AMTAS® - Automated Method for Testing Auditory Sensitivity Development and Validation

Robert H. Margolis, Ph.D.
President, Audiology Incorporated

George L. Saly, Sc.D.
Software Engineer

LINKED TABLE OF CONTENTS

Version020108

INTRODUCTION

AMTAS® is an automated method for obtaining a diagnostic pure-tone audiogram. It is designed for routine clinical testing in any environment in which pure-tone audiometry is performed, including Audiology clinics, otology offices, hearing aid offices, primary care offices, schools, and industrial settings. Air-conduction and bone-conduction thresholds are acquired at standard audiometric frequencies. Masking noise is always presented to the non-test ear. An air-conduction-only audiogram can be obtained if desired. A child version (**KIDTAS™)** is designed for children who can be tested by manual behavioral audiometry (approx. 5-12 years of age). This report summarizes the rationale, development, design, and validation of AMTAS.

Back to Table of Contents

RATIONALE

The primary motivating consideration for the development of AMTAS was that the use of automated technology for routine pure-tone audiometry would benefit the profession of Audiology and the patients it serves by increasing efficiency and accuracy while decreasing costs. As Audiology transitions to a doctoral profession it is becoming increasingly inefficient to use highly trained, doctoral personnel to perform routine testing on patients who are capable of listening to and following instructions and responding to an automated protocol. This allows the reallocation of professional time to activities that require doctoral skill. If a significant proportion of patients can be tested with an automated method, access to Audiology will be enhanced. In addition, modern technology provides that capability to improve our test methods by incorporating quality assessment into the procedure and providing standardization of the procedure. Standardization would ensure that a hearing test performed anywhere would be the same eliminating repetition of tests, and for the first time providing for consistency that is not possible when tests are performed by individuals with a wide range of training and skill levels.

Back to Table of Contents

DEVELOPMENT

AMTAS was developed through a partnership between Audiology Incorporated (www.audiologyincorporated.com) and three research institutions, the University of Minnesota, the University of Utah, and the James H. Quillen Veterans Administration Medical Center. The development was funded by the National Institutes of Health Small Business Technology Transfer (STTR) Program. STTR grants were awarded in 2001 (Phase I) and 2002 (Phase II). The development work was performed by Audiology Incorporated; evaluation and testing by the three research institutions. In 2002, U.S. patent No. 6,496,585 was awarded for AMTAS. Patents are pending for the quality assessment method and the audiogram classification system that are incorporated into AMTAS.

AMTAS was designed to be administered by any clinical audiometer that can be controlled by computer. This approach was adopted in order to take advantage of the excellent engineering features of modern audiometers and to provide the capabilities of automated and manual audiometry with the same equipment. By using existing audiometers, the development of hardware platforms was avoided and the development work was primarily related to software. To date, the system has been implemented on four commercial audiometers (Madsen Conera, Madsen Aurical, Madsen Itera, Grason Stadler GSI-61, and Otovation Amplitude A1 (air-conduction only).

Table 1. Collaborators

Audiology Incorporated
Robert H. Margolis, President
George Saly, Software Engineer
University of Minnesota
Robert H. Margolis, Ph.D., Principal Investigator
Allison Kohtz, M.A., Research Audiologist
Chap Le, Ph.D., Biostatistician
University of Utah
Lisa L. Hunter, Ph.D., Principal Investigator
M. Victor Barrett, Ph.D., Research Audiologist
James H. Quillen Veteran Hospital, Mountain Home, Tennessee
Richard H. Wilson, Ph.D., Principal Investigator
Deborah Weakley, M.A., Research Audiologist
Sherri Smith, Ph.D., Research Audiologist

The development and evaluation of AMTAS could not be possible without an extraordinarily competent, dedicated, and skilled group of collaborators. The collaboration team is shown in Table 1.

The development process consisted of developing the phase I software system, performing a feasibility study at the University of Minnesota, refinement of the software system, and performing a multicenter trial at the three research institutions.

Back to Table of Contents

DESCRIPTION

1. Psychophysical Method

AMTAS was designed to provide the same test results that would be obtained by an expert audiologist with a psychophysical procedure that has advantages over the commonly used clinical ("Hughson-Westlake") method. The psychophysical procedure is a single-interval, Yes-No, forced choice procedure with feedback. This means that there is a defined period in which the stimulus may occur (the observation) interval, after which the patient votes Yes or No. A predetermined proportion of trials are "catch trials", observation intervals in which no stimulus is presented. A Yes response to a catch trial is a "false alarm". When a false alarm occurs, the patient is informed and instructed to be sure to respond Yes only when there is a tone (feedback).

Back to Table of Contents

2. Masking

In routine clinical testing it is common to begin the test without masking, discover that masking is needed, and retest with masking. AMTAS avoids this inefficiency by always masking. A proprietary method is used to select the masker level based on the information that is available at the time. When the test is completed, an analysis is performed of masker levels and thresholds of both ears to determine if overmasking or undermasking may have occurred. When this occurs, the thresholds are identified with "masker alerts". An analysis of our clinical

trial data indicated that there were almost no cases where masker alerts occurred where a different masker level would have provided a more accurate result. In other words, almost all the masker alerts were “masking dilemmas” where there was no masker level that was audible in the non-test ear that was not also audible in the test ear. Nevertheless, when masker alerts occur, those thresholds should be retested manually.

Back to Table of Contents

3. Earphones

The most commonly used earphones for pure-tone audiometry are the supra-aural earphones manufactured by Telephonics. The cushion used to couple the earphone to the head causes some problems that complicated the test. These include a) a large low frequency occlusion effect that requires that the earphone be placed off the ear during bone conduction testing, b) poor ambient noise exclusion; c) ear canal collapse, and d) discomfort. A separate project for the development of an improved earphone is under way. In the mean time, the Sennheiser HDA200 earphones that are commonly used for extended high-frequency testing can be used for the conventional frequencies. That earphone provides a reasonable solution to the problems mentioned, except that they produce a significant occlusion effect at 250 Hz.

Back to Table of Contents

4. Bone Conduction

Although bone conduction testing is usually performed with the bone vibrator on the mastoid, there are many advantages to placing the vibrator on the forehead. These include a) eliminating the need to move the transducer during the test, b) lower intersubject variability, c) more stable placement, and d) less influence by middle ear conditions. Forehead testing requires more output from the audiometer which has been a limitation but modern instruments are capable of providing adequate stimulation levels for forehead testing. The elimination of the need to move the transducer to test the other ear is an efficiency that makes it possible to obtain a complete air- and bone-conduction audiogram with masking without moving the transducers.

Back to Table of Contents

5. Qualind™ - Quality Assessment Method (patent pending).

The quality assessment method is a critical feature of AMTAS. When expert audiologists observe behaviors that indicate that the results may be inaccurate, they use a variety of techniques to insure test accuracy such as reinstructing the patient, modification of the procedure, and retesting. But these quality control methods are not taught, documented, or formally incorporated into the test. Their use is highly dependent on the skill level of the tester. Many of the behaviors that are used by audiologists for quality assessment can be tracked and measured by computer and can be formally incorporated into a validated prediction of test accuracy. AMTAS incorporates eight quality indicators into the procedure and uses them to predict test accuracy.

Back to Table of Contents

6. AMCLASS™ - Audiogram Classification System (patent pending)

When an audiogram has been completed, the air- and bone-conduction thresholds are analyzed and the audiogram is given three descriptive categories based on configuration, severity, and site of lesion. In addition, the audiogram is analyzed for interaural asymmetry. Despite the fact that they are used widely for the clinical description of test results, there are no standard definitions for these categories. A validation study was conducted to derive valid category definitions. The classification system is shown in Figure 1.

Figure 1. AMCLASS

AUDIOLOGY INCORPORATED

AMCLASS™ - AUDIOGRAM CLASSIFICATION SYSTEM

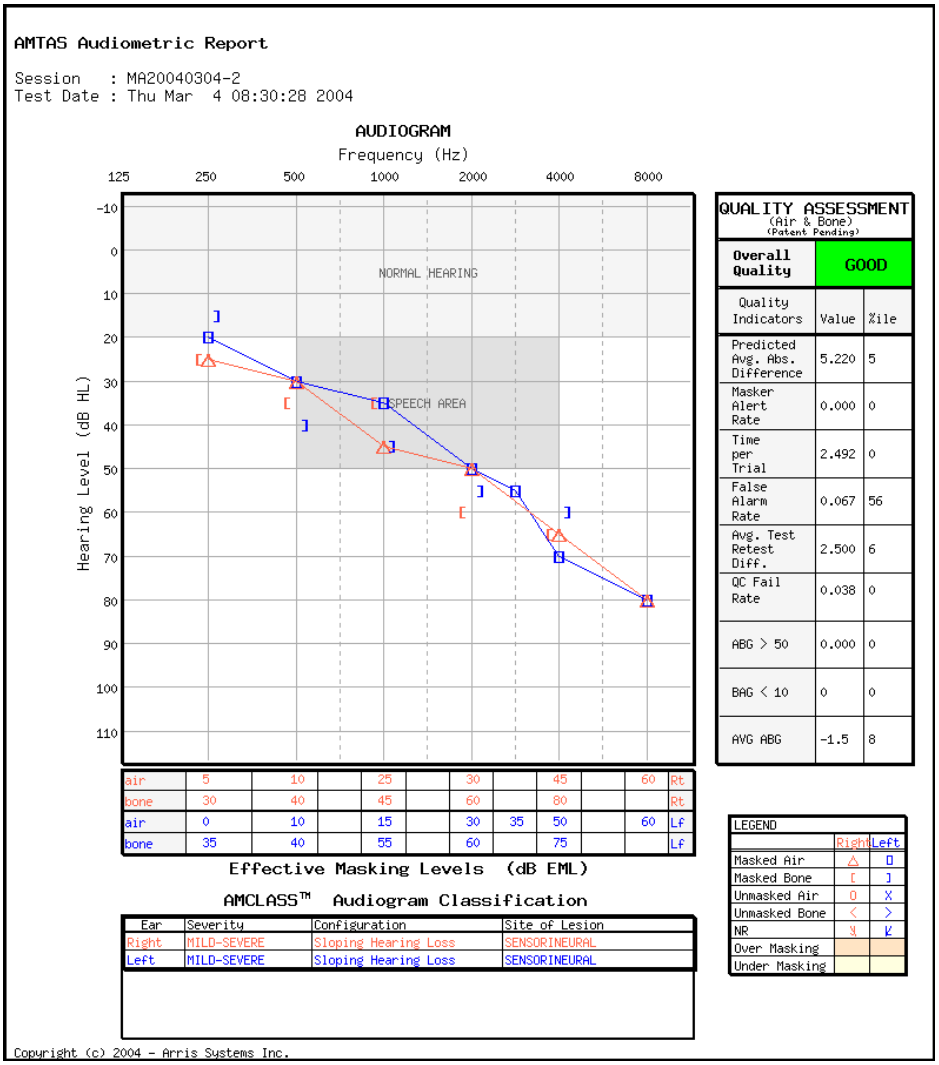
Configuration	Severity	Site of Lesion	Symmetry
Normal Hearing		Conductive	Symmetrical Hearing Loss
Flat Hearing Loss	Mild Moderate Severe Profound	Sensorineural Mixed Sensorineural or Mixed	Asymmetrical Hearing Loss
Sloping Hearing Loss	Normal-Mild Normal-Moderate Normal-Severe Mild-Moderate Mild-Severe Moderate-Severe Severe-Profound		
Rising Hearing Loss	Mild-Normal Moderate-Normal Moderate-Mild Severe-Normal Severe-Mild Severe-Moderate		
Trough-shaped Hearing Loss	Mild Moderate Severe		
Peaked Hearing Loss	Mild Moderate Severe		
Other	Mild Moderate Severe		

Back to Table of Contents

7. Report

The AMTAS report presents the audiogram, quality indicators, overall quality category, masker alerts, and AMCLASS category. An example is shown in Figure 2.

Figure 2. AMTAS Report



Note that all the symbols on the audiogram are masked symbols because masking is always presented to the non-test ear. The audiogram also indicates the “Normal Hearing” area and the “Speech Area”. We have found these to be very useful in explaining to patients how their hearing relates to the normal-hearing population and to hearing for speech. The speech area is intended to capture the range of “soft speech” as well as “normal speech”.

The Quality Assessment box presents the overall quality category (“Good”), the Predicted Average Absolute Difference, the value for each quality indicator, and a percentile for each. The predicted average absolute difference is the predicted average difference between AMTAS thresholds and thresholds that would be obtained by an expert audiologist. In this case the predicted average was 5.22 dB with an associated percentile of 5. This indicates

that 95% of cases are expected to have accuracies that are poorer (greater) than 5.22 dB. It is useful to scan the percentiles for values that may suggest inaccuracy. Any value greater than 90 would be suspect, even if the overall quality is “Good”.

Below the audiogram is a box that shows masker levels for each threshold. When masker alerts occur, the cell is shaded to indicate possible undermasking or possible overmasking as indicated in the legend.

Below the masker levels is the AMCLASS box which provides the configuration, severity, and site of lesion categories for each ear. If there is interaural asymmetry it is noted just below the AMCLASS categories.

Back to Table of Contents

VALIDATION STUDIES

1. Pilot Studies

The STTR Phase I project was comprised of pilot studies conducted on ten subjects with normal hearing, ten subjects with simulated hearing loss, and six subjects with sensorineural hearing loss. Each subject was tested by AMTAS and manual testing using TDH-49 earphones and mastoid bone conduction for both methods.

Figure 3 shows average air conduction and bone conduction thresholds for twenty ears of ten subjects for manual pure tone audiometry and AMTAS, employing three threshold rules, identified as AMTAS0, AMTAS-5, and AMTAS+5. The top two panels show the results on audiogram coordinates. The bottom panels show the results on an expanded scale to better illustrate the differences.

The three methods provide essentially equivalent average thresholds. The differences between the manual method and each AMTAS estimate were less than 5 dB with AMTAS producing consistently lower thresholds than the manual method. This is consistent with other studies that show lower thresholds for adaptive methods compared to the clinical method (Marshall and Jesteadt, 1986; Marshall et al., 1996). It is likely that the subject's knowledge of the observation interval provided by AMTAS provides an advantage compared to the indefinite observation interval in the clinical method. It is well known in the psychoacoustics literature that the more information about signal characteristics that is provided to the subject, the better the subject's performance on a sensitivity task (Green and Swets, 1974, pp. 265-271).

Of the three AMTAS threshold rules, AMTAS+5 provided the best agreement with manual audiometry, an average threshold difference of 1.5 dB for air conduction and 2.5 dB for bone conduction.

The largest mean difference between methods occurs at 8 kHz where the difference between manual and AMTAS+5 was 4.5 dB. This difference is mostly attributable to one subject who showed a 35 dB difference for the two methods. That subject showed the largest differences at all frequencies and probably wasn't performing the AMTAS task correctly, as suggested by a number of quality indicators, which are discussed below.

Figure 4 shows the distribution of air conduction threshold differences for manual audiometry and AMTAS+5. The majority of threshold differences (83%) are 5 dB or less. The proportion of threshold differences exceeded 10 dB was 10% and one third of those occurred for one subject, N10, who is discussed below. That subject had four occurrences of a >10 dB threshold difference, compared to an average of 1.2. Without subject N10, the proportion of differences exceeding 10 dB was 7%.

Figure 3. Average Air and Bone Conduction Thresholds for 20 ears of 10 normal-hearing subjects.

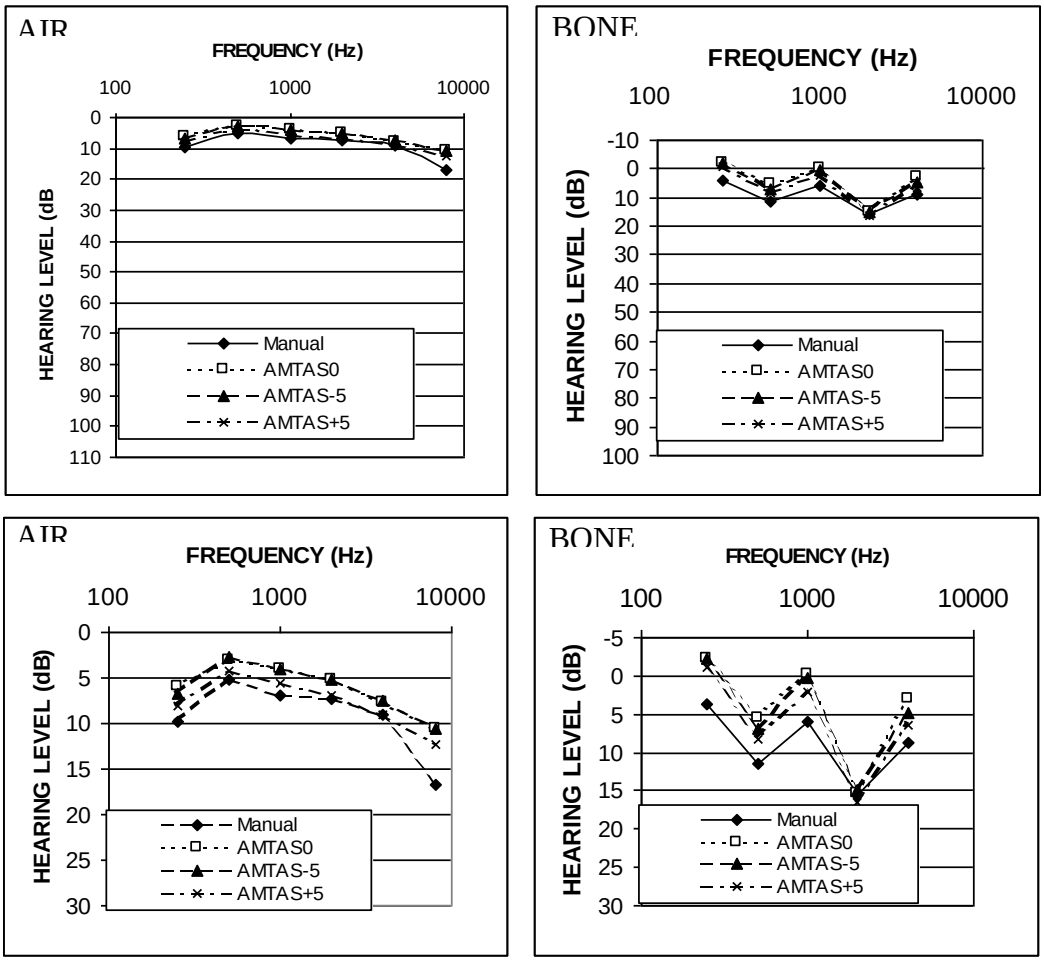


Figure 5 shows the distribution of bone conduction threshold differences for manual audiometry and AMTAS+5. The proportion of threshold differences that are 5 dB or less is 63%, lower than the corresponding proportion for air conduction, due to a larger number of 10

dB differences. The proportion of threshold differences exceeding 10 dB was 10%, identical to the corresponding proportion for air conduction. The differences in the distribution of differences for air conduction and bone conduction probably results from the greater variability of bone conduction threshold testing that is largely due to variability in coupling of the bone vibrator to the skull and transmission of vibratory stimuli to the inner ear.

Figure 4. Distribution of Air Conduction Threshold Differences for Normal Subjects

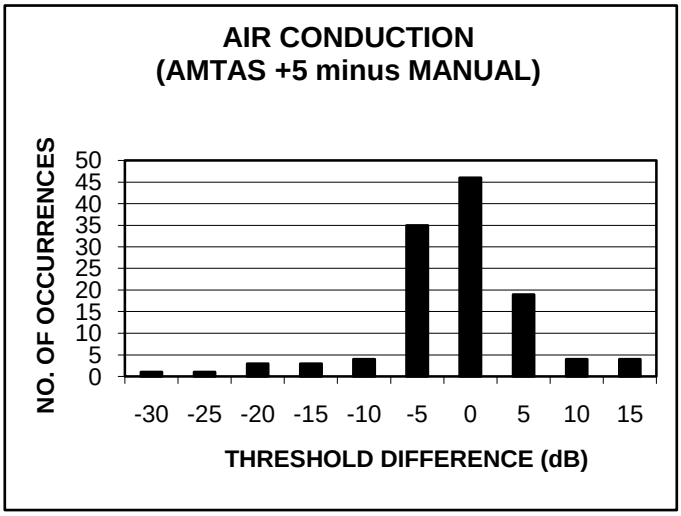
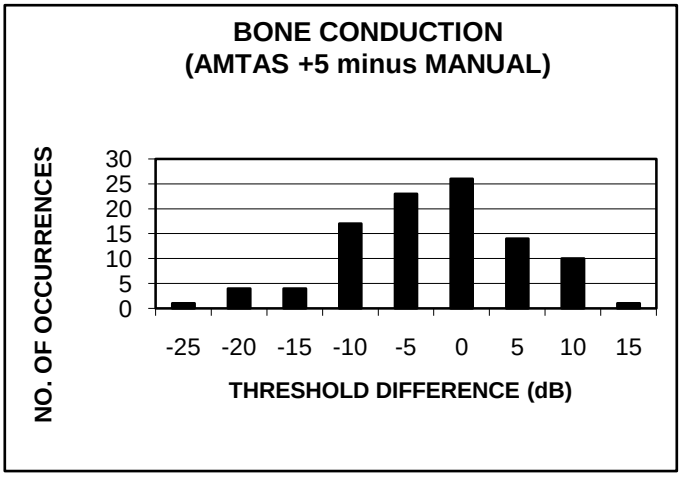


Figure 5. Distribution of Bone Conduction Threshold Differences for Normal Subjects



Ten normal-hearing subjects were tested with simulated unilateral hearing loss produced by placing an earplug in the ear canal. One subject was discarded because the

earplug became dislodged between tests. Two subjects were discarded who reported they were responding to the masking noise rather than the signal. Average thresholds for the plugged and unplugged ears are shown in Figure 6. Differences between mean manual and AMTAS thresholds are comparable to those for normal subjects with AMTAS producing slightly lower thresholds. The distribution of threshold differences for the two methods is shown in Figure 7. 71% of the differences were between –5 and 5 dB; 10% exceeded ±10 dB. The proportion of "serious errors" (10%) is identical to the results for normal subjects.

Average air and bone conduction audiograms for twelve ears of six hearing-impaired subjects are shown in Figure 8. The apparent air-bone gap in the averaged audiograms (difference between air conduction thresholds in left panel and bone conduction thresholds in right panel) occurred because the subject with the greatest hearing loss had bone conduction thresholds that were beyond the limits of the audiometer and are not averaged. Similar to the other subject groups, AMTAS tended to produce slightly lower thresholds with AMTAS+5 more closely approximating manual audiometry than AMTAS0. The averaged audiograms show very good agreement between the manual and automated methods.

Figure 6. Average thresholds for subjects with simulated unilateral hearing loss.

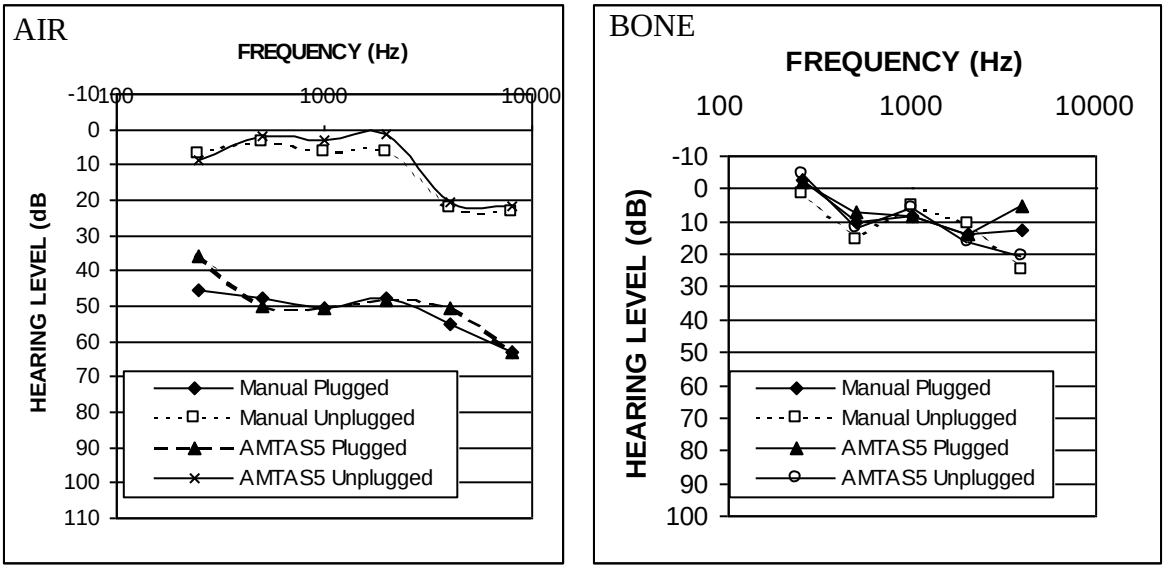


Figure 7. Distribution of Air Conduction Threshold Differences for Subjects with Simulated Unilateral Hearing Loss

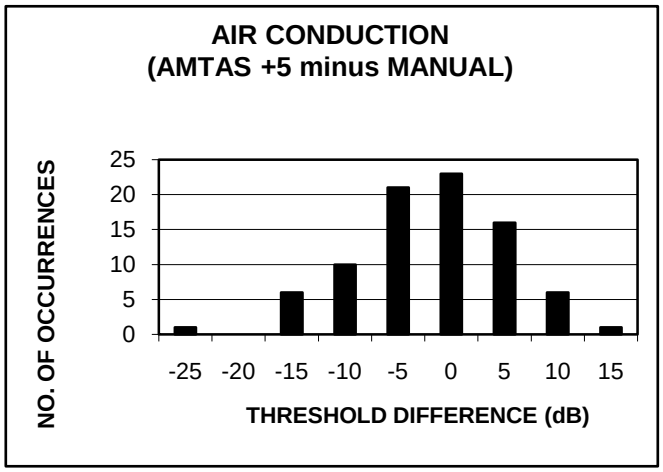


Figure 9 shows the distribution of threshold difference between manual and AMTAS+5 air conduction thresholds. 83% of the differences are $\leq$ 5 dB and 10% are > 10 dB. The proportion of differences that exceed 10 dB in the three subject groups is remarkably constant at 10%. A major emphasis of Phase II development will be to reduce this number.

Figure 8. Average thresholds for subjects with hearing loss

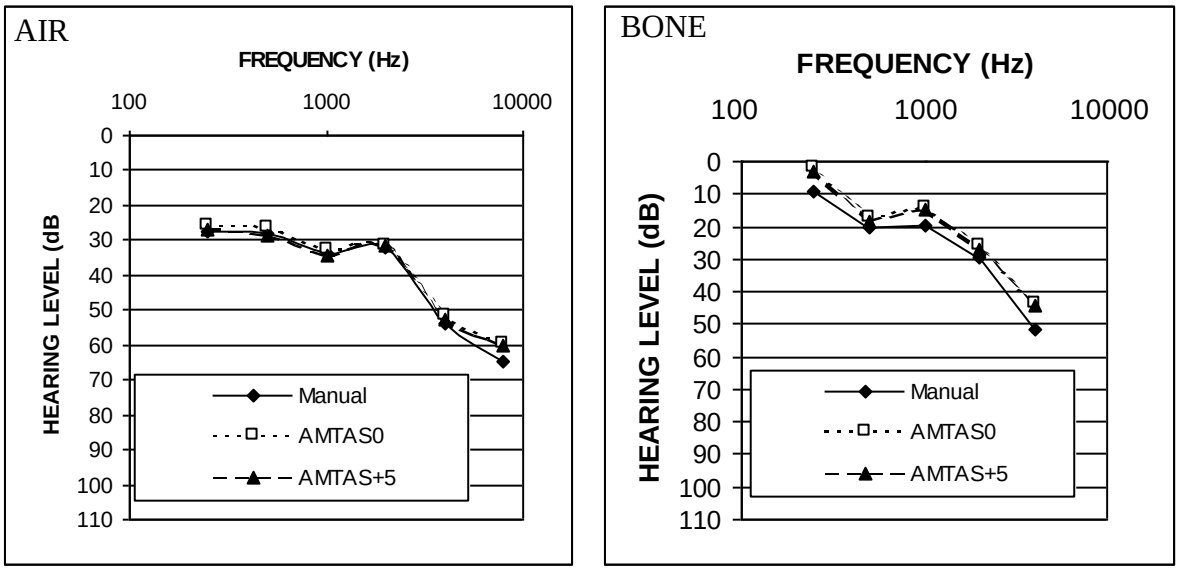
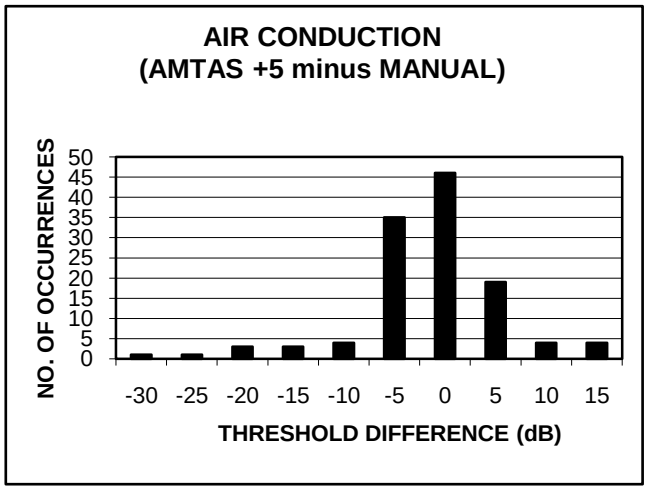


Figure 9. Distribution of Air Conduction Threshold Differences for Subjects with Hearing Loss



Back to Table of Contents

2. Comparison of AMTAS and Manual Thresholds

The Phase II project was a multicenter trial conducted at the three research institutions. An expert audiologist tested patients with AMTAS and manual testing at each site. AMTAS testing was performed with a prototype headset that was designed to minimize the occlusion effect, permitting complete air- and bone-conduction testing without repositioning transducers. Bone conduction was tested with the vibrator placed on the forehead secured by an elastic band. Manual testing was performed with TDH-49 earphones. The bone vibrator was placed on each mastoid and contralateral masking was used when deemed necessary by the audiologist. Each audiologist was validated against a fourth audiologist to insure high inter-tester agreement. Inter-tester correlation coefficients were 0.95, 0.97 and 0.97 for the three test sites.

One measure of accuracy that we employed for comparison of AMTAS and manual thresholds is the average absolute difference, the average of the absolute values of the differences between AMTAS and manual thresholds at each frequency. If the average absolute differences between AMTAS and manual testing are the same as the differences for two audiologists, we can claim that AMTAS is as accurate as an expert audiologist. Our audiologist validation study indicated that the average absolute difference between manual

audiograms obtained by two expert audiologists was 4.62 dB (s.d. = 4.13 dB). These values were used to construct the “Good”, “Fair”, and “Poor” agreement regions as described in the next section. “Good” represents and agreement within one standard deviation of the average agreement between two expert audiologists; “fair” is the region between one and two standard deviations, and “poor” is agreement that is beyond two standard deviations.

The number of cases in each category from the Quality Indicators study described below is shown in Table 2. The fourth column shows the percentage of cases in each group that are expected if the distribution of average absolute differences between AMTAS and manual testing is a normally distributed variable. The actual percentages are similar to those expected from the normal distribution assumption. Ten percent of cases were in the “Poor” category. These are cases that would probably require retesting with manual audiometry.

Table 2. Number of Case in Each Quality Category

Category	No. of Occurrences	%	% Normal
Good	83	69	68
Fair	25	21	26
Poor	12	10	5

Back to Table of Contents

3. Qualind™

Qualind is a method for predicting the accuracy of any test that meets certain requirements. The steps for the development of a prediction of test accuracy are the following.

1. Identify a set measurable behaviors and characteristics that may be related to test accuracy. For example, the test-retest difference at 1000 Hz is commonly used as a quality indicator in manual testing.

2. Identify an independent measure of the dimension being testing. For automated audiometry, the independent measure is the audiogram obtained by an expert audiologist.
3. Test a group of patients with both tests (automated and manual audiometry) that meet the expected requirements for the test. For automated audiometry, this would be adults and children (5-12 years old) who have no difficulty following verbal instructions.
4. Calculate a measure of agreement between the two tests. We used the average absolute difference, that is, the average differences between thresholds obtained with both methods.
5. Derive an equation that predicts the average absolute difference from the quality indicators in step 1. This can be done with a multiple regression technique. The strength of the regression (like a correlation coefficient) indicates how accurately accuracy can be predicted.
6. The predicted accuracy can be converted to categorical scale consisting of descriptive terms like “good”, “fair”, and “poor”.

The multiple regression also allows us to select quality indicators that are predictive of accuracy and reject those that are not predictive. The quality indicators that were found to be predictive and are used by AMTAS are shown in Table 3.

Table 3. AMTAS Quality Indicators

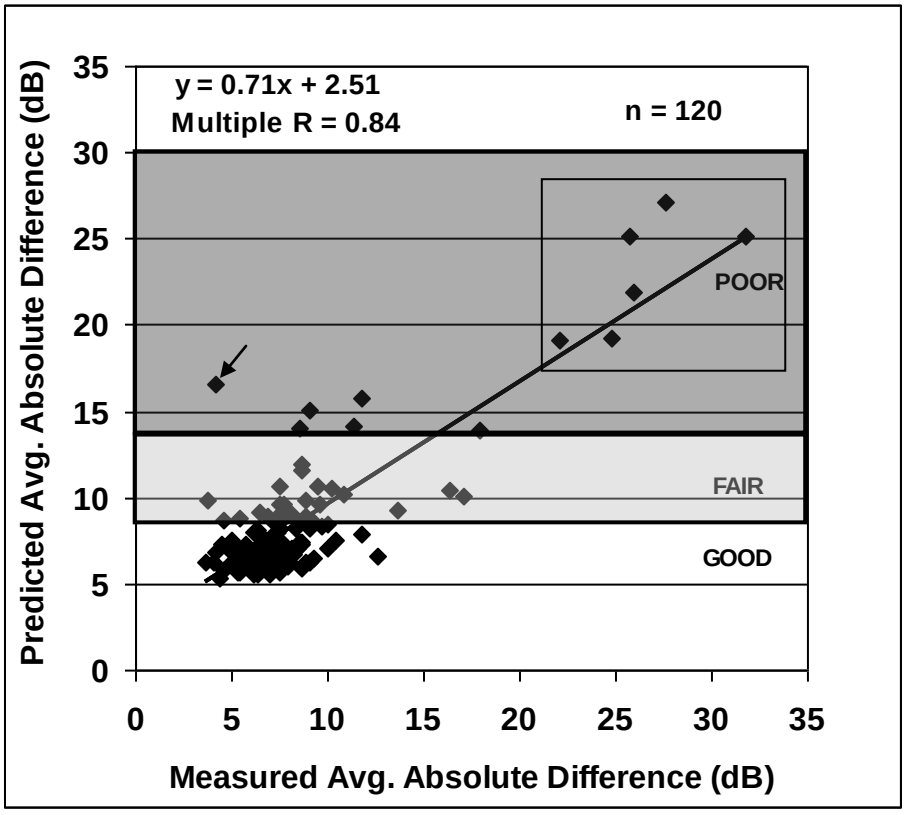
Masker Alert Rate
Time per Trial
False Alarm Rate
Test –Retest Difference at 1000 Hz
Quality Check Fail Rate
Number of Air-Bone Gaps > 50 dB
Number of Air-Bone Gaps < -10 dB
Average Air-Bone Gap

In a study conducted on 123 patients with widely varying age and hearing loss, the regression coefficient was 0.84 indicating that 70% of the variance in thresholds was predicted

by the quality indicators. This is a very strong relationship and indicates that the quality indicators are effective for predicting accuracy.

The predicted accuracy can be categorized into groups like “good”, “fair”, and “poor” by constructing a scale based on the differences that occur when the same patients are tested by two experienced audiologists. Ideally, the differences in threshold between AMTAS and an expert audiologist should be similar to those that occur between thresholds obtained by two audiologists. That would indicate that AMTAS is as accurate as an expert audiologist. To construct this scale, a group of patients were tested by two audiologists and the average absolute differences in thresholds were determined. Then the “good”, “fair”, and “poor” ranges were defined by the distribution of differences between the thresholds obtained by two audiologists. If the average absolute difference is within one standard deviation of the average difference for two audiologists, it is considered to be “good”. “Fair” is between one and two standard deviations, and “Poor” is beyond two standard deviations.

Figure 10. Prediction of Accuracy of AMTAS Audiograms



The scale is illustrated in the Figure 10 along with the predicted and measured accuracy for the 123 patients. For 78% of the cases, the predicted category was identical to the observed category. For example, the six worst cases (in terms of agreement between AMTAS and manual testing) were all categorized as “Poor”. These are the data points in the rectangle. Only one case (arrow) was placed in the poor range but the accuracy was actually good.

Back to Table of Contents

4. Bone Conduction Headband Force Levels

The ANSI audiometer standard (ANSI S3.21-2004) specifies that the bone vibrator should be placed on the head with a coupling force of 5.4 N ± 0.5 N (551 g ± 51 g). An elastic headband was designed for forehead bone conduction that provides the specified force for average head sizes.

Table 4 shows force measurements for three headbands of the preferred size for ten adult subjects. Table 5 shows the 10-90 %ile range for males and females for 3 age groups. The average head size of our subjects (22.6 in) is within the 10-90 %ile ranges for male and female 18-year olds.

Table 4. Bone Conduction Headband Force Measurements

				Headband			
Subject	Age	Sex	Circum (in)	1	2	3	Avg
JM	57	F	23.00	506	470	590	522
BNM	16	F	20.25	520	452	528	500
JCM	18	F	23.00	620	602	616	613
JB	30	M	24.00	650	720	680	683
AK	39	F	22.00	470	540	470	493
KW	18	F	21.25	520	480	506	502
NW	18	F	23.00	648	614	636	633
MZ	70	M	23.00	658	520	630	603
LB	60	M	24.50	700	658	684	681
CK	45	F	22.00	456	444	460	453
Mean	37		22.60	575	550	580	568
S.D.	20		1.26	89	95	83	84

Force was measured by placing the bone vibrator (B71) on the forehead held in place by the headband. The force required to just separate the headband from the head was measured. The mean force was 568 g (s.d. = 84 g). This value is within 3% of the force level specified in the standard.

Table 5. Normative head circumference (from Roche et al., 1987, Pediatrics, 79, 706-712)

		Roche et al. 10th - 90th %ile			
		female		male	
		cm	in	cm	in
5 yr olds	10th	49.02	19.30	49.83	19.62
	90th	52.37	20.62	53.30	20.98
12 yr olds	10th	51.35	20.22	52.35	20.61
	90th	55.12	21.70	56.20	22.13
18 yr olds	10th	53.33	21.00	54.50	21.46
	90th	57.47	22.63	58.68	23.10

Back to Table of Contents

5. AMCLASS

There are no standard definitions of the descriptive terminology used to characterize the configuration, severity, site of lesion, and symmetry of an audiogram. Nor are there independent measures against which definitions of these terms can be validated. The best method for validating AMCLASS definitions is to compare AMCLASS categories against the judgments of a panel of expert judges. Two validation studies were conducted. In the first study the configuration, severity, and site of lesion categories were validated against the judgments of five experts (four audiologists and an otologist). A complete report of that study is provided in Margolis and Saly (2007). In the second study, the experts judgments of audiometric asymmetry were compared to those of AMCLASS. A complete report of that study has been submitted for publication.

We selected 231 audiograms from a clinical database and asked a panel of five expert audiologists to select a configuration, severity, and site of lesion category for each. From their responses, a consensus was determined, that is, the category chosen most frequently by the panel of judges. This made it possible to compare the judgments of each judge with the consensus and to compare the categories selected by AMCLASS with the consensus. If

AMCLASS agreement with the consensus is as high as the agreement between the average judge and the consensus, we can claim that AMCLASS is as good as an average expert audiologist.

Agreement among judges, consensus, and AMCLASS is summarized in Table 6. Interjudge agreement was surprisingly low. The average agreement between pairs of judges for configuration indicates that for one third of cases judges disagree. On the average, they disagreed on about a fifth of the cases for severity and a third of the cases for site of lesion. The low interjudge agreement indicates that there is no consistency in how audiograms are described, even among highly experienced audiologists.

Table 6. AMCLASS Agreement

	Interjudge Agreement	Judges v. AMCLASS	Judges v. Consensus	AMCLASS v. Consensus
Configuration	67.6	75.3	83.1	89.6
Severity	82.6	85.5	88.2	92.2
Site of Lesion	74.4	77.2	86.1	84.8

The average agreement between judges and AMCLASS was higher than the average interjudge agreement. This indicates that on average the judges agreed with AMCLASS more often than they agreed with each other.

The most important comparison for assessing AMCLASS is the last two columns of Table 6. For configuration and severity, agreement between AMCLASS and consensus is higher than the average agreement between judges and consensus. This indicates that AMCLASS performs better than the average expert. For site of lesion, agreement between AMCLASS and consensus was slightly lower than the average agreement between judges and consensus. This occurred because the judges did not use the Sensorineural or Mixed category as expected. This category was included for cases of profound hearing loss where the bone conduction thresholds cannot be measured due to the maximum output limits of audiometers.

The judges tended to categorize these as Sensorineural. Rather than adjust the rules to maximize agreement with judges, which was done for configuration and severity, we chose to retain this definition of Sensorineural or Mixed because a conductive component cannot be ruled out in these cases.

To validate AMCLASS determinations of audiometric asymmetry, the panel of five experts judged 199 audiograms as symmetrical or asymmetrical. The audiograms were selected to span a wide range of interaural differences while eliminating cases were the differences were so great that there is not likely to be disagreement among judges. The average interjudge agreement for symmetry and asymmetry was 77%. The average agreement between judges and AMCLASS was 83%. For each case a consensus was determined, that is, the category selected by the majority of judges. The average agreement between AMCLASS and the consensus of judges was 91%. The results indicate that a) judges agreement with AMCLASS is higher than agreement among judges, and b) agreement between the consensus of judges and AMCLASS is higher than the average agreement between each judge and the consensus.

Back to Table of Contents

6. Bone Conduction Variability – Comparison of Forehead and Mastoid Placement

In order to derive full benefit from automating pure-tone audiometry, it is desirable to place the transducers before the test begins in a manner that avoids the need to reposition them during the test. This requires that the bone vibrator be placed in a neutral location that is appropriate for testing each ear separately. Masking is presented to the non-test ear to isolate the responses of the ears. The forehead location has been used for that purpose.

Two early reviews of research on bone conduction concluded that forehead placement is preferable to mastoid placement for several reasons (Naunton, 1963; Dirks, 1973). The reasons include smaller intra-subject variability, smaller intersubject variability, and less effect of the middle ear associated with forehead placement. The authors point out that the differences in variability are small, but seem to favor forehead placement. However, the

measurements that were available at the time were not obtained with the currently-used bone conduction transducer.

This study was undertaken to compare bone conduction thresholds and variability for bone conduction stimuli presented to normal-hearing listeners by the most commonly-used bone vibrator (Radioear B-71).

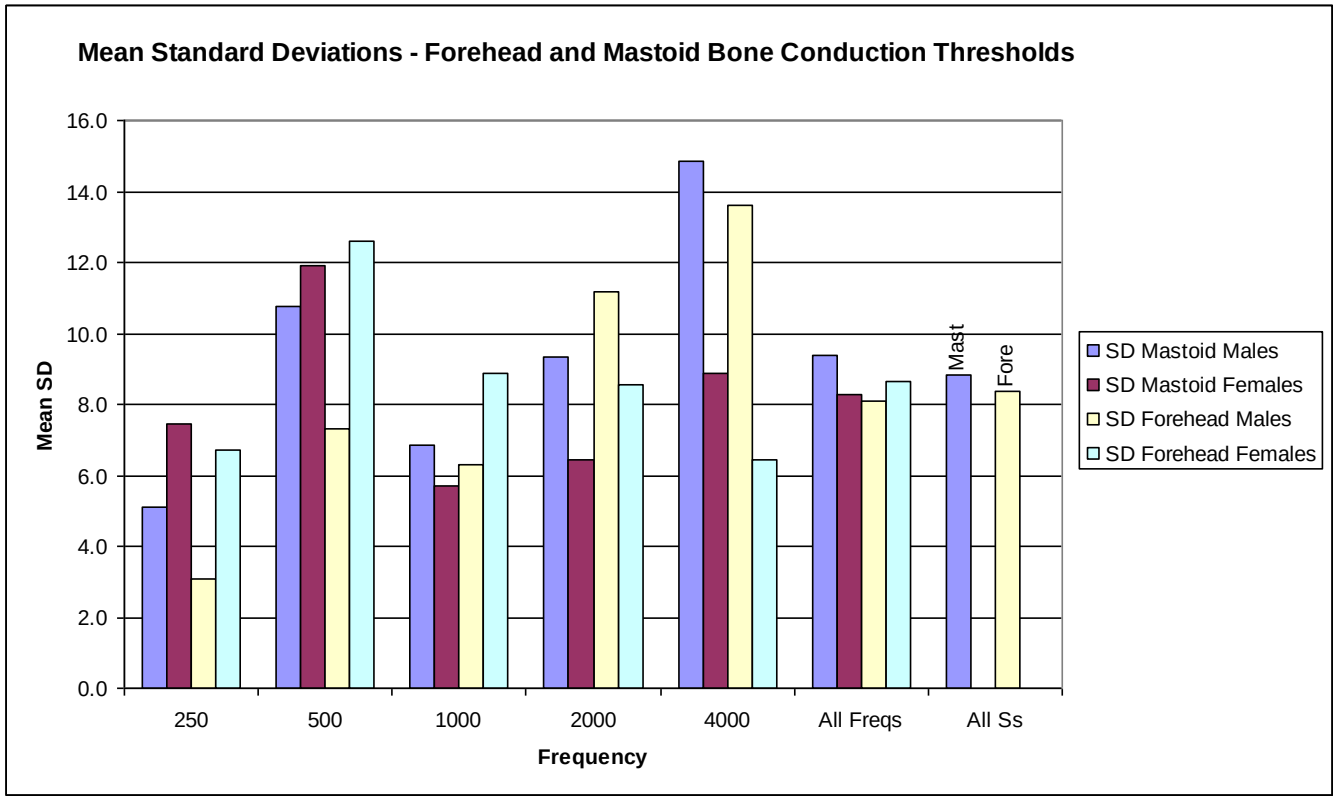
Method. Five female and five male subjects with clinically-normal hearing were tested in the Audiology Research Laboratory of the University of Minnesota Department of Otolaryngology. Two male subjects had mild, age-related, high-frequency sensorineural hearing losses. Subjects had no recent history of middle-ear disease and no suspicion of conductive hearing loss.

Air-conduction and bone-conduction thresholds were obtained by an automated method for obtaining a pure-tone audiogram (AMTAS® - see www.audiologyincorporated.com). Air-conduction stimuli were delivered by circumaural earphones (Sennheiser HDA200). Two complete tests were performed, one with mastoid placement and one with forehead placement of the bone vibrator. During the mastoid condition the test ear was uncovered and the vibrator was held on the mastoid with a spring headband used routinely in clinical audiometry that is designed to produce a coupling force of 5.4 N. The test was paused after bone conduction testing of the first ear was completed while the transducers were repositioned for testing the other ear. During testing with the forehead condition both ears were covered by the earphones and the bone vibrator was held in place by an elastic headband designed to produce the 5.4 N coupling force. Transducer were not moved during the test. Air-conduction thresholds were obtained for octave frequencies in the range 250 – 8000 Hz and at interoctave frequencies when the difference between two adjacent octave frequencies was $\geq$ 20 dB. Bone conduction thresholds were obtained for octave frequencies in the range 250 – 4000 Hz.

Results. Figure 1 shows standard deviations for each test frequency for bone conduction thresholds averaged for the right and left ears. There were no consistent differences associated with gender or stimulation site. The right most bars in the figure show

standard deviations averaged across all subjects and all frequencies for the two sites (8.8 and 8.4 dB for mastoid placement and forehead placement, respectively).

Figure 1.



Perhaps a more useful way to compare the two stimulation sites is to examine the air-bone gaps (the difference between air- and bone-conduction thresholds at each frequency). The air-bone gap removes the variability associated with cochlear sensitivity. Figure 2 presents the average standard deviations for air-bone gaps for each test frequency and for all test frequencies combined for female and male subjects. Because the variability associated with cochlear sensitivity is removed, standard deviations for air-bone gaps tend to be smaller than those for bone-conduction thresholds. The two right-most bars show the almost identical standard deviations averaged across all subjects and all frequencies for the two stimulation sites (7.8 and 8.0 dB for mastoid and forehead placement, respectively).

Bone conduction thresholds for normal listeners are subject to a “floor effect”, a lower limit on thresholds imposed by the internal noise of the instrumentation and the ear.

Consequently, the variability associated with the subjects tested in this study is expected to be less than that associated with subjects with sensorineural hearing loss. This would affect the results from the two placement sites equally. The results suggest that there is no clinically-important difference in variability associated with bone conduction testing at the two stimulation sites.

Figure 2.

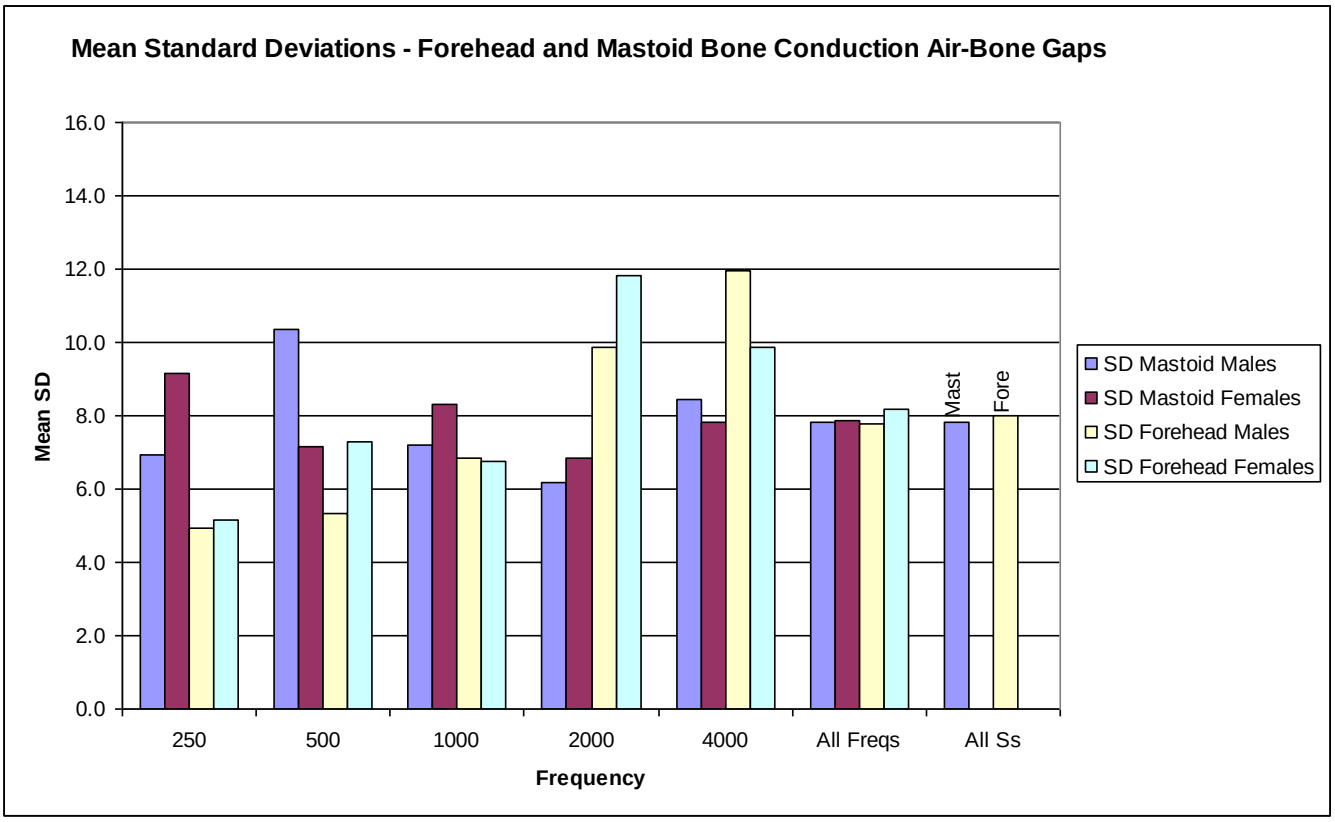


Table 1 shows the average differences between bone conduction thresholds for the two sites and the differences in Reference Equivalent Threshold Sound Pressure Levels (RETSPL) for the two sites from the International and American audiometer standards. The observed differences are less than those expected based on the standard RETSPLs. This may be partially attributable to the “floor effect” mentioned above because both measurements are limited on the low end of the range.

There has been a consistent observation in AMTAS trials and in Audiology clinics in the U.S. that erroneous air-bone gaps occur at 4000 Hz. This is probably due to an inappropriate

RETSPL for 4000 Hz bone conduction. The phenomenon was not observed in this study because the subjects had predominantly normal hearing, limiting the magnitude of the air-bone gap that could be observed. This is also due to the "floor effect."

Table 1. Mean threshold difference for mastoid and forehead placement.

	FREQUENCY (Hz)				
	250	500	1000	2000	4000
This Study	4.3	5.8	4.5	-1.0	5.3
Calibration Standards	12.0	14.0	8.5	11.5	8.0

Summary and Conclusion. Air and bone conduction thresholds were obtained with two bone conduction stimulation sites for 10 subjects with clinically-normal hearing (5 females and 5 males). The results suggest that there are no systematic differences associated with gender or bone-conduction stimulation site. The results are probably influenced by a "floor effect" that would affect both stimulation sites equally. Variability associated with bone conduction thresholds for subjects with sensorineural hearing losses is probably greater because the floor effect would not limit the variability in those subjects.

The differences in bone conduction thresholds for the two stimulation sites are less than the differences in RETSPLs in the international and American audiometer standards. This difference may be partially attributable to the floor effect that restricts the differences that can be observed at low stimulus levels.

References

Dirks DD (1973). Bone conduction measurements. In Jerger J (Ed.) *Modern Developments in Audiology*. Second Edition. New York: Academic Press.

Naunton RF (1963). The measurement of hearing by bone conduction. In Jerger J (Ed.) *Modern Developments in Audiology*, New York: Academic Press.

Back to Table of Contents

Lions International Convention Hearing Screening
Minneapolis Convention Center
6-7 July 2009

The annual convention of Lions International was held at the Minneapolis Convention Center on 6-10 July 2009. Free hearing screenings were offered all day on 6 July and a half day on 7 July. Testing was administered by audiology graduate students from the University of Minnesota Department of Speech-Language-Hearing Sciences in a meeting room designed for groups of about 150 persons. The room faced a main convention center pedestrian thoroughfare.

Two air-conduction AMTAS systems were used for testing. Each system was comprised of a touchscreen personal computer (Belview PP8510), an air-conduction, pc-controlled audiometer (Benson CCA-100), and circumaural earphones (Sennheiser HDA200). AMTAS was operated in the screening mode, which measures thresholds at three frequencies in each ear (1.0, 2.0, and 4.0 kHz). Manual screening equipment was available and was used when both AMTAS systems were in use.

Listeners were seated in front of the computer and verbally instructed to listen during the "LISTEN" message and respond by touching "YES" or "NO" during the "DID YOU HEAR THE BEEP" message.

One hundred ten listeners were tested by AMTAS during the 1.5 days. No demographic information was obtained. The subjects were predominantly male seniors. Many were not native English speakers. There were no tests that were attempted and not successfully completed. For each ear a "pass" or "fail" was automatically determined by AMTAS. A pass was defined as thresholds less than 25 dB at 1 and 2 kHz and less than 30 dB at 4 kHz.

RESULTS

Results are summarized in Table 1. About 70% of ears tested failed the screening test. Twenty-two percent of subjects tested passed the test in both ears.

Minimum thresholds ranged from -10 to 0 dB HL at the six frequency-ear combinations, indicating that the circumaural earphones provided adequate ambient noise attenuation to permit valid testing in the meeting room environment. Maximum thresholds ranged from 85 to >100 dB HL indicating that some subjects have severe hearing losses.

The average number of trials required for the test was 55 (9 per threshold). Because 20% of trials are catch trials (no stimulus) the average number of stimuli presented was about 45. The average test time was 3.7 min.

The mean false alarm rate was 0.04 with a standard deviation of 0.08. In the multicenter AMTAS validation study, the mean was 0.08 with a standard deviation of 0.10. This indicates that the subjects in this study had low false alarms rates.

The average number of Quality Check Fails was 0.05 with a minimum of 0 and a maximum of 3. This corresponds to a Quality Check fail rate of 0.01 (one fail per 100 threshold determinations). One subject had three Quality Check Fails, three subjects had one, and 106 subjects had none. The mean Quality Check Fail rate in the multicenter validation study was 0.02 indicating that the subjects in this study had a low rate of Quality Check Fails.

The low false alarm rates and low occurrence of Quality Check Fails suggest that the subjects had no difficulty performing the test accurately.

Table 1.

	Right			Left			Right & Left				Elapsed	False	No. QC
Freq	1k	2k	4k	1k	2k	4k	1k	2k	4k	No. Trials	Time(min)	Alarm Rate	Fails
Mean	23.5	25.1	38.1	21.3	24.2	39.3	22.4	24.7	38.7	55.2	3.7	0.04	0.05
SD	14.3	17.8	24.5	14.5	19.0	24.9	14.4	18.4	24.7	6.0	0.7	0.08	0.33
n	110	110	110	110	110	110	220	220	220				
Max	85	85	100	80	90	100	85	90	100	71	6.9	0.50	3.00
Min	-10	0	-5	-5	-10	-5	-10	-10	-5	36	2.5	0.00	0.00
% Pass		29			32			22					

SUMMARY AND CONCLUSION

Attendees of the Lions International annual convention, held in Minneapolis 6-7 July 2009 were offered free hearing screenings. This report summarizes results of 110 subjects. Automated testing was performed with a screening version of AMTAS that obtains air-conduction thresholds at 1.0, 2.0, and 4.0 kHz. Stimuli were delivered by circumaural earphones (Sennheiser HDA 200) that provide good ambient sound attenuation. Testing was performed in a convention center meeting room that is adjacent to a main pedestrian thoroughfare. We conclude the following from this study.

1. Subjects were able to perform the test successfully. Quality indicators (false alarm rate and Quality Check Fail rate) indicated better performance for this group of subjects when compared to the results of the AMTAS multicenter trial.

2. The majority of subjects failed the screening test. This is a common finding in programs like this in which older subjects are tested. Many subjects were aware that their hearing was not normal.
3. Minimum thresholds ranged from -10 to 0 dB HL for the six frequency-ear combinations indicating that ambient noise did not significantly affect test results.

4. Average test time was under 4 min per test indicating that the method is acceptable for screening programs.

ACKNOWLEDGEMENTS

We are grateful to the Lions 5M Hearing Foundation for making arrangements to offer hearing screening at the Lions International convention. The participation of students and faculty of the Department of Speech-Language-Hearing Sciences is appreciated. We are grateful for Chad Johnson's assistance in data analysis.

Back to Table of Contents

AUTOMATED AUDIOMETRY PUBLICATIONS

1. Margolis, R.H. Automated Audiometry – Progress or Pariah? Audiology Online (www.audiologyonline.com/articles). January 17, 2005.

2. Margolis, R.H. Automated Audiometry: Progress or Pariah. Audiology Today 17:21, 2005.

3. Margolis, R.H., Saly, G., Le, C., Laurence, J. Qualind™: A Method for Assessing the Accuracy of Automated Tests. J. Amer. Acad. Audiol., 18, 78-89, 2007 .

4. Margolis, R.H., Saly, G.S. Toward a standard description of hearing loss. Int. J. Audiology 46:746-758.

5. Margolis, R.H., Morgan, D.E. The Value of Automated Audiometry. Insights in Practice for Clinical Audiology, January 2008.

6. Margolis, R.H., Saly, G.L. Prevalence of hearing loss types in a clinical population. Ear and Hearing, 29, 524-532, 2008.

7. Margolis, R.H., Saly, G.L. Toward a standard description of hearing loss. Int. J. Audiology, 46, 746-758, 2007.

8. Margolis, R.H., Saly, G.L. Asymmetrical Hearing Loss: Definition, Validation, Prevalence. Otology & Neurotology, 29, 422-431, 2008.

9. Margolis, R.H., Morgan D.E. Automated Pure-Tone Audiometry – An Analysis of Capacity, Need, and Benefit. Amer. J. of Audiology, 17, 109-113, 2008.

10. Margolis, R.H., Glasberg, B.R., Creeke, S., Moore, B.C.J. AMTAS - Automated Method for Testing Auditory Sensitivity: Validation Studies. Int. J. Audiology, in press.

Back to Table of Contents