



Published in final edited form as:

*Ear Hear.* 2019 ; 40(6): 1267–1279. doi:10.1097/AUD.0000000000000711.

## Reliability of measures intended to assess threshold-independent hearing disorders.

Aryn M. Kameroner<sup>1</sup>, Judy G. Kopun<sup>1</sup>, Sara E. Fultz<sup>1</sup>, Stephen T. Neely<sup>1</sup>, Daniel M. Rasetshwane<sup>1</sup>

<sup>1</sup>Boys Town National Research Hospital, Omaha, NE 68131

### Abstract

**Objectives:** Recent animal studies have shown that noise exposure can cause cochlear synaptopathy without permanent threshold shift. Because the noise exposure preferentially damaged auditory nerve fibers that processed suprathreshold sounds (low-spontaneous rate fibers), it has been suggested that synaptopathy may underlie suprathreshold hearing deficits in humans. Recently, several researchers have suggested measures to identify the pathology or pathologies underlying suprathreshold hearing deficits in humans based on results from animal studies, however, the reliability of some of these measures have not been assessed. The purpose of this study was to assess the test-retest reliability of measures that may have the potential to relate suprathreshold hearing deficits to site(s)-of-lesion along the peripheral auditory system in humans.

**Design:** Adults with audiometric normal hearing were tested on a battery of behavioral and physiologic measures that included (1) thresholds in quiet (TIQ), (2) thresholds in noise (TIN), (3) frequency-modulation detection threshold (FMDT), (4) word recognition in four listening conditions, (5) distortion-product otoacoustic emissions (DPOAE), (6) middle ear muscle reflex (MEMR), (7) tone-burst-elicited auditory brainstem response (tbABR), and (8) speech-evoked ABR (sABR). Data collection for each measure was repeated over two visits separated by at least one week. The residuals of the correlation between the suprathreshold measures and TIQ serve as functional and quantitative proxies for threshold-independent hearing disorders because they represent the portion of the raw measures that is not dependent on thresholds in quiet. Reliability of the residual measures was assessed using intraclass correlation (ICC).

**Results:** Reliability for the residual measures was good (ICC = 0.75) for FMDT, DPOAEs, and MEMR. Residual measures showing moderate reliability (0.5 < ICC < 0.75) were tbABR wave I amplitude, TIN, and word recognition in quiet, noise, and time-compressed speech with reverberation. Wave V of the tbABR, waves of the sABR, and recognition of time-compressed words had poor test-retest reliability (ICC < 0.5).

**Conclusions:** Reliability of residual measures were mixed, suggesting that care should be taken when selecting measures for diagnostic tests of threshold-independent hearing disorders. Quantifying hidden hearing loss as the variance in suprathreshold measures of auditory function

Address correspondence to Aryn M. Kameroner, Boys Town National Research Hospital, 555 N. 30<sup>th</sup> St, Omaha, NE 68131, USA. aryn.kameroner@boystown.org, Telephone: 1+531-355-5070, Fax: 1+531-355-6351.

**Author contributions:** A.K. analyzed data and wrote the paper; J.K and S.F. collected and managed data; S.N. and D.R. conceptualized and designed the study, designed and wrote software for data collection, and analyzed data. All authors discussed the results and implications and commented on the manuscript at all stages.

that is not due to thresholds in quiet may provide a reliable estimate of threshold-independent hearing disorders in humans.

---

## INTRODUCTION

For decades, clinical audiologists have heard complaints from patients regarding their inability to understand speech in noise despite normal audiometric thresholds. While the existence of deficits in processing of suprathreshold stimuli with normal audiometric thresholds has been known to clinicians, the anatomical pathology/pathologies underlying this phenomenon is a mystery (Alvord, 1983). Recent animal studies have demonstrated a mechanism of noise-induced hearing loss which could explain such a clinical complaint. Experiments in which rodents were exposed to high levels of noise for a short period of time, or moderate levels over a longer duration, resulted in damage to the synapses (synaptopathy) between inner hair cells and auditory nerve fibers (Kujawa & Liberman, 2006, 2009, 2015; Lin et al., 2011; Maison et al., 2013; Sergeyenko et al., 2013; Furman et al., 2013; Liberman & Liberman, 2015; Fernandez et al., 2015). This synaptopathy was usually followed by neural degeneration (neuropathy). The loss of synapses surprisingly did not lead to substantial permanent change in auditory sensitivity but led to a reduction to wave I amplitude of the auditory brainstem response (ABR). In these studies, low spontaneous-rate auditory nerve fibers, which code mostly suprathreshold sounds, were more susceptible to noise damage than their high spontaneous-rate counterparts which code mostly threshold-level sounds (Liberman, 1978).

It has been suggested that synaptopathy may underlie suprathreshold hearing deficit in humans (Schaette & McAlpine, 2011; Plack et al., 2014; Liberman et al., 2016; Mehraei et al., 2016). However, since synaptopathy cannot be confirmed in live humans, researchers are evaluating whether several physiological measures can serve as proxies for synaptopathy in humans, based on results from animal studies. Several studies have evaluated the utility of ABR wave I, particularly its relation to noise exposure. A few studies demonstrated a relationship between ABR wave I and noise-exposure (e.g. Stamper & Johnson, 2015; Bramhall et al., 2017; Valderrama et al., 2018). A relationship has also been observed between the ratio of summing and action potentials (SP/AP; the AP is the electrocochleography equivalent of ABR wave I) and noise exposure (Liberman et al., 2016). However, other studies did not find evidence of this relationship (e.g. Fulbright et al., 2017; Prendergast et al., 2017; Guest et al., 2018a; 2018b).

The utility of the middle-ear muscle reflex (MEMR) as a proxy for synaptopathy has also been investigated (Valero et al., 2016; Wojtczak et al., 2017) because the MEMR has been suggested as an indicator the integrity of low spontaneous-rate auditory nerve fibers (Liberman & Kiang, 1984; Ryugo & Rouiller, 1988; Kobler et al., 1992). While acoustic stapedial reflex thresholds and steady-state decay elicited using tonal stimuli are commonly used in the clinic to identify ears with cochlear and retro-cochlear pathologies (Gelfand et al., 1990), there is evidence that the magnitude of the response or the response shift over time may have better diagnostic value (Chertoff et al., 2018). In addition, wideband-elicited MEMR may produce lower thresholds than clinical thresholds, at least when combined with

broadband elicitors, which may allow testing on populations with larger hearing loss (Keefe et al., 2010; Feeney et al., 2017).

Temporal coding deficits are also suspected to result from cochlear neuropathy (e.g. Bharadwaj et al., 2014; Bharadwaj et al., 2015). Speech ABR (sABR) is a physiologic measure that reflects temporal neural coding ability of both transient events (e.g. onset of the syllable) and sustained features (e.g. steady-state portion of the vowel; Johnson et al., 2005; Chandrasekaran & Kraus, 2006; Akhoun et al., 2008; Sinha & Basavaraj, 2010; Skoe & Kraus, 2010). The sABR has also been demonstrated to predict speech-in-noise perception difficulties more accurately than some behavioral measures (Anderson et al., 2013). Temporal coding abilities can also be assessed behaviorally using frequency-modulation detection threshold (FMDT), which has been shown to correlate with performance on speech-in-noise tasks (Strelcyk & Dau, 2009; Papakonstantinou et al., 2011; Bharadwaj et al., 2014; Johannesen et al., 2015, 2016).

Noise-induced synaptopathy would likely manifest in humans as suprathreshold hearing deficits in spite of normal sensitivity (Lieberman et al., 2016; Mehraei et al., 2016; Schaeffe & McAlpine, 2011). Word recognition scores in quiet and adverse listening conditions are a clinical tool used to gain insight into a patient's suprathreshold hearing deficits. Liberman et al. (2016) demonstrated correlation between NU-6 speech scores and the SP/AP ratio, suggesting a link between synaptopathy and speech understanding in difficult listening situations. (Lobrinias et al. (2016) showed that selective IHC loss leads to elevated behavioral thresholds in noise (TIN), while thresholds in quiet remain unchanged. Because IHC loss also leads to reduced ABR Wave I amplitude (Wang et al., 2002; El-Badry & McFadden, 2007), similar to noise-induced cochlear synaptopathy, TIN may reflect synaptopathy. Ridley et al. (2018) demonstrated a relationship between TIN (after residualizing it to remove the variability due to thresholds in quiet) and the SP/AP ratio and ABR waves I amplitude (measures that have been hypothesized as proxies for synaptopathy).

Although some promising results have been observed in the search for proxies for synaptopathy, the reliabilities of the measures that researchers are investigating as proxies has not been assessed. The present study assessed the test-retest reliability of the residuals of a number of physiologic measures that have been suggested as proxies for synaptopathy, as well as behavioral measures of suprathreshold hearing deficits. The measures include: (1) thresholds in noise (TIN), (2) frequency-modulation detection threshold (FMDT), (3) word recognition scores, (4) distortion-product otoacoustic emissions (DPOAEs), (5) middle-ear muscle reflex (MEMR), (6) tone-burst ABR (tbABR), and (7) speech ABR (sABR). Similar to most previous studies that investigate proxies for synaptopathy (e.g., Liberman et al., 2016; Mehraei et al., 2016), we included otoacoustic emissions to account for possible damage to outer hair cells unrelated to synaptopathy.

Data collection for each measure was repeated over two visits separated by at least one week. The residuals of the correlation between the measures of interest and threshold in quiet were calculated using multivariate linear regression analysis. The residuals served as functional and quantitative proxies for hidden hearing loss because they represent the

portion of the raw measures that is not dependent on thresholds in quiet. Reliability of the residuals was assessed using intraclass correlation. The results of this study will provide useful information as to which measures researchers should focus on in their quest to identify proxies for synaptopathy, as well as efforts to develop diagnostic procedures for suprathreshold hearing deficits.

## MATERIALS AND METHODS

### Participants

Seventeen adults with normal hearing participated in the study (10 female, mean age = 28, SD=8, range = 19 to 50 years). Standard audiometric and tympanometric procedures were used to determine inclusion into the study. Pure-tone air conduction thresholds (ER-3A insert earphones; Etymotic Research, Elk Grove, IL) at octave frequencies (0.25 to 8 kHz) and two inter-octave frequencies (3 and 6 kHz) were measured using an audiometer (GSI AudioStar Pro, Grason-Stadler) in 5 dB steps using the modified Hughson-Westlake procedure. Participants were required to have thresholds  $\leq 15$  dB hearing level (HL) at all frequencies. Additionally, participants were required to have normal middle ear function as assessed by 226 Hz tympanometry (Otoflex 100, Madsen) which included a middle ear pressure range of  $-100$  to  $+50$  daPa and static compliance between  $0.3$  and  $2.5$  cm<sup>3</sup>. All measures were made monaurally; if both ears met the inclusion criteria, the better ear was chosen for testing. If both ears had similar audiometric thresholds, the test ear was selected randomly. In total, there were nine right ears and eight left ears included in the study.

### Procedures

Participants completed all measures in random order over the course of two visits. Participants then returned to repeat all measures over another two visits, with at least a one-week difference in time between the first and second measurement. The test time required to collect all measures over two separate visits was approximately eight hours per participant. All procedures were approved by the Boys Town National Research Hospital Institutional Review Board, and informed consent was obtained from all participants. Participants were paid for their participation.

### Behavioral Measures

Audiometric thresholds represent the minimum level of a pure-tone stimulus that is audible to the participant. Common clinical methods of audiometry employ variants of the classical method of limits with characteristics of the staircase method. While the benefits of clinical audiometry are that it can be performed quickly and in a standardized manner across clinics, a major downfall of clinical audiometry is its susceptibility to a variety of response biases. These include the interval bias: entrainment to the stimulus interval; and effects of age: older people are more likely to wait until they are sure they heard something before they respond (Yost, 1978; Gelfand, 1982). A multiple alternative forced choice psychoacoustic method mitigates some of these biases. Stimulus parameters for the multiple alternative forced choice can mimic those of the standard clinical procedures. When three intervals are used for the procedure (3AFC), the three intervals are presented and only one interval, which is randomly assigned, contains the stimulus manipulation that is being investigated. The

participant is required to indicate which interval contained the stimulus with the manipulation under test or which interval was different from the other two. Stimulus adjustments for the subsequent interval can be determined based on the up-down adaptive tracking method. Any measure of psychometric threshold can use a variety of up-down methods to estimate various points on the psychometric function. The modified Hughson-Westlake procedure (Carhart & Jerger, 1959) is used most often in clinical audiometry, however the transformed up-down procedure has been shown to be more accurate (Levitt, 1971). Several researchers have compared the two methods and found that multiple-alternative forced-choice methods result in lower thresholds by up to 6 dB (Marshall & Jesteadt, 1986; Gatehouse Davis, 1992). In the present study, a 3AFC procedure was used for measurements of TIQ, TIN, and FMDT, using the AudioLab MATLAB package developed by Lopez-Poveda. For each measurement, a 2-up, 1-down adaptive procedure was used to track the 71% point on the psychometric function.

In addition to the psychoacoustic procedures, test-retest reliability for word recognition was assessed. All behavioral stimuli were presented monaurally via ER-3A insert earphones (Etymotic Research, Elk Grove, IL) in a soundproof room.

**Thresholds in quiet.**—A 3AFC adaptive procedure was used to measure auditory thresholds in quiet (TIQ). This test was done in addition to audiometric thresholds in order to obtain a more accurate estimate of hearing sensitivity, and consequently a more accurate estimate of the residual for the suprathreshold measures, such as thresholds in noise and frequency-modulation detection thresholds. Thresholds were assessed at 1.5 and 4 kHz. Participants completed one practice run to familiarize with the test procedure and then two performance runs which were averaged to determine TIQ. The initial stimulus level was 30 dB SPL. The initial step size, which lasted for two reversals, was 6 dB. The step size was then reduced to 3 dB. Data collection continued until a total of five reversals occurred. The mean of the last three reversals was used to determine threshold (dB SPL). Threshold in quiet was used for calculation of the residuals of suprathreshold measures. Our hypothesis was that the residual can serve as a proxy for hidden hearing loss because the residual is the portion of the variance of the suprathreshold measures that is independent of TIQ.

**Thresholds in Noise.**—The same 3AFC procedure was also used to measure thresholds in noise (TIN) at 1.5 and 4 kHz. Broadband noise from 0.2–8 kHz was set at a constant 70 dB SPL for both frequencies and the level of the tone was varied to determine threshold. The initial tone stimulus was 70 dB SPL. The initial step size was 6 dB which lasted for two reversals before reducing to 3 dB. The task continued until twelve reversals occurred. The mean of the last four reversals determined threshold. Participants completed one practice run to familiarize with the test procedure and then two performance runs. Thresholds from the performance runs were averaged to determine TIN. Threshold in noise in dB SPL was the variable used in the reliability analysis.

**Frequency-modulation detection threshold.**—Temporal processing abilities were assessed using frequency-modulation detection thresholds (Strelcyk & Dau (2009). FMDT was defined as the minimum detectable excursion in frequency (Johannesen et al., 2016). FMDT was measured using a 3AFC method, similar to that used for TIQ and TIN. The

experiment was identical to that of Johannesen et al. (2016): a pure tone of 1.5 kHz with a duration of 750 ms was presented at 70 dB SPL. In one interval, one tone was frequency-modulated with a variable maximum frequency excursion. The minimum detectable excursion in Hz was estimated and log-transformed to be used in the reliability analysis. In order to prevent the participants from using cues based on changes to excitation patterns in the cochlea, the tones in all three intervals were also sinusoidally-amplitude-modulated with a modulation depth of  $m = 0.333$  or  $20\log_{10}(1 + m/1 - m) = 6$  dB (Moore & Glasberg, 1989; Moore & Sek, 1996; Johannesen et al., 2016). Following Johannesen et al., 2016, the initial and final modulation rates were randomized in the interval between 1 and 3 Hz under the constraint that the modulation rate change was always above 1 Hz. The initial step size of the frequency excursion was  $\log_{10}(1.5)$ . This was decreased to  $\log_{10}(1.26)$  after four reversals. The adaptive procedure continued until a total of twelve reversals in frequency excursion had occurred. The mean of the last four reversals was used to determine FMDT ( $\log_{10}\text{Hz}$ ). One practice run and three performance runs were completed. A run was excluded and repeated if the standard deviation was  $>0.15$  (Strelcyk & Dau, 2009). Thresholds from the three performance runs were averaged and used in reliability analysis.

**Word Recognition.**—Word recognition scores were assessed in four listening conditions: 1) speech in quiet, 2) speech in the presence of noise, 3) speech that had been time-compressed by 45%, and 4) speech that had been time-compressed by 45% and a reverberation time of 0.3 sec (Noffsinger et al., 1994). The stimuli were four 50-word lists spoken by a male talker (NU-6; Auditec, Inc., St. Louis, MO). The words were presented at 60 dB SPL. The order of the four conditions was randomized for each participant. The noise condition consisted of steady noise, spectrally-weighted using the international long-term average speech spectrum for combined male and female talkers and presented at 0 dB SNR (Byrne et al., 1994). Performance was assessed by a single scorer, seated in the booth across from the participant and listening to the stimuli through a unilateral monitor headphone (DJ Pro 300 Stanton, Deerfield Beach, FL). This allowed the scorer to use lip reading cues to help clarify participant responses. The score, or number of correct words out of 50, for each condition, were the variables used in the reliability analysis.

### Physiologic Measures

**DPOAE.**—DPOAEs were measured monaurally using custom-designed software (EMAV, version 3.3; Neely & Liu, 1994) following the same protocol as Ridley et al. (2018). In short, two primary tones ( $f_1$  and  $f_2$ ) were generated by two separate channels of a 24-bit soundcard (Hammerfall DSP Multi-Face II, RME, Germany) and sent to two separate ER-3A insert headphones and then routed to sound ports housed in the ER-10B+ probe microphone system (Etymic Research, Elk Grove Village, IL, USA). DPOAEs were measured at  $f_2 = 1.5$  and 4 kHz. The  $f_2/f_1$  ratio was 1.22. The level of  $f_2$  was  $L_2 = 55$  dB SPL and the level of  $f_1$  was set at  $L_1 = 61$  dB SPL in accordance with Kummer et al. (2000). Stimulus levels were calibrated in-ear and DPOAEs were recorded via the ER-10B+ microphone housed in the probe system. DPOAEs were collected in two separate buffers. The  $2f_1 - f_2$  frequency bin (resolution of 3.9 Hz) of the two buffers was summed to determine the level of the DPOAE. Noise was calculated by subtracting the contents of the two buffers then averaging the  $2f_1 - f_2$  frequency bin and the five bins on either side. Data collection

ended when one of the three following criteria were met: 1) the noise floor was  $< -20$  dB SPL; 2) artifact-free averaging time was  $>65.6$  sec; or 3) SNR reached 60 dB.

**Middle ear muscle reflex.**—The procedure for wideband measurement of MEMR developed by Keefe et al. (2010) was used in the present study. This protocol uses a pulsed-activator stimulus set in which the probe stimulus consists of five clicks, interleaved with pulses of a broadband noise activator (see Keefe et al., [2010, 2017] for further detail). The click stimulus was a 100- $\mu$ s rectangular click. The activator was a 116-ms broadband noise stimulus containing equal levels for frequencies from 0 to 24 kHz. Both ipsilateral and contralateral reflexes were evoked in this study. In the ipsilateral condition, the equipment setup matched that used for DPOAE measurements, with the probe and activator stimuli sent to separate ER-3A insert headphones. In the contralateral condition, an additional ER-3A insert headphone was used to deliver the activator stimulus. MEMR were measured using custom designed software (Middle-ear muscle reflex [MEMR] version 1.0; Boys Town National Research Hospital, Omaha, NE) and a 24-bit soundcard (Hammerfall DSP Multi-Face II, RME, Germany). The first click was preceded by silence, resulting in a response that did not contain effects of MEMR. The click level was fixed at 95 dB pSPL, and activator level was varied from 45–90 dB SPL in 5-dB steps. Responses to the consecutive four clicks included effects of MEMR because these clicks were preceded by reflex activators. There was evidence of MEMR effects for responses to all clicks that were preceded by an activator. To simplify the analysis of reliability, only the first and last clicks were used to measure the MEMR. MEMR effect was calculated as the ratio of the peak response pressure magnitude (dB Pa) to the final click to the peak response pressure magnitude to the first click (baseline).

**Tone-burst ABR.**—Tone-burst-elicited ABR (tbABR) waveforms were collected using custom-designed software (Cochlear Response [CResp] version 1.0; Boys Town National Research Hospital, Omaha, NE) running on a computer equipped with a 24-bit soundcard (Babyface; RME, Germany). Two-channel electroencephalographic (EEG) responses were acquired simultaneously using surface electrodes placed at the high forehead (Fpz, ground), contralateral mastoid (inverting reference), and two noninverting active electrodes placed at the vertex (Cz) and in the ear canal (ER3–26A gold foil tiptrode). Stimulus and recording parameters can be found in Table 1. Stimuli were presented monaurally to an ER-3A insert earphone (Etymotic Research, Elk Grove, IL) connected to the soundcard. The stimulus level used in the present study was 103 dB SPL for 4 kHz and 110 dB SPL for 1.5 kHz, because these were the levels at which a response could be observed in most participants in the previous study (Ridley et al., 2018). Electrode impedances were  $\leq 5$  k $\Omega$  in all cases. The EEG signal was amplified (gain = 100,000) and filtered (0.1 to 1.5 kHz; Opti-Amp 8001; Intelligent Hearing Systems, Miami, FL) and directed to the computer via the soundcard for averaging. Responses were separated by even and odd recordings and stored in two buffers. Comparison of responses from the two buffers was used as a visual test of intra-session reliability. The average of the two buffers were used as the estimate of the signal and the difference between the two buffers as the estimate of the noise. Artifact rejection was based on the peak absolute differences between the buffers. Two electrode montages were used to highlight the two waves of interest. The tiptrode montage resulted in more observations and

larger amplitudes of wave I compared to the vertex montage (56 of 68 waveforms compared to 30 of 68). Wave V was observed equally on the two montages (59 of 68 for the tiptrode and 61 of 68 for the vertex montage). However, the vertex montage resulted in larger amplitudes of wave V (see Fig.1). Thus, wave I amplitudes were calculated from the tiptrode waveforms and wave V from the vertex waveforms.

Two examiners independently identified peaks and troughs of waves I and V and determined the amplitude and latency of the peaks and troughs. The software allowed for resolution of 0.02  $\mu\text{V}$  for amplitude and 0.02 ms for latency. Amplitude was calculated as the difference between the positive peak and the following trough. The processing delay of the soundcard was taken into account when analyzing the data for latency. Peak and trough latency were not used in the reliability analysis but rather to clarify disagreements between examiners. Disagreements occurred in less than 10% of the total 268 waveforms, and were resolved by a third examiner. The wave I amplitude in the tiptrode montage and the wave V amplitude of the vertex montage were used in the reliability analysis.

**Speech ABR.**—The same hardware and software used for the tbABR measurements was also used to record the ABR to a speech signal (sABR). A 40-ms synthetic /da/ was chosen as the stimulus because it has been used extensively in complex ABR research (see Skoe & Kraus [2010]). The stimulus used in the present study was developed by the Auditory Neuroscience Laboratory at Northwestern as part of their Brainstem toolbox. Stimulus and recording parameters can be found in Table 1. The processing delay of the soundcard was also taken into account when analyzing the data. The EEG was band-pass filtered at cutoff frequencies of 0.1–3 kHz. Three metrics were derived from the sABR for test-retest reliability: 1) stimulus-to-response correlations, 2) magnitude-squared coherence of the stimulus and response, 3) and discrete peak measures. Stimulus-to-response correlation assesses how well the frequency-following portion of the response represents the periodicity of the steady-state portion of the vowel. This analysis was performed over the 11–40 ms portion of the response. The reliability of the stimulus-to-response correlation was assessed between visits. Discrete peaks for waves V, A, C, D, E, F, and O were identified by two examiners (Banai et al., 2007). The amplitudes of the peaks, referenced to the 0 V baseline, as well as latencies, were quantified. Latency was used to assess disagreement between examiners. Disagreements occurred in less than 5% of the total 238 waveforms and were resolved by a third examiner. Occasionally, a wave could not be found and was therefore excluded from the data. An example waveform highlighting these peaks is shown in Fig. 2. Peaks V, A, C, and O encode transient events in the stimulus and peaks D, E, and F encode the periodicity of the vowel (Russo et al., 2004). Amplitude of each peak ( $\mu\text{V}$ ) were used for the reliability analysis.

### Reliability Analysis

The hearing measures included in this study are intended for use in a predictive model of suprathreshold hearing deficits. In this model, the variance of each measure that can be explained by thresholds in quiet is removed, such that each measure only represents what cannot be accounted for by hearing sensitivity alone, i.e. threshold-independent hearing disorder. To accomplish this, each measure is regressed against both TIQ at 1.5 and 4 kHz in



a multivariate linear model. All measures showed a linear relationship with TIQ, with the exception of weak relationships with speech ABR wave amplitudes. Robust regression with a bi-square weighting function was used to account for heteroscedasticity in the measures. Intraclass correlation coefficients (ICC) between measures for the first and second visits were used to examine the test-retest reliability of the residuals of the behavioral and physiologic measures. A two-way mixed effects absolute agreement model was chosen to account for repeated measurements (Portney & Watkins, 2000; Koo & Li, 2016; Liu et al., 2016). The ICC was calculated using the following formula:

$$\frac{MS_R - MS_E}{MS_R + (k - 1)MS_E + \frac{k}{n}(MS_C - MS_E)}$$

where  $MS_R$  = mean square for observations,  $MS_E$  = mean square for error,  $MS_C$  = mean square for visit,  $n$  = number of participants, and  $k$  = number of observations. Small between-subject variance can cause ICC to take on a negative value, even if there is small within-subject variance. Negative ICC values are theoretically impossible (Giraudeau, 1996), therefore, negative values were changed to zero and the only interpretation was that the measure was not reliable. Standardization of the residual measures is also important for interpretability for their future use in a predictive model, therefore residual measures in this study were standardized as well, though this does not affect reliability whatsoever. The strength of the reliability of each measure was categorized (Koo & Li, 2016), as indicated in Table 2.

## RESULTS

The test-retest reliability of the residual measures was based on the residuals post multivariate linear regression with TIQ at 1.5 and 4 kHz. These residuals represent the portion of each suprathreshold measure that cannot be explained by a participant's TIQ and therefore may be used as a proxy for hidden hearing loss. The test-retest reliability of the residual measures are reported here, and are summarized in Table 3. The test-retest reliability of the raw measures are reported in a table in the Supplemental Material (Supplemental Digital Content 1).

### Behavioral measures.

The test-retest reliability of TIN residuals was moderate with ICC (95% CI) = 0.68 (0.30–0.87) and 0.52 (0.05–0.80) for 1.5 and 4 kHz, respectively (Fig. 3). Frequency modulation detection threshold residuals showed good reliability with ICC = 0.79 (0.50–0.92; Fig. 4). Word recognition scores showed mostly poor or moderate test-retest reliability (Fig. 5). The residual of word recognition in quiet had moderate test-retest reliability with ICC = 0.63 (0.24–0.85). This is likely due to ceiling effects as described in the Discussion section. The residuals of word recognition scores in the presence of noise also showed moderate test-retest reliability (ICC = 0.55 [0.11–0.81]). Time-compressed word recognition residuals showed poor reliability (ICC = 0.39 [0–0.72]), but interestingly, residuals of time-compressed words with reverberation showed moderate test-retest reliability (ICC = 0.57 [0.14–0.82]).

### Physiologic measures.

The test-retest reliability of the residuals of the physiologic measurements were generally better compare to the reliability of behavioral measurements. Test-retest reliability of DPOAE residual was good (ICC = 0.78 [0.48–0.91]) at 1.5 kHz (Fig. 6 left) and excellent (ICC = 0.91 [0.76–0.97]) at 4 kHz (Fig. 6 right). The reliability of the MEMR residuals depended on activator level, revealing a growth in reliability that saturates at moderate-to-high levels (Fig 7). At the highest activator level (90 dB SPL), reliability was good for both the ipsilateral (ICC = 0.83 [0.60–0.94]; Fig. 8 left) and contralateral probe configurations (ICC = 0.87 [0.67–0.97]; Fig. 8 right). The tABR results were mixed with wave I residuals having a higher test-retest reliability than wave V residuals for both 1.5 and 4 kHz (Fig. 9). Wave I residuals showed moderate reliability (ICC = 0.74 [0.25–0.93] and 0.68 [0.19–0.90] for 1.5 and 4 kHz, respectively). Wave V residuals resulted in poor test-retest reliability with ICC = 0.47 (0–0.81) for 1.5 kHz and was negligible (ICC = 0 [0–0.15]) for 4 kHz.

Three metrics of the sABR were analyzed for test-retest reliability: 1) stimulus-to-response correlations, 2) magnitude-squared coherence of the stimulus and response, and 3) discrete peak measures. Stimulus-to-response correlation (STR) and magnitude-squared coherence (MSC) for the two electrode montages are shown in Table 4. Variability was smaller for the vertex montage and more reliable across visits than the tiprode recording montage. The residuals of the discrete wave amplitudes were also assessed and displayed in Table 3 and Figure 10. None of the residual wave amplitudes of the sABR were reliable. Waves that encode the transient events in the stimulus (V, A, C, & O) had ICC = 0.29 and waves encoding the vowel (D, E, & F) had ICC = 0.19. Some waves did not appear often, like wave C, where examiners only found two of the seventeen participants where the wave was evident in both visits. Wave C was excluded from reliability analysis because there were not enough data points.

## DISCUSSION

The goal of the present study was to evaluate the test-retest reliability of a number of behavioral and physiologic measures that may be used to diagnose hidden hearing loss. Residuals of the measures were obtained by regressing the raw data against the participant's TIQ at 1.5 and 4 kHz using multivariate linear regression analysis. Our working hypothesis is that these residuals serve as a proxy for subclinical –or hidden –hearing loss, as they represent the portion of the suprathreshold measure that cannot be explained by the participant's threshold in quiet. Some protocols and parameters used to obtain the responses in this study differ from those typically used in the clinic and therefore establishment of the test-retest reliability for the use of these measures in future research is important to draw appropriate conclusions.

### Behavioral measures

Several residual behavioral measures of hearing were assessed for reliability, including TIN, FMDT, and several word recognition tasks. TIN was a moderately reliable measure. Residualizing to TIQ did not have an effect on the reliability, with the exception of improving reliability at 4 kHz from poor to moderate (see table and figure, Supplemental

Digital Content 1 & 2, respectively). FMDT showed good reliability between visits. Residualizing with TIQ had no effect on the reliability (see table and figure, Supplemental Digital Content 1 & 3, respectively).

In understanding speech, humans use cues from both the speech envelope and temporal fine structure. The ability to efficiently code temporal fine structure may be related to performance in understanding speech in noise (Strelcyk & Dau, 2009): detection of frequency-modulation of carriers <5 kHz relies on the temporal, or phase-locking, ability of the auditory nerve (Moore & Sek, 1996), and would theoretically be affected by cochlear synaptopathy (Bharadwaj et al., 2014).

Word recognition scores using the NU-6 50-word lists were predicted to be fairly unreliable based on the literature (Thornton & Raffin, 1978; Grimes et al., 1984; Versfeld & Dreschler, 2002). The ICC for the residual scores of word recognition in quiet was weak; however, a closer look at the data (Fig. 5) reveals that ICC alone may not paint a good picture of the true reliability. In this study, all participants had normal hearing and therefore performed near or at ceiling on the word recognition task in a quiet background (see figure, Supplemental Digital Content 4). Since values are clustered near ceiling, between-subject variability is low, which will inevitably lead to a low ICC, even if measurement error is limited. A study that included participants with a large range of audiometric thresholds found word recognition test-retest reliability with Pearson correlation ranging from 0.92 to 0.96 (Causey & Hermanson, 1983). Therefore, word recognition test-retest reliability could be different with a broader hearing-threshold sample anticipating a larger range of word recognition scores. The reliability of word recognition in the three more difficult listening conditions: noise, time-compression, and time-compression plus reverberation resulted in larger score variability across normal-hearing participants. Reliability for words in noise and time-compressed words with reverberation was good, but the probability of an observed difference in two scores occurring by chance depends on the score itself. Thornton & Raffin (1978) determined critical differences for 50-word lists based on the confidence intervals of the true scores (Table 4 of Thornton & Raffin, 1978). When the second-visit word recognition scores for each participant were compared to the confidence intervals of their first visit scores, there were only 4 out of 68 scores in all conditions that were critically different between visits (their Fig. 5, filled squares). A poorer-than-expected score may be due to the unfamiliarity of the task itself. A solution to this may be to shorten session times, have a practice word list, or always start with the quiet condition (to provide training) and randomize the other three conditions. The test-retest reliability for time-compressed words was poor.

### Physiologic measures

The reliability of the residual DPOAE measures was moderate and very high for 1.5 and 4 kHz, respectively. This is commensurate with the literature which has shown that DPOAEs are a reliable measure (e.g. Franklin, 1992). Residualizing with TIQ did not affect the reliability (see table and figure, Supplemental Digital Content 1 & 5, respectively).

The test-retest reliability of the raw MEMR was dependent on activator level, with a trend of low reliability at low activator levels (45–50 dB SPL), increasing to moderate reliability at

moderate levels (50–65 dB SPL) and then to high reliability at moderate-to-high levels (70–90 dB SPL). Contralateral and ipsilateral activators had similar reliability. The poorer reliability of low-level MEMR cautions against the use of thresholds when measuring MEMR. The MEMR has been proposed to be a marker for auditory nerve degeneration, especially loss of low spontaneous-rate fibers (Valero et al., 2016; Wojtczak et al., 2017; Chertoff et al., 2018). Current clinical measurements of the acoustic reflex in humans only determine threshold of elicitation and decay for several frequencies. Other measures of the acoustic reflex may be necessary to capture features of hidden hearing loss at the level of the auditory nerve. Reflex strength has been shown to correlate with noise-induced tinnitus in humans (Wojtczak et al., 2017) and is reduced by noise-exposure in mice (Valero et al., 2016). Residualizing with TIQ did not have an effect on reliability (see table and figure, Supplemental Digital Material 1 & 6).

Features of both the tone-burst-elicited ABR and speech-evoked ABR were examined for reliability. The residuals of the amplitudes of waves I and wave V of the tbABR showed varied test-retest reliability. Residual wave I amplitude was highly reliable at 1.5 kHz and moderately reliable at 4 kHz. High reliability has been reported for wave I amplitude measured from the ear canal using click-stimuli (Mori et al., 1981) but may be different for tonal stimuli. Additionally, stimulus presentation rate can affect the visualization of wave I. In the present study, a rate of 11/sec was used, but a more optimal rate would be 7.7/sec (Butler & Stuart, 2017). In contrast, wave V had strong reliability at 1.5 kHz but poor reliability at 4 kHz. Residualizing ABR waves with TIQ had an interesting effect on reliability. For wave I at 1.5 kHz, reliability improved after residualizing with  $ICC = 0.38$  for the raw amplitude and  $ICC = 0.74$  for residual wave I amplitude. This was also true to a smaller extent for wave V at both frequencies. Wave I at 4 kHz was unaffected by residualizing. See Supplemental Digital Content 1 & 7 for ICC and a figure of the raw amplitudes.

The scalp-recorded ABR to complex stimuli, such as speech, allow us to examine more complex mechanisms of the auditory system. While the speech-evoked ABR has been extensively studied in the past decade (see reviews by Chandrasekaran & Kraus, 2006 and Skoe & Kraus, 2010), how best to analyze responses is still under consideration. In the present study we assessed the reliability of several metrics derived from the sABR. Since the tbABR was recorded simultaneously using two electrode montages, the sABR, which was recorded immediately following the tbABR, also recorded in these two montages. Recording from the ear canal is not typical for sABR, because the response occurs more centrally, but the results are reported nonetheless.

The sABR waveforms for each visit were compared for each participant. The median  $r$ -values showed a moderate test-retest reliability for the vertex recording and a poor reliability for the ear canal recording. In addition to the inter-response correlations, the stimulus-to-response correlations showed poor reliability for both the vertex and ear canal recordings. While the STR correlations may not be reliable within participants across visits, the median and interquartile range of the STR correlations (Table. 4) align well with previously published data (Song et al., 2011). We also explored the magnitude-squared coherence as another metric for assessing of the relationship between the stimulus and response (Table.

4). Interestingly coherence was higher for the ear canal recordings than the vertex recordings. It is possible that EEG responses from additional cortical functions, which can be considered noise in the response signal, contribute to the vertex recording more than the ear canal recording.

To produce a measure that would allow for the calculation of the residual for the speech ABR, as was done for other suprathreshold measures, the peak amplitudes were converted to residual peak amplitudes. The peaks chosen for the analysis include responses to both the transient and periodic portions of the stimulus. Data from Song and colleagues (2011), using the same 40 ms /da/ stimulus used here, showed high reliability for the peak latencies and amplitudes. The residual peak amplitudes of the sABR in the present study showed remarkably poor reliability which was not related to residualizing to TIQ. The reliability of the raw peak amplitudes (see figure, Supplemental Digital Content 8) were higher but not notably so. The poor reliability of E and F, which encode the periodicity of the vowel, may be due to the short duration steady-state segment of the 40-ms /da/. Several studies used a longer 170 ms /da/ stimulus. It is probable that a longer vowel production could produce better periodic responses to a longer stimulus and may be a more appropriate stimulus to measure the sABR.

Overall, there was a wide range of test-retest reliability in the measures. Some poor test-retest reliability was expected in some of the measures such as word recognition, however, reliability for some measures were unexpectedly poor such as wave V of the tABR. Conversion of the raw data to residuals of TIQ did not have an effect on interpretation of the reliability, with the exception of the sABR wave amplitudes. Test-retest reliability for the raw sABR amplitude was much higher than residual sABR amplitudes. This may be due to the lack of a linear relationship between TIQ and sABR wave amplitudes. Calculating residuals forces a regression between TIQ and wave amplitude where a relationship may not exist. If a poorly fit regression line changes between visits 1 and 2, the residual wave amplitudes will be poorly related between visits. Therefore, raw sABR wave amplitudes may be a better measure compared to the residual sABR measure.

### Limitations of the study

One of the major limitations of this study was the sample size. With a low number of subjects, the 95% confidence intervals for most of the measures are wide and reduce the interpretability of the ICC. With the exception of DPOAEs at 4 kHz and MEMR at 90 dB SPL, the lower bounds of the confidence intervals extend into the range of poor reliability. The number of subjects was limited due to the amount of testing required to generate such a large dataset of hearing measures (eight hours per participant).

Another limitation of using residual measures is the reliability of thresholds in quiet, on which all residual measures depend. Thresholds in quiet were used as the measure with which all other measure's residuals were calculated. Conclusions based on test-retest reliability of the residuals, therefore, requires test-retest reliability of TIQ to be high. Indeed, the test-retest reliability of TIQ was excellent for 4 kHz (ICC = 0.93) and good for 1.5 kHz (ICC = 0.88; see Supplemental Material 1 & 9). While a threshold obtained via standard audiometry can vary as much as 10 dB day-to-day (Atherley & Dingwall-Fordyce, 1963),

the TIQ measures obtained in the present study were much lower with a median (interquartile range) difference in threshold between visits of 1 (2.67) dB for 1.5 kHz and 1.33 (2.08) dB for 4 kHz. Nonetheless, there is a trade-off in using TIQ as opposed to audiometric thresholds. Obtaining psychometric thresholds using the 3AFC procedure requires long test times, which limited the number of different frequencies we could test. The choice of 1.5 and 4 kHz were made to capture the range of speech frequencies, and for future experiments, potential noise-induced “notches” in hearing sensitivity. Regardless of test time, we suggest using more than three reversals to calculate threshold to improve reliability. Additionally, it is possible that analysis utilizing TIQ at 1.5 and 4 kHz does not adequately remove the influence of high-frequency thresholds from suprathreshold measures. For example, Don and Eggermont, 1978 demonstrated that even frequencies greater than 8 kHz contributed to ABR responses at lower frequencies. However, this possibility is unlikely or only had an infinitesimal influence, as the reliability of the raw measures was comparable to the residual measures (see Supplemental Table).

The residuals of TIN were moderately reliable. In future studies, the reliability of this measure could be improved by imposing a set of rules for acceptance and exclusion of a trial run such as a maximum standard deviation within a run and maximum allowable dB difference between runs. These psychoacoustic tasks are tedious and long, thus attention and fatigue may have played a role in the reliability of this measure. However, FMDT was also performed under the same conditions and the test-retest reliability was higher than TIN.

## CONCLUSIONS

Residual measures that showed good test-retest reliability were FMDT, both ipsilateral and contralateral MEMR at high activator levels, and DPOAEs at 1.5 kHz, with DPOAEs at 4 kHz showing excellent reliability. Residual measures that were moderately reliable across visits were TIN, tbABR wave I, and word recognition in quiet, noise, time-compressed speech with reverberation. Poor reliability was found for wave V of the tbABR, time-compressed speech, and waves of the sABR. The results of this study suggest that care should be taken when selecting proxy measures for synaptopathy or measures of suprathreshold deficits in human.

## Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

## ACKNOWLEDGMENTS

This research was supported by the National Institutes of Health / National Institute on Deafness and Other Communication Disorders grants R01 DC016348 and P30 DC004662. We would like to thank Carissa Allen and Tessa Culbertson for their help with data collection and scoring of auditory brainstem response waveforms.

Financial Disclosures/Conflicts of Interest:

This research was funded by NIH 5R01DC016348-02

## REFERENCES

- Akhoun I, Gallégo S, Moulin A, Ménard M, Veuillet E, Berger-Vachon C, ... Thai-Van H (2008). The temporal relationship between speech auditory brainstem responses and the acoustic pattern of the phoneme /ba/ in normal-hearing adults. *Clinical Neurophysiology*, 119, 922–933. 10.1016/j.clinph.2007.12.010 [PubMed: 18291717]
- Alvord LS (1983). Cochlear dysfunction in “normal-hearing” patients with history of noise exposure. *Ear and Hearing*, 4(5), 247–250. 10.1097/00003446-198309000-00005 [PubMed: 6628849]
- Anderson S, Parbery-Clark A, White-Schwoch T, & Kraus N (2013). Auditory Brainstem Response to Complex Sounds Predicts Self- Reported Speech-in-Noise Performance. *Journal of Speech, Language and Hearing Research*, 56(1), 31–43. 10.1044/1092-4388(2012/12-0043).Auditory
- Atherley GRC, & Dingwall-Fordyce I (1963). the Reliability of Repeated Auditory Threshold Determination. *British Journal of Industrial Medicine*, 20(3), 231–235. 10.1136/oem.20.3.231 [PubMed: 14046161]
- Banai K, Abrams D, & Kraus N (2007). Sensory-based learning disability: Insights from brainstem processing of speech sounds. *International Journal of Audiology*, 46(9), 524–532. 10.1080/14992020701383035 [PubMed: 17828668]
- Bharadwaj HM, Masud S, Mehraei G, Verhulst S, & Shinn-Cunningham BG (2015). Individual Differences Reveal Correlates of Hidden Hearing Deficits. *Journal of Neuroscience*, 35(5), 2161–2172. 10.1523/JNEUROSCI.3915-14.2015 [PubMed: 25653371]
- Bramhall NF, Konrad-Martin D, McMillan GP, & Griest SE (2017). Auditory Brainstem Response Altered in Humans With Noise Exposure Despite Normal Outer Hair Cell Function, 1–12.
- Butler A, & Stuart A (2017). The Effect of Test, Electrode, and Rate on Electrocochleography Measures. *Journal of the American Academy of Audiology*, 13, 1–13. 10.3766/jaaa.17081
- Byrne D, Dillon H, Tran K, Wilbraham K, Cox R, Hagerman B, ... Ludvigsen C (1994). An international comparison of long-term average speech spectra. *J. Acoust. Soc. Am*, 96(4). Retrieved from <http://acousticalsociety.org/content/terms>.
- Carhart R, & Jerger J (1959). Preferred Method For Clinical Determination Of Pure-Tone Thresholds. *Journal of Speech and Hearing Disorders*, 24(4), 330 10.1044/jshd.2404.330
- Causey G, & Hermanson C (1983). A Comparative Evaluation of the Maryland NU 6 Auditory Test. *Journal of Speech and ...*, 48(2 1983), 62–69. 10.1044/jshd.4801.62
- Chandrasekaran B, & Kraus N (2006). The scalp-recorded brainstem response to speech: Neural origins and plasticity. *Psychophysiology*, 47(2), 236–246. 10.1111/j.1469-8986.2009.00928.x
- Chertoff ME, Martz A, Sakumura JT, Kamerer AM, & Diaz FJ (2018). The Middle Ear Muscle Reflex in Rat. *Ear and Hearing*, 39(3), 605–614. 10.1097/AUD.0000000000000520 [PubMed: 29189520]
- El-Badry MM, & McFadden SL (2007). Electrophysiological Correlates of Progressive Sensorineural Pathology in Carboplatin-Treated Chinchillas. *Brain Research*, 1134(1), 122–130. 10.1038/jid.2014.371 [PubMed: 17198689]
- Feeney MP, Keefe DH, Hunter LL, Fitzpatrick DF, Garinis AC, Putterman DB, & McMillan GP (2017). Normative wideband reflectance, equivalent admittance at the tympanic membrane, and acoustic stapedius reflex threshold in adults. *Ear and Hearing*, 38(3), e142–e160. 10.1097/AUD.0000000000000399 [PubMed: 28045835]
- Fernandez KA, Jeffers PWC, Lall K, Liberman MC, & Kujawa SG (2015). Aging after Noise Exposure: Acceleration of Cochlear Synaptopathy in “Recovered” Ears. *Journal of Neuroscience*, 35(19), 7509–7520. 10.1523/JNEUROSCI.5138-14.2015 [PubMed: 25972177]
- Franklin DJ (1992). Test retest reliability of distortion-product and transiently evoked otoacoustic emissions. *Ear & Hearing*, 13(6), 417–429. [PubMed: 1487104]
- Fulbright ANC, Le Prell CG, Griffiths SK, & Lobarinas E (2017). Effects of Recreational Noise on Threshold and Suprathreshold Measures of Auditory Function. *Seminars in Hearing*. 10.1055/s-0037-1606325
- Furman AC, Kujawa SG, & Liberman MC (2013). Noise-induced cochlear neuropathy is selective for fibers with low spontaneous rates. *Journal of Neurophysiology*, 110(3), 577–586. 10.1152/jn.00164.2013 [PubMed: 23596328]

- Davis Gatehouse. (1992). Clinical Pure-tone versus three-interval forced choice thresholds. *Audiology*, 31, 31–44. [PubMed: 1554331]
- Gelfand SA (1982). Hearing: An Introduction to Psychological and Physiological Acoustics. *Journal of Neurology, Neurosurgery & Psychiatry* (Vol. 45). 10.1136/jnnp.45.12.1175-b
- Gelfand SA, Schwander T, & Silman S (1990). Acoustic reflex thresholds in normal and cochlear-impaired ears: Effects of no-response rates on 90th percentiles in a large sample. *Journal of Speech and Hearing Disorders*, 55(2), 198–205. <https://doi.org/doi:10.1044/jshd.5502.198> [PubMed: 2329784]
- Giraudeau B (1996). Negative values of the intraclass correlation coefficient are not theoretically possible. *Journal of Clinical Epidemiology*, 49(10), 1205–1206. 10.1016/0895-4356(96)00053-4 [PubMed: 8827004]
- Grimes A, Mueller HG, & Williams D (1984). Clinical Considerations in the Use of Time-Compressed Speech. *Ear & Hearing*, 5(2), 114–117. Retrieved from <http://ovidsp.ovid.com/ovidweb.cgi?T=JS&PAGE=reference&D=emed3&NEWS=N&AN=1993336529> [PubMed: 6724171]
- Grose JH, Buss E, & Hall JW (2017). Loud Music Exposure and Cochlear Synaptopathy in Young Adults: Isolated Auditory Brainstem Response Effects but No Perceptual Consequences. *Trends in Hearing*, 21, 233121651773741. 10.1177/2331216517737417
- Guest H, Munro KJ, Prendergast G, Millman RE, & Plack CJ (2018). Impaired speech perception in noise with a normal audiogram: No evidence for cochlear synaptopathy and no relation to lifetime noise exposure. *Hearing Research*, 364, 142–151. 10.1016/j.heares.2018.03.008 [PubMed: 29680183]
- Guest H, Munro K, & Plack CJ (2018). Relations between speech perception in noise, high-frequency audiometry, and physiological measures of cochlear synaptopathy. *The Journal of the Acoustical Society of America*, 144(3), 1935–1935. 10.1121/1.5068467
- Johannesen PT, Pérez-González P, Kalluri S, Blanco JL, & Lopez-Poveda EA (2015). Predictors of supra-threshold speech-in-noise intelligibility by hearing-impaired listeners. *Proceedings of ISAAR 2015: Individual Hearing Loss – Characterization, Modelling, Compensation Strategies*. 5th Symposium on Auditory and Audiological Research, (8).
- Johannesen PT, Pérez-González P, Kalluri S, Blanco JL, & Lopez-Poveda EA (2016). The Influence of Cochlear Mechanical Dysfunction, Temporal Processing Deficits, and Age on the Intelligibility of Audible Speech in Noise for Hearing-Impaired Listeners. *Trends in Hearing*, 20, 1–14. 10.1177/2331216516641055
- Johnson KL, Nicol TG, & Kraus N (2005). Brain Stem Response to Speech: A Biological Marker of Auditory Processing. *Ear & Hearing*, 26, 424–434. Retrieved from <https://pdfs.semanticscholar.org/b4b3/3fd91b08a70077bea2878e4962aeb485345d.pdf> [PubMed: 16230893]
- Keefe DH, Feeney MP, Hunter LL, & Fitzpatrick DF (2017). Aural Acoustic Stapedius-Muscle Reflex Threshold Procedures to Test Human Infants and Adults. *JARO - Journal of the Association for Research in Otolaryngology*, 18(1), 65–88. 10.1007/s10162-016-0599-z [PubMed: 27957612]
- Keefe DH, Fitzpatrick DF, Liu Y, Sanford CA, Gorga MP, Town B, & Street N (2010). Wideband acoustic reflex test in a test battery to predict middle-ear dysfunction. *Hearing Research*, 263(0), 52–65. 10.1016/j.heares.2009.09.008.Wideband [PubMed: 19772907]
- Kobler JBB, Guinan JJ, Vacher SRR, & Norris BEE (1992). Acoustic reflex frequency selectivity in single stapedius motoneurons of the cat. *Journal of Neurophysiology*, 68(3), 807–817. 10.1152/jn.1992.68.3.807 [PubMed: 1432049]
- Koo TK, & Li MY (2016). A Guideline of Selecting and Reporting Intraclass Correlation Coefficients for Reliability Research. *Journal of Chiropractic Medicine*, 15(2), 155–163. 10.1016/j.jcm.2016.02.012 [PubMed: 27330520]
- Kujawa SG, & Liberman MC (2006). Acceleration of age-related hearing loss by early noise exposure: evidence of a missed youth. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 26(7), 2115–2123. 10.1523/JNEUROSCI.4985-05.2006 [PubMed: 16481444]
- Kujawa SG, & Liberman MC (2009). Adding Insult to Injury: Cochlear Nerve Degeneration after “Temporary” Noise-Induced Hearing Loss. *Journal of Neuroscience*, 29(45), 14077–14085. 10.1523/JNEUROSCI.2845-09.2009 [PubMed: 19906956]



- Kujawa SG, & Liberman MC (2015). Synaptopathy in the noise-exposed and aging cochlea: Primary neural degeneration in acquired sensorineural hearing loss. *Hearing Research*, 330, 191–199. 10.1016/j.heares.2015.02.009 [PubMed: 25769437]
- Kummer P, Janssen T, Hulin P, & Arnold W (2000). Optimal L1–L2 primary tone level separation remains independent of test frequency in humans. *Hearing Research*, 146(1–2), 47–56. 10.1016/S0378-5955(00)00097-6 [PubMed: 10913883]
- Levitt H (1971). Transformed Up-Down Methods in Psychoacoustics. *The Journal of the Acoustical Society of America*, 49(2B), 467–477. 10.1121/1.1912375
- Liberman LD, & Liberman MC (2015). Dynamics of cochlear synaptopathy after acoustic overexposure. *JARO - Journal of the Association for Research in Otolaryngology*, 16(2), 205–219. 10.1007/s10162-015-0510-3 [PubMed: 25676132]
- Liberman MC (1978). Auditory-nerve response from cats raised in a low-noise chamber. *The Journal of the Acoustical Society of America*, 63(2), 442–455. 10.1121/1.381736 [PubMed: 670542]
- Liberman MC, Epstein MJ, Cleveland SS, Wang H, & Maison SF (2016). Toward a differential diagnosis of hidden hearing loss in humans. *PLoS ONE*, 11(9), 1–15. 10.1371/journal.pone.0162726
- Liberman MC, & Kiang NY (1984). Single-neuron labeling and chronic cochlear pathology. IV. Stereocilia damage and alterations in rate- and phase-level functions. *Hearing Research*, 16(1), 75–90. [PubMed: 6511674]
- Liberman MC, & Kujawa SG (2017). Cochlear synaptopathy in acquired sensorineural hearing loss: Manifestations and mechanisms. *Hearing Research*, 349, 138–147. 10.1016/j.heares.2017.01.003 [PubMed: 28087419]
- Lin HW, Furman AC, Kujawa SG, & Liberman MC (2011). Primary neural degeneration in the guinea pig cochlea after reversible noise-induced threshold shift. *JARO - Journal of the Association for Research in Otolaryngology*, 12(5), 605–616. 10.1007/s10162-011-0277-0 [PubMed: 21688060]
- Liu J, Tang W, Chen G, Lu Y, Feng C, & Tu XM (2016). Correlation and agreement: overview and clarification of competing concepts and measures. *Shanghai Archives of Psychiatry*, 28(2), 115–120. 10.11919/j.issn.1002-0829.216045 [PubMed: 27605869]
- Lobarinas E, Salvi RJ, & Ding D (2016). Selective Inner Hair Cell Dysfunction in Chinchillas Impairs Hearing-in-Noise in the Absence of Outer Hair Cell Loss. *JARO - Journal of the Association for Research in Otolaryngology*, 17(2), 89–101. 10.1007/s10162-015-0550-8 [PubMed: 26691159]
- Maison SF, Usubuchi H, & Liberman MC (2013). Efferent Feedback Minimizes Cochlear Neuropathy from Moderate Noise Exposure. *Journal of Neuroscience*, 33(13), 5542–5552. 10.1523/JNEUROSCI.5027-12.2013 [PubMed: 23536069]
- Marshall L, & Jesteadt W (1986). Comparison of pure-tone audibility thresholds obtained with audiological and two-interval forced-choice procedures. *Journal Of Speech And Hearing Research*, 29, 82–91. [PubMed: 3702383]
- Mehraei G, Hickox AE, Bharadwaj HM, Goldberg H, Verhulst S, Liberman MC, & Shinn-Cunningham BG (2016). Auditory Brainstem Response Latency in Noise as a Marker of Cochlear Synaptopathy. *The Journal of Neuroscience*, 36(13), 3755–3764. 10.1523/JNEUROSCI.4460-15.2016 [PubMed: 27030760]
- Moore BCJ, & Glasberg BR (1989). Mechanisms underlying the frequency discrimination of pulsed tones and the detection of frequency modulation Effects of carrier frequency, modulation rate, and modulation waveform on the detection of modulation and the discrimination of modulation type (a. Citation: *The Journal of the Acoustical Society of America*, 86, 2468 10.1121/1.411967
- Moore BCJ, & Sek A (1996). Detection of frequency modulation at low modulation rates: Evidence for a mechanism based on phase locking. *The Journal of the Acoustical Society of America*, 100(4), 2320–2331. 10.1121/1.417941 [PubMed: 8865639]
- Mori N, Matsunaga T, & Asai H (1981). Intertest reliability in non-invasive electrocochleography. *International Journal of Audiology*, 20(4), 290–299. 10.3109/00206098109072702
- Neely S, & Liu Z (1994). BOYS TOWN NATIONAL RESEARCH HOSPITAL EMAV: Otoacoustic Emission Averager. Retrieved from <http://audres.org/downloads/emavtm.pdf>
- Noffsinger D, Wilson RH, & Musiek FE (1994). Department of Veterans Affairs compact disc recording for auditory perceptual assessment: background and introduction. *Journal of the*

- American Academy of Audiology, 5(4), 231–235. Retrieved from <http://www.ncbi.nlm.nih.gov/pubmed/7949294> [PubMed: 7949294]
- Papakonstantinou A, Strelcyk O, & Dau T (2011). Relations between perceptual measures of temporal processing, auditory-evoked brainstem responses and speech intelligibility in noise. *Hearing Research*, 280(1–2), 30–37. 10.1016/J.HEARES.2011.02.005 [PubMed: 21354285]
- Plack CJ, Barker D, & Prendergast G (2014). Perceptual consequences of “hidden” hearing loss. *Trends in Hearing*, 18, 1–11. 10.1177/2331216514550621
- Portney LG, & Watkins MP (2000). *Foundations of clinical research: Applications to practice*. New Jersey: Prentice Hall.
- Prendergast G, Guest H, Munro KJ, Kluk K, Léger A, Hall DA, ... Plack CJ (2017). Effects of noise exposure on young adults with normal audiograms I: Electrophysiology. *Hearing Research*, 344, 68–81. 10.1016/j.heares.2016.10.028 [PubMed: 27816499]
- Ridley CL, Kopun JG, Neely ST, Gorga MP, & Rasetshwane DM (2018). Using Thresholds in Noise to Identify Hidden Hearing Loss in Humans. *Ear and Hearing*, 1–16. 10.1097/AUD.0000000000000543 [PubMed: 28863035]
- Russo N, Nicol TG, Musacchia G, & Kraus N (2004). Brainstem responses to speech syllables. *Clin Neurophysiol*, 115(9), 2021–2030. 10.1016/j.clinph.2004.04.003 [PubMed: 15294204]
- Ryugo DK, & Rouiller EM (1988). Central projections of intracellularly labeled auditory nerve fibers in cats: Morphometric correlations with physiological properties. *Journal of Comparative Neurology*, 271(1), 130–142. 10.1002/cne.902710113 [PubMed: 3385008]
- Schaette R, & McAlpine D (2011). Tinnitus with a Normal Audiogram: Physiological Evidence for Hidden Hearing Loss and Computational Model. *Journal of Neuroscience*, 31(38), 13452–13457. 10.1523/JNEUROSCI.2156-11.2011 [PubMed: 21940438]
- Sergeyenko Y, Lall K, Liberman MC, & Kujawa SG (2013). Age-related cochlear synaptopathy: an early-onset contributor to auditory functional decline. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 33(34), 13686–13694. 10.1523/JNEUROSCI.1783-13.2013 [PubMed: 23966690]
- Sinha SK, & Basavaraj V (2010). Speech Evoked Auditory Brainstem Responses: A New Tool to Study Brainstem Encoding of Speech Sounds. *Indian Journal of Otolaryngology and Head and Neck Surgery*, 62(4), 395–399. 10.1007/s12070-010-0100-y [PubMed: 22319700]
- Skoe E, & Kraus N (2010). Auditory Brain Stem Response to Complex Sounds: A Tutorial. *Ear & Hearing*, 31(3), 302–324. 10.1097/AUD.0b013e3181c1db272 [PubMed: 20084007]
- Song JH, Nicol TG, & Kraus N (2011). Test-retest reliability of the speech-evoked auditory brainstem response. *Clinical Neurophysiology*, 122(2), 346–355. 10.1016/j.clinph.2010.07.009 [PubMed: 20719558]
- Stamper GC, & Johnson TA (2015). Auditory function in normal-hearing, noise-exposed human ears HHS Public Access. *Ear Hear*, 36(2), 172–184. 10.1097/AUD.0000000000000107 [PubMed: 25350405]
- Strelcyk O, & Dau T (2009). Relations between frequency selectivity, temporal fine-structure processing, and speech reception in impaired hearing. *Journal of the Acoustical Society of America*, 125(5), 3328–3345. 10.1121/1.3097469 [PubMed: 19425674]
- Thornton AR, & Raffin MJM (1978). Speech-discrimination scores modeled as a binomial variable. *Journal Of Speech And Hearing Research*, 21, 507–518. [PubMed: 713519]
- Valderrama JT, Beach EF, Yeend I, Sharma M, Van Dun B, & Dillon H (2018). Effects of lifetime noise exposure on the middle-age human auditory brainstem response, tinnitus and speech-in-noise intelligibility. *Hearing Research*, 365, 36–48. 10.1016/j.heares.2018.06.003 [PubMed: 29913342]
- Valero MD, Hancock KE, & Liberman MC (2016). The Middle Ear Muscle Reflex in the Diagnosis of Cochlear Neuropathy. *Hearing Research*2(332), 29–38. 10.1016/j.heares.2015.11.005.
- Versfeld NJ, & Dreschler WA (2002). The relationship between the intelligibility of time-compressed speech and speech in noise in young and elderly listeners. *The Journal of the Acoustical Society of America*, 111(1), 401–408. 10.1121/1.1426376 [PubMed: 11831813]
- Wang J, Ding D, & Salvi RJ (2002). Functional reorganization in chinchilla inferior colliculus associated with chronic and acute cochlear damage. *Hearing Research*, 168(1–2), 238–249. [PubMed: 12117524]

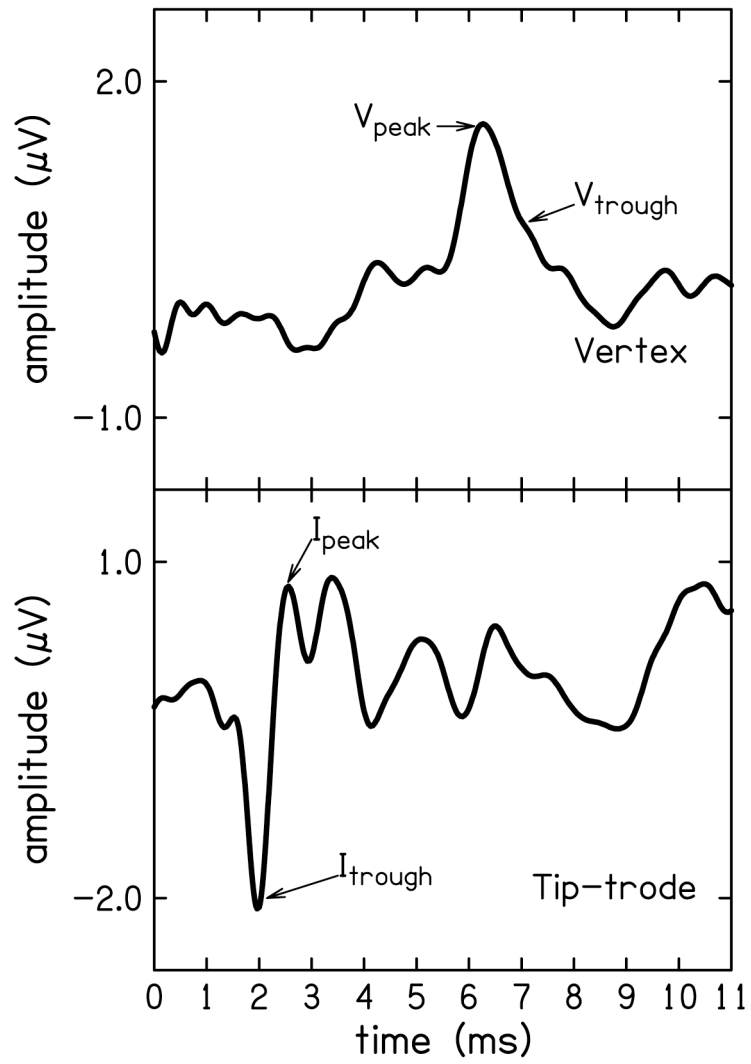
- Wojtczak M, Beim JA, & Oxenham AJ (2017). Weak Middle-Ear-Muscle Reflex in Humans with Noise-Induced Tinnitus and Normal Hearing May Reflect Cochlear Synaptopathy. *Eneuro*, 4(6), ENEURO.0363-17.2017. 10.1523/ENEURO.0363-17.2017
- Yost WA (1978). A forced-choice adaptive procedure for measuring auditory thresholds in children. *Behavior Research Methods & Instrumentation*, 10(5), 671-677. 10.3758/BF032f05369

Author Manuscript

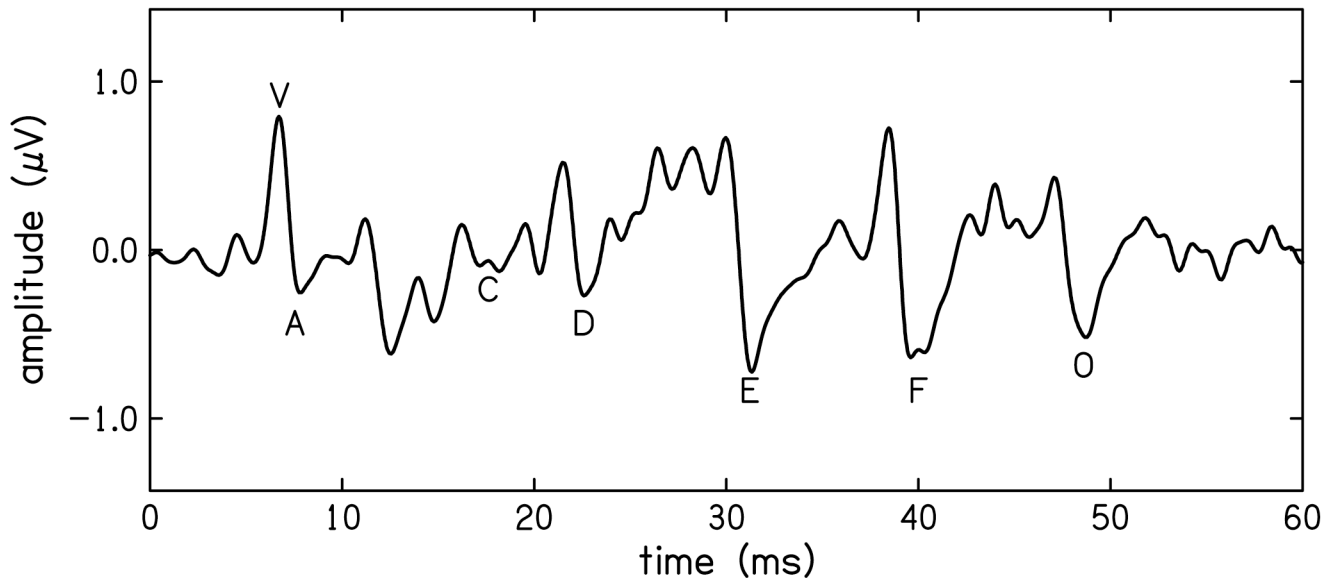
Author Manuscript

Author Manuscript

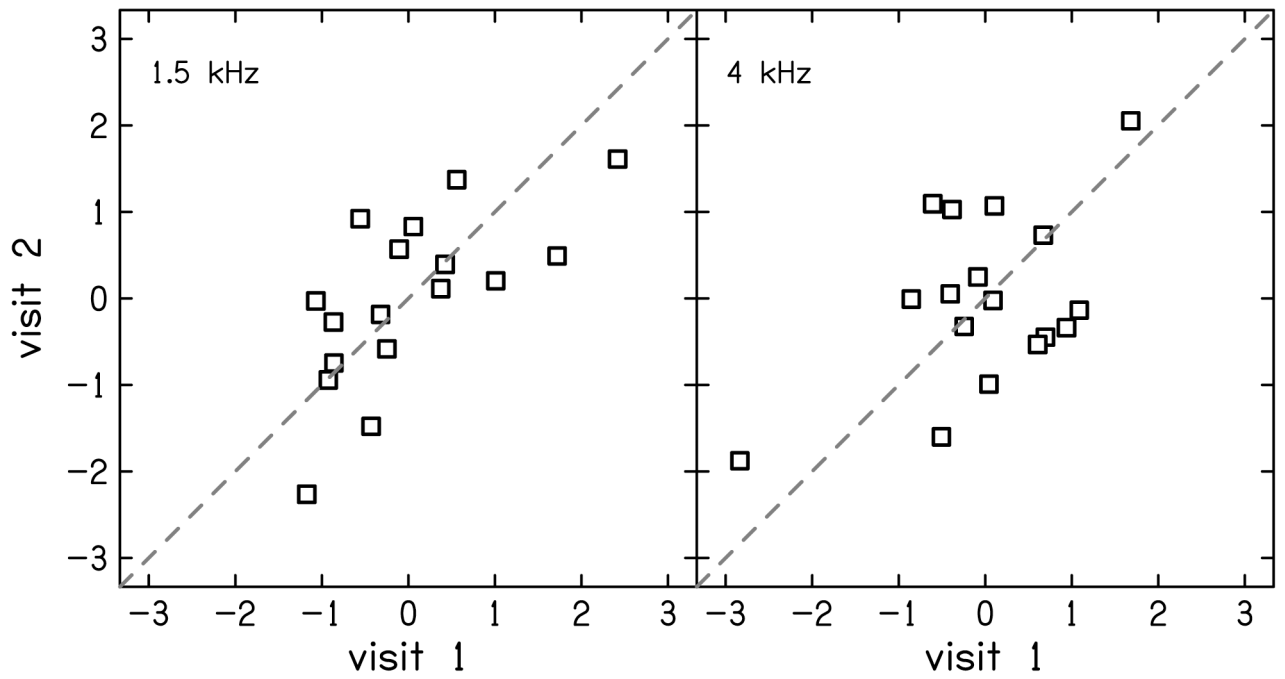
Author Manuscript



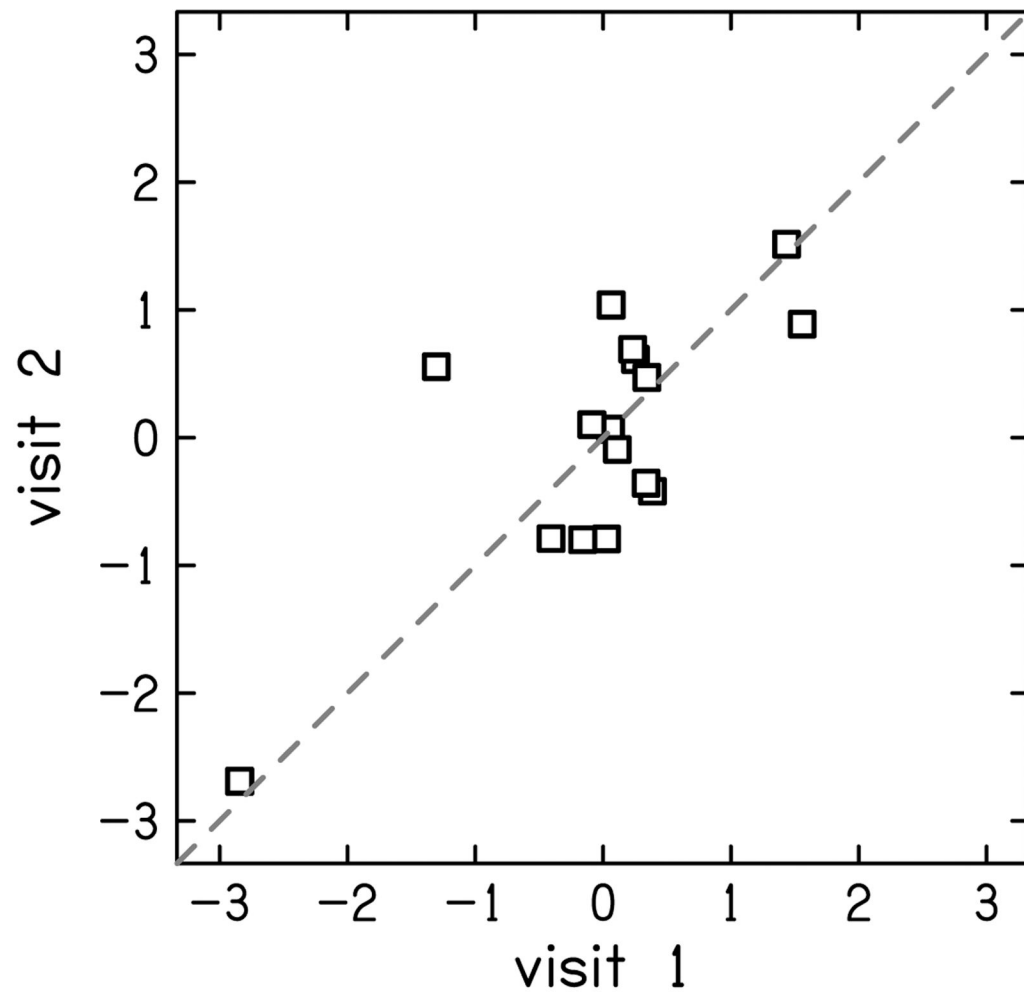
**Figure 1.** Exemplar tone-burst ABR waveforms from one participant for both electrode montages: Wave I amplitude was calculated from the tiptrode montage (top) as the absolute difference between the peak and the preceding trough. Wave V amplitude was calculated from the vertex montage (bottom) as the difference between the peak and the following trough.



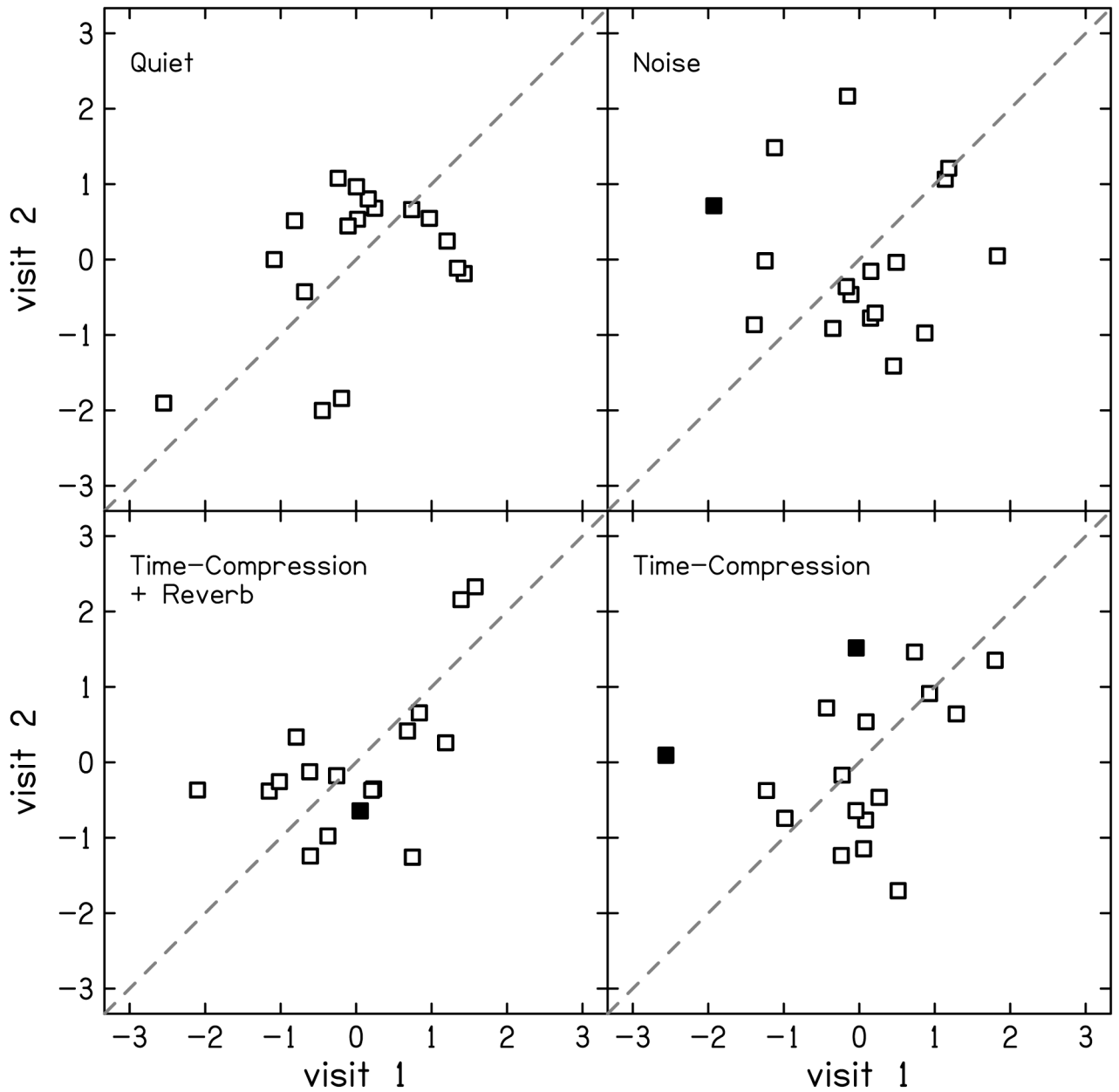
**Figure 2.**  
Exemplar speech-ABR waveform from one participant, where the discrete peaks (V, A, C, D, E, & F) are highlighted.



**Figure 3.** Correlation between visits 1 and 2 for z-scores of threshold in noise (TIN) residuals at 1.5 kHz (left) and 4 kHz (right). The dashed line indicates perfect reliability between visits.



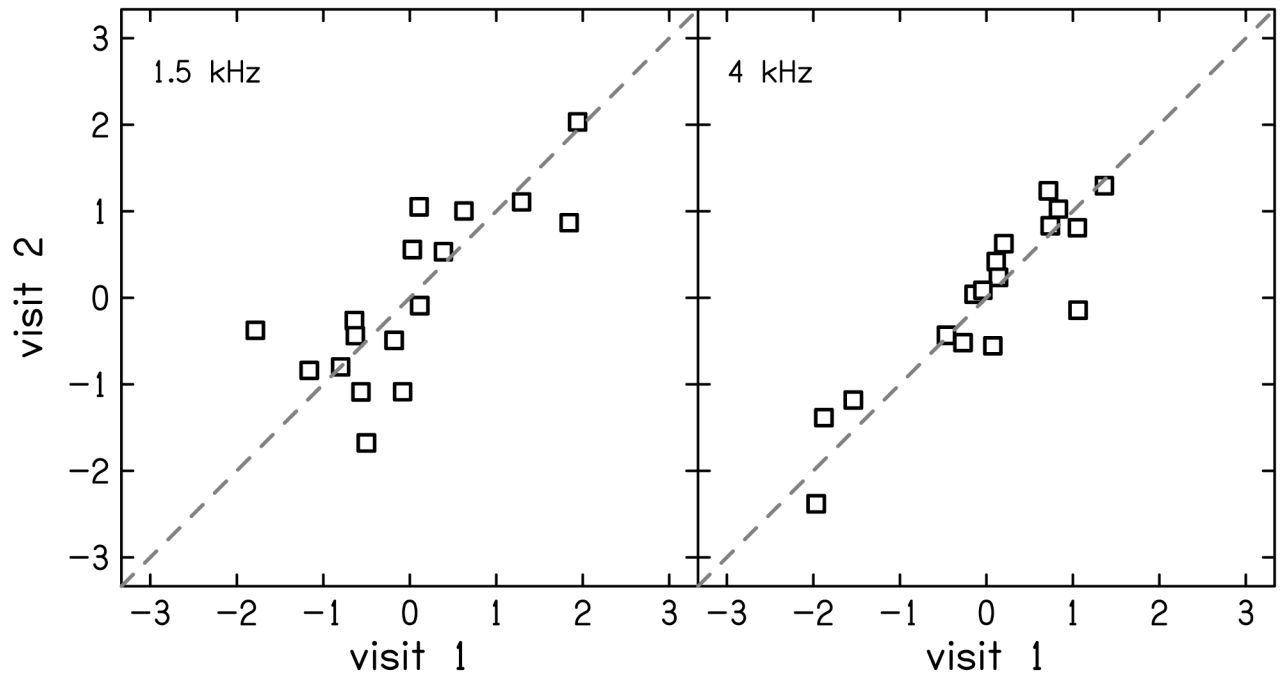
**Figure 4.** Correlation between visits 1 and 2 for z-scores of frequency-modulation detection threshold (FMDT) residuals. The dashed line indicates perfect reliability between visits.



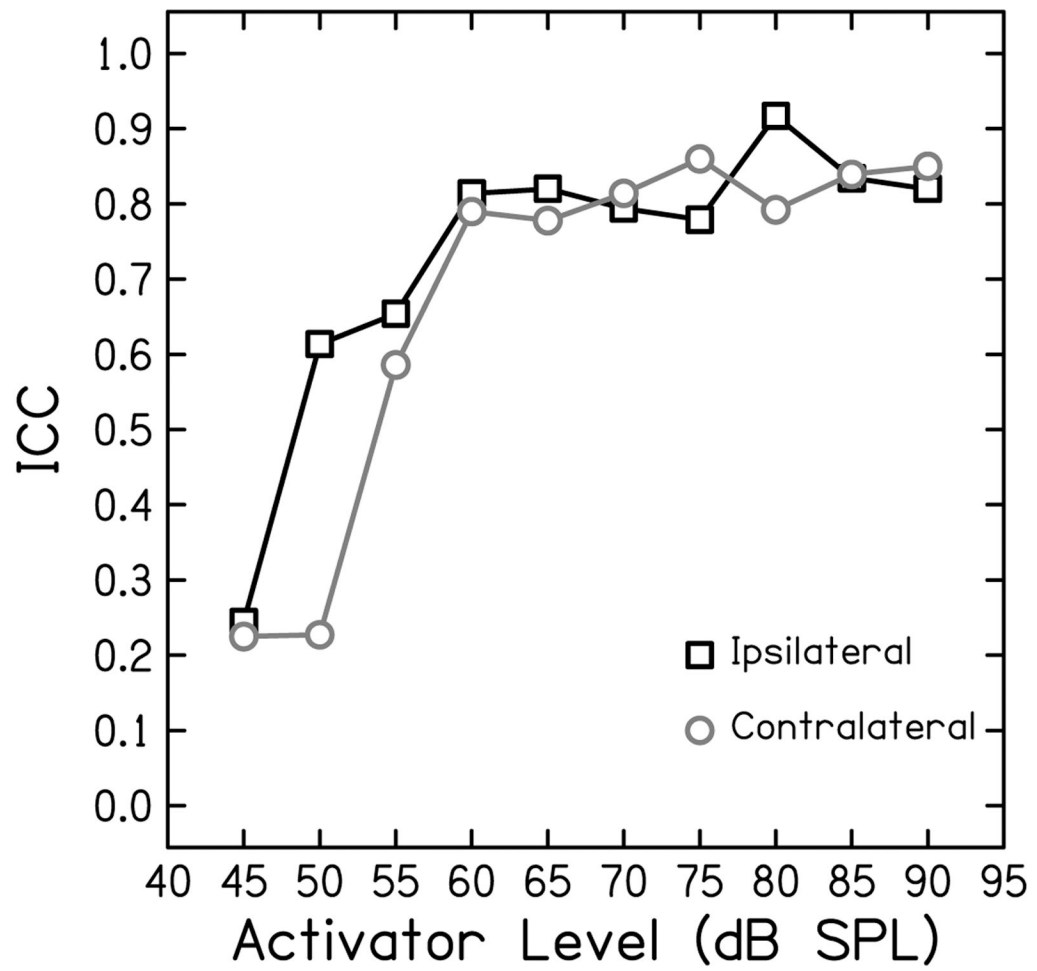
**Figure 5.**

Correlation between visits 1 and 2 for z-scores of word recognition residuals in four listening conditions: quiet (top-left), noise (top-right), time-compressed plus reverberation (bottom-left) and time-compressed alone (bottom-right). Filled squares indicate participants whose difference in score between visits was considered significant according to the 95% confidence intervals presented by Thornton & Raffin (1978). The dashed line indicates perfect reliability between visits.

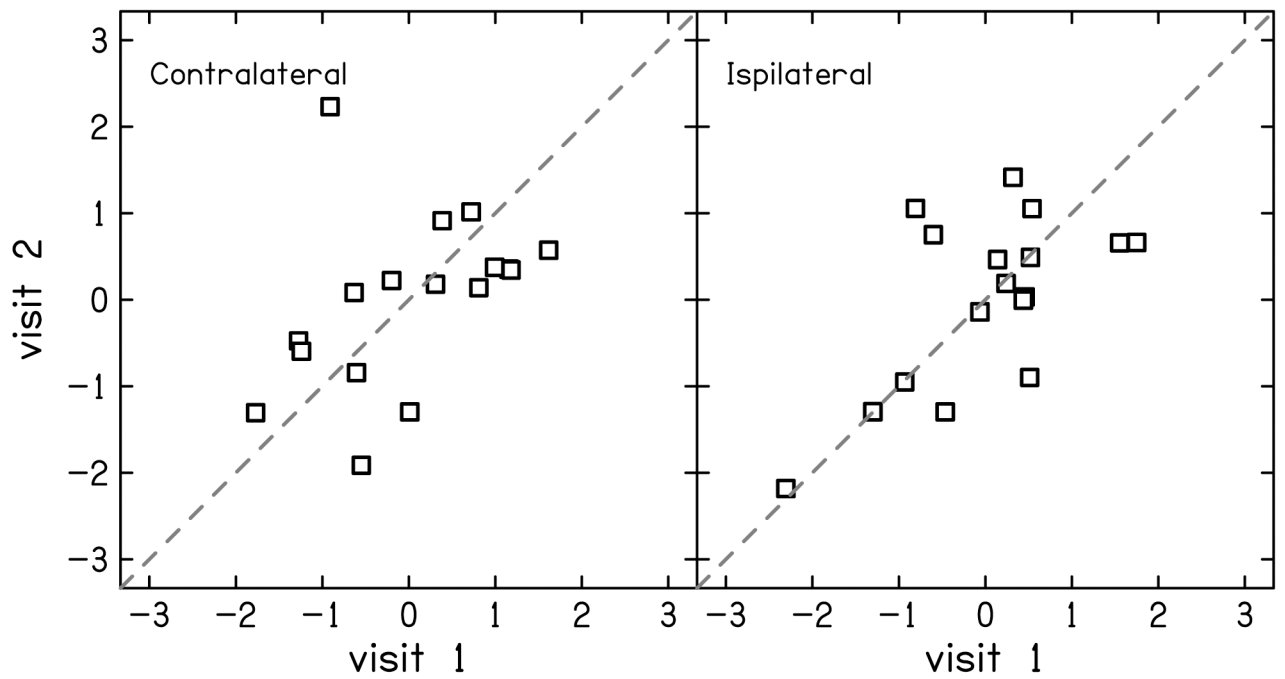




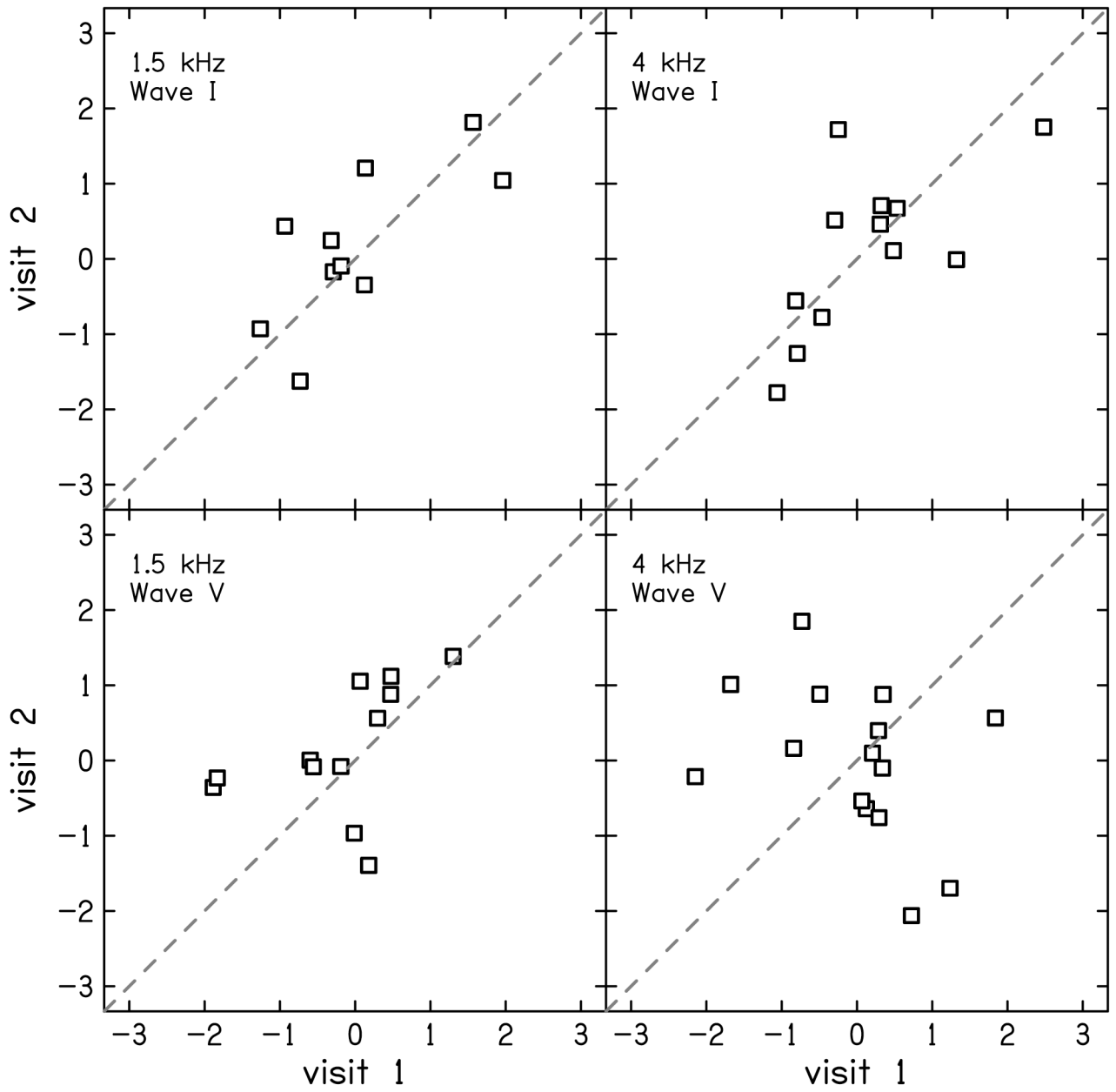
**Figure 6.** Correlation between visits 1 and 2 for z-scores of distortion-product otoacoustic emissions (DPOAE) residuals at 1.5 kHz (left) and 4 kHz (right). The dashed line indicates perfect reliability between visits.



**Figure 7.** Pearson correlation coefficients as a function of activator level for the middle ear muscle reflex. Black boxes indicate r-values for the ipsilateral probe and grey circles represent r-values for the contralateral probe.

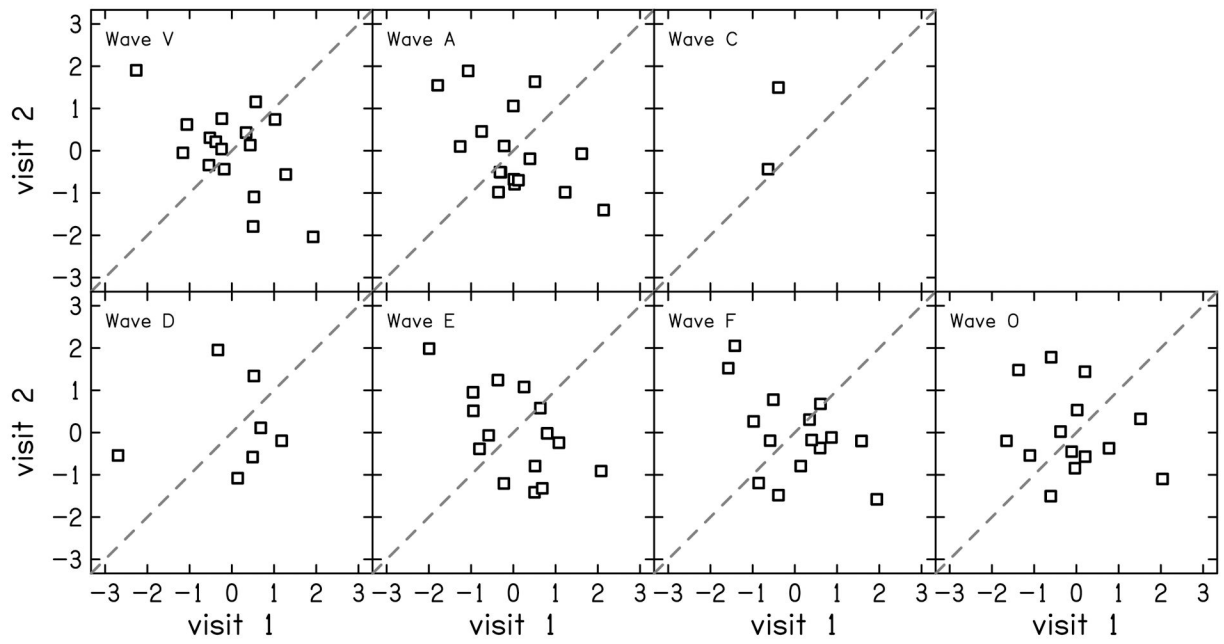


**Figure 8.** Correlation between visits 1 and 2 for z-scores of middle ear muscle reflex (MEMR) residuals in the contralateral (left) and ipsilateral (right) probe conditions. The activator level was 90 dB SPL for both conditions. The dashed line indicates perfect reliability between visits.



**Figure 9.**

Correlation between visits 1 and 2 for z-scores of wave amplitudes of the tone-burst ABR (tbABR) residuals at 1.5 kHz (left) and 4 kHz (right). Wave I amplitude was calculated from the tiptrode montage (top) and wave V from the vertex montage (bottom). The dashed line indicates perfect reliability between visits.



**Figure 10.**  
Correlation between visits 1 and 2 for z-scores of wave amplitudes of the speech ABR (sABR) residuals. The dashed line indicates perfect reliability between visits.

**TABLE 1.**

Stimulus parameters for auditory brainstem measurements

	<i>ibABR</i>		<i>sABR</i>
	1500 Hz	4000 Hz	/da/
<i>Stimulus</i>			
Level	110 dB peSPL	103 dB peSPL	100 dB SPL
Duration	1 ms	1 ms	40 ms
Envelope	Blackman	Blackman	
Rate	11/sec	11/sec	11/sec
Polarity	Condensation	Condensation	Condensation
<i>Recording</i>			
Total averages	1500	1500	2000
Filters	10-1500 Hz	10-1500 Hz	100-3000 Hz
Gain	100,000	100,000	100,000
Artifact Rejection	< ±20 µV	< ±20 µV	< ±20 µV
Sampling Rate	48000 Hz	48000 Hz	48000 Hz
<i>Montage</i>			
<i>Noninverting (+):</i>	Tiptrode & Vertex (Cz)	Tiptrode & Vertex (Cz)	Tiptrode & Vertex (Cz)
<i>Inverting (-):</i>	Opposite ear Mastoid (M2)	Opposite ear Mastoid (M2)	Opposite ear Mastoid (M2)
<i>Ground:</i>	Forehead (Fz)	Forehead (Fz)	Forehead (Fz)

**TABLE 2.**

Qualitative categorization of Reliability based on Intraclass Correlation Coefficient

---

0.9	$r < 1.0$	Excellent
0.75	$r < 0.9$	Good
0.5	$r < 0.75$	Moderate
0	$r < 0.5$	Poor

Author Manuscript

Author Manuscript

Author Manuscript

Author Manuscript

**TABLE 3.**

Reliability of Residual Measures: Absolute Agreement between Visits 1 &amp; 2

		Intraclass Correlation	95% CI		F Test			Sig.
			Lower	Upper	Value	df1	df2	
DPOAE	1.5 kHz	0.78	0.48	0.91	7.65	16	16	< 0.001
	4 kHz	0.91	0.76	0.97	19.29	16	16	< 0.001
MEMR	Contra <sub>90</sub>	0.87	0.67	0.95	13.23	16	16	< 0.001
	Ipsi <sub>90</sub>	0.83	0.60	0.94	10.39	16	16	< 0.001
tbABR	Wave I <sub>1.5 kHz</sub>	0.74	0.25	0.93	6.37	9	9.56	0.005
	Wave I <sub>4 kHz</sub>	0.68	0.19	0.90	4.98	11	11	0.007
	Wave V <sub>1.5 kHz</sub>	0.47	0.00*	0.81	2.89	11	12	0.041
	Wave V <sub>4 kHz</sub>	0.00*	0.00*	0.15	0.45	14	14	0.926
sABR	Wave V	0.00*	0.00*	0.46	0.95	16	16	0.541
	Wave A	0.00*	0.00*	0.25	0.59	16	16	0.853
	Wave C	-	-	-	0.02	1	0.01	0.988
	Wave D	0.06	0.00*	0.76	1.11	6	6	0.452
	Wave E	0.19	0.00*	0.64	1.44	14	14	0.253
	Wave F	0.02	0.00*	0.53	1.05	14	14	0.466
	Wave O	0.28	0.00*	0.70	1.73	13	13	0.169
TIN	1.5 kHz	0.68	0.30	0.87	5.01	16	16	0.001
	4 kHz	0.52	0.05	0.80	3.02	16	16	0.017
FMDT	1.5 kHz	0.79	0.50	0.92	8.20	15	15	<0.001
Words	Quiet	0.63	0.24	0.85	4.27	17	17	0.002
	Noise	0.55	0.11	0.81	3.31	17	17	0.009
	TC <sub>45</sub>	0.39	0.00*	0.72	2.23	17	17	0.054
	TC <sub>45</sub> + RT <sub>0.3</sub>	0.57	0.14	0.82	3.53	17	17	0.007

DPOAE, distortion-product otoacoustic emission; MEMR, middle ear muscle reflex activated at 90 dB SPL; tbABR, tone-burst auditory brainstem response; sABR, speech-evoked auditory brainstem response; TIN, threshold-in-noise; FMDT, frequency-modulated detection threshold; TC<sub>45</sub>, Time-compressed speech; TC<sub>45</sub>+RT<sub>0.3</sub>, Time-compressed speech plus reverberation.

\* ICC value erroneously negative, interpreted as zero



**TABLE 4.**

Median (IQR) of speech ABR stimulus-to-response correlations and coherence

		<b>Visit 1 Median (IQR)</b>	<b>Visit 2 Median (IQR)</b>
STR	Tip-trode	0.48 (0.29)	0.31 (0.24)
	Vertex	0.24 (0.12)	0.21 (0.09)
MSC	Tip-trode	0.21 (0.23)	0.28 (0.25)
	Vertex	0.13 (0.19)	0.13 (0.18)

IQR, inter-quartile range; STR, stimulus-to-response correlation; MSC, magnitude-squared coherence