studies showed a relationship between the two threshold measurements, these measurements have not been applied clinically, presumably because estimates of DPOAE threshold require several measurements above and below threshold, and the conditions under which these measurements would be made are characterized by poor signal-to-noise ratios (SNR) because DPOAE level is small. Thus, the reliability of direct DPOAE threshold measurements is reduced by the increased uncertainty in response measurements at threshold.

In other attempts to estimate pure-tone thresholds from DPOAE data, DPOAE level or SNR for suprathreshold eliciting stimuli have been correlated with behavioral thresholds for ears with normal hearing (Allen and Levitt, 1992; Dorn *et al.*, 1998) and for ears in which hearing loss existed (Martin *et al.*, 1990; Gorga *et al.*, 1996, 1997, 2002; Janssen *et al.*, 1998; Kummer *et al.*, 1998). Typically, DPOAE level was measured for fixed, moderate-level primaries that produced responses that were well above DPOAE threshold, at least for ears with normal hearing. These DPOAE levels and/or the SNR were then correlated with audiometric thresholds. While DPOAE level or SNR decreased as audio-

[a]Electronic mail: gorga@boystown.org

metric thresholds increased, the relationship was variable, thus reducing the accuracy with which predictions of behavioral thresholds could be made from DPOAE level. Efforts to include several variables in a prediction model did not result in significant improvements in predictive accuracy (Kimberley *et al.*, 1994, 1997).

More recently (Boege and Janssen, 2002; Oswald *et al.*, 2002), DPOAE input/output (I/O) functions have been used to provide estimates of DPOAE thresholds, which were then correlated with behavioral thresholds. In this approach, DPOAE level (in dB SPL) was measured for levels ranging from 20 to 65 dB SPL, with primary levels chosen to optimize response level (Kummer *et al.*, 1998; Janssen *et al.*, 1998). If an SNR inclusion criterion (6 dB) was met for at least three points on the I/O function, these data were converted into pressure (in μPa) and fit with a linear equation. If the linear solutions met some additional inclusion criteria (related to slope of the best fit line and variability), the data were used to determine DPOAE threshold (defined as the extrapolated DPOAE level at which the pressure equaled 0 μPa). Significant correlations were observed between pure-tone and DPOAE thresholds, although estimates of DPOAE threshold were not possible in a percentage of cases. Oswald *et al.* (2002) noted larger discrepancies between DPOAE and behavioral thresholds at 2000 and 8000 Hz, compared to other frequencies, although the effects of frequency were not described in detail. They also observed a tendency to underestimate behavioral threshold by increasing amounts as behavioral threshold increased.

The present study was designed to replicate and extend the work of Boege and Janssen (2002) and Oswald *et al.* (2002). Specifically, DPOAE I/O functions were analyzed using the same stimulus conditions and inclusion criteria as those used in these previous studies. Following a replication of the results observed by Boege and Janssen, the present study extended the previously reported findings by evaluating DPOAE predictions of pure-tone thresholds as a function of frequency. In addition, the audiometric thresholds of those ears not meeting inclusion criteria were evaluated in order to provide information regarding the conditions under which the Boege and Janssen approach did not fit. Finally, additional stimulus conditions and inclusion criteria were tested in order to determine if predictive accuracy could be improved.

II. METHODS

A. Subjects

Ninety-seven subjects with normal hearing and 130 subjects with hearing loss participated in these studies. Depending on frequency and level, DPOAE and audiometric data were available on as many as 278 ears from these subjects. All subjects had normal middle-ear function at the time of the DPOAE measurements, as determined by tympanometry. For the subjects with hearing loss, the site of lesion was assumed to be the cochlea, based on clinical history and other special audiological tests, including measures of speech reception, acoustic reflex thresholds, and auditory brainstem responses. However, specific etiology was frequently undetermined, which is not an uncommon occurrence for clinical studies.

B. DPOAE stimuli

Custom-designed software was used for data collection (EMAV, Neely and Liu, 1994). DPOAE stimuli were produced and responses were recorded by a high-quality soundcard (CardDeluxe, Digital Audio Labs) housed in a PC. The sampling rate was 32 kHz and the sample resolution was 24 bits. Separate channels of the soundcard were used to produce each of the two primary tones. The outputs from each channel were fed to separate loudspeakers housed in a probe-microphone system (Etymotic ER-10C), which had been modified to remove 20 dB of attenuation. A microphone housed in the same probe unit was used to calibrate stimulus level and to record ear-canal responses and noise. While this calibration procedure may introduce errors associated with standing waves (Siegel, 1994, 2002), it represents common practice for DPOAE measurements and was viewed as a reasonable compromise, given the current status of calibration procedures (Neely and Gorga, 1998).

DPOAEs were elicited in response to pairs of primary tones (f_1,f_2), with f_2/f_1 approximately equal to 1.22 for all test conditions, which differs slightly from the 1.20 frequency ratio used by Boege and Janssen (2002). The higher frequency in each primary pair (f_2) varied from 500 to 8000 Hz in $\frac{1}{2}$-octave steps. These frequencies were chosen, in part, because they correspond to the frequencies at which pure-tone audiometric thresholds were measured in the clinic. Given the sampling rate (32 kHz), the duration of the buffer (64 ms), and the number of points in each sample (2048), the frequency resolution was 15.6 Hz. The f_2 frequency was set exactly at the octave and interoctave frequencies described above. The frequency of f_1, however, did not always exactly equal the frequency defined by a primary ratio of 1.22, but was never more than 15.6 Hz away from that frequency. Primary levels (L_1,L_2) were set according to the equation, $L_1 = 0.4\ L_2 + 39$ dB (Kummer *et al.*, 1998; Janssen *et al.*, 1998) for L_2 levels up to 65 dB SPL. Beyond this level, equal-level primaries were used ($L_1 = L_2$). L_2 varied in 5-dB steps from 10 to a maximum of 85 dB SPL. Measurements in four different cavities, validated against measurements in subjects with cochlear implants, were used to determine the level of system distortion [see Dorn *et al.* (2001) for a more complete description of the approach that was taken to estimate system distortion].

C. Audiometric procedures

Pure-tone audiometric data were collected using routine clinical techniques. All measurements were made in a sound-treated booth, using either TDH39 supra-aural earphones or ER-3A insert earphones. Both earphones were calibrated according to their respective standards (ANSI, 1996). Audiometric test frequencies varied from 500 to 8000 Hz in $\frac{1}{2}$-octave steps. The manner in which audiometric stimuli were calibrated and the conditions under which audiometric thresholds were measured represent a potentially important distinction between the present results and those reported by Boege and Janssen (2002), who used the same system and approach to calibration for both DPOAE and audiometric measurements. Furthermore, they measured behavioral thresholds with 1-dB precision, in contrast to the present measurements, in which 5-dB steps were used.

D. DPOAE procedures

DPOAE data were collected in the form of I/O functions. For each f_2 frequency, stimulus level was initiated at the highest level ($L_2 = 85$ dB SPL). L_2 was then decreased in 5-dB steps until the response was no longer measurable above the noise floor (or the level at which system distortion occurred). Both DPOAE and noise level were measured in the same frequency bin, using a subtraction technique, in which data were collected into two separate buffers, alternating between the two on successive samples. The contents of the two buffers were summed in order to provide an estimate of DPOAE level. Their contents were subtracted in order to provide an estimate of the noise level. Both signal and noise were estimated at the $2f_1$-f_2 distortion frequency. This approach has the advantage of estimating signal and noise levels at the same frequency. However, it has the disadvantage of introducing greater variability in noise estimates, compared to paradigms in which noise is derived from the average level in several frequency bins on either side of $2f_1$-f_2.

Measurement-based stopping rules were used during data collection. Data collection stopped if either of two criteria were met. The first stopping criterion was met when the noise floor was less than -25 dB SPL for L_2 levels of 65 dB SPL or less. For higher-level primaries, the test stopped once the "noise level" was less than the level at which system distortion occurred (Dorn *et al.*, 2001). This level increased with primary levels above 65 dB SPL. In both cases, the aim was to measure DPOAEs over the widest range possible without running the risk of misinterpreting system distortion for biological distortion. The second stopping criterion was met when 32 s of artifact-free averaging time had been devoted to that condition, even if the noise-floor criterion was not met. This stopping criterion was necessary to avoid prolonged averaging times for any one condition, as there were many conditions for each subject.

E. Inclusion criteria

Boege and Janssen (2002) set four inclusion criteria for their analyses, one of which related to the reliability of the DPOAE measurements, the other three of which related to the characteristics of linear fits to the DPOAE I/O functions. First, there needed to be at least three points on the DPOAE I/O function with SNR$\geqslant$6 dB. If this criterion was not met, that function was not included in further analyses. If the SNR criterion was met, then DPOAE levels (dB SPL) were converted into pressure (μPa), and the data for each I/O function were fit with a linear equation. Data were included in the next level of analyses only if the slopes of the individual linear regressions were $\geqslant$0.2 μPa/dB,[1] the variance accounted for (r^2) was $\geqslant$0.8, and the standard error was $\leqslant$10 dB. If these criteria were met, the equations were solved for

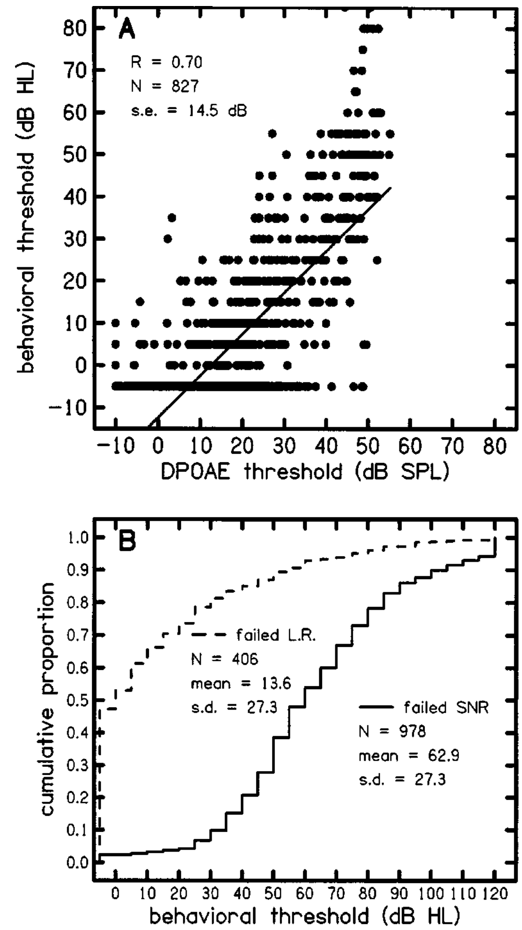

FIG. 1. (a) Behavioral threshold (dB HL) as a function of predicted DPOAE threshold (dB SPL). DPOAE thresholds were predicted from linear fits to individual DPOAE I/O functions (see text for details). The solid line in the figure represents the best-fit line to the behavioral and DPOAE thresholds. Also shown is the correlation coefficient, the number of threshold comparisons (i.e., the number of I/O functions meeting all inclusion criteria), and the standard error. (b) Cumulative proportions of the number of conditions that failed to meet the SNR criterion (solid line) or the inclusion criteria associated with the linear regressions (dashed line) as a function of behavioral threshold (dB HL). See text for details regarding the inclusion criteria. Also shown in this panel are the number of conditions represented on each distribution, the mean behavioral thresholds, and the standard deviations.

the DPOAE stimlus level (in dB SPL) at which the DPOAE amplitude equaled 0 μPa. This stimulus level was defined as DPOAE threshold, which was then correlated with behavioral thresholds. If these criteria associated with the linear fits to the DPOAE I/O functions were not met, the data were excluded from further analyses. We chose the same inclusion criteria in our first level of analyses in order to obtain results that were comparable to those observed by Boege and Janssen (2002). In additional analyses, we varied these inclusion criteria to determine if more accurate predictions of threshold could be achieved.

III. RESULTS

A. Comparison to Boege and Janssen (2002)

Figure 1(a) plots behavioral pure-tone thresholds (in dB HL) as a function of estimated DPOAE thresholds; recall that DPOAE thresholds (in dB SPL) were defined as the extrapolated stimulus level for which the DPOAE response

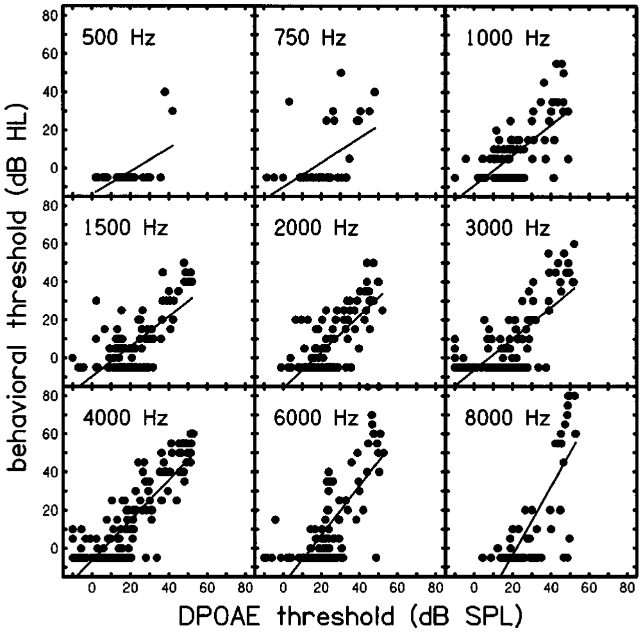

FIG. 2. Behavioral threshold (dB HL) as a function of DPOAE threshold (dB SPL) for each of nine frequencies. Solid lines represent the best-fit line to the data in each panel. See Table I for information regarding correlation coefficients, standard errors, and the number of observations in each panel.

was 0 μPa. For this analysis, primary levels (L_2) were restricted to the range from 20 to 65 dB SPL, which is the same range used by Boege and Janssen (2002). The data included in this figure were collapsed across all nine frequencies, and include only those cases meeting the inclusion criteria previously selected by Boege and Janssen and described above. Thus, these data were derived from DPOAE I/O functions in which at least three points had SNRs$\geqslant$6 dB, and for which the linear fits after converting DPOAE levels to pressure had slopes $\geqslant$0.2 μPa/dB, $r^2 \geqslant 0.8$, and standard errors $\leqslant$10 dB. After applying all of these inclusion criteria, only 827 of 2211 DPOAE I/O functions (37.4%) were selected for further analyses. This percentage is less than the percentage of I/O functions that met criteria in the Boege and Janssen study. One reason for this difference might relate to the distribution of thresholds in the present study, which extended to greater hearing losses, compared to the subjects in the Boege and Janssen study. It is likely that a high proportion of the cases with greater losses did not meet inclusion criteria, especially the SNR criterion. A linear fit to these data is shown as the solid line in Fig. 1(a). The correlation coefficient (r) for these data was 0.70, which is similar to the correlation previously observed by Boege and Janssen ($r = 0.65$). Thus, it would appear that, when analyzed in the same way, the present results are similar to those obtained by Boege and Janssen (2002).

Figure 1(b) plots the cumulative proportions of conditions in which either the SNR criterion was not met (solid line) or, after meeting the SNR criterion, the inclusion criteria associated with the linear regressions were not met (dashed line). These plots are noteworthy for several reasons. Of the total sample of 2211 DPOAE I/O functions, 978 (44.2%) failed to meet the SNR inclusion criterion (i.e., at least three points on the I/O function with a SNR of at least 6 dB). Of these 978 I/O functions, 90% (880) had behavioral thresholds exceeding 30 dB HL, with a mean threshold for these conditions of 62.9 dB HL (SD = 27.3 dB). Thus, it was highly likely that an ear failing the SNR inclusion criterion had hearing loss. An additional 406 DPOAE I/O functions (18.4% of the total sample of 2211) failed to meet the inclusion criteria associated with the linear regressions (slope, correlation coefficient, standard error). Behavioral thresholds exceeded 30 dB HL in only 19% of these cases (77 out of 406 I/O functions not meeting the linear regression criteria). This means that the linear regression inclusion criteria associated with these fits were not met in 329 cases in which normal or near-normal hearing existed. This number represents 14.9% of the total number of I/O functions that were available for analyses. In order to provide a number that shares some characteristics with the false-positive rate, these ears could be added to the number that failed the SNR criterion but had thresholds $\leqslant$30 dB HL (98). Thus, 427 conditions with normal hearing failed to meet criteria; if these cases can be viewed as false positives, then this translates into a false-positive rate of 19.3%.

B. Frequency effects

Figure 2 plots behavioral thresholds (in dB HL) as a function of the predicted DPOAE threshold (in dB SPL) when the data were separated by frequency. Each panel shows the data for a different frequency, going from 500 Hz (upper left panel) to 8000 Hz (lower right panel). In all other respects, the convention followed in Fig. 1(a) is followed here. As a general rule, there were fewer DPOAE data available (and, therefore, fewer opportunities for comparisons between behavioral thresholds and DPOAE thresholds) at lower frequencies. The reduced numbers at low frequencies is a direct result of the fact that noise levels increased as frequency decreased during DPOAE measurements. Thus, fewer DPOAE I/O functions were available that had three points meeting the SNR criterion for lower f_2 frequencies. As a consequence, data are sparse in some panels in Fig. 2 (in particular, those associated with 500 and 750 Hz). Because of this, caution should be exercised when interpreting data in these cases. Reliable predictions are not possible, given the paucity of data at 500 and 750 Hz. At other frequencies, however, a relationship exists between behavioral and DPOAE threshold estimates.

Table I provides a summary that includes the number of observations meeting inclusion criteria for each of the nine test frequencies, along with the slopes, correlations, intercepts, and the standard errors for the linear regressions of behavioral threshold onto DPOAE threshold. In addition to the increase in the number of conditions for which data were available, the correlation between behavioral thresholds and predicted DPOAE thresholds increased as frequency increased. With the exceptions of 500, 6000, and 8000 Hz, the standard errors were relatively constant as a function of frequency. The small standard error at 500 Hz likely results from the fact that, regardless of DPOAE threshold, the vast majority of the cases meeting inclusion criteria had normal behavioral thresholds at this frequency. The reasons for the higher standard errors at 6000 and 8000 Hz are less obvious.

TABLE I. The number of conditions meeting all inclusion criteria, along with the slopes, correlations, and standard errors when behavioral thresholds were predicted from DPOAE thresholds at each of nine frequencies (see Fig. 2).

Frequency (Hz)	No. of conditions meeting inclusion criteria	Slope	Intercept	Correlation	Standard error (dB)
500	27	0.6	−14	0.57	9.0
750	48	0.7	−10	0.49	13.8
1000	88	0.8	−9	0.66	11.6
1500	98	0.8	−10	0.68	11.2
2000	110	1.0	−16	0.74	10.6
3000	103	0.8	−6	0.74	12.5
4000	149	1.1	−6	0.85	11.2
6000	118	1.3	−19	0.74	16.3
8000	81	1.8	−39	0.76	19.2

TABLE II. The number of conditions failing to meet SNR criterion, the percentage of conditions for which behavioral thresholds were less than or equal to 30 dB HL, and the mean and standard deviations for these conditions at each of nine frequencies (see solid lines, Fig. 3).

Frequency (Hz)	No. of conditions failing SNR criterion	% of thresholds ≤30 dB HL	Mean thresholds	Standard deviation
500	88	33	44.2	31.3
750	73	12	57.7	26.3
1000	120	16	55.3	26.0
1500	122	7.5	59.4	24.6
2000	124	7.5	63.1	25.1
3000	122	5	67.5	23.8
4000	104	0	74.2	22.5
6000	120	3	73.3	25.0
8000	111	11	65.6	29.0

At 8000 Hz, this effect could be the result of the paucity of observations in which behavioral thresholds were between 20 and 40 dB HL. In addition, the larger standard errors at 6000 and 8000 Hz might have been due to standing-wave problems during DPOAE measurements, although one might predict that these problems would be more likely at 4000 Hz (Siegel, 1994, 2002), a frequency at which performance was particularly good. In general, the correlations increase as frequency increases, achieving a maximum of 0.85 at 4000 Hz, decreasing slightly at 6000 and 8000 Hz.

Following the convention used in Fig. 1(b), Fig. 3 plots the cumulative proportions for those cases failing to meet the SNR criterion and the inclusion criteria associated with the results of the linear regressions. Table II provides a summary of the number of conditions for which the SNR criterion was not met, the percentage of cases in which inclusion criteria were not met *and* threshold ≤30 dB HL, and the mean thresholds (and standard deviations) for these conditions. The trends in Fig. 3 and Table II are similar to those observed in Fig. 1(b), when all of the data were combined. The majority of ears that failed to produce at least three points on the DPOAE I/O function for which the SNR≥6 dB had behavioral pure-tone thresholds exceeding 30 dB HL. A smaller number of cases were not included because they failed to meet the inclusion criteria associated with the linear fits to individual I/O functions. Thus, hearing loss existed in the majority of conditions in which we could not reliably measure DPOAEs above the noise floor for at least three points on an I/O function.

This point is perhaps made more clearly in Fig. 4, in which the cumulative distributions describing the conditions in which the SNR criterion was not met are superimposed in one plot for the octave frequencies from 500 to 4000 Hz. These are the same distributions that were shown in the appropriate panels of Fig. 3. Note that there is systematic shift of these distributions towards higher behavioral thresholds as frequency increases. Thirty-three percent of the cases in which the SNR was not met at 500 Hz had thresholds better than 30 dB HL, reflecting the difficulty in making DPOAE measurements at such low frequencies. At 1000, 2000, and 4000 Hz, this number was 16%, 7.5%, and 0%, respectively.

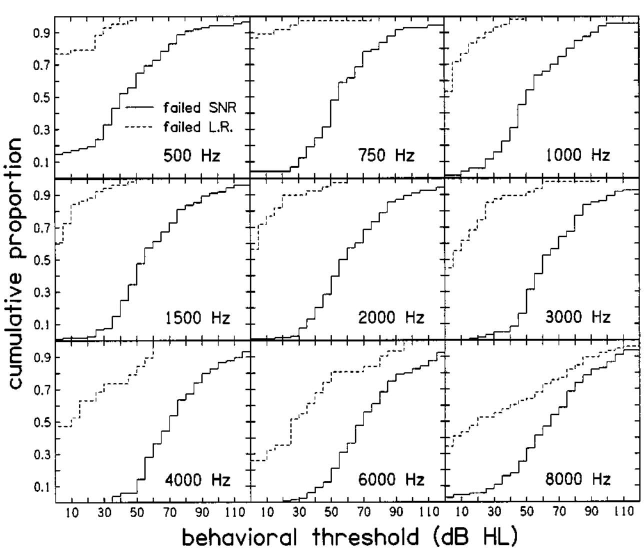

FIG. 3. Cumulative proportions of the number of conditions that failed to meet the SNR criterion (solid lines) or the inclusion criteria related to the linear regressions of individual DPOAE I/O functions (dashed lines) as a function of behavioral threshold (in dB HL). See Table II for a listing of the number of conditions represented on each function, along with the mean behavioral thresholds and their accompanying standard deviations.

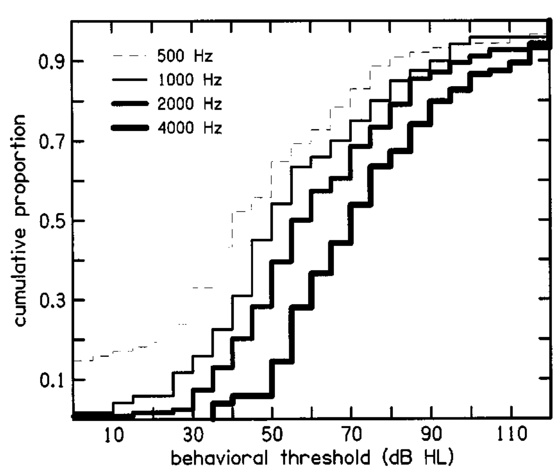

FIG. 4. Cumulative proportions of the behavioral thresholds for those conditions in which the SNR criterion (three points on the I/O function with SNRs≥6 dB) was not met. These functions are the same data as was shown in the appropriate panels of Fig. 3, only here they are superimposed for the four octave frequencies from 500 to 4000 Hz.

TABLE III. The number of conditions failing to meet criteria associated with linear regressions (LR) for individual DPOAE I/O functions, the percentage of conditions for which behavioral thresholds were less than or equal to 30 dB HL, and the mean and standard deviations for these conditions at each of nine frequencies (see dashed lines, Fig. 3).

Frequency (Hz)	No. of conditions failing LR criteria	% of thresholds ≤30 dB HL	Mean thresholds	Standard deviation
500	43	93.0	3.1	15.7
750	37	97.0	−0.3	14.9
1000	60	93.5	5.2	13.8
1500	51	94.0	3.9	13.4
2000	39	90.0	6.2	16.3
3000	47	87.0	12.4	21.9
4000	19	74.0	17.1	24.6
6000	31	55.0	32.1	31.3
8000	78	55.0	34.6	39.1

Thus, the proportion of time the SNR criterion was not met in normal ears decreased as frequency increased. At 4000 Hz, every condition failing to meet the SNR criterion had thresholds exceeding 30 dB HL. As would be expected from these distributions, there also was a systematic, monotonic increase in mean thresholds as frequency increased, going from 44.2 dB HL at 500 Hz to 74.2 dB HL at 4000 Hz (see Table II).

In comparison to the number of times the SNR criterion was not met, a smaller number of conditions did not meet the inclusion criteria associated with the linear regressions, and the majority of this subset of cases had normal hearing. Table III provides a summary of the number of times this occurred, the percentage of those cases with thresholds ≤30 dB HL, and the mean and standard deviations of the behavioral thresholds for these distributions at each of nine frequencies. In contrast to the behavioral thresholds for those DPOAE I/O functions not meeting the SNR criterion, the majority of the subset of cases not meeting the inclusion criteria associated with the linear regressions had normal hearing.

C. Extended analyses

Several additional analyses were performed in an effort to (1) improve upon the accuracy with which the data from DPOAE I/O functions predicted behavioral thresholds and (2) increase the number of conditions in which the inclusion criteria were met. In the two analyses described below, data were collapsed across the nine test frequencies. In this respect, these analyses are similar to the approach taken in Fig. 1. Using the same stimulus conditions and inclusion criteria that were used by Boege and Janssen (2002), data were reanalyzed with three added constraints: (1) any DPOAE threshold predictions that were less than 20 dB SPL were arbitrarily set to 20 dB SPL, (2) any behavioral thresholds less than 0 dB HL were arbitrarily set to 0 dB HL, and (3) any behavioral thresholds exceeding 60 dB HL were similarly set to 60 dB HL. The DPOAE threshold limit was based on the view that DPOAE thresholds lower than this would be difficult to measure because of problems associated with the noise floor. The lower behavioral threshold limit was based on the view that it is uncommon to measure behavioral thresholds less than 0 dB HL. The upper limit on behavioral thresholds was based on the hypothesis that DPOAEs are produced by OHCs, and that complete loss of OHCs (with completely normal inner hair cells) will produce no more than about 60 dB of hearing loss. The combined effect of these three constraints was to restrict the threshold predictions to the range where the relation between DPOAE thresholds and behavioral thresholds appears to be most linear. A scatter plot depicting this analysis is shown in Fig. 5(a). This figure follows the convention that was used in Fig. 1(a), with behavioral thresholds plotted as a function of predicted DPOAE thresholds in Fig. 5(a). Cumulative distributions for the cases not meeting either the SNR criterion or the inclusion criteria based on linear regressions are shown in Fig. 5(b). As expected, there was no change in the number of conditions that met the inclusion criteria, as these criteria were not changed in this analysis. Thus, the percentage of ears with normal hearing failing to meet criteria are identical to those described in association with Fig. 1(b). However, the

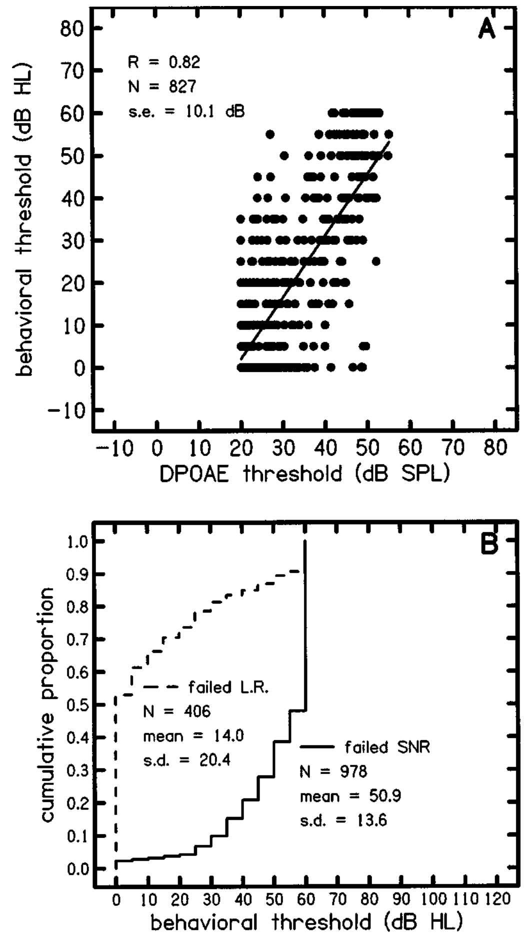

FIG. 5. (a) Behavioral threshold (dB HL) as a function of predicted DPOAE threshold (dB SPL). The analysis here used the same inclusion criteria as was used in the analysis shown in Fig. 1. However, any DPOAE threshold prediction less than 20 dB SPL was arbitrarily set to 20 dB SPL, any behavioral thresholds less than 0 dB HL were set to 0 dB HL, and any behavioral thresholds exceeding 60 dB HL were set to 60 dB HL. (b) Cumulative proportions for the conditions failing to meet the SNR criterion (solid line) or the inclusion criteria associated with the linear regressions (dashed line). Also shown are the number of conditions represented on each distribution, the mean behavioral thresholds and their standard deviations.

correlation increased to 0.82 and the standard error decreased to 10.1, both of which represent improvements over what was achieved in the analysis summarized in the top panel of Fig. 1, where there were no constraints on either behavioral or predicted DPOAE thresholds.

In the final analyses, behavioral thresholds and predicted DPOAE thresholds were restricted to the same range that was used for the analysis described in Fig. 5. However, the primary levels (L_2) were expanded to include the range from 10 to 85 dB SPL. This rule was modified so that, for any individual DPOAE I/O function, fits were performed for a maximum range of primary levels of 40 dB, starting at the lowest L_2 level that achieved a 10-dB SNR. The SNR criterion was increased to 10 dB, in part to assure that the DPOAE levels measured for high-level stimuli were not affected by system distortion. The increase in SNR criterion from 6 to 10 dB and the inclusion of higher primary levels might have resulted in a less frequency-specific response, compared to conditions in which the SNR=6 dB and primary levels were restricted to 65 dB SPL or less. However, as will be seen below, the correlation was unaffected by these changes. Finally, the criteria based on the linear regressions were altered, such that the slope had to be at least 0.1 μPa/dB, the correlation coefficient had to be $\geqslant$0.7, and the standard error had to be $\leqslant$9 dB. In total, these new criteria were selected such that more I/O functions would meet inclusion criteria while the correlation would remain the same or increase. Other criteria also were evaluated (such as returning to the SNR criterion of 6 dB). While other criteria resulted in the inclusion of more I/O functions, they did so at the expense of a reduction in the correlation.

The results of this analysis are summarized in Fig. 6, following the convention used in Figs. 1 and 5. This analysis resulted in the highest correlation (0.83) of any of the three analyses that evaluated the data collapsed across frequency, while the standard error for this condition (10.7) was slightly higher than the standard error observed in one of the previous two analyses (10.1). However, inclusion criteria were met for more than 100 additional conditions in this case. Thus, the inclusion of a wider range of stimulus levels and a slight alteration of the inclusion criteria resulted in an increase in the number of conditions in which the approach proposed by Boege and Janssen (2002) could be applied. Any loss in frequency specificity (due to the increased primary levels and SNR criterion) did not negatively impact the correlation.

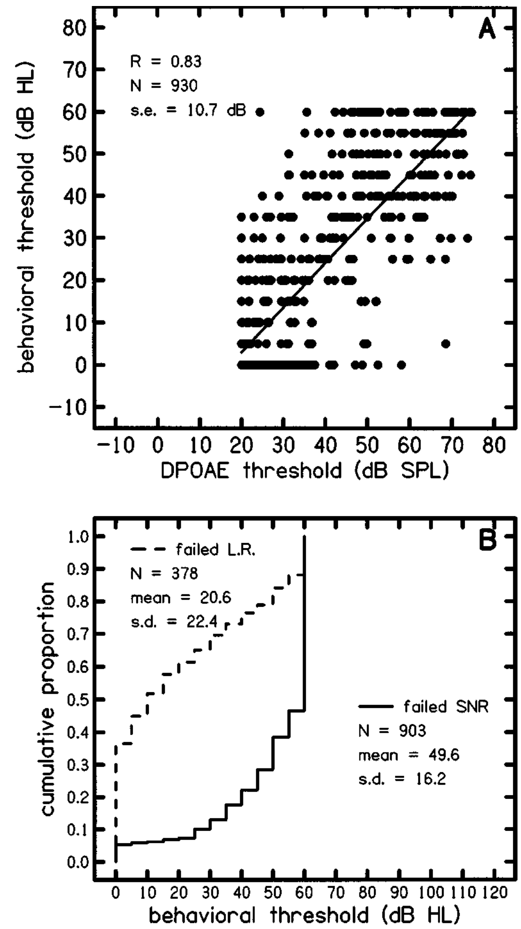

FIG. 6. (a) Behavioral threshold (dB HL) as a function of predicted DPOAE threshold (dB SPL). The range of stimulus levels was increased from 20–65 dB SPL to 10–85 dB SPL; however, only a 40-dB range of stimulus levels was used when fitting linear equations to individual DPOAE I/O functions. In addition, the SNR criterion was changed to 10 dB, and the criteria associated with the linear regressions were relaxed (see text for details). Both DPOAE threshold prediction and behavioral thresholds were restricted in the same way as they were restricted for the analyses shown in Fig. 5. (b) Cumulative proportions for the conditions failing to meet the SNR criterion (solid line) or the inclusion criteria associated with the linear regressions (dashed line). Also shown are the number of conditions represented on each distribution, the mean behavioral thresholds, and their standard deviations.

IV. DISCUSSION

To summarize the results of the first part of this study, the results reported by Boege and Janssen (2002) were essentially replicated, using the same stimulus conditions and inclusion criteria that were used by them. When evaluating the results collapsed across frequency, similar results were observed for the present data compared to the previously reported correlations between behavioral thresholds and predicted DPOAE thresholds, based on linear regressions of DPOAE amplitude (μPa) onto DPOAE stimulus level (dB SPL). The previous results were extended in a number of different ways. First, we evaluated the auditory thresholds among those ears that failed to meet inclusion criteria. The majority of the cases failing to meet a SNR criterion had hearing loss. A smaller number of ears failing to meet inclusion criteria associated with the linear regressions of individual DPOAE I/O functions also had hearing loss. In another extension of the previous work, we evaluated the effect of frequency on the accuracy with which audiometric thresholds could be estimated. The best performance was observed at 4000 Hz. Best performance was defined as the frequency for which the correlation between threshold estimates was highest, the standard error was the lowest, and the percentage of ears not meeting inclusion criteria that also had hearing loss was highest. At low frequencies, performance was particularly poor, but poor performance also was observed at 8000 Hz. Finally, inclusion criteria were adjusted in efforts to determine if better test performance could be achieved with a different set of rules than those proposed by Boege and Janssen (2002). Taking measurements for a wider range of levels and slightly altering the inclusion criteria resulted in an improvement in test performance. That is, a larger number of conditions met inclusion criteria and a higher correlation was observed between behavioral thresholds and predicted DPOAE thresholds.

It may be important to note that the present results could be specific to the conditions of these measurements, especially the way in which L_1 varied in relation to L_2 (Kummer *et al.*, 1998). It is unclear whether similar results would be observed if a constant primary-level difference of 0, 10, or 15 dB were used. The present data do not allow us to address this issue. In addition, audiometric threshold (in dB HL) was predicted in the present study, whereas Boege and Janssen (2002) predicted behavioral thresholds in dB SPL. Audiometric thresholds were chosen in the present study because of their use in clinical assessments. In addition, converting the present audiometric thresholds to dB SPL would have no influence on the correlations for individual frequencies, since the conversion would amount to adding the same constant to audiometric thresholds in each panel of Fig. 2, even though that constant would vary across frequency. The threshold reference could have had an influence on the correlations between behavioral and DPOAE thresholds when data were collapsed across frequency (Figs. 1, 5, and 6). However, in the condition in which the most direct comparisons could be made between the present results and those reported by Boege and Janssen (see Fig. 1), similar correlations were observed. Thus, there was no apparent detrimental affect of threshold reference on the correlation.

Boege and Janssen (2002) related their DPOAE I/O pressure functions to previous measurements of basilar-membrane motion, described by Ruggero *et al.* (1997). While it is attractive to relate the shape of the DPOAE functions to underlying physiological response properties, it is unclear that such effort was necessary or adds to the practical value of the observations they made. In the end, the success of their approach relies on an empirical evaluation of the extent to which behavioral thresholds can be predicted from DPOAE I/O functions. Their efforts in this regard make an important contribution to the continuing evolution of the clinical application of DPOAE measurements.

Our approach has been to follow a similar empirical evaluation of the extent to which DPOAE measurements can be used to predict behavioral pure-tone thresholds. We were able to essentially replicate the work of Boege and Janssen. If at least three points on the DPOAE I/O function are characterized by a SNR $\geqslant$6 dB, DPOAE levels, converted to pressure and fit with a linear equation of pressure onto stimulus level, can be used to extrapolate to a DPOAE threshold that itself can be used to predict behavioral pure-tone thresholds. An advantage of using supra-threshold values on the DPOAE I/O function is that the measurements can be made in conditions for which a favorable SNR might be observed. Contrast this case with one in which DPOAE thresholds are estimated from direct measurements in which stimulus levels are used that range from being just above to just below DPOAE threshold. Under these latter conditions, the SNR, by definition, will be low in every case, which will affect the reliability of the measurement and make response detection difficult. Thus, supra-threshold DPOAE measurements provide an opportunity for more reliable and potentially quicker measurements than DPOAE threshold measurements. In turn, this could lead to behavioral threshold predictions that might be accomplished under routine clinical conditions.

Our initial, combined-frequency correlation coefficient (0.70) exceeded the correlation reported by Boege and Janssen (0.65) despite our having two additional sources of variability. First, we did not use the same earphone for behavioral threshold and DPOAE measurements, as was done in the previous study. Second, our subjects had a wider range of hearing thresholds. Our observation of a higher correlation may have been due to our use of measurement-based stopping rules, which resulted in longer averaging times for those DPOAE measurements in which the noise level was high.

Several studies have examined the relationship between DPOAE measurements and behavioral thresholds (Martin *et al.*, 1990; Allen and Levitt, 1992; Kimberley *et al.*, 1994, 1997; Gorga *et al.*, 1996, 1997, 2002; Dorn *et al.*, 1998). These studies observed varying degrees of success in relating the two measures. In the end, however, DPOAE measurements have been used almost exclusively to make a dichotomous decision as to whether hearing is normal or impaired, without regard to the magnitude of the hearing loss (e.g., Gorga *et al.*, 1993, 1997, 2000; Stover *et al.*, 1996; Kim *et al.*, 1996). Most multivariate estimates also were concerned with determining if hearing was normal or impaired (Dorn *et al.*, 1999; Gorga *et al.*, 1999), although the work of Kimberley *et al.* (1994, 1997) represents an exception to that rule. The approach described by Boege and Janssen (2002) and Oswald *et al.* (2002) makes use of the entire DPOAE I/O function in deriving an estimate that goes beyond a dichotomous decision and predicts behavioral threshold. Although variable, it was the case that behavioral thresholds could be predicted from DPOAE data, an observation that was replicated in the first part of the present experiment.

Unfortunately, not all DPOAE data met the initial SNR inclusion criterion for a sufficient number of points. Clinically, it is of interest to understand what proportion of the time this occurred and, more importantly, what was the auditory status in those cases when the SNR criterion was not met. Using the same criterion that was used by Boege and Janssen (2002), 44.2% of all DPOAE I/O functions (collapsed across frequency) did not meet the SNR inclusion criterion. However, 90% of these cases had accompanying behavioral thresholds greater than 30 dB HL, and the mean threshold for the entire group failing to meet the SNR criterion was 62.9 dB HL (s.d. = 27.3 dB) [see Fig. 1(b)]. While it was not possible to predict behavioral threshold from the DPOAE data when the SNR criterion was not met, hearing loss was present in the majority of these cases. From a clinical perspective, this is important information in that the hearing loss was identified by the technique, even if it was not quantified. On the other hand, 18.4% of the total sample of DPOAE I/O functions failed to meet the inclusion criteria associated with the linear regression of DPOAE pressure (μPa) onto DPOAE stimulus level (dB SPL). Eighty-one percent of this subgroup (329 DPOAE I/O functions) had thresholds better than 30 dB HL. In a sense, these cases represent a ‘‘false-positive’’ condition, in which normal ears failed to meet criteria. When added to the number of conditions for which the SNR criterion was not met among ears with thresholds less than or equal to 30 dB HL (98 conditions), one derives an overall ''false-positive'' rate of 19.3% ((329+98)/2211). Obviously, reducing this number would be of clinical interest.

Errors in prediction were not uniformly distributed across frequency. The best performance was observed for mid-to-high frequencies, with poorer performance at lower frequencies, 8000 Hz, and perhaps 6000 Hz (see Figs. 2–4, and Tables I–III). For example, the correlations were higher and the standard errors were typically lower at 2000, 3000, and 4000 Hz, compared to higher and lower frequencies. Furthermore, the percentage of cases with behavioral thresholds less than or equal to 30 dB HL that did not meet the SNR criterion were lower at these frequencies (7.5% at 2000 Hz, 5% at 3000 Hz, and 0% at 4000 Hz). Thus, there was a smaller percentage of cases at 2000, 3000, and 4000 Hz for which the SNR criterion was not met *and* hearing was normal. Finally, one can also estimate something akin to a ''false-positive'' rate by adding the number of cases with behavioral thresholds ≤30 dB HL that failed to meet either the SNR criterion or the inclusion criteria associated with the linear regression, and divide this number by the total number of DPOAE I/O functions for these three frequencies. This results in a ''false-positive'' rate of 12.9%, which is less than the similarly calculated percentage collapsed across all frequencies (19.3%). These observations are not unexpected, given previous results that have demonstrated that DPOAE test performance in a dichotomous pass/fail decision is better at mid and high frequencies, compared to lower frequencies (e.g., Gorga *et al.*, 1993, 1997, 2000; Kim *et al.*, 1996; Stover *et al.*, 1996).

Finally, overall performance was improved if the range of predicted DPOAE thresholds and the range of behavioral thresholds were restricted. These restrictions resulted in an increase in the correlation from 0.70 to 0.82. Additionally, including a wider range of stimulus levels and altering the inclusion criteria associated with the linear regressions resulted in a slight further increase in the correlation between behavioral thresholds and predicted DPOAE thresholds. More importantly, these changes in inclusion criteria allowed for predictions of behavioral thresholds in a larger percentage of DPOAE measurements. Some of these manipulations are based in an understanding of both DPOAE responses and behavioral thresholds, and others are more arbitrary. However, these results suggest that further improvements in accuracy might be achieved through additional efforts to optimize predictions of behavioral thresholds from DPOAE data (see also Oswald *et al.*, 2002).

ACKNOWLEDGMENTS

This work was supported by a grant from the NIH (NIDCD R01 DC2251). Portions of this work were presented at the 2003 Midwinter Meeting of the Association for Research in Otolaryngology. We thank two anonymous reviewers for their helpful suggestions on an earlier version of this manuscript.

[1]This slope criterion value differs from the 0.1 μPa/dB criterion reported by Boege and Janssen (2002) in their Eq. (4), but agrees with the value shown in their Fig. 8(b). The slope criterion listed in their Eq. (4) (0.1 μPa/dB) represents a typographical error. The actual slope criterion was 0.2 μPa/dB, which is what was used in their Fig. 9 (Boege, 2002), which represents the summary to which the present data are compared.

Allen, J. B., and Levitt, H. (**1992**). ''A comparison of pure tone audiometric and distortion product otoacoustic emission thresholds,'' unpublished manuscript.

ANSI(**1996**). ANSI S3.6, ''Specifications for Audiometers'' (American Institute of Physics, New York).

Boege, P. (**2002**). Personal communication.

Boege, P., and 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.

Brownell, W. E. (**1990**). ''Outer hair cell electromotility and otoacoustic emissions,'' Ear Hear. **11**, 82–92.

Dallos, P. J., Harris, D. M., Relkin, E., and Cheatham, M. A. (**1980**). ''Two-tone suppression and intermodulation distortion in the cochlea: Effect of outer hair cell lesions,'' in *Psychophysical, Physiological and Behavioral Studies of Hearing*, edited by G. van den Brink and F. A. Bilsen (Delft U. P., Delft, The Netherlands), pp. 242–252.

Dorn, P. A., Piskorski, P., Gorga, M. P., Neely, S. T., and Keefe, D. H. (**1999**). ''Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses,'' Ear Hear. **20**, 149–163.

Dorn, P. A., Piskorski, P., Keefe, D. H., Neely, S. T., and Gorga, M. P. (**1998**). ''On the existence of an age/threshold/frequency interaction in distortion product otoacoustic emissions,'' J. Acoust. Soc. Am. **104**, 964–971.

Dorn, P. A., Konrad-Martin, D., Neely, S. T., Keefe, D. H., Cyr, E., and Gorga, M. P. (**2001**). ''Distortion product otoacoustic emission input/output functions in normal-hearing and hearing-impaired human ears,'' J. Acoust. Soc. Am. **110**, 3119–3131.

Gorga, M. P., Neely, S. T., and Dorn, P. A. (**1999**). ''DPOAE test performance for a priori criteria and for multifrequency audiometric standards,'' Ear Hear. **20**, 345–362.

Gorga, M. P., Neely, S. T., and Dorn, P. A. (**2002**). ''Distortion product otoacoustic emissions in relation to hearing loss,'' in *Otoaccoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 243–272.

Gorga, M. P., Stover, L. J., and Neely, S. T. (**1996**). ''The use of cumulative distributions to determine critical values and levels of confidence for clinical distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 968–977.

Gorga, M. P., Nelson, K., Davis, T., Dorn, P. A., and Neely, S. T. (**2000**). ''Distortion product otoacoustic emission test performance when both 2f1-f2 and 2f2-f1 are used to predict auditory status,'' J. Acoust. Soc. Am. **107**, 2128–2135.

Gorga, M. P., Neely, S. T., Ohlrich, B., Hoover, B., Redner, J., and Peters, J. (**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., Bergman, B. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., and Jesteadt, W. (**1993**). ''Otoacoustic emissions from normal-hearing and hearing-impaired subjects: Distortion product responses,'' J. Acoust. Soc. Am. **93**, 2050–2060.

Janssen, T., Kummer, P., and Arnold, W. (**1998**). ''Growth behavior of the 2f1-f1 distortion product otoacoustic emission in tinnitus,'' J. Acoust. Soc. Am. **94**, 2659–2669.

Kim, D. O., Paparello, J., Jung, M. D., Smursynski, J., and Sun, X. (**1996**). ''Distortion product otoacoustic emission test of sensorineural hearing loss: Performance regarding sensitivity, specificity, and receiver operating characteristics,'' Acta Octolaryngol. (Stokh.) **116**, 3–11.

Kimberley, B. P., Hernadi, I., Lee, A. M., and Brown, D. K. (**1994**). ''Predicting pure-tone thresholds in normal and hearing-impaired ears with distortion product emission and age,'' Ear Hear. **15**, 199–209.

Kimberley, B. P., Brown, D. K., and Allen, J. B. (**1997**). ''Distortion product emissions and sensorineural hearing loss,'' in *Otoacoustic Emissions: Clinical Applications*, edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 181–204.

Kummer, P., Janssen, T., and Arnold, W. (**1998**). ''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. **103**, 3431–3444.

Martin, G. K., Ohlms, L. A., Franklin, D. J., Harris, F. P., and Lonsbury-Martin, B. L. (**1990**). ''Distortion product emission in humans. III. Influence of sensorineural hearing loss,'' Ann. Otol. Rhinol. Laryngol. Suppl. **147**, 30–42.

Neely, S. T., and 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., and Liu, Z. (**1994**). ''EMAV: Otoacoustic emission averager,'' Tech. Memo No. 17 (Boys Town National Research Hospital, Omaha, NE).

Oswald, J. A., Muller, J., and Janssen, T. (**2002**). ''Audiometric threshold estimation in cochlear hearing loss ears by means of weighted extrapolated DPOAE I/O functions,'' presented at the Twenty-Fifth Annual Midwinter Research Meeting of the Association for Research in Otolaryngology, St. Petersburg Beach, FL.

Ruggero, M. A., Rich, N. C., Recio, A., Narayan, S. S., and Robles, L. (**1997**). ''Basilar-membrane responses to tones at the base of the chinchilla cochlea,'' J. Acoust. Soc. Am. **101**, 2151–2163.

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

Siegel, J. H. (**2002**). ''Calibrating Otoacoustic Emission Probes,'' in *Otoacoustic Emissions: Clinical Applications*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical, New York), pp. 416–441.

Stover, L., Gorga, M. P., Neely, S. T., and Montoya, D. (**1996**). ''Towards optimizing the clinical utility of distortion product otoacoustic emission measurements,'' J. Acoust. Soc. Am. **100**, 956–967.