Original Article

Reliability of real ear insertion gain in behind-the-ear hearing aids with different coupling systems to the ear canal

Charlotte Thunberg Jespersen & Kimi Nina Møller

GN ReSound A/S, Lautrupbjerg 7, 2750 Ballerup, Denmark

Abstract

Objective: The last decade has offered a multitude of instant fit coupling systems to be fitted with behind-the-ear (BTE) hearing aids. The impact of these designs on the reliability of real ear measurements (REMs) has not been reported. The purpose of this study was to obtain REM reliability data for instant fit coupling systems. **Design:** REM reliability data was obtained for four different instant-fit coupling systems and for standard size 13 tubing and custom earmolds. REMs were performed for all five coupling systems two times and by two examiners. **Study sample:** Ten normal-hearing individuals (20 ears) served as participants. **Results:** The REM test-retest reliability is high for the four instant fit coupling systems as well as for the custom earmolds. The REM inter-examiner reliability is high for three of the four instant fit coupling systems. **Conclusions:** Carrying out REMs with instant fit coupling systems appears to be fundamentally no different than performing REMs with conventional hearing aids. For either, care should be taken in probe tube placement in terms of insertion depth and maintaining the probe tube placement, and other best practices regarding test environment and test setup should be observed.

Key Words: Real ear measurements; reliability; instant fit coupling systems

Real ear measurements (REMs) provide clinicians with a valid and reliable way of verifying hearing-aid gain and output in situ. They are the only means by which the actual hearing aid and coupling system's gain or output in the ear of the individual hearing-aid user can be determined. Clinical applications of REMs are many and may include verification that prescriptive targets are met, documentation of the fitting, visual assistance in the fine-tuning process, counseling of the hearing-aid user and significant others, verification of signal processing algorithms, and as a baseline measure from which to work should the clinician need to troubleshoot hearing-aid user complaints related to sound processing features or general lack of benefit. One of the most compelling reasons to verify hearing-aid gain or output in situ is that benefit has been shown to be greater if the measured gain or output in the ear canal is well matched to the target (Baumfield & Dillon, 2001). In addition, REMs are part of best practice procedures for hearing-aid fittings today (British Society of Audiology and British Academy of Audiology guidelines, 2007; ISO 12124, 2001; Cox, 2005; Valente et al, 2006).

The last decade has offered a multitude of instant fit coupling systems to be fitted with BTE hearing aids. These instant fit coupling systems include thin tubes and receiver-in-the-ear designs combined with different dome designs (Kiessling et al, 2005; Alworth et al, 2010). The impact of these designs on the reliability of REMs has not been reported, but it is important for clinicians who fit hearing aids using these instant fit coupling systems to be aware of it. It is

anticipated that the reliability of REMs with these instant fit coupling systems is less as compared to the REM reliability with a standard size 13 tubing and a custom made earmold. The reasoning for this is that more variability is expected in how the instant fit coupling systems are placed in the ear canal relative to placement of custom made earmolds.

REM variability needs to be limited in order for the measurements to be useful in a clinical decision-making process, and should be smaller than changes resulting from modifications made to the hearing-aid fitting. Numerous studies on reliability of REMs have resulted in best practice recommendations for test set-up, equipment, and environment and procedures (Larson et al, 1977; Pedersen et al, 1982; Wetzell & Harford, 1983; Ringdahl & Leijon, 1984; Hawkins & Mueller, 1986; Dirks & Kincaid, 1987; Killion & Revit, 1987; Tecca, 1990; Ickes et al, 1991; Stone & Moore, 2004).

Inter-examiner reliability of REMs has been found to be high and independent of measurement equipment and hearing-aid type as long as a careful clinical procedure is followed (Hawkins, 1987; Valente et al, 1990; Hawkins et al, 1991).

Hawkins (1987) investigated the inter-examiner variability of real-ear-insertion-gain (REIG) using a BTE hearing aid combined with a soft earmold. Six examiners did each five measurements on one participant's ears using the Rastronics CCI-10 equipment in a fitting room and used their own clinical procedure. At 1000 and 2000 Hz the majority of the variability was within ± 3 to 4 dB. At 4000

Abbreviations

BTE	Behind-the-ear
IEC	International Electrotechnical Commission
ITC	In-the-canal
ITE	In-the-ear
REAR	Real-ear-aided-response
REIG	Real-ear-insertion-gain
REM(s)	Real ear measurement(s)
REUR	Real-ear-unaided-response
RITA	Receiver-in-the-aid
RITE	Receiver-in-the-ear
SPL	Sound pressure level

and 6000 Hz the variability increased, half of the REIG at 6000 Hz showed differences of more than 5 dB and 25% were greater than 10 dB. The author reported differences in the procedure, e.g. insertion depth of the probe tube, to be the reason for the low reliability in the high frequencies. Valente et al (1990) determined inter-examiner variability of REIG in in-the-canal (ITC) and in-the-ear (ITE) hearing aids. Two examiners jointly conducted measurements on either right or left ear of 28 participants, using the Frye 6500 unit and a careful procedure under anechoic conditions. The inter-examiner variability was insignificant and not reported. Hawkins et al (1991) investigated inter-examiner variability of real-ear-unaided-response (REUR), real-ear-aided-response (REAR), and REIG with a BTE hearing aid coupled to the ear with a Comply earmold. Measurements were conducted twice by each of the two examiners on 24 participants using the Rastronics CCI-10 and a careful clinical procedure under anechoic conditions. Inter-examiner variation was not reported specifically (combined with test-retest variation), but the reliability was reported to be high.

Test-retest reliability is also of importance, and is anticipated to be the largest source of variance when using instant fit coupling systems. Tecca (1990) refers to three different types of test-retest reliability: the immediate-, the short-term- and the long-term test-retest reliability. Immediate test-retest variability refers to differences found between measures completed consecutively within the same session and without removal of the earmold or hearing aid. Short-term test-retest variability refers to differences between measures conducted sequentially within the same session, but where the hearing aid or earmold and probe tube is removed and reinserted for subsequent measurements. Long-term test-retest variability refers to differences found between measures completed during two different sessions.

Multiple studies have been conducted on test-retest reliability of REIG. As can be seen in Table 1, studies have been conducted using different hearing-aid styles, different measurement equipment, and different measurement environments. The majority of studies have investigated the short-term test-retest reliability, but Tecca et al (1987) have also investigated the long-term test-retest reliability of REIG.

The short-term test-retest reliability of REIG conducted with ITC, ITE, and BTE hearing aids with conventional coupling systems has been shown to be high, but to decrease with frequency. The lowest test-retest reliability is reported by Hawkins (1987) and Humes et al (1988). Hawkins (1987) reports differences in the procedure, such as insertion depth of the probe tube, to be the reason for the low reliability. Valente et al (1990) mention a greater loudspeaker-to-test participant distance, the possible lack of precautions to control

tube movement between the unaided and aided measurements, and experience level of the examiner conducting the measurements as potential reasons for the low reliability in the study conducted by Humes et al (1988). A mean test-retest reliability of 2.0 dB standard deviation seems to be the short-term test-retest reliability reported in most studies independent of hearing-aid style, measurement equipment, and environment as long as a careful clinical procedure is followed. A mean long-term test-retest reliability of 3.3 dB standard deviation has been reported by Tecca et al (1987).

Tecca (1990) suggested the following clinical application of REM test-retest reliability. For immediate REM test-retest reliability, measurement differences up to 2 dB can be considered measurement variability and only differences in excess of 2 dB should be considered 'real' differences in the amplification. A clinical example of immediate test-retest measurements can be a clinician wanting to verify changes in hearing-aid settings without removing the device from the ear. For short-term test-retest variability, measurement differences in excess of 3–5 dB are suggested to be considered significant in clinical settings. A clinical example of short-term test-retest measurements can be verification that less low frequency gain or output has been achieved by modifying the vent in the hearing aid or earmold. For long-term variability, measurement differences in excess of 6 to 7 dB are considered clinically significant for clinicians. A clinician might want to do a long-term measurement comparison when the hearing-aid user complains of lack of benefit after having used the hearing aids for a longer period.

The purpose of this study was to obtain REM reliability data for instant fit hearing-aid coupling systems, and to compare these with measurement reliability using hearing aids fit with standard size 13 tubing and custom earmolds and following recommended clinical procedures. Specifically, three measures of reliability were examined: (1) reliability with the different coupling systems, (2) long-term test-retest reliability, and (3) inter-examiner reliability.

Method

Subjects

This study was carried out with a group of ten normal-hearing individuals (20 ears). Five individuals were male and five were female. The inclusion criteria for the participants was normal outer- and middle-ear conditions with a tympanic pressure within ± 50 daPa, a tympanic compliance between 0.3 and 1.6 mmho, and a tympanic volume between 0.6 and 2.0 ml. Tympanometry was done using a GN Otometrics Otoflex 100. Participants were not compensated for their participation.

Hearing aids and coupling systems

Three pairs of a commercially available receiver-in-the-aid (RITA) hearing aid (ReSound Live 71D devices), and two pairs of a receiver-in-the-ear (RITE) hearing aid (ReSound Live 70D RIE devices) were used for the REMs. All five pairs of hearing aids were fitted to amplify the IEC's standard moderate audiogram # 3 (Bisgaard et al, 2010).

The hearing aids were fitted using the Aventa 2.9.15.1 fitting software. Amplification was prescribed according to the manufacturer's proprietary fitting rule, Audiogram+, with the "experienced non-linear" profile selected. The Audiogram+ fitting rule prescribes insertion gain for narrowband signals at input levels of 50 and 80 dB SPL, so in this study the maximum real ear output has not exceeded 92 dB SPL at the eardrum level and can for this reason be considered acceptable for the normal-hearing participants. The hearing aids

Table 1. Overview of test-retest reliability of REIG results reported in the literature.

Short-term	Study	Hearing aids	Measurement system	Conditions AC = Anechoic conditions, FR = Fitting room		Test-retest differences	Confidence	Frequency range
Short-term	Hawkins (1987)	BTE coupled to a soft earmold	Rastronics CCI-10	FR		Standard deviation ranging from 2.2 to 7.0 dB		1000 to 6000 Hz
	Hawkins et al (1991)	BTE hearing aid coupled to the ear with a comply earmold	Rastronics CCI-10	AC		Standard deviation below 2.5 dB (mean standard deviation: 1.9 dB)	95% confidence intervals of ± 3 dB to ± 6 dB (mean 95% confidence interval: 3.8 dB)	1000 to 5000 Hz
	Humes et al (1988)	BTE hearing aid combined with a foam earplug	Frye 6500 real ear analyzer	AC		Standard deviation ranging from 1.0 to 6.0 dB (mean standard deviation: 3.5 dB)	95% confidence interval of ± 2 dB to ± 11.8 dB (mean 95% confidence interval: 6.7 dB)	500 to 4000 Hz
	Killion and Revit (1987)	BTE hearing aid combined with a foam earplug	Frye 6500 real ear analyzer	AC		Standard deviation of less than 3.5 dB (mean standard deviation: 1.7 dB)	95% confidence interval of ± 2.0 to ± 6.9 dB (mean 95% confidence interval: 3.3 dB)	200 to 6000 Hz
	Valente et al (1990)	ITC and ITE hearing aids	Frye 6500 real ear analyzer	AC		Standard deviations of less than 2.1 dB (mean standard deviation: 1.7 dB)	95% confidence interval of ± 4.1 dB (mean 95% confidence interval: 3.3 dB)	250 to 4000 Hz
Long-term	Tecca et al (1987)	BTE hearing aid combined with an individually fitted and fully occluding earmold	Starkey RE4 interfaced with the Fonix 5500Z hearing aid analyzer	AC		Standard deviation ranging from 0.6 to 4.0 dB (mean standard deviation: 2.0 dB)	95% confidence interval of ± 2 dB to ± 4.0 dB (mean 95% confidence interval: 3.9 dB)	200 to 4000 Hz
	Purdy et al (1989)	BTE hearing aid coupled to custom made earmold	Rastronics CCI-10	FR		Standard deviations ranging from 1.5 to 2.7 dB (mean standard deviation: 2.0 dB)	95% confidence interval of ± 2.8 dB to ± 5.3 dB (mean 95% confidence interval: 3.9 dB)	250 to 4000 Hz
	Tecca et al (1987)	BTE hearing aid combined with an individually fitted and fully occluding earmold	Starkey RE4 interfaced with the Fonix 5500Z hearing aid analyzer	AC		Standard deviations ranging from 1.4 to 6.0 dB (mean standard deviation: 3.3 dB)	95% confidence interval of ± 5 dB to ± 12.0 dB (mean 95% confidence interval: 6.4 dB)	200 to 4000 Hz

were fitted with an omnidirectional program with all features but the digital feedback suppression deactivated. The digital feedback suppression feature was calibrated and thereby activated in order to be able to do the measurements without feedback squeal. Digital feedback suppression is a pure feedback cancelling feature and no gain reduction was applied in any of the fittings.

The hearing-aid pairs were fitted with different coupling systems. The different coupling systems were a pair of RITA instruments and thin tubes combined with open domes (hereafter named RITA instruments with open domes), a pair of RITA instruments and thin tubes combined with tulip domes (RITA instruments with tulip domes), a pair of RITE instruments combined with open domes (RITE instruments with open domes), a pair of RITE instruments combined with double domes (RITE instruments with double domes), and a pair of RITA instruments and standard size 13 tubes combined with closed custom earmolds (RITA instruments with custom earmolds). Note that the terminology used to describe the different types of domes in this paper is consistent with that used by the manufacturer of the test hearing aids. The terms used to describe such domes differ among manufacturers. The thin tubes, receivers-in-the-ears, and domes were all of the appropriate sizes for the individual participants' ears.

Procedures

REMs were performed for all five hearing aid and coupling systems two times by two examiners. The two examiners were experienced in making REMs and knowledgeable concerning factors affecting reliability. The measurement order of the five different coupling methods as well as the order of the examiners doing the measurements was randomized. There was a median of one day (1st and 3rd quartile of one and six days) between the test-retest measurements.

Test room conditions, equipment, and participant orientation in the test room, probe tube calibration, and probe tube placement were consistent with the British Society of Audiology and British Academy of Audiology guidelines (2007). REMs were obtained in a quiet fitting room complying with the ISO 12124 (2001) standard and using an Aurical Plus from GN Otometrics that had been calibrated within the last 12 months prior to the measurements. The Aurical Plus complies with the IEC 61669 (2001) standard.

The loudspeaker was positioned with no large reflecting surfaces near the participant. The loudspeaker and the reference microphone when positioned on the participant were both one meter away from the nearest reflective surface.

Probe tube calibration was performed with beveled end probe tubes and by placing the probe tube so that it was close to the reference microphone aperture, without blocking either of the microphones. Probe tube calibration was done with the headset placed on the ears of the participant, who was seated 0.5 meters in front of the loudspeaker with the microphone and probe tube facing the loudspeaker at an angle of 0° azimuth. A probe tube calibration using a 65 dB SPL warble tone sweep stimulus was obtained every time a new probe tube was used, and was as a minimum done for each participant.

For REMs, the participants were seated in a chair at a distance of 0.5 meters facing (0 degrees azimuth) the loudspeaker and at a level with the center of the loudspeaker itself. The participants were told to visually fixate on a location on the loudspeaker during all measurements, but their head was not held stationary. The participants were asked to sit quietly during the measurements.

Correct insertion depth with the probe tube extending at least 5 mm beyond the tip of the domes and earmolds and being 6 mm or

less from the tympanic membrane was ensured in order to measure at the level of the tympanic membrane and to avoid standing waves affecting the measured SPL in the ear canal. Correct insertion depth was ensured by a dual approach. Initially, a visual positioning, using a marker length of 28 mm from the end of the probe tube for females and 30 mm from the end of the tube for males was used, and it was measured that the length of the probe tube from the marker to the tip of the probe tube ensured that the tip of the probe tube extended 5 mm past the sound outlet of the domes and earmolds. The marker was placed at the tragus once the probe tube was inserted. Secondly, an acoustic positioning check at 6000 Hz was done for the REUR, checking that the unaided gain was not below -5 dB. If the unaided gain at 6000 Hz was below -5 dB, the probe tube length was adjusted and/or the probe tube was repositioned and the measurement repeated until an acceptable measurement was obtained.

The measurement method used was the modified pressure method with stored equalization, as the hearing aids were fitted with a digital feedback suppression feature and as most of the coupling systems were open (Larsby & Arlinger, 1988; Hawkins & Mueller, 1992; Lantz et al, 2007). A concern with the modified pressure method with stored equalization is that movement of the participant's head can lead to changes in the sound field and thus to errors in REIG (Shaw, 2010). Aazh et al (2012) have shown that the modified pressure method with stored equalization can be used if the participants can keep their head still during the measurements.

The REUR was measured using a 65 dB SPL warble tone sweep stimulus. Although use of this signal is not representative of current clinical practice, this choice may have reduced variability due to small differences in the signal level or spectrum between measurements or sessions. For broadband signals, the measurement equipment adjusts the overall presentation level based on the level at the reference microphone but does not ensure that the desired signal spectrum is obtained. When a narrowband signal is used, the level of each frequency presented is controlled. Then the coupling system and hearing aid was carefully inserted into the ear. The real-ear-occluded-response was measured with the hearing aid turned off and with a warble tone sweep stimulus and it was checked that the gain/output was not reduced much in the low frequencies, if this was the case then the hearing aid was repositioned and the probe tube was checked for wax/moisture. The hearing aid was turned on and the REAR was measured. The measurements were conducted at 129 frequencies from 200 Hz to 8000 Hz. REIG was calculated by the measurement equipment, and this data was used for analysis. Data were electronically exported from the Aurical Plus software to a spreadsheet for analysis.

Results

The long-term REM test-retest and inter-examiner reliability of the different instant fit coupling systems was compared with the measurement reliability using RITA hearing aids fit with custom earmolds.

A multivariate analysis of variance (MANOVA) was used to assess the effect of factors and interactions on several response variables (129 frequencies). Three fixed factors were investigated: type of coupling systems (X1), examiner (X2) and number of measurements performed by the same examiner (X4). Participants are a random effect (X3). Two-way interactions between the fixed effects were included in the MANOVA. The mean response across right and left ear was analysed in the MANOVA. Multivariate normality was verified with the help of a Q-Q plot of the standardized residuals against

Table 2. MANOVA table. Columns denote the sources of variation (Coupling system = X1, Examiner = X2, Participants = X3, Number of measurement performed by the same examiner = X4), degrees of freedom (DF), Pillai, approx F, num Df, den Df and the probability of observing the F statistics under the null hypothesis ($Pr(> F)$).

Source	DF	Pillai	Approx F	num Df	den Df	$Pr(> F)$
X1: Coupling system	4	3.5631	3.1614	516	200	<2.2e-16
X2: Examiner	1	0.8358	1.8549	129	47	0.008455
X3: Participants	9	8.0484	3.6059	1161	495	<2.2e-16
X4: Number of measurement performed by same examiner	1	0.7106	0.8945	129	47	0.692476
X1*X2	4	2.7584	0.8611	516	200	0.903060
X1*X4	4	2.6384	0.7511	516	200	0.993685
X2*X4	1	0.8339	1.8295	129	47	0.009710
Residuals	175					

The “*” denotes interaction between factors.

their expected values under normality. The assumption holds. There were ten participants (20 ears) in each “Coupling system *Examiner *Participants *Number of measurement performed by the same examiner” combination. The “*” denotes interaction between factors. The results are summarized in Table 2. Whenever significant and meaningful effects were found, a Bonferroni correction was used to locate differences.

The approximate F statistics show significant differences on two of the four independent factors. The measured REIG values differ significantly with different coupling systems ($F = 3.16$, $p < 0.0001$). The measured REIG values obtained by different examiners differ significantly ($F = 1.85$, $p = 0.0084$). The measured REIG values do not differ significantly between the different measurements done by the same examiner ($F = 0.89$, $p = 0.6924$).

The interaction between the coupling system and examiner ($X1*X2$) is not significant ($F = 0.86$, $p = 0.9030$), indicating that the differences between the coupling systems are not due to differences between the measurements done by the two examiners. The interaction between coupling system and number of measurement done by the same examiner ($X1*X4$) is not significant ($F = 0.75$, $p = 0.9936$), indicating that the differences between the coupling systems are not due to differences between the different measurements done by the same examiner. The interaction between examiner and number of measurement ($X2*X4$) is significant ($F = 1.82$, $p = 0.0097$), but the

magnitude of the detected differences is negligible, therefore even though significant differences were detected, they have no clinical interest.

To limit the risk of false positive results, a Bonferroni correction was applied. Only significant coupling system differences at the corrected Bonferroni significance level were found.

The mean absolute long-term test-retest differences across frequencies with standard deviations for test-retest differences are shown for all four instant fit coupling systems compared to the mean absolute long-term test-retest differences with standard deviations for test-retest differences for the custom earmolds across frequencies in Table 3. The measurements for the two examiners have been combined. The table illustrates intra-frequency differences for the different coupling systems.

The mean absolute inter-examiner differences across frequencies with standard deviations for inter-examiner differences are shown for all four instant fit coupling systems and for the custom earmolds across frequencies in Table 4. The first and second measurements for the two examiners have been combined as no significant test-retest differences were found with any of the coupling systems. The table illustrates intra-frequency differences for the different coupling systems.

The mean REIG across frequencies for all four instant fit coupling systems compared to the mean REIG for the custom earmolds is

Table 3. Mean absolute test-retest differences in dB across frequencies with standard deviations for test-retest differences.

Frequency (Hz)	RITA w/earmold		RITA w/tulip dome		RITA w/dome		RITE w/double dome		RITE w/dome	
	Test-retest diff	SD	Test-retest diff	SD	Test-retest diff	SD	Test-retest diff	SD	Test-retest diff	SD
250	0.8	± 0.9	0.3	± 0.3	0.3	± 0.2	0.3	± 0.4	0.2	± 0.2
500	1.4	± 1.8	0.5	± 0.8	0.2	± 0.2	0.7	± 0.9	0.3	± 0.2
750	1.6	± 2.1	1.7	± 1.9	0.5	± 0.4	1.4	± 1.5	1.0	± 0.9
1000	1.3	± 1.7	2.0	± 2.2	0.8	± 0.6	1.8	± 1.6	1.3	± 1.1
1500	1.0	± 1.1	1.8	± 1.4	1.1	± 1.0	1.4	± 1.3	1.7	± 1.4
2000	1.3	± 1.1	1.8	± 1.3	1.1	± 1.1	1.3	± 1.3	1.6	± 1.3
3000	1.3	± 0.9	1.8	± 1.4	1.1	± 1.1	1.0	± 0.8	1.1	± 0.9
4000	1.1	± 1.0	1.8	± 1.5	1.2	± 1.4	1.1	± 0.8	1.2	± 0.8
5000	1.7	± 1.2	2.2	± 1.8	1.5	± 1.5	1.8	± 1.3	1.4	± 1.2
6000	1.9	± 1.5	2.4	± 1.9	2.3	± 2.1	2.2	± 1.6	2.3	± 1.7
8000	3.4	± 2.7	3.9	± 2.7	3.2	± 2.8	3.9	± 2.9	2.5	± 1.8

Table 4. Mean absolute inter-examiner differences in dB across frequencies with standard deviations for inter-examiner differences.

Frequency (Hz)	RITA w/earmold		RITA w/ tulip dome		RITA w/ dome		RITE w/ double dome		RITE w/ dome	
	Inter-examiner diff	SD	Inter-examiner diff	SD	Inter-examiner diff	SD	Inter-examiner diff	SD	Inter-examiner diff	SD
250	0.0	± 0.1	0.1	± 0.1	0.0	± 0.0	0.1	± 0.2	0.0	± 0.0
500	0.3	± 0.6	0.1	± 0.6	0.1	± 0.1	1.2	± 1.4	0.1	± 0.2
750	0.1	± 0.4	0.6	± 0.9	0.0	± 0.1	2.4	± 1.5	0.3	± 0.1
1000	0.2	± 0.2	0.9	± 1.1	0.1	± 0.2	2.4	± 0.8	0.5	± 0.2
1500	0.1	± 0.2	0.6	± 0.7	0.1	± 0.1	0.4	± 0.3	0.5	± 0.4
2000	0.2	± 0.2	0.4	± 0.4	0.3	± 0.5	0.6	± 0.2	0.3	± 0.5
3000	0.1	± 0.2	0.7	± 0.0	1.0	± 0.1	0.3	± 0.4	0.0	± 0.3
4000	0.1	± 0.1	0.9	± 0.0	1.3	± 0.3	0.3	± 0.3	0.0	± 0.2
5000	0.1	± 0.1	1.5	± 0.4	2.0	± 0.5	0.1	± 0.2	0.2	± 0.1
6000	0.0	± 0.1	1.1	± 0.7	2.1	± 1.0	1.0	± 0.1	1.5	± 0.1
8000	0.6	± 1.0	0.7	± 0.7	0.0	± 1.3	0.5	± 2.0	1.4	± 0.4

shown in Table 5. The mean has been calculated based on all four measurements.

Discussion

It was hypothesized that greater variability in how instant fit coupling systems can be placed in the ear canal relative to BTE fit with standard tubing and custom earmolds would result in lower reliability of REMs. This was mostly not supported by the current study. Except for small differences in one condition which are judged to have negligible clinical relevance, reliability was found to be equivalent across measurement conditions, between tests, and between the two examiners. In addition, reliability was as high or higher than that reported in the literature.

The long-term test-retest reliability is high for the four different instant fit coupling systems as well as for the custom earmold coupling systems, as long as a careful clinical procedure is followed. This means that the measurement variability between measures completed with the same hearing aid, hearing-aid settings and in the same ear but during two different sessions is small. This result supports that long-term test-retest reliability is high provided adherence to careful clinical procedures for probe tube placement and stability of probe tube placement between measurements (Dirks & Kincaid, 1987; Hawkins & Mueller, 1986). This result is in agreement with previous studies on both short-term and long-term test-retest

reliability, which have documented high test-retest reliability with different types of hearing aids when a careful clinical procedure was followed (Ringdahl & Leijon, 1984; Hawkins, 1987; Tecca et al, 1987; Killion & Revit, 1987; Valente et al, 1990; Hawkins et al, 1991). The data from the present study shows that high long-term test-retest reliability also applies to REMs conducted on RITA and RITE instruments with different instant fit coupling systems.

The standard deviations for long-term test-retest differences are all at or below ± 2.9 dB in this study and thereby smaller than the long-term standard deviations in the study by Tecca et al (1987). As clinically significant differences are often considered to be 2 SD, this means that measurement differences in excess of ± 6 dB can be considered clinically significant.

The test-retest reliability in this study was determined by one test measurement and one retest measurement (two REIG measurements) as in the studies by Tecca et al (1987), Humes et al (1988), Purdy et al (1989), Valente et al (1990), and Hawkins et al (1991). More retest measurements may have been desirable as in the studies by Hawkins (1987) and Killion and Revit (1987), wherein four retest measurements (five REIG measurements) were conducted. However, the test-retest reliability reported in the different studies is very close regardless of the number of retest measurements, with the exception of Humes et al (1988) and Hawkins (1987), where lower test-retest reliability likely was due to procedural factors. The high test-retest reliability found in this study is for this reason considered to be valid. REM reliability across frequencies was not the objective of this study, but in agreement with findings from previous studies (Larson et al, 1977; Wetzell & Harford, 1983; Ringdahl & Leijon, 1984) data from the present study (Table 3) shows greater variability for REMs at higher frequencies. Basic physical principles determine that at higher frequencies a larger accuracy in positioning the probe tube inside the ear canal is needed in order to measure the correct gain. Adherence to a careful clinical procedure is likely to continue to become even more important as hearing aids with broader frequency ranges are developed and REMs will be conducted up to higher frequencies than the 8000 Hz in this study and the 4000 and 6000 Hz in previous studies.

The data in Table 3 show a tendency for larger variability in the low frequencies with the RITA instruments with custom earmolds as compared to any of the instant fit coupling conditions. The difference between the RITA instruments with custom earmolds and the RITE instruments with double domes is smaller than with the other instant fit coupling systems. This tendency is likely to be due

Table 5. Mean REIG across frequencies for all four instant fit coupling systems compared to the mean REIG for the custom earmolds.

Frequency (Hz)	RITA w/earmold	RITA w/tulip dome	RITA w/dome	RITE w/double dome	RITE w/dome
	REIG	REIG	REIG	REIG	REIG
250	2.2	0.4	0.2	-0.2	-0.2
500	4.5	0.8	0.2	2.3	0.8
750	10.4	3.1	1.1	5.6	3.2
1000	14.5	5.4	2.5	10.9	8.5
1500	10.5	6.4	4.2	15.2	14.1
2000	9.0	5.7	6.6	14.5	16.6
3000	6.7	7.5	9.9	15.7	18.3
4000	6.2	7.4	8.8	10.8	13.7
5000	7.2	10.1	11.8	9.1	10.6
6000	10.2	13.8	14.9	13.5	12.5
8000	3.1	8.1	9.3	13.7	10.9

to greater variation in the way that the custom-made earmolds fit in each ear, and the amount of slit-leakage that may have occurred for each individual. The open domes and tulip domes provide virtually the same venting conditions in all normal adult ear canals. Therefore the response is not predominated by the electroacoustic gain but rather by the direct sound entering the ear canal. For the double dome coupling system, the degree of occlusion in individual ear canals can be expected to be more variable than the open and tulip domes. This would explain why the variability in this condition is greater than with the custom earmolds, but less than with the less occluding instant fit coupling systems.

High inter-examiner reliability of REMs is desirable, as it means that REMs can be conducted by different audiologists without it affecting the measurement result. The REM inter-examiner variability is on the order of 1.5 dB, which can be considered a high reliability for three of the four instant fit coupling systems: the RITA instruments with tulip domes, the RITE instruments with open domes, the RITE instruments with double domes as well as with the RITA instruments with custom earmolds. Slightly lower reliability with variability on the order of 2 dB at 5000 and 6000 Hz was found with the RITA instruments with open domes. This difference is statistically significant, and also higher than for custom earmolds, but is well within guidelines for clinical significance. The inter-examiner reliability results should be interpreted with caution due to the fact that there were only two examiners. Other studies with only two examiners also reported high inter-examiner reliability (Valente et al, 1990; Hawkins et al, 1991), while a greater number of examiners were associated with higher inter-examiner variability that may have been due to procedural factors (Hawkins et al, 1987). The inter-examiner reliability with instant fit coupling systems therefore merits further investigation.

The REIG obtained with the different coupling systems was significantly different, which is expected as the coupling systems affect the output of the systems differently e.g. thin tube versus thick tube, and dome versus earmold.

As when performing REMs with hearing aids having custom-fit earpieces, the same care should be taken in probe tube placement in terms of insertion depth and maintaining the probe tube placement between the REUR and REAR measurements in order to ensure that measurement differences are resulting from modifications made to the hearing-aid fitting rather than a different insertion depth of the instant fit coupling system during different measurements, or a failure in keeping the probe tube placement constant during REUR and REAR measurements.

Conclusions

The REM long-term test-retest reliability is high for the four instant fit coupling systems as well as for the custom earmolds. The REM inter-examiner variability is on the order of 1.5 dB, which can be considered a high reliability for three of the four instant fit coupling systems: the RITA instruments with tulip domes, the RITE instruments with open domes, the RITE instruments with double domes, as well as with the RITA instruments with custom earmolds. Slightly lower reliability with variability on the order of 2 dB is found at 5000 and 6000 Hz with the RITA instruments with open domes. This difference is statistically significant, and also higher than for custom earmolds, but is well within guidelines for clinical significance. In conclusion, carrying out REMs with instant fit coupling systems appears to be fundamentally no different than performing REMs with conventional hearing aids. For either, care should be taken in probe tube placement in terms of insertion depth and maintaining

the probe tube placement, and other best practices regarding test environment and test setup should be observed.

Acknowledgements

The authors would like to thank Jennifer Groth for providing input to the manuscript and Tobias Piechowiak for help with the statistical analysis of the data.

Declaration of interest: Charlotte Thunberg Jespersen and Kimi Nina Møller are employees of GN Resound A/S. The authors alone are responsible for the content and writing of the paper.

References

- Alworth L.N., Plyler P.N., Reber M.B. & Johnstone P.M. 2010. The effects of receiver placement on probe microphone, performance, and subjective measures with open canal hearing instruments. *J Am Acad Audiol*, 21, 249–266.
- Aazh H., Moore B.C.J. & Prasher D. 2012. Real ear measurement methods for open fit hearing aids: Modified pressure concurrent equalization (MPCE) versus modified pressure with stored equalization (MPSE). *Int J Audiol*, 51, 103–107.
- Baumfield A. & Dillon H. 2001. Factors affecting the use and perceived benefit of ITE and BTE hearing aids. *Brit J Audiol*, 18, 31–38.
- Bisgaard N., Vlaming M.S.M.G. & Dahlquist M. 2010. Standard audiograms for the IEC 60118-15 measurement procedure. *Trends Amplif*, 14, 113–120.
- British Society of Audiology and British Academy of Audiology. 2007. Guidance on the use of real ear measurements to verify the fitting of digital signal processing hearing aids. http://www.thebsa.org.uk/index.php?option=com_content&view=category&layout=blog&id=7&Itemid=16
- Cox R.M. 2005. Evidence-based practice in provision of amplification. *J Am Acad Audiol*, 16, 419–438.
- Dirks D.D. & Kincaid G.E. 1987. Basic acoustic considerations of ear canal probe measurements. *Ear Hear*, 8, 60–67.
- Hawkins D. & Mueller H. 1986. Some variables affecting probe-tube microphone measurements. *Hear Instrum*, 37, 8–12, 49.
- Hawkins D.B. 1987. Variability in clinical ear canal probe microphone measurements. *Hear Instrum*, 38, 30–32.
- Hawkins D.B., Alvarez E.D. & Houlihan J.R. 1991. Reliability of three types of probe tube microphone measurements. *Hear Instrum*, 42, 14–16.
- Hawkins D.B. & Mueller H.G. 1992. Procedural considerations in probe-microphone measurements. Mueller, Hawkins & Northern (eds.) *Probe Microphone Measurements. Hearing Aid Selection and Assessment*. San Diego: Singular Publishing Group, pp. 67–89.
- Humes L.E., Hipskind N.M. & Block M.G. 1988. Insertion gain measured with three probe tube systems. *Ear Hear*, 9, 108–112.
- Ickes M.A., Hawkins D.B. & Cooper W.A. 1991. Effect of reference microphone location and loudspeaker azimuth on probe tube microphone measurements. *J Am Acad Audiol*, 2, 156–165.
- IEC 61669. 2001. Electroacoustics: Equipment for the measurement of real-ear acoustical characteristics of hearing aids. Geneva: International Electrotechnical Commission.
- ISO 12124. 2001. Acoustics: Procedures for the measurement of real-ear acoustical characteristics of hearing aids. Geneva: International Organization for Standardization.
- Kiessling J., Brenner B., Jespersen C.T., Groth J. & Dyrland Jensen O. 2005. Occlusion effect of earmolds with different venting systems. *J Am Acad Audiol*, 16, 239–251.
- Killion M.C. & Revit L.J. 1987. Insertion gain repeatability versus loudspeaker location: You want me to put my loudspeaker WHERE? *Ear Hear*, 8, 68–73.
- Lantz J., Jensen O.D., Hastrup A. & Olsen S.Ø. 2007. Real-ear measurement verification for open, non-occluding hearing instruments. *Int J Audiol*, 46, 11–16.

- Larsby B. & Arlinger S. 1988. Comparison of different types of equipment for insertion gain measurement. Jensen (ed.), *Hearing Aid Fitting. Theoretical and Practical Views*. Copenhagen: 13th Danavox symposium, pp. 39–47.
- Larson V.D., Studebaker G.A. & Cox R.M. 1977. Sound levels in a 2 cc cavity, a Zwislocki coupler, and occluded ear canals. *J Am Audiol Soc*, 3, 63–70.
- Pedersen B., Lauridsen O. & Nielsen B. 1982. Clinical measurements of hearing aid insertion gain. *Scand Audiol*, 11, 181.
- Purdy S., Dodd D. & Keith W. 1989. Insertion gain measurements: Technical and methodological considerations. *Hear Instrum*, 40, 24–26.
- Ringdahl A. & Leijon A. 1984. The reliability of insertion gain measurements using probe microphones in the ear canal. *Scand Audiol*, 13, 173–178.
- Shaw P. 2010. Are real-ear measurements (REM) accurate when using the modified pressure with stored equalization (MPSE) method? *Int J Audiol*, 49, 463–466.
- Stone M.A. & Moore B.C.J. 2004. Estimated variability of real-ear insertion response (REIR) due to loudspeaker type and placement. *Int J Audiol*, 43, 271–275.
- Tecca J., Woodford C. & Kee D. 1987. Variability of insertion-gain measurements. *Hear J*, 2, 18–20.
- Tecca J. 1990. Clinical application of real ear probe tube measurement. In: Sandlin (ed.), *Handbook of Hearing Aid Amplification, Vol II: Clinical Considerations and Fitting Practices*. Boston: College-Hill Press, pp. 225–255.
- Valente M., Meister M., Smith P. & Goebel J. 1990. Intratester test-retest reliability of insertion gain measures. *Ear Hear*, 11, 181–184.
- Valente M., Abrams H. & Benson D. 2006. Guidelines for the audiological management of adult hearing impairment. *Audiol Today*, 18, 32–36.
- Wetzell C. & Harford E.R. 2003. Predictability of real ear hearing-aid performance from coupler measurements. *Ear Hear*, 4, 237–242.

Advances in Audiology: Tomorrow's Solutions for Today's Challenges

Catch up on what was presented and discussed at the latest clinical conference sponsored by Phonak.

"Advances in Audiology: Tomorrow's Solutions for Today's Challenges", Las Vegas, 3 – 5 December 2012.

The latest developments in the topical areas "Innovation in Hearing Care for Adults" and "Latest Developments in FM and Soundfield" were discussed by renowned speakers. The conference was chaired by Louise Hickson (AUS), Linda Thibodeau (USA) and Cheryl DeConde Johnson (USA).

To access live recordings of the conference presentations visit

www.phonakpro.com/advances-in-audiology

PHONAK
life is on