Reliability of distortion-product otoacoustic emissions and their relation to loudness

Megan J. Thorson,[a] Judy G. Kopun, Stephen T. Neely, Hongyang Tan, and Michael P. Gorga[b]

Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131

The reliability of distortion-product otoacoustic emission (DPOAE) measurements and their relation to loudness measurements was examined in 16 normal-hearing subjects and 58 subjects with hearing loss. The level of the distortion product (L_d) was compared across two sessions and resulted in correlations that exceeded 0.90. The reliability of DPOAEs was less when parameters from non-linear fits to the input/output (I/O) functions were compared across visits. Next, the relationship between DPOAE I/O parameters and the slope of the low-level portion of the categorical loudness scaling (CLS) function (soft slope) was assessed. Correlations of 0.65, 0.74, and 0.81 at 1, 2, and 4 kHz were observed between CLS soft slope and combined DPOAE parameters. Behavioral threshold had correlations of 0.82, 0.83, and 0.88 at 1, 2, and 4 kHz with CLS soft slope. Combining DPOAEs and behavioral threshold provided little additional information. Lastly, a multivariate approach utilizing the entire DPOAE I/O function was used to predict the CLS rating for each input level (dB SPL). Standard error of the estimate when using this method ranged from 2.4 to 3.0 categorical units (CU), suggesting that DPOAE I/O functions can predict CLS measures within the CU step size used in this study (5). © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3672654]

I. INTRODUCTION

Distortion-product otoacoustic emissions (DPOAE) result when two pure tones of slightly different frequency are played to the ear at the same time. The interaction of these tones results in the generation of several distortion products, all of which are mathematically related to the frequencies of the two primary tones. These low-level distortion products are apparently dependent on the status of OHCs (outer hair cells) (Brownell, 1990). Damage to the OHCs invariably results in hearing loss and a reduction in DPOAEs. As a result of these relationships, DPOAEs have been used as an objective and noninvasive way to assess cochlear status (e.g., Lonsbury-Martin and Martin, 1990; Gorga *et al.*, 1993, 1997). DPOAEs are currently used in newborn-hearing screenings, ototoxicity monitoring, and differential diagnoses, and although they are not routinely used to predict behavioral threshold, a relationship has been demonstrated between DPOAEs and audiometric threshold (Boege and Janssen, 2002; Gorga *et al.*, 2003; Oswald and Janssen, 2003; Rogers *et al.*, 2010).

The utility of DPOAEs requires that they are repeatable. One way to examine the reliability of DPOAEs is to compare the level of the distortion-product (L_d) for measurements when both L_1 and L_2 levels (the levels of the two frequencies making up the primary-frequency pair used to elicit the response) are set to moderate intensities (i.e., 50–75 dB SPL). These specific primary levels are commonly used in the clinic. Several studies have shown that DPOAEs elicited with moderate-level primaries can be measured reliably (Franklin *et al.*, 1992; Roede *et al.*, 1993; Beattie *et al.*, 2003; Wagner *et al.*, 2008; Keppler *et al.*, 2010). It may be important to note, however, that every subject included in these studies had normal hearing. We are unaware of data that describe the reliability of DPOAE measurements in subjects with hearing impairment. Potentially, this is an important limitation of previous work because subjects with normal hearing typically produce large responses that are well above the noise floor. Subjects with hearing loss produce smaller responses, resulting in DPOAE levels closer to the noise floor, which may affect the reliability of these measurements.

DPOAE reliability can also be examined through measurements of L_d for a wide range of input levels, including lower levels that are not in common clinical use, but would be applicable for cases in which the data are used to predict behavioral thresholds. Franklin *et al.* (1992) measured DPOAE input/output (I/O) functions in 12 normal-hearing adults across short-term and long-term retest periods. DPOAE reliability was high ($r = 0.90$ to 0.97) when responses were elicited from moderate and high primary levels ($L_2 > 55$ dB SPL), but repeatability decreased ($r = 0.30$ to 0.90) with lower primary levels ($L_2 < 45$ dB SPL), a consequence of the fact that L_d depends on primary level whereas noise levels are not directly related to primary levels. As a result, the signal-to-noise ratio (SNR) decreases as primary level decreases. Thus, variance in noise level exerts a greater relative influence on measured L_d at low primary levels, making the measured response less reliable, compared to the results for higher primary levels. This finding has been observed in other studies (Roede *et al.*, 1993; Lasky *et al.*, 1994; Stuart *et al.*, 2009).

[a] Also at: Department of Communication Disorders and Special Education, University of Nebraska–Lincoln, Lincoln, NE 68501.
[b] Author to whom correspondence should be addressed. Electronic mail: michael.gorga@boystown.org

However, like much of the previous work, every subject included in these studies had normal hearing. Furthermore, the observation of reduced repeatability at low primary levels in normal-hearing subjects may portend what occurs in subjects with hearing loss.

The reliability of DPOAE measurements can also be assessed by comparing a variety of parameters that describe I/O functions. Descriptions of the reliability of the entire DPOAE I/O function (including low-level conditions) are relevant to efforts to predict audiometric threshold from DPOAE measurements (Boege and Janssen, 2002; Gorga et al., 2003; Oswald and Janssen, 2003; Rogers et al., 2010). Typically, DPOAE I/O functions are fit with a simple linear regression (e.g., Kummer et al., 1998; Boege and Janssen, 2002; Gorga et al., 2003); however, fitting I/O functions with two separate lines may be more appropriate, given their shape. In fact, Neely et al. (2009) evaluated "best fits" to 1779 DPOAE I/O functions by comparing linear regression fits with nonlinear, two-slope functions. These I/O functions came from both normal-hearing and hearing-impaired subjects. Overall, Neely et al. found that the nonlinear, two-slope method provided a "better fit" to approximately 96% of the I/O functions. Although this method may provide the "best fit," the repeatability of the parameters obtained from this method has not been evaluated. Regardless of how it is assessed, the reliability of DPOAE measurements will have an impact on the extent to which these measurements can be used to identify hearing loss, predict threshold, and provide insights into underlying cochlear function in humans.

Another potential use of DPOAE measurements might be to predict auditory response growth, such as growth of loudness. In cases of mild hearing loss, such information may be of clinical use in helping to determine the appropriate signal-processing strategy in hearing aids, especially for infants, young children, and patients with developmental disabilities who may be unable to describe their percepts of loudness. This application would be of interest clinically because one consequence of cochlear hearing loss may be an abnormal growth of loudness, sometimes referred to as loudness recruitment.

Loudness recruitment occurs when the range of levels from threshold to uncomfortable loudness is reduced. The exact cause underlying this abnormal response growth is unknown. However, some studies (e.g., Scharf, 1978; Moore and Glasberg, 1997; Moore et al., 1999) have attributed loudness recruitment to changes in the basilar-membrane (BM) mechanical response. Direct BM measurements in lower mammals with normal auditory sensitivity have revealed I/O functions that grow approximately linearly in response to low-level inputs, but as sound levels increase, the response becomes compressive (Robles and Ruggero, 2001). This process allows the ear to respond to a wide range of levels. However, damage to the OHCs from cochlear insults, such as noise exposure (Zhang and Zwislocki, 1995) or furosemide treatment (Ruggero and Rich, 1991), alters the mechanical response of the cochlea causing it to behave more linearly. This linear response causes a reduction in the range of levels to which the ear can respond, and supports the notion that loudness recruitment is caused by damage to the OHCs. However, there are alternative theories about the underlying cause

of loudness recruitment. For example, Heinz et al. (2005) and Cai et al. (2009) have suggested that changes occur in the auditory system beyond its periphery and that these more central changes are responsible for loudness recruitment. Regardless of its source, loudness recruitment can be examined through measures of loudness (Allen et al., 1990; Hellman and Meiselman, 1993; Buss et al., 1998). Hearing aids can be used to compensate for elevated thresholds, while also compressing the normal range of levels into the reduced dynamic range of patients with hearing loss. Because abnormal perceptions of loudness can impact a patient's success with assistive technology (including the patient's willingness to use hearing aids), loudness measurements are sometimes made clinically to assist with hearing-aid fittings. Some patients (i.e., infants and patients with developmental delays), however, may be unable to provide reliable indications of the loudness of sound.

The compressive nonlinearity of the cochlea can be observed indirectly through OAE measurements. In fact, I/O measurements of both transient-evoked otoacoustic emissions (TEOAEs) (Epstein et al., 2004) and DPOAEs (Withnell and Yates, 1998) provide evidence of normal compression that is similar to what has been observed in direct BM I/O measurements (Rhode, 1971; Ruggero et al., 1996; Robles and Ruggero, 2001). In normal-hearing subjects, DPOAE I/O functions are characterized by nonlinear, compressive growth, but in ears with hearing loss, DPOAE I/O functions become increasingly linear (Dorn et al., 2001; Neely et al., 2003; Müller and Janssen, 2004; Neely et al., 2009). The change in response growth seen in DPOAE I/O functions when hearing loss exists is consistent with a loss of compressive nonlinearity that may underlie loudness recruitment.

It is these observations that have led to the hypothesis that there may be a relationship between growth of OAE level with primary level and growth of loudness. For example, it has been suggested that TEOAEs may provide a measure related to BM motion in humans (Epstein et al., 2004). In fact, Epstein and Florentine (2005) found that TEOAEs and psychoacoustic measurements resemble estimates of BM responses when they compared TEOAE I/O functions to data obtained from three separate psychoacoustic tasks (temporal-integration measurements, multitone loudness summation, and pulsation thresholds) in eight normal-hearing subjects. These findings were later supported by Epstein and Silva (2009), who compared TEOAE measurements of six normal-hearing subjects to a psychoacoustic loudness task known as cross-modality matching (CMM). While a relationship between TEOAE and CMM measures was observed at 1 kHz, a similar relationship was not observed at 4 kHz. Overall, the results suggest that TEOAE I/O functions may provide an objective way to measure BM response growth, at least for 1 kHz. However, it is important to note that these findings were based on a relatively small group of subjects, all of whom had normal hearing.

There is also evidence to suggest that DPOAEs may provide information related to behavioral response growth (loudness). In fact, Neely et al. (2003) described a relationship between DPOAEs and loudness when they compared DPOAE growth patterns to Fletcher and Munson (1933) loudness functions plotted on a logarithmic scale. Both the averaged DPOAE I/O data and the loudness functions in normal-hearing

subjects demonstrated linear growth near threshold. As stimulus levels increased, a more compressive growth was observed. Similarities in these functions suggested that the same peripheral nonlinearity may determine the growth of both of these measures of response growth, and that it may be possible to predict loudness from DPOAE I/O functions. However, Neely *et al.* observed intersubject variability in DPOAE measurements, suggesting that individual I/O measurements may not accurately predict loudness growth on a case-by-case basis.

Müller and Janssen (2004) extended the work of Neely *et al.* (2003) by evaluating the relationship between DPOAEs and loudness with the aim of investigating the use of DPOAE I/O functions as an objective tool in hearing-aid fittings. DPOAE I/O and categorical loudness scaling (CLS) functions were measured in both normal-hearing and hearing-impaired subjects. They observed a similar behavior between CLS and DPOAE I/O functions when plotted on a logarithmic scale, confirming the findings of Neely *et al.* (2003). In addition, small differences between gain estimates derived from normalized DPOAE and CLS data were interpreted to mean that DPOAE data relate to CLS measurements. Although these findings support the idea that DPOAE I/O functions may potentially be used as an objective tool in hearing-aid fittings, this relationship was based on data from ten normal-hearing subjects and nine subjects with hearing impairment whose losses were categorized as moderate. Further studies are needed to corroborate the proposed relationship between CLS and DPOAE measurements.

The ability to predict behavioral response growth (such as loudness) from OAE measurements may be useful in patients who cannot provide reliable voluntary responses about loudness. Although OAEs are not currently used to make loudness predictions, a relationship between these measures has been suggested (Neely *et al.*, 2003; Müller and Janssen, 2004; Epstein and Florentine, 2005; Epstein and Silva, 2009). Thus, it may be possible to use OAE data to objectively estimate growth of loudness. There may also be benefit in combining information about audiometric thresholds and OAE data to improve the accuracy of loudness predictions. Al-Salim *et al.* (2010) found that loudness growth (based on CLS measurements) may be predicted from audiometric threshold, accounting for nearly 80% of the variance in the slope of the CLS function. It is possible that combining behavioral threshold and OAE data may more accurately predict loudness growth than either measure alone.

The purposes of the present study were two-fold. First, the reliability of DPOAEs was assessed by comparing L_d levels for a wide range of primary levels and by comparing parameters derived from fits to the DPOAE I/O functions from repeated measurements. Second, the relationship between DPOAEs and CLS measurements was examined. Specifically, the ability to estimate loudness growth using DPOAEs, both independently and in combination with behavioral thresholds, was evaluated.

II. METHODS

A. Subjects

A total of 74 subjects, ages 11–76 years, participated in this study. Sixteen subjects had normal hearing and 58 sub-

jects had hearing loss. Data were collected during two sessions separated by as little as one week to as much as six months. All measurements were made monaurally. Pure-tone air-conduction thresholds were measured at octave and inter-octave frequencies from 0.25 to 8 kHz while pure-tone bone-conduction thresholds were measured at octave frequencies from 0.25 to 4 kHz. Subjects with air-conduction thresholds less than or equal to 15 dB HL were considered normal hearing, whereas those with thresholds greater than 15 dB HL were categorized as hearing impaired. Hearing loss did not exceed 75 dB HL at the test frequency for any subject. Subjects were excluded from the study if air-bone gaps were greater than 10 dB at any frequency. Middle-ear function was assessed using 226-Hz tympanometry prior to each session in which DPOAE measurements were made. Static compliance between 0.3 to 2.5 cm^3 and middle-ear pressure ranging from +25 to −75 daPa were considered normal. Subjects not meeting these criteria were excluded from the study.

B. DPOAE stimuli

DPOAE data were collected using custom-designed software (EMAV version 2.89, Neely and Liu, 1994). Stimuli were produced and responses were recorded by a 24-bit soundcard (CardDeluxe, Digital Audio Labs, Chanhassen, MN) housed in a PC. Separate channels of the soundcard were used to generate two primary tones (f_1 and f_2), which were sent to two separate loudspeakers that were housed in a probe-microphone system (ER-10C, Etymotic Research, Elk Grove Village, IL). The probe-microphone system was coupled to the ear using a foam tip. The ER-10C had been modified to remove 20 dB of attenuation in order to present stimuli up to 80 dB SPL during DPOAE measurements. Stimulus calibration in sound pressure level (SPL) was completed in the ear canal using the ER-10C microphone. Although there are concerns for potential errors associated with standing waves when using this calibration method (Siegel, 2002, 2007; Scheperle *et al.*, 2008), recent data suggest that DPOAE measurements are not affected by calibration method when DPOAEs are used either to make diagnostic predictions (normal-hearing versus hearing-impaired) (Burke *et al.*, 2010) or to predict behavioral thresholds (Rogers *et al.*, 2010).

DPOAE I/O functions were recorded at three f_2 frequencies (1, 2, and 4 kHz). The level of f_2 (L_2) ranged from −20 to 80 dB SPL. The level of f_1 (L_1) was determined using the formula suggested by Johnson *et al.* (2006a,b):

$$L_1 = 80 + 0.137 \cdot \log_2(18/f_2) \cdot (L_2 - 80).$$

The f_2/f_1 ratio was also determined using a formula suggested by Johnson *et al.* (2006a,b):

$$f_2/f_1 = 1.22 + \log_2(9.6/f_2) \cdot (L_2/415)^2.$$

These parameters have been shown to result in the largest DPOAE level, at least for normal-hearing subjects (Neely *et al.*, 2005; Johnson *et al.*, 2006a,b).

C. DPOAE measurement procedure

Data collection began at an L_2 equal to 15 dB below a subject's audiometric threshold for each f_2 frequency. Because behavioral thresholds were measured in dB HL, correction factors were applied to convert threshold data into dB SPL (ANSI, 2004). L_2 was increased in 5-dB steps until L_2 equaled 80 dB SPL. DPOAE data were collected into two separate buffers. Collection alternated between the two buffers on consecutive measurements. The level of the DPOAE (L_d) was based on the level in the $2f_1-f_2$ frequency bin and was estimated by summing the contents of the two buffers. The level of the noise (L_n) was estimated by subtracting the contents of the two buffers and then averaging the level in the $2f_1-f_2$ frequency bin along with the level in the five bins on either side of the $2f_1-f_2$ frequency bin.

Measurement-based stopping rules were used during data collection. Data collection was stopped if one of the three following criteria were met: the noise floor was less than or equal to -25 dB SPL, artifact-free averaging time exceeded 64 s, or the signal-to-noise ratio (SNR) was greater than 60 dB. These rules were selected so that data collection never stopped on the basis of the SNR criterion. The noise-floor criterion was the primary stopping rule and was chosen as a conservative estimate of the level at which system distortion occurred. In this way, its use helped eliminate the misinterpretation of system distortion as a biological response, while, at the same time, it maximized the dynamic range of reliable DPOAE measurements. Not all measurements terminated based on the noise criterion alone because, for some subjects, some conditions were characterized by high levels of noise; therefore, it was necessary to include an averaging-time stopping rule to avoid prolonged averaging times.

D. DPOAE analysis

A total of 442 I/O functions were available for analysis. DPOAE reliability was assessed using two methods. First, the L_d levels at each L_2 were compared across visits for each frequency. Analyses were performed with all L_d levels present and then with all L_d levels ≤ -25 dB SPL excluded because "responses" at these low levels may be a result of system distortion and not a true biological response. After eliminating those conditions for which $L_d \leq -25$ dB SPL, there were 2226 L_d levels available for comparison. The mean absolute differences and the mean signed differences in L_d levels, as well as the mean standard deviations for these differences, were calculated for each frequency as one way to assess the repeatability of these measurements. In addition, correlations were computed by combing L_d levels for each of the three frequencies.

DPOAE reliability was further examined by comparing the parameters obtained when nonlinear fits were applied to the DPOAE I/O functions, using a computer implementation of the model described by Neely et al. (2009). This model determines the "best" two-slope fit for the I/O functions. To be included in the analysis, DPOAE I/O functions had to have three or more consecutive points for which the L_d level was >-25 dB SPL and had a SNR ≥ 3 dB. The highest L_2 at which an acceptable SNR occurred marked the highest L_2
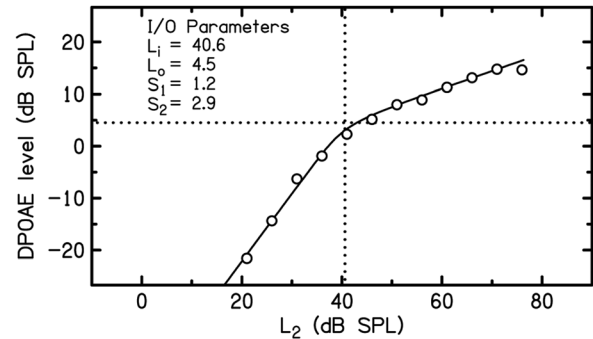

FIG. 1. Example of an I/O function fit by the two-slope, nonlinear model described by Neely et al. (2009). The two-slope fit (solid line) is superimposed on the DPOAE measurements for an $f_2 = 1$ kHz. S_1 represents the low-level slope and S_2 represents the additional compression that reduces the slope at high levels. The breakpoint between S_1 and S_2 is represented by two parameters: L_i and L_o. L_i represents the level of f_2 (L_2) that corresponds to the breakpoint between S_1 and S_2 and was estimated at 40.6 dB SPL (represented by the vertical dashed line). L_o represents the L_d at which the breakpoint occurred and was estimated at 4.5 dB SPL (represented by the horizontal dashed line).

level included in the fits. Any non-consecutive points below this L_2 level were excluded, including those for which the SNR exceeded 3 dB.

The stopping rules, described above, were chosen to assure that only biologically generated responses (as opposed to system distortion) were included in the analyses of DPOAE data. In a further effort to assure that only I/O functions representing true biological responses were analyzed, measurements from subjects with audiometric thresholds exceeding 50 dB HL were also excluded from this portion of the analysis. In total, there were 384 I/O functions that met all criteria and, therefore, were fit with the nonlinear, two-slope method described by Neely et al. (2009).

Figure 1 shows an example of an I/O function that was fit by this nonlinear, two-slope method. I/O parameters calculated for that specific I/O function are provided as insets. These parameters include two estimates related to the slopes of the function (S_1, the slope of the low level portion of the I/O function and S_2, the additional compression that was needed for the high level part of the function) and two values associated with the break between these two segments (L_i, the L_2 at the break point, and L_o, the L_d at the break point). Additional details regarding these four parameters and the extent to which they describe DPOAE I/O functions have been provided previously (Neely et al., 2009), in which over 1700 individual DPOAE I/O functions were analyzed. We are unaware of other models that account for more variability than this two-segment, nonlinear model. As a consequence, it was used to describe the present set of DPOAE I/O functions. The reliability of the DPOAE I/O parameters from the model fits were assessed across visits using correlation analyses.

E. CLS procedures

CLS measurements were obtained in the same groups of subjects. Measurement procedures will be only briefly summarized here because the data used in the present study were previously analyzed and described by Al-Salim et al. (2010).

Pure tones of varying intensity were generated using the same 24-bit sound card (CardDeluxe; Digital Audio Labs, Chanhassen, MN) used for DPOAE stimulus generation. For the CLS measurements, the sound card was controlled by custom-designed software (Behavioral Auditory Research Tests, v2.3.32). The stimuli were routed through a headphone buffer (HB7; Tucker-Davis Technologies, Alachua, FL) to a probe-microphone system (ER10-B+, Etymotic Research, Elk Grove Village, IL). This system was chosen because it allowed for the use of ER3A earphones which enabled the production of higher level stimuli (up to 105 dB SPL) which would not have been possible with the loudspeaker built into the ER-10C probe-microphone system used for DPOAE measurements. The stimuli were 1 s in duration with 20 millisecond rise/fall times. Stimulus presentations were separated by a 2-s interstimulus interval. In-the-ear calibrations were performed at each frequency before and after each measurement. Subjects were asked to rate the loudness of the tone immediately after each presentation by selecting, on a touch-screen monitor, one of 11 colored horizontal bars, which increased in length as the loudness category increased. Six bars contained meaningful labels (such as "Can't Hear," "Soft," "Loud," and "Too Loud") and alternated with five unlabeled bars. In addition, each bar also included one of a set of arbitrary numbers from 0 to 50 in steps of 5 that were used to represent categorical units (CU). Subjects were asked to make judgments of loudness based on each individual presentation. CLS measurements were obtained at 1, 2, and 4 kHz (the same frequencies for which DPOAE I/O functions were measured) and repeated three times for each frequency. All of these measurements were repeated during a second session that took place anywhere from a week to six months after their first visit. In order to construct the CLS functions, two steps were taken. First, the three trials from each frequency for each visit were averaged separately. Then, those values were averaged across both visits in order to obtain one dB SPL value for each loudness category. The CLS functions were constructed from those averages. The CLS functions were then separated into two portions. The low-level portion of the CLS function (referred to as "soft slope") was represented by CUs less than or equal to 20 whereas CUs greater than 20 represented the high-level portion. The reliability of CLS measurements and details regarding the parsing of the CLS function into two segments have been described previously in detail (Al-Salim et al., 2010).

The current paper examined the relationship between soft slope and DPOAE measurements in an effort to determine if objective measurements (such as measurements of DPOAE I/O functions) can be used to predict subjective judgments of loudness (such as CLS). To examine the relationship between soft slope and DPOAEs, two separate analyses were performed. For both analyses, L_d levels obtained from the DPOAE I/O functions were averaged across the two sessions. In the first analysis, DPOAE I/O functions were fit with two lines using the nonlinear method described by Neely et al. (2009) and reviewed in Fig. 1. The four DPOAE I/O parameters were used in several ways to assess the relationship between DPOAEs and CLS soft slope. First,

each individual parameter was correlated with CLS soft slope separately. Then, the DPOAE I/O parameters were combined and correlated with CLS soft slope using multiple-linear regressions. Next, the benefit of combining DPOAE parameters with audiometric thresholds to predict loudness was evaluated. Finally, threshold information was removed from the correlational analyses to further examine the extent to which DPOAE I/O parameters account for additional variance beyond what is accounted for by behavioral thresholds. In order to remove the influence of threshold, the predicted CLS soft slope was calculated and then subtracted from the actual CLS soft slope. This difference was referred to as residual soft slope. First, individual DPOAE I/O parameters were correlated with residual soft slope and then DPOAE I/O parameters were summed and correlations with residual soft slope were determined using multiple-linear regressions. The second analyses used to assess the relationship between DPOAEs and CLS measures utilized a multivariate approach in which the entire DPOAE I/O function was used to predict the CLS rating for each input level (dB SPL) for each individual subject.

III. RESULTS

A. Reliability of DPOAEs

One purpose of this study was to assess the reliability of DPOAE measurements in both normal-hearing and hearing-impaired subjects. The number of subjects per hearing loss (HL) category for each frequency included in the analysis is provided in Table I. DPOAE I/O functions were obtained at all three test frequencies for each of the 74 subjects. For a subject to be included in the analyses at a test frequency, I/O functions had to be available for both session 1 and session 2. At 1 kHz, there are more normal-hearing subjects than there are hearing-impaired subjects, whereas at 4 kHz, the pattern is

TABLE I. The number of subjects in each HL category for which DPOAE I/O functions were collected. Subjects were only included if DPOAE I/O functions were available for both session 1 and session 2. The number of subjects per HL category is shown separately for each frequency.

dB (HL)	1 kHz	2 kHz	4 kHz
−5	1	1	2
0	2	4	3
5	10	6	6
10	13	8	4
15	11	9	4
20	11	10	6
25	6	4	6
30	5	3	5
35	2	8	2
40	4	5	6
45	3	6	3
50	5	6	14
55	1	3	8
60	–	1	2
65	–	–	1
70	–	–	–
75	–	–	1
Total	74	74	73

TABLE II. Mean absolute and signed differences of the level of the distortion product (L_d) between sessions for each test frequency. Analyses were performed with all measured L_d levels and after all L_d levels ≤ -25 dB SPL were excluded. Mean absolute and mean signed differences, SDs of these differences, and correlations of L_d from the two test sessions are provided in the table.

	L_d no exclusion			$L_d \leq -25$ dB SPL excluded		
	1 kHz	2 kHz	4 kHz	1 kHz	2 kHz	4 kHz
Mean absolute L_d difference	4.59	3.66	4.08	3.51	2.62	2.88
SD of difference	4.90	3.95	4.47	3.53	2.62	2.75
Mean signed L_d difference	−0.06	−0.45	0.14	−0.30	−0.36	−0.01
SD of difference	6.71	5.34	6.05	4.97	3.69	3.98
Correlation coefficient	0.90	0.93	0.91	0.91	0.95	0.94

reversed, with 2 kHz falling in between these two "extremes." When combined across frequency, there were about the same number of observations from normal-hearing and hearing-impaired ears.

Reliability was examined by calculating the mean difference in L_d across a range of primary levels for each f_2 frequency and calculating SDs of the mean L_d difference. In addition, DPOAE reliability was assessed using correlations of L_d from the two test sessions. Both the mean absolute L_d difference and the mean signed L_d difference were calculated, and comparisons were made between measured response levels from session 1 and session 2 for each primary level (L_2). DPOAE analysis was first performed with all L_d levels. Then, the process was repeated after L_d levels ≤ -25 dB SPL were excluded from the analysis. These exclusions are justified because measured L_d levels at these low levels (even if the noise floor was much lower) may not be reliable. They may be a manifestation of system distortion and not a true biologic response. Including them, therefore, would not be meaningful.

The above estimates, all of which were designed to provide information about the reliability of DPOAE measurements for individual stimulus conditions, are summarized in Table II for each of the three test frequencies. Overall, DPOAE reliability was higher when all L_d levels ≤ -25 dB SPL were excluded from the analyses. Although the differences were small, the exclusion rule improved DPOAE reliability and therefore, the results from the analyses in which the exclusion rule was applied will be described below. The mean absolute L_d difference, which was calculated by averaging the difference between L_d levels measured at each L_2 in session 1 and L_d levels for corresponding L_2 levels in session 2, ranged from 2.6 to 3.5 dB, with SDs of the mean absolute difference also ranging from 2.6 to 3.5 dB. Not surprisingly, the mean signed L_d difference was less than the absolute difference and was <1 dB at all test frequencies, while the SDs of this difference ranged from 3.7 to 5.0 dB. Lower standard deviations were observed, as expected, with absolute differences because the sign of the difference was ignored during calculations. In addition, the L_d measured at each L_2 in session 1 was correlated with L_d measured for the same L_2 in session 2. Correlations were ≥ 0.90 even when the exclusion rule was not applied, as can be seen in the bottom row of Table II.

Figure 2 plots the L_d in session 2 as a function of the L_d from session 1 for equivalent L_2 levels. Data from both normal-hearing and hearing-impaired subjects were combined

for each frequency and displayed within this figure, just as they were in the analyses summarized in Table II. The correlations shown as insets in each panel are reproduced from the right columns, bottom row of Table II. This figure only includes data after the exclusion rule was applied, which is why the lower limits of the data on both the y-axis and x-axis terminate at -25 dB SPL. Data for a different frequency are represented in each panel. There is greater variability for conditions in which L_d was small (for example, <0 dB SPL) compared to the variability when L_d was large (>0 dB SPL). Assuming the noise level was relatively constant regardless of L_d, its influence on measured L_d is less for large responses and greater for low-level responses, resulting in greater variability for L_d levels <0 dB SPL.

The correlation between each of the previously described I/O parameters from session 1 and session 2 was used as another way to assess the reliability of DPOAE measurements. Figure 3 shows the repeatability of the two slope parameters; each data point represents a data pair from one subject. In Fig. 3(A), S_1 derived from data collected in session 2 is plotted as a function of S_1 derived from data obtained in session 1. In Fig. 3(B), S_2 derived from data obtained in session 2 is plotted as a function of S_2 based on data collected during session 1. The slope estimates from the two sessions are plotted separately for each frequency. The highest correlation in slope data was observed for S_2 at 1 kHz (r = 0.52), with lower correlations at 2 and 4 kHz (r = 0.44 and r = 0.20). S_1 correlations were the lowest at 2 kHz (r = 0.28), with higher correlations at 1 and 4 kHz (r = 0.38 and r = 0.39). Correlations of S_1 and S_2 are less than the correlations for individual points on the I/O functions (see Table II and Fig. 2 for comparison).
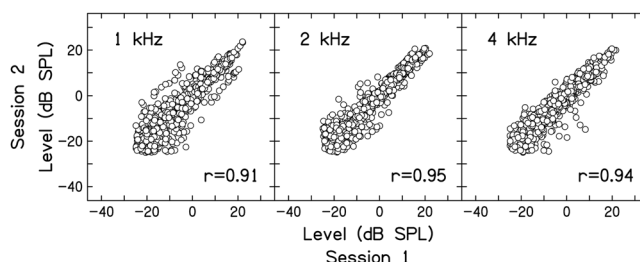

FIG. 2. L_d level in session 2 as a function of the L_d from session 1. Each data point represents an L_d recorded at one L_2. Data are provided for each frequency and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined in each panel.
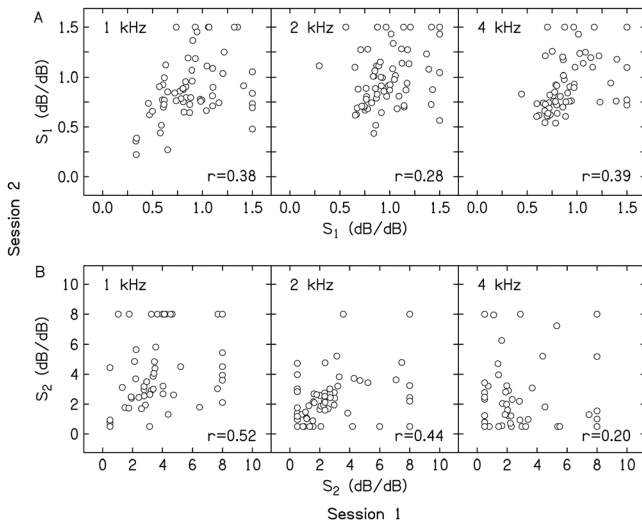
FIG. 3. Repeatability of the parameters obtained from nonlinear fits to DPOAE I/O functions. In the top row of panels (A), S_1 from session 2 is plotted as a function of S_1 obtained in session 1. In the bottom row of panels (B), S_2 from session 2 is plotted as a function of S_2 from session 1. In both A and B, each data point represents a pair of parameter values from one subject. Data are provided for each frequency and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined in each panel.

Correlations of L_i and L_o across the two sessions are shown in Figs. 4(A) and 4(B), respectively. The highest correlation was observed at 1 kHz ($r = 0.73$), which was the same for both L_i and L_o and can be found as insets in Figs. 4(A) and 4(B). Lower correlations were observed at 2 and 4 kHz for both L_i ($r = 0.60$ and $r = 0.57$) and L_o ($r = 0.56$ and $r = 0.54$). Like the correlations for S_1 and S_2, L_i and L_o had lower correlations compared to those observed from individual values of L_d on the DPOAE I/O function (see Table II and Fig. 2). However, L_i and L_o had stronger correlations than S_1 and S_2 (see Fig. 3).
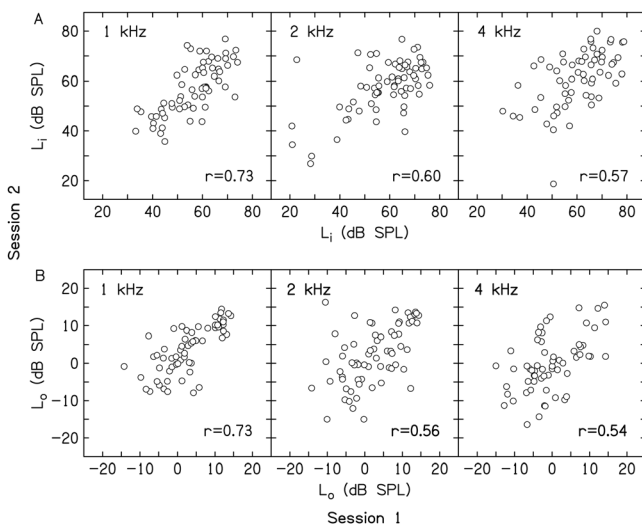

FIG. 4. Row A (top) plots L_i in session 2 as a function of L_i obtained in session 1. In the bottom panel (B), L_o in session 2 is plotted as a function of L_o from session 1. In both A and B, each data point represents a pair of parameter values from one subject. Data are provided for each frequency and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined in each panel.

Overall, the data summarized in Table II and Fig. 2 suggests that DPOAE measurements are reliable when the level of the distortion product is compared across sessions. However, the reliability of DPOAE measurements decreases when the parameters derived from nonlinear fits to I/O functions (S_1, S_2, L_i, L_o) are compared across sessions as evidenced by lower correlations presented in Figs. 3 and 4.

B. Relationship between CLS measurements and DPOAE I/O parameters

In addition to examining the reliability of DPOAE measurements, we also wanted to evaluate the relationship between DPOAE data and loudness. CLS and DPOAE data were obtained in the same groups of normal-hearing and hearing-impaired subjects from whom the data were obtained to estimate the reliability of DPOAE measurements. Because measurements were obtained during two sessions, the data from the two sessions were averaged prior to analyzing the relationship between CLS and DPOAE data.

Figure 5 shows the correlations between the four parameters describing DPOAE I/O functions and CLS soft slope for each frequency plotted on a logarithmic scale. By plotting soft slope on a log scale, the vertical spread of the slope values was reduced. In Fig. 5, each point represents the data from one subject. The left-most column describes the relationship between audiometric threshold and soft slope, based on data taken from Al-Salim *et al.* (2010). Audiometric data were included in this analysis because Al-Salim *et al.* found a strong correlation between loudness growth (based on CLS measurements) and audiometric threshold. Those data provide a baseline condition against which correlations between CLS and DPOAE data can be compared. Our first goal in this regard was to determine if the correlations between CLS soft slope and parameters from the nonlinear fits to DPOAE I/O functions were similar to those observed between CLS soft slope and behavioral threshold. Audiometric threshold had correlations with CLS soft slope of 0.82, 0.83, and 0.88 at 1, 2, and 4 kHz, respectively (see insets in the left column of each panel of Fig. 5) and this indicates that threshold accounts for as much as 77% of the variance in soft slope. Of the four parameters describing DPOAE I/O functions, L_i had the strongest positive correlation with CLS soft slope; however, these correlations were not as high as those observed with audiometric threshold. L_i correlations were 0.34, 0.47, and 0.72 at 1, 2, and 4 kHz, respectively. The remaining parameters (S_1, S_2, and L_o) were characterized by even lower correlations with soft slope. In general, DPAOE I/O parameters obtained at 1 kHz had poorer correlations with soft slope than those observed at 2 and 4 kHz. These correlations are provided as insets in each panel in Fig. 5. Overall, audiometric threshold had a stronger correlation with CLS soft slope than any one parameter from the nonlinear fits to the DPOAE I/O functions.

To further examine the predictive value of DPOAE I/O parameters, an additional analysis was performed with threshold dependence removed. To accomplish this, the predicted CLS soft slope was subtracted from the actual CLS soft slope. This difference was referred to as residual soft slope. Figure 6
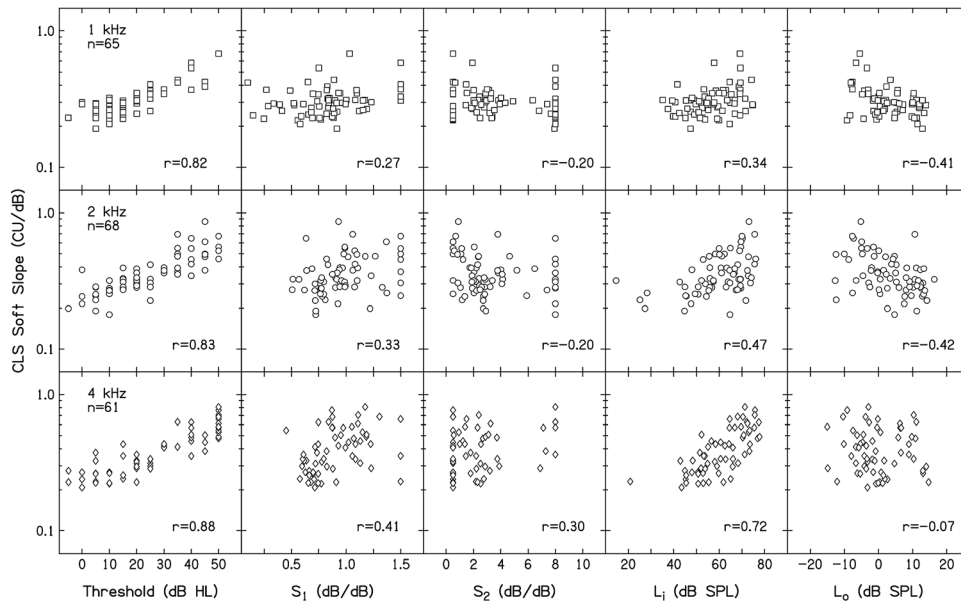
FIG. 5. CLS soft slope as a function of five predictor variables. The left column describes the relation between soft slope and audiometric threshold, based on data taken from Al-Salim *et al.* (2010). Correlations between CLS soft slope and the four parameters describing DPOAE I/O functions are provided in the remaining four columns. Each data point represents data from one subject. Data are provided for each frequency and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined within each panel.

shows correlations of residual soft slope, plotted in log units, to the four parameters describing the DPOAE I/O function as well as with audiometric threshold. As expected, correlations of zero were observed between residual soft slope and audiometric threshold because the calculation of residual soft slope removed any dependence on audiometric threshold. DPOAE I/O parameter correlations with residual soft slope at 4 kHz were greater than zero and greater than those at 1 and 2 kHz, with the exception of S_1. Also, correlations for S_1 at 1 kHz and for S_2, L_i, and L_o at 4 kHz were greater than 0.15, suggesting that parameters describing DPOAE I/O functions contain some information about CLS soft slope beyond what can be predicted from behavioral threshold alone. To further

examine this finding, we evaluated the predictive value of combining all four DPOAE I/O parameters when the variance accounted for by behavioral threshold was removed. Correlations of 0.37, 0.10, and 0.32 at 1, 2, and 4 kHz were observed between residual soft slope and combined DPOAE I/O parameters. These correlations are shown in Table III (bottom row). In particular, correlations between residual soft slope and combined DPOAE parameters observed at 1 and 4 kHz indicate that DPOAEs, based on these four parameters, account for approximately 10% to 14% of the variance in soft slope and supports the notion that DPOAE I/O parameters contain some information about CLS soft slope beyond that described by audiometric threshold.
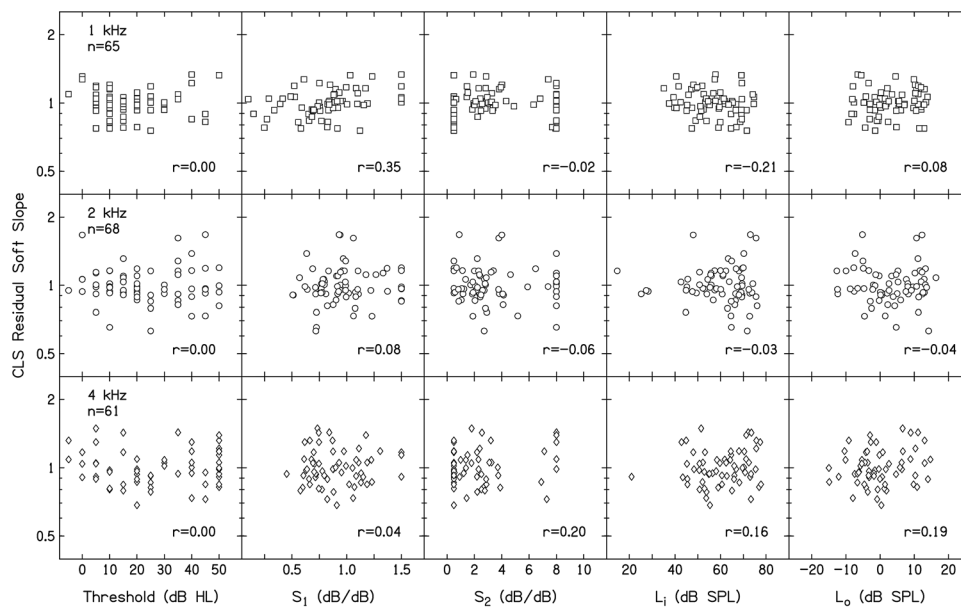

FIG. 6. Residual soft slope, displayed on the Y-axis, as a function of five predictor variables. CLS residual slope was determined by subtracting the predicted CLS soft slope (based on correlations with audiometric threshold) from the actual CLS soft slope. The first column describes the relation between audiometric threshold and residual soft slope. Correlations of zero were observed, as expected, because this manipulation was intended to remove any dependence on audiometric threshold. Relations of residual soft slope with the four parameters describing the DPOAE I/O functions are provided in the remaining four columns. Each data point represents data from one subject. Data are provided for each frequency and correlations are provided as insets in each panel. Data from normal-hearing and hearing-impaired subjects are combined within each panel.

TABLE III. Correlations between CLS soft slope, parameters describing DPOAE I/O functions, audiometric threshold, and combinations of these variables. The top row describes the relationship between audiometric threshold and CLS soft slope, based on data taken from Al-Salim *et al.* (2010). The second row (soft slope, DPOAEs only) provides correlations between soft slope and DPOAEs when all four DPOAE I/O parameters (S_1, S_2, L_i, and L_o) were combined. The third row (soft slope, DPOAEs + threshold) shows correlations between CLS soft slope and the combination of audiometric threshold and all of the parameters describing DPOAE I/O functions. The last row (residual, DPOAEs) describes the predictive value of combing all four parameters describing DPOAE I/O functions when the variance accounted for by behavioral threshold was removed (see Fig. 6).

	1 kHz	2 kHz	4 kHz
Soft slope, threshold only	0.82	0.83	0.88
Soft slope, DPOAEs only	0.65	0.74	0.81
Soft slope, DPOAEs + threshold	0.85	0.83	0.89
Residual, DPOAEs	0.37	0.10	0.32

Given that individual DPOAE I/O parameters have some correlation with CLS soft slope, we wanted to evaluate the relationship between CLS slope soft and DPOAEs when all four DPOAE I/O parameters were combined. Combining these parameters resulted in correlations of 0.65, 0.74, and 0.81 at 1, 2, and 4 kHz which are shown in Table III (the row referred to as soft slope, DPOAEs only). Although combined DPOAE I/O parameters resulted in higher correlations with CLS soft slope than any individual parameter, they were lower than those between CLS soft slope and audiometric threshold (see the first row of Table III).

Since audiometric threshold and DPOAE data both provide information about CLS soft slope independently, we evaluated the benefit of combining the data from both of these measurements. Together, behavioral thresholds and DPOAE data may more accurately predict loudness growth than either measure alone. However, when parameters describing the DPOAE I/O functions were combined with threshold data, there was little or no increase in correlation when compared to behavioral threshold alone. Correlations with CLS soft slope went from 0.82, 0.83, and 0.88 for audiometric threshold alone to 0.85, 0.83, and 0.89 at 1, 2, and 4 kHz when audiometric data were combined with

DPOAE data. These correlations are presented in Table III (row one, soft slope, threshold only and row three, soft slope, DPOAEs + thresholds). Overall, our results suggest that DPOAE I/O parameters correlate with CLS soft slope, but not as strongly as behavioral threshold, and that combining DPOAE I/O parameters and behavioral thresholds provides little benefit over what is achieved when behavioral threshold is used alone to predict CLS soft slope.

C. Relationship between CLS measurements and DPOAEs using multivariate analyses

Previous observations of the relationship between loudness measures and DPOAEs (Neely *et al.*, 2003; Müller and Janssen, 2004) as well as TEOAEs (Epstein and Florentine, 2005; Epstein and Silva, 2009) suggest that a stronger relationship exists between OAEs and loudness than was demonstrated in our correlation analyses. In further efforts to examine the relationship between DPOAEs and CLS measures, a multivariate approach was utilized in which the entire DPOAE I/O function was used to predict the CLS rating for each input level (dB SPL) for each subject.

Figure 7 shows examples of predicted CLS measures based on DPOAEs using multivariate analyses and the actual CLS measurements for 12 subjects. The first two columns provide examples for six normal-hearing subjects whereas the last two columns provide examples for six hearing-impaired subjects. These 12 subjects were selected randomly from the overall pool of 74 subjects. Behavioral thresholds for each subject are provided as insets in each panel. Filled symbols represent the measured CLS data for these 12 subjects, while the open symbols represent the predicted CLS function based on a multivariate linear regression in which all of the DPOAE data were used as inputs. It is important to note that the open symbols do NOT represent actual DPOAE levels but are predicted CLS units based on a multivariate prediction from DPOAE data. Qualitatively, it appears that using a multivariate approach allows for accurate predictions of CLS measures. The agreement between actual and predicted CLS units was further assessed by calculating the standard error of the estimate (SEE) for each of these 12
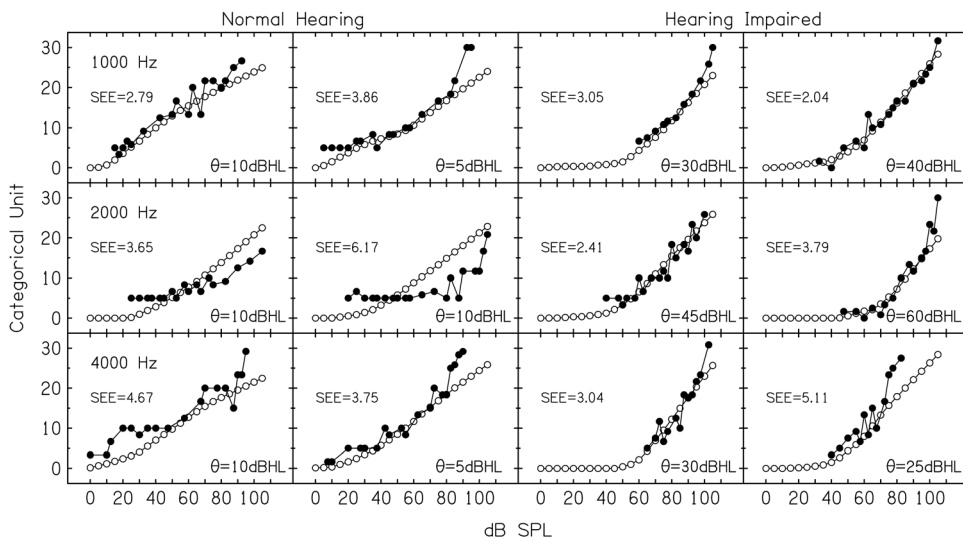

FIG. 7. Predicted categorical units (CU) based on DPOAE measurements using multivariate analyses (open circles) from 12 individual subjects plotted against actual CUs (closed circles) obtained in the same subjects. The first two columns provide examples for 6 normal-hearing subjects whereas the last two columns provide examples for 6 hearing-impaired subjects. Behavioral thresholds and the standard error of the estimate for individual subjects are provided as insets in each panel.

TABLE IV. Standard error of the estimate (SEE) of CLS functions from DPOAE I/O functions using multivariate analyses. The top row describes the SEEs when hearing was normal, the second row describes the SEEs when hearing loss existed, and the bottom row provides SEEs when data from all subjects were combined.

	1 kHz	2 kHz	4 kHz
Normal hearing	2.9	2.5	3.0
Hearing-impaired	2.4	2.6	2.9
All subjects	2.7	2.5	2.9

subjects. In order to calculate SEEs (which is equivalent to a standard deviation), a line was fit to the actual CLS function, and then SEEs were calculated using that line and the predicted CLS measurements. Consequently, the number of degrees of freedom in the SEE calculation was the number of observations minus 2. The calculated SEE for each subject is provided as insets in Fig. 7. Based on both visual inspection and the SEE for these 12 examples, it also appears that this method provides more accurate predictions for subjects with hearing impairment than it does for those with normal hearing.

Since the 12 examples shown in Fig. 7 suggest that DPOAEs can accurately predict CLS measures when using this multivariate approach, we further quantified the accuracy of the prediction by calculating the SEE for the entire group of 74 subjects. Table IV provides the mean calculated SEEs, which were 2.9, 2.5, and 3.0 CUs at 1, 2, and 4 kHz when hearing was normal and 2.4, 2.6, and 2.9 CUs at 1, 2, and 4 kHz when hearing loss existed. Overall, these results suggest that applying a multivariate analysis to DPOAE data results in an accurate estimate of loudness growth (as measured by CLS), and that this method is nearly equally accurate for both normal-hearing and hearing-impaired subjects. SEEs calculated with the data from all subjects combined were 2.7, 2.5, and 2.9 CUs at 1, 2, and 4 kHz (Table IV). Interestingly, CUs in this study were measured in steps of 5, which means that with this analytical technique, DPOAEs can predict CLS measurements, on average, within less than one CU step.

IV. DISCUSSION

A. Reliability of DPOAEs

DPOAE reliability was first determined by calculating the mean differences in L_d across a range of primary levels. Mean absolute differences fell within the range reported by Lasky et al. (1994) who evaluated both the within-session and between-session reliability of DPOAEs recorded at multiple frequencies and intensities. They reported between-session mean absolute differences that ranged from 2 to 7 dB, which was larger than their within-session reliability (2 to 4 dB). The mean absolute differences observed in the present study were closer to the within-session reliability reported by Lasky et al. (1994). In addition, they found that DPOAE reliability decreased as frequency decreased which is also demonstrated in the present study and can be observed in Table II. This effect is likely a consequence of the decrease in SNR as frequency decreases. Beattie et al. (2003) examined test-retest reliability of DPOAEs by calcu-

lating the standard error of measurement over three time intervals with fixed primary tones of 65 dB SPL ($L_1 = L_2$). They reported standard errors of approximately 2.5 dB at 1, 2, and 4 kHz when combined across frequency. Overall, Beattie et al. (2003) found that DPOAEs were reliable when measured on the same day and across test sessions. The repeatability of DPOAE measurements has been demonstrated in other studies (Franklin et al., 1992; Roede et al., 1993; Wagner et al., 2008; Keppler et al., 2010). All of these studies reported data from normal-hearing subjects. In the present study, data from normal-hearing and hearing-impaired subjects were combined in the assessment of DPOAE reliability. The small differences observed in the present study were similar to those measured in normal-hearing subjects and suggest that DPOAEs are reliable across sessions for both normal-hearing and hearing-impaired subjects for all three frequencies examined.

DPOAE reliability was further assessed using correlations of L_d from two test sessions. Franklin et al. (1992) reported similar reliability coefficients (0.90 to 0.97) to those observed here when they examined DPOAE reliability across short-term and long-term intervals. They measured DPOAEs using fixed primary tone levels ($L_1 = L_2$) at 55, 65, and 75 dB SPL across five test frequencies (1, 2, 4, 6, and 8 kHz). Note that they did not include data from low primary-level conditions like those included in the present study. Franklin et al. reported correlations that varied from 0.80 to 0.97 at 2 kHz and from 0.95 to 0.97 at 4 kHz, similar to values observed in the present study. In addition, they found that DPOAEs recorded at 1 kHz were less reliable, compared to higher test frequencies, with correlations ranging from 0.70 to 0.91. The lowest correlations for 1 kHz were observed for primary tones fixed at 55 dB, and as the level of the primaries increased, the reliability increased, a pattern also observed in the present study. The level effect, like the frequency effect, is a consequence of its effect on SNR. In general, Franklin et al. (1992) found that correlations decreased with decreasing frequency as well as decreasing primary level. Correlations obtained in the current study were at the top end of those reported by Franklin et al. (1992) and may have been a result of the measurement-based stopping rules utilized here, which likely resulted in relatively low noise levels that were more uniform across conditions and across subjects, compared to the case when other stopping rules are used. The noise-level stopping rule (which was selected as a conservative estimate of system distortion) increased the measured SNR and, by extension, the reliability of the measurements. Overall, these findings suggest that DPOAEs are reliable across sessions for both normal-hearing and hearing-impaired subjects.

DPOAE variability was small for conditions in which L_d was large (for example, > 0 dB SPL), compared to variability observed when L_d was small (< 0 dB SPL). In general, L_d decreases with decreasing intensity level and when hearing loss exists. In fact, many studies (Franklin et al., 1992; Roede et al., 1993; Lasky et al., 1994; Stuart et al., 2009) have demonstrated this effect. Efforts to predict audiometric threshold from DPOAE measurements depend on low-level responses; therefore, the greater variability observed for L_d levels < 0 dB SPL will affect estimates of threshold. DPOAEs produced

J. Acoust. Soc. Am., Vol. 131, No. 2, February 2012

Thorson *et al.*: Otoacoustic emissions and loudness   1291

with mid-to-high intensity primaries (i.e., 50–75 dB SPL) have high repeatability, but it is likely that this effect will be reduced in subjects with hearing loss. These specific primary-level conditions are commonly used in the clinic, where one would expect similar results. An L_d-dependent reliability was observed in the present study, in which these measurements were made in normal-hearing and hearing-impaired subjects. These results support the measurement of DPOAEs with moderate-level stimuli for clinical procedures such as newborn-hearing screenings, ototoxicity monitoring, and differential diagnoses.

This study also investigated the reliability of the slopes of DPOAE I/O functions. Traditionally, DPOAE I/O functions are fit with a single line (e.g., Kummer *et al.,* 1998; Boege and Janssen, 2002; Gorga *et al.,* 2003). Fitting DPOAE I/O functions with a simple linear regression would be appropriate if the function grew linearly; however, there are data to suggest that DPOAE I/O functions grow non-linearly (e.g., Nelson and Kimberley, 1992; Beattie *et al.,* 2004; Goldman *et al.*, 2006; Neely *et al.,* 2009). Beattie *et al.* (2004) examined DPOAE slopes derived from different segments of the I/O functions and found that the low-level segments (40–50 dB SPL) were steeper than slopes derived from moderate (50–65 dB SPL) and high-level (65–75 dB SPL) segments. Goldman *et al.* (2006) had proposed that fitting DPOAE I/O functions with segmented regressions may be superior to the linear-pressure model in common use. Segmented regressions may provide a better representation of DPOAE growth rate over all primary levels. Neely *et al.* (2009) examined the use of nonlinear, two-slope fits and found that this method provided a "better fit" over the typical single linear regression in most cases. Therefore, the nonlinear, two-slope method described by Neely *et al.* (2009) was used to fit DPOAE I/O functions in the current study. This method resulted in four parameters (S_1, S_2, L_i, and L_o), the values of which were correlated across the two test sessions. Correlations of slope data (S_1 and S_2) were less than the correlations for individual points (L_d) on the I/O functions. The variability seen in slope data was greater than the variability seen when L_d was compared across sessions, suggesting that estimates of slope have greater sensitivity to small changes in measured values. The distributions of S_1 and S_2 seen in Fig. 3 are consistent with those reported by Neely *et al.* (2009); however, repeatability of these parameters was not assessed in that study. Beattie *et al.* (2004) examined test-retest reliability of DPOAE slopes at three frequencies ($f_2 = 1$, 2, and 4 kHz) for three intensity-level segments (L_2–$L_1 = 40$–50, 50–65, and 65–75 dB) and found large intrasubject and intersubject variability in DPOAE slope. Overall, slopes obtained from DPOAE data appear to be less reliable than measurements of L_d.

Like the correlations seen with slope parameters, the parameters describing the breakpoint (L_i and L_o) had lower correlations compared to those observed for individual values of L_d on the I/O functions. However, breakpoint parameters had stronger correlations than slope parameters. Greater reliability was observed with breakpoint parameters because they represent a single point on the I/O function, and as stated previously, the slope parameters are more sensitive to small changes in measured values. It is surprising, however, that the breakpoint parameters had lower correlations than L_d. Apparently, the nonlinear fitting process has excessive sensitivity to outliers.

Differences observed in DPOAEs across test sessions may have been a result of changes in probe placement, middle-ear status, and/or noise levels. Although middle-ear status was assessed before each session, even small changes in pressure that fell within the normal range may have resulted in changes in DPOAEs (Naeve *et al.,* 1992; Zhang and Abbas, 1997). Measurements may have also been affected by stimulus calibration. Calibration was based on dB SPL measurements in the closed ear canal. Although recent data suggest that DPOAE test performance or predictions of threshold are not affected by calibration (Burke *et al.,* 2010; Rogers *et al.,* 2010), concerns have been noted for potential errors associated with standing waves when using the present calibration method (Siegel, 2002, 2007; Scheperle *et al.,* 2008; Kirby *et al.,* 2011). Therefore, it is possible that using either sound intensity level or forward pressure level during calibration may have reduced calibration errors (Neely and Gorga, 1998; Scheperle *et al.,* 2008; Kirby *et al.,* 2011; Richmond *et al.,* 2011) and resulted in greater reliability.

Overall, DPOAEs were reliably measured for both normal-hearing and hearing-impaired subjects when L_d was compared across test sessions. DPOAE reliability was greater for L_d levels > 0 dB SPL, while variability increased for L_d levels < 0 dB SPL. Reliability of DPOAEs decreased when the parameters obtained from the nonlinear, two-slope fits were compared across sessions, suggesting that this nonlinear fitting process is sensitive to outliers.

B. Relationship between CLS measurements and DPOAE I/O parameters

Cochlear compressive nonlinearity can be observed indirectly through psychoacoustic loudness tasks (Hellman and Meiselman, 1993; Buss *et al.,* 1998; Moore *et al.,* 1999) and OAE measurements (Dorn *et al.*, 2001; Epstein *et al.*, 2004; Epstein and Florentine, 2005; Withnell and Yates, 1998); therefore, it seemed reasonable to assume that DPOAEs may provide information related to behavioral response growth (loudness). In the present study, we wanted to determine if DPOAE I/O parameters could be used to predict behavioral response growth as measured by CLS.

In the same way as we described the reliability of DPOAEs, Al-Salim *et al.* (2010) described the reliability of CLS measurements by comparing both the difference in SPL assigned to each loudness category as well as the slope of the CLS functions. Overall, Al-Salim *et al.* found that CLS measurements were reliable from one session to the next. Al-Salim *et al.* also evaluated the relationship between CLS measurements and behavioral threshold, and found that loudness growth (up to medium loudness level, which corresponded to $CU = 20$) may be predicted from audiometric threshold, accounting for as much as 77% of the variance in soft slope of the CLS function.

Given the observations of Neely *et al.* (2003) and Müller and Janssen (2004), who were able to demonstrate a relationship between DPOAEs and loudness, we wanted to

further examine this relationship and determine whether DPOAEs, like behavioral threshold, could be used to predict loudness growth. Using DPOAE I/O parameters, we were able to demonstrate that DPOAEs correlate with CLS soft slope; however, the correlation was less than expected based on the relationship previously described between OAE measurements and loudness functions (Neely *et al.*, 2003; Müller and Janssen, 2004; Epstein *et al.*, 2004; Epstein and Florentine, 2005). In the present study, DPOAE I/O functions obtained from two separate visits were averaged and then fit with the nonlinear, two-slope method. In an effort to control for outliers that may impact the nonlinear fits, several analyses were performed, but regardless of the method, the correlation between DPOAE I/O parameters and CLS functions remained similar to those reported in this paper and were lower than originally expected.

Some correlation was observed between individual DPOAE I/O parameters and CLS soft slope. However, these correlations were poorer than those observed between behavioral threshold and CLS soft slope reported by Al-Salim *et al.* (2010) and reproduced in this paper. In fact, correlations between behavioral threshold and CLS soft slope were 0.82, 0.83, and 0.88 at 1, 2, and 4 kHz. Combining the DPOAE I/O parameters resulted in higher correlations (0.65, 0.74, and 0.81 at 1, 2, and 4 kHz) with CLS soft slope than individual parameters but were lower than those observed between behavioral threshold and CLS soft slope. Overall, our findings suggest that if behavioral threshold is known, it provides a better means of predicting loudness than either individual or combined DPOAE I/O parameters.

A documented relationship between audiometric thresholds and CLS functions (Lauer, 1995; Elberling, 1999; Al-Salim *et al.* 2010) suggests that behavioral-threshold measures can be used to predict growth of loudness. However, the variability observed in these studies also suggests that threshold data alone would not accurately predict response growth for every individual. In an effort to reduce the variability, we tried combining DPOAE I/O parameters and behavioral thresholds to determine if there was benefit in using both sets of measurements to predict loudness growth. Unfortunately, this provided little benefit over what was achieved when behavioral threshold was used alone to predict CLS soft slope. Thus, DPOAE I/O parameters and behavioral thresholds may be useful in predicting CLS soft slope independently, but when used collectively, little additional variance was accounted for.

C. Relationship between CLS measurements and DPOAEs using multivariate analyses

Although we were able to demonstrate a relationship between CLS measures and DPOAE I/O parameters, the strength of the relationship, based on a nonlinear fit to the slope of the DPOAE, was less than observed between behavioral thresholds and CLS slope. This is a disappointing finding because it was hoped that objective measures (like DPOAE I/O functions) could be used to predict loudness growth. In an effort to improve the predictive value of

DPOAE measurements, a multivariate analysis of DPOAE data was used to predict the CLS function.

Multivariate analyses were performed, in which all DPOAE data served as inputs, in an effort to provide an estimate of the CLS function. With this approach, actual and predicted CLS functions closely approximated each other, based on a sampling of data from six normal-hearing and six hearing-impaired subjects. Based on an analysis of the measured CLS function and the predicted CLS function using DPOAE data from all 74 subjects, SEEs suggested that it is possible to use DPOAE data to predict, on average, growth of loudness (based on CLS measures) to within one CU step. Multivariate analyses of DPOAEs have also been demonstrated to improve the accuracy with which OAEs identify auditory status for both DPOAEs (Dorn *et al.*, 1999; Gorga *et al.*, 1999, 2005; Kirby *et al.*, 2011) and TEOAEs (Hussain *et al.*, 1998; Vinck *et al.*, 1998). Our findings suggest that multivariate analyses may also be useful in efforts to predict growth of loudness from DPOAE measurements. We are surprised by this finding, given the relatively poor correlations that were observed between CLS slope and parameters from the fits to the DPOAE I/O functions. It is important to note, however, that multivariate analyses can be idiosyncratic, in that the choice of variables and coefficients may be unique to a particular set of data. As a consequence, additional studies are needed to validate the present findings.

The ability to use DPOAE measurements to predict response growth, such as growth of loudness, could be beneficial in the clinic. In cases of mild hearing loss, such information may be useful in determining individual signal-processing strategies (such as compression ratio) in hearing aids, especially for infants, young children and patients with developmental disabilities who may be unable to describe their percepts of loudness. This view is based on the assumption that both response growth and DPOAE growth depend on the same underlying compressive cochlear nonlinearity. While direct predictions of loudness from DPOAE data may be problematic, applying multivariate solutions to the same data may result in predictions that are sufficiently accurate to provide information that might be useful in setting hearing-aid characteristics. Additional work is needed to validate the multivariate approach and to determine the extent to which these approaches lead to predictions of loudness that can be used to set amplitude compression parameters for hearing aids. Even if the multivariate solutions are validated in future studies, it is likely that the application of this approach to predict loudness growth will be restricted to cases of mild-to-moderate hearing loss, as it is unlikely that DPOAEs will be present in greater degrees of hearing loss.

V. CONCLUSIONS

The correlation of repeated measurements of DPOAE level was >0.90 while lower correlations were observed between repeated estimates of parameters fit to DPOAE I/O functions. The correlation between CLS soft slope and DPOAE I/O parameters seems mainly to be due to their mutual dependence on behavioral threshold. Combining DPOAE I/O parameters with behavioral threshold provided little

additional predictive value. However, a multivariate prediction of CLS based on the entire DPOAE I/O function produced accurate predictions of CLS functions, with standard errors from 2.4 to 3.0 CUs, suggesting that DPOAE I/O functions can be useful in prediction of CLS measures.

Allen, J. B., Hall, J. L., and Jeng, P. S. (1990). "Loudness growth in 1/2-octave bands (LGOB)—A procedure for the assessment of loudness," J. Acoust. Soc. Am. 88, 745–753.

Al-Salim, S. C., Kopun, J. G., Neely, S. T., Jesteadt, W., Stiegemann, B., and Gorga, M. P. (2010). "Reliability of categorical loudness scaling and its relation to threshold," Ear Hear. 31, 567–578.

ANSI (2004). S3.6-2004, *Specifications for Audiometers* (AIP, New York).

Beattie, R. C., Kenworthy, O. T., and Luna, C. A. (2003). "Immediate and short-term reliability of distortion-product otoacoustic emissions," Int. J. Audiol. 42, 348–354.

Beattie, R. C., Kenworthy, O. T., and Neal-Johnson, C. M. (2004). "Distortion product otoacoustic emissions: Input-output functions for primary-tone pairs at 1000, 2000 and 4000 Hz in normal-hearing females," Aust. N. Z. J. Audiol. 26, 116–132.

Boege, P., and Janssen, T. (2002). "Pure-tone threshold estimation from extrapolated distortion product otoacoustic emissions 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.

Burke, S., Rogers, A. R., Neely, S. T., Kopun, J. G., Tan, H., and Gorga, M. P. (2010). "Influence of calibration method on distortion-product otoacoustic emission measurements: I. test performance," Ear Hear. 31, 533–545.

Buss, S., Musch, H., and Florentine, M. (1998). "On loudness at threshold," J. Acoust. Soc. Am. 104, 399–410.

Cai, S., Ma, W. D., and Young, E. D. (2009). "Encoding intensity in ventral cochlear nucleus following acoustic trauma: Implications for loudness recruitment," J. Assoc. Res. Otolaryngol. 10, 5–22.

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 and impaired ears," J. Acoust. Soc. Am. 110, 3119–3131.

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.

Elberling, C. (1999). "Loudness scaling revisited," J. Am. Acad. Audiol. 10, 248–260.

Epstein, M., Buus, S., and Florentine, M. (2004). "The effects of window delay, delinearization, and frequency on tone-burst otoacoustic emission input/output measurements," J. Acoust. Soc. Am. 116, 1160–1167.

Epstein, M., and Florentine, M. (2005). "Inferring basilar-membrane motion from tone-burst otoacoustic emissions and psychoacoustic measurements," J. Acoust. Soc. Am. 117, 263–274.

Epstein, M., and Silva, I. (2009). "Analysis of parameters for the estimation of loudness from tone-burst otoacoustic emissions," J. Acoust. Soc. Am. 125, 3855–3864.

Fletcher, H., and Munson, W. (1933). "Loudness: Its definition, measurement, and calculations," J. Acoust. Soc. Am. 9, 1–10.

Franklin, D. J., McCoy, M. J., Martin, G. K., and Lonsbury-Martin, B. L. (1992). "Test/retest reliability of distortion-product and transiently evoked otoacoustic emissions," Ear Hear. 13, 417–429.

Goldman, B., Sheppard, L., Kujawa, S. G., and Seixas, N. G. (2006). "Modeling distortion product otoacoustic emission input/output functions using segmented regression," J. Acoust. Soc. Am. 120, 2764–2776.

Gorga, M. P., Dierking, D. M., Johnson, T. A., Beauchaine, K. L., Garner, C. A., and Neely, S. T. (2005). "A validation and potential clinical application of multivariate analyses of distortion-product otoacoustic emission data," Ear Hear. 26, 593–607.

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

Gorga, M. P., Neely, S. T., and Dorn, P. A. (1999). "Distortion product otoacoustic emission test performance for a priori criteria and for multifrequency audiometric standards," Ear Hear. 20, 345–362.

Gorga, M. P., Neely, S. T., Dorn, P. A., and Hoover, B. M. (2003). "Further efforts to predict pure-tone thresholds from distortion product otoacoustic emission input/output functions," J. Acoust. Soc. Am. 113, 3275–3284.

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.

Heinz, M. G., Issa, J. B., and Young, E. D. (2005). "Auditory-nerve rate responses are inconsistent with common hypotheses for the neural correlates of loudness recruitment," J. Assoc. Res. Otolaryngol. 6, 91–105.

Hellman, R. P., and Meiselman, C. H. (1993). "Rate of loudness growth for pure tones in normal and impaired hearing," J. Acoust. Soc. Am. 93, 966–975.

Hussain, D. M., Gorga, M. P., Neely, S. T., Keefe, D. H., and Peters, J. (1998). "Transient evoked otoacoustic emissions in patients with normal hearing and in patients with hearing loss," Ear Hear. 19, 434–449.

Johnson, T. A., Neely, S. T., Garner, C. A., and Gorga, M. P. (2006a). "Influence of primary-level and primary-frequency ratios on human distortion product otoacoustic emissions," J. Acoust. Soc. Am. 119, 418–428.

Johnson, T. A., Neely, S. T., Kopun, J. D., and Gorga, M. P. (2006b). "Reducing reflected contributions to ear-canal distortion product otoacoustic emissions in humans," J. Acoust. Soc. Am. 119, 3896–3907.

Keppler, H., Dhooge, I., Maes, L., D'haenens, W., Bockstael, A., Philips, B., Swinnen, F., and Vinck, B. (2010). "Transient-evoked and distortion product otoacoustic emissions: A short-term test-retest reliability study," Int. J. Audiol. 49, 99–109.

Kirby, B. J., Kopun, J. G., Hongyang, T., Neely, S. T., and Gorga, M. P. (2011). "Do "optimal" conditions improve distortion product otoacoustic emission test performance?," Ear Hear. 32, 230–237.

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

Lasky, R. E., Snodgrass, E., and Hecox, K. (1994). "Distortion-product otoacoustic emission input/output functions as a function of frequency in human adults," J. Am. Acad. Audiol. 5, 183–194.

Lauer, S. (1995). "Loudness perception in listeners with sensorineural hearing impairment," Ph.D. thesis, Oldenburg University, Oldenburg, Germany.

Lonsbury-Martin, B. L., and Martin, G. K. (1990). "The clinical utility of distortion product otoacoustic emissions," Ear Hear. 11, 144–154.

Moore, B. C. J., and Glasberg, B. R. (1997). "A model of loudness perception applied to cochlear hearing loss," Aud. Neurosci. 3, 289–311.

Moore, B. C. J., Glasberg, B. R., and Vickers, D. A. (1999). "Further evaluation of a model of loudness perception applied to cochlear hearing loss," J. Acoust. Soc. Am. 106, 898–907.

Müller, J., and Janssen, T. (2004). "Similarity in loudness and distortion product otoacoustic emission input/output functions: Implications for an objective hearing aid adjustment," J. Acoust. Soc. Am. 115, 3081–3091.

Naeve, S. L., Margolis, R. H., Levine, S. C., and Fournier, E. M. (1992). "Effect of ear canal pressure on evoked otoacoustic emissions," J. Acoust. Soc. Am. 91, 2091–2095.

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

Neely, S. T., Gorga, M. P., and Dorn, P. A. (2003). "Cochlear compression estimates from measurements of distortion-product otoacoustic emissions," J. Acoust. Soc. Am. 114, 1499–1507.

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

Neely, S. T., Johnson, T. A., Kopun, J. G., Dierking, D. M., and Gorga, M. P. (2009). "Distortion-product otoacoustic emission input/output characteristics in normal-hearing and hearing-impaired human ears," J. Acoust. Soc. Am. 126, 728–738.

Neely, S. T., and Liu, Z. (1994). "EMAV: Otoacoustic emission averager," Technical Memo No. 17, Boys Town National Research Hospital, Omaha, NE.

1294   J. Acoust. Soc. Am., Vol. 131, No. 2, February 2012

Thorson *et al.*: Otoacoustic emissions and loudness

Nelson, D. A., and Kimberley, B. P. (**1992**). "Distortion-product emissions and auditory sensitivity in human ears with normal hearing and cochlear hearing loss," J. Speech Hear. Res. **35**, 1141–1159.

Oswald, J. A., and Janssen, T. (**2003**). "Weighted DPOAE input/output functions: A tool for automatic assessment of hearing loss in clinical application," Z. Med. Phys. **13**, 93–98.

Rhode, W. S. (**1971**). "Observations of the vibration of the basilar membrane in squirrel monkey using the Mössbauer effect," J. Acoust. Soc. Am. **49**, 1218–1231.

Richmond, S. A., Kopun, J. G., Neely, S. T., Tan, H., and Gorga, M. P. (**2011**). "Distribution of standing-wave errors in real-ear sound-level measurements," J. Acoust. Soc. Am. **129**, 3134–3140.

Robles, L., and Ruggero, M. A. (**2001**). "Mechanics of the mammalian cochlea," Physiol. Rev. **81**, 1305–1352.

Roede, J., Harris, F. P., Probst, R., and Xu, L. (**1993**). "Repeatability of distortion product otoacoustic emissions in normally hearing humans," Audiology **32**, 273–281.

Rogers, A. R., Burke, S., Tan, H., Neely, S. T., and Gorga, M. P. (**2010**). "Influence of calibration method on distortion-product otoacoustic emission measurements: II. Threshold predication," Ear Hear. **31**, 546–554.

Ruggero, M. A., and Rich, N. C. (**1991**). "Furosemide alters organ of corti mechanics: Evidence for feedback of outer hair cells upon the basilar membrane," J. Neurol. **11**, 1057–1067.

Ruggero, M. A., Rich, N. C., and Recio, A. (**1996**). "The effect of intense acoustic stimulation on basilar-membrane vibrations," Aud. Neurosci. **2**, 329–345.

Scharf, B. (**1978**). "Loudness," in *Handbook of Perception: Vol IV. Hearing*, edited by E. C. Carterette and M. P. Friedman (Academic Press, New York), pp. 187–242.

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

Siegel, J. H. (**2002**). "Calibrating otoacoustic emission probes," in *Otoacoustic Emissions: Clinical Application*, 2nd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical Publishers, New York), pp. 416–441.

Siegel, J. H. (**2007**). "Calibrating otoacoustic emission probes," in *Otoacoustic Emissions: Clinical Application*, 3rd ed., edited by M. S. Robinette and T. J. Glattke (Thieme Medical Publishers, New York), pp. 403–427.

Stuart, A., Passmore, A. L., Culbertson, D. S., and Jones, S. M. (**2009**). "Test-retest reliability of low-level evoked distortion product otoacoustic emissions," J. Speech Hear. Res. **52**, 671–681.

Vinck, B. M., Cauwenberge, P. B., Corthals, P., and Vel, E. D. (**1998**). "Multi-variant analysis of otoacoustic emissions and estimation of hearing threshold: Transient evoked otoacoustic emissions," Audiology **37**, 315–334.

Wagner, W., Heppelmann, G., Vonthein, R., and Zenner, H. P. (**2008**). "Test-retest repeatability of distortion product otoacoustic emissions," Ear Hear. **29**, 378–391.

Withnell, R. H., and Yates, G. K. (**1998**). "Onset of basilar membrane nonlinearity reflected in cubic distortion tone input-output functions," Hear. Res. **123**, 87–96.

Zhang, M., and Abbas, P. J. (**1997**). "Effects of middle ear pressure on otoacoustic emission measures," J. Acoust. Soc. Am. **102**, 1032–1037.

Zhang, M., and Zwislocki, J. J. (**1995**). "OHC response recruitment and its correlation with loudness recruitment," Hear. Res. **85**, 1–10.