# Detecting high-frequency hearing loss with click-evoked otoacoustic emissions

Douglas H. Keefe[a)]
*Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131*

Shawn S. Goodman
*Department of Communication Sciences and Disorders, University of Iowa, Iowa City, Iowa 52242*

John C. Ellison, Denis F. Fitzpatrick, and Michael P. Gorga
*Boys Town National Research Hospital, 555 North 30th Street, Omaha, Nebraska 68131*

In contrast to clinical click-evoked otoacoustic emission (CEOAE) tests that are inaccurate above 4–5 kHz, a research procedure measured CEOAEs up to 16 kHz in 446 ears and predicted the presence/absence of a sensorineural hearing loss. The behavioral threshold test that served as a reference to evaluate CEOAE test accuracy used a yes–no task in a maximum-likelihood adaptive procedure. This test was highly efficient between 0.5 and 12.7 kHz: Thresholds measured in 2 min per frequency had a median standard deviation (SD) of 1.2–1.5 dB across subjects. CEOAE test performance was assessed by the area under the receiver operating characteristic curve (AUC). The mean AUC from 1 to 10 kHz was 0.90 (SD = 0.016). AUC decreased to 0.86 at 12.7 kHz and to 0.7 at 0.5 and 16 kHz, possibly due in part to insufficient stimulus levels. Between 1 and 12.7 kHz, the medians of the magnitude difference in CEOAEs and in behavioral thresholds were <4 dB. The improved CEOAE test performance above 4–5 kHz was due to retaining the portion of the CEOAE response with latencies as short as 0.3 ms. Results have potential clinical significance in predicting hearing status from at least 1 to 10 kHz using a single CEOAE response.
© *2011 Acoustical Society of America.* [DOI: 10.1121/1.3514527]

## I. INTRODUCTION

Translating from the discovery of click-evoked otoacoustic emissions (CEOAEs) by Kemp (1978), CEOAEs are used to assess the integrity of cochlear function across frequency in a single wideband measurement. In an ear with normal middle-ear function, a relatively low-level or absent CEOAE response predicts an increased likelihood of a sensorineural hearing loss (SNHL). Avan *et al.* (1997) reported that the maximum frequency at which any CEOAEs were detected using a clinical system varied from 3.2 to 5.5 kHz across 43 normal-hearing subjects. Because of this limitation on the maximum CEOAE test frequency, clinical studies to assess the accuracy of CEOAE predictions of SNHL using clinical decision theory have not evaluated test performance above 4 kHz. Goodman *et al.* (2009) described a double-evoked method to measure high-frequency CEOAEs, which was able to detect CEOAEs in normal-hearing adult ears from 1 to 16 kHz. The measured CEOAE latencies had values in the expected physiological range, decreasing to about 1 ms at 16 kHz. Goodman *et al.* (2009) explained that the ability of the double-evoked procedure to measure CEOAEs above 4–5 kHz was because this procedure retained the portion of the CEOAE response with latencies shorter than 2.5 ms, whereas clinical procedures discard this short-latency portion of the response. The current study was performed to measure

the accuracy of this high-frequency CEOAE test in detecting SNHL in adults.

A non-invasive acoustic test that accurately and rapidly screens for SNHL across the normal range of human hearing would have clinical significance in testing both infants and adults. Otoacoustic emission (OAE) assessment at higher frequencies may be relevant to newborn hearing screening programs because of the possible importance of high-frequency cues in speech perception. Otoacoustic emission responses are noisier in infants than in adults below approximately 1.5 kHz, so that it may be easier to detect CEOAE responses that include higher frequencies. CEOAE testing also has applications in the assessment of adult hearing, e.g., the onset of age-related hearing loss (HL) often occurs at high frequencies so that an improved objective test of cochlear function above 4 kHz might be helpful. Goodman *et al.* (2009) summarized findings from various studies demonstrating that OAEs have potential applications to screen for ototoxic effects, which typically damage the high frequencies first. An improved high-frequency CEOAE test might be useful in such an application.

While the results in Goodman *et al.* (2009) showed the ability to measure CEOAEs in some ears up to 16 kHz, subjects were included only if their hearing was normal at all frequencies from 0.5 to 8 kHz. Goodman *et al.* also measured high-frequency audiograms for each subject at third-octave frequencies of 10.1, 12.7, and 16 kHz and found that some of the test ears had elevated thresholds at frequencies above 8 kHz consistent with SNHL. A gap in knowledge is the

---

[a)]Author to whom correspondence should be addressed. Electronic mail: Douglas.Keefe@boystown.org

accuracy with which this new wideband CEOAE test detects SNHL across its range of frequencies. The ability to identify differences between CEOAE responses recorded in a normal-hearing group compared to a group of ears with SNHL is a more stringent test of the double-evoked procedure than its known ability to measure CEOAEs in normal-hearing ears.

The present study addressed this issue by acquiring wideband CEOAE responses in adult subjects with SNHL and those with normal hearing. The main result described below is that the wideband CEOAE test was accurate at detecting SNHL from 1 to 10.1 kHz. The repeatability of CEOAE and audiometric testing was assessed in a sub-group of normal-hearing subjects and found to be adequate for clinical use. A small group of subjects with profound HL, who also use cochlear implants, were also tested using CEOAEs (with the implant powered off) to evaluate whether the acoustic response was of an expected small magnitude compared to the noise level. The findings from this study set the stage for future studies using wideband CEOAE testing in infant ears and in adults with more specific risk factors for SNHL.

An additional sidearm analysis assessed the efficiency of the behavioral threshold technique, which used a yes–no task and an adaptive procedure. This continues the goals of the Goodman et al. (2009) study to investigate relationships between CEOAE measurements and behavioral thresholds at frequencies above 5 kHz.

## II. METHODS

### A. Subjects

Subjects were included only if they had normal middle-ear function as assessed by tympanometry, which was performed based on a 226-Hz probe tone using a GSI Tympstar Middle-Ear Analyzer (Grason-Stadler, Eden Prairie, MN). Tympanometry was normal if responses were in the following ranges: ear-canal volume between 0.6 and 2.5 ml, compensated admittance magnitude between 0.3 and 1.4 mmho, and tympanometric peak pressure between $-83$ and 50 daPa. These screening criteria were based on findings from previous studies summarized in Ellison and Keefe (2005).

Audiograms were measured with a GSI 61 clinical audiometer (Grason-Stadler, Eden Prairie, MN) using insert earphones. Subjects were included in the normal-hearing group only if their pure-tone air-conduction thresholds were $\leq 15$ dB HL at octave frequencies from 0.5 to 8 kHz. Thus, the term "normal-hearing group" is restricted to ears with audiograms within normal limits up to 8 kHz and unspecified at higher frequencies. The nominal "SNHL group" included any test ear that did not qualify for the normal-hearing group, i.e., their air-conduction thresholds were $>15$ dB HL at one or more octave frequencies from 0.5 to 8 kHz. Some subjects in this nominal SNHL group had normal hearing at other frequencies, in which case their responses were included as part of the normal population at such frequencies. Therefore, the numbers of ears with normal hearing and with SNHL differed at each test frequency.

Subject data in 39 ears of 23 normal-hearing adults that were obtained in the Goodman et al. (2009) study were also included in the normal-hearing group in the present study.

The normal-hearing group in Goodman et al. (2009) was composed of young adult subjects and enabled the construction of a HL scale at frequencies above 8 kHz at which the clinical audiometer was uncalibrated. Additional normal-hearing subjects were included in the normal-hearing group, ranging in age from 14 to 38 years. The SNHL group included older subjects as a result of their higher prevalence of SNHL, and some of these older subjects had normal hearing at some test frequencies.

Because both ears of a subject were tested whenever possible, the counts for the overall groups were based on numbers of ears rather than subjects. Despite an expected correlation between ears within the same individual, it is clinically relevant to study ears rather than subjects because any remediation of a HL would be performed separately for each ear. Nevertheless, across-ear factors do exist when treating a patient, e.g., a patient with a binaural HL would typically receive the same type of hearing aid for both ears, and comfort levels would be adjusted using binaural listening tasks. Overall, the normal-hearing group included 113 ears (57 left and 56 right ears). These included 26 male and 87 female test ears with a mean age of 21 years [standard deviation (SD) of 6 years]. The SNHL group included 333 ears (167 left and 166 right ears). These included 174 male and 159 female test ears with a mean age of 58 years (SD = 17 years). Thus, the total number of test ears at each frequency was 446.

Data were acquired in a subset of 41 ears selected from the normal-hearing group on a different test date to evaluate test repeatability. These included 12 male and 29 female ears, and 21 left and 20 right ears. The subject ages ranged from 14 to 38 years with a mean age of 21 years and a SD of 6 years. One test ear per subject was used in the repeatability study.

An additional five subjects (aged 18, 19, 20, 56, and 59 years) with cochlear implants were tested with the implant powered off. These tests provided a biological model to test the expectation that CEOAEs would be absent in these subjects with profound HL. These five test ears were not included with the SNHL group in the main analyses. This group included two females and three males, and three left and two right ears.

During testing, each subject was seated comfortably inside a sound-attenuated booth. The experimental protocol was approved by the Institutional Review Board at Boys Town National Research Hospital, and written informed consent was obtained from all participants.

### B. Behavioral threshold measurement procedures

Because the clinical audiometer used in the study was limited to a maximum test frequency of 8 kHz and because one goal was to assess the ability of CEOAEs to predict SNHL up to 16 kHz, behavioral thresholds were measured in the research protocol at four octave frequencies from 0.5 to 4 kHz, and at six third-octave frequencies above 4 kHz up to 16 kHz. The research protocol was designed so that behavioral thresholds and CEOAEs were measured using the same insert probe. This insert probe assembly used an ER-10B+ microphone and a pair of ER-2 sound sources (Etymotic, Elk

Keefe et al.: Detecting hearing loss click otoacoustic emissions

Grove Village, IL), which were the same probe transducers used by Goodman *et al.* (2009). For the audiometric task, the stimulus was presented through one of the sound sources. Based on the design of these sound sources that limited their maximum output level, the maximum possible HL was approximately 90 dB in the research protocol, which was lower than the maximum possible HL of the clinical system. The maximum HL was further reduced at the highest test frequencies because the calibrated sound pressure level (SPL) corresponding to 0 dB HL was larger at higher frequencies, as discussed in Goodman *et al.* (2009).

Whatever might be the high-frequency characteristics of the sound field generated by the probe within the ear canal, the characteristics would be the same in the behavioral and CEOAE measurements in each particular test ear. These issues were discussed by Goodman *et al.* (2009), who compared audiograms measured using clinical and research protocols in subjects with normal hearing up to 8 kHz. The thresholds using the two protocols were about the same at 0.5, 1, and 2 kHz, but the thresholds measured using the research protocol were slightly lower than clinical thresholds at 4 and 8 kHz. This difference was thought to be due to differences in calibrating the audiometers at higher frequencies. Overall, Goodman *et al.* (2009) concluded that the research protocol was adequate for audiometric measurements up to 16 kHz. This was an important observation because it provided behavioral threshold data against which the performance of high-frequency CEOAEs could be tested.

Efficient procedures exist to measure behavioral thresholds using a yes–no task and maximum likelihood (ML) to sequentially adjust the stimulus level depending on the preceding responses (Green, 1993; Leek *et al.*, 2000). The protocol used in the present study adopted the same underlying psychometric model as in Green (1993). ML protocols using the yes–no task and a relatively limited number of trials can produce errors in the measured threshold, and two procedural refinements have been suggested to obtain more accurate results. One refinement is to present the stimulus at a number of fixed levels across a broad dynamic range for the initial four trials and use ML to adaptively set the stimulus level in later trials (Keefe *et al.*, 2009; Goodman *et al.*, 2009). A second refinement is to adjust the slope of the psychometric function toward steeper values in impaired listeners, and to possibly use fewer trials in subjects with steeper slopes (Lecluyse and Meddis, 2009). Both protocols use catch trials (Gu and Green, 1994), i.e., trials on which the stimulus is absent, to capture information on subject inattentiveness. Results obtained using these refined procedures were improved relative to standard ML yes–no procedures.

The ML adaptive procedure used in the present study was built on procedures used by Keefe *et al.* (2009) and Goodman *et al.* (2009). Each tonal stimulus had an overall duration of 250 ms, which included 25-ms onset and offset ramps using cosine-squared envelopes. The properties of the custom-built response box used by the subject were described in Keefe *et al.* (2009). The response box provided visual feedback at the beginning of each run, the yes or no response that the subject entered on each trial, a priming alert for each upcoming trial, and a message that a run had ended. Three runs of the

adaptive yes–no threshold procedure were performed with each run composed of 15 trials. The stimulus levels in the first four trials that were not assigned as catch trials were set independently of the subject's response. The dynamic range of the stimulus was split into quadrants and each of the four trials used a stimulus level randomly selected within each quadrant, with each quadrant selected once in random order. Unless a trial was a catch trial, the stimulus levels in the remaining 11 trials of a run were adaptively set at the ML estimate of threshold based on all preceding trials. Three catch trials with no stimulus presentation were included, at random, among the 14 trials after the first trial. In runs in which a subject responded *No* in response to the first seven trials that were not catch trials, the stimulus would have been presented at its maximum level on the last three of these trials. When this condition occurred, the run was halted before completing the remaining trials; the threshold was classified as higher than the maximum presentation level, i.e., as no response (NR).

The threshold was calculated at each frequency using the medians of the thresholds from three successive runs, which is termed a block of runs, if the SD of these thresholds was within 5 dB for the block. In contrast to the mean threshold statistic, the median threshold statistic remained well-defined when NR classifications were present. The SD was defined only for calculated thresholds (i.e., those thresholds other than NR). For tests with fewer than two calculated thresholds in the three runs, the SD was not calculated and the threshold result for the block was encoded as NR. A 5-dB criterion was considered adequate because the categories of HL varied in 5-dB steps (the actual SDs were typically much smaller than 5 dB). If the SD was >5 dB on the first block of runs, the tester reinstructed the subject and another block of three runs were performed for a maximum of two more times or until a SD $\leq$ 5 dB was attained (it was attained within three blocks for all subjects). The threshold and its SD were reported from the final block of runs.

The threshold was first measured at 1 kHz, which also served to train each subject in the research protocol. If the subject's responses were inconsistent at 1 kHz, the tester paused the computer-controlled test, provided feedback to the subject and answered any questions about the procedure, and then continued the automated test at other frequencies. The presentation order for all other test frequencies was randomized across subjects to control for any frequency-dependent bias in subject response. One underlying issue was that because high-frequency HL has the highest prevalence, many of the high-frequency stimuli might be inaudible for some subjects. The presentation order of frequencies was mixed so as to limit the durations of the time intervals for many subjects over which the stimulus might be inaudible. This procedure helped maintain subject attentiveness. Nevertheless, the experimenter noticed that subjects sometimes were occasionally sleepy during the test, resulting in longer measurement durations.

## C. CEOAE measurement procedures

CEOAE data were acquired using methods summarized below and more fully described in Goodman *et al.* (2009).

J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions    247

The computer-based measurement system was implemented using custom software in which sounds were generated and recorded with a sound card (CARDDELUXE) at a sample rate of 44.1 kHz per channel. The electrical stimulus was designed so that each ER-2 receiver generated an acoustic click with the following property: The voltage output waveform from the ER10B+ microphone approximated the impulse response of a finite impulse response filter with a pass band between 0.5 and 16 kHz and stop bands below 0.043 kHz and above 17 kHz. The microphone magnitude and phase calibration were subsequently applied to calculate the pressure waveform. Because the microphone sensitivity was reduced at high frequencies, the acoustic click had a 5 dB lower pressure level in the third octave at 16 kHz compared to a relatively flat sensitivity at lower frequencies from 0.5 kHz up to the third-octave frequency of 12.7 kHz. A "relatively flat sensitivity" refers to variability within $\pm 2$ dB except for a boost of 6 dB in a narrow bandwidth centered close to 11.7 kHz.

While the peak-to-peak equivalent sound pressure level (peSPL) of the incident "probe" click was varied in Goodman et al. (2009) in 6-dB steps from 43 to 73 dB peSPL, the stimulus levels in the present study were the three highest levels used in that study, 61, 67, and 73 dB peSPL, as well as an additional highest level of 76 dB peSPL. The fact that the 446 ears included 39 ears from the Goodman et al. study means that these 39 ears did not receive a CEOAE test at the 76 dB peSPL. Thus, the total number of ears at this highest CEOAE test level was 407 ears.

CEOAEs were measured using a double-evoked method (Keefe, 1998) that extracts a nonlinear acoustic residual from a set of three 25.5-ms pressure waveforms, each elicited using a different acoustic click. Measured CEOAEs can be analyzed at times as short as the onset time of the click stimulus, which results in large amounts of high-frequency signal at short times in normal-hearing ears (Keefe and Ling, 1998; Goodman et al., 2009). The first stimulus, which generated the click at the peSPLs specified above, was presented through a first receiver, and a first pressure response waveform $p_1$ was measured. The second stimulus, which was an identical waveform to the first stimulus except 15 dB higher in level, was presented through a second receiver, and a second pressure waveform $p_2$ was measured. The third stimulus used the simultaneous presentation of the first stimulus by the first receiver and the second stimulus by the second receiver, and a third pressure waveform $p_{12}$ was measured. For example, if the probe click level of the first stimulus was 76 dB peSPL, the click level of the second stimulus was 91 dB peSPL and of the third stimulus was approximately 92.4 dB peSPL . The nonlinear residual pressure waveform $p_d$ was defined as $p_d = p_1 + p_2 - p_{12}$. This nonlinear residual was interpreted as a biological response and termed the CEOAE pressure waveform. This interpretation was based on the absence of intermodulation distortion relative to the noise floor in the analyses of coupler recordings (Goodman et al., 2009).

The ability to measure CEOAEs at frequencies as high as 16 kHz resulted because the double-evoked measurement procedure retained the short-latency part of the CEOAE response ($<3$ ms) that is discarded in other procedures. A common clinical procedure to measure CEOAEs is the nonlinear derived procedure of Kemp et al. (1986). The procedure measures a nonlinear residual of the CEOAE waveform through calculating a linear difference of several click-evoked pressure waveforms. However, the nonlinear derived procedure discards the initial part of the CEOAE waveform because of measurement-system artifact. Kemp et al. (1986) windowed the CEOAE waveform to 3–20 ms using on and off ramps of 1.2 ms duration. The resulting nonlinear-residual CEOAE waveform would have full-scale amplitude at times from 4.2 to 18.8 ms. Kemp et al. (1990) windowed the CEOAE waveform to 2.5–20 ms using on and off ramps of 2.5 ms, so that the resulting nonlinear-residual CEOAE waveform would have full-scale amplitude at times from 5.0 to 17.5 ms.

Comparisons of CEOAE waveforms measured using these procedures showed lower system distortion for the double-evoked procedure compared to one version of a nonlinear-derived procedure (Keefe and Ling, 1998). As described in more detail in Sec. II D, the double-evoked technique used in the present study windowed the CEOAE waveform from 0.25 to 21.3 ms using an on ramp of 0.16 ms duration and an off ramp of 5 ms duration. The resulting CEOAE waveform had full-scale amplitude at times from 0.41 to 16.3 ms. This allowed the analysis of the early latency ($<3$ ms) part of the COEAE containing the high-frequency response. Its system distortion was much lower mainly because two receivers within the probe were used to generate the OAE stimuli, such that each receiver always presented its signal at a fixed amplitude. Any harmonic distortion associated with either signal was thereby cancelled, leaving only intermodulation distortion. The procedures of Kemp et al. (1986, 1990) use a single receiver that generates a signal at different amplitudes and polarities, and their harmonic distortion components do not fully cancel in the linear-difference calculation.

The ability to identify ears with sensorineural loss depends on an appropriately low level of system distortion. If the distortion were too high or if stimulus artifact remained, then the test would be unable to distinguish between normal and impaired ears. Thus, an evaluation of test performance in normal and (more importantly) impaired ears directly tests the linearity of the recording system and, by extension, the efficacy of the test paradigm in identifying ears with HL.

For each of four levels, $N = 4050$ independent buffers were collected (duration 5.16 minutes). Each buffer consisted of a set of three elementary intervals, each having a duration of 25.5 ms. During CEOAE testing within the booth, subjects were given the opportunity to view a closed-caption DVD (i.e., without audible sound), which helped them remain awake and alert. The probe was not adjusted in the ear during the test session unless the experimenter determined that it needed refitting.

After all data were acquired, intermittently noisy buffers were discarded using a slight improvement of the procedure described by Goodman et al. (2009), which excluded outliers based on the mean-squared CEOAE pressure exceeding 2.25 times the interquartile range (IQR). The alternative procedure used in the present work was a combination of median

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions

absolute deviation (MAD) tests applied to peak-to-peak amplitude, crest value and noise energy. The procedure and its implementation in a real-time data acquisition system are described in Liu *et al.* (2008), although the MAD test was applied in the present study only after all data had been acquired. Extensive numerical simulations were performed using both artifact-rejection procedures, in which a click-like signal was embedded in various combinations and levels of simulated intermittent noise. The probability of artifact in the form of intermittent noise was systematically varied. The tests were compared using standard definitions of near and far outliers, and the MAD test was also evaluated in its original form (Hoaglin *et al.*, 1983). As noted by Reimann *et al.* (2005) and confirmed in the present numerical simulations, the original MAD test tends to discard too many buffers. The modified MAD test described in Liu *et al.* has identical performance to the far-outlier form of the IQR procedure used by Goodman *et al.* when the noise is normally distributed. For example, the cumulative probability of a far outlier in a normally distributed set of artifacts is approximately 1e-6. Real measured data are much more likely to contain artifacts than would be predicted by the normal distribution; the modified MAD procedures appear well suited to practical use in aural acoustic measurements.

The MAD and IQR procedures (when calibrated in the same manner based on either near or far outliers) performed similarly in the simulations when the probability of artifact was low, but the MAD test outperformed the IQR test when the probability of artifact was high. Both procedures shared the desirable property that the fraction of excluded buffers cannot become unduly high when noise levels are high. The IQR test relies on the construction of the IQR, which inherently prevents the middle 50% of the responses from being excluded, and no buffer is rejected unless it is distant from the IQR. The MAD test relies on an absolute deviation from the median, which thus also prevents the responses sufficiently close to the median from being excluded. In practice, a rejection of more than about 25% of buffers would correspond to a contaminated acoustic recording and was not observed in the analyses of CEOAE recordings. The addition of a test to detect high levels of stationary noise alongside the MAD test would be helpful in detecting excessively noisy recordings that might completely mask the detection of the signal of interest.

Across all subjects, this procedure resulted in rejecting approximately 4–18% of the buffers as outliers, with the exact number of rejected buffers independently determined for each subject, ear, and stimulus level. The remaining buffers in each condition were averaged to provide a single CEOAE pressure waveform.

## D. Multi-window spectral analyses of CEOAEs

Each acoustic click presented in the ear canal propagates in the forward direction through the middle ear to the cochlea. The resulting CEOAE is generated on the basilar membrane and propagates in the reverse direction through the middle ear and back to the probe microphone in the ear canal. Because ear-canal and middle-ear transmission are

linear processes (except for effects related to the stapedius muscle reflex that were assumed to be negligible in the present study) and cochlear generation and transmission are nonlinear, the subtraction of click responses leading to the CEOAE pressure waveform extracts a signal that encodes the cochlear nonlinearity. The compressive nonlinearity associated with outer-hair cell functioning in cochlear mechanics results in a dominance of compressive nonlinearity in the measured CEOAE.

Cochlear function is tonotopically organized, i.e., regions closer to the cochlear base are tuned to higher frequencies and regions closer to the cochlear apex are tuned to lower frequencies. Latencies of higher-frequency CEOAE components are shorter than latencies of lower-frequency components. The CEOAE latency-frequency map in human ears is consistent with a place-fixed source of CEOAE generation, i.e., in the region of the peak amplitude response of the cochlear traveling wave at each frequency. Because CEOAE latency in young adult ears with normal hearing is less than 2.5 ms for frequencies at and above 5 kHz at a stimulus level of 73 dB peSPL (Goodman *et al.*, 2009), it is important to retain the short-latency CEOAE signal to accurately assess cochlear function above 4 kHz (Goodman *et al.* also described latencies recorded at stimulus levels as low as 55 dB peSPL). The fact that the CEOAE waveform typically has high-frequency energy at shorter latencies was used in the present study to improve the ability to detect a CEOAE. This general property is found across a range of stimulus levels. Whereas Goodman *et al.* calculated CEOAE spectra using a single temporal window of $p_d$, the present study used a set of three temporal windows that spanned the range of latencies out to 21.3 ms to calculate the CEOAE spectra. Using such multiple-window procedures increases the signal-to-noise ratio (SNR) in measurements of CEOAE spectra (Whitehead *et al.*, 1995). The basic idea is that the CEOAE signal in any particular frequency band is temporally concentrated within a relatively narrow range of latencies, while the noise in the same frequency band is distributed broadly over time. By appropriately limiting the duration of the temporal window, the SNR can be improved by retaining most of the CEOAE energy while rejecting much of the noise energy. An individual CEOAE spectrum was calculated based on discrete Fourier transforms (DFTs) of the product of the CEOAE pressure waveform and each of three windows that selected out the early, middle, and late portions of this waveform. A composite CEOAE spectrum was then constructed by combining the three individual spectra using a SNR weighting procedure.

An example of the windowing is shown for an individual ear (subject 1) in Fig. 1, which plots the entire CEOAE pressure waveform on a logarithmic-time axis in the top panel, and the three windowed waveforms in the bottom panel. Plotting the waveform as a function of the logarithm of time, rather than as a function of time, helps in observing waveform features within the initial 2.5 ms following the onset time of the click stimulus ($t = 0$ ms). Except for raised cosine ramps at the ends of each window, the unit amplitude extended from 0.25 to 2 ms, 2 to 8 ms, and 8 to 21.3 ms in the early, middle, and late windows, respectively. Preliminary analyses of
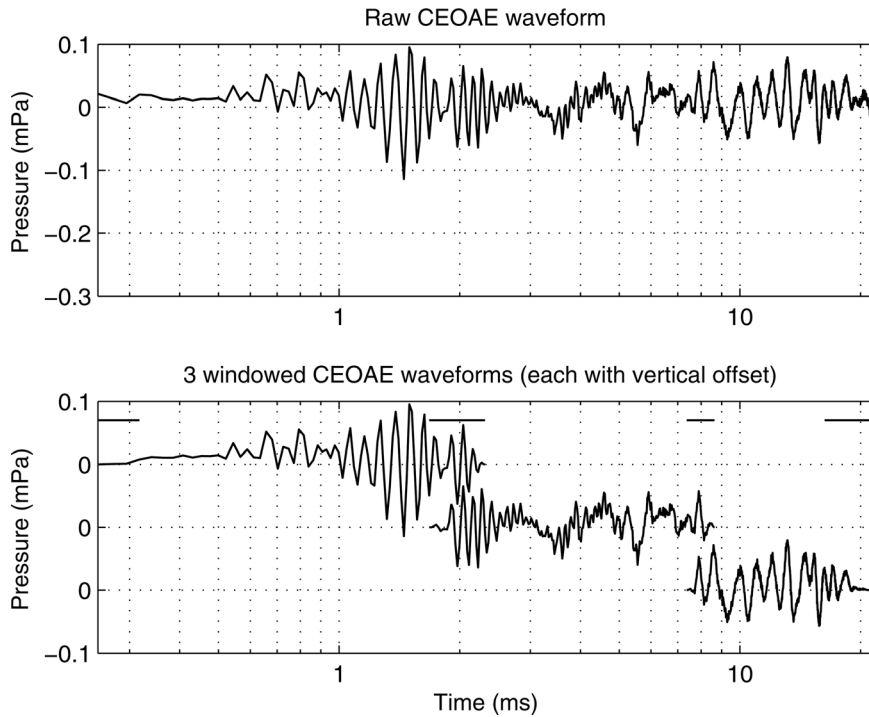
J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions     249

FIG. 1. Top panel: Mean CEOAE pressure waveform recorded in the test ear of subject 1. Bottom panel: The top, middle, and bottoms curves show the product of the early, middle, and late window, respectively, with the mean CEOAE pressure waveform. The horizontal lines at the top of the bottom panel indicate the times over which each pair of adjacent windows was ramped on or off. These three curves are separated vertically so that each is centered at 0 mPa, but each vertical tick mark corresponds to a change of 0.1 mPa.

CEOAEs recorded in young adults evaluated the effect of varying the onset of the early window between 0.02, 0.125, 0.25, and 0.5 ms. The delays at 0.02 and 0.125 ms were sufficiently short that some system intermodulation distortion, which was generated within the temporal duration of the acoustic click stimulus, was present in the $p_d$ waveform. Preliminary analyses confirmed that this distortion was below noise levels once the samples in $p_d$ earlier than 0.25 ms were zeroed, and this operation did not interfere with measuring CEOAEs up to 16 kHz. Zeroing the $p_d$ out to 0.5 ms attenuated the highest-frequency CEOAE components, so this was rejected because it excluded too much of the data of interest. An example of this system distortion is shown in the band-pass-filtered responses above 8 kHz in Fig. 6 of Goodman et al. (2009), which is consistent with the choice of the onset at 0.25 ms. It is important to note that the present temporal window represents an improvement over systems that remove the first 2.5–3 ms from the response waveform, thus discarding information at high frequencies. The value of including short-latency energy will be demonstrated in the main results of the present study.

Each window was of unit amplitude in its central portion and had cosine-squared onset and offset ramps. The pairs of adjacent windows at 2 and 8 ms overlapped in time such that the sum of their window amplitudes was one. The duration of the onset ramp of the early window was seven samples (0.16 ms). The duration of the offset ramp of the early window and the onset ramp of the middle window was 14 samples (0.32 ms). The duration of the offset ramp of the middle window and the onset ramp of the late window was 28 samples (0.63 ms). The duration of the offset ramp of the late window was 221 samples (5.01 ms). These window durations were selected to be approximately 2.5 periods of the frequency whose stimulus-frequency otoacoustic emission (SFOAE) group delay [based on the empirical fits of Shera et al. (2002)]

was equal to the temporal boundaries between the windows (i.e., at 0.25, 2, 8, and 25 ms). This procedure was expected to be insensitive to any errors in this model-based OAE group delay estimate, because the final spectrum included the spectral energy from all three windowed waveforms. The intent was to match the window ramp duration to the expected periodicity of the evoked OAE signal so as to minimize spectral splatter.

CEOAE spectra were calculated using a 1024-sample DFT of the corresponding zero-padded windowed CEOAE waveform. The magnitude spectra of the CEOAE signal and noise were separately calculated using the methods of Goodman et al. (2009). The duration of the non-zero energy varied across the three windows, which affected the overall magnitudes of the signal and noise. This factor was the same for each of the signal and noise spectra for a given window, so that the corresponding SNR was independent of the factor. The resulting SNR levels of the spectra of the early-, middle-, and late-windowed waveforms are plotted in Fig. 2 for the same subject whose waveforms are shown in Fig. 1. Aside from fluctuations across frequency, the SNR was approximately 5 dB or larger between 5 and 16 kHz for the early-windowed spectrum, 1–11 kHz for the middle-windowed spectrum, and 0.6–3.4 for the late-windowed spectrum. The spectral bandwidths were in the expected directions, but there was also substantial overlap, especially between early and middle windows, and middle and late windows. The sound exposure level (SEL, see Goodman et al., 2009) spectra of the early-, middle-, and late-windowed waveforms are plotted in Fig. 3 for the same subject whose data were shown in Figs. 1 and 2. The CEOAE SEL spectrum from the DFT calculated for the early window peaked at higher frequencies (4–16 kHz), that for the middle window peaked at intermediate frequencies (0.7–8 kHz), and that for the late window peaked at lower frequencies (0.5–4 kHz). The frequency-overlapped
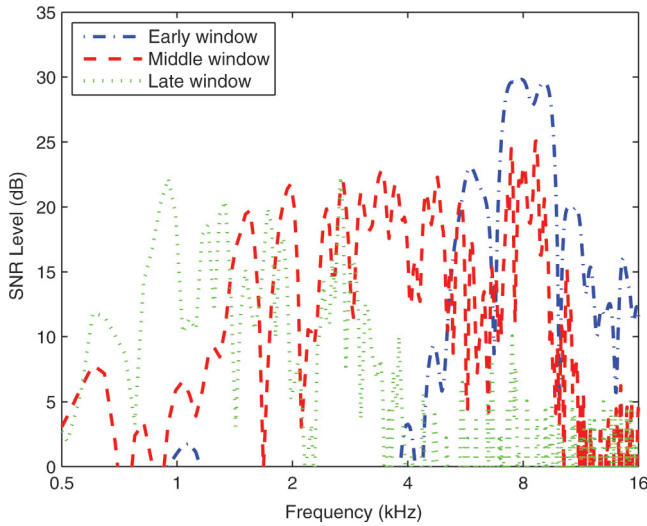
FIG. 2. (Color online) For the test ear of subject 1, the SNR levels are plotted for each windowed spectrum at the frequencies of the DFT. The early-windowed spectrum (dashed–dotted line) is peaked at high frequencies, the middle-windowed spectrum (dashed line) at intermediate frequencies, and the late-windowed spectrum (dotted) at low frequencies.

regions of these three spectra illustrate the need to include contributions of CEOAE energy from each windowed spectrum irrespective of the frequency of interest. For the case of absent CEOAEs and stationary noise throughout the entire waveform, the SEL spectral level would be largest for the late window, because its duration is the longest, and smallest for the early window because its duration is the shortest. This effect was present at 0.5 kHz in the SEL spectra and is consistent with a noise-dominated response at 0.5 kHz. It is also consistent with the low SNR levels near 0.5 kHz in Fig. 2. The fact that the SNR level in the early-windowed spectra was 0 dB at 1 kHz is consistent with the absence of a 1-kHz
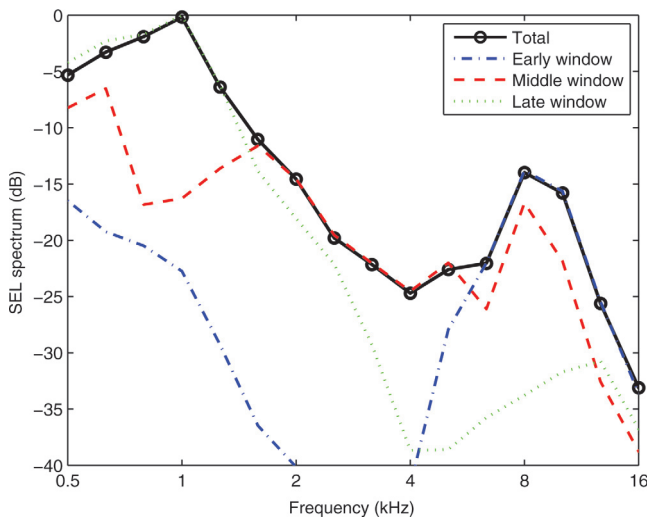


FIG. 3. (Color online) For the test ear of subject 1, the third-octave averaged signal levels are plotted as a SEL in decibels for each windowed spectrum and for the total spectrum (solid line with circles) calculated using SNR weightings. The early-windowed spectrum (dashed–dotted line) is peaked at high frequencies, the middle-windowed spectrum (dashed line) at intermediate frequencies, and the late-windowed spectrum (dotted) at low frequencies. The portion of the early-windowed SEL spectrum below −40 dB, which occurs between 2 and 4 kHz, is not plotted.

response at short times. Thus, the SEL spectral level of about −22 dB at 1 kHz for the early-windowed spectra means that this is a noise-dominated response at short times. Above 8 kHz, while the SEL spectrum for the middle-windowed CEOAE was nearly as large as the early-windowed CEOAE in Fig. 3, the SNR level for the middle-windowed CEOAE was smaller in relative terms than that for the early-windowed CEOAE in Fig. 2. These examples illustrate how the multi-window spectral approach was formulated based on SNR level.

The SNR was calculated as the ratio of the CEOAE signal magnitude to noise magnitude at each DFT frequency (see Fig. 2). A "total CEOAE" magnitude spectrum $S(f)$ was calculated as a normalized SNR-weighted sum of the CEOAE magnitude spectra calculated for each of the three windows ($S_i(f)$, $i = 1,2,3$ for early, middle, and late windows, respectively). With weighting coefficient $w_i(f)$ representing the SNR in the $i$th window, the total CEOAE spectrum is defined as

$$S(f) = \frac{w_1(f)S_1(f) + w_2(f)S_2(f) + w_3(f)S_3(f)}{w_1(f) + w_2(f) + w_3(f)}. \quad (1)$$

The total CEOAE spectrum was calculated at each DFT frequency, thereby taking into account the maxima and minima in each of the $w_i(f)$ $S_i(f)$, and averaged over each third octave. Each noise spectrum was calculated based on the incoherent spectrum (Schairer *et al.*, 2003; Goodman *et al.*, 2009). A total noise spectrum was calculated as the SNR-weighted sum of the noise spectra for each windowed spectrum at each DFT frequency based on an expression analogous to Eq. (1).

The resulting SEL spectrum of the total CEOAE is plotted as the solid line in Fig. 3. Due to the SNR weighting, the total spectrum included contributions from multiple windowed spectra. The detailed and third-octave averaged total SEL spectra are compared for the same subject in the top panel of Fig. 4 along with total noise levels (i.e., the "detailed" SEL spectrum was calculated at each frequency in the DFT). Averaging across each third octave smoothed the SEL spectrum around any narrow frequency band in which the level decreased, e.g., the spectral notch in Fig. 4 close to 10 kHz. The difference between the third-octave averaged signal and noise SEL spectra in the top panel defines the smoothed SNR level which is plotted in the bottom panel of Fig. 4. This may be compared to the SNR level plotted in Fig. 2 at each DFT frequency. This smoothed SNR level was used to predict hearing status at each octave or third-octave frequency with a larger SNR predicting an increased likelihood of normal hearing. The ear from which the data in Figs. 1–4 came had normal audiometric thresholds from 0.25 up to 8 kHz. The SNR of the CEOAE was at least 7 dB over all frequencies from 0.7 to 16 kHz. However, the CEOAE response was largely dominated by noise at 0.5 kHz.

Although detailed results are not plotted, the CEOAE SNR calculated by this multi-window spectral analysis was compared with the SNR calculated using a single-window spectral analysis for a set of ears with normal hearing. The main finding was a slight increase in the CEOAE SNR calculated using the multi-window spectral analysis, especially at
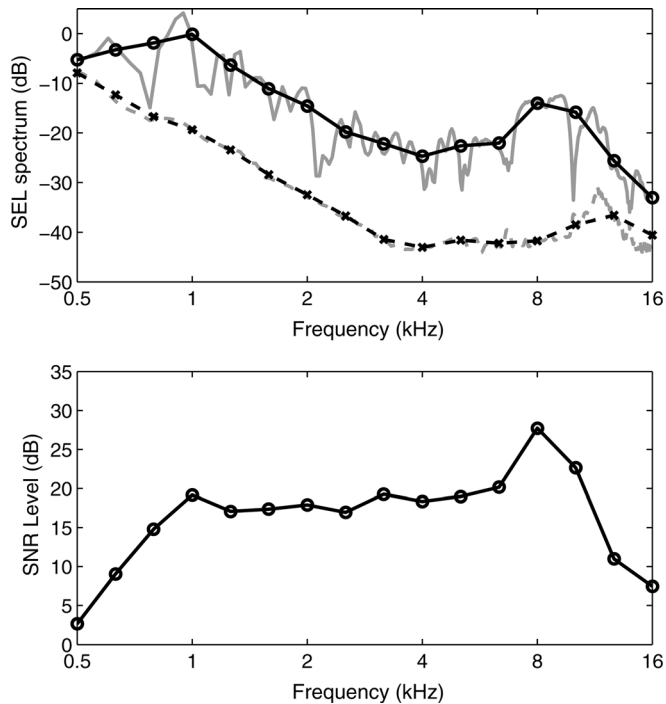
FIG. 4. Top panel: For the test ear of subject 1, the third-octave averaged total signal level (black solid line) and noise levels (black dashed line) are each plotted as a SEL in decibels. This total signal SEL is the same response plotted in Fig. 3. Also plotted are the detailed signal and noise levels without frequency averaging (gray lines with similar styles). Bottom panel: The CEOAE SNR level is plotted as the difference in the third-octave averaged signal and noise levels shown in the top panel.

frequencies for which the SNR was relatively low. Based on this finding, the multi-window spectral analysis was used in the main experiment.

## III. RESULTS

### A. Performance of the research protocol to measure behavioral threshold

A sidearm analysis evaluated the duration of the behavioral threshold test and the resulting SD in thresholds. There are few reports of such durations for this type of yes–no task, and no reports for the broad range of test frequencies that are the focus of this study. A first sample of threshold measurements from one ear of each of the 20 subjects was constructed by randomly sampling the normal-hearing group. Their group audiometric range was characterized on the basis of the pure-tone average audiogram (PTA) at 0.5, 1, and 2 kHz: The resulting group mean of the PTA was 5 dB HL with a SD of 3 dB. A second sample of threshold measurements in one ear of each of the 20 subjects was selected from the SNHL group. This selection was not based on a random sample. Instead, the PTA at 0.5, 1, and 2 kHz was calculated for each ear in the SNHL database, and subjects were selected based on the most elevated PTAs. The PTA for these 20 ears ranged from 58 to 90 dB HL with a group mean of 72 dB HL (and SD of 9 dB). Because HL at these intermediate frequencies was correlated with HL at higher frequencies, it was expected that these subjects would have elevated thresholds at higher frequencies,

although the high-frequency thresholds were not independently controlled.

The duration of each block of three runs, with each run composed of 15 trials, was measured as the time interval between the first stimulus presentation following the subject's initiation of the protocol and the final response acquired in each run. The measured durations of the first block of runs at each frequency are listed for the normal-hearing subjects in the top portion of Table I. The relevance of this duration is that each subject had at least a first block of runs at each frequency. The table lists the number of subjects ($N$) out of the 20 subjects with a threshold in the measurement range. In the normal-hearing group, all 20 subjects had a measured threshold (i.e., not NR) at all test frequencies except 16 kHz, at which 18 of the 20 subjects had a measured threshold. Of these 198 measured thresholds, 88% of them were obtained in the first block. The remaining 12% of thresholds had a SD that exceeded 5 dB, so that an additional second or third block of runs was completed until the criterion was met.

The duration of the first block ranged from 112 to 121 s over the ten frequencies in the normal group. The fact that the durations were similar at all frequencies suggests that subjects had no more difficulty completing the behavioral task at frequencies above 8 kHz than they did at lower frequencies. Because no SNHL subject had a measured threshold at 16 kHz, the median duration across frequency was calculated in the table for both groups over the frequency range 0.5–12.7 kHz. The median of the block 1 mean duration was 117 s in the normal group. Considering that each block included 45 yes–no trials, the subjects provided a yes–no response to the presentation of a tone at a rate of 2.6 s per trial. The block 1 SD of the test duration was calculated across the subjects at each frequency and varied from 12 to 39 s with a trend toward larger SDs in the upper half of frequencies. The median block 1 SD across frequency was 17 s.

Another measure of test duration was the total mean duration, which included the durations of additional blocks of runs that were needed in 12% of the conditions in the normal group. The total mean duration varied across frequency from 117 to 150 s, with a median of 127 s. This total duration was the sum of the durations of each test block but did not include the additional time between blocks during which the experimenter may have discussed the protocol with the subject, or in which the subject may have taken a rest break. The median total duration across all ten frequencies was 1270 s or about 21 min, which is within the total duration of protocols used in psychoacoustical research but too long for a clinical test.

A more critical measure of the research protocol was the variability in the thresholds across multiple runs for the same subject, in which the insertion of the probe in the ear canal was not altered. This was quantified by the SD of the threshold (decibels HL) calculated across the three runs in the final block of data, i.e., the block in which the subject had <5 dB SD. This SD for threshold varied from 1.06 to 2.60 dB in the normal group across frequency. There were no marked trends in the frequency dependence of this SD except for the larger SD of 2.60 dB at 0.5 kHz. In particular,

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions

TABLE I. Summary statistics for behavioral threshold measurements of test duration and threshold variability using maximum-likelihood yes–no task for normal and SNHL groups of subjects (1 ear per subject).

| Frequency (kHz) | N | Duration (s) Block 1 mean | Block 1 SD | Total | SD of threshold (dB) (Average across ears) |
|---|---|---|---|---|---|
| *Normal subjects (group mean PTA = 5 dB HL, SD = 3 dB)* | | | | | |
| 0.5 | 20 | 112 | 15 | 129 | 2.60 |
| 1 | 20 | 117 | 17 | 117 | 1.62 |
| 2 | 20 | 114 | 12 | 125 | 1.25 |
| 4 | 20 | 113 | 13 | 138 | 1.38 |
| 5 | 20 | 116 | 18 | 116 | 1.52 |
| 6 | 20 | 119 | 35 | 127 | 1.73 |
| 8 | 20 | 117 | 25 | 123 | 1.33 |
| 10.1 | 20 | 121 | 39 | 141 | 1.06 |
| 12.7 | 20 | 118 | 14 | 137 | 1.47 |
| 16 | 18 | 121 | 34 | 150 | 1.66 |
| Median (0.5–12.7) | | 117 | 17 | 127 | 1.47 |
| IQR (0.5–12.7) | | 4.3 | 11.3 | 14.2 | 0.29 |
| *SNHL subjects (group mean PTA = 72 dB HL, SD = 9 dB)* | | | | | |
| 0.5 | 19 | 122 | 15 | 154 | 1.15 |
| 1 | 20 | 172 | 98 | 177 | 1.67 |
| 2 | 14 | 123 | 15 | 141 | 0.92 |
| 4 | 13 | 118 | 15 | 127 | 0.80 |
| 5 | 12 | 118 | 12 | 146 | 1.12 |
| 6 | 12 | 118 | 38 | 118 | 1.55 |
| 8 | 11 | 120 | 13 | 174 | 1.57 |
| 10.1 | 8 | 120 | 14 | 120 | 1.11 |
| 12.7 | 2 | 133 | 4 | 133 | 1.45 |
| 16 | 0 | NR | NR | NR | NR |
| Median (0.5–12.7) | | 120 | 15 | 141 | 1.15 |
| IQR (0.5–12.7) | | 4.4 | 1.9 | 26.7 | 0.44 |
| | 111 | | | | |

N = Number of subjects out of 20 with threshold in measurement range.
Block 1 mean = Mean duration in first block of three trials.
Block 1 SD = SD of duration in first block of three trials.
Total = Mean total duration across all blocks of trials needed for each subject.
SD of threshold = SD of threshold (dB HL) averaged over N subjects.
Median = Median of each column variable across frequencies 0.5–12.7 kHz.
IQR = Inter-quartile range of each column variable across 0.5–12.7 kHz.
Notes: Medians and IQRs were calculated from 0.5 to 12.7 kHz because the SNHL group had no subjects with data at 16 kHz. Except for the Block 1 SD, the normal group responses at 16 kHz were within the IQR of the medians between 0.5 and 12.7 kHz.

the SDs did not appear larger at frequencies above 8 kHz. The median SD across frequency was 1.47 dB for the normal group with an IQR of 0.29 dB.

A crucial difference between the SNHL group and the normal group was the difference in the numbers of subjects (N) with measured thresholds. In the SNHL group, N was 19 and 20 at the lowest frequencies (0.5 and 1 kHz, respectively) but was in the range from 8 down to 0 above 8 kHz. Only 111 of the total of 200 conditions resulted in a measured threshold in the SNHL group. This is as expected given the selection criteria and would suggest lower confidence in the duration and threshold statistics at higher frequencies in the SNHL group.

Notwithstanding this limitation, the mean duration of the runs in block 1 across frequency ranged from 118 s to 172 s in the SNHL group. There were no pronounced trends in the frequency dependence of these durations, except for the longer duration (172 s) at 1 kHz. Of the 20 subjects tested at

1 kHz, the mean was increased primarily because one subject had a block-1 duration of 472 s and another had a duration of 342 s. Because thresholds were always measured initially at 1 kHz, this longer duration was probably due to a training effect in the SNHL group for these two subjects. The block-1 durations for these two subjects did not exceed 182 s in any subsequent test.

The median duration of block 1 in the SNHL group was 120 s, which was similar to that for the normal group (117 s) in relation to their group IQRs of 4.3–4.4 s. Aside from the larger SD of 98 s in the block-1 duration at 1 kHz, which occurred because of the effects described above for two subjects, the SD for block 1 had little variation across frequency. The fact that it was only 4 s at 12.7 kHz is of limited importance because the SD was based on an N of 2. The median SD for block 1 was 15 s for the SNHL group, which was within the IQRs of the median SD for the normal group.

Of the 111 conditions resulting in a measured threshold in the SNHL group, the threshold was measured in the first block in 103 conditions, or 93% of the time. For the other eight conditions, two blocks were required to measure the threshold in five conditions, and three blocks in three conditions. The total mean duration of all blocks was similar in the SNHL and normal groups, with a median of 141 s for the SNHL group, which was close to the total duration for the normal group to within their respective IQRs.

The SD of the audiometric thresholds in the SNHL group ranged from 0.80 to 1.67 dB with a median across frequency of 1.15 dB and an IQR of 0.44 dB. There was no effect of frequency on the SD. The median SD was similar in SNHL and normal groups to within their IQRs. Thus, the efficiency of the research protocol to measure thresholds was similar across frequency from 0.5 to 16 kHz, and similar between normal-hearing subjects and subjects with SNHL.

This level of efficiency appears slightly better than that reported by Lecluyse and Meddis (2009). Their results showed an average SD of threshold of 1.8 dB in nine normal-hearing subjects using their single-interval adaptive protocol, compared to 3.4 dB in a two-interval forced choice protocol. Their single-interval adaptive protocol was based on the SD across five runs, each composed of ten trials (50 trials total). The present procedure was based on three runs of 15 trials in the first block (45 trials total) followed by additional blocks in those subjects (12% of normal subjects and 7% of subjects with SNHL) whose SD exceeded 5 dB in the first block.

A more careful comparison of the single-interval adaptive protocol of Lecluyse and Meddis (2009) and the research protocol used in the present study, which was tested for reliability in Goodman *et al.* (2009), would require testing in the same group of subjects. Nevertheless, it appears that both protocols performed well for estimating behavioral thresholds. Variants of the research protocol in the present study have also been used to objectively and automatically estimate thresholds of SFOAE responses for tones in notched noise and in quiet conditions (Keefe *et al.*, 2009) and thresholds of the acoustic reflex (Keefe *et al.*, 2010).

## B. CEOAE predictions of SNHL

The SNR-weighted CEOAE signal and noise spectral levels were measured in each test ear at third-octave frequencies from 0.5 to 16 kHz, and the smoothed SNR was calculated as the difference in these levels at each frequency (see single-ear example in Fig. 4). Preliminary analyses confirmed that CEOAE SNR was a better univariate predictor of SNHL than CEOAE level. Therefore, CEOAE SNR was selected to predict whether the audiometric HL for that ear was within normal limits or impaired with octave frequency resolution from 0.5 to 4 kHz, and third-octave frequency resolution from 4 to 16 kHz. A criterion HL was defined as the highest HL included in the normal-hearing group and was varied from 10 to 40 dB in 5-dB steps. As an example, for a criterion HL of 20 dB, an ear was classified as having a SNHL if the audiometric HL exceeded 20 dB and was classified as normal if it was ≤20 dB HL. For both lower and higher frequencies, the third-octave averaged SNR of the CEOAE

centered at each audiometric frequency was used to predict audiometric status at that frequency. Test performance was analyzed using the non-parametric area under the receiver operating characteristic curve (AUC) at each of the four click stimulus levels (76, 73, 67, and 61 dB peSPL).

The receiver operating characteristic (ROC) curve is the plot of test sensitivity versus one minus test specificity, in which the sensitivity is the ratio of impaired ears correctly classified as impaired and the specificity is the ratio of normal ears correctly classified as normal. An ideal test would correctly identify all ears, resulting in a sensitivity and specificity equal to one. A "good" test is one that has a sufficiently high sensitivity paired with a sufficiently high specificity. The meaning of "sufficiently high" depends on many factors, including the prevalence of HL and cost to the health-care system of correctly or incorrectly identifying ears with normal function or SNHL. These factors vary with the target population (for example, babies in the well baby versus intensive care nursery) and financial considerations that may be idiosyncratic to individual health-care providers. Thus, it would be difficult, if not impossible, to develop definitions of sufficiently high that would be generally applicable.

An alternative measure of test performance to AUC is the point of symmetry (SYM) on the ROC curve at which the difference between the sensitivity and specificity is minimized. The SYM point on the ROC curve is close to one when both the specificity and sensitivity are relatively high; that is, SYM defines a test that is equally accurate at classifying normal ears as normal and impaired ears as impaired. Further comparisons of AUC and SYM are described elsewhere (Pepe, 2003; Sanford *et al.*, 2009).

Except for the 1-kHz predictor at the highest stimulus level, AUC decreased slightly with increasing HL criterion with a maximum AUC at 10 dB HL. The 1-kHz predictor had the largest AUC at a HL criterion of 20 dB at the higher stimulus levels (76 and 73 dB peSPL). This suggests that the present paradigm is sensitive to small amounts of threshold elevation, even those within the range that is typically considered normal. By extension, this suggests that the SNR decreases (as does CEOAE level) as behavioral threshold increases, even within the normal range. AUC was larger at higher stimulus levels (76 and 73 dB peSPL) than at lower levels (67 and 61 dB peSPL) across all criterion HLs, which focused interest on these higher stimulus levels. The AUC was lower at 0.5, 12.7, and 16 kHz than at intermediate frequencies (1–10.1 kHz) irrespective of the HL criterion or stimulus level. A test criterion of 20 dB HL or better for the normal group is commonly used clinically, in as much as 25 dB HL represents a borderline SNHL. Because AUC varied only weakly with HL criterion overall and because of its clinical utility (despite the observation of slightly better performance when the audiometric criterion was 10 dB HL), the main analyses were performed using this 20 dB HL criterion.

The average results for the present study are plotted in Fig. 5. The 95% range of AUC is shown as a dark-gray band for the CEOAE test at the highest stimulus level (76 dB peSPL) and as a light gray band for the next highest level (73 dB peSPL). The ranges for AUC were calculated using a bootstrap procedure based on 10 000 iterations. Such a large
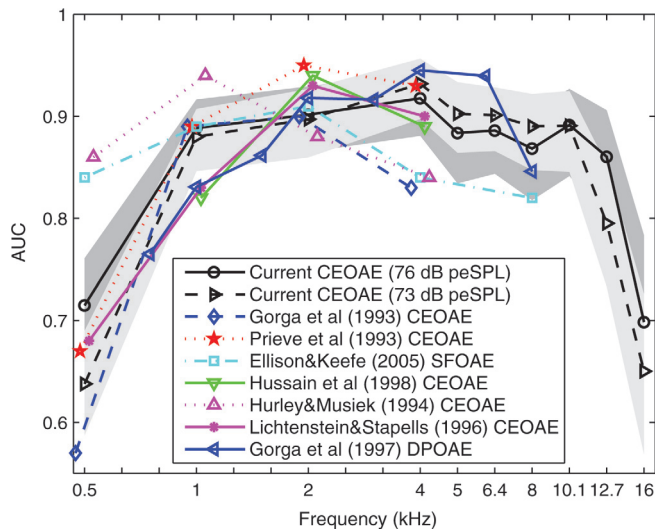
FIG. 5. (Color online) The AUC is plotted versus frequency for CEOAE predictions of auditory status from the present study at 76 and 73 dB peSPL stimulus levels, CEOAE predictions from Gorga *et al*. (1993), Prieve *et al*. (1993), Hurley and Musiek (1994), Hussain *et al*. (1998), and Lichtenstein and Stapells (1996), SFOAE predictions from Ellison and Keefe (2005), and DPOAE predictions from Gorga *et al*. (1997). For the present study and other studies where possible, the prediction was for audiometric status with normal ears having a HL of 20 dB or better. AUC values from the present CEOAE study are shown with their 95th percentiles in fill patterns of light gray (73 dB peSPL) and dark gray (76 dB peSPL). There is some overlap between these light- and dark-gray regions of AUC values. Each AUC function is shifted horizontally with respect to its nominal center frequency to improve visibility. See inset for line style and symbol corresponding to each study.
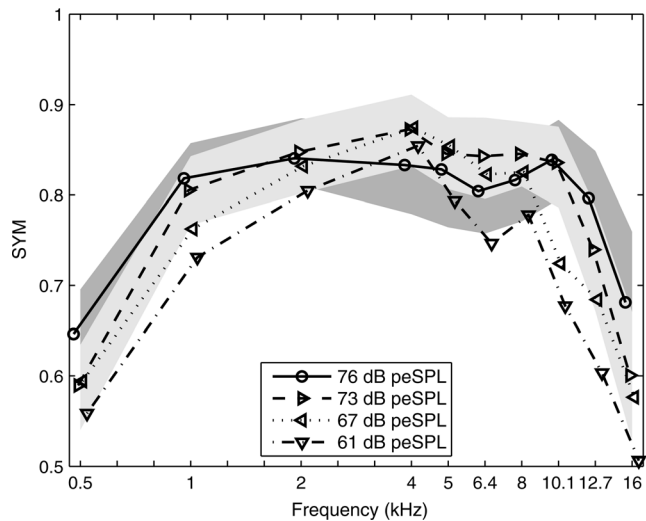


FIG. 6. The SYM is plotted versus frequency for CEOAE predictions of auditory status from the present study at 76, 73, 67, and 61 dB peSPL stimulus levels. The prediction was for audiometric status with normal ears having a HL of 20 dB or better. SYM values from the present CEOAE study are shown with their 95th percentiles in fill patterns of light gray (73 dB peSPL) and dark gray (76 dB peSPL). There is some overlap between these light- and dark-gray regions of SYM values. Each SYM function is shifted horizontally with respect to the nominal center frequency to improve visibility.

number of iterations are recommended to accurately calculate the 95th percentiles (Efron and Tibshirani, 1986).

The AUCs for stimulus levels of 73 and 76 dB peSPL were relatively constant between 1 and 10.1 kHz. The mean and SD of AUC over these frequencies was 0.89 (SD of 0.015) at 76 dB peSPL and 0.90 (SD of 0.0164) at 73 dB peSPL. The CEOAE test at the 76 dB stimulus level performed slightly better at the lowest (0.5 and 1 kHz) and highest (12.7 and 16 kHz) frequencies, and slightly worse at intermediate frequencies (4–8 kHz) compared to results for the 73 dB peSPL condition. The AUC at 12.7 kHz was only slightly reduced to 0.86 at 73 dB peSPL compared to its mean of 0.90 at 76 dB peSPL. At the lowest and highest test frequencies (0.5 and 16 kHz), the AUC values across stimulus level were between 0.63 and 0.72, which suggests that the test performance was inadequate at these frequencies.

The SYM values for the CEOAE test are plotted for the 20 dB HL criterion in Fig. 6 as a function of frequency for each of the four stimulus levels. These SYM values were estimated from the same ROC curves that produced the AUC values shown in Fig. 5. The 95% range of SYM based on 10 000 bootstrap iterations is also plotted but only for the two highest stimulus levels (73 and 76 dB peSPL). These SYM ranges were similar at the two lowest stimulus levels, although they are not plotted in order to emphasize the results at the highest levels. The qualitative shapes of the SYM curves were similar to the shapes of the AUC curves, e.g., a low SYM at 0.5 and 16 kHz. This is not surprising, since the data in Figs. 5 and 6 were derived from the same

ROC curves. SYM was relatively constant over the range from 1 to 10.1 kHz with the following means (and SDs): 0.83 (0.013) at 76 dB, 0.84 (0.020) at 73 dB, 0.81 (0.052) at 67 dB, and 0.77 (0.057) at 61 dB peSPL. SYM at each frequency increased relatively uniformly with increasing stimulus level from 61 to 73 dB peSPL, whereas SYM at 76 dB peSPL was larger than at any lower stimulus level at 0.5–1 kHz and 12.7–16 kHz and smaller at intermediate frequencies (2–8 kHz). This is probably related to the fact that the CEOAE predictions at 61, 67, and 73 dB peSPL were based on the same group of ears ($N = 446$), while the CEOAE prediction at 76 dB peSPL was based on 39 fewer ears in the normal group ($N = 407$).

## C. Repeatability

The repeatability of CEOAE and research audiometry tests were assessed in one ear of each of 41 subjects who completed all testing on two dates separated on average by 8.5 days (SD of 8 days, ranging from 1 to 31 days). An additional inclusion criterion was that the SNR level of the CEOAE test and retest responses had to be at least 6 dB, which paralleled an analysis of test repeatability of high-frequency distortion-product otoacoustic emission (DPOAE) responses (Dreisbach *et al*., 2006). This was done to avoid measuring variability when no emission was present. This criterion was imposed at each third-octave frequency; thus, responses from the same pair of subject tests might be included at some frequencies but not others.

The repeatability of measuring the CEOAE spectrum was assessed in terms of the magnitude difference ($\Delta$CEOAE) of the third-octave averaged SEL spectrum of the CEOAE measured in the test and retest, i.e., $\Delta$CEOAE was non-negative. The percentiles (median, 25th, 75th, and 90th) of
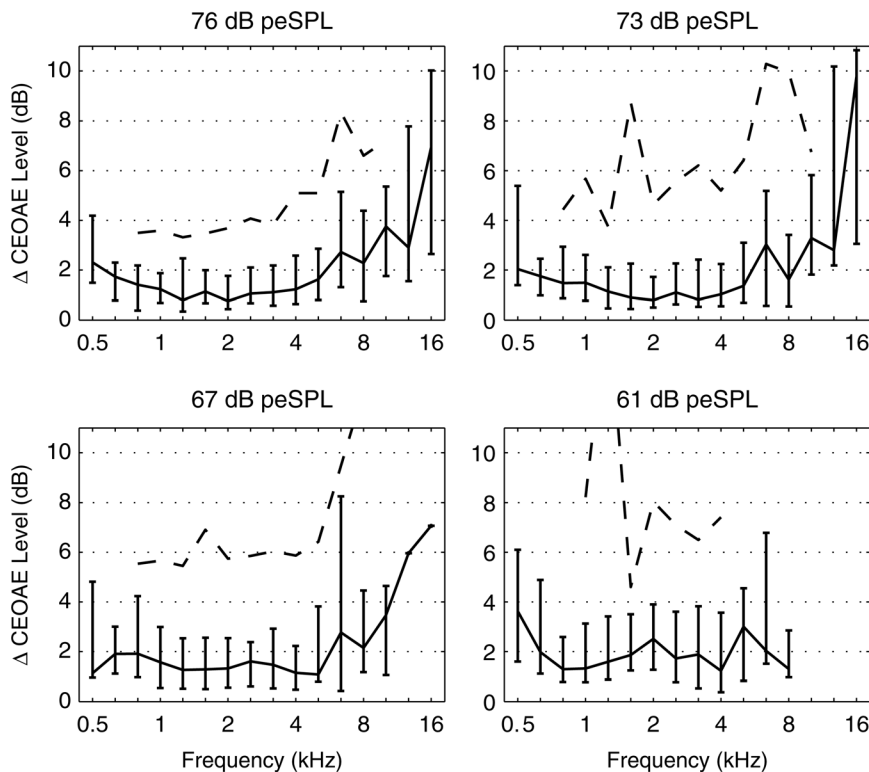
J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe *et al*.: Detecting hearing loss click otoacoustic emissions    255

76 dB peSPL      73 dB peSPL

67 dB peSPL      61 dB peSPL

FIG. 7. Test–retest analysis of the absolute value of the CEOAE level difference. Each panel shows the median level difference (line) with the lower and upper quartiles plotted as error bars. Each panel corresponds to the click stimulus level (decibels peSPL) shown at the top of the panel. The 90th percentile is plotted as a dashed line at frequencies that included at least 20 responses. Two 90th percentiles of CEOAE level differences were outside the plotting range at stimulus levels of 67 and 61 dB peSPL. Their values are as follows: 12.5 dB at 8 kHz in the lower left panel (67 dB peSPL), and 16.2 dB at 1.3 kHz in the lower right panel (61 dB peSPL).

this level difference are plotted in Fig. 7 across the four stimulus levels. At the two highest stimulus levels, the median CEOAE level difference had a minimum of approximately 1 dB from 1 to 4 kHz, was <3 dB from 0.5 to 8 kHz, and <4 dB from 0.5 to 12.7 kHz. At the two highest levels, the IQRs of the level difference (shown by the error bars in Fig. 7) were within 6 dB from 0.5 to 10.1 kHz, but ranged as high as 8 to 11 dB at 12.7 and 16 kHz. The 90th percentile of the CEOAE level difference is plotted only at those frequencies for which there were at least 20 responses, i.e., at frequencies from 0.8 to 10.1 kHz. The 90th percentile of the CEOAE level difference varied from approximately 4 dB at lower frequencies to 6–10 dB at higher frequencies.

The audiometric threshold was measured in both the test and retest using the research protocol, and its repeatability was assessed using the magnitude difference ($\Delta$HL) of the HL. It is relevant to investigate the extent to which audiometric data measured using a research protocol were repeatable, because the audiometric criterion is first selected and then used to evaluate CEOAE test performance. The top panel of Fig. 8 shows the percentiles of $\Delta$HL as a function of audiometric frequency. The median $\Delta$HL varied from 2 to 4 dB over frequencies from 0.5 to 12.7 kHz, with IQRs of 8 dB at 0.5 kHz and ranging from 4 to 7 dB from 1 to 12.7 kHz. The 90th percentile is plotted from 1 to 10.1 kHz, which are those frequencies at which both the test and retest CEOAE SNR levels exceeded 6 dB at the highest click stimulus level. The 90th percentile in $\Delta$HL was approximately 6–7 dB from 1 to 10.1 kHz and 11 dB at 12.7 kHz.

The CEOAE SNR level was used to predict hearing status rather than CEOAE level, so that its repeatability was independently assessed using the magnitude difference ($\Delta$SNR) of the SNR level of the CEOAE measured in test and retest.
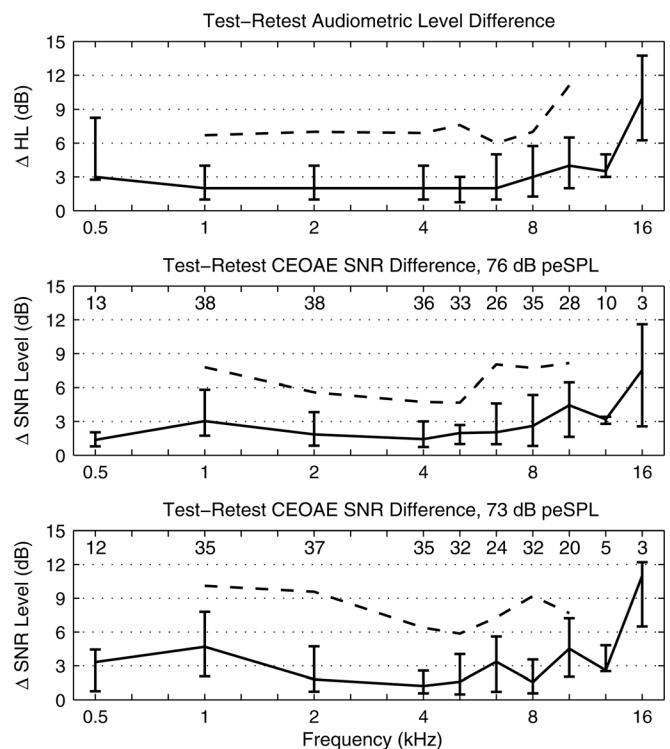


FIG. 8. Test–retest analysis of the absolute value of the difference in HL measured using the research protocol (top panel), and the absolute value of the difference in the CEOAE SNR level for a click stimulus level of 76 dB peSPL (middle panel) and 73 dB peSPL (bottom panel). Each panel shows the median level difference (line) with lower and upper quartiles shown as error bars. The numbers at each frequency at the top of the middle and lower panels are the number of responses that met the inclusion criterion (SNR level > 6 dB). The 90th percentiles are plotted with dashed lines at frequencies that included at least 20 responses. The HL difference in the top panel plots the percentiles based on the same sub-groups of subjects as in the middle panel.

The percentiles of ΔSNR are plotted for the highest click level (76 dB peSPL) in the middle panel of Fig. 8 and for the next highest click level (73 dB peSPL) in the lower panel. In addition, the numbers of subjects included in the percentiles at each frequency are given in each panel, which illustrates that the 90th percentiles are plotted only when this number was 20 or more. At 76 dB peSPL, the median ΔSNR of the CEOAE was approximately 1–4 dB between 0.5 and 12.7 kHz, and 7.4 dB at 16 kHz. The 75th percentiles of ΔSNR varied from 3 to 7 dB between 1 and 10.1 kHz, and the 90th percentiles from 5 to 8 dB. The percentiles of the ΔSNR at 73 dB peSPL generally overlapped the percentiles at 76 dB peSPL. Overall, the repeatability of the audiometric threshold and CEOAE SNR level were similar over the frequency range from 0.5 to 12.7 kHz, with increased variability in both responses at 16 kHz.

### D. Measurements in subjects with cochlear implants

CEOAE responses were measured in five subjects with profound HL who use a cochlear implant. These subjects were tested to analyze responses in ears that would not be expected to have a cochlear-generated signal, at least according to the outer-hair cell mechanisms by which CEOAEs are generated in ears with normal auditory function. These subjects were tested with the implant powered off. As explained in Sec. II, the "CEOAE" SNR level was calculated in terms of the nonlinear pressure residual derived from responses to three click stimuli at different presentation levels. The percentiles of this SNR level measured in the five ears with cochlear implants are shown in the top panel of Fig. 9 at the highest click level (76 dB peSPL) and in the bottom panel at the next highest click level (73 dB peSPL).
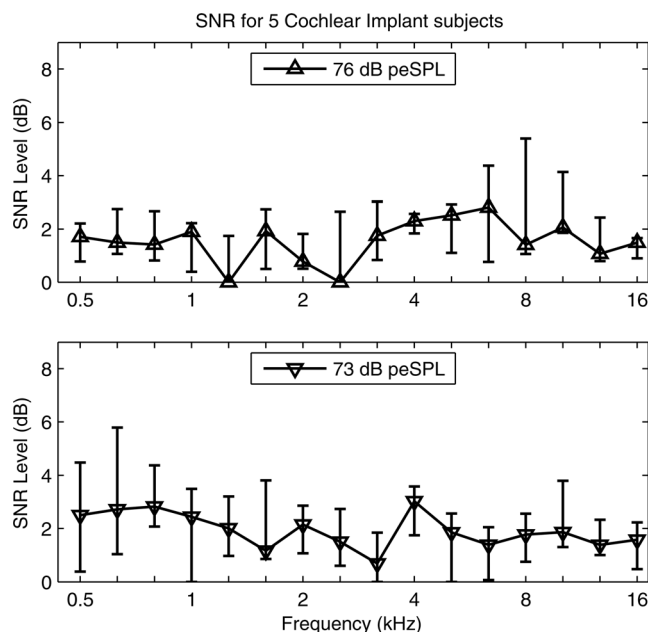


FIG. 9. The group statistics for the CEOAE SNR level are plotted for responses in five ears of five subjects with cochlear implant in the test ear. The error bar at each frequency shows the median and IQR. SNR levels are plotted for click stimulus levels of 76 dB peSPL (top panel) and 73 dB peSPL (bottom panel).

The median SNR level was approximately 2 dB at all frequencies (0.5–16 kHz) and all stimulus levels. The 75th percentiles of SNR level were bounded above by 4 dB at frequencies between 1 and 10.1 kHz and all stimulus levels, except that the 75th percentile was 5.5 dB at 8 kHz and 4.1 dB at 10.1 kHz at the highest stimulus level. This measurement of SNR in subjects with profound HL assessed the biological noise + distortion floor of the CEOAE measurement, including any biological sources of distortion within the cochlea. The resulting SNR was acceptably low for a CEOAE test procedure.

To the extent that any implant ear had a SNR level exceeding noise fluctuations, a click-evoked biological response might have arisen from either a nonlinear middle-ear or cochlear effect. Inasmuch as the normal-hearing ears in the Goodman et al. (2009) study showed no evidence of a short-latency middle-ear response, it appears unlikely that the middle ear would be the source of the response in implant ears. The fact that these subjects were using cochlear implants suggests that no active cochlear mechanism associated with outer-hair cell function was present to generate a biological response. The main outcome from the cochlear-implant measurements was that their median SNR levels were in the range of 0–3 dB. If measurement-system distortion had influenced these measurements, then the SNR level should have been larger at the higher stimulus level, but this did not generally appear to be the case. It is possible that some residual cochlear nonlinearity associated with the passive mechanics of the basilar membrane might have played a role in generating a response, or that there was additional measurement noise in the implant-ear measurements. Further study is needed.

## IV. DISCUSSION

### A. CEOAE test performance

The main finding is that CEOAEs accurately identified the presence of a SNHL at frequencies from 1 kHz up to at least 10 kHz. For the best performing stimulus level of 73 dB peSPL, the average AUC over this frequency range was 0.90 with an average SYM of 0.84. The latter means that the CEOAE test correctly identified SNHL at any frequency from 1 to 10 kHz in 84% of impaired ears, and correctly identified an ear with normal hearing in 84% of ears with normal hearing. The average ROC summary statistics were only slightly reduced at 76 dB peSPL, with an average AUC of 0.89 and average SYM of 0.83.

The test accuracy in terms of AUC was similar to that in other evoked-OAE studies, as illustrated by the results in Fig. 5. Each study was based on testing in adult ears and sometimes in ears of older children. For the highest pair of test levels at the 20 dB HL criterion, the frequency dependence of AUC in the present study is compared to published AUC results on predicting SNHL using CEOAEs (Gorga et al., 1993; Prieve et al., 1993; Hurley and Musiek, 1994; Lichtenstein and Stapells, 1996; Hussain et al., 1998), SFOAEs (Ellison and Keefe, 2005), and DPOAEs (Gorga et al., 1997). A frequency-specific evoked OAE was used to predict hearing status at the same frequency in each of these studies.

J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe et al.: Detecting hearing loss click otoacoustic emissions    257

Except for the DPOAE study of Gorga *et al.* (1997) and the SFOAE study of Ellison and Keefe (2005) that reported AUC up to 8 kHz, all of the other studies were limited to an upper test frequency of 4 kHz. In particular, no previous CEOAE study has examined the frequency range above 4 kHz using a ROC-based measure of test performance. If for no other reason, this may be because most available measurement systems provide no information at these higher frequencies due to the limits imposed by their choice of analysis windows and method of extracting a nonlinear CEOAE residual.

It is problematical to compare test performance of different OAE tests of SNHL across studies, mainly because different studies used different sample populations as well as different testing procedures. For example, different test durations were used in various studies, and AUC would tend to increase with increasing test duration. Nevertheless, the general similarity of AUC in different studies has been described (Fig. 5). The most interesting differences are those in which the AUCs from other published OAE studies in adults were outside the 95th percentiles of AUC in the present CEOAE study. None of the other studies reported a measure of variability of their AUC value. The CEOAE study of Hurley and Musiek (1994) achieved a larger AUC at 0.5 and 1 kHz, and a smaller AUC at 4 kHz, than in the present study. The SFOAE study of Ellison and Keefe (2005) achieved a larger AUC at 0.5 kHz and a smaller AUC at 4 and 8 kHz. The better performance of SFOAEs at 0.5 kHz may be due to increased stimulus energy at that frequency compared to the wideband stimulus used to elicit CEOAEs, and the fact that the physiological noise contributing to the OAE response has larger energy at lower frequencies. The CEOAE study of Prieve *et al.* (1993) had a larger AUC at 2 kHz, whereas the CEOAE study of Gorga *et al.* (1993) had a smaller AUC at 4 kHz. The DPOAE study of Gorga *et al.* (1997) had a smaller AUC at 1 kHz and larger AUC near the third-octave frequency of 6.3 kHz. A general tendency is that the absolute range of AUCs across all studies was largest at 0.5 kHz, smallest at 2 and 8 kHz, and intermediate at 1, 4, and 6.3 kHz. However, within the range of frequencies for which comparisons were possible, there was general agreement in test performance despite the variations in OAE types, subject inclusion, and other aspects of the measurement paradigms in the studies summarized in Fig. 5.

Each evoked OAE test has individual differences related to whether a wideband or frequency-specific stimulus is presented and the procedure by which the stimulus level is adjusted in the ear canal. While analyses of DPOAE stimulus conditions are outside the scope of this study, it is relevant to contrast CEOAE and SFOAE tests. A constraint of this high-frequency CEOAE study was the fact that a wide bandwidth of stimulus energy was presented as a short-duration signal. Assuming that the source mechanisms underlying SFOAEs and CEOAEs are similar, a larger stimulus level can be achieved in a single-frequency SFOAE stimulus than in a wideband click stimulus at the frequency of the SFOAE. Thus, when test performance is limited by insufficient stimulus level, this would be expected to be a more critical problem for measuring CEOAEs and would explain the better

test performance of SFOAEs at 0.5 kHz. This is likely related to the fact that middle-ear transmission is less efficient at 0.5 kHz and above 10 kHz. A corresponding advantage of the high-frequency CEOAE test is its ability to assess cochlear functions across a wide bandwidth in a single test. Thus, there is a tradeoff in test design between a more frequency-selective stimulus versus using a wide-bandwidth stimulus. The present results suggest that stimulus levels may have been insufficient below 1 kHz and above 10 kHz. The AUC and SYM values at 12.7 kHz at the highest stimulus level (76 dB peSPL) were larger than their respective values at the next highest stimulus level (73 dB peSPL) and were only slightly below their mean AUC and SYM values from 1 to 10 kHz (see Figs. 4 and 5). This suggests that an additional increase in stimulus level in the third octave at 12.7 kHz might further increase CEOAE test accuracy toward its accuracy in the 1–10 kHz range.

While AUC and SYM increased somewhat with increasing stimulus level at 0.5 kHz, the resulting accuracy remained low compared to that in the 1–10 kHz range. This may suggest that a larger stimulus level near 0.5 kHz may be needed to accurately predict auditory status at 0.5 kHz, or that other factors may be involved. For example, it is possible that the OAE test duration of 21.3 ms was too short to fully include the CEOAE signal at 0.5 kHz (Prieve *et al.*, 1993; Jedrzejczak *et al.*, 2009). However, the AUC values at 0.5 kHz reported by Lichtenstein and Stapells (1996) for CEOAEs using clicks were slightly lower for a 30-ms CEOAE than for a 20-ms CEOAE, which would suggest that added noise energy between 20 and 30 ms at 0.5 kHz diminished performance more than any added signal energy improved performance. Lichtenstein and Stapells also found that the AUC at 0.5 kHz improved for CEOAEs elicited using a brief tone burst at 0.5 kHz, which would have increased the stimulus level at 0.5 kHz compared to a click stimulus. Increasing the CEOAE duration would reduce the number of averages available within any fixed overall measurement duration and would not be advantageous for the high-frequency CEOAE measurements that were the most novel part of the present study.

It remains to be determined how different mechanisms interact to limit the magnitude of the recorded CEOAE in the ear canal at stimulus frequencies above 10 kHz and below 1 kHz, which correspond to CEOAE source locations in the most basal and apical regions of the cochlea. If forward transmission through the ear canal and middle ear is less efficient at these frequencies, then increasing the stimulus level should increase CEOAE level and thereby increase the accuracy in detecting SNHL. If reverse transmission through the ear canal and middle ear is less efficient, then an increased stimulus level must generate a correspondingly larger OAE signal to be detectable via an inefficient reverse pathway. A reverse transmission deficit may be more difficult to overcome than a forward transmission deficit due to the compressive growth of the OAE signal generated on the basilar membrane. If the cochlear nonlinearity related to normal outer-hair cell function is relatively weaker at 0.5 kHz and 12.7–16 kHz than at frequencies between 1 and 10 kHz, then the cochlear-source signal would also be weaker. Such a reduction in the OAE source strength on the basilar membrane might be expected in

the most basal region of the basilar membrane as the distance between base and tonotopic place becomes on the order of the wavelength of the cochlear traveling wave.

The stimulus was designed to have an approximately flat incident voltage spectrum at the output of the microphone preamplifier from 0.5 to 16 kHz. After including the frequency-dependent sensitivity of the microphone, the pressure spectrum level was flat to within $\pm 3$ dB from 0.5 to 8 kHz, but its level was reduced by approximately 5 dB at 12.7 kHz and 14 dB at 16 kHz relative to the level at 8 kHz (see related discussion of Fig. 1 in Goodman *et al.*, 2009). Future studies might compensate for this stimulus level reduction at 12.7 and 16 kHz and might further increase the level above the nominally flat ear-canal level at these extreme frequencies. Such increases would offset the reduced forward ear-canal and middle-ear transmission and improve the ability to measure evoked OAE responses above 10 kHz if a forward transmission deficit is a dominant mechanism limiting the strength of the CEOAE. The findings from the present study are insufficient to address the question of the mechanisms limiting test performance at 0.5 kHz and above 12 kHz.

The $\Delta$CEOAE repeatability data in normal-hearing adults (see Fig. 7) are compared to the $\Delta$DPOAE repeatability data (the magnitude difference in DPOAE level between test and retest) measured over frequencies from 2 to 15 kHz in normal-hearing adults (Dreisbach *et al.*, 2006). The DPOAE measurements were performed with a stimulus frequency ratio of $f_2 / f_1 = 1.2$, and with the level of the $f_1$ stimulus ranging from 60 to 70 dB SPL and the $f_2$ stimulus ranging from 45 to 60 dB SPL. Dreisbach *et al.* calibrated stimulus levels in the ear canal. The average $\Delta$DPOAE was 2.8 dB (with a SD of 2.7 dB) from 2 to 8 kHz, and 5.2 dB (with a SD of 4.4 dB) above 8 kHz up to 15 kHz. The median $\Delta$CEOAE in the present study was 1 dB from 1 to 4 kHz, $<3$ dB from 0.5 to 8 kHz, $<4$ dB from 0.5 to 12.7 kHz, and 7 dB at 16 kHz. Except for the increased $\Delta$CEOAE at 16 kHz, the distribution of $\Delta$CEOAE between 2 and 12.7 kHz overlapped but tended to be slightly lower in level overall compared to the distribution of $\Delta$DPOAE at similar frequencies in Dreisbach *et al.* (2006). Factors affecting these repeatability estimates include possible differences in subject populations, the type of evoked OAE, and the measurement protocols with variables such as averaging time, stimulus level, and calibration method. Both studies demonstrated a satisfactory level of repeatability for using high-frequency evoked OAEs to detect a SNHL. Applications such as ototoxicity monitoring that require longitudinal measurements of evoked OAEs are constrained by this baseline variability in repeated measurements of OAE level. A relevant observation based on Figs. 6 and 7 is that the repeatability in the OAE measurements was generally similar to the repeatability of the behavioral audiogram. This is important because the audiometric data defines the hearing status of the test ear, yet it also has inherent variability.

The duration of the CEOAE test was 5.16 min per ear for all frequencies. This duration may be too long for a clinical screening test of cochlear status. The CEOAE SNR for stationary white noise would decrease by 3 dB for each halving of test duration. Assuming that the noise in the actual

OAE test was adequately described by this theoretical model, a 1.3-min test duration would be achieved by accepting a SNR reduction of 6 dB. The SNR in a normal-functioning ear with robust CEOAEs was at least 18 dB between 1 and 10 kHz (see bottom panel of Fig. 4), so such a reduction in test duration might be achievable. More study is needed to assess the reduction in test performance (i.e., on AUC and SYM) that would result from a shorter duration. This trade-off exists for other types of OAE measurements as well.

While the accuracy of CEOAE test performance was analyzed using ROC analyses, such an approach does not explicitly provide the critical SNR levels that separated normal and hearing-impaired ears at each test frequency. One possible approach in defining a predictive test is to maintain a fixed value of the critical SNR level at each frequency. A drawback with this approach is that the accuracy of the predictive test will generally differ across frequency. An alternative approach is to set the critical SNR level based on some specified accuracy derived from the ROC curve. This alternative was explored by calculating the critical SNR level at the SYM point on the ROC curve at each frequency (this strategy could be used with other specific points on the ROC curve to take into account cost-benefit differences in identifying normal versus impaired ears). These critical SNRs, which are plotted in Fig. 10, were larger at the higher stimulus level (76 dB peSPL) than the lower stimulus level (73 dB peSPL), as expected. They attained maximum values exceeding 6 dB at 2 and 4 kHz, and minimal values less than 3 dB at frequencies where test performance was judged to be inadequate (0.5, 12.7, 16 kHz). These results illustrate how the SYM value or other properties of a measured ROC curve can assist in determining the desired pass-refer characteristics of a test in a clinical application.

Because an AUC value equally weights all points on the ROC curve, it is not clinically useful in determining a particular test criterion for a specific clinical application. In contrast,
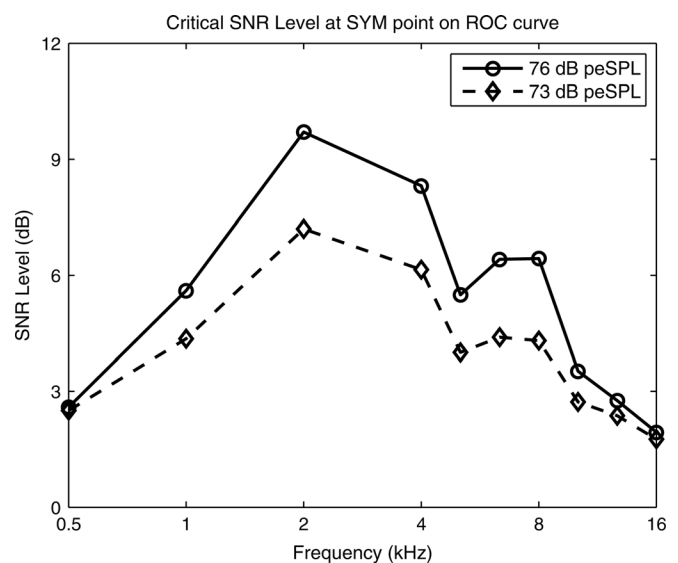


FIG. 10. The critical CEOAE SNR is plotted as a function of frequency for the SYM point on the ROC curve for click stimulus levels of 76 dB peSPL (circle symbols on solid line) and 73 dB peSPL (diamond symbol on dashed line).

J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions    259

the SYM point on the ROC curve lies in the region of high specificity and high sensitivity, so that it is likely to be reasonably close to possible criteria for a clinically useful test. A higher test sensitivity for any given test can be achieved by accepting a lower test specificity and vice versa. Such trade-offs may be important when the relative costs of incorrect detection of a normal-hearing or hearing-impaired ear are considered. The average SYM between 1 and 10 kHz for the CEOAE test ranged from 0.83 to 0.84. This value is within the range of the SYM points visually estimated from published ROC curves of CEOAE test performance in Prieve *et al.* (1993), Gorga *et al.* (1993), and Hussain *et al.* (1998). It would be expected that ROC curves with similar AUC values would also have similar SYM values. Given the general similarity of AUC across various OAE studies in Fig. 5, it would thus be expected that their SYM values would be similar as well. Exceptions to this general pattern of similarity were the higher SYM and AUC values at 0.5 and 1 kHz in Hurley and Musiek (1998), which were paralleled by lower SYM and AUC values in Hurley and Musiek at 4 kHz (see also the AUC comparison in Fig. 5). Lichtenstein and Stapells (1996) reported a ROC curve with slightly higher SYM for their half-octave analysis at 2 kHz.

## B. Efficiency of research protocol for audiometry

The research protocol to measure thresholds from 0.5 to 16 kHz was based on a sequential adaptive ML procedure using a yes–no task. The time to measure threshold was approximately 2 min per frequency. The variability of threshold was assessed in terms of the SD for the three runs in which threshold was reported. The median SD was 1.47 dB in normal-hearing ears and 1.15 dB in ears with SNHL. Along with the demonstration by Goodman *et al.* (2009) of the accuracy of the research protocol compared to a clinical protocol for frequencies up to 8 kHz, the present results quantified the efficiency of the research protocol. It is well suited for behavioral threshold measurements at frequencies up to 16 kHz. The calibration of the audiometric stimuli based on the incident pressure removed ambiguities that would have been introduced otherwise by standing waves in the ear canal (Goodman *et al.*, 2009), and thus improved the interpretation of the audiogram at high frequencies.

## V. CONCLUSIONS

CEOAE responses measured in normal ears and ears with SNHL conveyed information on cochlear status from 1 to 16 kHz that may be valuable in detecting ears with SNHL. The key property in predicting SNHL above 4 kHz was to retain the short-duration part of the CEOAE response, i.e., that part of the response within 2.5 ms of the peak of the click stimulus. Measurements of CEOAEs were highly repeatable on different test dates. The resulting CEOAE responses accurately detected the presence or absence of a frequency-specific SNHL from 1 to 10 kHz, with reduced performance at 0.5 kHz and at and above 12.7 kHz. Extending the accuracy of the upper frequency of CEOAE testing from 4 kHz up to 10 kHz without a large increase in test time adds a frequency range of assessment that is important for human auditory communication. These findings support the potential clinical use of this type of CEOAE test to screen for SNHL.

Avan, P., Elbez, M., and Bonfils, P. (**1997**). Click-evoked otoacoustic emissions and the influence of high-frequency hearing in humans. J. Acoust. Soc. Am. **101**, 2771–2777.

Dreisbach, L. E., Long, K. M., and Lees, S. E. (**2006**). Repeatability of high-frequency distortion-product otoacoustic emissions in normal-hearing adults. Ear Hear. **27**, 466–479.

Efron, B., and Tibshirani, R. (**1986**). Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. **1**, 54–77.

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

Goodman, S. S., Fitzpatrick, D. F., Ellison, J. C., Jesteadt, W., Keefe D. H. (**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. M., Beauchaine, K. L., Kaminski, J. R., Peters, J., Schulte, L., and Jesteadt, W. (**1993**). A comparison of transient-evoked and distortion product otoacoustic emissions in normal-hearing and hearing-impaired subjects. J. Acoust. Soc. Am. **94**, 2639–2648.

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.

Green, D. M. (**1993**). A maximum-likelihood method for estimating thresholds in a yes-no task. J. Acoust. Soc. Am. **93**, 2096–2105.

Gu, X., and Green, D. M. (**1994**). Further studies of a maximum-likelihood yes-no procedure. J. Acoust. Soc. Am. **96**, 93–101.

Hoaglin, D. C., Mosteller, F., and Tukey, J. W. (**1983**). *Understanding Robust and Exploratory Data Analysis* (Wiley, New York), pp. 404–414.

Hurley, R. M., and Musiek, F. E. (**1994**). Effectiveness of transient-evoked otoacoustic emissions (TEOAEs) in predicting hearing level. J. Am. Acad. Audiol. **5**, 195–203.

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.

Jedrzejczak, W. W., Lorens, A., Piotrowska, A., Kochanek, K., and Sharzynski, H. (**2009**). Otoacoustic emissions evoked by 0.5 kHz tone burst. J. Acoust. Soc. Am. **125**, 3158–3165.

Keefe, D. H. (**1998**). Double-evoked otoacoustic emissions: I. Measurement theory and nonlinear coherence. J. Acoust. Soc. Am. **103**, 3489–3498.

Keefe, D. H., Fitzpatrick, D. F., Liu, Y.-W., Sanford, C. A., and Gorga, M. P. (**2010**). Wideband acoustic reflex test in a test battery to predict middle-ear dysfunction. Hear. Res. **263**, 52–65.

Keefe, D. H., and Ling, R. (**1998**). Double-evoked otoacoustic emissions: II. Intermittent noise rejection, calibration, and ear-canal measurements. J. Acoust. Soc. Am. **103**, 3499–3508.

Keefe, D. H., Schairer, K. S., Ellison, J. C., Fitzpatrick, D. F., and Jesteadt, W. (**2009**). Use of stimulus-frequency otoacoustic emissions to investigate efferent and cochlear contributions to temporal overshoot. J. Acoust. Soc. Am. **125**, 1595–1604.

Kemp, D. T. (**1978**). Stimulated acoustic emissions from within the human auditory system. J. Acoust. Soc. Am. **64**, 1386–1391.

Kemp, D. T., Bray, P., Alexander, L., and Brown, A. M. (**1986**). Acoustic emission cochleography—Practical aspects. Scand. Audiol. Suppl. **25**, 71–95.

Kemp, D. T., Ryan, S., and Bray, P. (**1990**). A guide to the effective use of otoacoustic emissions. Ear Hear. **11**, 93–105.

Lecluyse, W., and Meddis, R. (**2009**). A simple single-interval adaptive procedure for estimating thresholds in normal and impaired listeners. J. Acoust. Soc. Am. **126**, 2570–2579.

Leek, M. R., Dubno, J. R., He, N., and Ahlstrom, J. B. (**2000**). Experience with a yes-no single-interval maximum-likelihood procedure. J. Acoust. Soc. Am. **107**, 2674–2684.

Lichtenstein, V., and Stapells, D. R. (**1996**). Frequency-specific identification of hearing loss using transient-evoked otoacoustic emissions to clicks and tones. Hear. Res. **98**, 125–136.

Liu, Y.-W., Sanford, C. A., Ellison, J. C., Fitzpatrick, D. F., and Gorga, M. P. (**2008**). Wideband absorbance tympanometry using pressure sweeps: System development and results on adults with normal hearing. J. Acoust. Soc. Am. **124**, 3708–3719.

Pepe, M. (**2003**). *The Statistical Evaluation of Medical Tests for Classification and Prediction* (Oxford University Press, Oxford, UK), pp. 76–80.

Prieve, B. A., Gorga, M. P., Schmidt, A., Neely, S., Peters, J., Schultes, L., and Jesteadt, W. (**1993**). Analysis of transient-evoked otoacoustic emissions in normal-hearing and hearing-impaired ears. J. Acoust. Soc. Am. **93**, 3308–3319.

Reimann, C., Filzmoser, P., and Garrett, R. G. (**2005**). Background and threshold: Critical comparison of methods of determination. Sci. Total Environ. **346**, 1–16.

Sanford, C. A., Keefe, D. H., Liu, Y.-W., Fitzpatrick, D. F., McCreery, R. W., Lewis, D. E., and Gorga, M. P. (**2009**). Sound-conduction effects on distortion-product otoacoustic emission screening outcomes in newborn infants: Test performance of wideband acoustic transfer functions and 1-kHz tympanometry. Ear Hear. **30**, 635–652.

Schairer, K. S., Fitzpatrick, D., and Keefe, D. H. (**2003**). Input-output functions for stimulus-frequency otoacoustic emissions in normal-hearing adult ears. J. Acoust. Soc. Am. **114**, 944–966.

Shera, C. A., Guinan, J. J., Jr. and Oxenham, A. J. (**2002**). Revised estimates of human cochlear tuning from otoacoustic and behavioral measurements. Proc. Natl. Acad. Sci. **99**, 3318–3323.

Whitehead, M. L., Jimenez, A. M., Stagner, B. B., McCoy, M. J., Lonsbury-Martin, B. L., and Martin, G. K. (**1995**). Time-windowing of click-evoked otoacoustic emissions to increase signal-to-noise ratio. Ear Hear. **16**, 599–611.

J. Acoust. Soc. Am., Vol. 129, No. 1, January 2011

Keefe *et al.*: Detecting hearing loss click otoacoustic emissions    261