

## Computer-automated tinnitus assessment using patient control of stimulus parameters

---

**James A. Henry, PhD; Betsy Rheinsburg, MS; Roger M. Ellingson, MSCSE**

*Department of Veterans Affairs (VA) Rehabilitation Research and Development (RR&D) National Center for Rehabilitative Auditory Research, VA Medical Center, Portland, OR; Department of Otolaryngology, Oregon Health and Science University, Portland, OR*

**Abstract**—A need exists for a standardized tinnitus evaluation tool to measure “acoustic” parameters of tinnitus. An automated tinnitus evaluation system has been developed in this laboratory, consisting of a patient-controlled touch-screen computer monitor, main computer, and Programmable Auditory Laboratory 3000 (PAL 3000) (a custom-built signal conditioning module). The system obtains and records data from tinnitus patients, including hearing thresholds and the pitch and loudness of their tinnitus. New features have been incorporated into the system to make it more user-friendly and practical for clinical application. Using the system, we evaluated 40 individuals with tinnitus to assess within- and between-session reliability of responses. Response reliability has been documented with previous versions of the system. Incorporating the new features has reduced testing time to an average of less than 19 minutes and has resulted in comparable loudness-match reliability and improved pitch-match reliability compared with the previous results. These data support the technique as suitable for clinical application, indicating the need to develop instrumentation for this purpose.

**Key words:** audiology, automated testing, hearing disorders, hearing thresholds, loudness perception, pitch perception, reliability of results, tinnitus.

### INTRODUCTION

Clinical measures of tinnitus loudness and pitch can provide data important for patient assessment and for counseling [1–5]. Traditional approaches to clinical tinni-

tus measurement require patients to perform the subjective task of balancing a tone generated by a clinical audiometer to either the loudness or the pitch of the perceived tinnitus. Using feedback from the patient, the clinician adjusts the tone until the patient reports a “match.” This general methodology quantifies tinnitus frequency and intensity, but no specific standards of testing exist.

Over 20 years ago, the establishment of standardized tinnitus evaluation was a focus of the Ciba Foundation in London [6] and the National Academy of Sciences [7]. These efforts resulted in recommendations for a specific battery of clinical tests for tinnitus: loudness-matching, pitch-matching, tinnitus maskability, and residual inhibition. Vernon and Meikle provided procedural details for

---

**Abbreviations:** ANOVA = analysis of variance, FCDS = forced-choice double-staircase, PAL 3000 = Programmable Auditory Laboratory 3000, SD = standard deviation, SL = sensation level, SPL = sound pressure level, VA = Department of Veterans Affairs, VAMC = VA Medical Center.

**This material was based on work supported by the Department of Veterans Affairs Rehabilitation Research and Development (RR&D) Service (grant C2300R).**

All correspondence should be addressed to James A. Henry, PhD; VA Medical Center (NCRAR), PO Box 1034, Portland, OR 97207; 503-220-8262, ext. 57466; fax: 503-402-2955; email: james.henry@med.va.gov.

DOI: 10.1682/JRRD.2003.10.0158

the tests [8], which established many basic parameters of tinnitus testing. These procedures have not, however, evolved to become the accepted clinical norm for tinnitus assessment, largely because of their requirement for specialized test equipment. Such equipment was developed and marketed at the time, but has since been discontinued by the manufacturers. Today, most audiologists who evaluate tinnitus patients use their clinical audiometers in some manner to perform some or all of these basic tinnitus tests.

Without procedures that are uniform and documented for response reliability, measurements indicating changes in the tinnitus percept do not possess clinical value. In addition, comparing tinnitus-matching data across clinics is not possible. The large number of individuals impacted by tinnitus underscores the need for standardized tinnitus assessment methods [9–10]. Audiological testing for hearing loss has been standardized for decades, and tinnitus testing needs to achieve comparable standardization.

One means to standardize tinnitus assessment is through computer automation. A computer-automated clinical procedure for tinnitus evaluation would be capable of consistently evaluating all parameters of tinnitus that are relevant to rehabilitation. A system to conduct computer-automated clinical procedures for tinnitus evaluation has been under development in our laboratory since 1995 [11]. Numerous studies have documented various aspects of the system's performance [3,12–16]. The system continues to undergo refinements, and although testing with the system has proven reliable for tinnitus loudness-matching, pitch-matching has continued to show considerable test-retest variability. In addition, overall testing time was too lengthy to be acceptable for clinical application. These two issues were addressed by the present study.

The previous iteration of the system was replaced with a redesigned system. The instruction and response screens were completely recreated, and new programming and hardware developed. With these changes, we intended to (1) reduce testing time by enabling more rapid responses by patients and (2) provide a testing platform that would result in more reliable pitch matches. The most notable difference is that a "knob" patient-interface device was created that allows patients increased control over the testing protocol. The initial phase of development and testing of the new automated system is described herein.

## METHODS

### Research Participants

We used the number of participants and the two testing trials to provide a sufficient number of observations to assess reliability of the different responses. Based on results of multiple studies conducted in our laboratory, we have established normal variability of responses for the different tests that are performed by the tinnitus test system. We used these values in a series of power analyses to determine that a power of at least 0.80 would be achieved with 40 participants to assess the reliability of hearing thresholds, loudness matches, and pitch matches.

Following pilot testing to ensure proper performance of the system, we recruited 40 individuals who reported constant tinnitus through a local newspaper advertisement and flyers posted at the Portland Department of Veterans Affairs (VA) Medical Center (VAMC), Oregon. Their mean age was 59 years (standard deviation [SD] = 12; range = 27 to 79). These individuals included 33 males (of which 31 were veterans) and 7 females (of which 2 were veterans). Five of the participants had hearing within normal limits (all thresholds within 25 dB HL), and the remainder had some degree of reduced hearing sensitivity.

Each participant completed a written questionnaire to describe his or her tinnitus. The participants reported the duration of their tinnitus (>20 yr:  $n = 19$ ; 11–20 yr:  $n = 7$ ; 6–10 yr:  $n = 3$ ; 3–5 yr:  $n = 1$ ; 1–2 yr:  $n = 8$ ; <1 yr:  $n = 2$ ). Their tinnitus was described as either binaural ( $n = 36$ ) or unilateral ( $n = 4$ ) and either tonal ( $n = 25$ ) or nontonal ( $n = 15$ ). Participants were asked the question, "How much of a problem is your tinnitus?" and responded "not a problem" ( $n = 2$ ), "small" ( $n = 7$ ), "moderate" ( $n = 18$ ), "big" ( $n = 8$ ), and "very big" ( $n = 8$ ) (one participant did not respond). The "number of tinnitus sounds" were also reported: 1 ( $n = 22$ ), 2 ( $n = 4$ ),  $\geq 3$  ( $n = 10$ ), and "unsure" ( $n = 4$ ).

Assessment of between-session response reliability required that participants be scheduled for repeat evaluations on separate days. The second appointment was scheduled as soon as possible after the first appointment. Most participants completed two evaluation sessions within 2 weeks (<1 wk:  $n = 28$ ; 1–2 wk:  $n = 8$ ; 2–4 wk:  $n = 3$ ; 12 wk:  $n = 1$ ). All participants confirmed that their tinnitus was essentially unchanged from the first to the second appointment. Further, an examination of the data confirmed that those participants who could not return within 1 week did not have increased variability of responses or any other anomalous results.

The Portland VAMC Institutional Review Board Committee approved all use of human participants for this research. Each participant signed an approved informed consent form prior to study enrollment. They received remuneration of \$20 for each test session.

### Equipment Development

Testing algorithms for determining hearing threshold, loudness match, and pitch match were completely redesigned. The previous system used a step response bracketing procedure, altered appropriately for the given task. The system would adjust the loudness or frequency of the stimulus, and then the patient's response would determine the next stimulus presented. The present system was designed to enable the patient to reduce test time by directly controlling the relevant stimulus parameters during testing. Patients, therefore, now have the ability to self-adjust stimulus parameters rather than being guided through the program at the computer's pace, which requires additional time and more test steps.

#### *PowerPoint 2000 Software Interface*

The previous version of the automated system used custom Microsoft Windows dialog and button objects as the interface [3,12,15–17]. Progression through the program was based on the patient's readiness to proceed. After reading the displayed instructions (on the computer monitor), the patient pressed the "buttons" on the touch screen to either advance through the program or to stop testing and contact the audiologist.

The previous test platform was extended for the present study to use PowerPoint presentation "slides" as the graphic user interface presented to the patient. For this study, the slides were launched and controlled via remote control from the tinnitus test platform. A custom-built dynamic link library was designed and developed to interface PowerPoint objects with the tinnitus test program and also to allow responses to be sent back to the program. Patient interface using the upgraded version is now accomplished via on-screen instructions presented using Microsoft PowerPoint 2000. While the patient progresses through the system in the same manner as with the past version, an added on-screen help option has been added. **Figure 1** shows the progression of slides that provide pretest general instructions.

#### *Hardware*

Because of the combining of two programmatically linked computers required for testing (previously, one

was used for patient control and one for main control of the system), programming advances have made the automated system more portable. The upgraded system uses one slate-type computer (Aqcess Technologies Qbe Personal Computing Tablet) for both patient and main control and supports user-input devices, including mouse, touch screen, and pen-pointing device (**Figure 2**).

The new simplified system includes the knob device, the slate computer, and the Programmable Auditory Laboratory 3000 (PAL 3000) (**Figure 3**) [18]. The PAL 3000 generates the stimuli that are then delivered to the patient via ER-4B Canal Phone™ insert earphones [15,17]. System control parameters and overall testing configuration continue to be database-driven.

The custom-built "knob" device was incorporated for direct patient control of auditory stimuli (**Figures 2 and 3**). The knob is used to control loudness of tones for testing hearing thresholds and loudness matches and testing frequency of tones for pitch-matching. Features of the knob device include (1) continuously variable (i.e., no minimum or maximum stopping points) and (2) turns without detents. Knob resolution is software programmable.

The "knob" is a 64-pulse-per-revolution optical encoder with a one-fourth-inch shaft. The encoder, digital conversion circuits, and Universal Serial Bus (USB) interface are housed in a 5- × 8-inch sloped-top instrument case. The physical knob itself has no indicator markings. The program uses a custom-built software module to track the knob's movements and direction, which the operating system reports to the control program as mouse-wheel movements.

The knob control software also has a "divisor" feature that allows the 64 pulses per revolution to be divided down to fewer pulses per revolution (effectively reducing the resolution per rotation distance). The actual divisor number is test-dependent—one divisor was used for obtaining thresholds and loudness matches and another for pitch-matching. Depending on the test, too many steps per revolution would make the stimulus change too rapidly, while too few steps could frustrate the patient because so much rotation distance would be needed to change the stimulus. Divisor numbers were selected to address these concerns.

#### *Calibration*

We performed automated system calibration as previously documented [12]. As before, a stand-alone Instrument Manager application was used to perform automated calibration and store the data in an instrumentation

<p><b>General Instructions</b></p> <ul style="list-style-type: none"> <li>To indicate your action, touch the "buttons" at the bottom of the screen.</li> <li>If you need further on-screen instructions touch the "?" button.</li> <li>If you need to talk to the audiologist, touch the "Contact Audiologist" button.</li> </ul>	<p><b>General Instructions</b></p> <ul style="list-style-type: none"> <li>When the computer is busy, an hourglass will be displayed.</li> </ul> <p>EXAMPLE </p> <ul style="list-style-type: none"> <li>You will be unable to proceed through the program while the hourglass is visible.</li> </ul>	<p><b>General Instructions</b></p> <ul style="list-style-type: none"> <li>You will control the sounds that you hear by turning the knob.</li> <li>Please try to use slow, steady movements when turning the knob.</li> </ul>
<p>  </p> <p>More Information      GO      Contact Audiologist</p>	<p>  </p> <p>Back      GO      Contact Audiologist</p>	<p>  </p> <p>Back      GO      Contact Audiologist</p>
<p><b>General Instructions</b></p> <ul style="list-style-type: none"> <li>You will be listening to and adjusting a variety of sounds.</li> <li>When a tone is presented, you will see this symbol appear at the top of the screen:</li> </ul> <p>EXAMPLE </p> <ul style="list-style-type: none"> <li>When you see this symbol, you can turn the knob to change the sound.</li> </ul>	<p><b>General Instructions</b></p> <ul style="list-style-type: none"> <li>During part of the test, you will just be reading instructions.</li> <li>At these times, no sound will be presented and you will see this symbol appear at the top of the screen:</li> </ul> <p>EXAMPLE </p> <ul style="list-style-type: none"> <li>You will not be able to adjust the sound when you see this symbol.</li> </ul>	<p><b>General Instructions</b> </p> <ul style="list-style-type: none"> <li>You now have the basic instructions to move through the program.</li> <li>Actual testing with further instructions will begin when you touch the "GO" button.</li> </ul>
<p>  </p> <p>Back      GO      Contact Audiologist</p>	<p>  </p> <p>Back      GO      Contact Audiologist</p>	<p>  </p> <p>Back      GO      Contact Audiologist</p>

**Figure 1.**  
On-screen general instructions for participants tested with automated system.

database. This instrumentation database was run-time accessed by the control program to provide calibrated stimulus levels.

*Procedures*

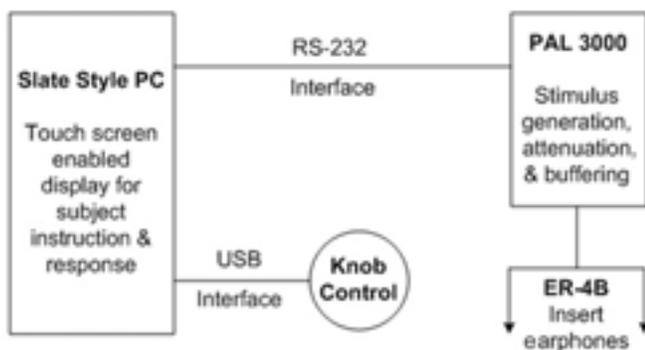
We conducted test procedures in a double-walled sound-attenuated suite constructed by Acoustic Systems (Model RE-245S). All testing was done with the automated tinnitus test system.

*Test Ear and Stimulus Ear*

Stimuli were delivered to the "test ear," which was determined randomly, unless any perceptible hearing asymmetry was found. If one ear had poorer hearing, the contralateral (better) ear was used as the test ear. The "tinnitus ear" was always contralateral to the test ear, and participants were instructed to match the tone presented to the test ear with the tinnitus in the opposite ear. Participants did not have their hearing tested before being



**Figure 2.**  
Participant using new automated tinnitus testing system.



**Figure 3.**  
Block diagram of patient-guided tinnitus measurement system. USB = Universal Serial Bus, PC = personal computer, PAL 3000 = Programmable Auditory Laboratory 3000.

tested with the automated system. The selection of “test ear” and “tinnitus ear” was thus based entirely on the participant’s reported impression of hearing ability between the two ears.

### *On-Screen Instructions*

Prior to testing, participants viewed the series of screens, describing general operation of the automated system (**Figure 1**). Specific instructions were then provided for determining hearing thresholds (**Figure 4**). Our experience (and the experience of others) has shown that patients often confuse the concepts of loudness and pitch [8,16]. Therefore, it is critical for patients to demonstrate that they understand how these concepts differ so that they will respond appropriately. A series of three instruction screens were developed to explain the words “loud-

ness” and “pitch” before participants performed tinnitus loudness- and pitch-matching (**Figure 5**).

### *Test Parameter Randomization*

Although participants had direct control over the adjustment of stimulus parameters, the computer program required algorithms to initially present tones for each new task. A critical feature of the automated system is that each time a new tone is presented, it is presented within a range of possible values to ensure that repetitive cues are not provided to the patient that would enable spuriously reliable responses.

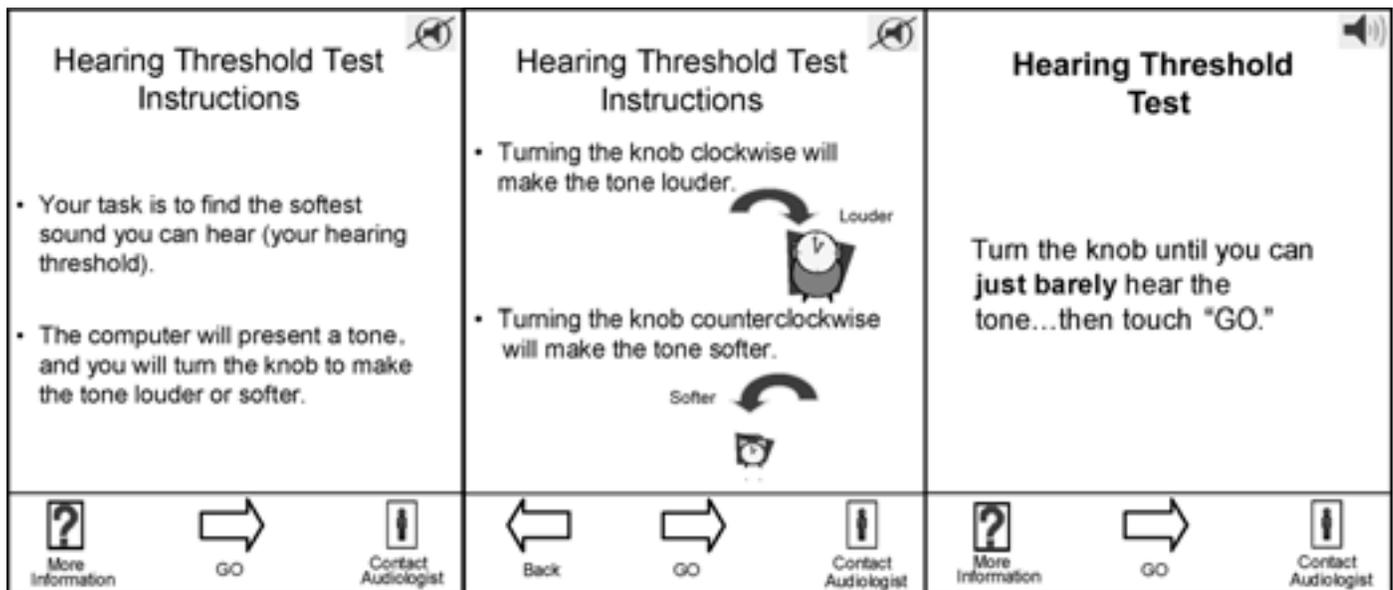
The Microsoft Visual Basic Rnd() function was used as a random number generator. The function returns a value  $\geq 0$  and  $< 1$ . The formula to determine the randomized level or frequency is

$$(\text{int})[(\text{highest value} - \text{lowest value} + 1) * \text{Rnd}() + \text{lowest value}] .$$

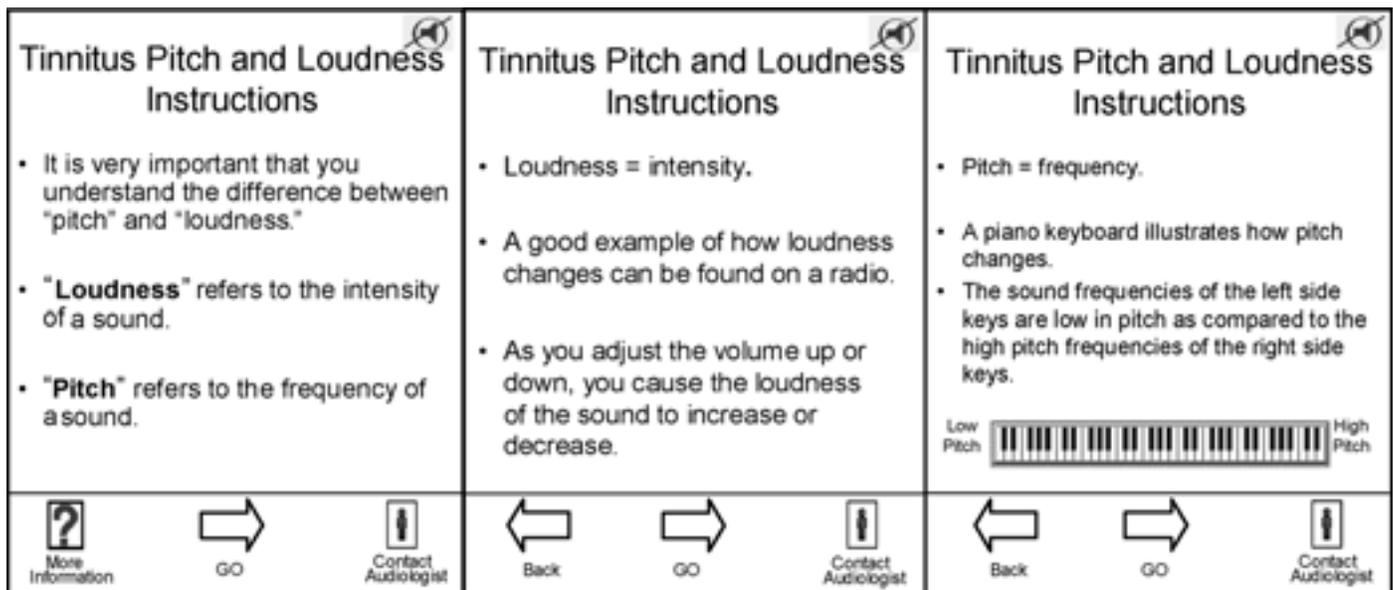
### *Hearing Thresholds*

When the computer obtained thresholds, participants were required to give two responses at each frequency. For the first run, the initial tone was 1,000 Hz presented at 60 dB sound pressure level (SPL). Participants rotated the knob to the point at which the test tone was “just barely” audible. Following a response, the computer presented the tone at a level randomly selected between the initial response and 10 dB above the response level. When a second threshold response was obtained, the computer averaged the two threshold responses to specify the threshold level at that frequency.

For hearing thresholds, the knob divisor was set to “2,” meaning that one complete turn of the knob equaled 32 steps in output-level change, at 1 dB per step. Clockwise turning of the knob increased loudness levels, and counter-clockwise decreased levels. If the knob was turned beyond the maximum calibrated output level (100 dB SPL), continued turning resulted in continuous presentation of the maximum output. With further turning in the same direction (in an attempt to achieve greater output), the program would respond with an instruction screen stating that the maximum output level had been reached. When this screen appeared, one of the response choices was to “try again.” Testing could be repeated up to three times in this manner before the computer would log the “max out” condition and resume testing at the next test frequency.



**Figure 4.**  
Screen displays for obtaining hearing thresholds (instructions and testing) with automated system.

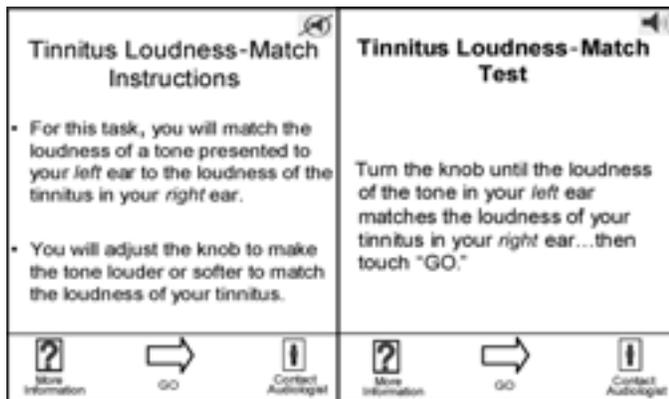


**Figure 5.**  
Screen displays that describe “pitch” and “loudness” prior to performing pitch and loudness-matching with automated system.

### *Loudness-Matching*

Threshold testing at any test frequency was always followed by loudness-matching at the same frequency. For loudness-matching, the same knob divisor of “2” was used as for hearing thresholds. Clockwise rotation always increased the loudness level, and counterclockwise always decreased the loudness level.

Following threshold testing at 1,000 Hz, we provided on-screen instructions to explain the tinnitus loudness-matching task (**Figure 6**). Following the instructions, we presented a 1,000 Hz tone at a level selected at random between the mean threshold level and 10 dB above the mean threshold level. The participant turned the knob and selected the level of the tone that matched the “loudness



**Figure 6.** Screen displays for obtaining tinnitus loudness matches (instructions and testing) with automated system.

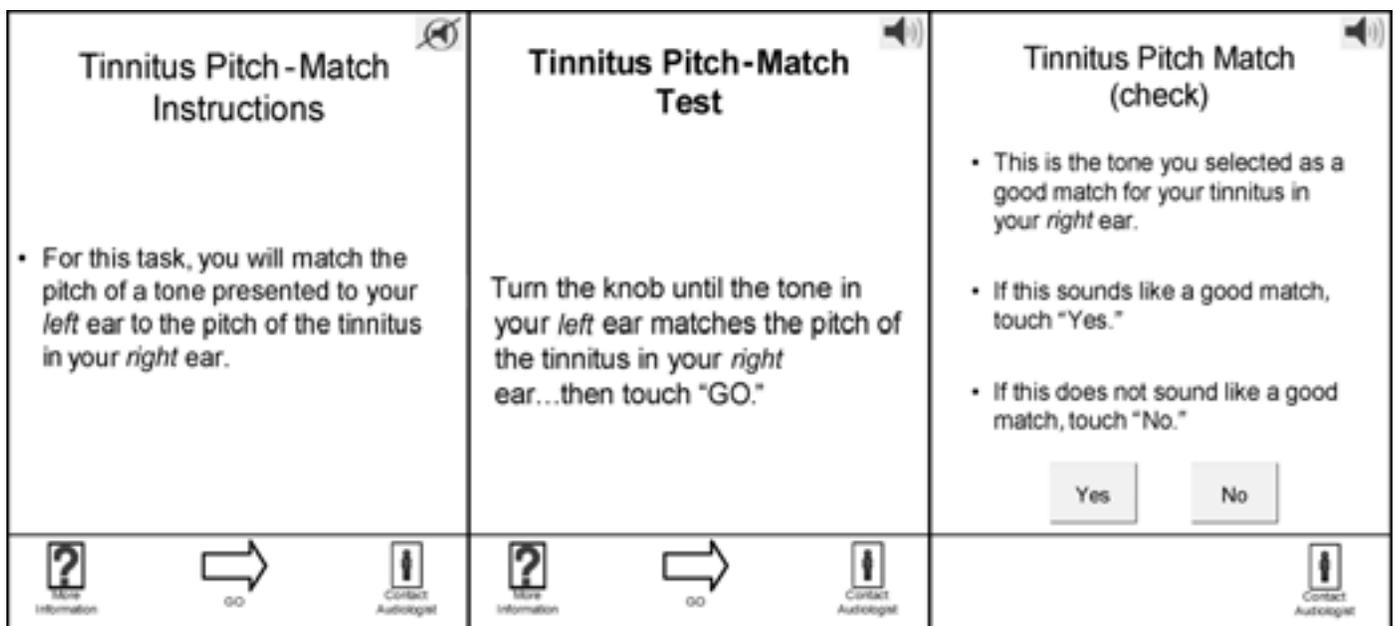
of the tinnitus.” The tone was again presented at a randomized output level between threshold and 10 dB above threshold, and a second loudness match was made. This sequence of testing (threshold followed by loudness match) was then repeated at the next test frequency (1,260 Hz), followed by the remainder of the test frequencies in ascending order. Thresholds and loudness matches were obtained in this manner at all 13 test frequencies, which included 1/3-octave frequencies ranging from 1,000 to 16,000 Hz.

### Pitch-Matching

For the system to perform the pitch-matching protocol, loudness matches were required to be first obtained at all test frequencies  $\leq 8$  kHz. All participants met this requirement. If loudness matches were not provided above 8 kHz, those frequencies were omitted from the pitch-match frequency test set.

For pitch-matching, the knob divisor was “8.” If the participant provided loudness matches at all 13 test frequencies, we required 1 5/8 rotations of the knob to cover the test frequency range. (Fewer test frequencies reduced the number of rotations accordingly.) Rotating the knob clockwise would raise the test frequencies until the maximum frequency was reached. Continued clockwise rotation would then descend the frequencies until the minimum was reached; whereupon, continued clockwise rotation would repeat this sawtooth-like frequency-stepping function. Reversal of the knob direction would always reverse the direction on the sawtooth function.

When hearing thresholds and tinnitus loudness matches had been obtained within the required range of test frequencies, on-screen instructions were presented for pitch-matching (**Figure 7**). Participants were instructed to use the “knob” to control frequency sweep across the test frequencies. Tones at each test frequency were presented at the levels that were previously matched in loudness to



**Figure 7.** Screen displays for obtaining tinnitus pitch matches (instructions and testing) with automated system.

the tinnitus. Participants turned the knob and selected the tone that provided the closest match to the pitch of their tinnitus. After a brief pause, the pitch-matched tone was presented again and the participant was asked if the tone was a “good match” with the tinnitus. Response options were “Yes” or “No.” This sequence of pitch-matching followed by confirmation was repeated five times.

#### Repeated Testing

After providing five pitch matches, participants took a short break, then returned to the sound booth to repeat the test protocol. Thus, two complete tests (hearing thresholds, loudness matches, and pitch matches) were performed during the first session. The second test session was identical to the first, resulting in a total of four complete tests for each participant.

## RESULTS

### Hearing Thresholds

#### Across-Subjects Mean Hearing Thresholds

**Table 1** shows the across-subjects mean hearing thresholds (in decibel SPL) for each of two tests conducted during each of two sessions. No data were missing from test frequencies below 8,000 Hz, i.e., each of the 40

participants provided a response for each of the four tests. At 8,000 Hz, one participant did not respond during the final test, thus the  $n = 39$  for 8,000 Hz. At frequencies above 8,000 Hz, a diminishing number of participants provided responses for each of the four tests (which we expected because of hearing loss that is usually associated with tinnitus). All four responses were obtained from 25 participants at 10,080 Hz, 18 at 12,700 Hz, and 9 at 16,000 Hz.

We performed a repeated measures analysis of variance (ANOVA) at each test frequency on the four hearing-threshold means (**Table 1**). Each ANOVA included only participants who provided a threshold response for each of the four tests (numbers [ $n$ ] at each frequency were just specified). Significant differences ( $p < 0.05$ ) were observed only for the three test frequencies below 2,000 Hz and at 2,520 Hz (**Table 2**). We conducted Sheffé post hoc testing to determine which pairs of means differed significantly. A summary of the post hoc results is shown in **Table 2**.

#### Intervals (Decibel) of Between-Session Difference Scores

Intertest differences in thresholds could be a positive or a negative value, or zero. Regardless of direction, values closest to zero reflected lowest response variability. Analyzing between-session threshold differences with respect to their proximity to zero was thus of interest.

**Table 1.**

Across-subjects mean hearing thresholds (decibel sound pressure level) at each test frequency for each of four tests.\*

Test Frequency (Hz)	Session 1		Session 2	
	Test 1 (Mean $\pm$ SD)	Test 2 (Mean $\pm$ SD)	Test 1 (Mean $\pm$ SD)	Test 2 (Mean $\pm$ SD)
1,000	26.2 $\pm$ 10.5	28.5 $\pm$ 11.4	29.9 $\pm$ 11.6	28.0 $\pm$ 11.1
1,260	28.1 $\pm$ 10.9	29.3 $\pm$ 10.0	32.0 $\pm$ 11.1	29.8 $\pm$ 11.3
1,585	31.7 $\pm$ 12.3	33.0 $\pm$ 11.4	34.1 $\pm$ 12.4	33.9 $\pm$ 12.4
2,000	38.5 $\pm$ 15.3	38.9 $\pm$ 15.4	40.2 $\pm$ 14.3	39.6 $\pm$ 15.0
2,520	44.3 $\pm$ 19.5	44.9 $\pm$ 18.5	46.2 $\pm$ 18.5	45.0 $\pm$ 19.1
3,175	49.6 $\pm$ 19.3	50.3 $\pm$ 18.3	50.9 $\pm$ 18.8	50.4 $\pm$ 19.5
4,000	52.8 $\pm$ 20.6	53.1 $\pm$ 20.0	53.9 $\pm$ 20.4	53.1 $\pm$ 20.7
5,040	56.0 $\pm$ 19.7	56.4 $\pm$ 19.8	56.9 $\pm$ 19.5	56.3 $\pm$ 20.3
6,350	58.4 $\pm$ 21.0	57.9 $\pm$ 21.5	58.8 $\pm$ 21.1	58.2 $\pm$ 21.5
8,000	60.5 $\pm$ 20.9	60.9 $\pm$ 20.9	61.7 $\pm$ 21.1	61.0 $\pm$ 21.4
10,080	76.5 $\pm$ 22.3	76.6 $\pm$ 21.4	75.7 $\pm$ 21.2	77.0 $\pm$ 21.5
12,700	81.0 $\pm$ 19.1	79.4 $\pm$ 22.1	81.3 $\pm$ 22.1	80.9 $\pm$ 20.7
16,000	82.1 $\pm$ 24.1	78.6 $\pm$ 22.3	75.7 $\pm$ 23.0	77.6 $\pm$ 19.9

\*For all frequencies <8,000 Hz  
 $N = 40$

$n = 39$  at 8,000 Hz  
 $n = 25$  at 10,800 Hz

$n = 18$  at 12,700 Hz  
 $n = 9$  at 16,000 Hz

**Table 2.**

Results of repeated measures analysis of variances (ANOVAs) and post hoc tests (Sheffé)\*

Test Frequency (Hz)	<i>p</i> -Value (Repeated Measures ANOVA)	Significantly Different Means (per Sheffé post hoc)	Difference Between Means	<i>p</i> -Value (Sheffé)
1,000	<0.0001	Session 1, Test 1, vs. Session 1, Test 2	2.3	0.0307
		Session 1, Test 1, vs. Session 2, Test 1	3.7	0.0001
1,260	<0.0001	Session 1, Test 1, vs. Session 2, Test 1	3.9	0.0001
		Session 1, Test 2, vs. Session 2, Test 1	2.7	0.0145
1,580	0.0034	Session 1, Test 1, vs. Session 2, Test 1	2.4	0.0124
		Session 1, Test 1, vs. Session 2, Test 2	2.2	0.0223
2,520	0.0194	Session 1, Test 1, vs. Session 2, Test 1	1.9	0.0232

\*Data shown are for only test frequencies where significant differences were observed by repeated measures ANOVAs ( $p < 0.05$ ) between means of hearing threshold as shown in **Table 1**.

**Table 3** displays progressively increasing decibel intervals for the differences to show the range of individual between-session differences in hearing thresholds. A total of 454 between-session threshold differences were found, and difference scores were grouped according to the indicated intervals ranging from  $\pm 1$  dB to  $\pm 30$  dB. Of the 454 differences, 340 (75%) were within  $\pm 5$  dB, 422 (93%) were within  $\pm 10$  dB, and 443 (98%) were within  $\pm 15$  dB. **Table 3** also shows cumulative percentages (in parentheses) from the group of 20 cochlear-impaired participants who were tested with the previous version of the system [17]. The present group of 40 participants apparently had overall greater between-session variability of responses than the group of 20 cochlear-impaired participants. To determine if this difference between the two data sets was

significant, we performed the nonparametric Kolmogorov-Smirnov test [19]. For this calculation, the actual percentages of differences were used (rather than the “cumulative” percentages as shown in **Table 3**). The two data sets were not significantly different ( $p > 0.05$ ).

The assessment of between-session decibel intervals was expanded to analyze between-session intervals at each of the 13 individual test frequencies. **Table 4** shows that, in general, between-session responses were most reliable at frequencies between 1,260 and 6,340 Hz (with respect to intersession differences that were 5 dB or less). **Table 4** also displays the corresponding cumulative percentages (in parentheses) obtained from the 20 cochlear-impaired participants [17] for a direct comparison of reliability between groups. Again, the between-session

**Table 3.**

Cumulative numbers and percentages of individual responses having between-sessions test minus retest threshold differences within various decibel intervals.

Interval (Decibel) in Which Between-Session Threshold Differences Occurred		Cumulative Number of Differences*	Cumulative Percentage of Differences†
From ( $\geq$ )	To ( $<$ )		
-1	1	64	14 (29)
-2	2	148	32 (49)
-3	3	240	53 (67)
-4	4	285	63 (79)
-5	5	340	75 (84)
-10	10	422	93 (96)
-15	15	443	98 (97)
-20	20	448	99 (100)
-30	30	454	100

\*Total number of between-session threshold differences = 454.

†Percentage of differences for cochlear-impaired participants from previous study shown in parentheses. Source: Henry JA, Flick CL, Gilbert A, Ellingson RM, Fausti SA. Reliability of computer-automated hearing thresholds in cochlear-impaired listeners using ER-4B Canal Phone earphones. *J Rehabil Res Dev.* 2003;40(3):253-64.

**Table 4.**

Cumulative percentages of individual responses having between-session test minus retest threshold differences within various decibel intervals.\*

From (≥)	To (<)	Frequency (kHz)												
		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	3 (20)	15 (30)	23 (25)	13 (40)	8 (45)	18 (35)	30 (35)	13 (35)	13 (16)	5 (21)	8 (9)	26 (57)	22 (17)
-2	2	25 (60)	40 (55)	40 (40)	28 (70)	28 (65)	40 (55)	48 (50)	28 (55)	38 (37)	15 (42)	19 (9)	37 (57)	56 (33)
-3	3	43 (75)	58 (70)	55 (60)	53 (75)	55 (75)	60 (85)	73 (75)	55 (65)	48 (63)	43 (53)	35 (27)	53 (71)	56 (83)
-4	4	45 (90)	65 (85)	63 (80)	65 (90)	68 (80)	75 (90)	78 (95)	70 (75)	60 (79)	50 (68)	50 (45)	63 (86)	56 (83)
-5	5	58 (90)	80 (85)	78 (80)	75 (90)	80 (85)	85 (90)	85 (95)	80 (90)	78 (89)	60 (68)	69 (64)	68 (86)	67 (83)
-10	10	90 (100)	93 (95)	95 (100)	88 (100)	93 (95)	98 (95)	100 (100)	90 (95)	95 (100)	95 (95)	85 (82)	95 (100)	89 (83)
-15	15	95 (—)	98 (95)	95 (—)	98 (—)	98 (95)	98 (95)	—	100 (100)	100 (—)	100 (95)	96 (91)	95 (—)	89 (100)
-20	20	98 (—)	100 (100)	98 (—)	98 (—)	100 (100)	100 (100)	—	—	—	— (100)	96 (100)	95 (100)	89 (100)
-30	30	100	—	100	100	—	—	—	—	—	—	100	100	100

\***1.** Each value represents percentage of responses that occurred for intervals indicated. **2.** Corresponding percentages from a similar study using the previous version of the automated system are shown in parentheses. *Source:* Henry JA, Flick CL, Gilbert A, Ellingson RM, Fausti SA. Reliability of computer-automated hearing thresholds in cochlear-impaired listeners using ER-4B Canal Phone earphones. *J Rehabil Res Dev.* 2003;40(3):253-64. **3.** Each percentage for present group is based on 40 responses, except at 10.08, 12.7, and 16 kHz, at which  $n = 26, 19,$  and  $9,$  respectively.

response reliability is apparently consistently better for the previous group of participants. A Kolmogorov-Smirnov test was performed between data sets at each of the 13 test frequencies. For this calculation, we used the actual percentages of differences (rather than the cumulative percentages as shown in **Table 4**). The two data sets were not significantly different at any of the test frequencies (all were  $p > 0.05$ ).

### Tinnitus Loudness Matches

Usually, clinical loudness-match data are expressed in decibel SL (sensation level), i.e., the difference in decibels between the hearing threshold and the loudness match at the same frequency. The decibel SPL loudness matches as obtained from the 40 participants were recalculated to decibel SL, and the recalculated data are shown in **Table 5**. Repeated measures ANOVAs were performed

**Table 5.**

Across-subjects means and standard deviations (SDs) of tinnitus loudness matches (decibel sensation level) made between "stimulus ear" and contralateral "tinnitus ear."\*

Test Frequency (Hz)	Session 1		Session 2	
	Test 1 (Mean ± SD)	Test 2 (Mean ± SD)	Test 1 (Mean ± SD)	Test 2 (Mean ± SD)
1,000	16.8 ± 13.8	14.3 ± 14.3	13.4 ± 12.0	16.5 ± 13.3
1,260	17.3 ± 15.6	15.0 ± 14.0	13.3 ± 11.4	15.3 ± 13.3
1,585	18.7 ± 18.4	16.3 ± 15.4	15.4 ± 11.6	15.5 ± 15.5
2,000	17.5 ± 15.3	15.6 ± 14.5	12.5 ± 11.7	14.4 ± 13.1
2,520	16.0 ± 16.0	14.9 ± 14.6	13.7 ± 12.6	15.4 ± 14.3
3,175	13.5 ± 13.0	12.0 ± 11.5	11.9 ± 11.8	11.7 ± 13.1
4,000	11.8 ± 12.7	11.7 ± 10.6	11.1 ± 12.7	12.1 ± 12.9
5,040	10.5 ± 10.3	10.7 ± 10.1	10.2 ± 12.0	11.7 ± 12.5
6,350	11.3 ± 10.7	10.8 ± 10.0	10.3 ± 12.0	11.9 ± 12.8
8,000	10.3 ± 9.7	9.9 ± 8.4	9.5 ± 9.4	10.5 ± 10.9
10,080	8.3 ± 10.2	9.6 ± 9.3	9.4 ± 9.5	10.0 ± 12.4
12,700	9.2 ± 5.0	8.8 ± 8.4	8.4 ± 9.3	9.2 ± 11.0
16,000	7.0 ± 3.3	8.7 ± 6.3	6.0 ± 5.0	9.6 ± 9.2

\*Each mean based on 40 responses, except at 8.0, 10.08, 12.7, and 16 kHz, at which  $n = 39, 19, 11,$  and  $5,$  respectively.

on the means at each frequency, and no significant differences were observed at any of the test frequencies ( $p > 0.05$ ) except at 2,000 Hz. Post hoc analysis (Sheffé) revealed a significant difference between loudness-match means at 2,000 Hz for only Session 1, Test 1, versus Session 2, Test 2.

#### *Test-Retest Differences in Loudness Matches*

The loudness-match data were analyzed for within-subjects, within- and between-sessions reliability of responses. Two tests were repeated during both Sessions 1 and 2. Within each session, the first response was subtracted from the second response. Positive values thus reflect loudness matches becoming larger when repeated, while negative values reflect the matches becoming smaller. The means of these differences are shown in **Table 6**.

#### *Test-Retest Differences in Loudness Matches: Recomputation of Means to Absolute Values*

We converted the actual values of the differences in the loudness matches to absolute values to reflect magnitude of the differences in responses. The means of the absolute values of the loudness-match differences are shown in **Table 7**.

#### *Pearson Product-Moment Correlations: Within- and Between-Session Reliability*

We calculated Pearson  $r$ 's for the two repeated loudness matches obtained during Session 1 and for the two

loudness matches obtained during Session 2 (**Table 8**). To evaluate between-session reliability, we calculated correlations for the first loudness matches obtained during Session 1 compared to the first loudness matches obtained during Session 2 (all were  $p < 0.05$  except at 16 kHz—within Session 1 and between sessions).

### **Tinnitus Pitch Matches**

#### *Conversion from Hertz to Frequency Positions*

Each of the 40 participants provided 20 individual pitch matches, consisting of five pitch matches selected during each of the four tests. Arithmetic averaging of within-subjects repeated pitch matches is problematic because hertz is a logarithmic frequency scale. For example, a test-retest difference of 1,000 Hz between 1,000 and 2,000 Hz is 1 octave difference, whereas the same difference of 1,000 Hz between 8,000 and 9,000 Hz is only 1/6 octave. We have previously addressed this concern by converting test frequencies in hertz to their *frequency position* in ascending order [3]. With this conversion, differences between frequencies are spaced equally—roughly equivalent to their relative spacing on the basilar membrane [20]. Each pair of adjacent test frequencies was spaced by 1/3 octave, allowing a natural order of frequency positions from 1 (1,000 Hz) to 13 (16,000 Hz). **Table 9** shows the test frequencies in hertz and their equivalent frequency positions.

**Table 6.**

Within- and between-session mean differences (actual values) for loudness matches in decibel sensation level ( $N = 40$ ).\*

Test Frequency (Hz)	Session 1 Difference	Session 2 Difference	Between-Session Difference
1,000	-2.5	3.1	-3.3
1,260	-2.3	2.0	-1.9
1,585	-2.4	0.1	-3.3
2,000	-1.9	2.0	-5.0
2,520	-1.1	1.7	2.4
3,175	-1.5	-0.2	-1.7
4,000	-0.1	1.0	-0.7
5,040	0.2	1.5	-0.3
6,350	-0.5	1.6	-1.0
8,000	-0.5	1.1	-0.8
10,080	1.7	0.1	1.7
12,700	1.6	1.7	1.0
16,000	1.7	3.6	-1.6
<b>Mean</b>	-0.6	1.5	-1.1

\*Each mean based on 40 responses, except at 8.0, 10.08, 12.7, and 16 kHz, at which  $n = 39, 19, 11,$  and  $5,$  respectively.

**Table 7.**Within- and between-session differences (absolute values) for loudness matches in decibel sensation level ( $N = 40$ ).\*

Test Frequency (Hz)	Session 1 Difference	Session 2 Difference	Within- and Between-Session Difference
1,000	7.8	7.1	7.8
1,260	5.1	6.9	6.8
1,585	4.3	5.2	8.0
2,000	4.5	5.1	9.2
2,520	4.3	4.0	6.2
3,175	4.2	4.2	5.5
4,000	4.2	2.7	5.7
5,040	3.6	2.8	4.8
6,350	3.9	2.6	5.4
8,000	4.4	2.5	4.2
10,080	3.9	2.7	4.1
12,700	4.5	2.1	4.8
16,000	3.7	3.6	4.8
<b>Grand Mean</b>	4.5	4.0	5.9

\*Each mean based on 40 responses, except at 8.0, 10.08, 12.7, and 16 kHz, at which  $n = 39, 19, 11,$  and 5, respectively.**Table 8.**Pearson  $r$ 's for within- and between-session loudness matches (decibel sensation level).\*

Test Frequency (Hz)	Session 1 Pearson $r$	Session 2 Pearson $r$	Between-Session Pearson $r$
1,000	0.74	0.76	0.67
1,260	0.90	0.71	0.61
1,585	0.97	0.88	0.66
2,000	0.92	0.76	0.48
2,520	0.91	0.92	0.79
3,175	0.89	0.87	0.81
4,000	0.86	0.96	0.80
5,040	0.83	0.96	0.78
6,350	0.77	0.97	0.67
8,000	0.77	0.95	0.75
10,080	0.78	0.96	0.81
12,700	0.60	0.97	0.73
16,000	0.61 <sup>†</sup>	0.95	0.29 <sup>†</sup>

\*Each mean based on 40 responses, except at 8.0, 10.08, 12.7, and 16 kHz, at which  $n = 39, 19, 11,$  and 5, respectively.<sup>†</sup> $p > 0.05$ 

### Means of Pitch Matches

For each test, the five pitch matches were averaged. Therefore, each participant provided four (averaged) pitch matches—one for each of the four tests. **Table 10** shows the across-subjects means and SDs of the pitch matches. The mean pitch matches ranged from 7.1 to 8.0 frequency positions across the four tests, which would be equivalent to approximately 4,000 to 5,040 Hz. A

repeated measures ANOVA determined that these four means were not significantly different ( $p > 0.05$ ).

### Ranges of Pitch Matches

**Table 10** also shows the means of the ranges of pitch matches. The means of the ranges are relevant to within-subject test-retest variability of the repeated pitch matches. The first of these means is 2.5 for Session 1,

**Table 9.**  
Conversion of test frequencies from hertz to “frequency position.”

Frequency (Hz)	Frequency Position
1,000	1
1,260	2
1,580	3
2,000	4
2,520	5
3,180	6
4,000	7
5,040	8
6,340	9
8,000	10
10,080	11
12,700	12
16,000	13

Test 1, which indicates that the 40 participants had an average range of pitch matches of 2.5 frequency positions during Test 1. Since each frequency position represents a range of 1/3 octave, a range of 2.5 frequency positions =  $2.5 \times 1/3$  or 5/6 octave. Thus, on average, the participants provided five pitch matches during the first test over a frequency range of 5/6 octave. The means of the ranges were also calculated for the two tests combined within each session. The combined means of the ranges were 4.4 (approximately 1 1/2 octaves) for Session 1 and 4.1 (approximately 1 1/3 octaves) for Session 2.

**Table 10** also provides across-session pitch-match data. The column “Test 1s Combined” refers to the Test 1 data combined across sessions (which include five pitch matches from Session 1, Test 1, and five pitch matches from Session 2, Test 1). These data would represent how patients would respond if tested only once (with five repeated pitch matches) during each of two sessions. The mean of the ranges when calculated this way is 4.2, indicating an across-session variability of approximately 1 1/3 octaves. Finally, the mean range was calculated across all four tests, resulting in a value of 5.8 frequency positions (approximately 2 octaves). Thus, the 20 pitch matches provided by these participants spanned an average range of about 2 octaves.

#### Confirmation of Pitch Matches

Following each pitch-match selection, the matched tone was again presented along with the next screen that posed the question, “This is the tone you selected as a good match for your tinnitus in your (*left/right*) ear. If this sounds like a good match, touch ‘Yes.’ If this does not sound like a good match, touch ‘No.’” Each participant responded to this question 20 times—once for each pitch match. **Table 11** shows the number of times a “No” response was obtained for each of the 20 tests. Since 40 participants participated, a response was selected 40 times for each test. For example, **Table 11** shows that the first

**Table 10.**  
Across-subjects means, standard deviations (SDs), and means of ranges of pitch matches (expressed in frequency position) ( $N = 40$ ).

Variable	Session 1			Session 2			Across Sessions	
	Test 1	Test 2	Tests 1 & 2 Combined	Test 1	Test 2	Tests 1 & 2 Combined	Test 1 Combined	All Tests Combined
Mean	7.1	7.6	7.4	7.5	8.0	7.8	7.3	7.6
SD	1.0	1.4	1.5	1.2	0.9	1.5	1.5	1.8
Mean of Range	2.5	3.3	4.4	2.7	2.2	4.1	4.2	5.8

**Table 11.**  
Numbers of times participants selected “No” (out of 40) when asked if the tone they selected sounded like a “good pitch match” with their tinnitus.\*

Test	Pitch Match (Session 1)					Pitch Match (Session 2)				
	1	2	3	4	5	1	2	3	4	5
1	10 (25%)	5 (13%)	7 (18%)	6 (15%)	4 (10%)	6 (15%)	1 (3%)	2 (5%)	2 (5%)	2 (5%)
2	6 (15%)	4 (10%)	3 (8%)	3 (8%)	3 (8%)	4 (10%)	2 (5%)	2 (5%)	2 (5%)	2 (5%)

\*The percentage of times participants selected “No” for each condition is indicated in parentheses.

pitch match from the first test had 10 “No” responses, thus 10 of the 40 participants (25%) indicated that the tone was not a good match the first time pitch-matching was conducted. For all the first Session 1, 13 percent of the responses were “No.” For the Session 2, 6 percent of the combined responses were “No.”

Of the 40 participants, 19 selected “Yes” each of the 20 times they were asked if the tone was a “good match.” Six participants selected “No” one time, and seven selected “No” two times. Six participants selected “No” between three and five times. One participant selected “No” 10 times, and one participant selected “No” 20 times. We analyzed these results to determine if any correspondence could be found between the number of times participants selected “No” and the variability of their pitch matches. No relationships were observed.

### Testing Time

Participants were timed for each complete test (single procedure of hearing thresholds and loudness matches at 13 test frequencies and 5 pitch matches). The computer program logged the overall testing time, which included the general instructions and instructions for each test. The mean time of testing across participants became smaller from the first to the second test during each session. Four tests were conducted (two tests during each session), and the mean testing times for each test were (in order of the four tests) 23.2 (7.4 SD), 18.2 (5.1 SD), 20.5 (6.4 SD), and 17.6 (4.8 SD) minutes (**Table 12**).

## DISCUSSION

Our computer-automated tinnitus evaluation system was modified so that we could enable greater patient control over stimuli presented during testing. With the new system, patients can directly control output levels during threshold testing and tinnitus loudness-matching.

**Table 12.**

Means, standard deviations (SDs), and ranges (in minutes) of time of testing for each of four repeated tests.

Session	Test	Mean	SD	Range
1	1	23.2	7.4	12–46
1	2	18.2	5.1	11–38
2	1	20.5	6.4	11–45
2	2	17.6	4.8	11–29

In addition, patients can control the frequency of tones presented during tinnitus pitch-matching. We designed these modifications to accomplish two primary objectives: (1) reduce testing time so that the automated testing technique will be more suitable for clinical application and (2) improve test-retest reliability of pitch matches. Testing was performed on 40 individuals with chronic tinnitus who would represent typical tinnitus patients. We repeatedly tested each of the participants with the automated system to assess test-retest reliability of hearing thresholds, loudness matches, and pitch matches.

### Hearing Thresholds

The 40 participants all had tinnitus, thus the majority of them also had hearing loss [9,21–23]. In spite of their hearing loss, all but one of the participants were able to provide hearing thresholds at all test frequencies up through 8,000 Hz. The three higher frequencies had progressively fewer responses.

Each participant provided thresholds a total of four times at each frequency, resulting in four across-subjects means that were analyzed by ANOVA at each of 13 test frequencies. For four of these frequencies, ANOVA detected significant differences. From these four frequencies, post hoc testing revealed statistically significant differences between seven pairs of means (out of a total of 78 possible pairings of means considering all test frequencies). Thus only 9 percent of all paired comparisons were “significantly” different. These significant differences ranged from 1.9 to 3.9 dB, which are small differences by clinical standards. Thus, although these differences were statistically significant, they are not clinically significant.

The reason hearing thresholds are obtained with this system is to provide a benchmark reference at each frequency for the corresponding tinnitus loudness matches. Loudness matches are normally expressed in decibel SL, which is the difference between the loudness match and the hearing threshold at the same frequency. The key concern is that the threshold responses are reliable. Within-subject reliability of responses for the present group of 40 participants was compared to a group of 20 cochlear-impaired participants who were evaluated with the previous version of the automated system [17]. Normal variability of hearing thresholds in the clinical setting is accepted as  $\pm 5$  dB [24–27]. For the previous group of 20 cochlear-impaired participants, 84 percent provided between-session repeated

thresholds within  $\pm 5$  dB (**Table 3**). For the present group of 40 tinnitus participants, 75 percent had repeated thresholds within  $\pm 5$  dB. Reliability differences between groups were not significantly different ( $p > 0.05$ ). These results are important because the threshold testing protocols were substantially different between studies. In the previous study, stimulus output was varied by the computer according to response algorithms [17]. Participants in the present study had direct control over stimulus output. The new protocol enables more rapid testing, and the slight decrement in response reliability can be tolerated in light of the time savings.

### Loudness-Matching

In this study, participants provided loudness matches at 13 test frequencies (except at the highest frequencies for some participants). Loudness matches obtained at a number of different frequencies yield *loudness-match functions*. Patterns of loudness-match functions vary individually and are seen most clearly when plotted in decibel SL. Previous analysis of such patterns, with the use of linear regression to obtain best-fit lines, revealed four distinct categories [28]. The most predominant patterns were “merging” (monotonically-decreasing function) and “parallel” (approximately equal decibel SL loudness matches at all frequencies). The mean data from the present group of participants would be described as “merging” (**Table 5**). Further analysis of these data is beyond the scope of this paper. Such analysis is planned, however, because shapes of loudness-match functions may ultimately be a factor that can help categorize patients diagnostically.

We analyzed the loudness-match data to determine within-subject reliability of responses. The within-sessions differences (absolute values showing magnitude of differences) averaged 4.5 dB for Session 1 and 4.0 dB for Session 2 (**Table 7**). The between-session differences averaged 5.9 dB. In a previous study in which an earlier version of our automated system was evaluated, 20 individuals with tinnitus provided repeated loudness matches over two sessions, with an average difference of 4.0 dB [12]. Thus, the present group had some additional variability in their repeated loudness matches. As for the hearing thresholds, this increased variability in repeated loudness matches points out the trade-off between testing time and response reliability. We will continue to attempt to achieve the shortest possible testing time, while achieving reliability that is acceptable for clinical use.

### Pitch-Matching

A total of 20 pitch matches were obtained from each participant, including five during each of two tests performed during each of two test sessions. This design allowed us to assess the reliability of multiple pitch matches both within and between sessions. Only one previous clinical study has provided multiple pitch matches over repeated sessions, which was the study conducted with the previous version of the automated system [3]. Relative to the previous study, the variability of pitch matches was improved with the new technique.

In the present study, five repeated pitch matches during a single test varied, on average, over a range of about 2/3 to 1 octave. The data also indicate that repeating the five pitch matches over two sessions results in an average range of 1 1/3 octaves. Most previous studies of tinnitus pitch-match reliability have generally ranged over several octaves [29–32]. Clearly, all of these studies cannot be directly compared because of substantial methodological differences. The present results, however, appear to indicate significantly improved pitch-match reliability compared to those studies.

A method that has been reported to obtain reliable tinnitus pitch matches is the forced-choice double-staircase (FCDS) procedure [33–34]. Because of the length of time required to perform this traditional psychoacoustic procedure, we had previously not considered the method to be tenable for rapid clinical testing. Although our present results reveal reduced variability of pitch matches relative to most other previous studies, we believe that test-retest reliability can be improved with the automated technique. We are planning, therefore, to evaluate the FCDS method in the next iteration of the computerized system, which is currently under development.

### Confirmation of Pitch Matches

When performing tinnitus pitch-matching, patients are asked to indicate which tone provides the “best match” to the pitch of their tinnitus. They inevitably select a tone, but it is unknown if they consider the tone to be a “good match.” Because of the seemingly inherent variability of pitch matches, we were interested in knowing if patients consider these matches to be “good” or not. We therefore added a new feature to the pitch-matching protocol: following each pitch match, the pitch-matched tone was again presented, along with a question asking if the tone and the tinnitus were a “good match.” In general, about 90 percent of the participants answered “yes” to this

question. The participants in this study therefore seemed to at least think that their pitch matches were “good” matches, although they continued to provide repeated matches that varied over about an octave, more or less.

Individuals with tinnitus seem to typically provide pitch matches over a range of frequencies, which might reflect the fact that their tinnitus does not sound like a pure tone, but rather that it has greater spectral content than for a tone [35]. Asking if the tone is a “good match” at least confirms that the patient thinks that the tone and the tinnitus have the same perceptual quality. Although participants confirmed a good pitch match 75 percent of the time at the initial test, the data in **Table 11** show that this percentage increased to 95 percent by the final tests. Thus, upon repeated testing, patients may be increasingly more certain that they have selected accurate pitch matches, despite their test-retest reliability not improving. This finding further confirms that tinnitus patients have inherent difficulty matching the frequency of pure tones to their tinnitus and would suggest that the bias that is observed can be ascribed to biological variability.

### Testing Time

In our most recent study evaluating different methods of pitch-matching, an average of 22 minutes was required to obtain the same measures as for the present study [3]. That method (“Binary” method), however, started testing at 4 kHz, and if the participant chose a higher frequency for a pitch match, no thresholds or loudness matches were obtained below 4 kHz. The computer algorithm was such that the participant’s selection to a higher frequency would lock in that higher frequency range. This algorithm worked fairly well for most participants, but they did not have the opportunity to select lower frequencies once they selected the higher range. As just mentioned, all studies to date that have performed clinical procedures for pitch-matching have shown large variability in repeated pitch matches. Because of this inherent variability, participants need to select from the entire frequency range when performing pitch-matching. For the present system, they could choose from any frequency between 1,000 and 16,000 Hz to select each of their 20 pitch matches. Although testing time was approximately the same between studies, more data were obtained with the present system than with the previous system and pitch matches were more reliable.

Tinnitus evaluation with the use of the updated automated system has provided useful data essential for us to determine if the system has the potential to be a clinically

useful tool. Testing time with the use of the previous prototype system ranged from 38 to 79 minutes, with a mean testing time of 52.6 minutes for Session 1 and 49.5 minutes for Session 2. Testing time with the use of the updated system ranged from 17 to 48 minutes, with a mean testing time of 19.7 minutes for Session 1 and 18.1 minutes for Session 2. This dramatically reduced testing time is promising regarding clinical use of the system.

Variability in testing time indicated that some patients might have difficulty performing testing with the automated system. The SD of 7.4 minutes for Test 1 from the Session 1 (**Table 12**) would indicate that about 68 percent of patients would be able to complete the test as described within about 30 minutes (95% within about 37 minutes). This finding represents variability that would be difficult to accommodate in routine clinical testing. Patients who have difficulty being tested with the automated system (as indicated by their longer testing times) may need special assistance by the clinician, or they may not be appropriate candidates for automated testing. Future efforts will focus on identifying why some patients take longer than others to perform the test and enacting changes that will improve testing time for all patients. We will further determine if a subcategory of patients exist who could be identified in advance as unsuitable for automated tinnitus testing.

### Evaluation of Tinnitus in U.S. Military Veterans

Veterans can claim tinnitus as a service-connected disability, and this is occurring at an ever-increasing rate. The United States has 132 VA audiology clinics, and clinicians of each of these clinics must perform compensation and pension exams to determine if reasonable evidence exists to support claims that tinnitus was incurred in or aggravated by military service. Veterans are also visiting VA clinics in increasing numbers to obtain treatment for their condition. Whether a veteran is claiming tinnitus as a disability or is seeking treatment, standardized assessment of tinnitus is essential to quantify the perceptual attributes of the disorder. At present, tinnitus measurement at VA audiology clinics is nonuniform and, in many cases, is not performed at all. This situation is of course not unique to the VA healthcare system, because very few non-VA clinics offer sophisticated testing of tinnitus.

Further development of the automated system is ongoing with the primary objective of providing a standardized tinnitus measurement technique that would be

suitable for all VA audiology clinics. This standardized technique would allow veterans to receive routine tinnitus testing in a manner that is uniform across clinics and would help assess the validity of claims submitted to receive compensation benefits. The standardized technique could further facilitate establishment of a VA-wide tinnitus data registry, which would provide a national resource of measurement and epidemiological data from veterans with tinnitus. These data would be invaluable in designing further studies to benefit veterans with tinnitus.

### Future Directions

All of our efforts thus far have been directed toward developing procedures that could be conducted in a clinical setting; thus they needed to be rapid and efficient. The present test requires an average of less than 19 minutes, and ways to reduce this testing time even further will be evaluated. For example, the present technique obtains repeated thresholds and loudness matches at each of 13 test frequencies. These measurements are made to ensure that pitch-matching is performed with tones that have been previously matched in loudness to the tinnitus. For routine clinical quantification of tinnitus, the essential matching measurements are the pitch match and the loudness match at the pitch-match frequency. Thus, the thresholds and loudness matches at the remaining frequencies are presently irrelevant for clinical application (although loudness-match functions obtained in this way may ultimately contain diagnostic information). We also asked participants to confirm each of their pitch matches. This was done to address a research question, and the task would be unnecessary for clinical application. These and other issues will be addressed with the future system to maximally reduce testing time. We are also adding tests that will require additional testing time; thus loudness and pitch-matching need to be as rapid as possible so that sufficient time for these additional tests would be allowed. Our goal is to develop a complete tinnitus assessment battery that can be completed within 20 to 30 minutes.

### CONCLUSIONS

The majority of clinics that provide tinnitus management perform tinnitus quantification using some form of tone- and/or noise-matching. However, the techniques that are used to conduct such testing vary considerably. This variation is not a new concern—over 20 years ago the

Ciba Foundation in London and the National Academy of Sciences advocated the formal establishment of standardized tinnitus evaluation procedures [6–7]. Vernon and Meikle provided procedural details for the tests that were advocated by the Ciba Symposium [8]. To date, standardized test procedures have still not been developed.

We need uniform tinnitus assessment procedures to enable quantification of a patient's tinnitus perception, to specify minimum noise bands that would be therapeutic in masking tinnitus, and to detect treatment-related changes in the tinnitus percept [4–5,14,36]. Development of standardized methodology to measure tinnitus would help determine if any “rules” exist that apply to tinnitus masking and other psychoacoustic effects of tinnitus. Such rules thus far have not been observed, which could be due in part to nonstandardized methodologies used for tinnitus quantification during these studies.

The present findings should result in further movement toward standardization of clinical tinnitus assessment. An automated clinical procedure for tinnitus evaluation should ultimately include the evaluation of all parameters of tinnitus that are relevant to rehabilitation as well as the assessment of tinnitus claims. Further work is underway to increase testing capabilities of the automated system to include other aspects of tinnitus measurement, such as determining bands of noise that match tinnitus and evaluating tinnitus maskability and “residual inhibition” (reduction of tinnitus following presentation of a masking noise). In addition, procedures are being developed for clinicians to assess whether or not a patient has tinnitus as claimed—such testing is expected to involve repeated testing of multiple tinnitus tests for assessing the patient's “response profile.” These improvements and additions are expected to be accomplished within the next 2 to 3 years.

### ACKNOWLEDGMENT

We acknowledge the helpful assistance of Akiko Kusumoto and Theresa Chisholm, PhD, in the preparation of this manuscript.

### REFERENCES

1. Jastreboff PJ. Tinnitus habituation therapy (THT) and tinnitus retraining therapy (TRT). In: Tyler RS, editor. Tinnitus handbook. San Diego (CA): Singular Publishing Group; 2000. p. 357–76.

2. Jastreboff PJ, Jastreboff MM. Tinnitus Retraining Therapy (TRT) as a method for treatment of tinnitus and hyperacusis patients. *J Am Acad Audiol*. 2000;11:162–77.
3. Henry JA, Flick CL, Gilbert A, Ellingson RM, Fausti SA. Comparison of manual and computer-automated procedures for tinnitus pitch-matching. *J Rehabil Res Dev*. 2004;41(2):121–38.
4. Henry JA, Meikle MB. Psychoacoustic measures of tinnitus. *J Am Acad Audiol*. 2000;11:138–55.
5. Tyler RS. The psychoacoustical measurement of tinnitus. In: Tyler RS, editor. *Tinnitus handbook*. San Diego (CA): Singular Publishing Group; 2000. p. 149–79.
6. Evered D, Lawrenson G, editors. *Tinnitus*. Ciba Foundation Symposium 85. London: Pitman Books, Ltd.; 1981.
7. McFadden D. *Tinnitus—Facts, theories and treatments*. Washington (DC): National Academy Press; 1982.
8. 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.
9. Davis A, Refaie AE. Epidemiology of tinnitus. In: Tyler R, editor. *Tinnitus Handbook*. San Diego (CA): Singular Publishing Group; 2000. p. 1–23.
10. Vernon JA, Sanders BT. *Tinnitus questions and answers*. Needham Heights (MA): Allyn & Bacon; 2001.
11. Henry JA, Fausti SA, Mitchell CR, Flick CL, Helt WJ. An automated technique for tinnitus evaluation. In: Reich GE, Vernon JA, editors. *Proceedings of the 5th International Tinnitus Seminar 1995*. Portland (OR): American Tinnitus Association; 1996. p. 325–26.
12. Henry JA, Flick CL, Gilbert AM, Ellingson RM, Fausti SA. Reliability of tinnitus loudness matches under procedural variation. *J Am Acad Audiol*. 1999;10:502–20.
13. Henry JA, Flick CL, Gilbert AM, Ellingson RM, Silaski GC, Fausti SA. Fully-automated system for tinnitus loudness and pitch matching. In: Hazell J, editor. *Proceedings of the Sixth International Tinnitus Seminar*. Cambridge, United Kingdom: The Tinnitus and Hyperacusis Centre; 1999. p. 520–21.
14. Henry JA, Fausti SA, Mitchell CR, Flick CL, Helt WJ, Ellingson RM. Computer-automated clinical technique for tinnitus quantification. *Am J Audiol*. 2000;9:36–49.
15. Henry JA, Flick CL, Gilbert A, Ellingson RM, Fausti SA. Reliability of hearing thresholds: computer-automated testing with ER-4B Canal Phone earphones. *J Rehabil Res Dev*. 2001;38(5):567–81.
16. Henry JA, Flick CL, Gilbert AM, Ellingson RM, Fausti SA. Comparison of two computer-automated procedures for tinnitus pitch-matching. *J Rehabil Res Dev*. 2001;38(5):557–66.
17. Henry JA, Flick CL, Gilbert A, Ellingson RM, Fausti SA. Reliability of computer-automated hearing thresholds in cochlear-impaired listeners using ER-4B Canal Phone earphones. *J Rehabil Res Dev*. 2003;40(3):253–64.
18. Flick CL, Fausti SA, Henry JA, Ellingson RM, Silaski GC. Development of a computerized system for conducting audiometric research. In: Ouslander JG, editor. *Second National Department of Veterans Affairs Rehabilitation Research and Development Conference Proceedings*. Washington (DC): VA Rehab R&D Service; 2000. p. 72.
19. Norman GR, Streiner DL. *PDQ Statistics*. 2d ed. Hamilton, Ontario: B.C. Decker, Inc.; 1999.
20. Gulick WL, Gescheider GA, Frisina RD. *Hearing: physiological acoustics, neural coding, and psychoacoustics*. New York: Oxford University Press; 1989.
21. Axelsson A, Ringdahl A. Tinnitus—a study of its prevalence and characteristics. *Br J Audiol*. 1989;23(1):53–62.
22. Henry JL, Wilson PH. *The psychological management of chronic tinnitus*. Needham Heights (MA): Allyn & Bacon; 2001.
23. Vernon JAE. *Tinnitus treatment and relief*. Needham Heights (MA): Allyn & Bacon; 1998.
24. Witting EG, Hughson W. Inherent accuracy at a series of repeated clinical audiograms. *Laryngoscope*. 1940;50:259–69.
25. Currier WD. Office noises and their effect on audiometry. *Arch Otolaryngol*. 1943;38:49–59.
26. Jerger J. Hearing tests in otologic diagnosis. *Asha*. 1962;4:139–43.
27. Newby HA. *Audiology*. New York: Appleton-Century-Crofts; 1972.
28. Meikle MB, Henry JA, Mitchell CM. Methods for matching the loudness of tinnitus using external tones. In: Reich GE, Vernon JA, editors. *Proceedings of the 5th International Tinnitus Seminar 1995*. Portland (OR): American Tinnitus Association; 1996. p. 178–79.
29. Burns EM. A comparison of variability among measurements of subjective tinnitus and objective stimuli. *Audiology*. 1984;23:426–40.
30. Norton SJ, Schmidt AR, Stover LJ. Tinnitus and otoacoustic emissions: is there a link? *Ear Hear*. 1990;11(2):159–66.
31. Penner MJ. Variability in matches to subjective tinnitus. *J Speech Hear Res*. 1983;26:263–67.
32. Tyler RS, Conrad-Arnes D. Tinnitus pitch: a comparison of three measurement methods. *Br J Audiol*. 1983;17:101–7.
33. Penner MJ, Bilger RC. Consistent within-session measures of tinnitus. *J Speech Hear Res*. 1992;35:694–700.
34. Penner MJ, Klafter EJ. Measures of tinnitus: step size, matches to imagined tones, and masking patterns. *Ear Hear*. 1992;13(6):410–16.
35. Penner MJ, Bilger RC. Psychophysical observations and the origin of tinnitus. In: Vernon JA, Moller AR, editors. *Mechanisms of tinnitus*. Needham Heights (MA): Allyn & Bacon; 1995. p. 219–30.
36. Tyler RS, Aran J-M, Dauman R. Recent advances in tinnitus. *Am J Audiol*. 1992;36–44.

Submitted for publication October 20, 2003. Accepted in revised form August 9, 2004.