



Journal of Rehabilitation Research and Development Vol. 38 No. 5, September/October 2001
Pages 567-581

Reliability of hearing thresholds: Computer-automated testing with ER-4B Canal Phone™ earphones

James A. Henry, PhD; Christopher L. Flick, BS; Alison Gilbert, MS; Roger M. Ellingson, MS; Stephen A. Fausti, PhD

National VA Rehabilitation Research and Development Center for Rehabilitative Auditory Research, Portland VA Medical Center, Portland, OR; Department of Otolaryngology, Oregon Health Sciences University, Portland, OR

Abstract—This study was conducted to document test-retest reliability of hearing thresholds using our computer-automated tinnitus matching technique and Etymotic ER-4B Canal Phone™ insert earphones. The research design involved repeated threshold measurements both within and between sessions, and testing to evaluate the potential effect of eartip removal and reinsertion. Twenty normal-hearing subjects were evaluated over two testing sessions with the use of a fully automated protocol for determining thresholds with 1-dB precision. Thresholds were first obtained at 0.5–16.0 kHz, in one-third octave frequency steps (16 test frequencies). The octave frequencies were then retested, first without removing the eartips, then after eartip removal and replacement. Responses between sessions differed by an average of 2.5 dB across all 16 test frequencies, and 91.5 percent of the repeated thresholds varied within ± 5 dB (98.1 percent within ± 10 dB). Reliability of within-sessions thresholds was also good, and there was no effect of eartip removal and replacement.

Key words: auditory threshold, hearing, reliability of results.

INTRODUCTION

Efforts are ongoing at the Rehabilitation Research and Development (RR&D) National Center for Rehabilitative Auditory Research (NCRAR) to develop clinical techniques for quantifying the phantom acoustical sensations that define tinnitus. A basic premise of this work is that patients, by “listening” to their tinnitus, can control the adjustment of acoustical parameters of external sounds to match these parameters to their tinnitus. By so doing, an acoustical image of the tinnitus can be created that can be useful for a variety of clinical and research purposes (1,2). Using our automated testing technique, individuals with essentially nonfluctuating tinnitus can match their tinnitus loudness very reliably to pure tones across the audible frequency range (3). Additional studies are in progress to develop automated methods for matching tinnitus pitch and for assessing other acoustical parameters of tinnitus, such as its maskability and spectral content.

This material is based upon work supported by the Veterans Affairs Rehabilitation Research and Development (RR&D) Service (C891-RA and RCTR 597-0160).

Address all correspondence and requests for reprints to James A. Henry, PhD, RR&D Center for Rehabilitative Auditory Research, VA Medical Center (NCRAR), PO Box 1034, Portland, OR 97207; email: henryj@ohsu.edu.

Historically, methodological variations for matching tinnitus loudness and pitch have been myriad. A common element of most methods, however, has been the requirement to obtain hearing thresholds. Each threshold serves as the level from which to begin matching tinnitus loudness at a given test frequency, and also as the point from which to calculate sensation levels of the loudness matches. Because loudness matches are usually determined to within 1 dB, hearing thresholds must also be obtained with 1-dB precision. Tinnitus pitch matching is often part of an interleaved testing protocol that involves evaluation of thresholds, loudness matches, and pitch matches (4–6).

Our automated tinnitus-matching protocol also involves measurement of hearing thresholds. With the automated system, a number of factors could affect test-retest reliability of the thresholds, including (1) a unique computer algorithm for obtaining thresholds, (2) the measurement of thresholds with 1-dB resolution, (3) the use of Etymotic Research (Elk Grove Village, IL) ER-4B Canal Phone™ insert earphones, and (4) reinsertion of eartips for the insert earphones. The present study was conducted, therefore, to demonstrate within-subject, within-session, and between-session reliability of hearing thresholds obtained with the automated tinnitus system, in a group of normal-hearing individuals.

METHOD

Subjects

Twenty subjects with normal-hearing sensitivity completed all testing. One ear was selected as the test ear for each subject, and only that ear was tested. For the test ear, the subjects were required to have hearing thresholds ≤ 25 dB hearing level (HL) at octave frequencies from 0.25–8.00 kHz, and at 3 and 6 kHz. Subjects consisted of 16 females and 4 males ranging in age from 19–54 y (mean=33.9 y; SD=10.8 y).

Computer-Automated Testing System

The equipment used for this study has been described in detail (3) and is described briefly herein. There were four major system components: (1) main computer, (2) subject computer, (3) signal-conditioning module, and (4) the ER-4B insert earphones. A block diagram of this system has been shown (refer to Figure 1, Henry et al. (3)). Both the main and subject computers used the Microsoft Windows 95 operating

system, and all custom software was Windows 95 compatible.

Main Computer

The main computer (Dell Dimension, 166 MHz Pentium CPU) resided in a control room and was used to control all testing functions. A 16-bit signal generator card (National Instruments, AT-DSP2200-128k) was installed in one of the peripheral card slots of the computer. A custom software application was developed to control all processes necessary for the delivery of pure tone signals to the earphones, including generation of pure tone signals from the signal generator card, and attenuation parameters for the signal conditioning module.

The main computer was connected to the subject computer via a local area network (LAN) interface with the use of standard networking protocols for two-way communication. The custom software application of the main computer communicated with the subject computer over the network. As pen-touch responses were made on the subject computer, the main computer received and analyzed these responses for program control and recorded the responses into data files. The software program of the main computer also provided dialog forms on the main-computer monitor for examiner entry of subject information, test session information, parameters for testing, and visual displays for monitoring testing status, progress, and results.

Subject Computer

The subject computer (Compaq Concerto 4/25) was selected specially to provide the testing interface between the individual being tested and the main computer. This notebook computer was enabled for Microsoft Windows for Pen; that is, the subject used a pen-pointing device to indicate responses by “pen-touching” the appropriate buttons on the touch-sensitive video screen.

The subject computer resided in the testing booth. A remote custom software application, under control of the main computer, displayed testing instructions for the subject, received the subject's responses during testing, and transmitted response information to the main computer. Acoustic and electrical noise emanating from the subject computer was not a concern because the computer was operated under battery power, there was no fan, and the hard drive was disabled during testing.

Signal-Conditioning Module

A signal-conditioning module (custom-built at Oregon Hearing Research Center, Portland, OR) was installed in-line between the signal generator of the main computer and the earphones and was used for signal mixing, attenuation, and earphone buffering.

ER-4B Canal Phone™ Insert Earphones

ER-4B Canal Phone™ insert earphones (www.etymotic.com) are designed to be used as high-fidelity studio monitor quality earphones. **Figure 1** provides photographs of the ER-4B earphone. The ER-4B utilizes an ear-level transducer, eliminating the long tubing associated with Etymotic Tubephone™ insert earphones. The ER-4B provides greater overall output and enhanced high-frequency response (above 6000 Hz) relative to the other insert phones (**Figure 2**). Sound output is >100 dB

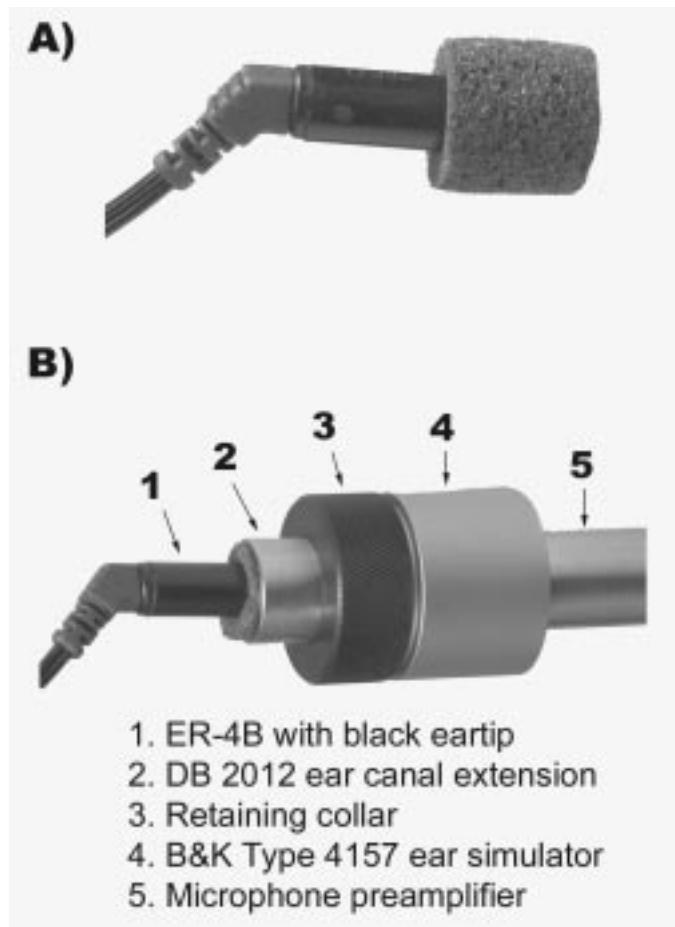


Figure 1.

ER-4B Canal Phone™ insert earphone (a) shown prior to insertion into human ear and (b) shown coupled to the B&K Type 4157 ear simulator for calibration.

sound pressure level (SPL) output from 1 to 16 kHz, with <3 percent harmonic distortion. Black foam eartips (ER4-14F) from Etymotic Research were used during both calibration and testing.

Instrumentation for Conventional Audiometry

Conventional-frequency (0.25–8.00 kHz) hearing thresholds were obtained using a Virtual Corporation (Portland, Oregon) Model 320 audiometer with TDH-50P earphones in MX-41/AR cushions. Instrumentation and procedures for manual threshold evaluation were as previously described (7). Tympanometric screening was performed with a Grason-Stadler GSI-37 Auto Tymp.

Calibration

Details of calibration have also been described (3). Briefly, output of all pure tones was calibrated at the beginning of each test day with the use of a custom automated-calibration application. The application used serial interface control of a Brüel and Kjaer (B&K) Instruments (Copenhagen) 2231 sound level meter with Type 1625 octave filter set. The ER-4B insert earphone was coupled to the sound level meter using a B&K Type 4157 ear simulator as shown in **Figure 1(b)**. A black foam eartip of the same type used for testing (Etymotic ER4-14F) was

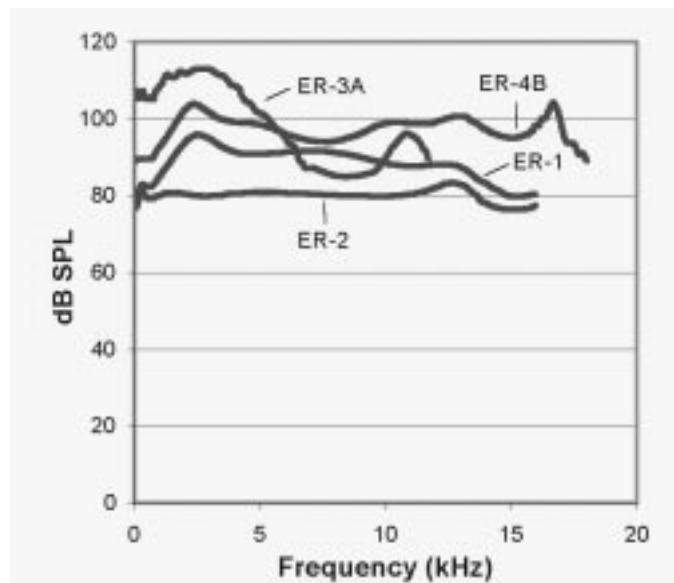


Figure 2.

Swept-frequency output, in dB SPL, for four types of Etymotic Research insert earphones, using a fixed voltage in a Zwischenstufen coupler. ER-4B Canal Phone™ had highest relative output for frequencies above 6000 Hz (data provided by Etymotic Research).

applied to the ER-4B earphone and inserted and aligned flush with the base of the B&K DB2012 Ear Canal Extension (this ensured consistent placement for calibration). Calibration values were stored in a database and later accessed, while testing, to provide precise attenuation settings.

The conventional-frequency earphones (TDH-50P) were calibrated in compliance with American National Standards Institute standards (8) with a B&K 2231 sound-level meter with a one-third-octave band filter set in an artificial ear (B&K 4153).

Procedures

For each subject, procedures were conducted over two test sessions that were separated by no more than 1 wk. Session 1 required 1.00–1.25 h of time, and Session 2 required less than 1 h.

Initial Evaluation (Session 1 Only)

At the start of the first session, a short case history was obtained to provide information regarding demographics, auditory and vestibular disorders, and family history of hearing loss. Subjects were also asked if they had been exposed to significant noise and, if so, they completed a noise exposure questionnaire.

Tympanometric screening was performed with the Auto Tymp to rule out active middle-ear pathology. Before we tested with the automated technique, hearing thresholds were obtained manually with the Virtual 320 audiometer at octave frequencies from 0.25–8.00 kHz, and at 3 and 6 kHz.

Selection of Test Ear

Subjects had little, if any, difference in hearing sensitivity between ears. If one ear appeared to have better sensitivity, it was chosen as the "test ear." If the ears were about equal in sensitivity, the test ear was selected randomly.

Experimental Protocol (Both Sessions)

In order to evaluate test-retest reliability of threshold responses of subjects both within and between sessions, we repeated thresholds within sessions and repeated all testing during a second session.

There were three stages of testing during each session. The first stage was to evaluate hearing thresholds at all frequencies in the frequency range 0.5–16.0 kHz, in one-third-octave steps (16 test frequencies). For the second stage, thresholds were repeated, but only at the

octave frequencies between 0.5–16.0 kHz (six test frequencies). This second stage was conducted immediately following the first stage and without removing the foam eartip from the subject's ear canal. Testing in the third stage was identical to the second stage, except the foam eartip was removed and reinserted before retesting. With the eartip removed, the subject was encouraged to take a short break, which usually consisted of 5–10 min outside of the testing booth.

Foam Eartip Insertion

The examiner inserted the eartip for the ER-4B earphone by making the outside eartip-surface flush with the concha bowl. If the eartip could not be inserted to that depth, it was inserted as far as possible without undue forcing.

Instructions to Subjects

Instructions for responding were presented at the beginning of each of the three testing stages. This was accomplished by displaying the instruction screen shown in **Figure 3(a)**. When subjects had read and understood the instructions, they touched the "Go" button on the screen with the pen device. The threshold-testing screen then appeared (**Figure 3(b)**), and testing proceeded.

Test Frequencies

Test frequencies for hearing thresholds obtained in Stage 1 included 0.5, 0.62, 0.8, 1.0, 1.26, 1.6, 2.0, 2.52, 3.18, 4.0, 5.04, 6.36, 8.0, 10.08, 12.7, and 16.0 kHz, and testing proceeded in a stepwise fashion, in this frequency order. For Stages 2 and 3, only octave frequencies were tested, which included 0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 kHz.

Operational Definition of Hearing Thresholds

The goal for obtaining hearing thresholds with the automated system was not to obtain hearing thresholds as defined normally (i.e., 50-percent response level). Rather, "threshold" was defined operationally as the average of two minimum response levels determined using an adaptation of the modified Hughson-Westlake audiometric test technique (9). The two responses defining threshold were obtained during presentation of tones in ascending 1-dB increments (i.e., during Stage 3).

Automated Testing for Hearing Thresholds

Details of the threshold-seeking algorithm were fully described in Henry et al. (3). Briefly, initial

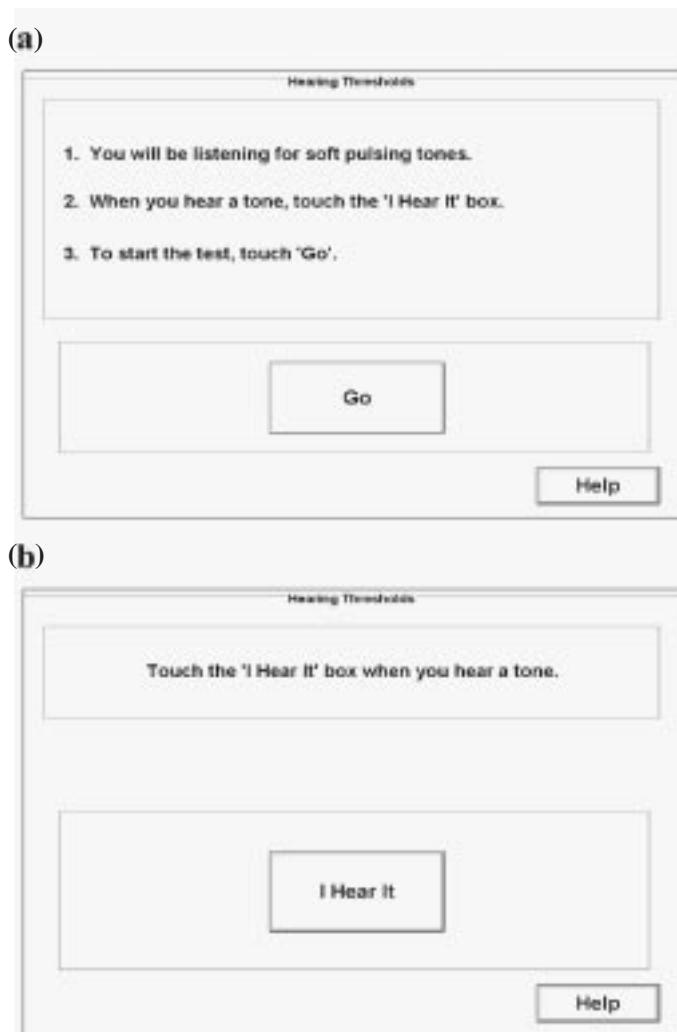


Figure 3.
Screen displays on subjects' notebook computer for hearing thresholds: (a) instructions and (b) response screen.

presentation levels were fixed at 60 dB SPL for each test frequency. Three series of bracketing procedures progressively reduced the step sizes to result in threshold responses with 1-dB resolution. For Series 1, step increments were up 10 dB, down 20 dB, and the first response initiated the Series 2 algorithm. Series 2 and Series 3 used, respectively, increments of up 5 dB, down 10 dB and up 1 dB, down 2 dB. Two responses were required for each of Series 2 and 3, and responses were averaged to obtain the minimum response level for a series.

RESULTS

Conventional Hearing Thresholds

Given an equivalent input voltage, the ER-4B earphones provide higher output and greater frequency response than other insert earphones used for conventional audiometry (see **Figure 2**). Thus, the ER-4B earphones offer advantages for audiometric testing and could be used for this purpose in the future. It was of interest, therefore, to make within-subject comparisons of thresholds obtained with the ER-4B earphones to thresholds obtained from the same subjects using the TDH-50 earphones. Such a comparison would provide preliminary normative threshold data for the ER-4B earphones.

Mean thresholds were compared between the Virtual 320 audiometer and the automated system at test frequencies that were common to both systems (octaves from 500 to 8000 Hz). The threshold measurements with the use of the Virtual 320 were obtained in dB HL. To compare between systems using the same dB metric, we adjusted the dB SPL thresholds obtained with the ER-4B earphones to dB HL using the reference equivalent threshold sound pressure levels (RETSPLs) for insert earphones calibrated in an occluded ear simulator (9). It should be noted that production of the same sound pressure level for both earphones in their respective calibration couplers did not ensure that the earphones produced the same sound pressure at the eardrum.

With this caveat in mind, **Table 1** shows that the threshold means (in dB HL) for the two systems differed by 1.0 to 10.2 dB at the different octave frequencies. To determine if these differences were significant, we calculated t-tests. Since multiple tests were performed on these data, Bonferroni corrections dictated significance levels to interpret the results ($p < 0.01$ to correspond with 0.05 level for a single t-test). The mean thresholds were significantly different at 2, 4, and 8 kHz. All further threshold data are reported in dB SPL with the use of the automated system and ER-4B Canal PhoneTM earphones.

Between-Session Reliability

Within-Group Reliability

Table 2 shows the across-subjects mean thresholds, in dB SPL, separated by test frequency, session, and stage of testing during each session. During Stage 1, the hearing threshold for each of the 16 test frequencies between 0.5 and 16.0 kHz was determined. For Stages 2 and 3, threshold testing was repeated, but only at the

Table 1.

Mean hearing thresholds, in dB HL, obtained with two systems: (1) Virtual 320 audiometer with TDH-50P supra-aural earphones and (2) automated system with ER-4B Canal Phone™ earphones.

Frequency (Hz)	TDH-50P Supra-aural Earphones	ER-4B Canal Phone™ Earphones	p-value*
500	2.0	4.7	0.0247
1,000	4.3	5.3	0.2207
2,000	4.3	7.0	0.0046
4,000	10.8	3.8	<0.0001
8,000	12.0	1.8	<0.0001

* Results of paired t-tests; comparisons at 2,000, 4,000, and 8,000 Hz were significant after corrections for multiple tests using Bonferroni's method.

Table 2.

Means of hearing thresholds, in dB SPL, obtained with automated system. Between Stages 2 and 3 during each session, foam eartips from insert earphones were removed and reinserted.

Freq (Hz)	Session 1			Session 2			p-value*
	Stage 1 (All Freqs)	Stage 2 (Octave Freqs)	Stage 3 (Octave Freqs)	Stage 1 (All Freqs)	Stage 2 (Octave Freqs)	Stage 3 (Octave Freqs)	
500	14.15	13.25	13.05	14.00	12.75	11.85	0.0141
620	11.55	—	—	11.05	—	—	0.3828
800	9.75	—	—	9.60	—	—	0.7858
1,000	10.75	10.50	10.75	10.25	10.60	9.85	0.3531
1,260	12.40	—	—	10.80	—	—	0.0252
1,580	13.00	—	—	12.35	—	—	0.4241
2,000	18.55	18.70	18.40	18.30	18.50	18.15	0.9181
2,520	20.05	—	—	19.65	—	—	0.5219
3,180	19.40	—	—	18.65	—	—	0.4321
4,000	18.80	18.25	18.05	18.40	18.00	16.95	0.2886
5,040	18.35	—	—	16.50	—	—	0.0506
6,340	19.45	—	—	18.40	—	—	0.3514
8,000	17.30	16.90	16.60	16.40	17.20	16.50	0.9116
10,080	35.70	—	—	35.00	—	—	0.5619
12,700	47.50	—	—	46.70	—	—	0.4271
16,000	66.00	66.06	67.00	65.18	61.88	65.00	0.1139

* Results of repeated measures ANOVAs at octave frequencies (0.5, 1.0, 2.0, 4.0, 8.0, 16.0 kHz); results of t-tests at nonoctave frequencies. None of the ANOVAs or t-tests was significant after corrections for multiple tests using Bonferroni's method.

octave frequencies (0.5, 1.0, 2.0, 4.0, 8.0, and 16.0 kHz). There were thus six means for each of the octave frequencies, and repeated measures ANOVAs were calculated on these means at each octave frequency. When there were only two means (i.e., at nonoctave frequencies), t-tests were calculated. The multiple tests required Bonferroni corrections to determine significance levels ($p < 0.008$ to correspond with 0.05 level for a single ANOVA; $p < 0.005$ to correspond with 0.05 level

for a single t-test). None of the ANOVAs or t-tests revealed significant differences.

Within-Subjects Reliability

Table 2 shows good reliability of threshold responses for the subjects as a group, both within and between sessions. To evaluate between-sessions reliability of responses, within subjects, differences were calculated between individual repeated thresholds at each frequency

(Session 2, Stage 1, threshold minus Session 1, Stage 1 threshold). The across-subjects means of these differences are shown in column 2 of **Table 3**. These are the means of the actual differences, and thus reflect the directionality of the responses between sessions.

It is noteworthy that all of these mean differences were negative, indicating a significant trend ($p < 0.05$, Wilcoxon matched-pairs signed ranks test) for Stage 1 mean threshold responses obtained at the second session to be less than those from the first session. The third column in **Table 3** shows how many of the individual differences were positive at each frequency, which averaged 6.56 (out of a possible 20 individual differences), while column 4 shows an average of 9.88 negative differences. There was an average of 3.38 times, per frequency, when the thresholds were identical between Sessions 1 and 2 (column 5). The standard deviations of the between-sessions differences are shown in the next column, where it can be seen that they ranged from 1.94 dB to 5.30 dB, with an average standard deviation across frequencies of 3.38 dB.

Pearson product-moment correlations were also evaluated for each frequency, and the Pearson r 's are shown in **Table 3**. Each of these r -values was ≥ 0.787

(average r across frequencies = 0.897), and all coefficients were significant at $p < 0.0001$. The square of the correlation coefficient (r^2) gives the proportion of the variance in the thresholds of the second session that is explained by the thresholds of the first session. These values ranged from 0.619–0.972, with a mean across frequencies of 0.809. Thus, approximately 81 percent of the variance in the Session 2 thresholds can be explained by the variance in the Session 1 thresholds. Put another way, 81 percent of the variance can be explained by the relationship between the Session 1 and Session 2 repeated thresholds, leaving an unaccountable variance of 19 percent.

The mean differences shown in column 2 of **Table 3** are based on the actual differences in thresholds between Session 1 and Session 2. These means show the directionality of the responses, as described above. It was also of interest to determine the average magnitude of the differences between sessions. To do that, the absolute value of the between-session threshold difference for each subject was calculated before determining the across-subjects means at each frequency. These means of the absolute values of the between-sessions threshold differences are shown in the last column of **Table 3**. The means ranged from 1.45–3.60 dB. For the entire dataset of

Table 3.

Means of individual differences in hearing thresholds, in dB, between Session 1 and Session 2. See body text for full explanation of each column's data.

Freq (Hz)	Mean (dB) of Actual Diffs	No. of Diffs > 0	No. of Diffs < 0	No. of Diffs = 0	Standard Deviation of Diff Scores (dB)	Pearson r^*	r^2	Mean of Absolute Values of Diffs (dB)
500	-0.15	9	6	5	2.93	0.853	0.728	2.15
620	-0.50	7	10	3	2.50	0.865	0.748	2.00
800	-0.15	10	8	2	2.43	0.859	0.738	2.05
1,000	-0.50	5	8	7	2.14	0.906	0.821	1.50
1,260	-1.60	6	13	1	2.95	0.806	0.650	3.00
1,580	-0.65	6	9	5	3.56	0.787	0.619	2.15
2,000	-0.25	6	10	4	1.94	0.971	0.943	1.45
2,520	-0.40	7	7	6	2.74	0.924	0.854	1.80
3,180	-0.75	6	10	4	4.18	0.837	0.701	2.55
4,000	-0.40	6	8	6	2.52	0.950	0.903	1.90
5,040	-1.85	5	15	0	3.96	0.879	0.773	2.85
6,340	-1.05	6	12	2	4.92	0.854	0.730	3.05
8,000	-0.90	5	14	1	3.64	0.949	0.901	3.10
10,080	-0.70	7	9	4	5.30	0.948	0.899	3.60
12,700	-0.80	9	10	1	4.41	0.986	0.972	3.50
16,000	-0.82	5	9	3	4.02	0.983	0.966	2.94
Average	-0.72	6.56	9.88	3.38	3.38	0.897	0.809	2.47

* All correlation coefficients significant at $p < 0.0001$.
(Diffs=differences)

differences in hearing thresholds between Sessions 1 and 2, the average difference, ignoring the direction of the differences, was 2.47 dB.

Confidence Intervals for Difference Scores

The above analyses are based on group comparisons, with the assumption that the individual subjects were reasonably representative of the group. Reporting confidence intervals best shows the range of individual between-sessions differences in hearing thresholds. These intervals are shown in **Table 4** with the numbers and percentages of difference scores falling within each

Table 4.

Confidence intervals (dB) for between-sessions differences in hearing thresholds.

From (\geq)	To (<)	Number of Differences*	Percent of Differences
-1	1	92	29.0
-2	2	166	52.4
-3	3	227	71.6
-4	4	269	84.9
-5	5	290	91.5
-10	10	311	98.1
-15	15	315	99.4
-20	20	317	100.0

* Total number of between-sessions threshold differences=317.

specified interval. Of the 317 between-sessions threshold differences that are represented in **Table 4**, 290 (91.5 percent) were within ± 5 dB, 311 (98.1 percent) were within ± 10 dB, and 315 (99.4 percent) were within ± 15 dB. Threshold differences equaled 15 dB on only two occasions, and never equaled 20 dB.

We also evaluated the confidence intervals at the individual test frequencies. These results are shown in **Table 5**, which is similar to **Table 4** except that the percentages of responses for each dB interval are shown separately for each test frequency. These data indicate that, in general, between-session responses were more reliable at frequencies up to 1.26 kHz, with less reliable responses at the higher test frequencies.

Within-Session Reliability

During each session, three thresholds were obtained at each of the octave frequencies. This protocol enabled analyses of (1) within-subject, within-session response reliability, and (2) the potential effect of removing and reinserting the foam eartip of the insert earphone before repeating the threshold measurement. **Table 6** shows the means of the threshold differences between each possible pair of tests (Stages 1, 2, and 3 as also shown in **Table 2**) during each session.

Stage 1 involved the baseline measurements (hearing thresholds at all 16 frequencies). For Stage 2,

Table 5.

Confidence intervals for between-sessions differences in hearing thresholds. Each value represents the percentage of responses which occurred for each interval indicated.

Interval (dB) in Which Between- Sessions Threshold Differences Occurred	From (\geq)	To (<)	Frequency (kHz)															
			0.5	0.62	0.8	1.0	1.26	1.58	2.0	2.52	3.18	4.0	5.04	6.34	8.0	10.08	12.7	16.0
-1	1	30	25	15	45	10	30	55	35	40	30	21	42	25	25	20	30	
-2	2	45	65	45	70	20	60	70	60	60	60	63	59	45	40	30	53	
-3	3	70	75	80	85	50	90	85	90	80	75	79	68	50	60	50	65	
4	4	85	90	95	95	85	90	95	90	85	95	95	95	70	65	65	71	
-5	5	90	100	100	100	95	100	95	100	95	85	95	100	95	85	80	75	76
-10	10	100	—	—	—	—	95	—	100	95	100	—	100	100	90	100	100	
-15	15	—	—	—	—	—	100	—	—	100	—	—	—	—	100	—	—	

Table 6.

Means of actual values of individual differences in hearing thresholds. All means shown are for various combinations of within-session differences.

Freq (Hz)	Session 1			Session 2		
	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2
500	-0.90	-1.10	-0.20	-1.30	-2.15	-0.85
1,000	-0.25	0.00	-0.25	0.35	-0.40	-0.75
2,000	-0.15	-0.10	-0.25	0.20	-0.15	-0.35
4,000	-0.55	-0.75	-0.20	-0.80	-1.45	-1.05
8,000	-0.40	-0.70	-0.30	0.80	0.10	-0.70
16,000	-0.06	1.00	0.94	-3.29	-0.18	3.12
Average	-0.32	-0.28	-0.43	-0.67	-0.71	-0.97

repeated thresholds were obtained at octave frequencies only, with the eartip left in place. Stage 3 involved repeated thresholds at octave frequencies only, with the eartip removed and reinserted.

Each difference score was calculated by subtracting an earlier response from a later response. The mean differences shown in **Table 3** revealed a trend of Session 2 thresholds being lower than Session 1 thresholds, significantly more often than the reverse case. The within-session differences in **Table 6** reflect the same trend (Wilcoxon, $p<0.05$). Of the 36 means shown in **Table 6**, 9 are positive and 26 are negative (with 1 mean being 0). The mean differences are again very small, with the average difference across the six conditions being less than 1 dB.

Table 6 shows the means of the actual differences in thresholds between the various within-sessions conditions, and because differences could be positive or negative, **Table 6** reflects the directionality of the paired responses. To reveal the magnitude of the individual differences in thresholds, we calculated the absolute value of each difference and determined the means of the absolute values (**Table 7**). The averages of these mean differences ranged from 1.28 to 2.93 dB.

It was a primary objective of the within-session study design to establish whether there was any effect on hearing thresholds when the foam eartip was removed and reinserted. To evaluate for that potential effect, t-tests were calculated, at each frequency, between the following means: "Stage 2 minus Stage 1" versus "Stage 3 minus Stage 1." This was done for both Session 1 and Session 2 pairs of means. None of these t-tests was significant (all p 's <0.05). For completeness, t-tests were also calculated to examine for potential

differences in thresholds between Stage 1 *versus* Stage 2, and Stage 1 *versus* Stage 3. Again, none of the t-tests was significant.

Confidence Intervals for Difference Scores

The range of individual between-sessions differences in hearing thresholds is shown by reporting confidence intervals, seen in **Table 4**. Similarly, the range of within-sessions differences is shown in **Table 8**. There were, however, multiple combinations of differences to be reported for the within-sessions repeated thresholds. For each session, three thresholds were obtained at each of the octave frequencies, which allowed three difference scores to be calculated from each session: (1) Stage 2 threshold minus Stage 1 threshold, (2) Stage 3 threshold minus Stage 1 threshold, and (3) Stage 3 threshold minus Stage 2 threshold.

Table 8 shows the percentages of difference scores for the various combinations within each specified confidence interval. For Session 1, the Stage 2 minus Stage 1 column shows the percentages of differences when testing was repeated without removing the eartips. The eartips were removed and replaced between Stages 2 and 3; thus, the next two columns in **Table 8** (Stage 3 minus Stage 1, and Stage 3 minus Stage 2) reflect eartip replacement. In general, the percentages of differences were slightly higher for the no-replacement condition than for the replacement condition for each session.

Table 8 also shows that within-session reliability was somewhat better during Session 1 than during Session 2. During Session 1, 97.4 percent of the differences occurred within ± 5 dB for the no-replacement condition, and 94.9 percent of the differences occurred within ± 5 dB for each of the replacement conditions.

Table 7.

Means of absolute values of individual differences in hearing thresholds. All means shown are for the various combinations of within-session differences.

Freq (Hz)	Session 1			Session 2		
	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2
	1.70	2.50	1.30	2.30	3.35	2.85
500	1.15	1.30	1.45	1.65	1.90	1.75
2000	1.05	1.30	0.55	1.70	2.25	1.85
4000	1.25	1.45	1.00	2.60	3.05	2.55
8000	1.50	2.70	3.10	2.00	2.60	3.00
16,000	1.00	1.47	1.77	4.24	3.00	5.59
Average	1.28	1.79	1.53	2.42	2.69	2.93

Table 8.

Confidence intervals for within-sessions differences in hearing thresholds.

Interval (dB) in which Within-Sessions Threshold Differences Occurred		Percent of differences*					
		Session 1			Session 2		
From (\geq)	To (<)	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2	Stage 2 minus Stage 1	Stage 3 minus Stage 1	Stage 3 minus Stage 2
-1	1	51.3	40.2	49.6	33.3	29.1	30.8
-2	2	76.1	59.8	70.9	59.8	45.3	53.0
-3	3	88.9	77.8	82.9	80.3	69.2	69.2
-4	4	95.7	93.2	91.4	88.9	88.0	81.2
-5	5	97.4	94.9	94.9	94.0	92.3	88.9
-10	10	100.0	100.0	100.0	98.3	99.1	98.3
-15	15	—	—	—	99.1	99.1	99.1
-20	20	—	—	—	100.0	100.0	100.0

* Total number of within-sessions threshold differences=117.

The Session 2 respective percentages were 94.0 percent, 92.3 percent and 88.9 percent. For Session 1, 100 percent of the differences were within ± 10 dB, while a few differences were between 10 and 20 dB for Session 2.

DISCUSSION

Our ultimate goal is to develop tinnitus assessment methodology suitable for routine clinical application. Attaining this goal will require the ability to conduct all testing rapidly, while maintaining a high level of test-retest response reliability. The automated method was

developed specifically for quantification of acoustical parameters of tinnitus, and an essential component of such testing is the measurement of hearing thresholds. Although test-retest reliability of hearing thresholds is well documented, the unique features of the automated system required a system-specific analysis of threshold reliability. The purpose of the present study was, therefore, to demonstrate reliability of auditory thresholds with the use of our computer-automated method.

Test-Retest Reliability of Pure Tone Thresholds

Pure tone audiometry involves routine procedures that have been thoroughly documented for response reliability by studies dating back to the 1930s (10–13). Since

that time, many studies have shown good reliability of repeated threshold measurements in the conventional-frequency ($\leq 8\text{ kHz}$) range (13–20). For high-frequency ($> 8\text{ kHz}$) pure tone testing, standing waves have often been cited as a concern (19,21–24). At frequencies $> 8\text{ kHz}$, the quarter wavelength is short enough to produce nodes and antinodes in the ear canal, resulting in varied sound pressure across the surface of the tympanic membrane (21). Thus, changes in the position of a transducer, unavoidable with repeated testing, would be expected to have greater effects on higher frequency tones in the ear canal than on lower frequency tones. Therefore, investigators have compared threshold reliability between conventional- and high-frequency ranges and have reported that reliability is equally good in both ranges (7,14,24–29).

1-dB Threshold Resolution

For most audiological applications, hearing thresholds are obtained with 5-dB resolution; therefore, use of 5-dB step sizes was adopted for the majority of reliability studies cited above. In the absence of organic or nonorganic change between tests, the standard error of the estimated threshold (a measure of the intrasubject consistency) is considered to be approximately 5 dB for both air- and bone-conduction measurements (30,31). Clinical audiologists thus operate under the assumption that repeated thresholds within $\pm 5\text{ dB}$ reflect normal tolerance for clinical error (13,32). Tinnitus loudness-matching, however, requires step changes of 1 dB to obtain precise loudness matches. Because the loudness matches are referenced to hearing thresholds at each test frequency, the thresholds must also be obtained with 1-dB precision. Thus, for the present study it was necessary to obtain all thresholds to the nearest decibel.

Means of the actual differences in hearing thresholds were shown, both between sessions (**Table 3**) and within sessions (**Table 6**). These analyses reveal whether the thresholds trended higher or lower upon repeated testing (discussed in the next paragraph). The absolute values of these differences were also calculated, the means of which reveal the magnitude of the differences across subjects. These means generally ranged between 1 and 3 dB. For audiologists, the expected $\pm 5\text{ dB}$ test-retest variability of hearing thresholds is predicated upon testing in 5-dB steps. Hearing thresholds are not normally obtained with 1-dB precision, thus there are no clinical norms for the variability of these measurements. However, results of this study indicate that the performance of this automated technique for obtaining reliable hearing thresholds is well

within a clinically acceptable range. The mean differences between responses across all subjects and conditions were 1–3 dB, and 91.5 percent and 99.4 percent of the between-sessions threshold differences were within, respectively, $\pm 5\text{ dB}$ and $\pm 10\text{ dB}$. Our finding that 91.5 percent of differences are within $\pm 5\text{ dB}$ indicates an improvement in test-retest reliability compared to previously reported data (33–35). Our data, therefore, suggest that greater precision of clinical thresholds may be achieved using a 1-dB step procedure as compared to the traditional use of 5-dB steps.

Learning/Practice Effect

There was a significant trend for the threshold measurements to improve with repeated testing. All the between-sessions mean differences were negative (**Table 3**). These mean differences, however, were small—all were less than 2 dB, and the average of the means across the 16 test frequencies was only -0.72 dB . Improvements in mean thresholds were also observed within sessions (**Table 6**). These differences were again very small and averaged less than -1 dB . The systematic improvement in absolute auditory thresholds after repeated measurements has been reported previously (36). Zwislocki et al. studied this effect under various experimental conditions and concluded that the threshold of audibility improves with practice. The improvement was attributed to effects of practice and motivation, and thresholds were noted to improve during several experimental sessions. The effect was also postulated to be due to the discrimination of tones against a background of physiological noise, and, with practice, this discrimination ability becomes more sensitive. Improvements in thresholds with repeated testing have been reported by additional investigations (10,18,19,37–39). Other studies, however, have shown no improvement in thresholds with repeated testing (15,40–43).

Although the practice/learning effect for thresholds is equivocal in the literature, the present data suggest that there is such an improvement in normal-hearing individuals. Our results agree with those of the one study that tested systematically for this effect (36). Although not stated in the study by Zwislocki et al., it is likely that their listeners also had normal hearing. The data from the present study, along with those from the Zwislocki et al. study, together argue strongly that this effect occurs when hearing sensitivity is normal. There is yet the need to determine if this effect also occurs for subjects with cochlear hearing loss.

Automatic Audiometry

The present study was a component of a larger project that is designed to develop automated methodology for obtaining tinnitus-matching measurements. Thus, development of computer automation to obtain hearing thresholds was not an end in itself. However, because of the history of attempts to develop automatic audiometry as an alternative to traditional manual audiometry, these data contribute to this area and some relevant comments are warranted.

The defining characteristics of automatic methods for pure tone audiometry are that the listener maintains control over the level of stimulus presentation and that at least some of the procedures are automated (44). The first automatic, self-recording audiometer was described by von Bekesy in 1947 (45). That audiometer produced a sweep-frequency tone, and in 1956, a fixed-frequency version appeared, inviting direct comparison with manual audiometers. A number of studies were conducted subsequently to compare hearing thresholds, in the same individuals, between manual and self-recording audiometers. Most generally, these studies showed that self-recording audiometry resulted in slightly more sensitive thresholds than manual audiometry (46). Using 1-dB step sizes, most studies have shown an improved sensitivity of 1–2 dB with automatic audiometry, while an average difference of about 3 dB was reported by Robinson and Whittle (39).

For the present study, mean thresholds were compared between the conventional audiometer and the automated system at octave frequencies between 500–8000 Hz (**Table 1**). The use of different headphones (TDH-50P supra-aural *versus* ER-4B insert) required the caveat that, although dB HL was matched between earphones (8), the pressure produced at the eardrum was not necessarily equal because of the different acoustic characteristics of both the earphones and the couplers.

The advent of microcomputers provided another method for conducting automatic audiometry, and the potential advantages of computerized audiometry were recognized as long ago as 1971 (47). At that time, it was considered a “foregone conclusion” that computer-driven audiometry would supplant manual audiometry. Such a transition has obviously never occurred, but automatic audiometry has found utility for certain applications, especially industrial audiometric testing. The use of automated testing in the audiology clinic would require programming of computer algorithms to perform testing at the level of a skilled audiologist. This may be feasible for

unmasked pure tone air conduction audiometry, but sophisticated masking and bone conduction techniques may never be adaptable to automation. Nevertheless, just as automated testing is used for industrial monitoring purposes, it could also have application for ototoxicity monitoring.

The main problem with ototoxicity monitoring is the difficulty obtaining audiometric data from patients at repeated intervals. Whether these patients are in the hospital or in their homes, scheduling their repeated audiometric exams has proven to be cumbersome, and impossible in many cases. Consequently, many patients who are included in an ototoxicity-monitoring program do not receive the level of service that is available to prevent significant ototoxic effects. These kinds of difficulties make ototoxicity-monitoring programs difficult to operate effectively and may be the reason such programs are scarce, despite published standards for early detection of ototoxicity (48).

In the present study, the differences in hearing thresholds between-sessions did not produce a single value that would have met the ASHA (1994) criteria for ototoxicity (48). Thus, this technique has the potential to reduce false positive responses that are associated with ototoxicity monitoring. For this to be investigated further, a threshold reliability evaluation of the automated technique should be conducted in a population of patients not receiving ototoxic drugs.

Etymotic Research ER-4B Canal PhoneTM Earphones

When faced with the decision of selecting earphones for use with the automated testing system, we were primarily concerned with using earphones that were capable of reproducing tones at high levels throughout the frequency range of 0.5–16 kHz for tinnitus matching. Testing at high frequencies (>8 kHz) requires high output capability due to the gradual reduction in human auditory sensitivity in this frequency range. After evaluating the commercial possibilities, we found that the Etymotic ER-4B Canal PhoneTM earphones appeared to provide the best performance characteristics for our application. Considering the availability of a variety of circum-aural and insert earphones that are intended specifically for audiometric application, selection of an in-the-ear transducer that was designed for listening to binaural recordings was unexpected. The present study has demonstrated that the ER-4B has practical application for use as a single earphone transducer to evaluate an extended range of auditory sensitivity.

In addition to its utility for full-frequency testing capability, the ER-4B shares in the advantages that are offered by any type of insert earphone. Some of the most obvious advantages include the reduction of ambient noise during testing (49,50) and the significant increase in interaural attenuation (51,52). Lilly and Purdy (53) have described other advantages of insert earphones relative to supra-aural/circumaural earphones.

The present study has further documented that test-retest reliability of threshold sensitivity using the ER-4B is at least as good as that shown for other insert earphones and for traditional audiometric earphones. Studies have compared test-retest reliability of hearing thresholds using the Etymotic ER-3A Tubephone™ *versus* standard supra-aural earphones, including the TDH-50 (34,54), TDH-39 (55) and TDH-49 (56). These studies all showed that reliability of thresholds for frequencies up to 8 kHz was at least as good for the ER-3A as for the standard audiometric earphones.

Other studies have shown good threshold reliability with insert earphones for frequencies above 8 kHz. Tang and Letowski (57) obtained repeated thresholds at 10–16 kHz using the Sennheiser HD-250 circumaural earphone and Etymotic ER-1 Tubephone™. Their results revealed significantly smaller variability with the insert earphones. Valente, Valente, and Goebel (58) compared test-retest reliability of high-frequency thresholds up to 18 kHz using the Koss HV/1A+ *versus* the Etymotic ER-2 Tubephone™. Intrasubject response variability was comparable between the two earphones.

The present study adds to this literature by showing that the Etymotic ER-4B earphones can provide response reliability that is comparable to all earphones that have been demonstrated to be reliable for testing auditory sensitivity.

Reinsertion of Foam Eartips

An additional concern addressed by this study was whether removal and replacement of the ER-4B foam eartips might have an effect on threshold reliability. This is a particularly important question when testing at higher frequencies where standing waves might be affected by earphone placement, with the potential to significantly affect sound pressure level at the eardrum.

Hickling (19) found that when TDH-39 supra-aural earphones were removed and replaced between tests, the reliability of 6 and 8 kHz thresholds was significantly poorer than when earphones were left in place during repeated testing. Earphone replacement did not have an

effect at 1 and 2 kHz. The effect at 6 and 8 kHz was attributed to standing wave formation at these frequencies. Erlandsson et al. (40) found greater variability of repeated auditory thresholds when a circumaural earphone was repeatedly replaced versus when thresholds were retested with the earphone fixed in position for each repetition. The authors suggested that a circumaural earphone deforms the pinna, which can affect the transmission of sound pressure to the ear canal. Gauz, Robinson, and Peters (59) found no effect on threshold measurements when a circumaural earphone was replaced.

Stelmachowicz et al. (60) compared reliability of high-frequency (8–20 kHz) thresholds using two systems. One was a prototype high-frequency audiometer, originally described by Stevens et al. (21), which used a 60-cm plastic tube to couple the high-frequency transducer to the ear of the subject. The other system used Koss HV/X supra-aural earphones. Repeated thresholds were obtained without moving the earphones. The earphones were then removed and replaced, and a third set of thresholds was obtained. For both systems, replacement of earphones resulted in a slightly higher standard error of measurement (SEM) than when earphones were left in place, with the supra-aural earphones having the best response reliability with replacement.

Larson et al. (34) conducted test-retest measurements using the ER-3A Tubephone™. A component of that study was to conduct two retests, one with the ER-3A eartips left in place and the second after removal and replacement of the eartips. They found no significant effect on test reliability when eartips were replaced. Larson's study is the only one we know of that evaluated threshold reliability between the two conditions of eartips fixed *versus* replaced. The present study confirms the results of Larson for reliability of thresholds using insert-style earphones.

CONCLUSION

These data validate use of our automated technique for obtaining reliable hearing thresholds. Results of this study may have further generalized application, including (1) confirmation that the ER-4B Canal Phone™ earphones can be used for clinical audiology, (2) the ER-4B eartips can be removed and reinserted without appreciably affecting the measurements, and (3) 1-dB step sizes can be used for obtaining precise tinnitus matching measurements. In addition to using the tech-

nique for tinnitus matching, there may be further uses for applications requiring serial monitoring of auditory thresholds, such as hearing conservation and ototoxicity monitoring.

REFERENCES

1. Henry JA, Meikle MB. Psychoacoustical measures of tinnitus. *J Amer Acad Audiol* 2000;11:138–55.
2. Henry JA, Meikle MB, Gilbert A. Audiometric correlates of tinnitus pitch: insights from the tinnitus data registry. In: Hazell J, editor. 6th International Tinnitus Seminar 1999; Cambridge, England. London: Tinnitus and Hyperacusis Centre; 1999. p. 51–7.
3. Henry JA, Flick CL, Gilbert AM, Ellingson RM, Fausti SA. Reliability of tinnitus loudness matches under procedural variation. *J Amer Acad Audiol* 1999;10:502–20.
4. Vernon JA, Meikle MB. Tinnitus masking: unresolved problems. In: Evered D, Lawrenson G, editors. CIBA Foundation Symposium 85. Tinnitus. London: Pitman Books, Ltd.; 1981. p. 239–56.
5. Vernon JA, Meikle MB. Measurement of tinnitus: an update. In: Kitahara M, editor. Tinnitus. Pathophysiology and management. Tokyo: Igaku-Shoin; 1988. p. 36–52.
6. Vernon J, Fenwick J. Identification of tinnitus: a plea for standardization. *J Laryngol Otol* 1984;45 Suppl 9:45–53.
7. Fausti SA, Frey RH, Henry JA, Knutson JL, Olson DJ. Reliability and validity of high-frequency (8–20 kHz) thresholds obtained on a computer-based audiometer as compared to a documented laboratory system. *J Amer Acad Audiol* 1990;1:162–70.
8. American National Standards Institute (ANSI). American National Standard: specification for audiometers. American National Standards Institute; 1996. ANSI S3.6-1996.
9. Carhart R, Jerger JF. Preferred method for clinical determination of pure-tone thresholds. *J Speech Hear Disord* 1959; 24:330–45.
10. Corso JF, Cohen A. Methodological aspects of auditory threshold measurements. *J Exp Psychol* 1958;55:8–12.
11. Ciocco A. Audiometric studies on school children. III. Variations in the auditory acuity of 543 school children reexamined after an average interval of three years. *Ann Otol Rhinol Laryngol* 1937;46:55.
12. Lifschitz S. Fluctuation of the hearing threshold. *J Acoust Soc Am* 1939;11:118–21.
13. Witting EG, Hughson W. Inherent accuracy at a series of repeated clinical audiograms. *Laryngoscope* 1940;50:259–69.
14. Fletcher JL. Reliability of high-frequency thresholds. *J Audit Res* 1965;5:133–7.
15. Brown REC. Experimental studies on the reliability of audiometry. *J Laryngol Otol* 1948;42:487–524.
16. High WS, Glorig A, Nixon J. Estimating the reliability of auditory threshold measurements. *J Audit Res* 1961;1:247–62.
17. Atherley GS, Dingwall-Fordyce I. The reliability of repeated auditory threshold determination. *Br J Ind Med* 1963;20:231–5.
18. Hickling S. The validity and reliability of pure tone clinical audiometry. *New Zeal Med J* 1964;63:379–82.
19. Hickling S. Studies on the reliability of auditory threshold values. *J Audit Res* 1966;6:39–46.
20. Tyler RS, Wood EJ. A comparison of manual methods for measuring hearing levels. *Audiol* 1980;19:316–29.
21. Stevens KN, Berkovitz R, Kidd G, Jr., Green DM. Calibration of ear canals for audiometry at high frequencies. *J Acoust Soc Am* 1987;81:470–84.
22. Stinson MR, Shaw EAG. Wave effects and pressure distribution in the ear canal near the tympanic membrane. *J Acoust Soc Am* 1982;71 Suppl 1:S88.
23. Stinson JR, Lawton BW. Specification of the geometry of the human ear canal for the prediction of sound-pressure level distribution. *J Acoust Soc Am* 1989;85:2492–502.
24. Zhou B, Green DM. Reliability of pure-tone thresholds at high frequencies. *J Acoust Soc Am* 1995;98:828–36.
25. Dreschler WA, van der Hulst RJAM, Tange RA, Urbanus NAM. The role of high frequency audiometry in early detection of ototoxicity. *Audiol* 1985;24:387–95.
26. Frank R, Dreisbach LE. Repeatability of high-frequency thresholds. *Ear Hear* 1991;12:294–5.
27. Laukli E, Mair IWS. High-frequency audiometry: normative studies and preliminary experiences. *Scand Audiol* 1985;14:151–8.
28. Fausti SA, Frey RH, Erickson DA, Rappaport BZ. 2AFC versus standard clinical measurement of high frequency auditory sensitivity (8–20 kc/s). *J Audit Res* 1979;19:151–7.
29. Matthews LJ, Lee F-S, Mills JH, Dubno JR. Extended high-frequency thresholds in older adults. *J Speech Lang Hear Res* 1997;40:208–14.
30. Jerger J. Hearing tests in otologic diagnosis. *ASHA* 1962; 4:139–43.
31. Newby HA. Audiology. New York: Appleton-Century-Crofts; 1972.
32. Currier WD. Office noises and their effect on audiometry. *Arch Otolaryngol* 1943;38:49–59.
33. Studebaker GA. Intertest variability and the air-bone gap. *J Speech Hear Disord* 1967;32:82–6.
34. Larson VD, Cooper WA, Talbott RE, Schwartz DM, Ahlstrom C, DeChicchis AR. Reference threshold sound-pressure levels for the TDH-50 and ER-3A earphones. *J Acoust Soc Am* 1988; 84:46–51.
35. Frank T. High-frequency hearing thresholds in young adults using a commercially available audiometer. *Ear Hear* 1990; 11:450–4.
36. Zwislocki J, Maire F, Feldman AS, Rubin H. On the effect of practice and motivation on the threshold of audibility. *J Acoust Soc Am* 1958;30:254–62.
37. High WS, Glorig A. The reliability of industrial audiometry. *J Audit Res* 1962;2:56–65.
38. Burns W, Hinchcliffe R. Comparison of the auditory threshold as measured by individual pure tone and by Bekesy audiometry. *J Acoust Soc Am* 1957;29:1274.
39. Robinson DW, Whittle LS. A comparison of self-recording and manual audiometry: some systematic effects shown by unpractised subjects. *J Sound Vibr* 1973;26:41–62.
40. Erlandsson B, Hakanson H, Ivarsson A, Nilsson P. The reliability of Bekesy sweep audiometry recording and effects of the earphone position. *Acta Oto-Laryngol (Stockh)* 1980;366:99–112.
41. Harris JD, Myers CK. Experiments on the fluctuation of auditory acuity. *J Gen Psychol* 1954;50:87–109.

42. Herman G. Variability of the absolute auditory threshold—A psychophysical study. *J Acoust Soc Am* 1953;25:822.
43. Munson WA, Wiener FM. Sound measurements for psychophysical tests. *J Acoust Soc Am* 1950;22:382–6.
44. Sieminski L. Automatic audiometry. *Otolaryngol Clin North Am* 1978;11:701–8.
45. Bekesy GV. A new audiometer. *Acta Oto-Laryngol (Stockh)* 1947;35:411–22.
46. Lutman ME, Cane MA, Smith PA. Comparison of manual and computer-controlled self-recorded audiometric methods for serial monitoring of hearing. *Br J Audiol* 1989;23:305–15.
47. Wood TJ, Wittich WW, Mahaffey RB. An application of digital computer logic to pure-tone and speech audiometric procedures. *J Acoust Soc Am* 1971;49:131(A).
48. American Speech-Language-Hearing Association. Guidelines for the audiology management of individuals receiving cochleotoxic drug therapy. *ASHA* 1994;36(March,Suppl.12):11–9.
49. Clemis JD, Ballad WJ, Killion MC. Clinical use of an insert earphone. *Ann Otol Rhinol Laryngol* 1986;95:520–4.
50. Berger EH, Killion MC. Comparison of the noise attenuation of three audiometric earphones, with additional data on masking near threshold. *J Acoust Soc Am* 1989;86:1392–403.
51. Killion MC, Wilber LA, Gudmundsen GI. Insert earphones for more interaural attenuation. *Hear Instr* 1985;36:34–6.
52. Sklare DA, Denenberg LJ. Interaural attenuation for Tubephone insert earphones. *Ear Hear* 1987;8:298–300.
53. Lilly DJ, Purdy JK. On the routine use of Tubephone insert earphones. *Amer J Audiol* 1993;2:17–20.
54. Stuart A, Stenstrom R, Tompkins C, Vandenhoff S. Test-retest variability in audiometric threshold with supraaural and insert earphones among children and adults. *Audiol* 1991;30:82–90.
55. Borton TE, Nolen BL, Luks SB, Meline NC. Clinical applicability of insert earphones for audiometry. *Audiol* 1989;28:61–70.
56. Lindgren F. A comparison of the variability in thresholds measured with insert and conventional supra-aural earphones. *Scand Audiol* 1990;19:19–23.
57. Tang H, Letowski T. High-frequency threshold measurements using insert earphones. *Ear Hear* 1992;13:378–9.
58. Valente M, Valente M, Goebel J. High-frequency thresholds: circumaural earphone *versus* insert earphone. *J Amer Acad Audiol* 1992;3:410–8.
59. Gauz MT, Robinson DO, Peters GM. High-frequency Békésy audiometry: III. Reliability and validity revisited. *J Audit Res* 1981;21:167–80.
60. Stelmachowicz PG, Beauchaine KA, Kalberer A, Kelly WJ, Jesteadt W. High-frequency audiometry: test reliability and procedural considerations. *J Acoust Soc Am* 1989;85:879–87.

Submitted for publication October 19, 2000. Accepted in revised form January 19, 2001.

