Cochlear Reflectance and Otoacoustic Emission Predictions of Hearing Loss

Stephen T. Neely, Sara E. Fultz, Judy G. Kopun, Natalie M. Lenzen, and Daniel M. Rasetshwane

Objectives: Cochlear reflectance (CR) is the cochlear contribution to ear-canal reflectance. CR is a type of otoacoustic emission (OAE) that is calculated as a transfer function between forward pressure and reflected pressure. The purpose of this study was to compare wideband CR to distortion-product (DP) OAEs in two ways: (1) in a clinical-screening paradigm where the task is to determine whether an ear is normal or has hearing loss and (2) in the prediction of audiometric thresholds. The goal of the study was to assess the clinical utility of CR.

Design: Data were collected from 32 normal-hearing and 124 hearing-impaired participants. A wideband noise stimulus presented at 3 stimulus levels (30, 40, 50 dB sound pressure level) was used to elicit the CR. DPOAEs were elicited using primary tones spanning a wide frequency range (1 to 16 kHz). Predictions of auditory status (i.e., hearing-threshold category) and predictions of audiometric threshold were based on regression analysis. Test performance (identification of normal versus impaired hearing) was evaluated using clinical decision theory.

Results: When regressions were based only on physiological measurements near the audiometric frequency, the accuracy of CR predictions of auditory status and audiometric threshold was less than reported in previous studies using DPOAE measurements. CR predictions were improved when regressions were based on measurements obtained at many frequencies. CR predictions were further improved when regressions were performed on males and females separately.

Conclusions: Compared with CR measurements, DPOAE measurements have the advantages in a screening paradigm of better test performance and shorter test time. The full potential of CR measurements to predict audiometric thresholds may require further improvements in signal-processing methods to increase its signal to noise ratio. CR measurements have theoretical significance in revealing the number of cycles of delay at each frequency that is most sensitive to hearing loss.

Key words: Cochlear reflectance, Distortion product otoacoustic emissions, Sensorineural hearing loss.

(Ear & Hearing 2018;XX;00–00)

INTRODUCTION

Otoacoustic emissions (OAEs) are acoustic signals that originate within the cochlea as by-products of its normal function (Kemp 2002). OAEs are generated within the cochlea by either (1) intermodulation distortion due to nonlinearity in outer hair cells (OHCs) or (2) wave reflection due to slight irregularities in the mechanical properties of the basilar membrane (Shera & Guinan 1999). Both of these mechanisms are dependent on the viability of OHCs (Brownell 1990) and both generate retrograde pressure waves that travel toward the base of the cochlea, through the middle ear, and into the ear canal where they may be recorded by a microphone. In some individuals, OAEs are produced spontaneously (SOAEs), in the absence of a stimulus (Burns et al. 1992). In most normal-hearing (NH) ears, OAEs can be evoked using one of several different types of stimuli. Click-evoked OAEs (CEOAEs) are measured using clicks and, thus, provide information for a wide range of frequencies. Tone burst–evoked OAEs (TBOAEs) are measured using short-duration sinusoids, and thus, provide information that spans the frequency range of the tone burst stimulus. CEOAEs and TBOAEs are examples of transient-evoked OAEs. Band-limited noise has been used to evoke OAEs that mimic CEOAEs (Maat et al. 2000). Stimulus-frequency OAEs (SFOAEs) typically use sinusoidal stimuli and provide information from a narrow frequency range around the frequency of the stimulus. Distortion-product OAEs (DPOAEs) are evoked by a pair of primary tones (f_1 and f_2) that interact within the cochlea to produce other frequencies that are arithmetic combinations of the primary tones, such as $2f_1 - f_2$. DPOAEs are mostly generated near the cochlear place that responds best to f_2, and so provide information about that frequency, but contributions to DPOAEs from other places may not be negligible.

Damage to the OHCs results in a reduction in OAEs (Brownell 1990). In several studies, relationships have been demonstrated between audiometric threshold and OAE level. For example, such relationships have been observed for DPOAEs (Gorga et al. 1993; Stover et al. 1996; Boege & Janssen 2002; Johnson et al. 2010; Kirby et al. 2011), CEOAEs (Gorga et al. 1993; Prieve et al. 1993; Hussain et al. 1998; Goodman et al. 2009; Mertes & Goodman 2013), SFOAEs (Ellison & Keefe 2005), and TBOAEs (McPherson et al. 2006; Jedrzejczak et al. 2012). OAEs are clinically useful, including in newborn hearing screening, because of their noninvasive nature. The most commonly used OAE types in the clinic are DPOAEs and CEOAEs. Clinical benefits of OAE measurement includes the short test time and low cost. OAEs can provide an assessment of hearing status when behavioral responses are unobtainable or a cross-check for validity when behavioral responses are unreliable. Furthermore, subclinical reduction in OAEs may be observed before seeing permanent threshold shifts in the audiogram.

Cochlear reflectance (CR) is an alternative measure of cochlear response that was suggested by Allen (1997). Specifically, CR is the cochlear contribution to ear-canal reflectance (ECR)*. Alternatively, CR may be defined by a transfer function between the forward-propagating component of pressure measured in the ear canal and the component of the reflected pressure that comes from the cochlea. The CR properties that allow its separation from contributions that come from the middle ear, ear canal, and measurement system are its longer latency and dependence on stimulus level (Rasetshwane & Neely 2012; Rasetshwane et al. 2015). Our procedure for measuring CR is described in the Materials and Methods section. The use of

Center for Hearing Research, Boys Town National Research Hospital, Omaha, Nebraska, USA.

* ECR is the complex ratio of reflected acoustic pressure to forward pressure measured in the ear canal (Voss & Allen 1994). ECR has been shown to differentiate between normal ears and ears with specific middle ear disorders (Voss et al. 2012).

0196/0202/2018/XXXX-00/0 • Ear & Hearing •

<zdoi; 10.1097/AUD.0000000000000677>

TABLE 1. Distribution of auditory status by frequency

Threshold (dB HL)	Frequency (kHz)										
	0.5	0.75	1	1.5	2	3	4	6	8	11.2	16
≤20	95	95	86	80	72	53	55	50	49	30	8
>20	61	61	70	76	84	103	100	103	101	113	79
NR	0	0	0	0	0	0	1	3	6	13	69

The number of participants is given for each threshold category at each frequency. NR indicates the number of participants who did not hear the tone at the highest level that the audiometer could produce. These participants were not included in any frequency-specific analysis.

wideband noise (WBN) to evoke CR prevents the entrainment of SOAEs (Harte & Elliott 2005). In addition, the use of a WBN allows for the invocation of de Boer's (1997) nonlinear equivalence (EQ-NL) theorem. According to the EQ-NL theorem, for a given class of nonlinear systems (of which the cochlea is an example), there is an equivalent linear system that has the same WBN response. Having a stimulus and a system (i.e., the cochlea) that conform to the EQ-NL theorem is advantageous because it supports inferences based on a linear model. Thus, a potential advantage of WBN CR measurements over other types of OAEs is an interpretation based on a linear cochlear model.

Clinical decision theory (Swets 1988; Fawcett 2006) has been used previously to test the accuracy of dichotomous decisions regarding auditory status based on OAE measurements (Gorga et al. 1993; Prieve et al. 1993; Stover et al. 1996; Boege & Janssen 2002; Ellison & Keefe 2005; Goodman et al. 2009; Johnson et al. 2010). Our use of multivariate analyses parallels previous efforts in which these techniques were applied to CEOAEs (Hussain et al. 1998; Mertes & Goodman 2013), DPOAEs (Dorn et al. 1999; Gorga et al. 2005; Kirby et al. 2011), and SFOAEs (Ellison & Keefe 2005).

This study extends our previous study on the relationship between CR and auditory status (Rasetshwane et al. 2015). The main goals of this study were to (1) compare the accuracy of CR and DPOAEs in terms of identification of auditory status (i.e., normal versus impaired), which we refer to as test performance and (2) compare the accuracy of CR and DPOAEs in their prediction of audiometric thresholds. Another useful result of this study was refinement of the region in time-frequency space in which CR appears to be most sensitive to hearing threshold.

MATERIALS AND METHODS

Participants

A total of 156 participants were enrolled in this study (91 females and 65 males). All participants were recruited from a database of potential research participants that is maintained by Boys Town National Research Hospital. Participants were paid for their participation. Data collection was conducted under a protocol that had been approved by the Boys Town National Research Hospital's Institutional Review Board and informed consent was obtained before testing each participant.

Audiometric thresholds were measured at 12 octave and interoctave frequencies (0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6, 8, 11.2, and 16 kHz) using Sennheiser HDA 300 headphones (Wedemark, Germany). Thresholds were measured by an audiometer (GSI AudioStar Pro, Grason-Stadler) following a Modified Hughson-Westlake procedure that used a 2-dB step size (instead of the usual 5 dB) for the final determination of threshold. The range of observed thresholds was −10 to 102 dB hearing level (HL); however, some participants had no response at higher frequencies due to equipment limitations. The headphones were calibrated by a commercial service (Audiology Systems Inc.). The maximum output of the audiometer was 94 dB HL at 11.2 kHz and 60 dB HL at 16.0 kHz. All participants had normal middle ear function based on otoscopic inspection, tympanometry, and bone conduction threshold. The inclusion criteria for tympanometry (Otoflex 100, Madsen) required a peak-compensated static acoustic admittance between 0.3 and 2.5 mmhos and peak tympanometric pressure between −100 and +50 daPa. The upper limit of the admittance range is higher than typically used but was selected for consistency with previous studies in our laboratory (Rasetshwane et al. 2015; Sieck et al. 2016). The inclusion criterion for bone conduction (B71, Radioear) required air-bone gaps ≤10 dB.

The auditory status at each frequency was said to be hearing impaired (HI) when the threshold at that frequency was >20 dB HL. Otherwise, the auditory status was said to be NH. The number of HI and NH participants at each frequency is shown in Table 1. An effort was made to represent a wide range of thresholds. Table 2 shows the number of participants at each frequency that had thresholds in each of 5 ranges of hearing loss: ≤10, >10 to ≤30, >30 to ≤50, >50 to ≤70, and >70 dB HL. Note that the midpoints of these threshold ranges (with approximations at the highest and lowest ranges) are 0, 20, 40, 60, and 80 dB HL.

TABLE 2. Distribution of hearing threshold by frequency

Threshold (dB HL)		Frequency (kHz)										
Range	Midpoint	0.5	0.75	1	1.5	2	3	4	6	8	11.2	16
≤10	0	54	59	58	62	49	33	36	36	34	23	5
>10 to 30	20	62	54	50	40	42	45	33	28	25	14	9
>30 to 50	40	22	20	27	26	36	41	41	30	31	17	10
>50 to 70	60	12	17	15	19	15	23	30	39	37	37	57
>70	80	6	6	6	9	14	14	15	20	23	52	6

The number of participants is given for each hearing-threshold category at each frequency. Each category is defined by its range; however, its midpoint provides a more convenient label.

Participants were classified as overall HI (OHI) when their thresholds were >20 dB HL for at least 1 frequency in the range from 1 to 8 kHz. There were 124 OHI participants (72 females and 52 males). The average age for OHI participants was 60.2 years (range = 19 to 75 years, SD = 12.3 years). Participants who were not OHI were classified as overall NH (ONH). There were 32 ONH participants (19 females and 13 males). The average age for ONH participants was 39.1 years (range = 19 to 66 years, SD = 12.8 years).

All CR measurements were made monaurally. If both ears met the inclusion criteria in ONH participants, the ear with better hearing was selected for testing. Thresholds at all frequencies were considered when determining the better ear. If thresholds differed by more than about 10 dB at more than one frequency, then the ear with better thresholds was selected. Otherwise, an ear was selected randomly, although there was some attempt to equalize the number of left and right ears. If both ears met the inclusion criteria in OHI participants, the ear with audiometric thresholds in the mild-to-moderate range above 1 kHz was tested to increase the likelihood of a CR response above the noise floor. If both ears had thresholds in this range, the test ear was selected randomly. Only 1 ear of each participant was tested. In all, data were collected from 17 left and 15 right ears of ONH participants and 64 left and 60 right ears of OHI participants.

Equipment

The sound-delivery system for CR and DPOAE data collection was an ER-10X probe-microphone system (Etymōtic Research, Elk Grove Village, IL) and a 24-bit sound card (Hammerfall DSP Multi-Face II, RME, Germany). For CR, each stimulus condition was presented on each of two receivers (which we will refer to as sound sources) consecutively, providing two independent measurements. The ER-10X is well suited for measuring CR at high frequencies because it has a wide bandwidth, low distortion, and low cross-talk compared with other emission probes. The measurement system was calibrated daily before data collection to determine the Thévenin-equivalent source impedance and pressure (Allen 1986; Keefe et al. 1992; Rasetshwane & Neely 2011). Thévenin-equivalent source parameters are required in the transformation of ear-canal pressure to CR and in calculation of forward pressure level (FPL) used in DPOAE stimulus calibration. Evanescent waves in the ER-10X calibration waveguide were reduced by modifying the multilumen tubes that are inserted into the ER-10X eartips such that the front surface was beveled at approximately a 45° angle instead of the normally perpendicular front surface (Siegel & Neely 2017). Further compensation for the possible presence of evanescent waves was included in the calculation of the Thévenin-equivalent source parameters (Nørgaard et al. 2017). Stimulus delivery and data collection were controlled using locally developed software (EMAV v 3.33; Neely & Liu 1994).

Measurement of CR

The CR stimuli were (1) a WBN and (2) a wideband linear -sweep chirp (LSC), both digitally generated at a sampling rate of 48 kHz. The LSC and WBN digital stimuli both contained equal levels of all frequencies from 0 to 24 kHz. The duration of each stimulus/response buffer was 171 msec. Data were collected in response to the LSC stimulus at 60 dB sound pressure level (SPL) and in response to the WBN stimulus presented at 3 levels (30, 40, 50 dB SPL). These stimulus levels were determined using a sound level meter (System 824, Larson Davis, Provo, UT) with C weighting. Following Rasetshwane and Neely (2012), longer averaging times were used at the lower stimulus levels to improve the signal to noise ratio (SNR) of the measured response. The stimulus conditions are summarized in Table 3. Artifact rejection based on the root-mean-square (RMS) amplitude of the recordings of time-domain pressure was utilized. In this procedure, if two consecutive sweeps differed by a criterion amount, both sweeps were rejected, a procedure that is effective for rejecting transient artifacts. In a single visit, data collection using the complete set of stimulus conditions was repeated three times on each of the two sound sources. The three sets of responses were combined into a single average for all analyses. The measurement probe was not reseated at the beginning of each set of measurements unless the previous measurement showed evidence of an acoustic leak (Groon et al. 2015). The test times listed in Table 3 only include artifact-free responses. Additional time was required for (1) artifact rejection, (2) software inefficiency, and (3) probe reinsertion to correct air leaks (or other issues). CR data-collection time was, on average, 35 min per session.

TABLE 3. Stimulus type, level, number of averages per sound source, and test time for CR measurements

Stimulus Condition	Type	Level (dB SPL)	Number of Averages	Test Time (s)
1	WBN	30	32	96
2	WBN	40	16	48
3	WBN	50	8	24
4	LSC	60	2	6

The stimulus type was either WBN or LSC. Note that the test time listed only includes responses that did not contain artifacts.
CR, cochlear reflectance; LSC, linear-sweep chirp; WBN, wideband noise.

CR measurements were made following procedures described by Rasetshwane et al. (2015). Thévenin-equivalent source parameters for the ER-10X were determined before data collection. The source parameters allow ear-canal impedance to be computed from the pressure recorded in the ear canal in response to each stimulus (LSC and WBN). ECR Γ_{ec} is calculated from ear-canal impedance Z_{ec} and the characteristic impedance of the ear canal Z_0, which can be estimated from Z_{ec} (Scheperle et al. 2011):

$$\Gamma_{ec} = \frac{Z_{ec} - Z_0}{Z_{ec} + Z_0}. \quad (1)$$

The ECR in response to the LSC stimulus is expected to have little, if any, contribution from the cochlea for two reasons: (1) the LSC stimulus level is at least 10 dB higher than the WBN stimuli and (2) the chirp is more frequency specific within short time intervals, which is expected to produce larger local responses within the cochlea that are more likely to saturate OHC feedback forces. CR Γ_c is calculated as the difference between ECR in response to the WBN stimulus $\Gamma_{ec,wbn}$ and ECR in response to the LSC stimulus $\Gamma_{ec,lsc}$:

$$\Gamma_c = \Gamma_{ec,wbn} - \Gamma_{ec,lsc}. \quad (2)$$

A time-domain representation of CR was computed as the inverse Fourier transform of Γ_c. To allow removal of residual ear-canal and middle ear contributions, a time-frequency analysis was performed by using a complex gammatone filterbank with 49 channels (Patterson & Holdsworth 1996; Hohmann 2002). The center frequencies were logarithmically spaced from $f_c = 0.0625$ to 16 kHz in 1/6-octave steps, and the tuning of each filter was $Q_{erb} = 3$, where Q_{erb} is defined as the center frequency divided by the equivalent rectangular bandwidth of the filter (Shera et al. 2010). The focus of our algorithm on the extraction of signal properties that are characteristic of the cochlea increases the likelihood that the CR we observe is of biologic origin; however, observation of its sensitivity to hearing-threshold category provides additional evidence for this conclusion. For additional theoretical and measurement details, please see Rasetshwane and Neely (2012).

To improve SNR, the CR magnitude (CRM) at each frequency was calculated over a time span that was defined by (1) minimum and maximum time-delays (τ_H and τ_L) and (2) minimum and maximum cycle-delays (N_L and N_H). The region in time-frequency space that is delimited by these four parameters is illustrated in Figure 1. The values of these four analysis parameters were adjusted for the present study before the reported data analysis to yield better test performance, which was defined as having larger receiver operating characteristic (ROC) area averaged over all frequencies and all participants. Adjustments were made by increasing or decreasing each parameter one at a time (manually) in small increments, which was a time-consuming and somewhat inefficient process. Subsequent to this adjustment, the same parameter values were used to quantify all CR data reported here.

The steps outlined above for the calculation of CR were carried out independently for measurements from the two sound sources. The mean of the two measurements was used to estimate the CRM and the mean of the difference between the two measurements was used to estimate the CR noise (CRN). CR SNR was subsequently calculated using CRM and CRN.

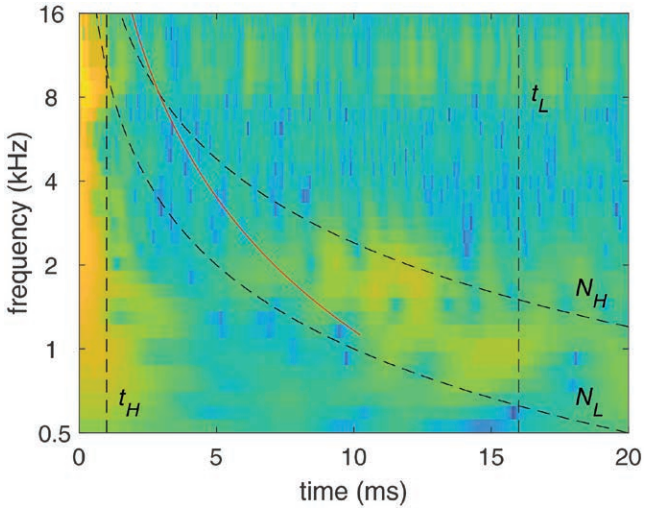

Figure 1. Time-frequency representation of cochlear reflectance. The region that was determined to be most sensitive to hearing loss is delimited (dashed lines) by (1) minimum and maximum time-delays (τ_H and τ_L) and (2) minimum and maximum cycle-delays (N_L and N_H). For comparison, estimated stimulus-frequency otoacoustic emission delay (Shera & Guinan 2003) is superimposed (solid line).

Measurement of DPOAEs

The same measurement system was used for DPOAE and CR data collection. For DPOAE data collection, the two primary tones, f_1 and f_2 were routed through separate channels of the sound card and presented from separate sound sources of the Etymotic ER-10X probe-microphone system. The buffer length was 8192 samples and the sampling rate was 32 kHz. This resulted in a 3.9 kHz frequency resolution. DPOAE responses were collected in two separate buffers. The level of the DPOAE was obtained by summing the contents of the two buffers in the frequency bin containing the $2f_1 - f_2$ frequency. The noise was calculated by averaging the level in the $2f_1 - f_2$ frequency bin and the five bins on either side of the $2f_1 - f_2$ frequency bin for each buffer and then subtracting the noise estimates in the two buffers.

Before DPOAE measurement, in-situ calibration was used to determine stimulus levels. To minimize the effects of standing waves, all stimuli were specified using FPL rather than SPL (Neely & Gorga 1998; Scheperle et al. 2008). In FPL measurements, the forward wave is isolated from the reflected wave thus preventing inaccuracies in specifying the level of the stimulus at the eardrum that can result from destructive or constructive interactions of the forward and reflected waves in the ear canal. To specify stimulus level in FPL, Thévenin-equivalent source characteristics (impedance and pressure) were determined using the techniques previously described in the Equipment section. Then, in-situ sound pressure measurements were made in each participant's ear canal. Those measurements, combined with source impedance and pressure, were used to calculate ear-canal impedance. Then ear-canal impedance, ear-canal pressure, and source impedance were used to convert SPL to FPL. The EMAV software then adjusted the voltage applied to the sound sources to produce the desired level at the eardrum.

DPOAEs were elicited using the primary frequencies, f_1 and f_2. The primary frequency ratio (f_2/f_1) was set to 1.22. The following equation was used to set the level of f_1 (L_1) for each level of f_2 (L_2; Kirby et al. 2011):

$$L_1 = 80 + 0.1 \cdot (L_2 - 80) \cdot \log_2(64 / f_2) \quad (3)$$

Data collection continued until 1 of 2 measurement-based stopping rules were met: (1) the noise floor reached a level less than or equal to −20 dB SPL or (2) 32 sec of artifact-free averaging time was completed. The noise-floor stopping rule was chosen to ensure that DPOAE levels were above the level at which system distortion could compromise the data. The measurement-time stopping rule ensured that the total time did not exceed what was considered an acceptable amount of time for data collection. DPOAE data were collected for a single L_2 of 55 dB FPL for 9 f_2 frequencies (1, 1.414, 2, 2.828, 4, 5.656, 8, 1.1312, 16 kHz).

Analyses

Single-frequency analyses utilized ear-canal measurements at a single frequency for predictions of either auditory status (i.e., greater or less than 20 dB HL) or auditory threshold at the matching audiometric frequency. Logistic regressions were used for auditory-status predictions and linear regressions for auditory-threshold predictions. For DPOAE measurements, the two predictor variables were DPOAE response level and DPOAE noise level. For CR measurements, the three predictor variables were ECR, CRM, and CRN. Because CR test performance was similar across the three stimulus levels, only one representative level is

reported. The representative level was selected to be 50 dB SPL, the highest level, because this level had the smallest p value for threshold when a three-way analysis of variance (ANOVA; frequency, threshold, sex) was performed on each level separately†.

Clinical decision theory (Swets 1988; Fawcett 2006) was used for the assessment of test performance (identification of normal versus impaired hearing). A description of CR test performance was obtained by computing hit rates (sensitivity), which is the proportion of HI ears that were correctly identified, and corresponding false-alarm rates (1 − specificity), which is the proportion of NH ears incorrectly identified as HI. The assignment of hearing category (NH or HI) was conducted on a frequency-by-frequency basis. The number of ears classified as NH and HI at each frequency are provided in Table 1. ROC curves (plots of hit rate versus false-alarm rate) were constructed, and the area under each ROC curve (A_{ROC}) was computed. A_{ROC} provides a single estimate of hit rate averaged over all possible false-alarm rates. A value of $A_{ROC} = 0.5$ indicates that hit and false-alarm rates are equal (chance performance), while a value of $A_{ROC} = 1.0$ indicates that the hit rate is 100% for all false-alarm rates, including a false-alarm rate of 0% (perfect test performance). A_{ROC} values were calculated for all audiometric frequencies at the stimulus level with the smallest RMS dB prediction error.

Multifrequency analyses incorporated ear-canal measurements at many frequencies to predict the auditory status or auditory threshold at each audiometric frequency. The prediction of auditory threshold (dB HL) used an ordinal five-category logistic regression model (mnrfit, MATLAB). Multifrequency analyses are reported only for CR measurements. The 5 threshold categories were the same as described in Table 2: ≤10, >10 to ≤30, >30 to ≤50, >50 to ≤70, >70. The CR variables included in the regression were ECR, CRM at the highest two levels and CRN at the highest two levels. ECR was quantified as a response to the LSC stimulus, which was only presented at one level. The inclusion of multiple levels of CRM and CRN allowed representation in the regression of growth rate. Each of these 5 variables was represented at 10 frequencies, which were half-octave intervals from 0.7 to 16 kHz, for a total of 50 predictor variables. Because some regressions failed to converge with this large number of predictor variables, which was not unexpected, a principal component analysis (Jolliffe 2002) was performed and regressions were subsequently based on only the 15 most significant components. This number of components was selected because it is less than a tenth of the number of participants (156), which reduces the risk of overfitting the data. Another advantage of using principal component analysis is that it avoids problems with multiple collinearity in the data because the principal components are orthogonal by definition. Probabilities were obtained for each of the five hearing-threshold categories at each frequency. The algorithm for obtaining median threshold predictions and interquartile ranges (IQR) was to compute a cumulative sum of probabilities across categories and then perform linear interpolation between category boundaries to estimate thresholds associated with 25th, 50th, and 75th percentiles.

RESULTS

DP and CR Levels

Figure 2 shows the dependence of CR and DPOAE measurements on hearing loss. Each symbol represents an average

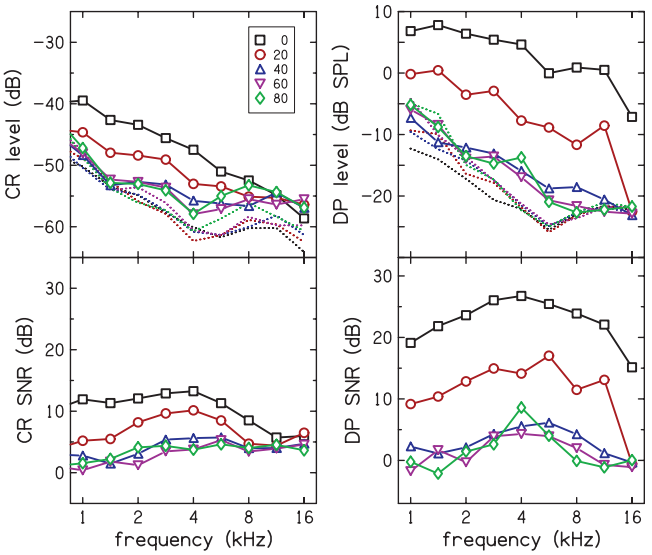

Figure 2. CR and DPOAE measurements. Response levels (upper panels) and SNR (lower panels) of CR (left panels) and DPOAE (right panels) measurements averaged across participants in five hearing-threshold categories. The dotted lines in the upper two panels represent the corresponding noise levels. CR is shown for the 50 dB SPL stimulus level. The legend identifies the hearing-threshold categories by their midpoints in dB HL. CR indicates cochlear reflectance; SNR, signal to noise ratio.

over all participants in a particular hearing-threshold category. For brevity, the hearing-threshold categories are labeled in the legend by their midpoints. CR is shown only for the highest stimulus level (50 dB SPL). The lower stimulus levels produced higher response levels but had similar frequency dependence and similar SNR. The two upper panels show a trend for the response levels of both CR (left) and DPOAE (right) to decrease with either increasing hearing-threshold category or increasing frequency. The lower panels show a similar trend for CR and DPOAE SNR to decrease with hearing-threshold category, but the SNR was greater in the midfrequencies due to the spectrum of the noise floor. The SNR for DPOAE was higher than the SNR for CR for participants in the first 2 hearing-threshold categories (0 and 20 dB HL). The SNRs for the other hearing-threshold categories were close to zero for both CR and DPOAEs. Overall, DPOAE level provided a better separation of hearing-threshold categories and had higher SNR than CR.

The dotted lines in the upper two panels of Figure 2 represent the noise levels for each of the hearing-threshold categories. There appears to be a trend for the CRN level to increase with hearing-threshold category at frequencies below 1.5 kHz and above 4 kHz. Also, there is a trend for DPOAE noise to increase with hearing-threshold category at frequencies below 4 kHz. The measurement-based stopping rule for noise level prevented observation of this trend for DPOAE measurements at frequencies above 3 kHz. Although the tendency for noise levels to increase with hearing-threshold category has been observed previously (Rasetshwane and Neely 2012), the reasons for this trend are not understood.

Figure 3 replots the data from Figure 2 to show the sensitivity of CR and DPOAE measurements to hearing-threshold category. Sensitivity is defined as the difference in CR or DPOAE measures between adjacent hearing-threshold categories divided by 20 dB, which is the difference between the

† In general, an ANOVA partitions the total variance in a data set into the variance contributed by each of the factors and the p value indicates the probability that the variance contributed by a particular factor occurred by chance.

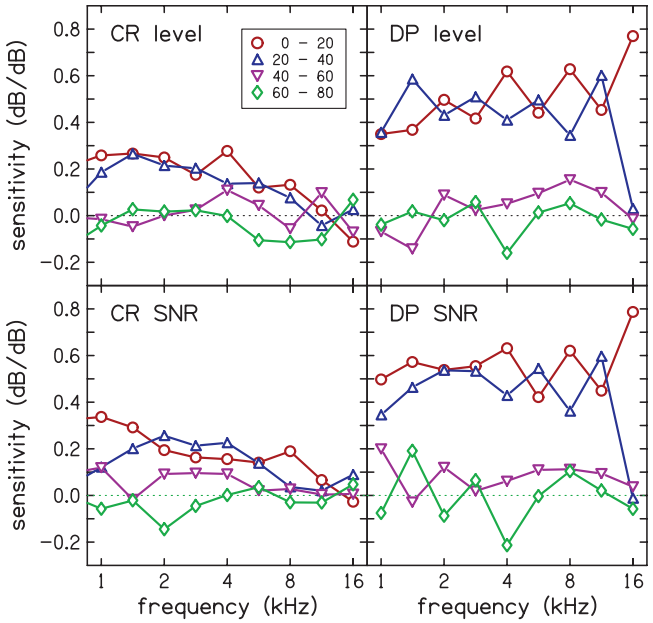

Figure 3. Sensitivity to hearing threshold for response levels (upper panels) and SNR (lower panels) of CR (left panels) and DPOAE (right panels) measurements. Sensitivity is the difference in CR or DPOAE measurements between adjacent hearing-threshold categories (shown in Fig. 2) divided by 20 dB. The legend identifies pairs of hearing-threshold categories by their midpoints. CR indicates cochlear reflectance; DPOAE, distortion-product otoacoustic emission; SNR, signal to noise ratio.

midpoints of the categories. For example, the sensitivity of CR response level between the hearing-threshold categories 0 and 20 is computed as

$$S_{\mathrm{CR},0-20} = \frac{\mathrm{CR}_0 - \mathrm{CR}_{20}}{20}, \tag{4}$$

where CR_0 and CR_{20} represent the average CR response levels (shown in Fig. 2) for hearing-threshold categories 0 and 20, respectively. The largest DP response-level sensitivities are $S_{\mathrm{DP},0-20}$ and $S_{\mathrm{DP},20-40}$ and are about 0.5 dB/dB in the frequency range from 1 to 11 kHz. A sensitivity of 0.5 indicates that the response level decreases about 1 dB for every 2 dB increase in hearing threshold. CR response levels were (on average) less than half as sensitive to hearing threshold compared with DP response levels. Neither DP nor CR showed consistent hearing-threshold sensitivity for transitions between the other hearing-threshold categories, which is presumably because of poorer SNR (Fig. 2). The lower panels of Figure 3 show SNR sensitivities. Note that although SNR sensitivities are similar to their corresponding response-level sensitivities, they are not identical because the estimated noise levels were not independent of stimulus level. Comparison of the hearing-threshold sensitivities in Figure 3 suggests that DP measures might provide more accurate predictions of auditory status and audiometric thresholds compared with CR measures. However, the clinically relevant quantification of this expectation is the ROC-area statistic, which is described next.

CR and DP Predictions Based on Single-Frequency Measurements

Prediction of auditory status at a particular frequency as HI was defined by predicted threshold (THR) >20. Single-frequency

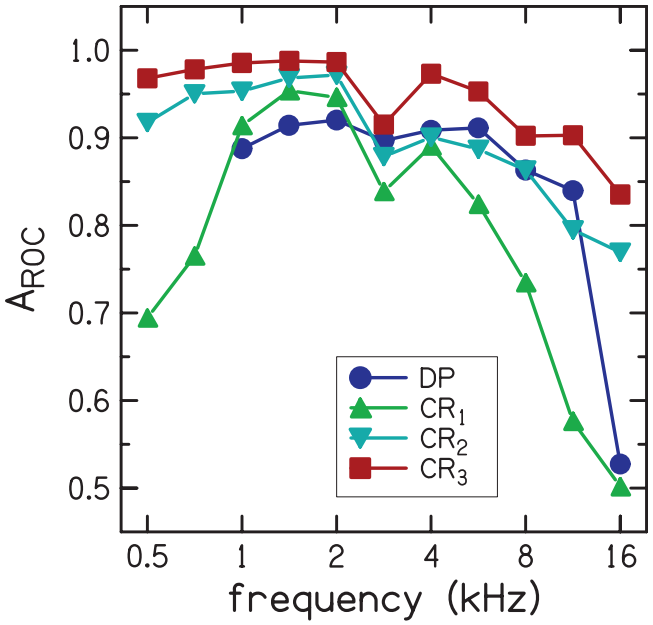

Figure 4. ROC areas (A_{ROC}) as functions of frequency for prediction of frequency-specific auditory status based on DPOAE and CR measurements. Higher is better. The DPOAE prediction (circles, DP) was based on linear regression of the DP variables (DP level and DP noise) at each frequency onto threshold at the matching frequency. For comparison, the first CR prediction (upward triangles, CR_1) was based on linear regression of the reflectance variables (ear-canal reflectance, CR magnitude, and CR noise) at each frequency onto threshold at the matching frequency. The second CR prediction (downward triangles, CR_2) was based on logistic regression of the reflectance variables at 10 frequencies onto threshold at each frequency. The third CR prediction (squares, CR_3) is the same as the second prediction except that the logistic regressions were performed separately for males and females. A_{ROC} indicates area under each receiver operating characteristic curve; CR, cochlear reflectance; DPOAE, distortion-product otoacoustic emission.

DP predictions were based on the linear regression model THR ~ DPL + DPN, where DPL is the DPOAE level and DPN is the DPOAE noise at a measured f_2 frequency that matches the audiometric frequency. Similarly, CR predictions were based on the linear regression model THR ~ ECR + CRM + CRN, where ECR is the ear-canal reflectance, CRM is the CR magnitude, and CRN is the CRN. CR predictions were most accurate at the highest of the three stimulus levels (i.e., 50 dB SPL), so only results at that level are presented.

Three types of CR predictions are considered. Predictions based on CR at a single frequency that approximated the audiometric frequency always combined males and females and are labeled as CR_1. Predictions based on CR at multiple frequencies used a principal component are labeled as CR_2 when males and females were combined and as CR_3 when independent predictions were made for males and females.

Figure 4 compares test performance of DP (circles) and CR (upward triangles) for single-frequency prediction of auditory status. CR predictions based on single-frequency measurements are labeled as CR_1 in the figure legend. The A_{ROC} average across frequency was 0.852 for DP prediction and 0.758 for CR_1 prediction. Neither type of measurement performed as well at 16 kHz as at 8 kHz. In the frequency range from 3 to 11 kHz, DP prediction was better than CR_1 prediction, which is consistent with DP measures having higher hearing-threshold sensitivity.

In the 1 to 2 kHz range, CR_1 prediction is better than DP prediction, which is unexpected because the DP measures had higher hearing-threshold sensitivity in this frequency range. Below 1 kHz, CR_1 test performance is not as good as in the 1 to 2 kHz range. However, comparison with DPOAE prediction cannot be made due to the lack of DPOAE measurements below 1 kHz. Figure 4 also contains two other types of CR predictions labeled CR_2 and CR_3 that will be discussed below.

Figure 5 compares the accuracy of DP (circles) and CR (upward triangles) single-frequency prediction of audiometric threshold. CR predictions based on single-frequency measurements are labeled as CR_1 in the figure legend. Prediction accuracy is quantified by its RMS error (RMSE) at each frequency in dB. The RMSE average across frequency was 18.7 for DP prediction and 20.2 for CR_1 prediction. The RMSE of CR predictions was similar to DP predictions in the frequency range from 1 to 6 kHz and at 16 kHz but was much larger at 8 and 11 kHz.

CR Predictions Based on Multiple-Frequency Measurements

The test performance advantage of using many frequency measurements in the prediction of each audiometric status has been described previously for DPOAE measurements (Dorn et al. 1999; Gorga et al. 2005). Multiple-frequency predictions are reported here only CR predictions to demonstrate their advantage over single-frequency CR predictions.

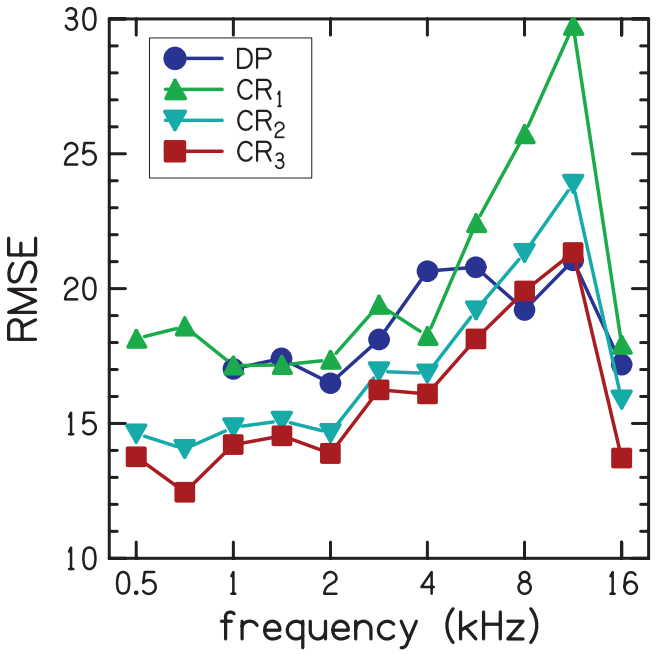

Figure 5. RMSE as functions of frequency for prediction of frequency-specific hearing threshold based on DPOAE and CR measurements. Lower is better. The DPOAE prediction (circles, DP) was based on linear regression of the DP variables (DP level and DP noise) at each frequency onto threshold at the matching frequency. For comparison, the first CR prediction (upward triangles, CR_1) was based on linear regression of the reflectance variables (ear-canal reflectance, CR magnitude, and CR noise) at each frequency onto threshold at the matching frequency. The second CR prediction (downward triangles, CR_2) was based on logistic regression of the reflectance variables at 10 frequencies onto threshold at each frequency. The third CR prediction (squares, CR_3) is the same as the second prediction except that the logistic regressions were performed separately for males and females. CR indicates cochlear reflectance; DPOAE, distortion-product otoacoustic emission; RMSE, root-mean-squared-error.

Test performance of CR for multiple-frequency prediction of auditory status is superimposed on the single-frequency predictions in Figure 4. CR predictions based on multiple-frequency measurements are labeled as CR_2 (downward triangles) for regressions that combined males and females and CR_3 (squares) for regressions that separated males and females. The A_{ROC} average across frequency was 0.896 for CR_2 prediction and 0.944 for CR_3 prediction. At every frequency, A_{ROC} for CR_2 prediction was higher than for CR_1 prediction, which is evidence of the expected advantage of multiple-frequency predictions over single-frequency predictions. At every frequency, A_{ROC} for CR_3 prediction was higher than for CR_2 prediction, which indicates the importance of considering males and females separately when making predictions about hearing status. A_{ROC} for CR_3 prediction was not only higher than for CR_2 prediction, it was also higher than for DP single-frequency prediction. So, the advantage of using multiple frequencies for prediction of auditory status was more than sufficient to compensate for the disadvantage of CR measurements having less hearing-threshold sensitivity compared with DP measurements.

Accuracy of CR for multiple-frequency prediction of audiometric threshold is superimposed on the single-frequency predictions in Figure 5. CR predictions based on multiple-frequency measurements are labeled as CR_2 (downward triangles) for regressions that combined males and females and CR_3 (squares) for regressions that separated males and females. The RMSE average across frequency was 17.0 for CR_2 prediction and 15.8 for CR_3 prediction. At every frequency, RMSE for CR_2 prediction was lower than for CR_1 prediction, which is evidence of the expected advantage of multiple-frequency predictions over single-frequency predictions. At every frequency, RMSE for CR_3 prediction was lower than for CR_2 prediction, which indicates the importance of considering males and females separately when making predictions about hearing status. RMSE for CR_3 prediction was nearly equal to RMSE for DP prediction at 8 and 11 kHz and was lower at other frequencies. So, the advantage of using multiple-frequencies for prediction of auditory status was sufficient to compensate for the disadvantage of CR measurements having less hearing-threshold sensitivity compared with DP single-frequency measurements.

Logistic regressions not only provided a prediction of audiometric threshold as the median of the hearing-threshold-category probability distribution but they also provided an estimate of the IQR with each threshold prediction. The IQR could be clinically useful as a means of quantifying the uncertainty associated with the predicted threshold. This idea is illustrated in Figure 6 by plotting the IQR of threshold predictions for 2 NH participants (upper panels) and 2 HI participants (lower panels). The IQR in each panel is shown as a shaded region that is bisected by the median of the distribution, which is the best single estimate of threshold. The actual audiometric thresholds for these 4 participants are superimposed over the IQR plots as circles. Although the accuracy of these threshold predictions may be less than required for selecting hearing aid gain, they could provide useful information beyond auditory status whenever a traditional audiogram is not possible or not reliable. Such cases could include infants, toddlers, adults with cognitive impairments, and suspected malingerers.

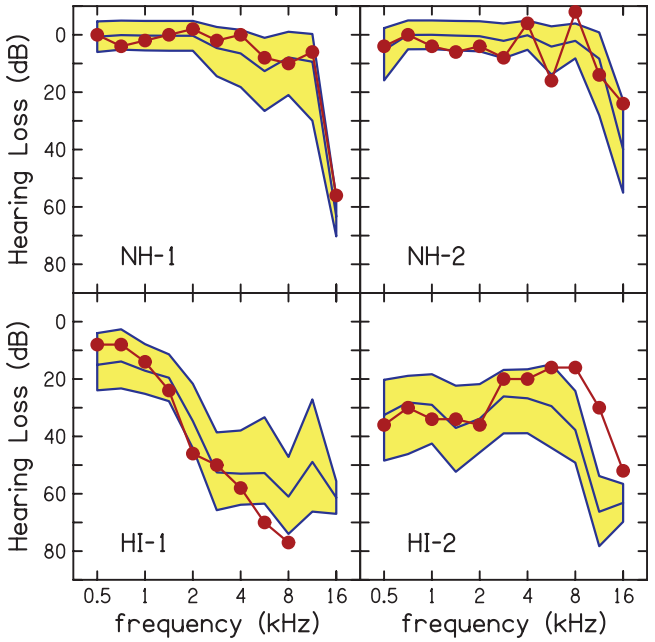

Figure 6. Examples of threshold predictions based on cochlear reflectance measurements for 2 NH (upper panels) and 2 HI (lower panels) participants. The shaded region represents the interquartile range (i.e., from 0.25 to 0.75 cumulative probability) of the threshold prediction at each frequency. The line that bisects the shaded region represents the median of the prediction. For comparison, the actual thresholds (circles) for these participants are superimposed. HI indicates hearing impaired; NH, normal hearing.

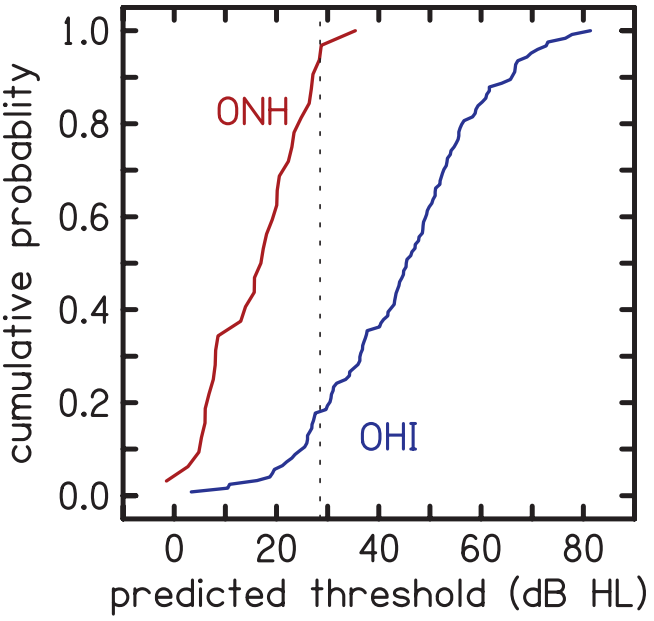

Figure 7. Cumulative distributions for ONH and OHI participants of predicted thresholds at 6 kHz. These predictions were based on distortion-product otoacoustic emission level and noise measurements at 6 kHz. The vertical dashed line at 28.8 dB HL intersects the ONH distribution at 0.95 and intersects the OHI distribution at 0.18. OHI indicates overall hearing impaired; ONH, overall normal hearing.

Prediction of Overall Auditory Status

The hearing-screening paradigm requires a single pass/refer decision, which needs a definition of overall auditory status. If we adopt the definition of OHI described previously (i.e., HL >20 at any frequency from 1 to 8 kHz), then the DPOAE prediction of OHI only requires measurement at a single frequency. By comparing the accuracy of DPOAE predictions of overall auditory status at every f_2 frequency, it was determined that the frequency with the best DPOAE prediction of overall auditory status was 6 kHz. The linear regression equation for the estimated threshold at 6 kHz is

$$\mathrm{THR} = 38.03 - 1.506 \cdot \mathrm{DPL} + 0.772 \cdot \mathrm{DPN}. \quad (5)$$

For example, if the measured DPOAE response level when f_2 = 6 kHz is DPL = 5 dB SPL and the noise level is DPN = −10 dB SPL, then the regression equation gives us THR = 22.8. This estimated threshold is predictive of OHI; however, the criterion value for the OHI prediction depends on the desired trade-off between specificity and sensitivity. The average sensitivity over all possible specificities is A_{ROC}, which for the OHI prediction (from a DPOAE measurement at 6 kHz) is 0.942. For comparison, A_{ROC} for multiple-frequency CR prediction of OHI is 0.908 when males and females are combined and 0.964 when males and females are separated. DPOAE prediction has the distinct advantage (over CR prediction) of shorter test time, which may be only 4 sec at 6 kHz compared with 4 or 5 min for the CR measurement.

The cumulative probabilities in Figure 7 show the distributions of single-frequency DPOAE predictions for OHI and ONH participants. These are the distributions for which the A_{ROC} is 0.942. These distributions illustrate the trade-off between sensitivity and specificity when the pass/refer criterion is set to any particular predicted threshold. For example, if the pass/refer criterion is set to 28.8 dB, then 95% of the ONH participants and 18% of the OHI participants would pass. So, a single DPOAE measurement at 6 kHz, which only requires a few seconds, could achieve a 95% specificity and 82% sensitivity in a hearing screening paradigm. The significance of this observation is that there is sufficient information at a single DPOAE frequency to allow predictions to be made regarding audiometric status defined over a range of frequencies.

DISCUSSION

Although the main goal of this study was to assess clinical utility, the results also reveal basic properties of cochlear mechanics by identifying the region in time-frequency space that is most sensitive to hearing loss. This region (Fig. 1) is delimited by (1) minimum and maximum time-delays (τ_{H} and τ_{L}) and (2) minimum and maximum cycle-delays (N_{L} and N_{H}). The limits that produced the greatest sensitivity to hearing loss were $\tau_{\mathrm{H}} = 1$ msec , $\tau_{\mathrm{L}} = 16$ msec , $N_{\mathrm{L}} = 10$ cycles, and $N_{\mathrm{H}} = 24$ cycles. A line that bisects this region would have a delay of $N_{\mathrm{CR}} = 17$ cycles at all frequencies. For comparison, the delay of SFOAEs is estimated to be $N_{\mathrm{SFOAE}} = 11 f^{-0.37}$ (Shera & Guinan 2003), which equals 11, 18, and 31 cycles at 1, 4, and 16 kHz, respectively. The agreement between CR and SFOAE latency is good at 4 kHz. The mismatch at higher and lower frequencies suggests that it might be possible to obtain better hearing-threshold sensitivity from CR measurements using a different region for time-frequency analysis.

The repeatability of CR measurements was established in a previous study (Rasetshwane et al. 2015). Initial assessments

of CR test performance, however, were disappointing in comparison to DPOAE test performance. The present study replicates the finding of poorer test performance for single-frequency predictions but emphasizes the potential of multiple-frequency CR prediction of audiometric threshold. IQR plots based on CR measurements (Fig. 6) would provide more information to the audiologist regarding auditory status than what is currently provided from DPOAE measurements. The test time required for the CR measurements from which the IQR plot is determined would require about 4 or 5 min, which would be a reasonable amount of time for a clinical measurement. However, for the prediction of overall auditory status in a hearing-screening paradigm, a DPOAE measurement at 6 kHz achieves better test performance in less time compared with CR measurements.

In current clinical practice, hearing thresholds are sometimes estimated from auditory brainstem response (ABR) measurements. For example, estimates of threshold at 1 kHz from a tone burst–evoked ABR at the same frequency have been reported to have an error with a standard deviation of about 9.5 dB (Gorga et al. 2006). Suppose that the error distribution for ABR prediction is approximately Gaussian, which means 2 SDs span 68% of the distribution. A rough estimate of the IQR for ABR prediction, which spans 50% of a distribution, would be a little less than 19 dB. For comparison, the average IQR across all frequencies and all four of the examples of CR prediction shown in Figure 6 is about 20 dB, which is not very different than the span of the ABR prediction error. Thus, with further refinement, it might be reasonable to expect the accuracy of CR threshold prediction to be similar to that of ABR threshold prediction.

Because the utility of DPOAE measurements in the prediction of auditory status has already been established (Johnson et al. 2010; Kirby et al. 2011), the role that DPOAE measurements served in the present study was mainly as a familiar reference for comparing the accuracy of predictions based on CR measurements. The fact that the DPOAE measurements were made in the same ears helped to ensure the validity of such comparisons. Single-frequency DPOAE prediction of auditory status outperformed single-frequency CR prediction. The average (over frequency) was $A_{ROC} = 0.852$ for DPOAE prediction and $A_{ROC} = 0.785$ for CR prediction (for the participants in this study). Although multifrequency, separate-sex CR prediction (average $A_{ROC} = 0.944$) exceeded single-frequency, combined-sex DPOAE prediction, it did not exceed multifrequency, separate-sex DPOAE prediction (average $A_{ROC} = 0.975$). Regression coefficients have been previously published for multifrequency DPOAE prediction (Dorn et al. 1999; Gorga et al. 2005), but this approach has not been adopted in standard clinical practice. With further development of signal processing and analysis methods, CR predictions may have clinical utility in prediction of audiometric thresholds. However, the longer test time required to make CR measurements makes them unlikely to ever replace DPOAE measurements for the prediction of auditory status.

Predictions based on CR measurements of both auditory status and audiometric thresholds improved when predictions were made separately for males and females. Similar improvements were observed for predictions of auditory status based on DPOAE measurements. These results suggest that sex should be considered whenever hearing assessments are based on physiological measurements. Sex differences have been previously reported for prevalence of SOAEs (Bilger et al. 1990), but have not been reported for OAE test performance. It has been suggested that the greater prevalence of SOAEs in females is due to reduced MOC efferent activity (McFadden 1993). In our measurements, CR SNR was larger for females, which is consistent with having larger cochlear-amplifier gain, which conceivably could be due to reduced MOC efferent activity.

The reason for the poorer performance of single-frequency predictions at 16 kHz for both CR and DP measurements is not well understood. DPOAE primary levels might not have been optimal for achieving the highest test performance. The temporal limits for the computation of CRM might not have been optimal. Measurement of audiometric threshold was more difficult at 16 kHz, so measurement error was likely to be higher. Finally, the equipment could not produce a 16 kHz tone at a level high enough to be heard by some subjects, which reduced the number of data points at this frequency.

Some aspects of the signal-processing methods used to extract information from CR measurements should be reassessed. Using a gammatone filterbank to perform time-frequency analysis has some desirable properties but is unlikely to be optimal for this purpose because it is impossible to exactly match filterbank properties to individual cochlear properties. Furthermore, instead of limiting the CRM calculation to a specific region in time-frequency space, a more efficient approach might be to perform a principal component analysis on the entire time-frequency space. Further research is needed to address these signal processing issues.

CONCLUSIONS

CR measurements were made at three WBN stimulus levels to determine which level was best for distinguishing ear with NH from ears with hearing loss. Test performance, quantified by A_{ROC}, was similar for the three levels. We recommend using the highest level, 50 dB SPL, based on a comparison of ANOVAs at each level. However, test performance at the lower levels (30 and 40 dB SPL) may be similar.

For audiometric status predictions based on single-frequency measurements, CR did not perform as well as DPOAE measurements. However, there is potential for improvement of CR sensitivity to hearing threshold by defining CRM differently. One possible alternative would be to compute principal components that represent a larger region of time-frequency space instead of the horn-shaped region used in the present study.

An unexpected outcome of this study was the additional improvement in test performance when the sex of the listener was considered. This observation serves as a reminder that sex should be included as a biological variable in future research studies. Furthermore, clinical assessments of hearing based on physiological measurements may benefit from having different criteria for each sex.

CR measurements offer greater potential when the entire frequency range of the response is included in predictions of auditory thresholds. Further improvements in CR sensitivity could increase the reliability of hearing-threshold predictions, which could facilitate the fitting of hearing aids in cases when a traditional audiogram is not possible. In addition, CR has the potential to be simulated by a linear cochlear model, which has the potential to provide new approaches to difference diagnosis of cochlear dysfunction.

ACKNOWLEDGMENTS

Assistance with the calibration software provided by Matthew Waid is gratefully acknowledged.

This research was supported by grants R01 DC008318 and P30 DC004662 from the National Institute of Health.

The authors have no conflicts of interest to disclose.

Address for correspondence: Stephen T. Neely, Center for Hearing Research, Boys Town National Research Hospital, 555 N 30th St. Omaha, NE 68131, USA. E-mail: Stephen.Neely@boystown.org

Received March 27, 2018; accepted September 18, 2018.

REFERENCES

Allen, J. B. (1986). Measurement of eardrum acoustic impedance. In J. B. Allen, J. L. Hall, A. Hubbard, S. T. Neely, A. Tubis (Eds.), *Peripheral Auditory Mechanisms*. New York, NY: Springer-Verlag.

Allen, J. B. (1997). Derecruitment by multiband compression in hearing aids. In W. Jesteadt (Ed.), *Modeling Sensorineural Hearing Loss* (pp. 99–112). Mahwah, NJ: Lawrence Erlbaum Associates.

Bilger, R. C., Matthies, M. L., Hammel, D. R., et al. (1990). Genetic implications of gender differences in the prevalence of spontaneous otoacoustic emissions. *J Speech Hear Res*, *33*, 418–432.

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

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

Burns, E. M., Arehart, K. H., Campbell, S. L. (1992). Prevalence of spontaneous otoacoustic emissions in neonates. *J Acoust Soc Am*, *91*, 1571–1575.

de Boer, E. (1997). Connecting frequency selectivity and nonlinearity for models of the cochlea. *Audit Neurosci*, *3*, 377–388.

Dorn, P. A., Piskorski, P., Gorga, M. P., et al. (1999). Predicting audiometric status from distortion product otoacoustic emissions using multivariate analyses. *Ear Hear*, *20*, 149–163.

Ellison, J. C., & Keefe, D. H. (2005). Audiometric predictions using stimulus-frequency otoacoustic emissions and middle ear measurements. *Ear Hear*, *26*, 487–503.

Fawcett, T. (2006). An introduction to ROC analysis. *Pattern Recognit Lett*, *27*, 861–874.

Goodman, S. S., Fitzpatrick, D. F., Ellison, J. C., et al. (2009). High-frequency click-evoked otoacoustic emissions and behavioral thresholds in humans. *J Acoust Soc Am*, *125*, 1014–1032.

Gorga, M.P., Neely, S.T., Bergman, B., et al. (1993). A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired participants. *J Acoust Soc Am*, *94*, 2639–2648.

Gorga, M. P., Dierking, D. M., Johnson, T. A., et al. (2005). A validation and potential clinical application of multivariate analyses of distortion-product otoacoustic emission data. *Ear Hear*, *26*, 593–607.

Gorga, M. P., Johnson, T. A., Kaminski, J. K., et al. (2006). Using a combination of click-and toneburst-evoked auditory brainstem response measurements to estimate pure-tone thresholds. *Ear Hear*, *27*, 60.

Groon, K. A., Rasetshwane, D. M., Kopun, J. G., et al. (2015). Air-leak effects on ear-canal acoustic absorbance. *Ear Hear*, *36*, 155–163.

Harte, J. M., & Elliott, S. J. (2005). Using the short-time correlation coefficient to compare transient- and derived, noise-evoked otoacoustic emission temporal waveforms. *J Acoust Soc Am*, *117*, 2989–2998.

Hohmann, V. (2002). Frequency analysis and synthesis using a gammatone filterbank. *Acustica*, *88*, 433–442.

Hussain, D. M., Gorga, M. P., Neely, S. T., et al. (1998). Transient evoked otoacoustic emissions in patients with normal hearing and in patients with hearing loss. *Ear Hear*, *19*, 434–449.

Jedrzejczak, W. W., Kochanek, K., Trzaskowski, B., et al. (2012). Tone-burst and click-evoked otoacoustic emissions in subjects with hearing loss above 0.25, 0.5, and 1 kHz. *Ear Hear*, *33*, 757–767.

Johnson, T. A., Neely, S. T., Kopun, J. G., et al. (2010). Clinical test performance of distortion-product otoacoustic emissions using new stimulus conditions. *Ear Hear*, *31*, 74–83.

Jolliffe, I. T. (2002). *Principal Component Analysis* (2nd ed.). New York, NY: Springer-Verlag.

Keefe, D. H., Ling, R., Bulen, J. C. (1992). Method to measure acoustic impedance and reflection coefficient. *J Acoust Soc Am*, *91*, 470–485.

Kemp, D. T. (2002). Otoacoustic emissions, their origin in cochlear function, and use. *Br Med Bull*, *63*, 223–241.

Kirby, B. J., Kopun, J. G., Tan, H., et al. (2011). Do "optimal" conditions improve distortion product otoacoustic emission test performance? *Ear Hear*, *32*, 230–237.

Maat, B., Wit, H. P., van Dijk, P. (2000). Noise-evoked otoacoustic emissions in humans. *J Acoust Soc Am*, *108*(5 Pt 1), 2272–2280.

McFadden, D. (1993). A speculation about the parallel ear asymmetries and sex differences in hearing sensitivity and otoacoustic emissions. *Hear Res*, *68*, 143–151.

McPherson, B., Li, S. F., Shi, B. X., et al. (2006). Neonatal hearing screening: Evaluation of tone-burst and click-evoked otoacoustic emission test criteria. *Ear Hear*, *27*, 256–262.

Mertes, I. B., & Goodman, S. S. (2013). Short-latency transient-evoked otoacoustic emissions as predictors of hearing status and thresholds. *J Acoust Soc Am*, *134*, 2127–2135.

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

Nørgaard, K. R., Fernandez-Grande, E., Laugesen, S. (2017). Compensating for evanescent modes and estimating characteristic impedance in waveguide acoustic impedance measurements. *J Acoust Soc Am*, *142*, 3497.

Patterson, R. D., & Holdsworth, J. (1996). A functional model of neural activity patterns and auditory images. In A.W. Ainsworth (Ed.), *Advances in Speech, Hearing and Language Processing* (pp. 547–563). London: JAI Press.

Prieve, B. A., Gorga, M. P., Schmidt, A., et al. (1993). Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears. *J Acoust Soc Am*, *93*, 3308–3319.

Rasetshwane, D. M., & Neely, S. T. (2011). Inverse solution of ear-canal area function from reflectance. *J Acoust Soc Am*, *130*, 3873–3881.

Rasetshwane, D. M., & Neely, S. T. (2012). Measurements of wide-band cochlear reflectance in humans. *J Assoc Res Otolaryngol*, *13*, 591–607.

Rasetshwane, D. M., Fultz, S. E., Kopun, J. G., et al. (2015). Reliability and clinical test performance of cochlear reflectance. *Ear Hear*, *36*, 111–124.

Scheperle, R. A., Goodman, S. S., Neely, S. T. (2011). Further assessment of forward pressure level for in situ calibration. *J Acoust Soc Am*, *130*, 3882–3892.

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

Shera, C. A., & Guinan, J. J. Jr. (1999). Evoked otoacoustic emissions arise by two fundamentally different mechanisms: A taxonomy for mammalian OAEs. *J Acoust Soc Am*, *105*(2 Pt 1), 782–798.

Shera, C. A., & Guinan, J. J. Jr. (2003). Stimulus-frequency-emission group delay: A test of coherent reflection filtering and a window on cochlear tuning. *J Acoust Soc Am*, *113*, 2762–2772.

Shera, C. A., Guinan, J. J. Jr, Oxenham, A. J. (2010). Otoacoustic estimation of cochlear tuning: Validation in the chinchilla. *J Assoc Res Otolaryngol*, *11*, 343–365.

Sieck, N. E., Rasetshwane, D. M., Kopun, J. G., et al. (2016). Multi-tone suppression of distortion-product otoacoustic emissions in humans. *J Acoust Soc Am*, *139*, 2299–2309.

Siegel, J. H. & Neely, S. T. (2017). Eartip modification greatly reduces evanescent waves. In Proceedings of the 40th Midwinter Meeting (Association for Research in Otolaryngology).

Stover, L., Gorga, M. P., Neely, S. T., et al. (1996). Toward optimizing the clinical utility of distortion product otoacoustic emission measurements. *J Acoust Soc Am*, *100*(2 Pt 1), 956–967.

Swets, J. A. (1988). Measuring the accuracy of diagnostic systems. *Science*, *240*, 1285–1293.

Voss, S. E., & Allen, J. B. (1994). Measurement of acoustic impedance and reflectance in the human ear canal. *J Acoust Soc Am*, *95*, 372–384.

Voss, S. E., Merchant, G. R., Horton, N. J. (2012). Effects of middle-ear disorders on power reflectance measured in cadaveric ear canals. *Ear Hear*, *33*, 195–208.