Validity and reliability of in-situ air conduction thresholds measured through hearing aids coupled to closed and open instant-fit tips.

Anna O'Brien, Gitte Keidser, Ingrid Yeend, Lisa Hartley, and Harvey Dillon

National Acoustic Laboratories

Corresponding author:

Gitte Keidser, NAL, 126 Greville St, Chatswood, NSW 2067, Australia

e-mail: gitte.keidser@nal.gov.au

Keywords:

In-situ audiometry, test-retest reliability, open fitting, insert earphones

Abbreviations:

3FA: three-frequency-average (average threshold measured across 0.5, 1, and 2 kHz); BTE: behind-the-ear; Closed: in-situ measurement through hearing aid coupled to a closed dome tip; FRT: fine resolution threshold (1.5 dB step size); GUI: graphical user interface; HL: hearing level; HTL: hearing threshold level; Insert: measurement obtained with conventional insert earphones; KEMAR: Knowles electronic mannequin for acoustic research; LRT: large resolution threshold (4.5 dB step size); Open: in-situ measurement through hearing aid coupled to an open dome tip; PC: personal computer; RECD: real-ear-to-coupler difference; REDD: real-ear-to-dial difference, REOR: real-ear occluded response; REUR: real-ear unaided response; RITE: receiver-in-the-ear; SD: standard deviation; SPL: sound pressure level; SQRT: square root; TSD: true standard deviation

Abstract

Audiometric measurements through a hearing aid ("in-situ") may facilitate provision of hearing services where these are limited. This study investigated the validity and reliability of in-situ air conduction hearing thresholds measured with closed and open domes relative to thresholds measured with insert earphones, and explored sources of variability in the measures. Twenty-four adults with sensorineural hearing impairment attended two sessions in which thresholds and real-ear-to-dial-difference (REDD) values were measured. Without correction, significantly higher low-frequency thresholds in dB HL were measured in-situ than with insert earphones. Differences were due predominantly to differences in ear canal SPL, as measured with the REDD, which were attributed to leaking low-frequency energy.   Test-retest data yielded higher variability with the closed dome coupling due to inconsistent seals achieved with this tip. For all three conditions, inter-participant variability in the REDD values was greater than intra-participant variability. Overall, in-situ audiometry is as valid and reliable as conventional audiometry provided appropriate REDD corrections are made and ambient sound in the test environment is controlled.

Introduction

Hearing aid fitting typically involves prescription of hearing aid output levels expressed in dB SPL, or gain in dB, based on measurement of frequency dependent hearing threshold data obtained in dB HL. To get from dB HL to dB SPL in the ear canal or in a coupler, average real-ear-to-dial difference (REDD) or real-ear-to-coupler difference (RECD) data are typically applied to the threshold data (e.g. Revit, 1997; Dillon, 2001, pg 407). For more precise calculations, the use of individual RECD or REDD values rather than average values is recommended. This is because these measurements have been found to vary greatly between both individuals and transducers due to placement of the transducer and differences in ear canal volume (e.g. Valente et al., 1994; Munro and Salisbury, 2002; Saunders and Morgan, 2003). The validity and reliability of using individually measured transforms to get from HL to SPL have been verified by Scollie et al. (1998). An alternative to using individual RECD or REDD values is to measure audiometric data directly in real ear SPL using a probe tube; i.e. "in-situ" (e.g. Kiessling, 1987; Gagné et al., 1991; Gauthier and Rapisardi, 1992; Valente et al., 1997). A potential problem, however, with using probe tube equipment for threshold measurements is the effect of the equipment's noise floor, which typically is about 40 dB SPL, on assessment of lesser degrees of hearing loss (Scollie et al., 1998).

Today, most hearing aids have an on-board signal generator, which in many cases is used to produce a brief tone, or series of tones, to guide the wearer through program and volume changes in the hearing aid and to alert the wearer when the battery needs changing. Some manufacturers have further used the hearing aid generated signals for in-situ threshold measurements (Ludvigsen & Topholm, 1997; Bostock et al., 2004), and even for probe microphone measurements (Yanz et al., 2007; Bohnert and Brantzen, 2004). In both cases, measurements are obtained with the hearing aid (and mould) in place using the hearing aid receiver as the transducer for delivering sound. In the case of the probe microphone measurements, the hearing aid microphone or direct audio input temporarily becomes the probe tube microphone. Control of the signal presentation is done either from the manufacturer's fitting software installed on a PC or on a proprietary hand-held programmer (e.g. Ludvigsen and Topholm, 1997). During in-situ audiometry, as in conventional audiometry, the hearing aid user indicates to the clinician when the sound is heard.

Smith-Olinde et al. (2006) studied test-retest reliability of in-situ thresholds measured with a commercial product in normal-hearing adults. In this study, pulsed frequency-modulated tone bursts were generated by and delivered through a Widex Diva SD-9 behind-the-ear (BTE) device coupled to the ear canal using ER-3A insert earphone tips. For each of 43 adults, threshold measurements were obtained in 5 dB steps at four audiometric frequencies in a sound-treated test booth. The measurements were obtained twice during the same appointment with the insert tip completely removed and reinserted between trials. It was concluded that in this population, the test-retest reliability of in-situ thresholds was equivalent to that of conventional audiometric procedures. Relative to conventional audiometry using pure tones, the average in-situ thresholds in dB HL varied 7, 7.5, 5, and -3.5 dB at 0.5, 1, 2, and 4 kHz, respectively, with individual threshold agreements ranging from -10 to +10 dB in at least 91% of measurements at each frequency (Smith-Olinde, personal communication 2007). It was speculated that the average discrepancies were mainly due to differences in stimuli and acoustic factors, especially related to tubing resonance. Further, Winter and Kuk (1998) reported that in-situ thresholds measured with a similar system on 14 children with moderately-severe hearing loss were within 5-10 dB of conventional audiometric thresholds for every child.

If valid and reliable, in-situ audiometry through hearing aids has great cost-saving potential as less equipment and space are needed, which may facilitate provision of hearing services in areas where these are limited. It should be made clear though that in-situ audiometry performed through hearing aids is currently restricted to non-complex air conduction threshold measurements as masking is not offered and measurement of bone conduction thresholds is not possible. However, in cases for whom it is likely that the hearing problem is relatively symmetrical and sensorineural in nature, in-situ audiometry could become a replacement for conventional air-conduction threshold measurements. The clinician's time may also be saved and potential errors reduced in transferring threshold data between different test modules.

A problem similar to the noise floor issue encountered in the probe tube equipment is inherent in in-situ audometry. Many hearing aid users, especially with mild-moderate sensorineural hearing loss, are fitted with vents or open moulds. Vents, or any leakage around the hearing aid or mould, will allow ambient low-frequency noise to directly enter into the ear canal and allow low-frequency sound to escape the ear canal (e.g. Dillon, 2001; pg 126-127), both of which may affect low-level threshold measurements particularly at low frequencies. To overcome this problem, it has been recommended that vents are blocked during in-situ threshold measurements (Bostock et al., 2004). This method, however, could pose some problems with the newer style open dome tips.

The overall aim of this study was to obtain a better understanding of the validity and reliability of in-situ audiometry measured through hearing aids on hearing-impaired persons fitted with instant-fit open and closed domes. Specifically, the study aimed at determining:

1) The validity of air conduction thresholds obtained with in-situ audiometry using closed and open dome instant-fit tips against conventional thresholds measured using insert earphones; both in dB HL and converted to dB SPL using individually measured REDD values,
2) The reliability of in-situ audiometry between sessions spaced about 3 weeks apart,
3) An understanding of the sources of any test-retest variability.

Methodology

Participants

Seventeen female and eight male hearing-impaired adults with known sensorineural hearing loss falling within the stimulus range of the test transducers completed the protocol. No participant had more than 30 dB asymmetry between the ears at any frequency, bypassing the need for masking. One participant was later omitted from the analysis due to fluctuating low-frequency hearing. The remaining 24 participants were aged between 43 and 82 years, with a mean age of 73. The mean three frequency average (3FA) for the ear used in the study (average hearing threshold level (HTL) at 0.5, 1 and 2 kHz) as measured with insert earphones at the first appointment was 50 dB HL [range: 36 – 72 dB HL]. Twenty-one participants were hearing aid wearers and three participants did not wear amplification. The right ear was selected as the test ear for half the participants and the left ear for the other half.

Instrumentation

In-situ audiometry: Siemens' Centra HP BTE hearing instruments that include on-board pure tone generators were used to obtain in-situ thresholds. With the hearing instrument microphone muted, activation of the tones was controlled via a Matlab Graphical User Interface (GUI) developed at NAL using a library of hearing instrument commands supplied by Siemens that enabled activation of the tones and controlled the duration of the presentation. Communication with the hearing instrument during testing was via an Aurical-integrated HiPro, which caused a 3.5 – 5 second delay between when the tone was elicited by PC and when it was heard. The level of the in-situ stimuli could be adjusted in 1.5 dB steps. For calibration of the hearing aid transducer, see below.

The hearing instruments were coupled to the test ear using the Siemens Life instant fit tubing of appropriate length so that the bend of the tubing sat against the top of the ear canal opening. Instant-fit domes were attached to the medial end of the tube. Open domes of 8 and 10 mm diameter were fitted to seven and 17 participants, respectively. Double-seal closed domes of 8-10 mm (eight participants) or 10-12 mm diameter (16 participants) were used. Henceforth, these two test conditions are in short referred to as "Open" and "Closed".

Conventional audiometry: To ensure consistency of the methodology across the conventional and in-situ threshold measurements, the Matlab GUI was also used to drive insert earphones for the conventional audiometry. In this case, the stimuli were pure tone wave files generated in Audition with 50 ms fade-in and fade-out to prevent spectral splatter from rapid signal onset/offset. The wave files were produced by a Lynx One sound card, and amplified through a Sony Source Direct Circuit Super Legato Linear F242 amplifier with a left/right switch box to eliminate cross-talk. The stimulus was then routed to Etymotic EAR 3A insert earphones. Unlike the in-situ conditions, the stimulus presentation time for the insert earphones was instantaneous. Henceforth this test condition is in short referred to as “Insert”.

Audiometry

Threshold measurements: Audiometry was conducted as per ISO 8253-1 using pure tones of 1 sec duration (computer controlled) presented in 4.5 dB steps. Thresholds were obtained for the following frequencies in the order: 1, 1.5, 2, 3, 4, 0.5, and 0.25 kHz followed by a 1 kHz re-check. Participants responded by pressing a button when they heard a tone, which illuminated a small light in front of the researcher. Threshold was defined as the lowest level at which the participant responded twice on an ascending presentation. Once a threshold was established using 4.5 dB steps (large resolution threshold - LRT), tones were presented 3 and 1.5 dB below that level to obtain a fine resolution threshold (FRT).

Noise floor: Threshold measurements were performed in a sound treated test booth with the computer positioned outside the booth to reduce ambient equipment noise. Ambient noise in the test-booth was ~ 20 dB A Leq.

Calibration: The hearing instruments and insert earphones were calibrated in a 2cc coupler with the reference equivalent sound pressure levels (RETSPLs) specified in ISO 389-2 (1994, E) using an 81 dB HL stimulus level. The hearing instruments were connected to the HA 1 coupler via a #2 thin tube using putty to seal around the tube. The inserts were coupled to the HA 2 2cc coupler with a short length of tube just long enough to accommodate the insert adapter, as specified in the standard. Subsequent to the study, a correction was made for the difference in the length of tube resulting from using two different couplers. To determine these correction values, the output of the insert phone (connected using zero-length tubing) in the HA2 coupler, minus the HA1 output (connected using the black 25 mm tubing sealed to the coupler using putty) was calculated. This difference was:

- Added to all in-situ HL values measured in the experiment.
- Subtracted from all in-situ REDD values measured in the experiment.

Real Ear to Dial Difference measurements

The REDD, used to convert the threshold measurements from dB HL to dB SPL, was calculated for each test condition by measuring the ear canal SPL and subtracting the HL dial setting. Pure tone stimuli of 10 sec duration were presented at a dial setting of 72 dB HL. The ear canal SPLs were measured using a Madsen Aurical real-ear measurement system. Care was taken to ensure that the tip of the probe tube was positioned within 6 mm of the eardrum, using the 6 kHz notch method (Dillon, 2001; pg 91-94). For each test condition, REDD values were obtained for each of the frequencies 0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, and 6 kHz.

Procedure

Each participant attended two appointments spaced about 3 weeks apart. At the first appointment a full audiogram was obtained using conventional audiometry to confirm candidacy for the project. Screening tympanometry was also assessed. To reduce test time and the likelihood of participant fatigue, all subsequent testing was conducted unilaterally. For all except four participants, the test ear was decided based on the order in which the participants was recruited, with the left ear tested for odd-numbered participants and the right ear tested for even-numbered participants. The remaining four participants were changed to the other ear for reasons such as excess cerumen, history of fluctuating hearing, and recent presence of otitis externa. Pure tone air conduction audiometry was repeated in the selected ear for the in-situ conditions (Open and Closed). A different instrument was used for each participant to include any device-related variability in the measurements. The order of open and closed testing was balanced across participants. REDD was measured for each test condition.

At the second appointment, tympanometry, conventional (Insert) and in-situ audiometry (Closed and Open), and REDD measurements were repeated in the test ear.

Results

REDD

Figure 1 shows the average REDD values measured across appointments for each test condition at each frequency. For ready use, the actual REDD values are also provided in Table I.   A repeated measures ANOVA was performed using the REDD values as observations and appointment (1, 2), condition (Insert, Closed, Open) and frequency

(0.25, 0.5, 0.75, 1, 1.5, 2, 3, 4, 6 kHz) as repeated measures, with ear (left, right) as a grouping variable. There was no significant effect of ear ($F_{1,22} = 0.48$, $p = 0.49$) or appointment ($F_{1,22} = 0.31$, $p = 0.58$), nor any significant interactions including ear or appointment ($F_{1,22} = 4.81$, $p = 0.053$), confirming that the REDD measurements were reliable over time and unaffected by ear. Test condition and frequency both showed significance ($F_{2,44} = 263.07$, $p < 0.000001$, and $F_{8,176} = 120.66$, $p < 0.000001$) as did the interaction between test condition and frequency ($F_{16,352} = 130.23$, $p < 0.000001$). Tukey's post hoc analysis showed that Insert REDD values were significantly higher than Closed REDD values at 0.25 and 0.5 kHz ($p = 0.00002$), indicating that greater occlusion was achieved with the foam eartip than with the closed dome. Further, the Open REDD values were significantly lower than both Insert and Closed REDD values at 0.25, 0.5, 0.75, 1, and 1.5 kHz ($p = 0.00002$), confirming that a substantial loss of low-frequency energy results from the open dome. At 3 kHz the Open REDD value was, on the other hand, significantly higher than the REDD value measured with the occluding tips ($p < 0.0002$), a result that is likely caused by some residual resonance of the ear canal enhancing this frequency.

Threshold measurements

Figure 2 shows the LRTs in dB HL obtained in each condition when averaged across participants and appointments, revealing substantial variability. These thresholds would be equivalent to those obtained in a clinical environment when no corrections are made for the use of each particular transducer and coupling. A repeated measures ANOVA was performed to determine the effect of retest. The ANOVA revealed no significant differences between test and retest data ($F_{1,23} = 0.17$, $p = 0.69$), and there were no significant interactions involving appointment time ($p > 0.71$). Specifically, across frequencies, 96%, 88%, and 96% of test-retest data were within ± 4.5 dB, and 99%, 98%, and 100% of test-retest data were within ± 9 dB for the Insert, Closed, and Open conditions, respectively. The statistical analysis further showed significant effects of test condition ($F_{2,46} = 267.1$, $p < 0.0000001$) and frequency ($F_{6,138} = 6.4$, $p = 0.000006$), and a significant interaction between test condition and frequency ($F_{12,276} = 98.6$, $p < 0.0000001$). These effects mirrored the effects measured in the REDD data, suggesting that the differences between the average thresholds seen in Figure 2 are predominantly due to loss of low-frequency energy, and hence that the thresholds in dB SPL (dB HL + REDD) will be comparable.

To more accurately assess the psychoacoustic variability in hearing threshold in dB SPL, each individual's REDD values measured at each appointment were added to their FRTs in dB HL measured at the same appointment. A repeated measures ANOVA conducted on the dB SPL data revealed a significant main effect of frequency ($F_{6,138} = 28.88$, $p < 0.000001$) and a significant interaction between test condition and frequency ($F_{12,276} = 5.55$, $p < 0.000001$). No other main effects or interactions were significant ($p > 0.42$). Tukey's post hoc analysis of the interaction between test condtion and frequency showed that at 0.25 kHz the Open threshold was significantly better (lower) than the Insert threshold by 2.5 dB ($p = 0.00005$). Further, at 2 kHz the Closed threshold was significantly poorer (higher) than the Insert threshold by 2.0 dB ($p = 0.002$), see Figure 3. Although statistically significant, the differences in FRTs (dB SPL) are clinically negligible, i.e. well within acceptable measurement errors for threshold values, and would not significantly impact fitting prescription targets. These results indicate that in-situ audiometry using either closed or open domes yields valid audiometric results when combined with individually-measured REDD values.

To investigate the accuracy with which hearing thresholds (SPL) may be inferred using average REDD data to convert from dB HL to dB SPL at the eardrum, the inter-participant standard deviations (SDs) of measured REDD values were inspected (cf. Table I). The variability in REDD measurements was particularly high for the Closed condition, especially below 1.5 kHz, suggesting the possibility that the positioning and/or seal of the closed dome in the ear canal varied considerably among participants. Subsequent measurements made on KEMAR revealed only a linear 4 dB increase in SPL in the ear canal across frequencies when the closed dome was inserted over a 12 mm range of insertion depth. These measurements indicate that residual ear canal volume makes only a minor contribution to the variability in REDD values obtained for the Closed condition; the major factor consequently being the extent of leakage that occurred around the ear tip. Overall, these findings suggest that while average REDD corrections may be adequate for the Insert and Open conditions, the variability in levels produced at the eardrum in the Closed condition warrants individual measurement of the REDD, at least below 1.5 kHz.

Sources of variability in measurements

Although there were no significant test-retest effects demonstrated in the average REDD or threshold data, some intra-participant variation was observed. To explore this further, the SD of the mean test-retest difference across participants was calculated to capture measurement variability between appointments. To isolate for each test condition whether intra-participant variability was caused more by acoustic factors (e.g. coupling of the transducer to the ear), psychoacoustic factors (e.g. responses to the threshold measurements), or a combination of these, the SD of the test-retest difference values was calculated separately for each frequency for the FRT (dB HL), REDD, and FRT (dB SPL) measurements, respectively. These values are shown in Figure 4.

For the Insert and Open conditions, the test-retest SDs calculated for FRT (dB HL) and FRT (dB SPL) tended to be slightly higher than the test-retest SD calculated for REDD, indicating that intra-participant variability was more of psychoacoustic origin than caused by acoustic factors. On the other hand, for the Closed condition, the test-retest SD calculated for FRT (dB HL) and REDD tended to show higher values than for the FRT (dB SPL), particularly at 0.25 and 0.5 kHz, indicating that for this condition the dominant source of intra-participant variability arose from inconsistent coupling of the closed dome to the ear. The latter finding supports the suggestion above that the seal of the closed dome is the most likely cause for the higher inter-participant SD values reported in table I. All conditions showed similar test-retest SDs in FRT (dB SPL) averaged across frequencies (2.6 – 3 dB), suggesting that once individual differences in REDD have been accounted for, the transducer and coupling combination do not influence the reliability of hearing thresholds obtained.

To further break down whether the variability in REDD measurements was due to true variation in the REDD values (i.e. due to variability in ear canal volume or placement/seal of the tip in the ear canal) or to variation resulting from the measurement procedure, the true inter-participant REDD SD was calculated according to formula 1, which subtracts from the apparent inter-subject variance, the variance that is due to random test-retest error within individuals.

$$TSD_{inter} = \sqrt{(SD^2_{inter} - (SD^2_{test\text{-}retest}/2))} \qquad (1)$$

Where:

TSD_{inter} = true inter-participant SD of REDD
SD_{inter} = inter-participant SD of REDD data (appointment 1)
$SD_{test\text{-}retest}$ = SD of the REDD test-retest difference values.

The comparison of the true inter-participant SD and the intra-participant SD ($SD_{test\text{-}retest}/\sqrt{2}$) can be seen in Figure 5. It is clear that for all three conditions the primary source of REDD variability is true inter-participant variability, that is true differences in ear canal acoustics and leakage between individuals, rather than measurement variability.

Discussion

The current study explored the validity and reliability of air conduction thresholds obtained using in-situ audiometry (with closed and open instant-fit domes) against conventional audiometry using insert earphones. Thresholds were measured in dB HL and converted to dB SPL using individually measured REDD values. Reliability was observed from measurements spaced about 3 weeks apart.

When using a step size of 4.5 dB, this study found differences of up to 10 and 30 dB between average air conduction thresholds measured in dB HL obtained with insert earphones and the Closed and Open conditions, respectively. The greatest discrepancy for both in-situ conditions was observed in the low frequencies (Figure 2). That is, the less occluded the coupling to the ear was, the more the in-situ thresholds deviated from the conventionally measured thresholds across the low frequencies. The difference in dB HL observed between the Closed and the Insert conditions is consistent with the up to 10 dB difference between in-situ and conventional thresholds reported by Winter and Kuk (1998) and Smith-Olinde et al (2006). Specifically, Smith-Olinde et al (2006) used unvented foam tips to couple the hearing aid to the ear and, although Winter and Kuk (1998) do not describe the earmold characteristics used in their study, the fact that the study was performed on children with average hearing loss in the better ear of moderately-severe degree across all frequencies suggests that venting would have been minimal.

Two factors come into play as the coupling from the hearing instrument to the ear becomes less occluded; the inward transmission of ambient noise, which could mask low-level test tones, and the loss of low-frequency energy from the ear canal. The real ear occluded response (REOR) was also measured in this study for each participant and each transducer and coupling combination using a 65 dB SPL input. This measurement showed that, on average, the insert earphone provided 10-15 dB attenuation of ambient noise across the frequency range, which is about half that claimed by the manufacturer when using deep insertion (Etymotic, 2003). The closed dome tip provided, on average, no attenuation from 2 - 4 kHz and only about 5 dB attenuation outside these frequencies. The open dome tip provided no attenuation of ambient noise, but maintained the passive amplification provided by pinna and ear canal resonances (e.g. Mueller and Ricketts, 2006; Kiessling et al., 2006). The implication of these observations is that both the Closed and Open in-situ conditions require respectively more stringent control of ambient noise than the more occluding conventional insert earphone to avoid inadvertent masking of test-tones. One way to ensure that the true threshold is measured would be for the hearing instrument microphone to first measure the level and spectrum of the ambient noise. These measurements can then be used to alert the clinician when the hearing thresholds measured are similar to those expected were they to be masked thresholds. Because of the low-noise test space used in this study, and the fact that the participants were known to have elevated thresholds, inward transmission of ambient noise is not considered to have influenced the results reported in this paper. Consistent with this is the uniformity of the dB SPL thresholds shown in Figure 3.

The loss of low-frequency energy from the ear canal was quantified by the measured REDDs (Figure 1), and almost perfectly mirrored the impact of coupling on the dB HL thresholds in Figure 2. Relative to the insert earphone, which in this study was used as the reference transducer for occluding the ear canal, both the Closed and Open in-situ conditions allowed leakage of low-frequency energy from the ear canal. Comparison of average REDDs measured in the Closed condition showed loss of 8 – 10 dB below 0.75 kHz relative to the insert earphone. Above this frequency, differences between Insert and Closed REDDs were clinically minor (≤ 3 dB). Not surprisingly, the average REDD for the Open condition showed much greater loss of low-frequency energy, increasing from 8 dB at 1.5 kHz to 30 dB at 250 Hz. This relative loss of low-frequency energy is similar to the difference in real-ear-to-coupler difference (RECD) measurements obtained with open and closed moulds seen in Figure 1 of Hoover et al. (2000). Interestingly, the Open REDD at 3 kHz was significantly higher than for the other conditions, which likely reflects the preservation of the ear canal resonance at this frequency.

As demonstrated in Figure 3, converting the in-situ thresholds measured in dB HL to dB SPL, by applying the individually measured REDD values applicable to the choice of transducer and coupling, will effectively address the problem with loss of low-frequency energy from the ear canal and present valid in-situ thresholds. Consequently, in-situ thresholds should not be directly entered into fitting prescription rules without somehow being corrected for the differences in REDD values, relative to the transducer that the tone generator is calibrated for.

With respect to reliability, it was found that across frequencies for each test condition, 88-96% of test-retest measurements for LRTs (dB HL) were within ±4.5 dB and that 98-100% were within ±9 dB. This is in good agreement with Schmuziger et al. (2004) who found the test-retest variability of threshold measurements on 138 ears, using insert earphones, to be within 5 dB for 89-99% of ears for each frequency, and within 10 dB for 95-100% of ears. For all conditions, except the Closed at 0.25 kHz, the test-retest SD values obtained for the FRT in dB HL (see Figure 4) are, comparable to the test-retest variability measured by Jerlvall and Arlinger (1986) on 10 listeners with cochlear hearing loss when using insert earphones and 2-dB intensity increments. They are, however, with the exception of 4 kHz, higher, by 1 to 2 dB, than the variability measured for in-situ thresholds on normal-hearing listeners using frequency-modulated tone bursts and a 5-dB step size in Smith-Olinde et al. (2006). At 4 kHz, about 1.5 dB lower variability was obtained with the smaller step size used for the FRT in this study.   Jerlvall and Arlinger (1986), who presented threshold data measured in both 5-dB and 2-dB step size on normal-hearing and hearing-impaired listeners, found a significant effect of step size only on test-retest variability measured on hearing-impaired listeners at 3 and 4 kHz. In agreement with the observation above, the variability was lower when using a smaller step size, and lower than the variability measured on normal-hearing listeners. They suggested that this finding could be explained by a lower difference limen for intensity measured on hearing-impaired listeners at these frequencies.

Finally, the test-retest variability in the REDD data obtained in this study (Figure 4) falls between the variability in REDD and RECD data obtained with TDH and insert earphones, respectively, by Scollie et al. (1998), and that reported by Munro and Lazenby (2001) and Munro and Davis (2003), who measured REDD and RECD values, respectively, using insert earphones. Overall, these findings suggest that in-situ thresholds can be measured as reliably as with conventional audiometry.

Of note was a noticeable higher test-retest variability of REDD at 6 kHz, evident in Figures 4 and 5, which is in agreement with data by Munro and Lazenby (2001) and Munro and Davis (2003). The most likely explanation for this would be variations in the position of the probe microphone used to measure the REDD, which may cause standing wave errors in the high frequencies. Consistent with this explanation is the SD of the real-ear unaided response (REUR) test-retest difference (not used in this paper), which was also highest at 6 kHz (3.5 dB compared with 0.73 – 2.5 dB at all other frequencies). These variations occurred despite the attempt in this study to deliberately control the standing wave error by using the 6 kHz notch method to position the probe microphone 5 mm from the eardrum, which should yield standing wave errors at 6 kHz of less than 2 dB (Dillon, 2001). Consequently, we currently have no explanation for the relatively high REDD variability obtained at 6 kHz.

The inter-participant SDs calculated for the REDD measurements, which were found to be due to actual differences between participants' REDD values more than variability in the measurement itself (Figure 5), were higher than the 2 dB inter-participant SD measured on 16 participants, for all frequencies below 6 kHz using insert earphones, by Munro and Lazenby (2001). On the other hand, they were generally lower than, or comparable to, the inter-participant variability in real-ear SPL and RECD measurements reported on larger populations (30 – 1814 ears) using insert earphones by Hawkins et al. (1990), Valente et al. (1994), and Saunders and Morgan (2003). The inter-participant variability in REDD values for the Open condition was comparable to, and even better at the low frequencies than, those obtained for insert earphones (Figure 5). This would indicate that at low frequencies the Open condition produced more consistent levels in the ear canal than the conventional insert earphone, presumably because the leakage of low-frequency energy from the open dome was more consistent, irrespective of the actual placement in the ear canal or diameter of the ear canal. It would also imply that the Open condition is the most sensitive to changes in hearing threshold over time. For both the in-situ conditions, the inter-participant variability was generally lower than that measured with supra-aural earphones (e.g. Hawkins et al., 1990; Valente et al., 1994). This indicates that in-situ threshold measurements using both closed and open dome tips appear to be more reliable than thresholds measured with supra-aural earphones. Further, threshold measurements using the open dome tip are as reliable as thresholds measured with insert earphones. It is speculated that the lower SD values reported by

Munro and Lazenby (2001) could be related to the age of the participants, which in their study ranged between 22 and 41 years, and hence was lower than the age range used in this study and may have been narrower than the age range targeted in the other studies. The relevance of age is that the cartilaginous portion of the ear canal in younger participants is likely to be firmer, which may facilitate deeper insertion and/or better retention of the insert tip.

This study has demonstrated that applying individually measured REDD values to thresholds in dB HL yields both reliable and valid air-conduction thresholds in dB SPL relative to conventionally measured thresholds. However, individual REDD measurements are probably only worth doing where the inter-participant variability is high (e.g. > 3 dB) and sufficiently exceeds intra-participant variability (e.g. by 1 dB). These conditions were met at most frequencies for the Closed condition, at 0.25 and 0.5 kHz for the Insert condition, and at 3 - 6 kHz for the Open condition. This means it is possible that the use of average REDD values would be sufficient for converting in-situ threshold measurements obtained with open dome tips, while individually measured REDD values are recommended to be used with in-situ thresholds obtained with a closed dome tip. The higher variability measured for the Closed condition has been attributed to a poorer, or less consistent, seal achieved with the closed dome tip than with the insert or open dome tips.

The REDD values measured for the Insert condition in the current study were similar to empirically-derived REDD data published by Munro and Lazenby (2001), but tended to be lower by 2-7 dB than the average adult REDDs published by Bentler and Pavlovic (1989; columns F (2cc to eardrum) + H (HL to SPL in 2cc)); cf. Figure 1. It could be speculated that the discrepancy with the Bentler and Pavlovic values may be due, at least in part, to a) the lower number of observations supporting some of the Bentler and Pavlovic (1989) data, and/or b) Bentler and Pavlovic's need to adjust individual transfer functions to produce a set of self-consistent transfer functions, and/or c) a shallower insertion depth of the insert earphone in the current study (and in Munro and Lazenby, 2001) than for the transducers on which the Bentler and Pavlovic (1989) data are based.

In passing, a borderline significant effect of gender was observed on the REDD values, with males having, on average, lower REDDs than females across conditions. This is in agreement with Liu and Chen (2000) who demonstrated that the average female ear canal volume is smaller than the average male ear canal volume. On this basis, gender should perhaps be considered whenever average REDD values are used, such as in gain prescription targets, and by hearing instrument manufacturers when implementing in-situ audiometry, or simulating real-ear hearing instrument responses in their fitting software.

It would be instructive for future research to confirm the average REDD correction values obtained in this study on another sample of listeners tested with open and closed domes, and to further explore any effect of gender on REDD and RECD values. Although it may be speculated that inter-aural attenuation of in-situ transducers may be similar to those seen in insert earphones, quantification of this would inform the validity of in-situ audiometry results when an interaural asymmetry is found.

Clinical implications

Due to the need for less equipment and space, the use of in-situ audiometry may facilitate provision of services where these currently are limited. Data from this study suggest that measuring the air conduction hearing threshold using the hearing aid as transducer and either an open or a closed coupling to the ear is a viable option when REDD values are applied to the threshold data. Individually measured REDD values by the clinician must be applied when the clinician is uncertain about the calibration of the transducer or when the selected coupling (e.g. the closed dome tip) provides an insufficiently consistent seal. Average REDD values may be applied, either automatically by the hearing aid or by the clinician, when the calibration of the transducer is controlled and the seal of the coupling is consistent (e.g. the open dome tip). Note that the average REDD values listed in table I is only applicable to devices, including receiver-in-the-ear (RITE) devices, for which the transducer has been calibrated to the RETSPLs of ISO 389-2 (1994-E). While the closed dome will enable threshold measurements with less risk of low-frequency masking from ambient noise than the open dome, it is interesting to note that the open dome will produce more reliable thresholds than the closed dome because the leakage of low-frequency energy from the open dome is most consistent. Whether the clinician chooses to obtain threshold measurements with an open coupling for better reliability of measurements or with a closed coupling to reduce interference from ambient noise, the clinician should be prepared to change the coupling afterwards to the one that is most appropriate for the client’s hearing loss. Finally, as long as masking and bone conduction measurements cannot be performed through the hearing aid, in-situ audiometry should only be relied upon for clients with expected sensorineural hearing loss.

Conclusions

Based on the results of the current study, it may be concluded that, relative to conventional audiometry with an insert earphone, in-situ audiometry using open or closed dome tips yields valid and reliable measures of air conduction hearing thresholds as long as the following conditions are met: 1) the REDD is corrected to account for different methods of coupling the hearing aid to the ear (closed or open), 2) individually-measured REDDs are used when the clinician is uncertain about the calibration of the transducer, and when coupling hearing aids with closed dome tips. The latter is because variability with the closed dome tip was high across most frequencies due to a variable seal achieved in the ear canal, both between participants and between sessions, and 3) the amount of ambient sound in the test environment is controlled such that its ingress to the ear canal with the open, and to a lesser degree the closed, domes does not mask thresholds. Fitting software may use the hearing instrument microphone as a sound level meter/spectral analyser and flag to the clinician when ambient sound is too high for accurate testing at any given frequency.

Acknowledgement

This study was partly funded by Siemens, however, the data analyses and interpretations presented in this paper are solely that of NAL. The authors would like to thank our colleagues Jorge Mejia for developing the Matlab GUIs and for assistance with calibration and troubleshooting and Margot McLelland for performing the insertion depth measurements on KEMAR. We also like to thank Matthias Latzel and Matthias Mueller-Wehlau from Siemens Audiologische Technik (SAT) for co-ordinating the project at SAT and for assistance with trouble-shooting the distortion in the in-situ stimuli, respectively.

References

Bentler,R.A. & Pavlovic,C.V. 1989. Transfer functions and correction factors used in hearing aid evaluation and research. *Ear Hear*, 10(1), 58-63.

Bohnert,A. & Brantzen,P. 2004. RECD and clinical verification in children. *Hear Rev*, 11(5), 50-52.

Bostock,S., Reber,M. & Scheller,L. 2004. In situ audiometry and open fitting: can venting affect test results? *Audiology Insight*, 2, 26-31.

Dillon,H. 2001. *Hearing Aids*. Sydney: Boomerang Press. pp 91-94, 126, 407.

Etymotic. 2003. *Specifications for ER-3A insert earphones for audiometry*. www.etymotic.com

Gagné,J., Seewald,R.C., Zelisko,D.L.C & Hudson,S.P. 1991. Procedure for defining the auditory area of hearing impaired adolescents with severe/profound hearing loss I: Detection thresholds. *J Speech Lang Path Audiol* , 15(3), 13-20.

Gauthier,E.A. & Rapisardi,D.A. 1992. A threshold is a threshold … or is it? *Hear Inst*, 43(3), 26-27.

Hawkins,D.B., Cooper,W.A. & Thompson,D.J. 1990. Comparisons among SPLs in real ears 2 cm3 and 6 cm3 couplers. *J Am Acad Audiol*, 1, 154-161.

Hoover,B.M., Stelmachowicz,P.G. & Lewis,D.E . 2000. Effect of earmold fit on predicted real ear SPL using a real ear to coupler difference procedure. *Ear Hear* 21(4), 31—17.

International Standard. 1994. *Acoustics – Reference zero for the calibration of audiometric equipment – Part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones. ISO 389-2. 1994(E).* Switzerland: International Organization for Standardization.

International Standard. 1989. *Acoustics – Audiometric test methods – Part 1: Basic pure tone air and bone conduction thresholds audiometry. ISO 8253-1. 1989(E).* Switzerland: International Organization for Standardization.

Jerlvall,L. & Arlinger,S. 1986. A comparison of 2-dB and 5-dB step size in pure-tone audiometry. *Scand Audiol*, 15, 51-56.

Kiessling,J. 1987. In situ audiometry (ISA): A new frontier in hearing aid selection. 1987. *Hear Inst*, 38(1), 28-29.

Kiessling,J., Brenner,B., Jespersen,C.T., Groth,J. & Jensen,O.D. 2006. Occlusion effect of earmolds with different venting systems. *J Am Acad Audiol*, 16, 237-249.

Ludvigsen,C. & Topholm,J. 1997. Fitting a wide range compression hearing instrument using real ear threshold data: a new strategy. High *Performance Hearing Solutions*, 2, 37-39.

Liu,T. & Chen,Y. 2000. Aging and the external ear response. *Audiol,* 39, 235-237.

Mueller,H.G. & Ricketts,T.A. 2006. Open-canal fittings: Ten take-home tips. *Hear J*, 59(11), 24-39.

Munro,K.J. & Davis,J. 2003. Deriving the real-ear SPL of audiometric data using the "coupler to dial difference" and the "real to coupler difference". *Ear Hear*, 24(2), 100-109.

Munro,K.J. & Lazenby,A. 2001. Use of the 'real-ear to dial difference' to derive real-ear SPL from hearing level obtained with insert earphones. *Br J Audiol*, 35, 297-306.

Munro,K.J. & Salisbury,V.A. 2002. Is the real-ear to coupler difference independent of the measurement earphone? *Int J Audiol*, 41, 408-413.

Revit,L. 1997. The circle of decibels: relating the hearing test, to the hearing instrument, to real ear response. *Hear Rev*, 14(5), 35-38.

Saunders,G.H. & Morgan,D.E. 2003. Impact on hearing aid targets of measuring thresholds in dB HL versus dB SPL. *Int J Audiol*, 42, 319-326.

Schmuziger,N., Probst,R. & Smurzynski,J. 2004. Test-retest reliability of pure-tone thresholds from 0.5 to 16 kHz using Sennheiser HDA 200 and Etymotic Research ER-2 headphones. *Ear Hear*, 25(2), 127-32.

Scollie,S.D., Seewald,R.C., Cornelisse,L.E. & Jenstad,L.M. 1998. Validity and repeatability of level-independent HL to SPL transforms. *Ear Hear*, 407-413.

Smith-Olinde,L., Nicholson,N., Chivers,C. & Highley,P. 2006. Test-retest reliability of in situ unaided thresholds in adults. *Am J Audiol*, 15, 75-80.

Valente, M., Potts, L.G. & Valente, M. 1997. Differences and intersubject variability of loudness discomfort levels measured in sound pressure level and hearing level for TDH-509P and ER-3A earphones. *J Am Acad Audiol*, 8, 59-67.

Valente, M., Potts, L.G., Valente, M., Vass, W., & Goebel, J. 1994. Intersubject variability of real-ear sound pressure level: Conventional and Insert Earphones. *J Am Acad Audiol*, 5, 390-398.

Winter,M. & Kuk,F. 1998. Using a DSP instrument fitting protocol for pediatric cases. *Hear Rev*, 15(11), 27,71-73 .

Yanz,J., Pisa,J.F.D. & Olson,L. 2007. Integrated REM: real-ear measurement from a hearing aid. *Hear Rev*, 14(5), 44-51.

Table I: Mean REED and inter-participant standard deviation values (shown in brackets) in dB measured for each transducer.

	Frequency in kHz								
Transducer	0.25	0.5	0.75	1	1.5	2	3	4	6
Insert	15.4 (5.0)	10.7 (3.2)	7.5 (3.0)	6.0 (2.7)	11.3 (2.5)	13.7 (2.9)	11.8 (2.2)	10.7 (2.9)	10.0 (4.1)
Closed	5.7 (8.8)	2.9 (5.2)	5.1 (3.8)	4.5 (3.3)	9.7 (3.0)	12.7 (3.6)	13.4 (3.6)	9.1 (4.1)	9.9 (5.0)
Open	-15.4 (2.1)	-13.4 (2.1)	-8.1 (2.1)	-6.3 (2.2)	2.9 (2.8)	10.9 (2.4)	17.3 (3.3)	10.9 (3.8)	10.1 (4.8)

Figure legends

Figure 1. Average REDD values measured for Insert, Closed, and Open conditions, pooled across appointments 1 and 2. Asterisks indicate the average adult REDD from Bentler and Pavlovic (1989). Filled squares indicate the average adult REDD from Munro and Lazenby (2001). Data points are shown shifted for clarity. Error bars indicate ± one standard deviation.

Figure 2. Large resolution thresholds (dB HL) measured in the Insert, Closed, and Open conditions, averaged across participants and appointments. Vertical bars indicate ± one standard deviation.

Figure 3. Fine resolution thresholds in dB SPL at the eardrum, averaged across appointments. Vertical bars indicate ± one standard deviation; asterisks indicate in-situ measurements yielding significantly different thresholds to the insert earphone at that frequency.

Figure 4. Standard deviation of test-retest differences in FRT (dB HL), REDD and FRT (dB SPL) for each frequency and across frequencies (“avg”) measured for a) conventional insert earphones, b) in-situ closed domes, and c) in-situ open domes. FRT data were not collected for 0.75 and 6 kHz.

Figure 5. True inter-participant and intra-participant standard deviation values of REDD measurements for each frequency for a) conventional insert earphones, b) in-situ closed domes, and c) in-situ open domes.

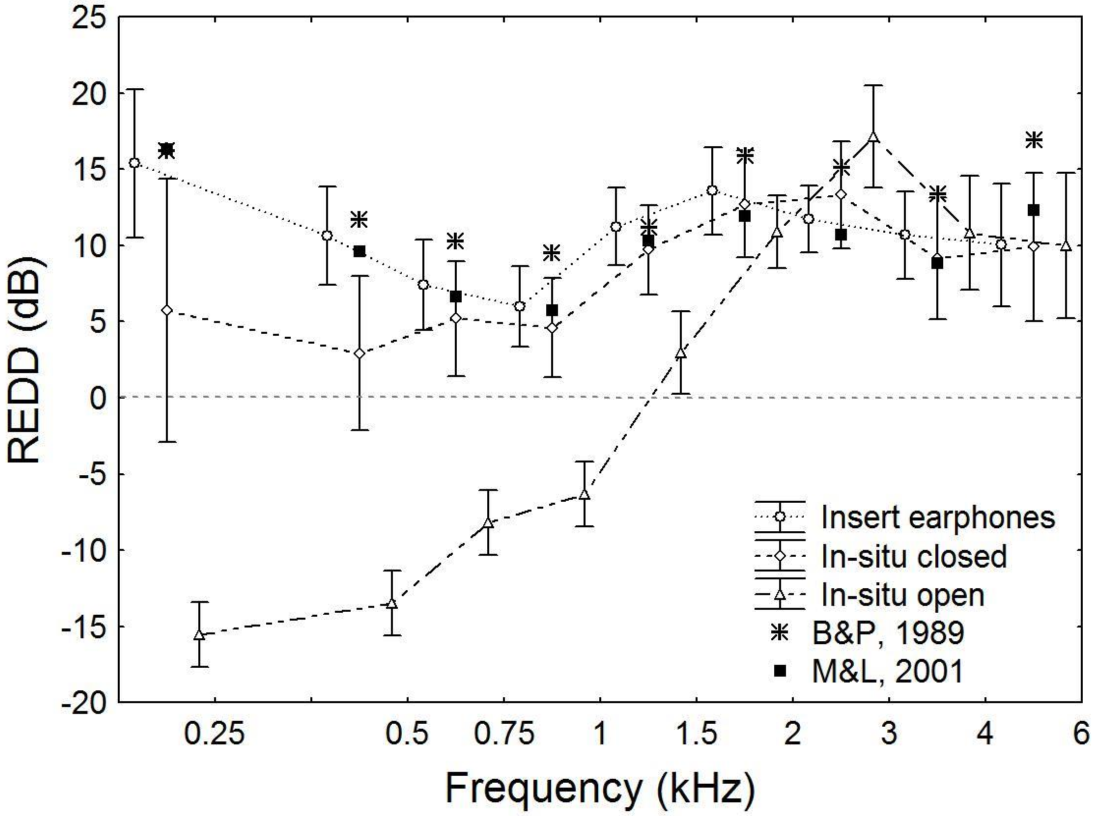
25
20
15
10
5
0
-5
-10
-15
-20
REDD (dB)
0.25
0.5
0.75
1
1.5
2
3
4
6
Frequency (kHz)
Insert earphones
In-situ closed
In-situ open
B&P, 1989
M&L, 2001

Figure 1

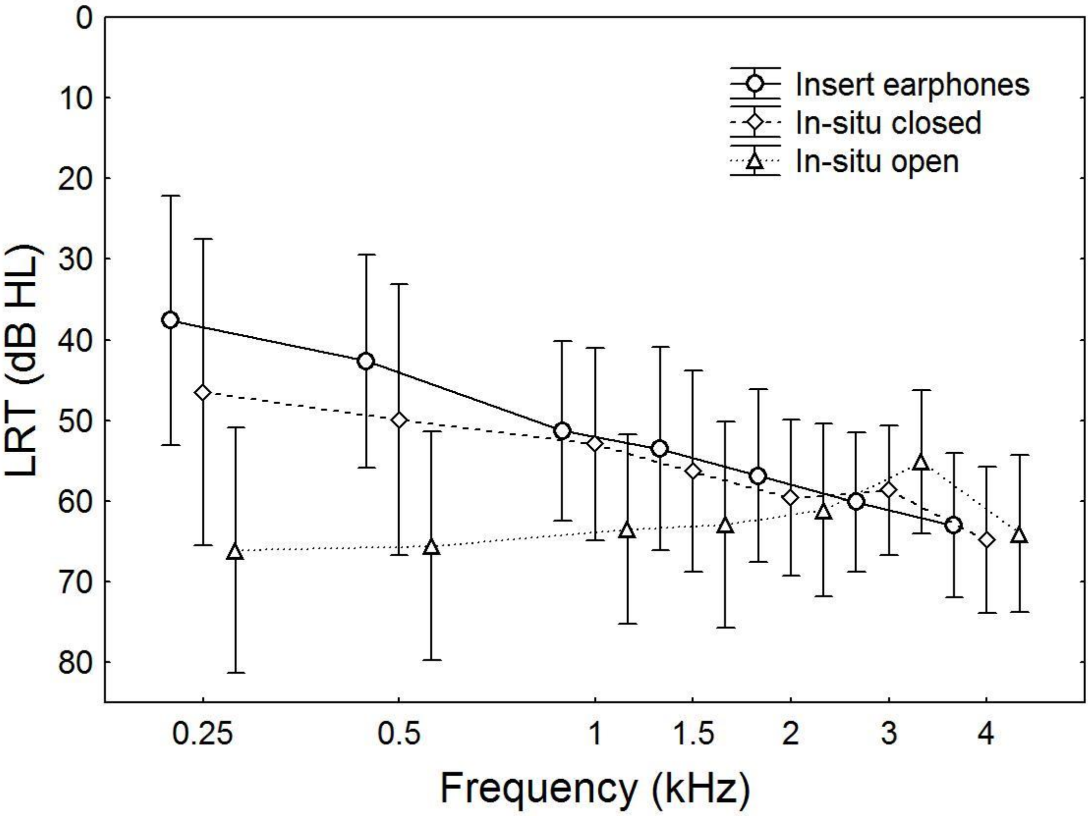
LRT (dB HL)
0
10
20
30
40
50
60
70
80
Insert earphones
In-situ closed
In-situ open
0.25
0.5
1
1.5
2
3
4
Frequency (kHz)

Figure 2

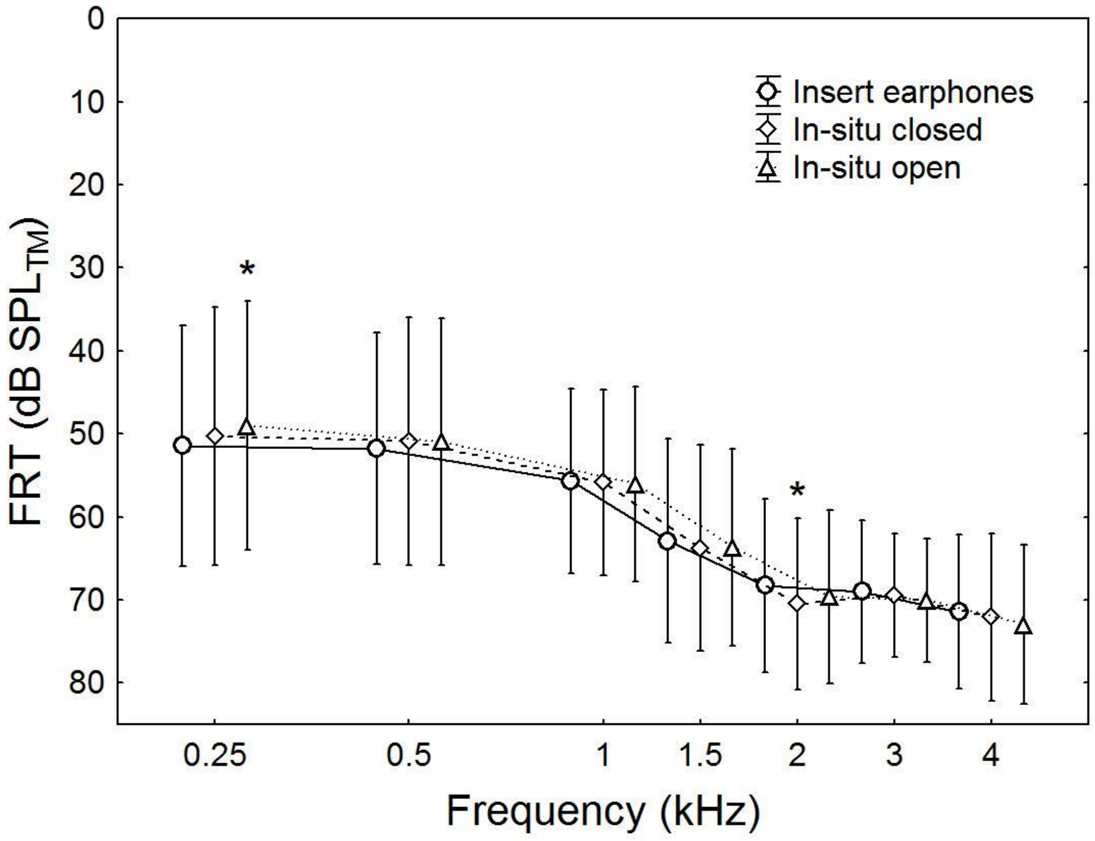
FRT (dB SPL$_{TM}$)
Frequency (kHz)
0
10
20
30
40
50
60
70
80
0.25
0.5
1
1.5
2
3
4
Insert earphones
In-situ closed
In-situ open
*
*

Figure 3

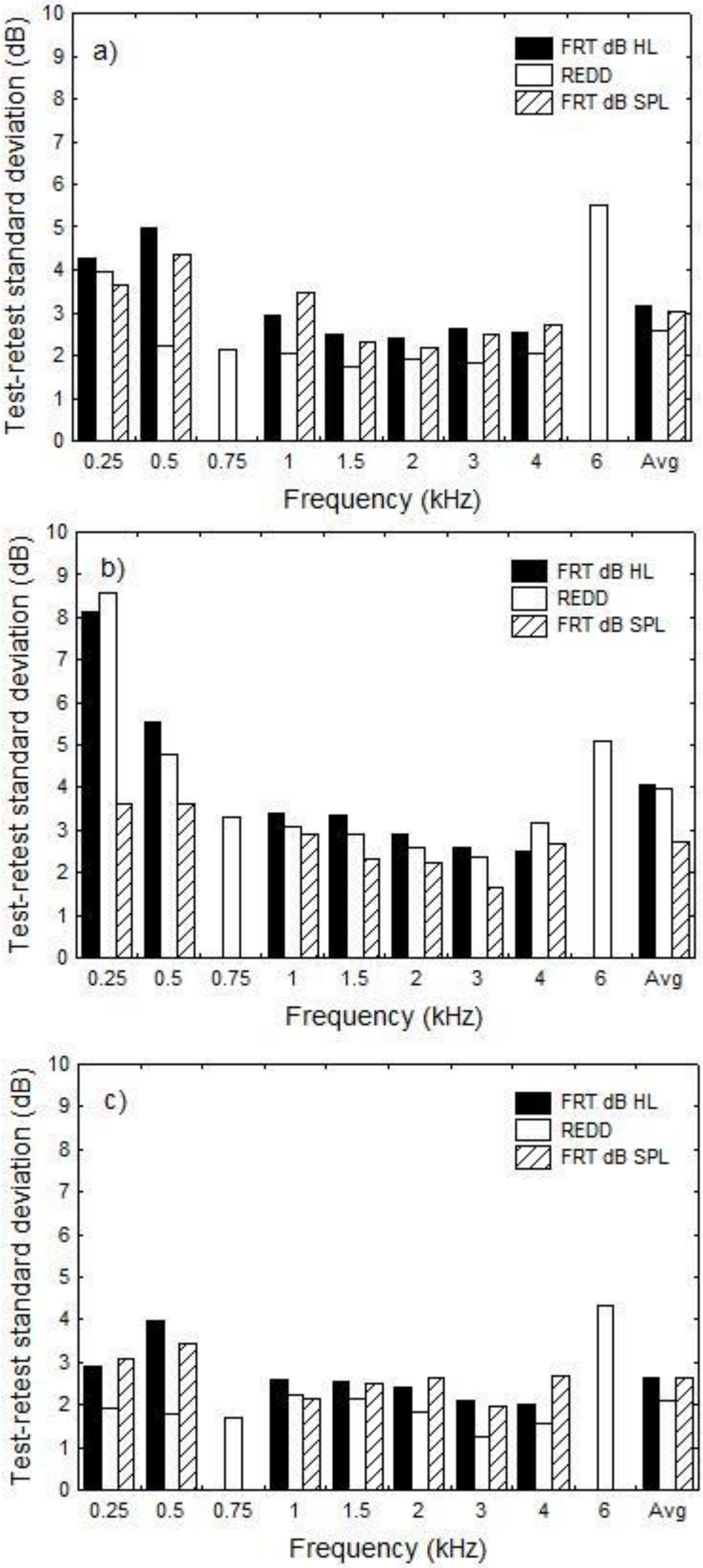
a)
b)
c)
FRT dB HL
REDD
FRT dB SPL
Test-retest standard deviation (dB)
0
1
2
3
4
5
6
7
8
9
10
0.25
0.5
0.75
1
1.5
2
3
4
6
Avg
Frequency (kHz)

Figure 4

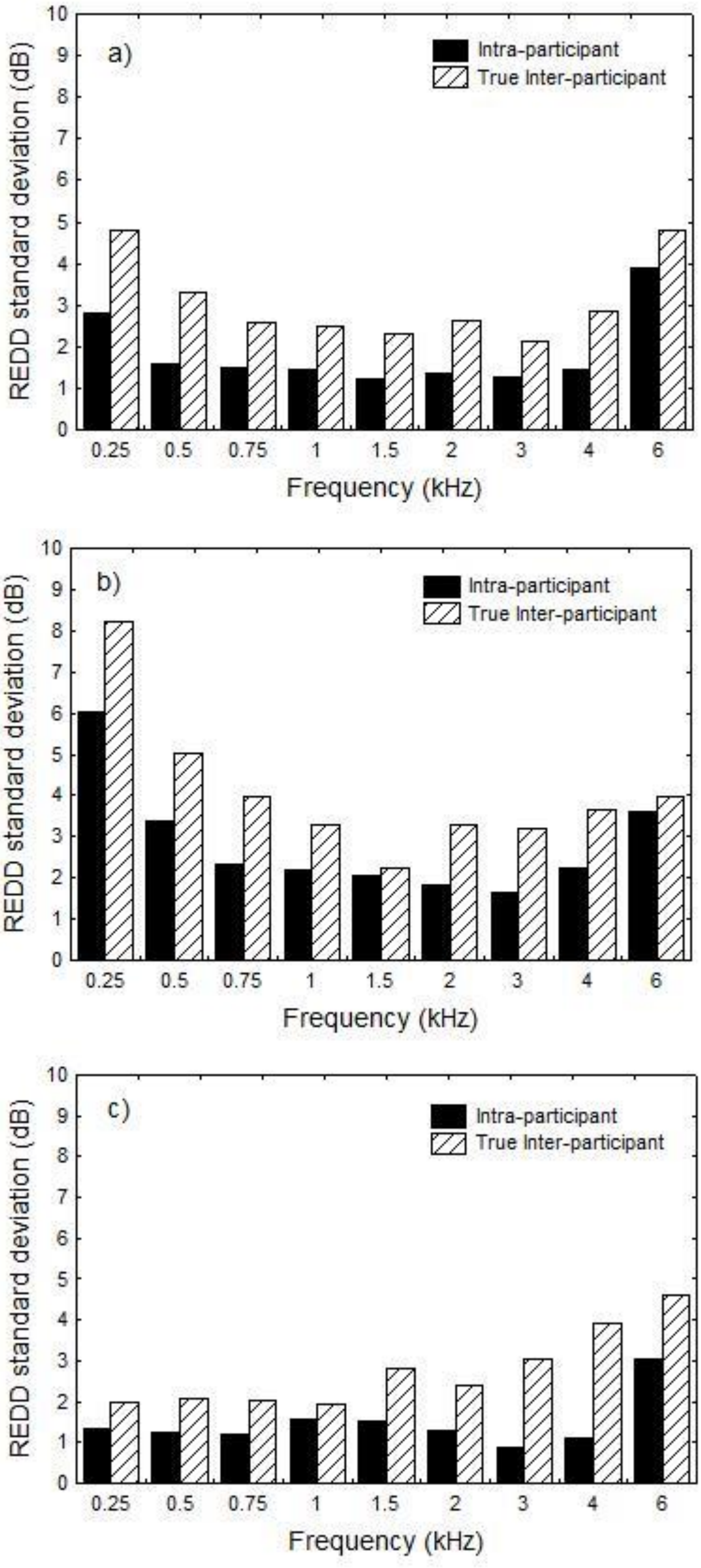
a)
b)
c)
Intra-participant
True Inter-participant
REDD standard deviation (dB)
0
1
2
3
4
5
6
7
8
9
10
0.25
0.5
0.75
1
1.5
2
3
4
6
Frequency (kHz)

Figure 5