Clinical Validation of the AMTAS Automated Audiometer

Robert H Eikelboom[1,2,3], De Wet Swanepoel[1,2,3], Shahpar Motakef[1,2,4], Gemma S Upson[1,2]

1. Ear Science Institute Australia, Subiaco, Australia
2. Ear Sciences Centre, School of Surgery, The University of Western Australia, Nedlands, Australia
3. Department of Communication Pathology, University of Pretoria, Pretoria, South Africa
4. Department of Otorhinolaryngology and Head & Neck Surgery, Witwatersrand University, Johannesburg, South Africa

Acronyms and abbreviations:
AMTAS: Automated Method for Testing Auditory Sensitivity
dB: decibel
Hz: Hertz
SD: Standard deviation

Address for correspondence:
Adj Prof RH Eikelboom, Ear Science Institute Australia, Suite 1 Level 2, 1 Salvado Rd, Subiaco WA 6008
Email: rob.eikelboom@earscience.org.au

Abstract

Objective: To validate the air and bone conduction AMTAS automated audiometry system.
Design: Prospective study. Test-retest reliability was determined by assessing adults with AMTAS air and bone conduction audiometry. Accuracy was determined by comparing AMTAS and manual audiometry conducted on adults. AMTAS testing was conducted in a quiet room and manual audiometry in a sound booth.
Study Sample: Ten participants for test-retest reliability tests and 44 participants to determine accuracy were included. Participants had varying degrees of hearing loss.
Results: For test-retest reliability the overall difference in air conduction hearing thresholds (n=119) was 0.5dB. The spread of differences (standard deviation of absolute differences) was 4.9dB. For bone conduction thresholds (n=99) the overall difference was -0.2dB, and the spread of differences 4.5dB. For accuracy the overall difference in air conduction hearing thresholds (n=509) between the two techniques was 0.1dB. The spread of differences was 6.4dB. For bone conduction thresholds (n=295) the overall difference was 0dB, and the spread of differences 7.7dB.
Conclusions: Variations between air and bone conduction audiometry for automated and manual audiometry were within normally accepted limits for audiometry. However, AMTAS thresholds were elevated but not significantly different compared to other contemporary studies that included an automated audiometer.

Keywords: automated audiometry, air conduction, bone conduction, audiometry

Introduction

There is an increasing demand worldwide for more accessible hearing services to overcrowded clinics and remote areas (Swanepoel et al., 2010). The expense of training more audiologists in general and intrinsic issues related to living in remote areas including social and professional isolation mean the problem of under-servicing is likely to persist if not increase (Goulios & Patuzzi, 2008). It has been estimated that in the United States alone in 2000 there were 8 million more audiograms needed than could be managed by the available audiologists, and this number will increase to more than 15 million by 2050 (Margolis & Morgan, 2008).

The protocol for pure-tone audiometry is well defined, and therefore logically suited to being automated. Although no automated audiometer can replace an audiologist, a system that can determine pure-tone hearing thresholds with similar accuracy to that of manual audiometry may be very beneficial to address the demand for hearing health services. Optimizing limited professionals time by utilizing automation may improve overall effectiveness in hearing health care (Swanepoel et al., 2010).

Automated pure-tone audiometry is no new phenomenon however. In 1947, Georg von Békésy (von Békésy, 1947) was the first to develop an automated audiometer to measure hearing thresholds with frequency and intensity sweep stimuli. Since then many such instruments have been developed, particularly for screening as part of hearing conservation programmes. For the most part these self-recording hearing tests rarely measure bone conduction and are not suitable for diagnostic audiology.

The Automated Method for Testing Auditory Sensitivity (AMTAS) is a prototype computer-based audiometer capable of both air and bone conduction testing with masking to the non-tested ear and incorporated quality control features (Margolis et al., 2007). It is suitable for those patients who are able to follow instructions in conventional manual audiometry. Several validation studies on this system have been published in recent years by the developers and their collaborators (Margolis et al., 2010; Margolis et al., 2011; Margolis & Moore, 2011). The reported evidence indicates that AMTAS compared to manual audiometry thresholds are similar to inter-tester variability in air conduction but slightly higher for bone conduction. This was partly attributed to reference equivalent threshold force levels on forehead and mastoid bone conduction, specifically at 4000 Hz, that may be incorrect (Margolis, Glasberg et al., 2010; Margolis & Moore, 2011). The authors have argued that this is a problem in the standards, not the application of the standards by the AMTAS.

Limitations of published validation studies in automated audiometry include omission of bone conduction testing in some cases and a lack of patients with hearing loss included in study samples (Swanepoel, Mngemane et al., 2010). Even if hearing losses are included, the descriptions and influence of the hearing loss on outcomes are not reported (Margolis, Glasberg et al., 2010; Margolis, Frisina et al., 2011). The purpose of this study was to independently assess the accuracy and reliability of AMTAS on a cohort of clinical subjects with a range of hearing losses, and to compare the findings to previous AMTAS studies and a selection of other studies on the accuracy and reliability of automated and manual audiometry.

Materials and Methods

Validity, in the context of audiometry refers to the ability of a new method to measure what it is supposed to, which means the method must be compared with a well-established one in terms of accuracy and test-retest reliability (Dobie, 1983). For pure-tone threshold determination, any system or technique to be validated has to be compared with conventional manual pure-tone audiometry as the gold standard (Burns & Hinchcliffe, 1957). For this reason, we performed both manual pure tone audiometry and AMTAS automated audiometry on the same subjects to compare the hearing thresholds obtained with the two methods. Test-retest reliability refers to the ability of a test to give similar results when applied more than once on the same subject under the same conditions (Dobie, 1983). To assess the reliability, we repeated the test with a number of subjects with the AMTAS hearing test on the same subjects under similar circumstances.

Apparatus

AMTAS is a computer programme that automates the standard procedure for established pure-tone air and bone conduction thresholds. A module in the Noah platform is utilised to control an Auricle/Conera audiometer (GNOtometrics, Copenhagen, Denmark) via a USB connection. A touch-screen monitor is used to provide written instructions to the subject and receive input from the subject. Sennheiser HDA 200 (Wennebostel, Germany) circumaural transducers are used, providing a level of ambient noise attenuation (approximately 2 to 3dB). This potentially allows testing to be undertaken in quiet rooms rather than in a sound treated room or booth.

The testing procedure has some differences from conventional manual audiometry. The subject is seated at the front of the touch screen, the headphones are placed over the ears, and a bone conductor is placed on the centre of the forehead. The subject is asked to follow the instructions on the screen; these are also presented through the headphones. At the start the volume is raised automatically if the subject cannot hear the instructions; this is determined by an initial question to the subject, asking them to respond if they heard the question.

The instructions explain the testing procedure, and inform the subject that the presentation of a stimulus will be followed by a question appearing on the screen, asking them to touch the large YES or NO button on the screen to indicate if they have or have not heard the stimulus. Using a forced-choice method of participant response is thereby different to conventional audiometry where a response is required only when the stimulus is heard. Although the AMTAS method is

similar to the modified Hughson-Westlake method, a more accurate description would be a bracketing procedure that converges on the 50% detection threshold in a single-interval, YES-NO sequence of stimuli with a 5-dB step size. In contrast to conventional manual testing, AMTAS is always masking during bone conduction testing. The bone conduction thresholds are calibrated for forehead bone conduction. AMTAS includes two methods to monitor subject reliability. On some occasions no stimulus is presented, but yet the subject is asked if sound was heard. A positive response to this generates a 'False Alarm' and the subject is warned via a message on the screen. Furthermore, after each threshold determination, a stimulus with intensity higher than threshold is presented. A 'NO' response is "Quality Check fail", which is recorded and reported at the end of the test.

Reporting of results includes a graphical plot of the audiogram, an assessment of quality (Good, Fair, Poor or Fail) based on a number of test variables including the number of false positives and false negatives, test-retest differences and air-bone gap, and also an automatic classification of the audiogram according to the type, degree and configuration of the loss (Margolis, Saly et al., 2007).

Subjects

Subjects for the study were recruited from the public and from patients at Lions Hearing Clinics, Subiaco and Nedlands, Western Australia. It was intended to have a range of hearing sensitivity amongst subjects, including those with hearing loss. Ethical approval for the study was received from The University of Western Australia Human Research Ethics Office. Participants provided informed consent before testing commenced.

Test Procedures

Manual audiometry was conducted by either of two experienced audiologists in a certified sound booth using either an Interacoustics AC 40 (Assens, Denmark) or GSI 61 (Eden Prairie, MN, USA) audiometer with TDH39 headphones, and the modified Hughson-Westlake threshold seeking method. Bone conduction audiometry was performed with mastoid placement of the bone oscillator. The AMTAS was administered by one of the authors in a quiet room following the protocols as the instrument would be normally used: seating the participants comfortably, and asking the participant to follow the instructions that are delivered through the headphones. The average ambient sound level in this room was measured for a 60 minute period of time, representative of the room's ambient noise levels during all testing, using a Type 2 Sound Level Meter (Lutron SL-4011, Lutron Electronic Enterprise, Taipei, Taiwan) and found to be 38.5 dBA. Prior to testing, otoscopy and tympanometry were performed to ensure no apparent middle ear pathology.

The order of testing was varied so that some of the participants (64%) were tested by manual audiometry first and then by AMTAS. Some subjects had been previously tested for hearing loss and were therefore familiar with the procedure whilst for others it was the first time their hearing was tested.

Both tests were performed within a short time period of each other, with a 10 to 20 minute break in between. A separate group of people was requested to undertake three consecutive AMTAS tests, with a 10 minute break between tests.

Analysis

The validation analysis assessed the air conduction thresholds for 250Hz, 500Hz, 1000Hz, 2000Hz, 4000Hz and 8000Hz, and the bone conduction thresholds for 500Hz, 1000Hz, 2000Hz and 4000Hz. Air and bone conduction thresholds recorded as reaching the output limits of either of the audiometers were not included in the analysis. 4000Hz bone conduction thresholds were corrected by -15dB to account for the perceived errors specifically (Margolis, Glasberg et al., 2010; Margolis & Moore, 2011).

Univariate analysis of variance was performed against sub-sets of thresholds, age, gender, and time to complete test. Mean and standard deviations were calculated for each frequency, as well as the mean differences (real and absolute) between the manual and AMTAS thresholds, and also for the test-retest tests with the AMTAS. To calculate the difference, the manual threshold was subtracted from the AMTAS threshold (Margolis et al., 2007). For the test-retest analysis, to conform to the method of analysis used by other studies, only data from two tests

were included in the analysis rather than all three (Frank & Ragland, 1987; Laukli & Fjermedal, 1990; Margolis, Glasberg et al., 2010; Swanepoel, Mngemane et al., 2010; Swanepoel & Biagio, 2011). The results from the second and third set of thresholds were used, thereby minimising the possible learning effect resulting from the first session of automated audiometry.

SPSSv19 and Microsoft Excel were used for the statistical analysis.

Results

Study 1: Test-retest reliability

Ten participants were tested three times by AMTAS automated audiometry; six males and four females, aged 54.7 ± 11.5 (range 30.1 to 68.6) years. There was no significant age difference between males and females.

Thresholds from both ears were included in the analysis, resulting in data from 20 ears for each test frequency, except for the air conduction thresholds at 8000Hz and the bone conduction thresholds at 4000Hz where the limit of testing was reached for one ear of one participant in each case. Except for this one instance, all participants had normal hearing in both ears (average of thresholds at 500, 1000, 2000 and 4000 Hz <20dB), and the reliability of the test was 'Good' in each case based on the AMTAS test reliability measure (Margolis, Saly et al., 2007).

The mean difference between the test-retest thresholds, a measure of the offset, was low (Table 1), ranging from -1.5dB at 250Hz to 1.8dB at 8000Hz for air conduction, and from -4.0dB at 250Hz to 2.3dB at 500Hz for bone conduction.

The standard deviations (SD) of the absolute differences were used to examine the distribution of differences as they were by previous studies (Margolis, Glasberg et al., 2010; Swanepoel & Biagio, 2011). The air conduction SD varied from 4.1 to 6.0dB, the highest being for 2000Hz (Table 1). For bone conduction SD ranged from 2.9 to 5.8dB, the highest being for 250Hz.

Study 2: Accuracy

Participants: 47 participants were tested by manual and AMTAS audiometry. Three participants were excluded from the primary analysis: one (aged 91 years) did not complete the automated test, and two tests were classified as having a 'Poor' reliability according to the AMTAS test reliability quality indicator. The tests of four of the participants were classified as 'Fair' and the remainder as 'Good' (Margolis, Saly et al., 2007).

Analysis was conducted on the remaining 44 participants, 23 males and 21 females. The mean age of the males was 61.7 years (SD 13.8, range 32.2 to 81.1 years), and the mean age of the females was 65.7 years (SD 17.1, range 21.4 to 88.4 years). There was no significant difference between the ages of males and females.

Half of all the ears had a mild hearing loss, 31% had no hearing loss, 18% moderate hearing loss, and one ear had a severe hearing loss (Table 2).

Both ears of all the subjects were included in the analysis, providing air conduction data from 88 ears, and bone conduction data for 74 ears. However, missing data resulted in a range of 72 to 87 data points per air conduction test frequency, and a range of 60 to 70 data points per bone conduction test frequency.

The mean difference between air conduction thresholds from the AMTAS and manual testing ranged from -3.3dB to 3.6dB (Table 3). Whilst these mean values indicate the overall offset between the two techniques, the true variability is masked because positive differences are counteracted by negative differences; using absolute variations is a better approach. The absolute mean difference ranged from 4.9 to 7.0dB, and the SDs from 5.9 to 6.9dB (Table 3). The percentage of AMTAS air conduction thresholds falling within 5, 10 and 15dB of manual thresholds are shown in Table 3 with 86.1% of thresholds within 10dB, with the best outcome for 500Hz and the worst for 8000Hz.

Univariate analysis of variance showed that there were no significant relationships between ear, frequency or severity of hearing loss and the absolute differences (including interactions between the independent variables) between manual and automated air conduction thresholds. For differences between manual and automated air conduction thresholds, there were significant relationships to ear ($F=4.351$, $p=0.038$, partial $ETA^2=0.009$), frequency ($F=3.93$, $p=0.002$, partial $ETA^2=0.040$) and

severity of hearing loss (F=3.26, p=0.021, partial ETA^2=0.020). However, although the observed power was 0.938 for frequency, in each case the effect size was very small, meaning that these variables have little influence on the difference between techniques. A post-hoc analysis of frequency versus the difference between manual and automated audiometer thresholds showed that the 250, 500 and 4000 Hz thresholds were most different between techniques.

A similar analysis of bone conduction data (Table 4) showed an overall offset of zero, mean absolute differences ranging from 7.2 to 11.9dB, and SD ranging from 6.2 to 8.8dB. 86.1% of all AMTAS bone conduction thresholds fell within 15dB of manual thresholds, with the worst outcome at 4000Hz.

Univariate analysis of variance was conducted on these data, showing a significant statistical relationship between the differences of manual and AMTAS bone conduction thresholds, and frequency (F=2.6; p=0.037) and severity of hearing loss (F=3.5; p=0.015). There was also a significant link between the absolute difference between manual and AMTAS bone conduction thresholds, and ear (F=3.9; p=0.048). However, in all of these cases the observed power was below 0.8 and the effect sizes were less than 4%.

The average test time for automated audiometry for the 44 participants was 16.1 minutes (SD: 3.9; range 9.9 to 30.9 minutes). The test times for the four participants graded as 'Fair' (19.9 minutes: SD: 4.96; range 15.91 to 27.15) appeared to be significantly longer (p=0.049 using univariate analysis of variance) than the time for the remaining 40 participants with a 'Good' reliability grading (15.82 minutes; SD: 3.71; range 11.9 to 30.91). However, the small sample size means the study has insufficient power to support this finding.

Discussion

Comparing two audiometric techniques can be viewed in a number of ways. In clinical practice a variation in threshold of less than 10dB is usually considered sub-clinical. Occupational regulators advise that a change of 10dB or more in the mid-frequencies of 2000, 3000 and 4000Hz can be classified as a Standard Threshold Shift if this change is confirmed by a retest (OSHA, 1983). A threshold shift of 15dB or more at any frequency from 500 to 6000Hz confirmed by a retest is considered as a "Significant Threshold Shift" in other jurisdictions (NIOSH, 1980). These values are derived from the fact that thresholds are determined to the nearest 5dB, and that calibration standards allow for a ±3dB variation (ANSI, 1996), besides variations which may be attributable to earphone placement and human factors (e.g. concentration, familiarity with the task). Acceptable variability in behavioural threshold audiometry may be defined by the typical variability in measures of test-retest or even inter-tester reliability (Margolis et al. 2010; Swanepoel, Mngemane et al. 2010). Whilst these measures of "normal variation" provide a useful reference, the exact way in which this variation or correspondence is expressed may vary. Bone conduction audiometry is inherently more variable than air conduction audiometry due to the complexity of the pathway conducting sound through the skull and ossicles to the cochlea, the effect of transducer placement and force, and the vibro-tactile sensations that may be present at lower test frequencies (Stuart et al., 1991; Smith-Olinde et al., 2006; Margolis, Glasberg et al., 2010; Swanepoel, Mngemane et al., 2010; Swanepoel & Biagio, 2011).

A simple method, that has often been used to compare the thresholds obtained by two techniques or in two sessions, is by calculating a correlation coefficient. This technique, however, is considered inappropriate to measure agreement between clinical measures (Bland & Altman, 1986), as it is a measure of relation rather than correspondence. Most published studies have used the standard deviation (SD) of the real difference or absolute difference of automated compared to manual thresholds to measure variability. Bland and Altman (1986) recommend using the SD of the absolute differences. They comment further that the repeatability coefficient should be used, in line with the British Standards Institution, which is defined as twice the standard deviation of differences between repeated measures. For the sake of comparison to other studies, one SD has been used in this paper as an indicator of variability.

Test-retest reliability

In this study the test-retest SDs for air and bone conduction thresholds for the AMTAS were all below 5dB, except at 1000Hz and 2000Hz air-conduction 250Hz bone conduction test frequencies. The SDs have been compared to those reported by a selection of other reports comparing two techniques or two instruments (Table 5). To determine if there was a significant difference in SD, an analysis of variance test was applied (http://statpages.org/anova1sm.html) utilising the values of SD and the sizes of groups (n).

These comparisons show that the AMTAS air conduction test-retest was not significantly different in most cases to that reported by other studies; AMTAS test-retest variation was significantly greater than some other studies at 2000Hz (Swanepoel, Mngemane et al., 2010) and at 8000Hz (Swanepoel, Mngemane et al., 2010), and when considering all frequencies (Swanepoel, Mngemane et al., 2010). Swanepoel et al. (2010) conducted two test-retest studies in a sound treated booth, with normal hearing participants, firstly with a KUDUwave audiometer in manual mode and then in automated mode. Furthermore, the KUDUwave audiometer utilises insert earphones covered by circumaural earcups, which provide a high level of passive attenuation of ambient sound (Clark & Roeser, 1988; Berger & Killion, 1989; Frank & Wright, 1990). Our study was conducted in a quiet room, with participants having a range of hearing losses that may have contributed to the slightly larger variation. However, of note is that our variation was not significantly greater than that reported by Margolis (2010) who also used the AMTAS but in a sound treated booth.

The test-retest variability for bone conduction was low, being over 5dB only at 250Hz. However, in general the variation is not significantly different to reports by others. Only at 250Hz is the variation significantly greater than that reported by Laukli (Laukli & Fjermedal, 1990). In two cases the variation is significantly less than that reported by Swanepoel (2011) for manual operation of the KUDUWave audiometer at 2000Hz and overall.

Accuracy

When comparing the AMTAS to manual thresholds the absolute threshold difference SD varied between 5.88 and 6.87dB for air conduction thresholds, and between 6.16 and 8.76dB for bone conduction. In most cases the variations are not significantly different to test-retest variations (Table 6); only the variation of the bone conduction thresholds at 2000Hz is significantly greater than the test-retest variation and overall SDs for both air and bone conduction. However, whilst the variations are greater than 5dB, they are all still within the 10dB variation that is normally considered acceptable (NIOSH, 1980; OSHA, 1983).

The overall difference between AMTAS and manual thresholds is almost zero. There is a tendency for the AMTAS to record a higher value for air and bone conduction thresholds in the lower frequencies, and lower values in the higher frequencies, but in all cases the mean difference is not greater than 4dB.

When comparing our results to those of others (Table 6) many of these variations are significantly greater ($p<0.01$) than those previously reported, especially by Swanepoel (Swanepoel, Mngemane et al., 2010; Swanepoel & Biagio, 2011) and Margolis (2010). As already noted, contributing to this greater variation may be that both these studies conducted all their assessments in a sound treated booth. Furthermore, Swanepoel & Biagio (2011) conducted their study with the KUDUwave in a manual mode, with insert earphones, and the majority of participants had normal hearing. Our study in contrast tested in two different ambient sound environments, with more than one audiologist conducting the manual audiometry, and the AMTAS audiometer tested in an environment that would better replicate how it may be used: in a quiet room with an ambient noise level less than 70dB. In addition the type of transducers used in Swanepoel & Biagio (2011) and Swanepoel, Mngemane et al. (2010) were the same (i.e. insert earphones) between test sessions. In the current study the AMTAS used circumaural earphones and manual audiometry was conducted with supra-aural earphones. Despite reference equivalent calibration levels between earphones aimed at equivalent threshold findings it may still introduce some level of variability due to placement of earphones on individual ears. Placement of

supra-aural earphones, unlike circumaural earphones used by the AMTAS, may more easily influence threshold results and may even result in ear canal collapse in some cases (Mahoney & Luxon, 1996).

Another variable we explored was the test quality as determined by the AMTAS ((Margolis, Saly et al., 2007). When those with a 'Fair' test quality were excluded from the analysis, reducing the number of participants by four to 43, there is little overall change (Table 7). There is, however, a lower overall variation resulting in it now no longer being significantly different from the test-retest variation.

Our study has shown that greater variation from manual audiometry, and automated audiometry with the KUDUwave (Swanepoel, Mngemane et al., 2010; Swanepoel & Biagio, 2011) or AMTAS in a sound treated environment (Margolis, Frisina et al., 2011; Margolis & Moore, 2011) can be expected when the AMTAS is used in an environment in which the ambient sound levels may not be controlled. Thresholds may be slightly higher in the low frequencies and lower in the high frequencies. Of note is that variation between manual and AMTAS thresholds is quite apparent at 4000Hz, the frequency at which the validity of air bone gaps has been questioned previously (Margolis, Glasberg et al., 2010; Margolis & Moore, 2011; Nondahl et al., 2012).

This study also provides some evidence that the accuracy of the audiograms determined objectively by the AMTAS to have a 'Fair' reliability rating may be lower than those judged to have a 'Good' reliability rating (Margolis et al. 2007). If adopted in clinical practice, audiograms produced by the AMTAS with a 'Fair' reliability rating should be scrutinised more closely than those with a 'Good' rating. Additional quality control features such as continuous monitoring of ambient noise levels during testing, as employed by the KUDUwave audiometer (Swanepoel, Mngemane et al., 2010; Swanepoel & Biagio, 2011) may also be useful when testing outside a sound treated booth.

Conclusion

Test-retest results show that the AMTAS automated audiometer used in a quiet room provides acceptable variation in air and bone conduction thresholds, and that for the most part the variations are not significantly different to that reported by others using different audiometers, both manual and automated, and using the AMTAS audiometer in a sound treated booth. The variations between AMTAS thresholds acquired in a quiet room and those determined manually in a standard audiological session are also within the test-retest variation of the AMTAS in most cases.

References

ANSI 1996. Specification for audiometers. (ANSI S3.6-1996). New York: American National Standards Institute.

Berger E.H. & Killion M.C. 1989. Comparison of the noise attenuation of three audiometric earphones, with additional data on masking near threshold. *J Acoust Soc Am*, 86, 1392-1403.

Bland J.M. & Altman D.G. 1986. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet*, 1, 307-310.

Burns W. & Hinchcliffe R. 1957. Comparison of the auditory threshold as measured by individual pure tone and Bekesy audiometry. *Journal of the Acoustical Society of America*, 29, 1274-1277.

Clark J.L. & Roeser R.J. 1988. Three studies comparing performance of the ER-3A tubephone with the TDH-50P earphone. *Ear Hear*, 9, 268-274.

Dobie R.A. 1983. Reliability and validity of industrial audiometry: implications for hearing conservation program design. *Laryngoscope*, 93, 906-927.

Frank T. & Ragland A.E. 1987. Repeatability of high-frequency bone conduction thresholds. *Ear Hear*, 8, 343-346.

Frank T. & Wright D.C. 1990. Attenuation provided by four different audiometric earphone systems. *Ear Hear*, 11, 70-78.

Goulios H. & Patuzzi R.B. 2008. Audiology education and practice from an international perspective. *Int J Audiol*, 47, 647-664.

Laukli E. & Fjermedal O. 1990. Reproducibility of hearing threshold measurements. Supplementary data on bone-conduction and speech audiometry. *Scand Audiol*, 19, 187-190.

Mahoney C.F. & Luxon L.M. 1996. Misdiagnosis of hearing loss due to ear canal collapse: a report of two cases. *J Laryngol Otol*, 110, 561-566.

Margolis R.H., Frisina R. & Walton J.P. 2011. AMTAS((R)): automated method for testing auditory sensitivity: II. air conduction audiograms in children and adults. *Int J Audiol*, 50, 434-439.

Margolis R.H., Glasberg B.R., Creeke S. & Moore B.C. 2010. AMTAS: automated method for testing auditory sensitivity: validation studies. *Int J Audiol*, 49, 185-194.

Margolis R.H. & Moore B.C. 2011. AMTAS((R)): automated method for testing auditory sensitivity: III.

sensorineural hearing loss and air-bone gaps. *Int J Audiol*, 50, 440-447.
Margolis R.H. & Morgan D.E. 2008. Automated pure-tone audiometry: an analysis of capacity, need, and benefit. *Am J Audiol*, 17, 109-113.
Margolis R.H., Saly G.L., Le C. & Laurence J. 2007. Qualind: A method for assessing the accuracy of automated tests. *J Am Acad Audiol*, 18, 78-89.
NIOSH 1980. Compendium of Materials for Noise Control. NTIS Stock No. PB 298307. Pub. No. 80-116. . Cincinnati, OH: National Institute for Occupational Safety and Health.
Nondahl D.M., Tweed T.S., Cruickshanks K.J., Wiley T.L. & Dalton D.S. 2012. Aging and the 4-kHz Air-Bone Gap. *J Speech Lang Hear Res*, 55, 1128-1134.
OSHA 1983. Occupational Noise Exposure: Hearing Conservation Amendment; Final Rule. Fed. Reg. 48: 9738 - 9785 (8 March 1983).
Smith-Olinde L., Nicholson N., Chivers C., Highley P. & Williams D.K. 2006. Test-retest reliability of in situ unaided thresholds in adults. *Am J Audiol*, 15, 75-80.
Stuart A., Stenstrom R., Tompkins C. & Vandenhoff S. 1991. Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults. *Audiology*, 30, 82-90.
Swanepoel D. & Biagio L. 2011. Validity of diagnostic computer-based air and forehead bone conduction audiometry. *J Occup Environ Hyg*, 8, 210-214.
Swanepoel D., Clark J.L., Koekemoer D., Hall J.W., 3rd, Krumm M., et al. 2010. Telehealth in audiology: the need and potential to reach underserved communities. *Int J Audiol*, 49, 195-202.
Swanepoel D., Mngemane S., Molemong S., Mkwanazi H. & Tutshini S. 2010. Hearing assessment-reliability, accuracy, and efficiency of automated audiometry. *Telemed J E Health*, 16, 557-563.
von Békésy G. 1947. A new audiometer. *Acta Oto-Laryngologica*, 35, 411-422.

Table 1: Mean differences and absolute mean differences and their standard deviations of air and bone conduction thresholds from AMTAS automated audiometry for each test frequency with 10 normal hearing participants. SD: standard deviation.

	Frequency (Hz)	250	500	1000	2000	4000	8000	All
Air	Mean difference	-1.5	1.3	0.3	0.8	0.3	1.8	0.5
	SD of differences	5.4	6.3	7.9	7.1	5.7	12.2	7.7
	Mean absolute difference	3.5	4.8	5.3	3.8	3.8	4.7	4.3
	SD of absolute differences	4.3	4.1	5.7	6.0	4.3	4.9	4.9
	n	20	20	20	20	20	19	119
Bone	Mean difference	-4.0	2.3	1.0	1.3	-1.3		-0.2
	SD of differences	7.0	6.0	5.5	4.3	5.7		6.1
	Mean absolute difference	5.5	4.8	3.5	3.3	3.4		4.1
	SD of absolute differences	5.8	4.1	4.3	2.9	4.7		4.5
	n	20	20	20	20	19		99

Table 2: Number (and percentage) of ears in the study by severity of hearing loss (HL) as determined by Four Frequency Average (4FA) - mean of thresholds at 500, 1000, 2000 and 4000Hz. No HL: 4FA<20dB; mild HL: >=20db 4FA <40dB; moderate HL: >=40dB 4FA <60dB; severe HL: 4FA >=60dB.

	Normal hearing	Mild HL	Moderate HL	Severe HL
Right ears	15 (33.3)	21 (46.7)	8 (17.8)	1 (2.2)
Left ears	12 (27.9)	23 (53.5)	8 (18.6)	0 (0)
All ears	27 (30.7)	44 (50.0)	16 (18.2)	1 (1.1)

Table 3: A description of air conduction differences (mean, standard deviation, range of correspondence) between manual and AMTAS thresholds by degree of hearing loss. Manual thresholds were subtracted from AMTAS thresholds.

Type of HL			Frequency (Hz)						
			250	500	1000	2000	4000	8000	All
All ears	n		88	88	87	87	87	72	509
	Difference	Mean	-3.0	0.7	-2.8	1.9	1.1	3.6	0.1
		SD	9.2	7.7	7.7	8.7	9.4	8.6	8.8
	Absolute difference	Mean	6.8	4.9	5.6	5.6	7.0	6.8	6.1
		SD	6.8	5.9	6.1	6.9	6.3	6.4	6.4
	within +/- 5dB (%)		64.4	77.0	81.6	81.6	69.0	68.1	73.8
	within +/- 10dB (%)		83.9	93.1	88.5	88.5	83.9	77.8	86.1
	within +/- 15dB (%)		96.6	96.6	94.3	92.0	92.0	95.8	94.5
No HL	n		27	26	27	27	27	24	158
	Difference	Mean	-2.4	3.3	0.7	4.1	1.8	4.0	1.9
		SD	7.4	8.2	7.4	5.9	7.6	7.5	7.6
	Absolute difference	Mean	6.1	6.0	4.4	5.2	5.9	5.6	5.5
		SD	4.7	6.5	5.9	4.9	5.0	6.3	5.5
Mild HL	n		44	44	16	44	44	39	259
	Difference	Mean	-4.2	-0.1	-3.6	1.0	0.5	3.2	-0.6
		SD	9.0	7.9	7.3	10.1	9.9	8.9	9.2
	Absolute difference	Mean	6.7	4.9	5.5	5.8	7.3	7.3	6.2
		SD	7.2	6.1	6.1	8.3	6.7	5.9	6.8
Moderate HL	n		15	15	16	16	16	9	87
	Difference	Mean	-1.3	-1.3	-6.3	0.6	2.0	4.4	-0.7
		SD	12.2	5.3	7.4	8.3	10.1	11.0	9.6
	Absolute difference	Mean	8.0	3.8	7.5	6.3	8.0	7.8	6.8
		SD	9.0	3.9	6.1	5.3	7.5	8.7	6.8

Table 4: A description of bone conduction differences (mean, standard deviation, range of correspondence) between manual and AMTAS thresholds by degree of hearing loss. Manual thresholds were subtracted from AMTAS thresholds.

Type of HL			Frequency (Hz)					
			250	500	1000	2000	4000	All
All ears	n		36	66	70	60	63	295
	Difference	Mean	3.3	-3.0	-0.7	-1.8	3.9	0.0
		SD	14.3	9.3	10.1	12.0	12.3	11.7
	Absolute difference	Mean	11.9	7.2	8.0	8.5	9.4	8.7
		SD	8.4	6.5	6.2	8.6	8.8	7.7
	within +/- 5dB (%)		33.3	60.6	57.1	55.0	55.6	54.2
	within +/- 10dB (%)		61.1	78.8	82.9	83.3	71.4	76.9
	within +/- 15dB (%)		75.0	90.9	88.6	88.3	82.5	86.1
No HL	n		9	22	22	21	23	97
	Difference	Mean	3.90	-1.10	3.20	0.80	7.60	3.50
		SD	13.60	8.90	8.20	11.90	12.80	11.10
	Absolute difference	Mean	11.70	6.60	6.80	6.70	11.10	8.20
		SD	7.10	5.90	5.50	10.50	9.80	8.20
Mild HL	n		21	34	35	32	34	156
	Difference	Mean	1.70	-4.00	-2.10	-4.40	2.80	-1.40
		SD	14.10	9.00	11.30	11.40	12.20	11.70
	Absolute difference	Mean	11.20	7.20	9.30	9.70	9.30	9.20
		SD	8.40	6.70	6.70	7.30	8.30	7.40
Moderate HL	n		6	9	12	7	6	40
	Difference	Mean	8.30	-3.90	-3.30	-7.10	-4.20	-2.50
		SD	17.50	11.90	8.30	9.50	5.80	11.40
	Absolute difference	Mean	15.00	9.40	6.70	8.60	4.20	8.50
		SD	11.00	7.70	5.80	8.00	6.80	7.90

Table 5: Test-retest variability of AMTAS automated audiometry, and comparison to other studies. ** this study's assessment of AMTAS test-retest variability is significantly greater (p<0.01) than that in reference study; * this study's assessment of AMTAS test-retest variability is significantly less (p<0.01) than that in reference study.

	Study	Instrumentation	n (ears)	Frequency						All frequencies	
				250	500	1000	2000	4000	8000	n (ears)	SD
Air conduction	Swanepoel et al. (2010) Table 1	KuduWave manual	60	3.70	3.70	3.40	3.60**	3.50	4.30	420	3.90*
	Swanepoel et al. (2010) Table 1	KuduWave automated	60	3.50	3.60	3.20	4.10	3.00	3.20**	420	3.80**
	Margolis et al. (2010) Table 2	GSI-61	36	3.50	2.80	4.00	3.10	4.70	4.50	216	3.80
	This study	AMTAS	20	4.32	4.13	5.73	6.04	4.25	4.85	119	4.88
Bone conduction	Laukli and Fjermedal (1990)	Madsen OB822	26	3.20**	3.70	3.30	4.20	4.80			
	Swanepoel et al. (2011) Table II	KuduWave manual	20	5.20	5.40	7.50	6.00*	6.10		100	6.40*
	Swanepoel et al. (2011) Table II	GSI-61	20	5.50	4.40	7.40	3.60	4.70		100	5.30
	Margolis et al. (2010) Table 2	GSI-61	36	6.80	4.90	4.90	4.10	6.40		180	5.40
	This study	AMTAS	20	5.83	4.13	4.32	2.94	4.73		99	4.48

Table 6: Accuracy of AMTAS automated audiometry, and comparison to other studies. **: this study's assessment of AMTAS variation (SD) is significantly greater (p<0.01) than that in reference study.

	Study	Instrumentation	n (ears)	Frequency						All Frequencies	
				250	500	1000	2000	4000	8000	n (ears)	SD
Air conduction	Swanepoel et al. (2011) Table II	KuduWave (manual mode) and GSI audiometer	60	3.30**	4.50	3.60**	3.00**	3.40**	4.40**	420	3.80**
	Margolis et al. (2010) Table 3	AMTAS and GSI-61	60	4.00**	3.20**	4.40**	4.40**	3.20**	5.80	360	4.20**
	This study	AMTAS test-retest	20	4.32	4.13	5.73	6.04	4.25	4.85	119	4.88**
	This study	AMTAS and manual	87	6.10	5.88	6.05	6.87	6.30	6.41	511	6.41
Bone conduction	Frank and Ragland (1987)	Practronic bone conductor; 5x in 10 days	30			7.30		8.80	7.30		
	Swanepoel et al. (2011) Table II	KuduWave (manual mode) and GSI audiometer	60	5.60**	4.90	5.30	3.50**	4.60**		300	4.90**
	Margolis et al. (2010) Table 4	AMTAS and GSI-61	60		6.50	5.80	5.10**	6.20**		300	5.90**
	This study	AMTAS test-retest	20	5.83	4.13	4.32	2.94**	4.73		99	4.48**
	This study	AMTAS and manual	70	8.39	6.51	6.16	8.60	8.76		295	7.73

Table 7: Accuracy of AMTAS automated audiometry, and comparison to other studies as per Table 6, but excluding four participants with a 'Fair' test quality. ** this study's assessment of AMTAS variation (SD) is significantly greater (p<0.01) than that in the reference study; * this study's assessment of AMTAS variation is significantly less (p<0.01) than that in reference study.

	Study	Instrumentation	n (ears)	Frequency						All Frequencies	
				250	500	1000	2000	4000	8000	n (ears)	SD
Air conduction	Swanepoel et al. (2011) Table II	KuduWave (manual mode) and GSI audiometer	60	3.30**	4.50	3.60	3.00**	3.40**	4.40**	420	3.80**
	Margolis et al. (2010) Table 3	AMTAS and GSI-61	60	4.00**	3.20**	4.40	4.40**	3.20	5.80**	360	4.20**
	This study	AMTAS test-retest	20	4.32	4.13	5.73	6.04	4.25	4.85	119	4.88
	This study	AMTAS and manual - GOOD and FAIR	87	6.10	5.88	6.05*	6.87	6.30	6.41	511	6.41*
	This study	AMTAS and manual - GOOD only	78	4.87	5.16	4.46	6.42	5.68	6.14	457	5.68