Published in final edited form as:

Ear Hear. 2018 ; 39(5): 906–909. doi:10.1097/AUD.0000000000000551.

Home Hearing Test™: Within Subjects Threshold Variability

Robert H. Margolis¹, Gene Bratt², M. Patrick Feeney³, Mead C. Killion⁴, and George L. Saly¹

¹Audiology Incorporated, Arden Hills, Minnesota

²Tennessee Valley Healthcare System, Nashville, Tennessee

³Oregon Health and Science University and VA National Center for Rehabilitative Auditory Research, Portland OR

⁴Etymotic Research, Elk Grove Village, Illinois

Abstract

Background—The Home Hearing Test (HHT) is an automated pure-tone threshold test that obtains an air-conduction audiogram at five test frequencies. It was developed to provide increased access to hearing testing and support home telehealth programs.

Purpose—Test and retest thresholds for 1000-Hz stimuli were analyzed to determine intra-subject variability from two independent data sets.

Research Design—Prospective, repeated measures.

Study Sample—In the VA Study results from twenty-six subjects, aged 44–88 years (mean = 65) recruited from the Nashville VA audiology clinic were analyzed. Subjects were required to have a Windows PC in the home and were self-reported to be comfortable with using computers. Two subjects had normal hearing and 24 had hearing losses of various severities and configurations.

The NCRAR sample included 100 subjects (68 males; 32 females) with a complaint of hearing difficulty recruited from the local community and veteran population. Subjects ranged in age from 32 to 87 years (mean = 63.7 years). They were tested in a quiet room at the NCRAR.

Data Collection and Analysis—Subjects in the VA study were provided kits for installing HHT on their home computers. HHT was installed on a computer at NCRAR to test subjects in the NCRAR study.

HHT obtains a five-frequency air-conduction audiogram with a retest of 1000 Hz in both ears. Only the 1000-Hz test-retest results are analyzed in this report. Six statistical measures of test-retest variability are reported.

Results—Test and retest thresholds were highly correlated in both studies ($r = 0.96$). Test-retest differences were within ± 5 dB 92 % of the time in the two studies. Standard deviations of absolute test-retest difference were 3.5 dB in the two studies.

Conclusions—Intra-subject variability is comparable to that obtained with manual testing by audiologists in sound-treated test rooms.

Keywords

Pure-tone audiometry; Home Hearing Test; AMTAS; telehealth

INTRODUCTION

The poor penetration of hearing services to those in need of services, a topic of discussion among hearing professionals for many years, is gaining traction among government agencies. A recent report (National Academies of Sciences, Engineering, and Medicine, 2016) estimated that 67 to 86 percent of the 30 million Americans with hearing loss do not use hearing aids or other hearing assistive technologies. The report cites several reasons for this shortfall including cost, overall dissatisfaction with hearing healthcare, and lack of access to services. Recommendation 10 of that report calls for the development of innovative models for hearing health care including “self administered hearing health care”. The wide gap between the capacity and need for hearing services (Margolis & Morgan, 2008; Windmill & Freeman, 2013) is palpable evidence that innovative approaches are needed to reduce the burden of hearing impairment. The Home Hearing Test™ (HHT) was developed to provide an accurate assessment of hearing sensitivity that can be administered on a home computer.

In a previous report (Margolis et al. 2016) we described an evaluation of the HHT that was conducted as part of a project to evaluate automated hearing testing for the U.S. Department of Veterans Affairs (VA). HHT audiograms were compared to audiograms that had been obtained by manual audiometry in a VA hospital clinic 0–145 days before the HHT was performed. In this article we expand the analysis to include an examination of within-subjects variability for the VA data set and for audiograms obtained in the course of a separate project aimed at exploring methods for increasing the participation of hearing-impaired individuals in the hearing care process. The latter project is a multicenter study based at Oregon Health and Science University and the National Center for Rehabilitative Auditory Research (NCRAR) at the VA Portland Health Care System in Portland, OR.

METHODS

Methods were approved by the Institutional Review Board at each institution and subjects signed informed consent forms.

The Home Hearing Test—The Home Hearing Test, distributed by Etymotic Research (Elk Grove Village, Illinois), consists of a kit that includes the following components:

- USB sound device that self-installs on Window computers
- pair of insert earphones (modified Etymotic Research mc5) with a variety of ear tips
- USB memory drive that stores the AMTAS software and the installation program

- user manual
- educational booklet that explains hearing testing and hearing loss.

The hearing test is conducted by AMTAS[®], a validated method for measuring pure-tone thresholds. AMTAS (Automated Method for Testing of Auditory Sensitivity) is a single-interval, forced-choice, adaptive psychophysical procedure. AMTAS has been found to produce threshold measurements that are equivalent to those obtained by experienced audiologists (Mahomed et al., 2013; Margolis et al., 2010; Margolis, Frisina, & Walton, J.P., 2011; Margolis & Moore, 2011). The AMTAS version employed by HHT measures air-conduction thresholds at octave frequencies from 500 – 8000 Hz with a retest at 1000 Hz in each ear. Step sizes are initially 10-dB and then 5-dB near threshold. The 1000-Hz test and retest thresholds are analyzed in this report to examine within-subjects variability.

Subjects are instructed to take the test in a quiet room that is free of distractions. The insert earphones provide some attenuation of ambient noise. The instructions that precede the administration of HHT are communicated by a non-verbal, animated video that demonstrates correct and incorrect responses to sound stimuli.

A video is provided that instructs the listener on selection of tips and placement of the earphones. The tips that are provided vary in size and accommodate all normal adult ears. Subjects in the VA study selected the tips and placed the earphones as instructed in the video. In the NCRAR study the tips were selected and earphones were placed by the audiologist.

A method for estimating the accuracy of the test (Qualind[®]) is embedded in AMTAS (Margolis et al., 2007). In the air-conduction version of AMTAS, Qualind tracks four variables (false alarm rate, test-retest difference at 1000 Hz, time per trial, and quality check fails) that have been found in a previous trial to correlate to test accuracy (defined as agreement between thresholds obtained by AMTAS and by expert audiologists). The values of the four variables are entered into a pre-determined regression formula to calculate the predicted average absolute difference, the predicted value of the mean of the absolute differences between AMTAS and clinical thresholds, averaged over the entire audiogram. The predicted average absolute differences are stratified into three statistically-defined categories – Good, Fair, and Poor.

Analysis—Within-subjects variability was evaluated by six statistical methods.

1. Test-retest reliability was examined with a Pearson product-moment correlation coefficient.
2. Test and retest thresholds were compared by *t* tests (paired two-sample for means, Microsoft Excel).
3. The mean and standard deviation of the test-retest threshold differences were calculated.
4. The percentage of test-retest differences in the range ± 5 dB was determined.

5. The mean and standard deviation of the absolute test-retest threshold differences were calculated.
6. The average within-subjects standard deviation was calculated by the following formula:

$$SD = \sqrt{\left(\frac{\sum (d/\sqrt{2})^2}{n} \right)} \quad \text{Eq 1}$$

Where d is the difference between test and retest thresholds and n is the number of threshold pairs.

VA Study

Subjects—Twenty-eight subjects (27 males; 1 female) were recruited from the Audiology Clinic of the Tennessee Valley Health System VA hospital in Nashville, Tennessee. Subjects were required to have had a hearing evaluation in the clinic within the past six months, have a personal computer at home running Windows 2007 or higher, and express a comfort level with computer usage. Age and hearing loss characteristics were not considered in subject selection. The subjects ranged in age from 44 to 88 years (mean = 65 years). Subjects received a HHT kit for their personal use and were not otherwise compensated. Due to technical reasons the data from two subjects were lost leaving results from 52 ears of 26 subjects available for analysis. HHT tests were obtained between February and November of 2014. The average air-conduction audiogram for these subjects (right and left ears combined) is shown in Figure 1.

Method—Hearing evaluations performed in the clinic were conducted by four experienced licensed audiologists in sound-attenuating rooms using the standard clinical protocols that are employed in that clinic. The audiologists had 10–37 years of clinical audiology experience. After obtaining written informed consent, an HHT kit was provided to the subject with instructions on how to install the software on their home computer. Results were saved on a USB flash drive and mailed to the VA Audiology Clinic for analysis.

NCRAR Study

Subjects—One hundred subjects (68 males; 32 females) with a complaint of hearing difficulty were recruited from the local community and Veteran population. Subjects ranged in age from 32 to 87 years (mean = 63.7 years). Fifty-three of the subjects were non-veterans. Subjects were excluded from participation if they had a hearing evaluation in the past three years, or had ever worn hearing aids. Subjects were paid for their participation in the study and received a version of the HHT for their personal use. The average air-conduction audiogram for these subjects (right and left ears combined) is shown in Figure 1.

Method—This work was a part of a larger study to evaluate several automated hearing screening systems. Study procedures were performed by two licensed audiologists in a quiet room. After obtaining written informed consent, otoscopy was performed by the study

audiologist who then inserted the subject's earphones for the HHT. Subjects then watched a video instructing them on how to use the HHT. The researcher then started the evaluation. Following the HHT, each subject watched a counseling video that explained their test results. The tests were performed between September 2016 and March 2017.

RESULTS

VA Study

Of the 26 audiograms that were analyzed, 23 were judged by Qualind to be Good, two were Fair, and one was Poor. The distribution of 1000-Hz test-retest threshold differences for the 52 ears of 26 subjects is shown in Figure 2. Table 1 summarizes the results of statistical analyses. Test and retest thresholds were highly correlated ($r = 0.96$) and were not statistically different ($P = 0.17$). The mean test-retest difference was less than 1.0 dB ($SD = 4.51$). Ninety-two percent of the test-retest difference were ≤ 5 dB. Because test and retest thresholds were not statistically different, the order of testing and thus the sign of the test-retest difference is unimportant. A better indication of the dispersion of differences is the standard deviation of the absolute differences (3.5 dB) which is very close to the average standard deviation calculated from Eq. 1 (3.2 dB).

NCRAR Study

Of the 100 audiograms that were analyzed, 85 were judged by Qualind to be Good, 10 were Fair, and 5 were Poor. The distribution of threshold differences for the 200 ears is shown in the bottom panel of Figure 2. Table 1 summarizes the results of statistical analyses. Like the VA study data, test and retest thresholds were highly correlated ($r = 0.97$) and were not statistically different ($P = 0.29$). The mean test-retest difference was 1.48 dB ($SD = 3.68$). Test-retest differences were ≤ 5 dB 93.5% of the time. The standard deviation of the absolute differences (3.2 dB) is very close to the average standard deviation calculated from Eq. 1 (2.8 dB).

DISCUSSION

There is a surprising lack of intrasubject variability data for audiometric thresholds. Skinner (1988) presented a summary table of repeated threshold measures from nine studies. The standard deviations of repeated measures ranged from 1.9 – 4.6 dB, consistent with the results of this report. Various threshold measurement procedures were used, not including typical clinical methods. Step sizes of 1, 2, and 2.5 dB were used in various studies and Bekesy audiometry was used in others. Nevertheless the range of standard deviations encompasses the values reported here.

Schmurziger, Probst, and Smurzynski (2004) obtained test and retest thresholds using a clinical (Hughson-Westlake) method for two earphones (circumaural and insert) for conventional audiometric frequencies (0.5 – 8.0 kHz) and extended high frequencies (9 – 16 kHz). They reported the percent of differences in the -5 to 5 dB range and the -10 to 10 dB range. In the conventional frequency range the percent of differences that were -5 to 5 dB ranged from 92 to 99 % for the circumaural earphone and 96 – 99 % for the insert earphone. These are similar to the results of the two studies reported here. Schmurziger et al. reported

that test-retest differences did not appear to be related to frequency suggesting that the results at 1 kHz reported here are generalizable to other frequencies in the conventional audiometric range.

Two limitations of these findings deserve mention. First, the different methods for placing the earphones used in the two studies is a potentially complicating variable. The tips were selected and the earphones were placed by the subjects in the VA study and by the audiologist in the NCRAR study. The fact that earphones were not replaced between test and retest thresholds and the agreement between the two studies evident in Table 1 suggests that the method of placement did not affect the results. Second, ambient noise was controlled in the NCRAR study but not in the VA study. The agreement in results suggests that the subjects in the VA study followed the instruction to perform the test in a quiet room that is free of distractions. Future users of HHT may take the test in less optimal conditions that could significantly affect intra-subject variability.

SUMMARY AND CONCLUSIONS

Two studies are reported on intra-subject variability of audiometric thresholds obtained with the Home Hearing Test™. In one study subjects installed and ran HHT on their home computers. In the other study testing was conducted in a sound booth. The test obtains a test and retest threshold at 1000Hz on each ear. Results indicate that a) test and retest thresholds are highly correlated ($r = 0.96$); b) test and retest thresholds are not statistically different ($p = 0.17$); c) mean test-retest differences are -1.5 dB; d) test-retest differences fall within the range ± 5 dB 92% of the time; e) mean absolute test-retest differences are 3 dB; f) standard deviations of test-retest differences are 4.5 dB; and g) standard deviations of absolute test-retest differences are 3.5 dB. Test-retest variability for the two studies reported here are similar to variability of repeated threshold measurements obtained under laboratory conditions. HHT appears to be a viable tool for increasing access to hearing testing.

Acknowledgments

This work was supported by contract VA118-12-C-0029 from the U.S. Department of Veterans Affairs and Grant No. 4R33DC011769 from the National Institute on Deafness and Other Communicative Disorders

Drs. Josephine Helmbrecht, Rachel Tomasek, and Jay Vachhani provided valuable assistance. The first, fourth, and fifth authors have commercial interests in the Home Hearing Test™. The content of this report is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health or the Department of Veterans Affairs.

The Home Hearing Test was developed in response to the recommendations from the Workshop on Accessible and Affordable Hearing Health Care for Adults with Mild to Moderate Hearing Loss, sponsored by the National Institute on Deafness and Other Communications Disorders, held on August 25 – 27, 2009, in Bethesda, Maryland. See Donahue, Dubno, and Beck (2010) for a report of the workshop.

This work was supported by contract no. VA-14-0005253 from the U.S. Department of Veterans Affairs and grant no. R33DC011769 from the National Institute on Deafness and Other Communication Disorders.

References

- Donahue A, Dubno JR, Beck L. Accessible and affordable hearing health care for adults with mild to moderate hearing loss. *Ear and Hearing*. 2010; 31:2–6. [PubMed: 20040828]
- Mahomed F, Swanepoel DW, Eikelboom RH, Soer M. Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear and Hearing*. 2013; 34:745–752. [PubMed: 24165302]
- Margolis RH, Frisina R, Walton JP. Automated method for testing auditory sensitivity: II. Air Conduction Audiograms in Children and Adults. *Int J Audiology*. 2011; 50:434–439.
- Margolis RH, Glasberg BR, Creeke S, Moore BCJ. AMTAS[®] - Automated method for testing auditory sensitivity: validation studies. *Int J Audiology*. 2010; 49:185–194.
- Margolis RH, Killion MC, Bratt GW, Saly GL. Validation of the Home Hearing Test. *J Amer Acad Audiol*. 2016; 27:416–420. [PubMed: 27179261]
- Margolis RH, Moore BCJ. Automated method for testing auditory sensitivity: III. Sensorineural hearing loss and air-bone gaps. *Int J Audiology*. 2011; 50:440–447.
- Margolis RH, Morgan DE. Automated pure-tone audiometry: an analysis of capacity, need, and benefit. *Amer J Audiology*. 2008; 17:109–113.
- Margolis RH, Saly G, Le C, Laurence J. Qualind[™]: A method for assessing the accuracy of automated tests. *J Amer Acad Audiol*. 2007; 18:78–89. [PubMed: 17252960]
- National Academies of Sciences, Engineering, and Medicine. Hearing Health Care for Adults: Priorities for Improving Access and Affordability. 2016. <http://nationalacademies.org/hmd/reports/2016/hearing-health-care-for-adults.aspx>
- Schmurziger N, Probst R, Smurzynski J. Test retest reliability of pure-tone thresholds from 0.5 to 16 kHz using Sennheiser HDA 200 and Etymotic Research ER-2 earphones. *Ear and Hearing*. 2004; 25:127–132. [PubMed: 15064657]
- Skinner MW. *Hearing Aid Evaluation*. Englewood Cliffs, NJ: Prentice Hall; 1988.
- Windmill IM, Freeman BA. Demand for audiology services: 30-Yr projections and impact on academic programs. *J Am Acad Audiol*. 2013; 24:407–416. [PubMed: 23739060]

Figure 1. Average audiograms for subjects in the VA study (52 ears of 26 subjects) and the NCRAR study (200 ears of 100 subjects). Vertical bars are standard deviations.

Figure 2. Distribution of test-retest differences for 52 ears of 26 subjects in the VA study and 200 ears of 100 subjects in the NCRAR study.

Table 1

Intra-subject variability of test and retest thresholds at 1000 Hz.

	n (cars)	Mean Test Threshold	Mean Retest Threshold	r	t test P (2 tail)	Mean Test-Retest Diff	SD of Test-Retest Diffs	% Diffs $\pm$ 5 dB	Mean Abs Test-Retest Diff	SD of Abs Diffs	Avg SD
VA12	52	29.0	29.9	0.96	0.17	-0.87	4.51	92	3.0	3.5	3.2
R33	200	22.2	20.7	0.97	0.29	1.48	3.68	93.5	2.4	3.2	2.8