

# Otoacoustic-emission-based medial-olivocochlear reflex assays for humans

Lynne Marshall<sup>a)</sup> and Judi A. Lapsley Miller

Naval Submarine Medical Research Laboratory, Box 900, Subase NLON, Box 900, Groton, Connecticut 06349-5900

John J. Guinan and Christopher A. Shera

Eaton-Peabody Laboratory of Auditory Physiology, Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Massachusetts 02114

Charlotte M. Reed, Zachary D. Perez, and Lorraine A. Delhorne

Research Laboratory of Electronics, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139-4307

Paul Boege

Consultant, Hirschanger 17, D-82319, Starnberg, Germany

(Received 28 December 2013; revised 15 September 2014; accepted 18 September 2014)

Otoacoustic emission (OAE) tests of the medial-olivocochlear reflex (MOCR) in humans were assessed for viability as clinical assays. Two reflection-source OAEs [TEOAEs: transient-evoked otoacoustic emissions evoked by a 47 dB sound pressure level (SPL) chirp; and discrete-tone SFOAEs: stimulus-frequency otoacoustic emissions evoked by 40 dB SPL tones, and assessed with a 60 dB SPL suppressor] were compared in 27 normal-hearing adults. The MOCR elicitor was a 60 dB SPL contralateral broadband noise. An estimate of MOCR strength, MOCR%, was defined as the vector difference between OAEs measured with and without the elicitor, normalized by OAE magnitude (without elicitor). An MOCR was reliably detected in most ears. Within subjects, MOCR strength was correlated across frequency bands and across OAE type. The ratio of across-subject variability to within-subject variability ranged from 2 to 15, with wideband TEOAEs and averaged SFOAEs giving the highest ratios. MOCR strength in *individual* ears was reliably classified into low, normal, and high groups. SFOAEs using 1.5 to 2 kHz tones and TEOAEs in the 0.5 to 2.5 kHz band gave the best statistical results. TEOAEs had more clinical advantages. Both assays could be made faster for clinical applications, such as screening for individual susceptibility to acoustic trauma in a hearing-conservation program. [<http://dx.doi.org/10.1121/1.4896745>]

PACS number(s): 43.64.Jb, 43.64.Wn, 43.64.Yp [CA]

Pages: 2697–2713

## I. INTRODUCTION

In mammals, the medial-olivocochlear reflex (MOCR) helps to prevent permanent noise-induced hearing loss (NIHL) and cochlear neuropathy (or cochlear synaptopathy) (e.g., [Maison and Liberman, 2000](#); [Luebke and Foster, 2002](#); [Maison et al., 2013](#); [Liberman et al., 2014](#); [Luebke et al., 2014](#)). A test for MOCR strength therefore might provide a useful predictor of human susceptibility to noise-induced damage. Before embarking on expensive field trials, an important step is to develop a non-invasive method to measure MOCR strength in humans that is fast and accurate enough to collect large amounts of data on noise-exposed populations, and has sufficient statistical properties to differentiate people with strong or weak MOCR. Such a test might be derived from changes in otoacoustic emissions (OAEs) due to the presence of an MOCR-eliciting acoustical stimulus (e.g., [Guinan et al., 2003](#)). If MOCR strength is a predictor of NIHL or cochlear-neuropathy risk in humans, hearing-

conservation programs (HCPs) could use an MOCR assay to identify those individuals most susceptible and take preemptive steps to prevent NIHL from occurring.

A weak MOCR is associated with an increased risk for permanent NIHL in guinea pigs ([Maison and Liberman, 2000](#); [Luebke and Foster, 2002](#)) and rabbits ([Luebke et al., 2014](#)), and with an increased risk for cochlear neuropathy in mice (e.g., [Maison et al., 2013](#); [Liberman et al., 2014](#)). The same is presumably true for humans, but it is more difficult to study, ultimately requiring experiments in populations already being exposed to high noise levels. The exact mechanism underlying the protective abilities of the medial olivocochlear (MOC) efferent system is unknown, but there are a number of possibilities (reviewed in [Marshall and Lapsley Miller, 2014](#)).

There are several OAE-based paradigms for measuring MOCR strength in humans, which usually involve measuring OAEs with and without the presence of an MOCR-eliciting stimulus, and comparing the difference. This has been done for the three common OAE evoking methods (as summarized by [Guinan et al., 2003](#)): transient-evoked otoacoustic emissions (TEOAEs), distortion-product otoacoustic

<sup>a)</sup>Author to whom correspondence should be addressed. Electronic mail: [Lynne.Marshall@med.navy.mil](mailto:Lynne.Marshall@med.navy.mil)

emissions (DPOAEs), and stimulus-frequency otoacoustic emissions (SFOAEs).

In choosing which OAE type for a human MOCR assay we first need to consider their underlying generation mechanisms. Evoked OAEs arise from two mechanisms in the cochlea: coherent reflection and distortion (e.g., [Shera and Guinan, 1999](#)). The coherent-reflection source emanates from random irregularities along the length of the cochlea, and is thought to be the primary source of SFOAEs and TEOAEs evoked with medium and low stimulus-levels. The distortion source arises from nonlinear interactions of traveling waves along the basilar membrane. DPOAEs in humans are a mix of reflection and distortion sources ([Shera and Guinan, 1999](#)) that may combine constructively or destructively.

This paper focuses on reflection-source OAEs for several reasons. The reflection component appears to be more sensitive to the MOCR than the distortion component (e.g., [Abdala et al., 2009](#); [Henin et al., 2011](#)). Although DPOAE methods exist that can separate out the two DPOAE components ([Talmadge et al., 1999](#); [Kalluri and Shera, 2001](#); [Long et al., 2008](#)), the reflection-component generation is dependent on the distortion component at an unknown and uncontrolled level. This potential confound, along with other complications (see [Guinan, 2006](#); [Wagner and Heyd, 2011](#); [Abdala et al., 2013](#); [Kumar et al., 2013](#)), led us to decide that DPOAEs were not our top choice for a clinical MOCR assay for humans. Although much of the animal work has been done with DPOAE adaptation paradigms, this approach is not yet suitable for human clinical use ([Meinke et al., 2005](#); [Guinan, 2006](#)).

There are pros and cons to using TEOAEs or discrete-tone SFOAEs as the underlying OAE test in an MOCR assay, and there is little research directly comparing the two methods. Discrete-tone SFOAEs offer a good signal-to-noise ratio (SNR); they are thought to arise from primarily a single place on the basilar membrane; and their production by coherent reflection is understood ([Guinan et al., 2003](#)). Using a 40 dB sound pressure level (SPL) stimulus level, little, if any, efferent activity is evoked by the stimulus ([Guinan et al., 2003](#)). However, the best SFOAE test frequencies need to be determined individually for each ear to ensure testing is not done at a null in the SFOAE spectrum and not near a large spontaneous otoacoustic emission (SOAE). [Backus and Guinan \(2007\)](#) warn that measuring the MOCR with individual SFOAE frequencies is not sufficient, and that reliable results are achievable only when averaging across multiple SFOAE frequencies, which significantly adds to the test time. TEOAEs offer a much wider frequency range in the same amount of test time, with the ability to analyze individual bands *post hoc*, so the same test parameters can be used for everyone. However, their SNR is typically not as high as for SFOAEs; there can be issues with stimulus artifact; and the TEOAE stimulus itself may be more likely to produce efferent activity.

An MOCR may be acoustically elicited contralaterally, ipsilaterally, or bilaterally. A contralateral elicitor is the simpler choice for clinical use in humans, because it does not contaminate the OAE stimulus and response ([Veuillet et al., 1991](#); [Guinan et al., 2003](#)).

For a clinical assay, it is critical to establish the statistical properties of the MOCR metric ([Backus and Guinan, 2007](#);

[Goodman et al., 2013](#)). The assay should produce accurate and consistent results with repeated testing in an individual. Because MOCR tests involve taking the difference between two OAE measurements, each OAE measurement itself must have low variability. An MOCR assay also should show a wide range of MOCR strengths in the population, relative to the test–retest variability, which is crucial if ears are to be accurately classified into various categories of MOCR strength. A key statistic we use is the ratio of within-subject measurement variability to the across-subject variability, referred to here as the *variance ratio*. As summarized by [Backus and Guinan \(2007\)](#), there is limited information about the population variation of MOCR strength because earlier studies did not also estimate within-subject variability.

MOCR measurements not only have all the issues associated with measuring OAEs (e.g., [Lapsley Miller et al., 2006](#)), but also have the problem that the middle-ear-muscle reflex (MEMR) can be inadvertently elicited during MOCR testing by the broadband MOCR elicitor in the contralateral ear or even by the OAE stimulus in the test ear (e.g., [Guinan et al., 2003](#)). A second problem is that the MOCR can be elicited by the OAE stimulus itself. [Guinan et al. \(2003\)](#) have summarized stimulus parameters for TEOAEs, DPOAEs, and discrete-tone SFOAEs that result in the OAE test stimuli themselves activating the MOCR, thereby obscuring the effect caused by an additional noise activator. Finally, the effect of SOAEs must be considered.

Humans and animals both show wide across-subject MOCR variability (as summarized in [Guinan, 1996](#)), so it is promising that an OAE-based MOCR assay could be used to screen for susceptibility to permanent NIHL in humans ([Maison and Liberman, 2000](#)). Recently, [Wolpert et al. \(2014\)](#) showed that the difference in DPOAE growth functions with and without a contralateral elicitor was moderately predictive of temporary threshold shift (TTS) magnitude. Although promising, this finding might not generalize to permanent threshold shift (PTS) risk, because PTS and TTS do not have the same physiological underpinnings (e.g., [Nordmann et al., 2000](#)), and in its current form this test takes too long for a clinical assay. Earlier human studies were not well-gearred to reveal whether MOCR strength is predictive of permanent NIHL in humans. In a prospective study, [Shupak et al. \(2007\)](#), using a TEOAE-based MOCR test, found no evidence that MOCR strength predicted PTS in humans. However, their MOCR test was possibly confounded with crossover noise from the headphones, and the MEMR could have been activated by the high (65 dB SPL) contralateral noise elicitor (e.g., [Guinan et al., 2003](#)). A DPOAE-based MOCR test did not predict temporary NIHL ([Muller and Janssen, 2008](#); [Muller et al., 2010](#)). In addition to the complicating factor of using DPOAEs, the MOCR was measured at the nulls in the DPOAE spectrum, which although yielding larger MOCR changes ([Wagner et al., 2007](#)), does not provide an accurate measure of MOCR activity ([Abdala et al., 2009](#); [Guinan, 2012](#)). [Abdala et al. \(2009\)](#) argue and [Henin et al. \(2011\)](#) show that the bi-directional changes seen in human MOCR-DPOAE are due to the changing phase relationship between the two DPOAE sources and do not accurately represent the

underlying MOCR strength. A factor that could have affected these studies is using amplitude differences rather than vector differences in OAEs, because amplitude differences produce smaller MOCR effects (Henin *et al.*, 2011).

To answer basic and applied research questions about the MOCR in humans requires testing many at-risk people in real-world settings. This study describes our first attempt toward defining a suitable MOCR assay, where we compare and contrast low-stimulus-level TEOAEs and SFOAEs as the underlying OAE measurement in a clinically-focused MOCR assay, and consider the viability of each.

## II. METHOD

### A. Subjects

Twenty-seven subjects completed the experiment (13 females, mean age 22 yrs, range 18 to 32 yrs; 14 males, mean age 26 yrs, range 18 to 38 yrs old; 13 left ears, 14 right ears).<sup>1</sup> Subjects had no ear pathologies and no regular exposure to loud noise. They avoided noise while participating in the study. No subject reported tinnitus either during testing or on a regular basis. One ear was chosen as the test ear.

All testing was conducted at the Massachusetts Institute of Technology. The experimental protocol was approved by that institute's internal review board and conducted in compliance with regulations and ethical guidelines on experimentation with human subjects. Subjects provided informed consent and were paid for their participation in the experiments.

### B. Equipment

Audiograms were measured with an audiometer with TDH-39P earphones and MX41/AR cushions (Model AD2293, Interacoustics Diagnostics, Assens, Denmark). Tympanometry was conducted with a middle-ear analyzer (model 1733, Grason-Stadler, Inc., Eden Prairie, MN).

TEOAEs were measured using HearID R3.2 (pre-release) or R3.3 (custom modified) systems (Mimosa Acoustics, Inc., Champaign, IL). SFOAEs were measured with the SFOAE-SG (v3.0.11) system (Mimosa Acoustics, Inc., Champaign, IL). Both HearID and SFOAE-SG systems used a 24-bit digital-signal-processing PC-card, running on an IBM Think Pad T43 laptop computer.<sup>2</sup> OAE stimuli were delivered and responses recorded with an ER-10C probe (Etymotic Research, Inc., Elk Grove Village, IL) using foam ear-tips.

The contralateral broadband noise (BBN) (0.01 to 10 kHz) was generated by an analog noise generator (Model 901B, Grason Stadler Corporation, Eden Prairie, MN), attenuated by a programmable attenuator (Model PA4, Tucker-Davis Technologies, Inc., Alachua, FL), amplified with an integrated stereo amplifier (model 31-1955, Realistic SA-150), and output to an ER-2 tube-phone (Etymotic Research, Inc., Elk Grove Village, IL). The timing of the noise was controlled manually by the tester.

### C. Procedures and screening

One ear of each subject was designated as the test ear. Determination of the test ear depended on normal hearing and tympanograms, clear ear canals, adequate OAE levels,

any SOAEs being sufficiently far from the SFOAE test frequencies, and sufficiently high MEMR thresholds. Selection of test ears was arranged to ensure half the test ears had SOAEs and half did not. When both ears met criteria, the ear with the best hearing, SOAE status, and best OAE amplitude was selected somewhat arbitrarily. The contralateral ear did not need to meet as stringent criteria as the test ear: It did not need to meet OAE criteria, and hearing could be slightly worse at 8 kHz.

For this initial work on MOCR test development, it was more important to attempt to get good data than it was to test everyone. Therefore, strict criteria were used for hearing thresholds, otoscopy, tympanograms, TEOAEs, SFOAEs, and ability to measure MEMR thresholds reliably at low levels.

Hearing thresholds were measured manually using the Modified Hughson-Westlake procedure with pulsed tones (ANSI, 2004). Subjects were required to have hearing thresholds  $\leq 15$  dB hearing level (HL) in their test ear at 0.5, 1, 2, 3, 4, 6, and 8 kHz; and thresholds  $\leq 15$  dB HL in their contralateral ear at 0.5, 1, 2, 3, 4, and 6 kHz; and  $\leq 20$  dB HL at 8 kHz.

Otoscopic exams were performed to check for abnormalities and clear ear canals. Excess cerumen was removed if it was near the ear canal entrance; otherwise the subject did not continue in the study unless they had the cerumen professionally removed.

Tympanometric peaks for both ears were required to be within  $0 \pm 50$  daPa when measured using a slow sweep speed (12.5 daPa/s). Subjects could Valsalva to achieve this. At the beginning of each test session, tympanometric-peak pressures were rescreened; the test ear was required to be within  $0 \pm 50$  daPa, and the contralateral ear was required to be within  $0 \pm 100$  daPa.

TEOAEs were required to be present in the test ear in the frequency bands 1 to 1.5, 1.5 to 2, and 2 to 2.5 kHz with  $\text{SNR} \geq 6$  dB. The TEOAE stimulus used for screening and for MOCR testing was a 1 to 5 kHz bandpass chirp (see "Shera chirp" in Lapsley Miller *et al.*, 2004a; Mimosa Acoustics, 2007). Each chirp had an absolute duration of 10.5 ms with an effective duration of approximately 6.5 ms. A chirp was presented every 32.5 ms.<sup>3</sup> In-the-ear spectrum calibration was used with the intention to flatten the stimulus to accommodate variations in ear-canal acoustics across individuals; however, a missing microphone equalization meant that the actual stimulus presented was not flat.<sup>4</sup> The measurement was made in non-linear mode (to avoid stimulus artifact), where a stimulus ensemble of four chirps was presented in a series. The first three chirps were 47 dB SPL (approximately 65 dB pSPL), and the fourth chirp was 9.5 dB higher in level and with opposite polarity. The responses to 500 presentations of this stimulus ensemble, all of which met noise-rejection criteria, were derived using the method developed by Bray (1989). A 14 ms response window (including 2.5 ms onset and offset ramps), which started 2 ms after the end of the digital stimulus to minimize the effects of stimulus ringing, and a 0.75 to 5 kHz bandpass response filter were used to extract the TEOAE from the averaged waveform (Mimosa Acoustics, 2007). There was

no artifactual response to this stimulus in a B&K 4157 artificial ear (Brüel & Kjær, Nærum, Denmark) or in severely hearing-impaired ears (Lapsley Miller *et al.*, 2004a). Microphone equalization was applied to the TEOAE responses *post hoc*.

SFOAEs were measured using the method described by Shera and Guinan (1999). The probe tone,  $f_p$ , was 40 dB SPL, and the ipsilateral suppressor tone,  $f_s$ , was 60 dB SPL and 47 Hz higher than  $f_p$  (Lilaonitkul and Guinan, 2009a,b). The probe tone played continuously while the suppressor cycled on and off every 170 ms (excluding ramps), taking a minimum of 8 s if no frames were rejected due to high noise, otherwise up to a maximum of 35 s. The measurement stopped as soon as 16 low-noise samples at  $f_p$  were obtained (both with and without the suppressor). These samples were then averaged and the SFOAE was derived by taking the vector difference of the average at  $f_p$  with and without the suppressor.

At screening, discrete-tone SFOAEs were measured sequentially across 71 frequencies between 0.970 and 2.530 kHz, spaced using a power-law function with exponent 1.64 (as best possible given the minimum frequency resolution of 11.7 Hz), and group delay was used to establish that a valid SFOAE had been measured (Lapsley Miller *et al.*, 2004b). For each ear, from the 71-point SFOAE-gram, 2 three-frequency stimulus ensembles were selected for MOCR testing. The *optimal-frequency* ensemble included one test frequency from each band of interest (1 to 1.5, 1.5 to 2, and 2 to 2.5 kHz) that had good SFOAE amplitude and was away from any SOAEs.<sup>5</sup> The *cluster* ensemble included one test frequency from the *optimal-frequency ensemble* (usually in the 1.5 to 2 kHz band) and the two neighboring frequencies that were 23 Hz above and below.<sup>6</sup> This cluster was used to calculate an averaged MOCR strength estimate (referred to as the *cluster-average*). If an SFOAE solution could not be found, the other ear was considered providing it met all criteria for being a test ear, otherwise the subject was not enrolled.

SOAEs were measured using the *SOAE50* test protocol in the TEOAE module (Mimosa Acoustics, 2007), which measures synchronized SOAEs. In-the-ear spectrum calibration was used to flatten the 50 dB SPL 1 to 5 kHz click stimulus.<sup>4</sup> The click was presented every 64 ms. Responses to 1000 stimuli were recorded in linear mode (where all stimuli were at the same level). The 20 ms response window (including 2.5 ms onset and offset ramps) started 20 ms after the end of the digital stimulus, which allowed time for the TEOAE response to dissipate, leaving only the SOAEs that were phase-locked to the stimulus. In the frequency domain, SOAEs were defined as present if a frequency bin (resolution 11.7 Hz) had a response greater than  $-20$  dB SPL, an SNR greater than 12 dB, and a noise level lower than  $-30$  dB SPL, between 0.5 and 3 kHz. Setting a criterion on SOAE level meant that small but clearly detectable SOAEs were not considered present as they were unlikely to be of sufficient strength to affect the MOCR result.

MEMR thresholds to the BBN were estimated using a novel procedure (Lilaonitkul and Guinan, 2009a,b; Lapsley Miller and Marshall, 2014) because measurements with

clinical audiological equipment are insensitive (e.g., Zhao and Dhar, 2010). It was important to ensure that the contralateral BBN did not elicit a MEMR; otherwise the MOCR measurements could be confounded by the MEMR. From previous experience, we expect 5% to 10% of subjects to have some reflex for a 60 dB SPL BBN, although it is unlikely to be substantive. To estimate the MEMR threshold, changes in a 40 dB SPL 1 kHz tone to a contralateral BBN elicitor were estimated in a series of measurements (Method 1 in Lapsley Miller and Marshall, 2014). The SFOAE generated by the tone was continually suppressed by a nearby suppressor tone at 60 dB SPL. Because the SFOAE itself was suppressed, any change in the response was primarily due to the activation of the middle-ear muscles. Four contralateral BBN noise levels were used: 50, 55, 60, and 65 dB SPL. The lowest BBN stimulus that evoked a change larger than 0 dB SPL was considered the MEMR threshold for that ear. All subjects included in the following analyses had thresholds  $\geq 65$  dB SPL, which was at least 5 dB higher than the 60 dB SPL BBN level used in the MOCR measurements.

We also tested whether the TEOAE chirp stimuli could trigger the MEMR at 47 and 52 dB SPL. The TEOAE stimulus was presented to the OAE test ear, and the MEMR measurement was made in the contralateral ear. There was a small MEMR in Subject 81 at 52 dB SPL and possibly a tiny MEMR at 47 dB SPL. Subject 81 met all our criteria for study inclusion so was included in the data analyses (so it is possible their results could have been affected by a MEMR from the combined effects of the TE stimulus and contralateral elicitor, but we were not able to determine this *a priori*).

Very small MEMRs can be difficult to distinguish from measurement variability in some subjects, and further refinement of this procedure is needed (Lapsley Miller and Marshall, 2014).

#### D. MOCR test procedure

Subjects were tested in four 2-h sessions, usually on separate days, but at least separated by 2 h. The first session was for screening. The second session consisted of a TEOAE-based measurement series then three SFOAE-based measurement series (one series for each of the three optimal frequencies). The third session consisted of a TEOAE-based measurement series followed by the three SFOAE-based measurement series (one series for each of the three clustered frequencies). The fourth session consisted of three SFOAE-based measurement series for the three optimal frequencies then three SFOAE-based measurement series for the three clustered frequencies. Each SFOAE test frequency was tested individually in its own measurement series, and the test time was divided up by the tester to ensure roughly equal time was allocated to each. The order of the three SFOAE frequencies within a session was decided randomly by tossing a die.

At the beginning of each session subjects were checked to ensure that they had not been recently noise exposed, had clear ear canals, and to ensure that their tympanometric peak pressure met inclusion criteria (Sec. II C). Subjects were not allowed to doze or sleep during data collection as sleep can

decrease efferent activity (Froehlich *et al.*, 1993). They were, however, allowed to read, watch DVDs, or engage in other quiet activities. A short break was given half-way through a test session, and subjects could take breaks at other times if necessary.

For each subject and each experimental condition, a series of MOCR measurements was made to establish MOCR statistics. An MOCR measurement series consisted of around 15 trial-pairs (median 15, range 10–20 trial-pairs), where each trial pair consisted of one OAE measurement made without the contralateral BBN (*Q* or “Quiet” trial) and one measurement made shortly after with the contralateral BBN (*N* or “Noise” trial). On an *N* trial, the tester turned on the noise and waited at least 2 s before starting the OAE test. At the end of the *N* trial, the tester turned off the BBN and waited at least 10 s for the MOCR to reset before starting the next trial pair. In-the-ear calibration was performed at the beginning of the measurement series, and again when deemed necessary throughout the test. If the tester noted a change in stimulus levels, or saw the probe move or fall out, the current measurement series was entirely rerun if there was time, otherwise the tester stopped testing that series. The number of trial pairs in each series for each subject/session thus depended on probe stability and time elapsed within the session.

For each MOCR trial-pair series involving SFOAEs, testing was done at individual frequencies with the entire series measured before changing frequency. This minimized the time elapsed between *Q* and *N* trials.

MOCR calculations were done offsite after the session, so were not available to the tester during testing. The tester concentrated on achieving good OAE measurements with low noise floors, good calibrations, and stable stimulus levels. Artifact rejection was achieved during data collection with a threshold technique where data frames were discarded when the instantaneous wideband noise level was above a criterion level. This level was adjusted by the tester as necessary to achieve good measurements.

## E. Definitions of MOCR strength estimates

MOCR strength is most often assessed by considering just the arithmetic difference between the two OAE amplitudes in dB SPL. This approach loses all phase information. The vector difference between SFOAE complex pressures, with and without MOC stimulation, uses both amplitude and phase to derive what we call here the “raw MOCR” (which is in Pascals and can be converted to dB SPL the same as the measured complex sound pressures). Backus and Guinan (2007) normalized the raw MOCR amplitude by the SFOAE amplitude to obtain an MOCR strength estimate in percent, which is referred to here as MOCR%. Normalization produces an MOCR measure that is uncorrelated with (i.e., unconfounded by) OAE amplitude. They did this for SFOAEs,<sup>7</sup> and we applied the same principles to TEOAEs. All the calculations were done in the frequency domain from FFTs.

Specifically, the normalized SFOAE MOCR% ( $\text{MOCR}_{\text{SF}}$ ) for a single *Q-N* trial-pair was defined as

$\text{MOCR}_{\text{SF}} = 100 (|P_N - P_Q| / |P_Q|)$ , i.e., the magnitude of the vector difference (in Pascals) between the complex-valued OAE measured on the *Q* trial ( $P_Q$ ) and the complex-valued OAE measured on the *N* trial ( $P_N$ ), normalized by the magnitude of the OAE measured on the *Q* trial, and expressed as a percentage.

The normalized TEOAE MOCR% ( $\text{MOCR}_{\text{TE}}$ ) was developed analogously to  $\text{MOCR}_{\text{SF}}$  by (a) at each FFT frequency, taking the vector difference of the *Q* and *N* TEOAE complex spectral densities (in Pascals) to obtain the raw MOCR spectrum, (b) summing the total power in the frequency bands of interest from the raw MOCR spectrum and the TEOAE *Q* spectrum (i.e., Pascals squared), (c) taking the square root, which makes these root-mean-square values, and (d) normalizing by dividing the MOCR total power by the TEOAE *Q* total power in each band, and expressing as a percentage:  $\text{MOCR}_{\text{TE}} = 100 (\sqrt{\sum |P_N - P_Q|^2} / \sqrt{\sum |P_Q|^2})$ .

For each trial-pair,  $\text{MOCR}_{\text{TE}}$  was estimated for four frequency bands (all from the same TEOAE measurements): 1 to 1.5 kHz, 1.5 to 2 kHz, 2 to 2.5 kHz, and the overall wideband response (0.5 to 6 kHz), also known as the “whole-response.”

For both  $\text{MOCR}_{\text{TE}}$  and  $\text{MOCR}_{\text{SF}}$ , the *aggregate MOCR%* was defined as the mean MOCR% across each trial-pair in an individual measurement series. An aggregate was calculated for each subject, condition, and frequency for which there were at least five individual MOCR% estimates that passed screening for OAE quality (described below).

The inherent across-trial variability in the MOCR measurement was estimated by pairing the *Q* and *N* trials in an MOCR test series into adjacent *Q-Q* and *N-N* trial-pairs and applying the same MOCR calculations as above for *Q-N* pairs. The resulting quantity is referred to here as *MOCR% variability*, which encapsulates noise from the individual OAE measurements, variability of the underlying MOCR, and across-trial variation. Because the time difference between *Q-Q* and *N-N* trials is longer than between *Q-N* trials, the MOCR% variability estimate can be considered an upper-bound as more variation is expected when the duration between trials increases.

The average  $\text{MOCR}_{\text{SF}}$  for the cluster ensemble was calculated trial-by-trial by averaging the three individual  $\text{MOCR}_{\text{SF}}$  values from each frequency (which were measured separately), in order of testing, after removing any flagged as bad quality (as described in Sec. II F). The number of trials for this *cluster-average* series was defined by the test frequency in the cluster with the fewest trials. Some trials were therefore not used. Most  $\text{MOCR}_{\text{SF}}$  cluster-average series had 15 trials, and all had at least 13 trials.

## F. Screening criteria for MOCR quality

MOCR results were screened for data quality by applying criteria to the SNR and noise level of the underlying OAE measurements. These criteria were determined in an exploratory manner. One criterion was varied at a time while visually inspecting MOCR variability and then trading off data quality with data quantity until a comparable set of

criteria were found that decreased variability without removing data for too many subjects.

For  $\text{MOCR}_{\text{TE}}$ , the 0.5 kHz band quality control criteria for each measurement were: OAE  $\text{SNR} \geq 9$  dB, and OAE  $\text{NF} < -9$  dB SPL; and the wideband criteria were: OAE  $\text{SNR} \geq 3$  dB, and OAE  $\text{NF} < 0$  dB SPL (wideband criteria do not need to be as strict because reliability increases with bandwidth, Marshall and Heller, 1996). At least 400 out of the maximum 500 averages in a trial were required, and the stimulus level had to be within 3 dB of target. For  $\text{MOCR}_{\text{SF}}$ , the criteria were the same for all frequencies: OAE  $\text{SNR} \geq 9$  dB (see Lapsley Miller *et al.*, 2004b) and OAE  $\text{NF} < -9$  dB SPL. The SFOAE and suppressor stimulus levels were required to be within 3 dB of target, and the difference between the two stimulus levels no more than 3 dB from the target difference.

Both OAE measurements in a  $Q$ - $N$  trial-pair had to pass quality control. For each individual MOCR series, at least five trial-pairs with OAEs that passed quality control screening were required for the data to be used, ensuring there was some information gained about the MOCR distribution (variance) while not eliminating too many datasets. All SFOAE measurement series had at least five trial-pairs per series—a direct result of choosing test frequencies corresponding to high amplitude SFOAEs. For TEOAEs, only the wideband condition produced sufficient data for all 27 subjects. The “Quality OAE” column in Table I shows the percentage of subjects with at least five trial-pairs where the OAEs were of sufficient quality to calculate MOCR.

### III. RESULTS

To establish the statistical utility of  $\text{MOCR}\%$ , we consider: contributions to within-subject variability; whether or not each participant’s  $\text{MOCR}\%$  estimate is separable from the inherent noise in their test; and how many trial-pairs are needed to ensure a stable  $\text{MOCR}\%$ . We then compare the SFOAE and TEOAE-based  $\text{MOCR}\%$  for the various

TABLE I. The percentage of subjects (out of 27) for each OAE type and frequency band where (a) the MOCR series had at least five trial-pairs where OAEs were of sufficient *quality* to calculate the  $\text{MOCR}\%$  (from Fig. 1 for TEOAEs and Fig. 2 for SFOAEs; and also Sec. II F), (b)  $\text{MOCR}\%$  was *separable* from the inherent test variability (Sec. III D), and (c)  $\text{MOCR}\%$  was *usable*, which included all ears with *separable*  $\text{MOCR}\%$  and also those ears with low-level  $\text{MOCR}\%$  undetectable from inherent variability and which fell in the bottom 10th percentile of  $\text{MOCR}\%$  (Sec. III E).

OAE type	Frequency band (kHz)	OAE quality screen passed	MOCR separable	MOCR usable
SFOAE	1 to 1.5	100%	70%	81%
	1.5 to 2	100%	89%	93%
	2 to 2.5	100%	78%	89%
	Cluster-average	100%	85%	85%
TEOAE	1 to 1.5	89%	78%	89%
	1.5 to 2	93%	81%	93%
	2 to 2.5	85%	74%	81%
	0.5 to 2.5	96%	81%	89%
	1 to 3	100%	89%	96%
	Wideband	100%	74%	81%

frequency bands. Comparisons include test–retest reliability, correlations across frequency bands and across OAE types, and variance ratios. Finally, we consider the ability of each condition to reliably classify ears into levels of MOCR strength, which is ultimately how the test could be used clinically.

For each subject, each OAE type and stimulus ensemble was measured twice, in two separate sessions (Sec. II D). The data from the first measurement session of each OAE type was used for most analyses presented here in order to simplify the presentation of the results. Analyses using more than one measurement session are specified. The data from the repeated measurement was mainly used to assess across-session reliability and correlations.

#### A. MOCR strength metrics ( $\text{MOCR}\%$ )

Our MOC-strength metrics all involve an MOC-induced change in an OAE normalized by the original value of the OAE, expressed as an  $\text{MOCR}\%$ . Because  $\text{MOCR}\%$  is defined from the magnitude of a vector difference, it must always be positive, and has a true minimum at 0%. However, due to inherent noise in the OAE measurements, the lower bound is never met, and there is an effective noise floor of approximately 10% for  $\text{MOCR}_{\text{SF}}$  and 21% for  $\text{MOCR}_{\text{TE}}$  (estimated from the median  $\text{MOCR}\%$  variability).  $\text{MOCR}_{\text{SF}}$  estimates vary between 14% and 45% (defined by the 10th and 90th percentiles) with a median of 26%, and  $\text{MOCR}_{\text{TE}}$  estimates vary between 27% and 59% with a median of 41%. Overall,  $\text{MOCR}_{\text{TE}}$  is systematically about 15 percentage points (pp) higher than  $\text{MOCR}_{\text{SF}}$  but with a similar range of  $\sim 32$  pp.

The  $\text{MOCR}\%$  distributions, represented as box-plots,<sup>8</sup> for each subject, OAE type, and frequency band was plotted against the  $\text{MOCR}\%$  variability distributions (see Fig. 1 for TEOAEs and Fig. 2 for SFOAEs). Ears were ranked by median  $\text{MOCR}\%$ .

#### B. $\text{MOCR}\%$ correlated with OAE quantities

It is important to have a measure of MOCR strength uncorrelated with OAE amplitude. Normalizing the raw MOCR-induced OAE change with OAE amplitude removes the correlation (Backus and Guinan, 2007). For each OAE type, the average  $\text{MOCR}\%$  was tested for its correlation with the average  $Q$ -trial OAE amplitude (dB SPL), SNR, and noise level (dB SPL).  $\text{MOCR}\%$  and the OAE characteristics for each subject were paired within the same frequency bands, and the data pairings were then pooled across all the bands to calculate an overall correlation. There were no significant correlations between  $\text{MOCR}\%$  and TEOAEs (amplitude:  $r = -0.09$ , ns; SNR:  $r = 0.03$ , ns; noise level:  $r = -0.20$ , ns) or SFOAEs (amplitude:  $r = -0.05$ , ns; SNR:  $r = -0.10$ , ns; noise level:  $r = -0.07$ , ns). By way of comparison, the correlation between the raw MOCR amplitude (vector difference converted into dB SPL but unscaled by OAE magnitude) and OAE amplitude revealed significant overall correlations for both TEOAEs (amplitude:  $r = 0.88$ ,  $p < 0.05$ ) and SFOAEs (amplitude:  $r = 0.76$ ,  $p < 0.05$ ). These results support the finding by Backus and Guinan (2007) that

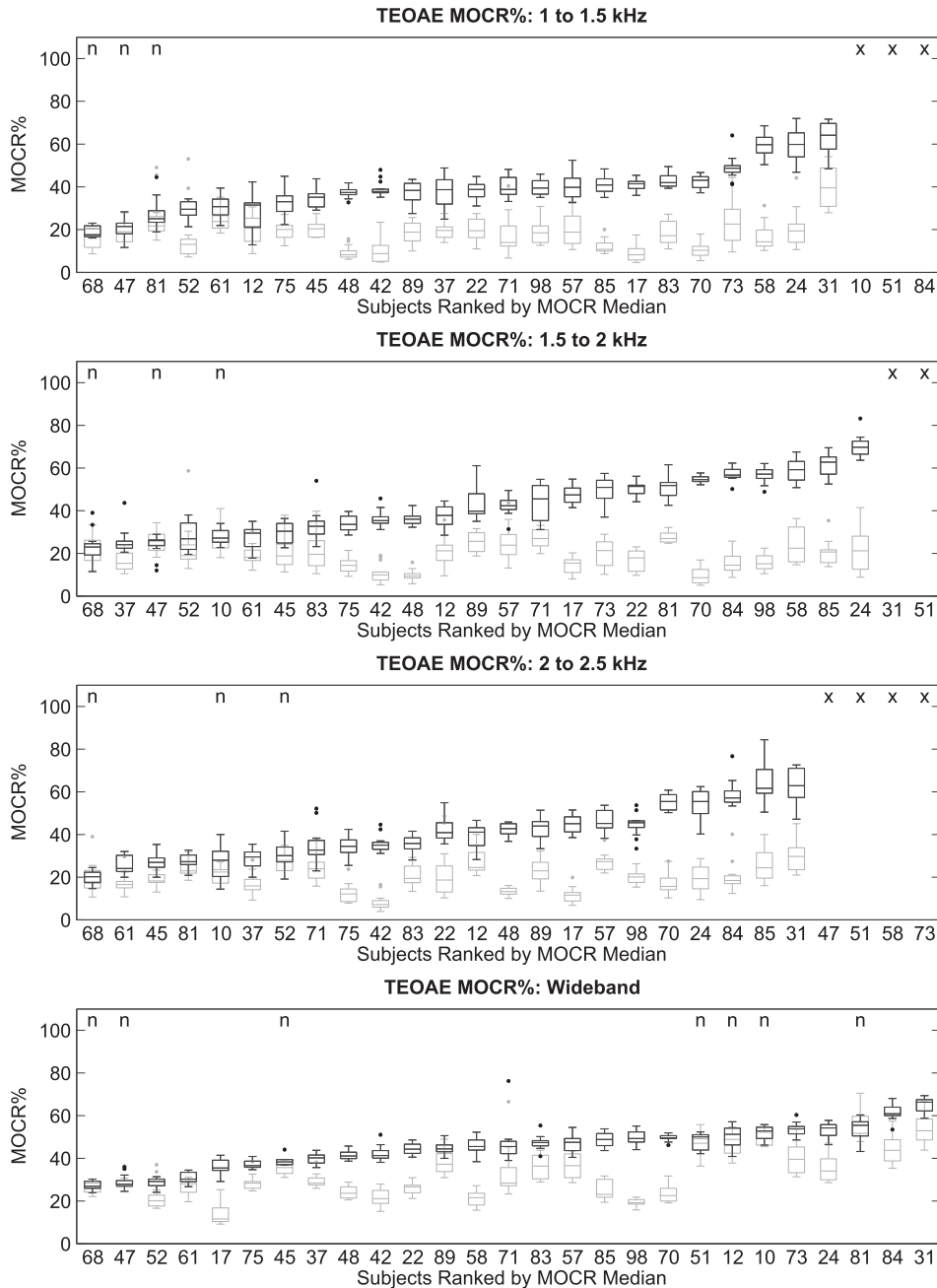


FIG. 1. MOCR<sub>TE</sub> distributions for the 27 subjects, represented as box-plots, and ordered by increasing median. From top to bottom: 1 to 1.5 kHz, 1.5 to 2 kHz, 2 to 2.5 kHz, and the wide-band response. Subjects with fewer than five trial-pairs meeting OAE quality criteria, are marked with an “x.” The remaining subjects provided at least 5 trial pairs of usable data, with most providing 15 trials pairs. The black plots represent the MOCR% distribution (based on the *Q-N* trial-pairs) and the gray plots are an estimate of the inherent variability in the MOCR measurement (based on the re-paired *Q-Q* and *N-N* trial-pairs). The “n” indicates those cases where the MOCR% estimate was not significantly different from the variability.

normalizing by OAE magnitude provides an MOCR metric that is not directly confounded by the OAE amplitude.

### C. MOCR% variability correlated with OAE quantities

Even after OAE quality-control screening (Sec. II F), the degree of within-subject MOCR variability differed across subjects (as seen in Figs. 1 and 2), with some subjects showing high consistency within a frequency band (e.g., subjects 42 and 70) and others showing a great deal of variability (e.g., subject 31). The high-variability cases were examined to establish what factors might be contributing to the variability. The biggest factors were OAE amplitude and SNR—the stronger the OAE and the lower the noise level, the more stable the MOCR%, even when the MOCR% itself was small. As an example, Fig. 3 shows the across-trial MOCR% standard deviation (SD) as a function of across-

trial average OAE SNR for the 1.5 to 2 kHz band for TEOAEs and SFOAEs (other frequency bands were similar). Lower SNRs were associated with higher variability, with a correlation of  $-0.78$  for SFOAEs and  $-0.63$  for TEOAEs (both  $p < 0.05$ ). This result is not surprising as the MOCR calculation subtracts two OAE measurements thereby doubling the effect of the OAE variance. Stimulus instability, OAE instability, temporal effects, number of trial-pairs, electrical interference, and the presence of SOAEs were not obviously related to increased variability.

To tease out if OAE amplitude, SNR, or noise level contributed most to MOCR% variability, the MOCR% SD was tested for its correlation with OAE amplitude, SNR, and noise level. MOCR% SD and the OAE characteristics for each subject were paired within the same frequency bands, and the data pairings were then pooled across all the bands to calculate an overall correlation. SNR was the more

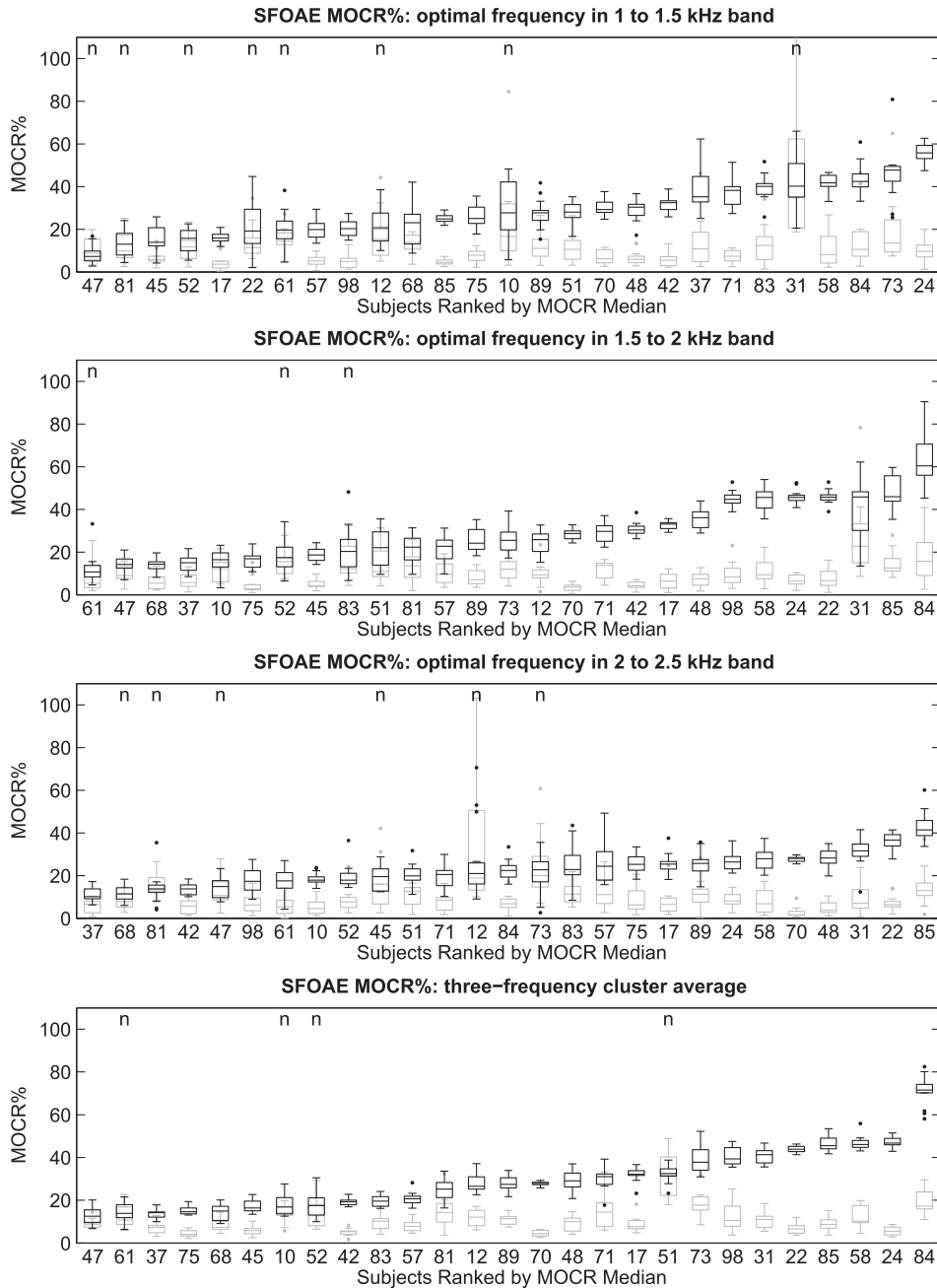


FIG. 2. MOCR<sub>SF</sub> distributions for the 27 subjects, represented as box-plots, and ordered by increasing median. From top to bottom: 1 to 1.5 kHz, 1.5 to 2 kHz, 2 to 2.5 kHz, and the three-frequency-cluster average. All subjects provided at least 10 trial pairs of usable data, with most providing 15 trials pairs. The black plots represent the MOCR% distribution (based on the *Q-Q* trial-pairs) and the gray plots are an estimate of the inherent variability in the MOCR measurement (based on the re-paired *Q-Q* and *N-N* trial-pairs). The “n” indicates those cases where the MOCR% estimate was not significantly different from the variability.

important factor for MOCR<sub>TE</sub> (amplitude:  $r = -0.50$ ,  $p < 0.05$ ; SNR:  $r = -0.62$ ,  $p < 0.05$ ; noise level:  $r = -0.05$ , ns) and MOCR<sub>SF</sub> (amplitude:  $r = -0.47$ ,  $p < 0.05$ ; SNR:  $r = -0.65$ ,  $p < 0.05$ ; noise level:  $r = 0.25$ , ns). The correlation of MOCR% with SNR was present even after using SNR as an initial quality control. SNR is derived from the OAE amplitude and the noise level; the OAE amplitude component is also significantly correlated with MOCR variability, but the OAE noise level is not. In part, this may be because the OAEs were screened to ensure sufficiently low noise levels.

#### D. Ensuring an MOCR was present

In most cases the distributions of MOCR% and MOCR% variability for each subject and condition (as shown in Figs. 1 and 2) showed separation, indicating that

the MOCR% was separable (i.e., detectable) from the inherent trial-by-trial variability in the measurement. Signal-detectability statistics such as the Area under the Receiver-Operating-Characteristic (ROC) curve are an appropriate method to consider how separated two distributions are from one another (Bamber, 1975). ROC curves were derived by treating the MOCR% variability as the “noise-alone”-event distribution and the MOCR% as the “signal-plus-noise”-event distribution. The Area under the ROC curve was calculated, and Bamber’s test for significance applied (Bamber, 1975). The MOCR% distributions marked with an “n” in Figs. 1 and 2 were not significantly different from the inherent variability; most were associated with high within-subject variability and/or low MOCR%. The percentage of subjects with a separable MOCR% is shown in Table I, and ranges from 70% to 89% for SFOAEs and from 74% to 89% for TEOAEs.



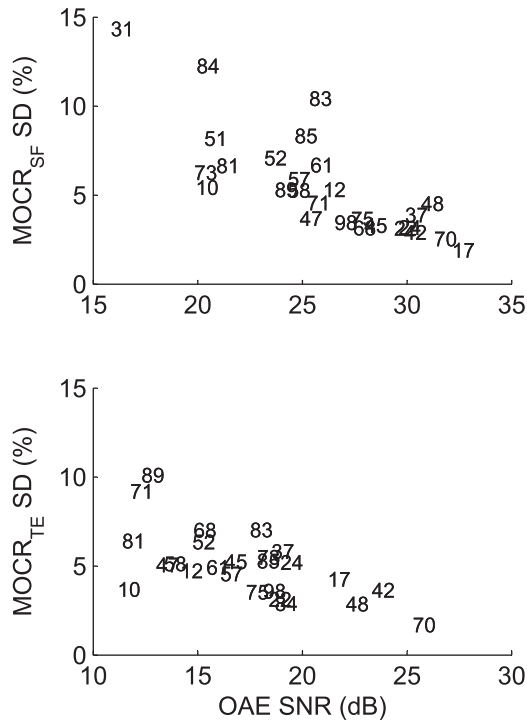


FIG. 3. Average MOCR within-subject variability (SD) as a function of average OAE SNR (dB) from the  $Q$ -trials for the 1.5 to 2 kHz band. The subject numbers represent the points, and can be related back to the second panels in Figs. 1 and 2.

### E. Including low-level MOCR% values that were not separable from noise

An important category of ears are those with low MOCR strength that may not produce MOCR% estimates that are statistically separable from the inherent noise in the test. For these ears, if the aggregate MOCR% was below the tenth percentile of separable MOCR values and if at least five trial-pairs passed OAE quality control, the aggregate MOCR% was considered to be *usable* and included in further analyses.

Table I shows the percentage of subjects with *usable* MOCR measurement series. For all but one condition, the percentage of acceptable data increased by including the “usable” data category. The best conditions were the 1.5 to 2 kHz band for both SFOAEs and TEOAEs where 93% of the subjects produced usable MOCR data (but curiously it was not the same two subjects who had unusable data), and the 1 to 3 kHz TEOAE band where 96% of the subjects produced usable data.

### F. Ratio of MOCR% across-subject variance to within-subject variance

A useful statistic is the ratio of the variance of the across-subject MOCR% mean to the mean of the within-subject across-trial MOCR% variance. This *variance ratio* can be used to compare the performance of the various MOCR strengths from the different OAE types and frequency bands. Large ratios indicate a good spread of values across ears relative to the test–retest variability for individual ears. Variance ratios were calculated for each

OAE type and frequency band for the usable MOCR data (defined in Sec. III E), and separately for the initial and repeat sessions. Table II shows the variance ratios for each frequency band under consideration. The MOCR<sub>TE</sub> wideband and the MOCR<sub>SF</sub> cluster-average produced the largest ratios. The variance ratios were similar between sessions, with differences attributable to the slightly different ears with usable data contributing to the ratios across sessions.

### G. Number of trial-pairs needed to establish within-session MOCR% stability

The MOCR% is stable across multiple measurements for some subjects but not others (Sec. III A and Figs. 1 and 2). This is primarily due to low OAE amplitude or SNR (Sec. III B and Fig. 3). Because time is limited in clinical testing, it is useful to know how many trial-pairs are needed to establish that the MOCR% estimate is stable.

A rather arbitrary threshold of an MOCR% SD of no more than 7.5 pp across the measurement series was used to define a stable MOCR% estimate. Table III shows the percentage of subjects per condition who achieved this criterion (relative to the number of subjects who met quality control criteria for OAEs; Table I). Then, for those stable cases, the number of trial-pairs that were actually needed to establish stability was estimated using a sampling technique.<sup>9</sup> The first trial-pair-number to achieve a mean MOCR SD within 1 pp of the final SD was defined as the minimum sufficient number of trial-pairs for that condition. For all but one condition, the median sufficient number of trial pairs for the group was three (Table III). It was not possible to estimate stability for the SFOAE cluster-average because the three measurements underlying it were collected in different measurement series. For this analysis, all MOCR% estimates that passed OAE quality control were included. Separability was not considered because it is calculated from the overall measurement series.

TABLE II. A key aim is to establish which MOCR assays have the biggest range in the population relative to the individual test–retest variability, defined as the ratio of the variance of the across-subject MOCR% mean to the mean of the within-subject, across-trial MOCR% variance, for each OAE type and frequency band under consideration. The larger the ratio, the better the test performance. Variance ratios were calculated separately for the two sessions to gauge repeatability.

OAE type	Frequency band (kHz)	Initial session		Repeat session	
		Subjects	Ratio	Subjects	Ratio
SFOAE	1 to 1.5	22	3.9	25	2.8
	1.5 to 2	25	5.0	25	4.9
	2 to 2.5	24	2.1	26	2.1
	Cluster-average	23	15.4	25	10.8
TEOAE	1 to 1.5	24	4.4	24	6.7
	1.5 to 2	25	6.1	23	7.0
	2 to 2.5	22	4.8	22	6.9
	0.5 to 2.5	24	9.4	26	11.0
	1 to 3	26	6.2	26	9.7
	Wideband	22	8.7	25	12.1

TABLE III. The minimum, median, and maximum number of trial-pairs needed to establish a stable MOCR% series, and the percentage of subjects who achieved stable results, excluding subjects with fewer than five trial pairs of data meeting the OAE quality control criteria. Note that subjects achieving stable results may not have had an MOCR% statistically separable from the inherent variability.

OAE type	Frequency band (kHz)	Subjects	Trial-pairs to stability			Percent stable
			Minimum	Median	Maximum	
SFOAE	1 to 1.5	27	2	4	6	74%
	1.5 to 2	27	2	3	8	81%
	2 to 2.5	27	2	3	8	85%
	Cluster-average	27	n/a	n/a	n/a	n/a
TEOAE	1 to 1.5	24	2	3	7	92%
	1.5 to 2	25	2	3	9	92%
	2 to 2.5	23	2	3	6	83%
	0.5 to 2.5	26	2	3	4	100%
	1 to 3	27	2	3	6	96%
	Wideband	27	2	3	14	100%

### H. Within-session MOCR% correlations among frequency bands and between OAE types

Table IV presents the within-session correlations for the aggregate MOCR% (averaged over the ~15 MOCR% estimates from each measurement series). Shown are the across-frequency-band correlations within an OAE type and the same-frequency-band correlations across OAE types (Pearson product-moment correlations with pair-wise data deletion; all are significant at  $p < 0.05$  except for one). Most of the correlations are moderately strong. TEOAE-based MOCR measurements produced higher across-band correlations than SFOAE-based measurements. This is probably because, for SFOAEs, the probe may have been refitted between frequencies. In contrast, for TEOAEs, MOCR% for each frequency band were derived from the same measurements, and the probe was not refitted during a measurement series.

Across OAE types, the highest correlations were for the 1.5 to 2 kHz TEOAE band against the 1.5 to 2 kHz SFOAE band ( $r = 0.81$ ,  $p < 0.05$ ; plotted in Fig. 4) and for the cluster-average SFOAE ( $r = 0.86$ ,  $p < 0.05$ ), where the cluster ensemble frequencies were also mostly in the 1.5 to

2 kHz band. Within OAE types, the highest correlations were those for adjacent or the same frequency bands.

### I. Across-session MOCR% correlations/test-retest reliability

Table V shows the across-session correlations (Pearson's product-moment correlations with pair-wise data deletion) for the average MOCR% (averaged over the ~15 MOCR% estimates from each measurement series) for each OAE type and frequency band. In all cases the correlation was high and significant, indicating good reliability. The time elapsing between sessions extended from at least 2 h to a few weeks, but for most subjects it was 1 day for TEOAEs and 2 days for SFOAEs. Also provided in Table V is the standard error of measurement, which is an estimate of test-retest reliability for individual measurements, based on the group (Ghiselli, 1964). It is derived from the correlation between test and retest and weighted by the average variance.  $MOCR_{SF}$  had slightly higher reliability than  $MOCR_{TE}$ , and the 2 to 2.5 kHz band for both OAE types produced slightly higher reliability than the other bands.

TABLE IV. Within-session aggregate MOCR% correlations were all significant except for the one indicated with an asterisk (Pearson product-moment correlations;  $p < 0.05$ ;  $N = 18$  to 24; pair-wise deletion). Comparisons across-OAE type for the same frequency ranges are in the bottom left quadrant, and comparisons within-OAE type across frequency ranges are in the top left and bottom right quadrants. The data were screened to include only usable MOCR measurements (unscreened data produced similar correlations).

OAE type	Frequency band (kHz)	SFOAE				TEOAE			
		1 to 1.5	1.5 to 2	2 to 2.5	Cluster-average	1 to 1.5	1.5 to 2	2 to 2.5	Wideband
SFOAE	1 to 1.5	1.00							
	1.5 to 2	0.47	1.00						
	2 to 2.5	0.21 <sup>a</sup>	0.59	1.00					
	Cluster-average	0.50	0.98	0.50	1.00				
TEOAE	1 to 1.5	0.78				1.00			
	1.5 to 2		0.81		0.86	0.73	1.00		
	2 to 2.5			0.75		0.75	0.83	1.00	
	0.5 to 2.5	0.74	0.84	0.53	0.80				
	1 to 3	0.70	0.81	0.58	0.79				
	Wideband	0.70	0.72	0.50	0.66	0.86	0.79	0.83	1.00

<sup>a</sup>Not significant

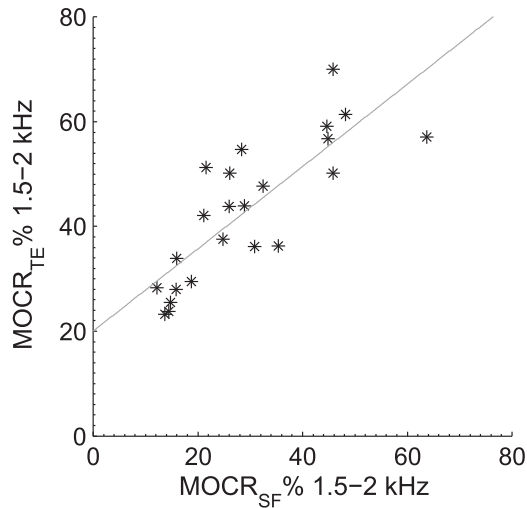


FIG. 4. Scatterplot of aggregate 1.5 to 2 kHz  $MOCR_{TE}$  against aggregate 1.5 to 2 kHz  $MOCR_{SF}$ , indicating the different measurement techniques are tapping into the same underlying phenomenon. The correlation coefficient is 0.86 (20 ears), and a least-squares fit is  $MOCR_{TE} = 0.8 MOCR_{SF} + 20$ .

### J. Effect of SOAEs on MOCR%

Mean MOCR% strength for the group of ears with no SOAEs ( $N = 12$  for TEOAEs, and  $N = 10$  for SFOAEs) and the group with SOAEs ( $N = 12$  for both; 9 of which had direct SOAE influence in the 1 to 1.5 kHz band) did not differ significantly for 1.0 to 1.5 kHz  $MOCR_{TE}$  ( $t$ -test:  $t_{22} = 0.75$ , ns, means 37.0% vs 40.4%) or the 1.0 to 1.5 kHz  $MOCR_{SF}$  ( $t$ -test:  $t_{20} = 1.80$ , ns, means 32.8% vs 23.8%).

### K. Optimal frequency region for $MOCR_{TE}$

*A priori*, the frequency bands of interest were three 0.5 kHz bands between 1 and 2.5 kHz. Individual SFOAE test frequencies were chosen to fall into each band for comparison with TEOAEs. These ranges were chosen because we wanted to consider the trade-off between the higher MOCR strength at lower frequencies and the higher noise floor typically found at lower frequencies. However, our *a priori* bands might not be the best.

TABLE V. Across-session aggregate MOCR% correlations for each OAE type and frequency band were all significant (Pearson product-moment correlations;  $p < 0.05$ ,  $N = 21$  to 23 for each cell; pair-wise deletion). The unit for the standard error of measurement is pp.

OAE type	Frequency band (kHz)	Across-session correlation	Standard error of measurement (pp)
SFOAE	1 to 1.5	0.94	2.7
	1.5 to 2	0.96	3.0
	2 to 2.5	0.95	1.8
	Cluster-average	0.96	2.9
TEOAE	1 to 1.5	0.80	5.1
	1.5 to 2	0.94	3.5
	2 to 2.5	0.96	2.4
	0.5 to 2.5	0.86	4.8
	1 to 3	0.89	4.0
	Wideband	0.89	3.8

Unlike SFOAEs, the frequency bands for TEOAEs can be *post hoc* analyzed into arbitrary bands. To investigate the optimal frequency range for  $MOCR_{TE}$  we calculated the average MOCR% and the MOCR variance ratio (introduced in Sec. III F above) in 0.5 kHz bands for center frequencies 0.75 to 5.75 kHz in 0.25 kHz steps, 1 kHz bands from 1 to 5 kHz in 0.5 kHz steps, 2 kHz bands from 1.5 to 5 kHz in 0.5 kHz steps, 3 kHz bands from 2 to 5 kHz in 0.5 kHz steps, and a 6 kHz band (wideband) at 3.5 kHz.

For each TEOAE frequency band, data were analyzed if at least 20 out of 27 subjects produced at least 5 trial pairs of screened data. A criterion of 30% was used to define low-level undetectable MOCR% for all frequency bands (as in Sec. III E), which was the approximate average level found in the earlier analyses. The underlying OAE data were screened similarly to before (Sec. II F) except that the noise level criterion was increased with increasing bandwidth using the 3-dB rule for each doubling, and the SNR criterion was decreased evenly with increasing bandwidth from 9 dB for 0.5 kHz bands, 7.5 dB for 1 kHz bands, 6 dB for 2 kHz bands, 5 dB for 3 kHz bands, and 3 dB for the 6 kHz band (because reliability increases with bandwidth, Marshall and Heller, 1996). For each subject, data from both sessions were analyzed, and the session that produced the most bands with usable data for that subject was chosen. This was done to increase the size of the data set. Some of the variation in the resulting analysis is due to different subjects contributing data (ranging from 21 to 27 ears).

Figure 5(a) shows the average group MOCR% for each bandwidth-frequency combination. For each bar, the height represents MOCR strength; the width represents the bandwidth; and the position on the abscissa represents the center frequency. The largest MOCR% (43.7%) is for the widest bandwidth. There is a clustering of MOCR strength around 36% to 40%, with most of these bandwidths including frequencies below 2 kHz. The 0.5 kHz frequency bands used extensively in our analyses, namely 1 to 1.5, 1.5 to 2, and 2 to 2.5 kHz, also have fairly high strength.

Figure 5(b) shows the MOCR variance ratio (across-subject variance to within-subject variance). The highest ratio (12.8) is for the widest bandwidth. The 0.5 to 2.5 kHz band ( $N = 26$ ) and the 1 to 3 kHz band ( $N = 27$ ) also had high variance ratios. Although not as high as the wideband result ( $N = 24$ ), they included more ears with usable data, so may provide a more reasonable trade-off between maximizing the number of ears and maximizing the variance ratio. This is why results for these bands were included in many of the preceding analyses (including Tables I–VI, using TEOAE screening criteria of  $\geq 6$  dB SNR and  $< -3$  dB SPL noise level). In comparison, the narrow bandwidths do not fare well. The higher variance ratios are for the wider bandwidths that include the 1.5 to 2 kHz band, unlike the MOCR strength above where higher strength was seen when lower frequencies were included.

### L. Test speed

The test time for a typical trial-pair was estimated by calculating how many trial-pairs were collected per 2-h

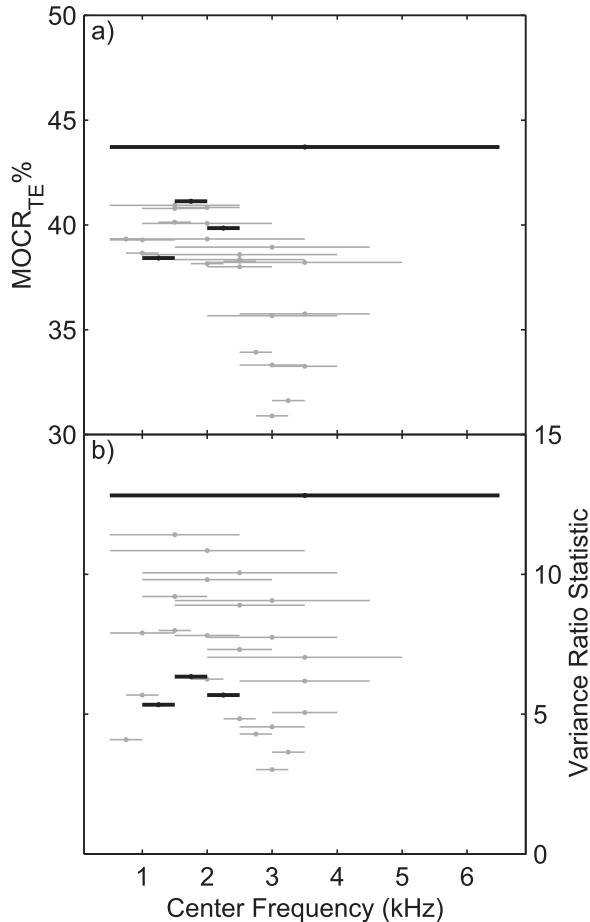


FIG. 5. (a)  $MOCR_{TE}$  strength and (b) MOCR variance ratio statistic for various analyzing bandwidths and center frequencies. The width of each line indicates the bandwidth and its position on the abscissa indicates the center frequency. The height of the line indicates the (a) MOCR% or (b) variance ratio. The four bands used in the other analyses are indicated in black. Conditions with insufficient data are not plotted. The MOCR variance ratio (an indicator of test performance) is not directly related to those conditions producing higher MOCR strength. Best performance is for wider bands and for bands that include the lower frequencies.

session, where as many trial-pairs were collected as possible in the allotted time (subtracting 10 min for test-day screening and short breaks). This gives an estimate that includes overhead associated with re-calibrating and re-fitting probes. For

TABLE VI. Repeatability of MOCR strength classifications. For each OAE type and bandwidth, ears (out of a possible 27) with usable MOCR% in both the first and second session were classified by the 10th and 90th percentiles into Low, Normal, and High MOCR strength groups. Shown are the number of ears receiving the same classification, a “Near Miss” (classifications were different, but MOCR strength differed by less than 5 pp across sessions), or a different classification. “Repeatable classifications” represents the percentage of ears in the second session that were correctly identified or were a near miss, weighted by the number of ears with usable data in both sessions.

OAE type	Frequency band (kHz)	Ears	Same	Near miss	Different	Repeatable classifications
SFOAE	1 to 1.5	22	21	0	1	95%
	1.5 to 2	21	19	1	1	95%
	2 to 2.5	23	19	3	1	96%
	Cluster-average	23	20	3	0	100%
TEOAE	1 to 1.5	22	17	2	3	86%
	1.5 to 2	23	18	3	2	91%
	2 to 2.5	21	19	0	2	90%
	0.5 to 2.5	24	21	2	1	96%
	1 to 3	25	21	1	3	88%
	Wideband	22	19	1	2	91%

27 subjects, 863 TEOAE trial-pairs were collected and 5197 SFOAE trial-pairs. This equates to 3 min per trial-pair for TEOAEs and 1 min per trial-pair for SFOAEs (one frequency).

### M. Classification into risk groups

A key question is whether MOCR strength classifications into low, normal, or high *risk* are repeatable. With the current data set, all we can do is assess whether it is possible to accurately repeat an arbitrary classification. Ears with usable MOCR in *both* the first and second session (21 to 23 ears out of a possible 27, with at least 5 trial-pairs contributing to the aggregate) were classified by the 10th and 90th percentiles into Low, Normal, and High MOCR strength, for the initial and repeated sessions, separately. Table VI shows the repeatability of the classification. For those ears gaining a different classification, indicated is whether the second session result was within 5 pp of their first session result (i.e., a near miss). For all of the SFOAE conditions, classifications were repeatable (correct or near-miss) for at least 95% of the ears with usable data. Repeatable classifications for TEOAEs ranged from 82% to 91%. Given the small number of subjects involved, this analysis can only be seen as indicative.

## IV. DISCUSSION

Our aim was to consider the viability of a clinical MOCR assay using OAEs: how to define MOCR strength, establishing the statistical properties and relationships of MOCR strength metrics, and establishing how best to measure the MOCR with OAEs, including working around confounds (MEMR and SOAEs). We conclude there is enough evidence to say that an MOCR assay is viable, but there are issues still to address. Finally we discuss how an MOCR assay might be incorporated into a HCP, what is needed for clinical trials, and other clinical uses of an MOCR assay.

### A. Which is the best MOCR assay statistically?

$MOCR_{TE}$  and  $MOCR_{SF}$  distributions differed; however, correlations were high (Table IV) in the same frequency

regions in the same ears, indicating that both measures were tapping into the same underlying phenomenon. Backus and Guinan (2007) reported normally-distributed MOCR% for test frequencies near 1 kHz (mean 36.6%, SD 11.7%, 25 ears). Our closest comparison is for MOCR<sub>SF</sub> in the 1 to 1.5 kHz band. For the 22 ears with usable MOCR data, the MOCR% was consistent with a normal distribution (mean of 29%, SD 12.2%; Shapiro-Wilk test,  $W = 0.98$ ,  $p = 0.82$ ).

Cluster-average SFOAEs and wideband TEOAEs produced the highest MOCR variance ratios (Table II), which was a key performance metric. For these two conditions, SFOAEs produced usable data in slightly more ears, had higher reliability, and produced slightly more repeatable classifications, but TEOAEs were more stable. (Other conditions did better on these secondary statistical benchmarks, but did not produce high variance ratios.) So cluster-average SFOAEs have an advantage, statistically, over TEOAEs. This can be attributed in part to individually selecting optimal test frequencies for SFOAEs. TEOAEs may have performed better had the stimulus been equalized.

For TEOAEs, it is also worth considering the two 2 kHz wide bands: 0.5 to 2.5 kHz and 1 to 3 kHz. Although these 2 kHz bands had slightly lower variance ratios than the wideband condition, they were measurable in most ears, and compared favorably in measurability with SFOAEs. In Table I, the percentage of usable data was 89% for the 0.5 to 2.5 kHz band and 96% for the 1 to 3 kHz band, which was higher than the other conditions. For both 2 kHz TEOAE bands, within-session correlations with the SFOAE conditions were higher than for wideband TEOAEs. Across-session test-retest reliability for the 2 kHz bands was comparable to wideband TEOAEs, but was lower than for SFOAEs. The 2 kHz bands were the most stable of all conditions: median trials to stability remained at 3, and the maximum reduced from 14 trials for wideband TEOAEs to 4 to 6 trials, for the 2 kHz bands. Classifications with 2 kHz TEOAE bands were repeatable for 96% of ears in the 0.5 to 2.5 kHz band, which compares favorably with SFOAEs. It also may be worthwhile in future efforts to individually choose TEOAE bandwidths to exclude large SOAEs. MOCR<sub>TE</sub> reliability may improve if TEOAEs are analyzed into bands individually determined for each ear, and in so doing, the correlation with MOCR<sub>SF</sub> might increase.

Either OAE type could form the basis of a clinical MOCR assay. The next iteration of the assays could specifically address the factors contributing to stability and variability. Data from more ears also are needed to better clarify conditions resulting in the best variance ratios while maintaining a high level of quality data.

## B. Pragmatics of MOCR measurements

### 1. Ensuring measurement quality

One of the key determinates of a quality MOCR measurement was a high OAE SNR. An SNR that is sufficient for an OAE test is not enough for an MOCR test because the noise floor increases when the two OAE measurements are subtracted, and because a small change in the signal is to be detected. There is a trade-off between quality and the number of ears that meet the quality criteria (e.g., if low-SNR

OAEs are included, it could produce a lower variance ratio in the group). One possibility is to increase the OAE stimulus levels to produce a higher SNR. Doing so may indeed increase SNR, but the higher level may also elicit a MEMR, or even be an ipsilateral elicitor for the MOCR. Further, the resulting OAEs may be less sensitive to the MOCR. More averages can be taken too, but this takes time, and, aside from the practical considerations, can result in measurement drift.<sup>10</sup>

We do not recommend any one set of criteria to screen the OAE data for good quality. Differences in methodology, stimuli, and equipment may also mean that recommendations do not generalize. Instead we recommend that SNR criteria are carefully considered, and the rationale for their choice clearly explained. Although 6 dB is a popular choice for an acceptable SNR (Goodman *et al.*, 2013; Mishra and Lutman, 2013), this may be much too low if the aim is to accurately detect a very small MOCR (Guinan, 2012; Goodman *et al.*, 2013).

### 2. Reducing test time

The test times from this study suggest an upper bound on the necessary length of a clinical test. The measurements were made with an improvised system with timings between trials done manually. An automated test could be much faster. Here, a TEOAE trial-pair took 3 min and a single-frequency SFOAE trial-pair took 1 min, on average, but for SFOAEs to yield comparable variance ratios to TEOAEs, trials should be averaged over three frequencies thereby requiring 3 min, too. To assess stability of the measurement, three repeats (median from Table III) typically are needed for either SFOAEs or TEOAEs. This amounts to 9 min per ear, which is a lengthy test. We anticipate that test time can be sped up considerably by the use of stopping rules for the OAE tests.

TEOAEs offer a practical advantage over discrete-tone SFOAEs. Although test time is roughly similar to cluster-averaged SFOAEs, TEOAEs overall are faster because they do not require the additional tests needed to ascertain the optimal test frequencies for SFOAEs (i.e., a TEOAE test or SFOAE-gram to find high-amplitude regions, and an SOAE test to avoid affected frequency regions, which is not always possible to find in some ears). Here we selected SFOAE test frequencies individually for each ear using a time-consuming SFOAE-gram that took about 20 min; however, analyses (not reported here) indicate good frequencies can also be selected from a faster TEOAE test (adding approximately 1 min to the test), because it is a close match to the SFOAE spectrum (Kalluri and Shera, 2007).

We found higher variance ratios for MOCR<sub>SF</sub> if trials were averaged across frequency. However, for clinical purposes, testing at one frequency is much easier than testing across frequency because choosing one good SFOAE frequency is much easier than choosing multiple frequencies (when trying to find frequencies with good amplitude that are away from SOAEs, etc.). This is another area for optimization.

### 3. Avoiding a MEMR

For each ear individually, the MEMR threshold was estimated using an OAE-based measurement that is more

sensitive than the usual clinical test (Lapsley Miller and Marshall, 2014). No ear had a substantive MEMR at 65 dB SPL or below, thus the MOCR measurements were most likely due to the MOC system and not the MEMR. However, our MEMR test was time-consuming and did not give clear-cut results in many ears due to high inherent variability in the MEMR test. People could have small MEMR responses that might not be detectable even with our more sensitive measurement. This is a problem if the raw MOCR response is also small because the change from fluctuating middle-ear muscle activity may be close to the same magnitude as the MOCR. It is also probable that the wideband TEOAE stimulus is a better MEMR elicitor than the tonal SFOAE stimuli, especially in combination with the wideband BBN contralateral elicitor, though we had little indication that this was an issue in the current experiment at the levels used (although this may have accounted for the inconsistent results from Subject 81).

Although Luebke *et al.* (2014) found that MOCR strength measurements containing some contamination by MEMR activity nevertheless was predictive of PTS in rabbits, it is not yet known if the same is true in humans. Future efforts are needed to increase the sensitivity and reliability of the MEMR measurement in humans, preferably derived from the MOCR measurement itself (Goodman *et al.*, 2013; Henin *et al.*, 2014; Lapsley Miller and Marshall, 2014).

#### 4. Influence of SOAEs

Although SOAEs are affected by the MOCR (e.g., Zhao and Dhar, 2010), our results showed that the presence of SOAEs did not have any significant effect (Sec. III J) on MOCR strength. This was expected for  $MOCR_{SF}$  because we tested away from SOAEs (it is already known that SOAEs affect  $MOCR_{SF}$ ). For  $MOCR_{TE}$ , it was not possible to test away from SOAE frequencies as the measurement is broadband in nature, but averaging over frequency may have diminished their effects.

#### C. Improvements over earlier studies

Measuring MOCR strength with OAEs is a delicate business, and we have attempted to overcome or mitigate some of the methodological problems present in earlier human studies. There is still much room for improvement.

Cross-over noise was avoided by using insert earphones rather than headphones to supply the contralateral stimulus. However, in our system the ER2 earphone contained only a speaker and not a microphone so it could only be calibrated in a coupler and not further adjusted in-the-ear. Thus the actual levels presented in the ear canal may have differed from the target level, and reliability may have been affected if the ear-tip depth of insertion differed across sessions (this may affect SFOAE cluster-averages more than TEOAEs in the same band because the probe may have been refitted and recalibrated between each SFOAE frequency).

A concern we have with some earlier published studies is that not enough time elapsed between the end of a trial with contralateral stimulation and the next trial with no stimulation. If the reset time is too short, the efferent activity

may still be partially activated at the beginning of the next trial without the contralateral elicitor, and thus the difference between the OAE level with and without the contralateral elicitor would be underestimated. In the current study, at least 10 s elapsed after trials with contralateral elicitation to ensure any efferent activity had decayed so that it would not affect the next trial. Similarly, the contralateral stimulus was on for at least 2 s before the OAE measurement began, to ensure the efferent system had engaged and stabilized. Our methodology measures MOC activity on a fast-time scale, which is positively correlated in humans to the MOC activity that occurs on a slower time scale. The magnitude of the faster MOC activity is substantially larger than the slower MOC activity (Zhao and Dhar, 2010), so it seems clinically expedient to focus on the faster activity. It remains to be seen if a 10-s reset time is overkill or whether it could be reduced substantially, which would be desirable for a clinical test that takes multiple measurements.

Some earlier human studies took only one measurement pair per ear. By making repeated measurements, we were able to gauge within and across-trial reliability. Our results show that one trial is not enough. Indeed in some ears, even after many trials, it is not always possible to separate MOCR activity from the inherent test variability.

#### D. A clinical assay for HCPs

Individuals in the same noise-hazardous environment will not incur the same hearing loss—some may not get any loss, yet others may suffer significant disability (e.g., Martin, 1976; Maison and Liberman, 2000; Maison *et al.*, 2002). For normal-hearing ears with normal OAEs, an MOCR assay may indicate susceptibility to NIHL or cochlear neuropathy before any damage has occurred. Although we do not yet have an optimal clinical test, our MOCR assays have sufficient statistical properties to make field trials feasible. We have established some parameters that result in a clinically viable test—both statistically and pragmatically—such that it is worth moving to clinical trials, albeit with some optimization to speed up the test. Our assays produce valid MOCR measurements that can be used to reliably classify ears according to MOCR strength. We need to establish if any aspect of this MOCR measurement is predictive of NIHL.

Field trials would take baseline MOCR measurements in a large group of people at risk for NIHL, and then follow them longitudinally to establish risk factors for NIHL. Either SFOAEs or TEOAEs perform well enough for field research. The required time for either TEOAEs or SFOAEs is from 6 to 27 min per ear for testing if stability is to be verified (9 min per ear for most). This is too long for many HCPs, so a clinical assay would need further optimization (e.g., by using stopping rules). Some newer OAE tests might be better, such as swept-tone SFOAEs (Henin *et al.*, 2011; Kalluri and Shera, 2013).

Swept OAEs provide the best of both the TEOAE- and SFOAE-based tests, but were not available at the time of the experiment. They have the potential to give high-quality estimates of the reflection component OAE, with a wide-spectral response, and a higher SNR. At present, this

test is a laboratory technique and needs a fast clinical implementation. Using linear TEOAEs also may be a better choice (with faster data collection and higher SNR) if the stimulus artifact can be sufficiently removed.

Irrespective of which OAE test is used, ears with relatively normal hearing and low-level or absent OAEs are not good candidates for an MOCR assay. However, ears in this category are already known to be more at risk for incipient NIHL from both continuous-noise and impulse-noise sources (Lapsley Miller *et al.*, 2006; Job *et al.*, 2009; Marshall *et al.*, 2009). Our working presumption is that these low-level OAEs are due to subclinical damage to the inner ear that has not yet shown a significant change on an audiogram. As such, testing for low-level OAEs in HCPs should become part of the standard test battery. The OAE test needed can be derived from the Q trials of an MOCR test so there is no additional overhead for OAEs independently of the MOCR measurement—new enrollees can have both OAE and MOCR magnitude evaluated at entry in the same test battery.

One of the most important principles to keep in mind is how easy it is to make classification mistakes in MOCR% if the underlying measurements and/or statistics are not carefully constructed. The variance of each OAE test (i.e., with and without the contralateral elicitor) as well as the variance of the MOCR% measurement makes it difficult to do this well. If only one test series was used, or if no quality controls were used for additional test series, it would be possible, for example, for an individual with a strong MOCR to show a low MOCR% value (or vice versa) simply because the measurement was noisy. It is important to have a minimum of three trial pairs in an MOCR series, each of which must pass quality-control criteria, and to continue the trial series if necessary until a stability criterion is reached. It is also important to determine whether the measurement clearly is distinguishable from the measurement variability.

## E. A general clinical test

Our primary interest is in hearing conservation, and our aim is to establish a good clinical test for use in that context. However, a clinical MOCR assay also is of more general interest, and could be used for other applications such as diagnostic testing for various pathologies (summarized in Kumar *et al.*, 2013), tracking improvements in listening strategies due to increased MOC functioning (reviewed in Guinan, 2012), and identifying those who could benefit from auditory training (summarized in Guinan, 2011). It could also be part of a test battery to select people with exceptional performance potential (e.g., Andeol *et al.*, 2011 show that MOCR strength is correlated with localization ability at low SNRs). Requirements for classification of ears according to MOCR strength in these contexts are not so severe, and there is likely more time for making quality measurements compared to the fast, mass-screening done in HCPs. Therefore, our current test could be used for some applications even before it is optimized.

## ACKNOWLEDGMENTS

Thanks to Mike Katzenbach from Mimosa Acoustics for technical support, Linton Miller for useful discussions about

the analyses, and Tim Villabona from MIT for assistance with data collection. Thanks to Kelly Watts from NSMRL for providing many helpful, constructive comments. The work was supported by ONR Grant/Agreement No. N0001412WX20904 “Susceptibility to Noise-Induced Hearing Loss using Otoacoustic Emissions,” NSMRL Work Unit No. 50904. L.M. is an employee of the U.S. Government. This work was prepared as part of her official duties. Title 17 U.S.C. §105 provides that ‘Copyright protection under this title is not available for any work of the United States Government.’ Title 17 U.S.C. §101 defines a U.S. Government work as a work prepared by a military service member or employee of the U.S. Government as part of that person’s official duties. The views expressed in this article are those of the authors and do not reflect the official policy or position of the Department of the Navy, the Department of Defense, or the United States Government.

<sup>1</sup>Eighty-seven subjects were screened. Thirty completed the experiment, but three were excluded from the final dataset because *post hoc* analyses of their MEMR thresholds yielded indeterminate results. Of the remaining 57 subjects, 19 passed screening but did not complete the experiment and 38 did not meet the screening criteria (9 did not meet hearing threshold criteria, 13 met hearing threshold criteria but did not have sufficient TEOAE levels, and 16 met hearing threshold and TEOAE criteria but did not have an SFOAE test-frequency solution, usually due to too many SOAEs and/or insufficient SFOAE amplitude at the frequencies away from SOAEs).

<sup>2</sup>The HearID and SFOAE-SG software systems were custom modified by Mimosa Acoustics part way through the experiment to provide the tester prompts for turning the contralateral noise on and off, to expedite file naming and saving, and to time the minimum duration between trials. No modifications were made to the OAE measurement routines.

<sup>3</sup>Chirps were chosen over the more commonly-used click stimulus because they showed fewer artifacts in ears with no OAEs (Lapsley Miller *et al.*, 2004a). Their longer duration and thus slower presentation rate also reduced the chances of unintended MOCR activity elicited by the stimulus (Veuille *et al.*, 1991; Guinan *et al.*, 2003).

<sup>4</sup>After data collection had finished, we found the TEOAE module in HearID R3 using the HearID PC-card did not provide sufficient microphone equalization. Because we used spectrum calibration to flatten the TEOAE stimulus in-the-ear, the missing equalization affected the shape of that spectrum and it was not flat as intended. The ER-10C microphone does not have a particularly flat frequency response, so without it the TEOAE stimulus spectrum and TEOAE response spectrum were slightly enhanced at 2.5 kHz and diminished by about 6 to 9 dB at 4 kHz. It should be noted that no other TEOAE measurement system that we know of attempts to flatten the stimulus in individual ears, and the spectrum presented in many ears would be far from flat. It should also be noted that we can only measure the spectrum at the probe microphone; what is delivered at the tympanic membrane can be far from flat due to standing waves. So although not intended, and not ideal, the non-flat spectrum does not compromise the findings in this paper. A *post hoc* equalization fix was applied to the TEOAE data. It could not correct the stimulus, but it could correct the TEOAE response. Informal testing was done on a group of ears: (a) with equalization for both stimulus calibration and response, (b) without equalization on the stimulus but with *post hoc* equalization on the response, and (c) with no equalization. In most cases, the *post hoc* equalization produced results close to those made with equalization, compared to those without any equalization. So despite the non-flat stimulus, the main effect of the missing equalization was on the TEOAE response. The MOCR strength estimate was not unduly compromised because the MOCR effect is greatest below 2 kHz, and because the MOCR measurement is a relative difference. For SOAEs, only those below 2.5 kHz were of interest so although the measurement was affected by the missing equalization, the effect was minimal.

<sup>5</sup>The *optimal-frequency* ensemble included one test frequency from each band of interest: 1 to 1.5, 1.5 to 2, and 2 to 2.5 kHz. They were chosen using an algorithm that considered stimulus quality (measured levels

- within 3 dB of target), SFOAE magnitude ( $> -5$  dB SPL), SFOAE SNR ( $>15$  dB), noise floor ( $< -15$  dB SPL), measurement validity (group delay based on at least 3 of the 5 surrounding frequencies, including the test frequency, with SNR  $>6$  dB and resulting correlation coefficient  $-0.96$ ; Lapsley Miller *et al.*, 2004b), flatness of the SFOAE spectrum (nearest neighbors within 2.5 dB), spacing (at least 0.35 kHz apart), and proximity of SOAEs (at least 1 microstructure period away from SOAEs; Shera, 2003). All potential frequency triplets were determined, and the triplet with the largest average amplitude was chosen.
- <sup>6</sup>In one case, the two additional frequencies in the cluster were both above the test frequency (Subject 70), and in two cases, a cluster in the 1 to 1.5 kHz band (Subject 73) or 2 to 2.5 kHz band (Subject 31) was chosen. This was typically because of SOAE proximity or because the SFOAE spectrum in the test region was not flat enough to find appropriate neighboring frequencies.
- <sup>7</sup>A percent change is easily converted to a dB change. MOC-induced OAE changes are usually measured as the difference in OAE level without and with a MOCR elicitor with both in dB SPL, which yields a dB change.  $MOCR_{SF}$  is different because it takes phase into account, and to help indicate this difference, Backus and Guinan (2007) used percent instead of dB for  $MOCR_{SF}$ . Since  $MOCR_{TE}$  was also calculated from a vector difference, we also used percent instead of dB for it too.
- <sup>8</sup>In these box-plots, the median is represented by the horizontal line within the box. The top and bottom of the box represents the interquartile range. The top and bottom whisker heights are the most extreme values within 1.5 times the interquartile range from the ends of the box. The circles represent outliers that fall outside the whiskers.
- <sup>9</sup>Each individual MOCR measurement series (consisting of  $\sim 15$  MOCR% estimates, and excluding those series with fewer than 5 trial-pairs meeting OAE quality control criteria), was subsampled without replacement to obtain a number of trial-pairs varying from 1 to the maximum number of trial-pairs available, and the aggregate MOCR% of the resulting dataset was calculated. Each random subsampling was repeated 250 times (with replacement) to build up a distribution of the MOCR% SD for each trial-pair length. As the number of trial-pairs increased, the mean of the MOCR SD tended to increase asymptotically, converging to the SD for all trial-pairs. This brute force method was considered sufficient for the problem at hand, compared with a more accurate, but computationally more resource-intensive combinatorial method where all combinations are accounted for.
- <sup>10</sup>Goodman *et al.* (2013) has suggested de-trending techniques to nullify the effects of slow stimulus level drift, which are not completely controlled by interleaving the noise-on condition with the noise-off condition. This was not possible with our implementation. Visual inspection of the data did not indicate drift was a major contributor to MOCR variability, possibly because our MOCR% derivation is applied on a trial-by-trial basis, and not calculated over the entire interleaved measurement series.
- Abdala, C., Mishra, S., and Garinis, A. (2013). "Maturation of the human medial efferent reflex revisited," *J. Acoust. Soc. Am.* **133**, 938–950.
- Abdala, C., Mishra, S. K., and Williams, T. L. (2009). "Considering distortion product otoacoustic emission fine structure in measurements of the medial olivocochlear reflex," *J. Acoust. Soc. Am.* **125**, 1584–1594.
- Andeol, G., Guillaume, A., Micheyl, C., Savel, S., Pellioux, L., and Moulin, A. (2011). "Auditory efferents facilitate sound localization in noise in humans," *J. Neurosci.* **31**, 6759–6763.
- ANSI (2004). S3.21, *Methods for Manual Pure-Tone Threshold Audiometry* (American National Standards Institute, New York).
- Backus, B. C., and Guinan, J. J. (2007). "Measurement of the distribution of medial olivocochlear acoustic reflex strengths across normal-hearing individuals via otoacoustic emissions," *J. Assoc. Res. Otolaryngol.* **8**, 484–496.
- Bamber, D. (1975). "The area above the ordinal dominance graph and the area below the receiver operating characteristic graph," *J. Math. Psychol.* **12**, 387–415.
- Bray, P. J. (1989). "Click evoked otoacoustic emissions and the development of a clinical otoacoustic hearing test instrument," Ph.D. thesis, Institute of Laryngology and Otolaryngology, University College and Middlesex School of Medicine, London.
- Froehlich, P., Collet, L., Valatx, J. L., and Morgon, A. (1993). "Sleep and active cochlear micromechanical properties in human subjects," *Hear. Res.* **66**, 1–7.
- Ghiselli, E. E. (1964). *Theory of Psychological Measurement* (McGraw-Hill, New York), p. 228.
- Goodman, S. S., Mertes, I. B., Lewis, J. D., and Weissbeck, D. K. (2013). "Medial olivocochlear-induced transient-evoked otoacoustic emission amplitude shifts in individual subjects," *J. Assoc. Res. Otolaryngol.* **14**, 829–842.
- Guinan, J. J., Jr. (1996). "Physiology of olivocochlear efferents," in *The Cochlea*, edited by P. Dallos, A. N. Popper, and R. R. Fay (Springer-Verlag, New York), pp. 435–502.
- Guinan, J. J., Jr. (2006). "Olivocochlear efferents: Anatomy, physiology, function, and the measurement of efferent effects in humans," *Ear Hear.* **27**, 589–607.
- Guinan, J. J., Jr. (2011). "Physiology of the medial and lateral olivocochlear systems," in *Auditory and Vestibular Efferents*, edited by D. K. Ryugo, R. R. Fay, and A. N. Popper (Springer Science+Business Media, LLC, New York), pp. 39–81.
- Guinan, J. J., Jr. (2012). "Efferent System," in *Translational Perspectives in Auditory Neuroscience: Normal Aspects of Hearing*, edited by K. Tremblay and R. Burkard (Plural Publishing, San Diego, CA), pp. 283–323.
- Guinan, J. J., Jr., Backus, B. C., Lilaonitkul, W., and Aharonson, V. (2003). "Medial olivocochlear efferent reflex in humans: Otoacoustic emission (OAE) measurement issues and the advantages of stimulus frequency OAEs," *J. Assoc. Res. Otolaryngol.* **4**, 521–540.
- Henin, S., Long, G. R., and Thompson, S. (2014). "Wideband detection of middle ear muscle activation using swept-tone distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **136**, 272–283.
- Henin, S., Thompson, S., Abdelrazeq, S., and Long, G. R. (2011). "Changes in amplitude and phase of distortion-product otoacoustic emission fine-structure and separated components during efferent activation," *J. Acoust. Soc. Am.* **129**, 2068–2079.
- Job, A., Raynal, M., Kossowski, M., Studler, M., Ghernaoui, C., Baffioni-Venturi, A., Roux, A., Darolles, C., and Guelorget, A. (2009). "Otoacoustic detection of risk of early hearing loss in ears with normal audiograms: A 3-year follow-up study," *Hear. Res.* **251**, 10–16.
- Kalluri, R., and Shera, C. A. (2001). "Distortion-product source unmixing: a test of the two-mechanism model for DPOAE generation," *J. Acoust. Soc. Am.* **109**, 622–637.
- Kalluri, R., and Shera, C. A. (2007). "Near equivalence of human click-evoked and stimulus-frequency otoacoustic emissions," *J. Acoust. Soc. Am.* **121**, 2097–2110.
- Kalluri, R., and Shera, C. A. (2013). "Measuring stimulus-frequency otoacoustic emissions using swept tones," *J. Acoust. Soc. Am.* **134**, 356–368.
- Kumar, U. A., Methi, R., and Avinash, M. C. (2013). "Test/retest repeatability of effect contralateral acoustic stimulation on the magnitudes of distortion product otoacoustic emissions," *Laryngoscope* **123**, 463–471.
- Lapsley Miller, J. A., Boege, P., Marshall, L., and Jeng, P. S. (2004a). "Transient-evoked otoacoustic emissions: Preliminary results for validity of TEOAEs implemented on Mimosa Acoustics' T2K measurement system v3.1.3," NSMRL Technical Report No. 1232 (Naval Submarine Medical Research Laboratory, Groton, CT).
- Lapsley Miller, J. A., Boege, P., Marshall, L., Shera, C., and Jeng, P. S. (2004b). "Stimulus-frequency otoacoustic emissions: Validity and reliability of SFOAEs implemented on Mimosa Acoustics' SFOAE measurement system v2.1.18," NSMRL Technical Report No. 1231 (Naval Submarine Medical Research Laboratory, Groton, CT).
- Lapsley Miller, J. A., and Marshall, L. (2014). "Three methods for estimating the middle-ear muscle reflex (MEMR) using otoacoustic emission (OAE) measurement systems," NSMRL Technical Report No. 1310 (Naval Submarine Medical Research Laboratory, Groton, CT).
- Lapsley Miller, J. A., Marshall, L., Heller, L. M., and Hughes, L. M. (2006). "Low-level otoacoustic emissions may predict susceptibility to noise-induced hearing loss," *J. Acoust. Soc. Am.* **120**, 280–296.
- Liberman, M. C., Liberman, L. D., and Maison, S. F. (2014). "Efferent feedback slows cochlear aging," *J. Neurosci.* **34**, 4599–4607.
- Lilaonitkul, W., and Guinan, J. J., Jr. (2009a). "Human medial olivocochlear reflex: Effects as functions of contralateral, ipsilateral, and bilateral elicitor bandwidths," *J. Assoc. Res. Otolaryngol.* **10**, 459–470.
- Lilaonitkul, W., and Guinan, J. J., Jr. (2009b). "Reflex control of the human inner ear: A half-octave offset in medial efferent feedback that is consistent with an efferent role in the control of masking," *J. Neurophysiol.* **101**, 1394–1406.
- Long, G. R., Talmadge, C. L., and Lee, J. (2008). "Measuring distortion product otoacoustic emissions using continuously sweeping primaries," *J. Acoust. Soc. Am.* **124**, 1613–1626.



- Luebke, A. E., and Foster, P. K. (2002). "Variation in inter-animal susceptibility to noise damage is associated with alpha 9 acetylcholine receptor subunit expression level," *J. Neurosci.* **22**, 4241–4247.
- Luebke, A. E., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (2014). "Adaptation of distortion product otoacoustic emissions predicts susceptibility to acoustic over-exposure in alert rabbits," *J. Acoust. Soc. Am.* **135**, 1941–1949.
- Maison, S. F., and Liberman, M. C. (2000). "Predicting vulnerability to acoustic injury with a noninvasive assay of olivocochlear reflex strength," *J. Neurosci.* **20**, 4701–4707.
- Maison, S. F., Luebke, A. E., Liberman, M. C., and Zuo, J. (2002). "Efferent protection from acoustic injury is mediated via alpha9 nicotinic acetylcholine receptors on outer hair cells," *J. Neurosci.* **22**, 10838–10846.
- Maison, S. F., Usubuchi, H., and Liberman, M. C. (2013). "Efferent feedback minimizes cochlear neuropathy from moderate noise exposure," *J. Neurosci.* **33**, 5542–5552.
- Marshall, L., and Heller, L. M. (1996). "Reliability of transient-evoked otoacoustic emissions," *Ear Hear.* **17**, 237–254.
- Marshall, L., and Lapsley Miller, J. A. (2014). "How can the auditory efferent system protect our ears from noise-induced hearing loss? Let us count the ways," in *Proceedings of the 12th International Mechanics of Hearing Workshop*, June 23, Attica, Greece.
- Marshall, L., Lapsley Miller, J. A., Heller, L. M., Wolgemuth, K. S., Hughes, L. M., Smith, S., and Kopke, R. (2009). "Detecting incipient inner-ear damage from impulse noise with otoacoustic emissions," *J. Acoust. Soc. Am.* **125**, 995–1013.
- Martin, A. (1976). "The equal energy concept applied to impulse noise," in *Effects of Noise on Hearing*, edited by D. Henderson, R. Hamernik, D. S. Dosanjh, and J. H. Mills (Raven Press, New York), pp. 421–454.
- Meinke, D. K., Stagner, B. B., Martin, G. K., and Lonsbury-Martin, B. L. (2005). "Human efferent adaptation of DPOAEs in the L(1),L(2) space," *Hear. Res.* **208**, 89–100.
- Mimosa Acoustics (2007). "TE Manual: Transient-evoked otoacoustic emissions measurement module manual for HearID R3.3," User Manual (Mimosa Acoustics, Inc., Champaign, IL).
- Mishra, S. K., and Lutman, M. E. (2013). "Repeatability of click-evoked otoacoustic emission-based medial olivocochlear efferent assay," *Ear Hear.* **34**, 789–798.
- Muller, J., Dietrich, S., and Janssen, T. (2010). "Impact of three hours of discotheque music on pure-tone thresholds and distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **128**, 1853–1869.
- Muller, J., and Janssen, T. (2008). "Impact of occupational noise on pure-tone threshold and distortion product otoacoustic emissions after one workday," *Hear. Res.* **246**, 9–22.
- Nordmann, A. S., Bohne, B. A., and Harding, G. W. (2000). "Histopathological differences between temporary and permanent threshold shift," *Hear. Res.* **139**, 13–30.
- Shera, C. A. (2003). "Mammalian spontaneous otoacoustic emissions are amplitude-stabilized cochlear standing waves," *J. Acoust. Soc. Am.* **114**, 244–262.
- Shera, C. A., and Guinan, J. J., Jr. (1999). "Evoked otoacoustic emissions arise by two fundamentally different mechanisms: Taxonomy for mammalian OAEs," *J. Acoust. Soc. Am.* **105**, 782–798.
- Shupak, A., Tal, D., Sharoni, Z., Oren, M., Ravid, A., and Pratt, H. (2007). "Otoacoustic emissions in early noise-induced hearing loss," *Otol. Neurotol.* **28**, 745–752.
- Talmadge, C. L., Long, G. R., Tubis, A., and Dhar, S. (1999). "Experimental confirmation of the two-source interference model for the fine structure of distortion product otoacoustic emissions," *J. Acoust. Soc. Am.* **105**, 275–292.
- VeUILLET, E., Collet, L., and Duclaux, R. (1991). "Effect of contralateral acoustic stimulation on active cochlear micromechanical properties in human subjects: Dependence on stimulus variables," *J. Neurophysiol.* **65**, 724–735.
- Wagner, W., Heppelmann, G., Muller, J., Janssen, T., and Zenner, H. P. (2007). "Olivocochlear reflex effect on human distortion product otoacoustic emissions is largest at frequencies with distinct fine structure dips," *Hear. Res.* **223**, 83–92.
- Wagner, W., and Heyd, A. (2011). "Measurement of medial olivocochlear efferent activity in humans: comparison of different distortion product otoacoustic emission-based paradigms," *Otol. Neurotol.* **32**, 1379–1388.
- Wolpert, S., Heyd, A., and Wagner, W. (2014). "Assessment of the noise-protective action of the olivocochlear efferents in humans," *Audiol. Neuro-Otol.* **19**, 31–40.
- Zhao, W., and Dhar, S. (2010). "The effect of contralateral acoustic stimulation on spontaneous otoacoustic emissions," *J. Assoc. Res. Otolaryngol.* **11**, 53–67.