Department of Veterans Affairs

JRRD Volume 41, Number 2, Pages 121–138
March/April 2004
Journal of Rehabilitation Research & Development

Comparison of manual and computer-automated procedures for tinnitus pitch-matching

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

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

Abstract—Clinical assessment of tinnitus usually includes pitch-matching between the tinnitus and a pure tone. Although such testing is performed routinely, response reliability has not been demonstrated yet. The present study continues a series of studies designed to develop automated methodology for quantifying tinnitus perceptual characteristics. Three methods for tinnitus pitch-matching were performed in a group of 42 subjects. Two methods were computer-automated (Binary and Subject-Guided) and the third method was a traditional manual technique. Each method provided excellent response reliability for about half of the subjects. The most reliable subjects, however, differed widely between the different methods. Each subject provided a total of 14 pitch matches using the three different methods. Analyses based on each subject's total of 14 pitch matches revealed the range of pitch matches for each subject. About half of the subjects selected pitch matches over a range of 2 1/3 octaves. Results of this study suggest that specifying the range of tinnitus pitch matches rather than attempting to identify a single pitch match may be more appropriate.

Key words: hearing disorders, pitch perception, reliability of results, tinnitus.

INTRODUCTION

Tinnitus is the perception of sound that has no source outside of the auditory nervous system. The condition is symptomatic of an auditory disorder and is not a disease in itself. People with chronic tinnitus must somehow adapt to the constant internal sound, which is only curable in a very small percentage of patients [1]. The majority of individuals with intractable tinnitus are apparently able to adapt quite successfully without the need for clinical intervention. For others, tinnitus becomes problematic. A review of the epidemiological literature reveals that 6 to 17 percent of the general population experience chronic tinnitus and that 1 to 2 percent of the population experience severe distress caused by the condition [2]. These figures are consistent with prevalence estimates by the American Tinnitus Association: 40

Abbreviations: ANOVA = analysis of variance, HL = hearing level, SD = standard deviation, TRT = tinnitus retraining therapy, 2AFC = two-alternative forced choice, VA = Department of Veterans Affairs, VAMC = Department of Veterans Affairs Medical Center.

This material was based upon work supported by the Department of Veterans Affairs Rehabilitation Research and Development (RR&D) Service (C891-RA, C2300R, and RCTR 597-0160).

Address all correspondence to James A. Henry, PhD; RR&D National Center for Rehabilitative Auditory Research, VA Medical Center (NCRAR), P.O. Box 1034, Portland, OR 97207; 503-220-8262, ext. 57466; fax: 503-402-2955; email: james.henry@med.va.gov.

to 50 million individuals in the United States experience chronic tinnitus, of which 10 to 12 million seek professional help and 2.5 million are debilitated to some degree [3].

Tinnitus is becoming an increasingly significant problem for veterans and for the Veterans Health Administration. The latest figures reveal that 196,541 veterans with service-connected tinnitus receive a total of $214,953,048 a year in tinnitus disability compensation [4]. This represents a 1-year increase of $42,232,048 for 34,132 veterans who received a new tinnitus service connection.

For veterans and nonveterans alike, the consequences of problematic tinnitus can include cognitive, emotional, and sleep disorders—all of which can impact the performance of everyday activities [1,5,6]. The consequences of tinnitus are extremely varied in both degree and scope. Unfortunately, individuals requiring treatment are faced with the realization that no standards of clinical management for the condition exist. Tinnitus management is offered in relatively few clinics in this country, and clinicians who provide such services are not accountable to any practice guidelines. Tinnitus patients are thus at a disadvantage in their efforts to pursue the most efficacious course of treatment.

We are working toward developing clinical management procedures for tinnitus that can be used in a standardized fashion. A necessary component of the intake assessment is to quantify the tinnitus perception with respect to its "acoustic" characteristics. Tinnitus quantification is done, using some form of tone and/or noise-matching, in most clinics that provide tinnitus management. The techniques for conducting such testing, however, vary considerably between clinics. There is a long overdue need to provide uniform methodology for conducting tinnitus matching.

For one to understand the clinical utility of tinnitus matching, two basic approaches to providing tinnitus management need to be clarified. The first approach is an attempt to reduce the intensity of the tinnitus symptom, i.e., to alter the tinnitus perceptual aspect(s) in such a way as to reduce its bothersome characteristics. (The ultimate objective of this approach would be to render the tinnitus inaudible.) Such efforts would include pharmacological intervention, surgery, masking, and electrical stimulation. The second approach ignores the tinnitus percept itself and focuses on reducing the patient's *reaction* to the perception. Methods for achieving this include psychological or psychiatric intervention, tinnitus retraining therapy (TRT), and cognitive-behavioral therapy. Each of these methods listed for the two approaches (and there are others) is not necessarily exclusive to one or the other approach. For example, the masking method is used to partially or completely cover the tinnitus sound, but concomitant counseling can be helpful in reducing a patient's reactions. Also, drugs may be used to reduce the perceived loudness of tinnitus, but these drugs can be further effective in alleviating the patient's emotional distress. Often, drugs are used specifically to alleviate distress without any intent to reduce the tinnitus loudness.

When treating a patient specifically to reduce the tinnitus perceptual characteristics, one can only assess treatment efficacy if the tinnitus characteristics are quantified at baseline and during treatment. The primary "acoustic" attributes of tinnitus that should be quantified are its loudness, pitch, and minimum masking level. These measures are often obtained, but without uniform methodology. The measures obtained from a given patient can in fact be fairly useless because of the lack of valid and reliable reference measures. This lack of uniformity further poses a problem for clinical research because the presentation of research evidence requires the use of accepted techniques and known standards for measurement.

Our efforts have been directed toward developing computerized methods to perform tinnitus matching. We have reported a series of studies that have documented the feasibility of such an approach [7–9]. In those studies, the computer-automated testing algorithm was based upon the clinical testing procedure described by Vernon and Meikle [10]. With their method (that is conducted manually by the examiner), the testing protocol starts with tone presentations at 1,000 Hz. Testing proceeds in ascending 1,000 Hz steps to gradually approach the test frequency that most closely matches the patient's perceived tinnitus pitch. Such testing can be time-consuming to determine a pitch match that most often occurs in the 4,000 to 8,000 Hz frequency range [11–13]. It was thus necessary to modify the algorithm to increase the efficiency of testing. We evaluated two variations of testing: the "Octave" and "Binary" methods [14]. With the Octave method, matching tones started at 1,000 Hz, but then progressed in octave intervals to bracket the tinnitus pitch. The Binary method started at a middle audiometric frequency (3,180 Hz), and the order of frequencies was designed to bracket the tinnitus pitch to within a quartile of the test-frequency range. Results of this study revealed that the Binary method provided more reliable pitch matches than did the Octave method. Time of testing was also shorter with the Binary method, but it was still longer than would be practical for clinical application.

In addition to the need to shorten testing time for pitch-matching, there is the concern that these pitch-match methods provide cues that could result in spuriously reliable responses. For both the Octave and Binary methods, testing starts at a fixed frequency and then follows a specific progression of frequency changes according to the individual's responses. It thus might be relatively easy for a patient to provide reliable intertest pitch matches by memorizing and repeating the same sequence of responses. There is no reason to believe that the subjects from the previous study consciously used such cues to provide responses. If a patient so desired, however, such cues would be available with these tests. It is also a possibility that patients who are not skilled at pitch-matching might subconsciously use such cues to respond more reliably than they might otherwise.

To eliminate the testing cues that could lead to specious responses and to further shorten testing time, we designed a new pitch-match protocol, termed the "Subject-Guided" method. With this method, testing always started at a test frequency chosen at random by the computer, thus eliminating any possibility of following a fixed order of frequency presentation. Another innovation with this technique was to enable subject control over test-frequency changes to approach and match the tinnitus pitch. With the control given to the subjects, testing time was expected to be significantly reduced.

Our primary objective of this study was to evaluate the new Subject-Guided automated method to determine if pitch-match reliability could be improved and if testing time could be reduced relative to the Binary method. As a secondary objective, we compared pitch-match reliability using automated testing to the reliability of pitch matches using a conventional manual method.

METHOD

Subjects

Forty-two subjects completed this experiment, including eight females and thirty-four males (mean age 59.9 years; range 23 to 84; standard deviation [SD] 11.9). As with the previous study [14], we selected subjects on the basis of having tonal, stable tinnitus to minimize any variability in the tinnitus that might confound interpretation of the reliability analyses. Twenty-seven of the subjects were previously patients at the Oregon Tinnitus Clinic located at Oregon Health and Science University. Of these 27 patients, 14 were veterans. An additional eight subjects were veterans recruited from the Portland Department of Veterans Affairs (VA) Medical Center (VAMC) Audiology Clinic. Seven subjects were recruited from the local community. The Institutional Review Board Committee at the Portland VAMC approved all use of human subjects for this research. Each subject signed an approved informed consent form prior to study enrollment.

Procedures

Each participant was evaluated with three tinnitus-matching procedures during each of two sessions that were conducted on separate days and within 2 weeks of each other. Two computer-automated procedures ("Binary" and "Subject-Guided") and a "Manual" procedure were used.

Audiological Evaluation

Each participant received an initial audiological evaluation that included hearing thresholds at audiometric frequencies (250 to 8,000 Hz), immittance measures, otoscopy, and case history. Instrumentation and procedures used for the audiological evaluation were as previously described [8]. Briefly, we performed tympanometric screening to rule out active middle-ear pathology using a Grason-Stadler GSI-37 Auto Tymp. Conventional hearing thresholds were obtained manually with a clinical, high-frequency-capable audiometer (Virtual Corp., Model 320) at octave frequencies from 0.25 to 0.8 kHz and at 1.5, 3.0, 6.0, 9.0, 10.0, 11.2, and 12.5 kHz.

Experimental Protocols

Selection of Ear for Stimulus Presentation. For each participant, tinnitus-matching tones were delivered to one ear ("stimulus ear") and tones were matched to the tinnitus in the contralateral ear ("tinnitus ear"). Contralateral presentation of tones was chosen for this study because it is considered less challenging to the patient than ipsilateral matching [8,15]. To determine the tinnitus ear, we queried participants about the location of their most noticeable tinnitus—the ear with the predominant tinnitus was considered the "tinnitus ear." If tinnitus was symmetrical between ears, the examiner selected the tinnitus ear arbitrarily.

Pretest to Evaluate Participants' Understanding of Pitch and Loudness. Prior to obtaining tinnitus measures, we evaluated participants to determine their understanding of the concepts of pitch and loudness. A description of the pretest protocol was recently published and is presented in the **Appendix**, which can be found on-line [14]. The pretest protocol included training if participants demonstrated difficulty distinguishing between pitch and loudness.

Instructions to Participants. Three pitch-matching methods were used in this study. Two were automated methods for which instructions appeared on the participant's video screen, and the third was a method in which the examiner manually provided instructions verbally to the participant through use of a talk-over system. The Binary automated method had three response tasks: threshold testing, loudness matching, and pitch-matching. Instructions for responding were displayed on the subject's video screen each time the task changed. These instruction screens have been shown previously, including those for (1) threshold testing [8], (2) tinnitus loudness matching [8], and (3) tinnitus pitch-matching using the Binary procedure [14]. The instruction screen for the Binary procedure is also shown in **Figure 1**. For the new Subject-Guided pitch-match procedure, the instruction screen is shown in **Figure 2**.

Instructions for Manual testing were not provided on a video screen as for Binary testing. During Manual testing, participants received instructions verbally from the

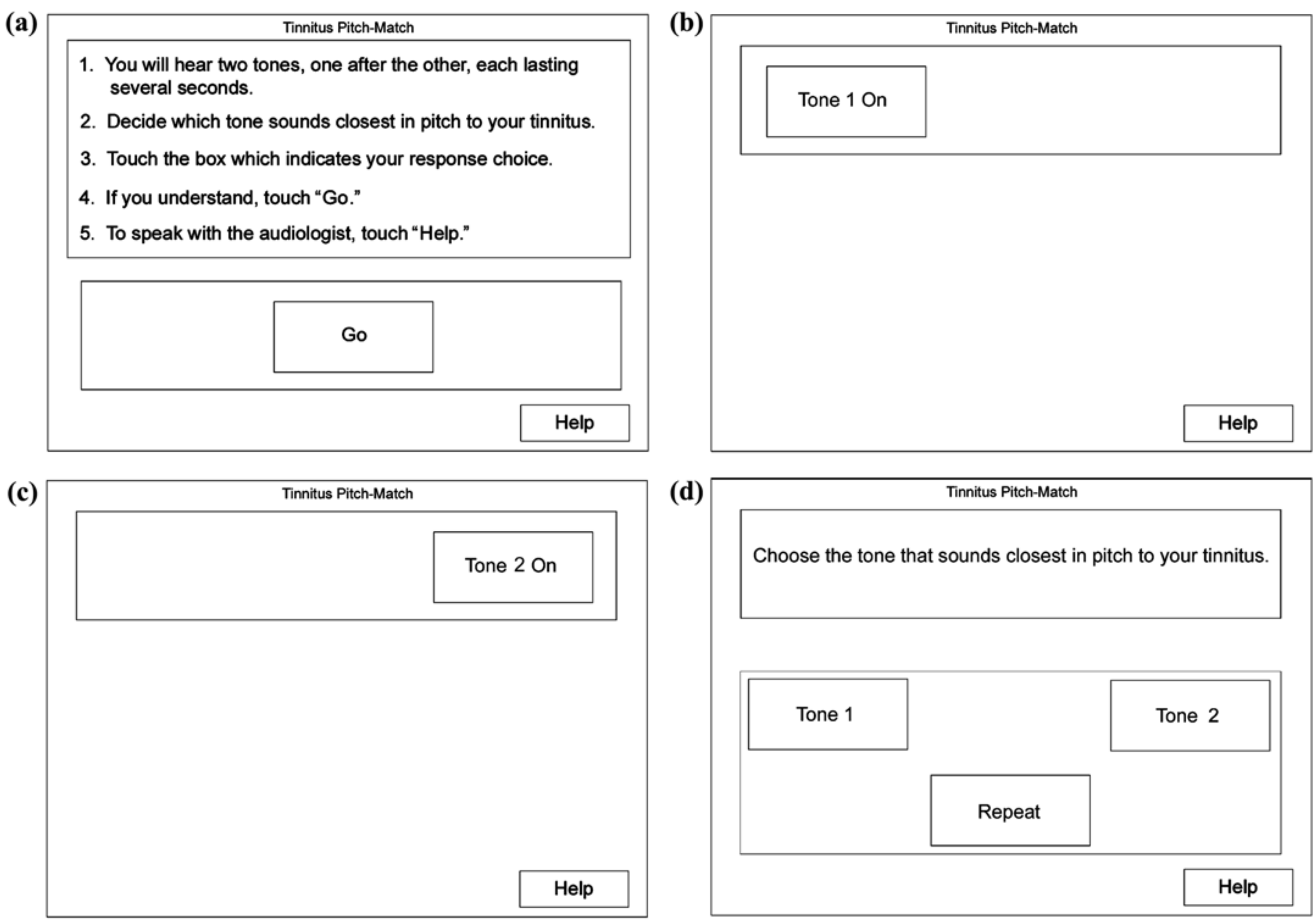

Figure 1.
Screen displays on subject's notebook computer for tinnitus pitch matching: **(a)** instructions; **(b)** indicator that first tone is being presented; **(c)** indicator that second tone is being presented; **(d)** response screen.

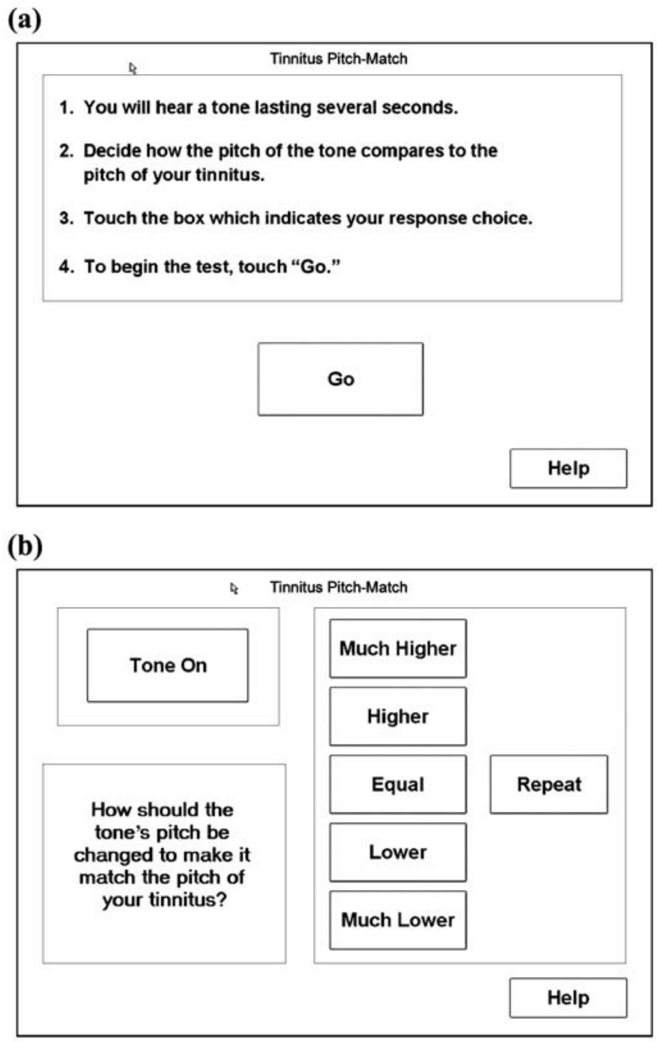

Figure 2.
Screen displays on participants' touch-sensitive video screen for Subject-guided tinnitus pitch matching: **(a)** instructions; **(b)** response screen.

examiner through the talk-over system. For threshold testing, the examiner told participants "raise your hand when you hear a beeping tone." For loudness matching, they were told "listen to the tone and report whether the tone should be made louder or softer to match the loudness of your tinnitus." For pitch-matching, they were told to "listen to two tones, one following the other, then report which of the two tones sounded closest in pitch to your tinnitus." Thus, for loudness-matching and pitch-matching, participants' responses were verbal.

Test Frequencies. For each procedure, a set of test frequencies was available for pitch-matching. Not all of these frequencies were used when obtaining individual pitch matches for the Binary and Manual methods, but they were all used for the Subject-Guided method. Test frequencies for the automated procedures included frequencies from 500 to 16,000 Hz, each separated by 1/3 octave. Test frequencies for Manual testing included 1,000 to 16,000 Hz, in increments of 1,000 Hz.

Order of Testing. The order in which the three pitch-match tests were conducted was counterbalanced between study participants and between sessions. For successive new participants, the first test administered was alternated between the Manual and Binary methods. The Subject-Guided method always followed the Binary method because it used the tinnitus loudness matches that had been obtained during testing with the Binary method. For each participant, the order of testing was reversed at the second session.

"Binary" Method. The Binary method was conducted as described in the previous study [14]. More briefly, we adhered to the premise set for forth by Vernon et al. [16,17] that patients easily confuse loudness and pitch; thus tinnitus pitch-matching should only be done using tones that have been previously matched in loudness to the tinnitus.

The testing protocol involved testing for hearing thresholds, loudness matches, and pitch matches. These three different tasks were "intermingled" according to the following procedural algorithm. At each test frequency, the hearing threshold was obtained to the nearest 1 dB, followed by a loudness match to the nearest 1 dB. An adaptive two-alternative forced-choice (2AFC) procedure was used to present pairs of tones from which the individual chose the one that was "closest in pitch" to his or her tinnitus. Except for octave-confusion testing (see paragraph after next), tone pairs for 2AFC pitch-matching were always separated in frequency by 1/3 octave.

Testing started at 3,180 Hz, followed by testing at 4,000 Hz. When loudness matches had been obtained at these two frequencies, the 2AFC procedure was used to identify which of these two tones was closest in pitch to the person's tinnitus. (The computer screens presented to subjects during pitch-match testing with the Binary method are shown in **Figure 1**.) This selection bracketed the tinnitus pitch to either above 4,000 Hz or below 3,180 Hz (binary bracketing). The computer algorithm then directed testing to, respectively, an octave above or an octave below the initial pair of test frequencies. A new pair of loudness matches was obtained, and the 2AFC procedure again identified whether the tinnitus pitch was above or below the pair of frequencies. Testing occurred in this fashion until the tinnitus pitch was bracketed to within 1 octave. Testing then occurred for tone pairs in the octave-frequency range, starting with the lowest frequency and progressing toward the highest.

When a final pitch match had been selected, "octave-confusion" testing was done. Octave confusion is thought to be a source of error when patients incorrectly identify their tinnitus pitch by selecting a frequency that is an octave above or an octave below their actual tinnitus pitch. Graham and Newby first reported this phenomenon [18], and testing for this error was later advocated as an integral part of any tinnitus-testing battery [19,20]. For the present study, octave-confusion testing was done an octave above and an octave below the pitch match whenever such frequencies were available. Completion of the octave-confusion task identified the "octave-confirmed" pitch match, which was obtained only one time with the Binary procedure.

"Subject-Guided" Method. When the octave-confirmed pitch match was obtained with the Binary method, loudness matches had not been acquired at all test frequencies. Since loudness matches at each frequency were necessary for conducting the Subject-Guided method (which immediately followed the Binary method), the computer then obtained thresholds and loudness matches at all remaining 1/3-octave frequencies, from 500 to 16,000 Hz. The computer then selected 1 of these 17 frequencies at random and presented it at the loudness-matched level. Using the response buttons on the response screen shown in **Figure 2**, the subject directed the computer to make the tone "higher," "much higher," "lower," or "much lower" in frequency until a tone was presented that was selected as a pitch match. When the subject selected "higher" or "lower," the test frequency was increased or decreased, respectively, by 1/3 octave. When the subject selected "much higher" or "much lower," the test frequency was increased or decreased, respectively, by 1 octave.

When the subject selected "equal" to indicate a pitch match, the computer switched to the octave-confusion mode. An instruction screen for octave-confusion testing appeared, and the testing was done as previously described for the Binary method. The Subject-Guided method was repeated five times during each session. However, loudness matches were unnecessary to repeat—only the pitch-matching was repeated five times, resulting in five pitch matches for this procedure during each session.

"Manual" Procedure. We used a tinnitus pitch-matching method that has been well documented for clinical tinnitus evaluation for this portion of the study [10]. We programmed and configured the automated system to facilitate manual evaluation of hearing thresholds, tinnitus loudness matches, and tinnitus pitch matches. The procedures for manual evaluation of tinnitus loudness and pitch have been described in greater detail elsewhere [21].

For Manual testing, the initial test frequency was 1,000 Hz, and frequencies were stepped up in 1,000 Hz increments. Threshold and loudness matches were obtained at each frequency, and the 2AFC procedure was used for subjects to make pitch choices between pairs of frequencies. When the subject chose a final pitch match, octave-confusion testing ensued.

The Manual procedure resembled a typical audiological testing paradigm, in that the examiner provided verbal instructions and presented stimuli to the subject using manual control. Starting at 1,000 Hz, a hearing threshold was obtained to the closest 1 dB followed by a tinnitus loudness match to the closest 1 dB. The test frequency was then changed to 2,000 Hz and a threshold and loudness match were similarly obtained. The 2AFC procedure was then conducted in which the subject was asked to choose the tone that was closest in pitch to his or her tinnitus. If the subject chose the lower frequency (1,000 Hz), then we considered 1,000 Hz the final pitch match, since octave-confusion testing would then be superfluous at 2,000 Hz (it had already been done with the 2AFC procedure) and could not be done below 1,000 Hz (which was the lowest available test frequency).

If the subject chose 2,000 Hz (which was usual) when presented the 1,000 and 2,000 Hz tones as a 2AFC for pitch-matching, a threshold and loudness match were then obtained at 3,000 Hz, followed by 2AFC pitch-matching using 2,000 and 3,000 Hz. Testing progressed in this manner, stepping up in 1,000 Hz steps (which was possible up to 16,000 Hz), until the subject chose the lower of the two tones during pitch-matching. At that point, octave-confusion testing was done at an octave higher frequency and an octave lower frequency (if such frequencies were available). Results of octave-confusion testing determined the octave-confirmed pitch match, which was obtained only once with this method during a test session.

RESULTS

Group Data

Means of the octave-confirmed pitch matches from both sessions are shown in **Table 1** (with SDs shown parenthetically). A single pitch match was obtained during each session for the Binary and Manual methods; thus each of the means for those methods is based on 42 pitch matches. With the Subject-Guided method, five pitch matches were obtained during each session. We calculated means shown in **Table 1** for the Subject-Guided pitch matches using the first of the five pitch matches from each session, as well as the average of the five pitch matches during each session. These means and SDs reveal that the distributions of pitch matches for the group of 42 subjects, regardless of the method used, were typical of a clinical population of tinnitus patients [12,13,22]. Note that the means were approximately bimodal—they were either close to 6,500 Hz or close to 5,000 Hz.

We calculated a repeated measures analysis of variance (ANOVA) on the mean pitch matches shown in **Table 1** to determine if there were significant differences between means. Because multiple pitch matches were obtained from subjects during the Subject-Guided method, only the means for the first pitch matches from each session were used for this ANOVA. Since only one pitch match was obtained during each session with the Manual and Binary methods, using the first pitch match from the Subject-Guided method made the between-group data more comparable. The repeated measures ANOVA revealed that significant differences were found between means [$F(5,41) = 39.8$, $p < 0.05$]. Post hoc analyses (Student-Newman-Keuls) showed that significant differences existed for all paired combinations between the higher three means (Binary session 1 and 2 and Manual session 2) and the lower three means (Manual, session 1, and Subject-Guided, session 1 and 2) (all p's < 0.05).

Between-Session Differences

"Actual" Differences

The individual differences in tinnitus pitch matches between sessions were determined for each method, and the means of these across-subject differences are shown in **Table 1** under the actual average differences. We calculated each difference score for a given subject by subtracting a Session 1 pitch match from the corresponding Session 2 pitch match. Thus a negative difference would reflect a lower-frequency pitch match at Session 2 than at Session 1. Conversely, a positive difference reflects a higher-frequency pitch match at Session 2. The differences varied between positive and negative fairly randomly between subjects for each method. Thus, although the subjects as a group had higher mean pitch matches during Session 2 than Session 1 for each of the methods, no significant trend was found for individual subjects to choose a higher or lower frequency during Session 2 (Wilcoxon test, $p > 0.05$). The mean pitch matches shown in **Table 1** reveal average increases in frequency from Session 1 to Session 2 for each of the three methods. A paired t-test for each method revealed that only the mean increase of 1,572 Hz for the Manual method was significant [$t(41) = -3.1$, $p < 0.05$]. These results were consistent with the Student-Newman-Keuls post hoc tests just described.

Absolute Values of Differences

While the previous analysis was based on the *actual* differences in pitch matches between sessions, **Table 1** also shows the *absolute values* of the same pitch-match differences. The absolute values of the differences reflect

Table 1.
Summary of group pitch-match results, expressed in hertz. Actual average difference provides across-subjects means of actual differences (i.e., including positive and negative values) in pitch matches between sessions. Absolute values of these differences were then determined (to provide all positive values) and absolute value average difference shows means of those absolute values.

Pitch-Match Method	Mean Pitch Match (Hz)		Average Difference		Pearson r
	Session 1 (SD)	Session 2 (SD)	Actual	Absolute Value	
Binary	6,436 (4,294)	6,643 (4,188)	207	1,640	0.807
Manual	4,762 (3,267)	6,334 (4,159)	1,572*	2,382	0.641
Subject-Guided, First Pitch Match	4,859 (3,733)	5,030 (3,421)	171	2,351	0.596
Subject-Guided, Mean of 5 Pitch Matches	4,942 (3,109)	5,438 (3,312)	497	1,518	0.785

*Significantly different (paired t-test, $p = 0.003$)

the magnitude of each difference, ignoring positive or negative changes in frequency. **Table 1** shows that the Manual method had the largest mean difference and that the Subject-Guided method (average of five responses) had the smallest difference. A repeated measures ANOVA was calculated across these four means, and no significant differences [$F(3,41) = 2.2$, $p > 0.05$] were found.

Correlation Analysis

We calculated Pearson product-moment correlation coefficients between Session 1 pitch matches and Session 2 pitch matches for each method. The Pearson *r*'s ranged from 0.596 to 0.807 (**Table 1**). Each Pearson *r* was significant at $p < 0.0001$.

Individual Data

To analyze individual data, we converted the test frequencies in hertz to their *frequency position* in ascending order so that differences between frequencies would be spaced logarithmically—roughly equivalent to their relative spacing on the basilar membrane [23]. **Table 2** shows how the frequency positions were derived. For the two automated methods (Binary and Subject-Guided), each pair of adjacent test frequencies was spaced by 1/3 octave, allowing a natural order of frequency positions from 1 to 16. Test frequencies using the Manual method were available in 1,000 Hz increments from 1,000 to 16,000 Hz. **Table 2** shows conversions to frequency position only for those frequencies that were actually chosen as octave-confirmed pitch matches during Manual testing. The frequency positions assigned to the Manual test frequencies were selected to correspond closely to the same logarithmic spacing of the automated procedures.

Using frequency positions to indicate pitch matches, **Figure 3** shows the pitch matches made by each subject during each of the two sessions. **Figure 3** corresponds to, respectively, the Binary, Manual, and Subject-Guided methods. The pitch matches shown for the Subject-Guided method were the first of the five pitch matches made by each subject during each session with that method. The vertical lines between points in **Figure 3** connect the first- and second-session responses for each subject. For each graph, subjects are ordered left to right on the abscissa from least variability between responses to most variability. The unconnected points on the left portion of each graph indicate that the first and second session responses were identical for those subjects. To examine further the variability of pitch matches, we combined all 14 pitch matches that were obtained from each subject (2 each from the Binary and Manual methods; 10

Table 2.
Conversion of test frequencies from hertz to "frequency position."

Automated Pitch-Match Methods (Binary and Subject-Guided)		Manual Pitch-Match Method	
Frequency (Hz)	Frequency Position	Frequency (Hz)*	Frequency Position
500	1	1,000	4.0
620	2	2,000	7.0
800	3	3,000	9.0
1,000	4	4,000	10.0
1,260	5	5,000	11.0
1,580	6	6,000	12.0
2,000	7	7,000	12.5
2,520	8	8,000	13.0
3,180	9	9,000	13.5
4,000	10	10,000	14.0
5,040	11	12,000	15.0
6,340	12	15,000	16.0
8,000	13	16,000	16.0
10,080	14	—	—
12,700	15	—	—
16,000	16	—	—

*Only test frequencies that were chosen as a pitch match with Manual method are included in this table.

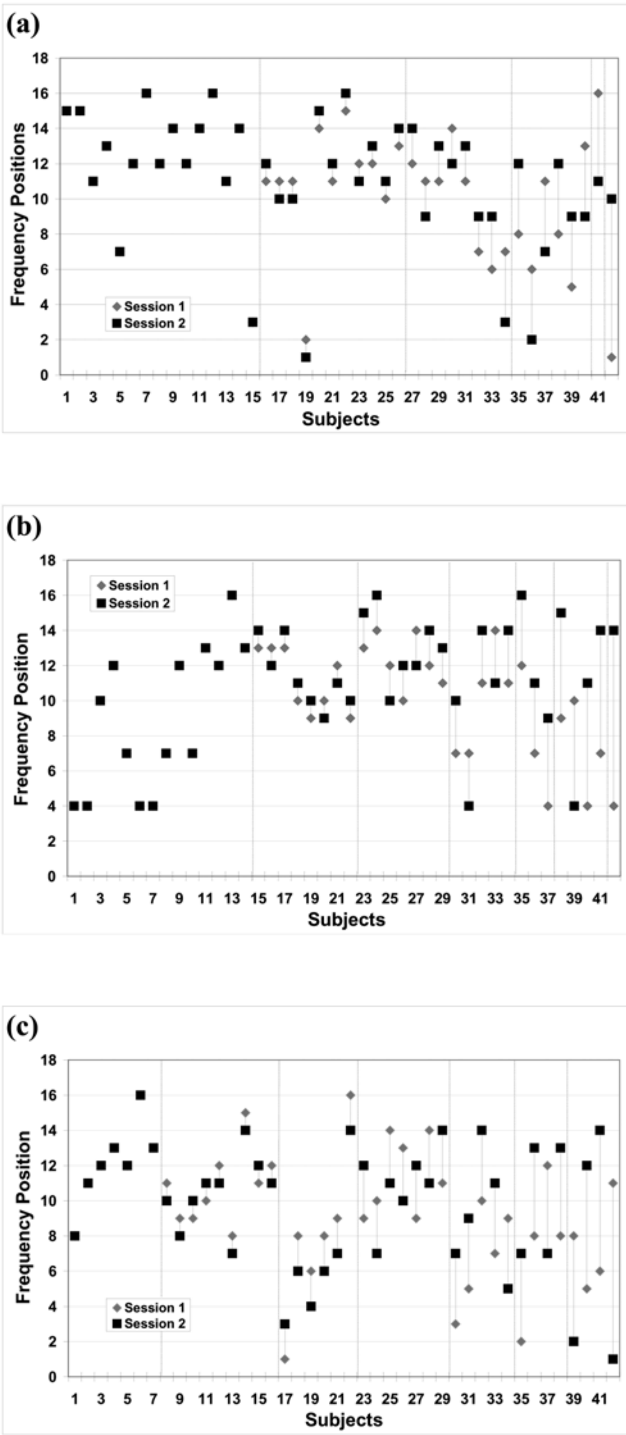

Figure 3.
Scattergrams of methods showing reliability between sessions of pitch matches made by each subject, with vertical lines connecting pitch matches between sessions reliability (single points indicate responses for both sessions were identical): **(a)** Binary, **(b)** Manual, and **(c)** Subject-Guided (first pitch match from each session).

from the Subject-Guided method). The "grand mean" for all of the pitch matches (total of 588 pitch matches made by 42 subjects) was 5,434 Hz, with an SD of 2,147 Hz. With the test frequencies converted to frequency position, the grand mean was 10.1, with a 2.1 SD.

The overall variability of pitch matches is further demonstrated in **Figure 4**, which is a frequency distribution of the range of each subject's 14 pitch matches. As the figure shows, only one subject had a range of zero; thus each of that subject's 14 pitch matches were identical. On the right of the graph, one subject had a range of two frequency positions, indicating a range of approximately 2/3 octave. For further clarification, **Table 3** shows the numbers of subjects for each range of pitch matches. The mean range of pitch matches was seven, corresponding approximately to an average range of 2 1/3 octaves for the 42 subjects.

Within-Subjects, Between-Methods Reliability

The preceding analysis shows that subjects generally provided pitch matches over a range of frequencies, averaging 2 1/3 octaves (**Table 3**). **Figures 3** and **4** reveal how subjects differed from each other with respect to the reliability of their pitch matches. **Figure 3** shows that more than half of the subjects for the Binary and Manual methods provided repeated pitch matches that did not differ by more than 1/3 octave. Also, more than half of the subjects when tested with the Subject-Guided method, provided repeated pitch matches within 2/3 octave. The

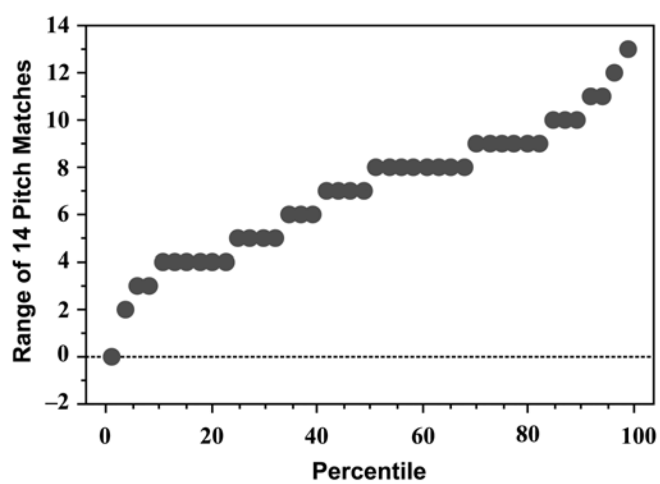

Figure 4.
Cumulative percentiles plot indicating numbers of subjects for each range of pitch matches, based on 14 pitch matches for each subject.

Table 3.
Frequency distribution for range of pitch matches based upon 14 pitch matches for each of 42 subjects.

Range (Frequency Positions)	Range (Octaves)	Count	Cumulative Percent
0	0	1	2.4
1	1/3	0	—
2	2/3	1	4.8
3	1	2	9.5
4	1 1/3	6	23.8
5	1 2/3	4	33.3
6	2	3	40.5
7	2 1/3	4	50.0
8	2 2/3	8	69.0
9	3	6	83.3
10	3 1/3	3	90.5
11	3 2/3	2	95.2
12	4	1	97.6
13	4 1/3	1	100.0
Total —	—	42	100.0

same 42 subjects are shown in each of the graphs, but whether the subjects were consistent in their degree of pitch-match reliability across the three methods was unknown. The analyses described in the subsequent paragraphs address this question.

We calculated Pearson product-moment correlation coefficients on the repeated pitch matches for the 21 most reliable subjects in each group. These analyses revealed that the first 21 subjects shown in each graph of **Figure 3** (not necessarily the same subjects for each analysis) had Pearson *r*'s of 0.99 (Binary), 0.99 (Manual), and 0.95 (Subject-Guided). Thus at least half of the subjects were "highly reliable" for each of the three methods. These same calculations were performed on the remaining 21 (more variable) subjects, with resulting Pearson *r*'s of 0.47 (Binary), 0.32 (Manual), and 0.01 (Subject-Guided).

We calculated a 3 × 3 "correlation matrix" of Pearson product-moment correlation coefficients to determine correlations between individual subjects' pitch-match reliability (using their differences in frequency positions between sessions for each method). Thus we calculated Pearson *r*'s to determine subjects' reliability with each method in relation to each of the other two methods. The Pearson *r*'s in this matrix were all less than 0.3, suggesting that subjects were inconsistent in their reliability across methods.

We conducted further analyses to determine if the same subjects were most reliable across all three methods. Each subject's SD in pitch matches across the three methods was determined. These SDs were calculated separately for Session 1 and Session 2. Each SD was based on three pitch matches: the pitch match from each of the Manual and Binary methods and the first pitch match from the Subject-Guided method. **Table 4** shows the frequency distributions for these SDs. The data shown in **Table 4** are based on frequency positions of the pitch matches, as just described and depicted in **Table 2**. **Table 4** again reveals the high intersubject variability across the different methods. Thus some subjects were clearly more reliable than others across the three methods. The question remaining to be answered was "Are subjects consistent in their degree of between-methods reliability from the first to the second session?"

Calculating the SDs separately for Session 1 and for Session 2 enabled Pearson product-moment correlation coefficients to be calculated between these two groups of data. This analysis determined that the Pearson *r* was 0.09, indicating no correlation between the two groups of SDs. Thus subjects were not consistent in their degree of between-methods reliability from the Session 1 to Session 2.

Analyses of Factors Potentially Affecting Pitch-Match Reliability

We conducted analyses to examine factors that might be related to subjects' degree of pitch-match reliability. A

Table 4.
Frequency distributions for individual standard deviations (SDs) of pitch matches across three methods, calculated separately for each session. All data are shown with respect to frequency positions of pitch matches (explained in **Table 2**).

Interval (Frequency Position) in which Between-Methods SD's Occurred		Number of Subjects	
From (≥)	To (<)	Session 1	Session 2
0.0	0.5	2	3
0.5	1.0	7	8
1.0	1.5	8	6
1.5	2.0	5	7
2.0	2.5	3	8
2.5	3.0	4	6
3.0	3.5	3	0
3.5	4.0	6	2
4.0	4.5	2	2
4.5	5.0	1	0
5.0	5.5	0	0
5.5	6.0	1	0
Total	—	42	42

given subject's overall pitch-match reliability was best described by the SD of the 14 total pitch matches (with respect to the "frequency position" of the pitch matches as just described). These SDs ranged from 0.0 to 4.2 (mean = 2.2) across the 42 subjects. Potential factors were identified and the analyses are described in the following section.

Fluctuating Tinnitus

On the questionnaire, subjects reported whether their tinnitus "tends to fluctuate." In response, 30 subjects reported "no" and 11 reported "yes" in response to this question (one subject did not answer the question). Paired *t*-tests were performed between these two groups to compare the means of the SDs of their 14 pitch matches. Results were not significant [$t(39) = 0.97$, $p > 0.05$].

Number of Tinnitus Sounds

Subjects reported whether they perceived their tinnitus as a "single sound" or as "multiple sounds." Only five subjects reported that their tinnitus consisted of two sounds, while the remainder reported that their tinnitus was perceived as one sound. For the five multiple-sound subjects, the mean of the SDs of pitch matches was 2.22 (frequency positions), and the mean was 2.14 for the 37 subjects who reported tinnitus as a single sound. A paired *t*-test revealed no significant difference between means [$t(40) = –0.18$, $p > 0.05$].

Age

For this group of 42 subjects, only one subject (age 23) was younger than 43 years. The rest of the subjects were fairly evenly disbursed between 43 and 78 years (one subject was 84 years of age). Although this group would be considered generally middle age and older, age may have been a factor influencing pitch-match reliability. To test for this effect, we divided subjects into three age groups of 14 subjects each: "lower" (24 to 53 years), "middle" (54 to 65), and "higher" (66 to 84). The means of their SDs of frequency positions increased for the older age groups from 1.8 to 2.3 to 2.4. We performed a one-way ANOVA to determine if these means differed between groups, revealing no significant differences [$F(2,39) = 1.7$, $p > 0.05$].

Hearing Loss

Another factor that could have affected pitch-match reliability was hearing loss. Test tones for pitch-matching were delivered to the "stimulus" ears of subjects. It seemed possible that pitch-match reliability could be reduced as a function of hearing loss in the stimulus ear. Thirty-six of the forty-two subjects had some degree of hearing loss (defined as at least one hearing threshold of 30 dB or greater). An unpaired *t*-test showed no significant difference with respect to the means of the SDs of pitch matches between the two groups [$t(40) = –0.12$, $p > 0.05$).

We performed a one-way ANOVA on the subjects to further look for an effect of hearing loss on the reliability of pitch matches. To conduct this analysis, we rank-ordered the 42 subjects with respect to the means of their conventional hearing thresholds (in decibels hearing level [HL]) across the audiometric frequencies (250 to 8,000 Hz). We then divided them into four groups (each consisting of 10 or 11 subjects), and their amount of hearing loss was considered "slight" (mean = 3.9 to 18.9 dB HL), "low" (22.2 to 30.6 dB HL) "medium," (31.1 to 35.6 dB HL), and "high" (36.1 to 70.0 dB HL). The ANOVA found no significant differences between these groups [$F(3,38) = 0.85$, $p > 0.05$].

Finally, we conducted a Pearson product-moment correlation analysis between the average of subjects' hearing thresholds (as defined previously for the ANOVA) and the SDs of pitch matches. The Pearson *r* for this analysis was 0.1.

DISCUSSION

Group Data

This experiment has provided reliability data on three distinctly different tinnitus pitch-matching protocols. In general, both of the automated methods provided greater intersession group reliability than did the Manual method. Results show that regardless of which pitch-match method was used, large differences were found between subjects in their ability to produce reliable pitch matches from session to session. Mean pitch matches across the 42 subjects are shown in the second and third columns of **Table 1**, revealing a bimodal distribution of the means. Some of the means centered around 6,500 Hz, while the others were closer to 5,000 Hz.

The bimodal distribution of mean pitch matches was seen between the two automated methods. The Binary method resulted in means of 6,436 and 6,643 Hz from the Session 1 to Session 2. The Subject-Guided method had respective means of 4,859 and 5,030 Hz. While the pairs of means within methods did not differ significantly, the means between methods were significantly different. There were important differences between the automated methods that might have explained this discrepancy. With the Binary method, pitch-matching always involved the presentation of pairs of tones and a selection of one of the tones as "closest in pitch" to the tinnitus. The testing algorithm was restrictive, in that each pitch-match selection determined the direction of testing (to higher or lower frequencies), leaving no opportunity to change the direction following a selection. With the Subject-Guided method, subjects were given control to listen to any tone in the range of test frequencies to select the best pitch match. All other testing parameters were the same for both protocols.

The mean pitch matches from the Manual method were consistent with the bimodal distribution of means. Session 1 mean pitch match (4,762 Hz) for the Manual method was similar to both means from the Subject-Guided method, while the Session 2 mean (6,334 Hz) was similar to both means from the Binary method. Only the between-sessions means for the Manual method were significantly different. Among the three methods, the Manual method provided the least restrictions on procedural variation. Otherwise, the operator, subjects, basic testing protocol, and equipment were consistent from test to test. Other than the greater potential for varying the procedure, no other reason can explain why only the Manual method showed a significant difference between means across sessions.

Between-Subjects Reliability

If any one of the three methods used in this study had been evaluated independently, the findings would have indicated that approximately half of the subjects were highly reliable in their pitch-matching ability. That is, for each of the methods, about half of the subjects provided pitch matches that were repeatable within 1/3 octave from session to session, as shown in **Figure 3**. For each method, the subjects were divided equally between the half who were most reliable and the half who were least reliable. The correlation coefficients were 0.95 or greater for the subjects who showed greater reliability. The remaining 21 subjects were less reliable to various degrees (Pearson *r*'s of 0.47 or less), with between-sessions differences up to 3 1/3 octaves for some subjects.

By combining the data across methods to include all 14 pitch matches for each subject, we were able to show that subjects generally produced repeated pitch matches that varied over a frequency range of 1 or more octaves (**Table 3**). Fifty percent of the subjects made pitch matches over a range of 2 1/3 octaves or less. An additional 40 percent ranged between 2 1/3 and 3 1/3 octaves. The last 10 percent made repeated pitch matches that ranged up to 4 1/3 octaves. With a sample of 42 subjects making 14 pitch matches each, these data suggest that a very small percentage of individuals can match their tinnitus to the same pure tone repeatedly. The majority of individuals will provide matches within a fairly wide range of frequencies. Such results would be consistent with previous studies that have reported lack of test-retest reliability when tinnitus pitch matches are repeated [22,24–26].

Within-Subjects, Between-Methods Reliability

The findings just described would seem to indicate that some individuals are inherently more reliable at pitch-matching than others. Further analysis revealed, however, that subjects who were most reliable with one method were not necessarily the same subjects who were most reliable with another method. For example, subject 1 in **Figure 3** provided identical pitch matches between sessions for each of the three methods. The pitch matches were very different, however, between the methods—note the pitch matches for subject 1 at frequency position 15 for Binary, at 4 for Manual, and at 8 for Subject-Guided. The analyses showed that there is more within-subjects pitch-match variability than would be indicated by repeated testing using only one method. These findings raise the issue of how each method differed with regard to procedures used to obtain a pitch match.

Comparison of Procedures Used for Each Method

The three methods used procedures that were alike in some respects and different in others. Procedures were similar, in that each method presented pure tones to subjects who were instructed to select the tone that most closely matched his or her tinnitus. Tones used for pitch-matching were always presented at the same level that had been previously matched to the loudness of the subject's tinnitus. Presenting tones at loudness-matched levels is thought to make the task easier for patients who often confuse the concepts of pitch and loudness [16,17]. Also, each method provided a mechanism whereby subjects could listen to tones that were higher or lower in frequency in order to select the best match. The mechanism for raising and lowering tonal frequencies, however, represented a major difference between methods. For the Manual and Binary methods, we used the 2AFC procedure. Test tones were presented in pairs of adjacent frequencies, and subjects were instructed to select the one tone that was closest to their tinnitus pitch. The tone that they selected determined whether subsequent tones would be presented at higher or lower frequencies. With the Manual protocol, testing started at a lower frequency and typically progressed to higher frequencies until the subject selected the lower of the two frequencies. The lower-frequency selection determined that octave-confusion testing would take place, which ended the pitch-matching procedure. The Binary procedure functioned similarly, except testing started at a higher frequency and the tinnitus pitch was bracketed to within an octave prior to performing more precise testing within the octave. Thus, both the Manual and Binary methods followed testing algorithms that moved progressively in a certain frequency direction dependent on how subjects selected between the pairs of tones.

Use of the 2AFC protocol for pitch-matching clearly places constraints on response choices provided to subjects. Given the degree of pitch-match variability that is seen for most subjects, it would seem likely that there would be comparable variability in tone selection from each tone pair used in a 2AFC protocol. Each selection specifies the subsequent range of available test frequencies, and once testing moves to a certain range, it cannot go outside of that range except for octave-confusion testing. Thus, with these protocols, pitch-matching follows algorithms that always impose a restricted range of test frequencies.

Because of this limitation, the Subject-Guided protocol was developed. The Subject-Guided protocol allowed subjects to raise or lower the test frequency at will to arrive at a pitch match. Response choices were not based on previous responses, thus they were never limited. Because of this lack of restrictions, the new protocol was expected to result in greater test-retest reliability. Results, however, indicated that responses were just as variable even when the testing limitations were removed.

Potential Sources of Variability

Although the task of matching tinnitus to a pure tone might appear straightforward, several factors can complicate such testing. In general, it has been a consistent observation that the auditory perception of tinnitus does not share the same psychophysical characteristics as does external acoustic stimuli [27–31]. Thus no known rules can be found with regard to effects of external sounds on the perception of tinnitus as can be found in the field of psychoacoustics, which has defined many rules for the perception of sound. If there are rules that relate to tinnitus, those rules will only be discovered when a method is found to reliably quantify the "acoustic" characteristics of tinnitus perception. The reliable quantification of tinnitus pitch remains an elusive goal.

Pitch-match reliability can be influenced by changes in the tinnitus sound itself [26,32]. Meikle and Walsh reported that almost half of a large clinical population of tinnitus patients reported fluctuating tinnitus [12]. These fluctuations could include changes in pitch, loudness and/or timbre. From the present study, 30 of the 42 subjects reported that their tinnitus did not fluctuate. Data analyses determined that the 11 subjects who reported tinnitus that "tends to fluctuate" (one subject did not provide a response to this question) did not have reduced pitch-match reliability relative to the remainder of subjects who did not report fluctuating tinnitus.

Another finding was from a large group of tinnitus patients that showed about half of the patients reported tinnitus consisting of multiple sounds [33]. Patients with such complex tinnitus must identify one of the sounds consistently to perform pitch-matching. The perception of multiple sounds thus might be expected to reduce the reliability of pitch matches. From the present study, 37 of the 42 subjects reported that they perceived only one tinnitus sound. The five subjects who reported hearing more than one tinnitus sound provided pitch matches of comparable reliability. We therefore determined that the potential confounding factors of complex or fluctuating tinnitus did not affect results of the present study.

Another factor potentially influencing pitch-match reliability is hearing loss, which is well documented to reduce frequency-resolving ability [34]. From the present group of subjects, all but six had at least one hearing threshold that exceeded 30 dB HL. The subjects generally had mild to severe high-frequency hearing losses (the mean hearing threshold at 3,000 Hz was 39 dB HL). We determined that the six subjects with "normal" hearing sensitivity (all thresholds 30 dB HL or better) did not demonstrate any better pitch-match reliability than the subjects who had poorer hearing sensitivity. The data were further analyzed in other ways to determine if hearing loss had any effect on pitch-match reliability and no effects could be found.

The analyses of potential human factors that might be associated with reliability showed that none of the factors evaluated, including age, were related in any consistent way to the degree of reliability exhibited by subjects. Future studies should obtain information from subjects concerning other factors that might affect reliability of pitch matches, such as musical ability and knowledge of acoustics.

Testing Time

We recently reported results of a study that evaluated two variations of automated testing (Binary and Octave) to obtain tinnitus pitch matches [14]. The Binary method was designed to shorten time of testing without compromising response reliability. For most clinical and research protocols that have been reported for pitch-matching, testing typically has started at a low frequency (often 1,000 Hz) with progression upward in frequency to gradually approach the tinnitus frequency. For the Binary protocol, testing started at a frequency that is within the frequency range identified as a pitch match by the majority of clinical patients. Since 4,000 Hz is in the middle of that frequency range [12], starting testing at 4,000 Hz eliminated the testing (thresholds, loudness matches, and pitch "preferences") below 4,000 Hz. Our results showed that the Binary protocol in fact only saved about 3 minutes on average (22 minutes for Binary versus 25 minutes for Octave) [14].

We determined time of testing for each of the three methods in the present study. An average of 22 minutes was again required to obtain a pitch match with the Binary method. The Manual method required an average of 12 minutes. The Subject-Guided method required an average of 2.7 minutes to obtain the first pitch match within a session and an average of 1.9 minutes to obtain each repeated pitch match within the same session. Thus obtaining the five Subject-Guided pitch matches within a session required an average of 10.3 minutes. However, note that for the Subject-Guided method, the time required to obtain thresholds and loudness matches at each frequency was not factored into the time of testing. Thus obtaining a pitch match with this method would take much longer if total testing time was combined.

With each of these methods, the progression of test frequencies required thresholds and loudness matches to be obtained at each frequency with 1 dB precision. Threshold and loudness-match measures consumed the majority of testing time for each technique, yet these measures were only used in the process of determining a pitch match. If obtaining a pitch match is the only objective, then the time required to determine thresholds and loudness matches in the process must be significantly shortened. We are presently developing new methodology to shorten testing time, as well as to add other tinnitus measurements that would be necessary for a comprehensive tinnitus evaluation [35]. For clinical application, the complete evaluation should require no more than about 30 minutes per test ear. This is a major objective of the present effort.

Importance of Tinnitus Psychoacoustic Assessment

Why is it important to quantify the psychoacoustic characteristics of tinnitus? This question has already been answered in a number of contexts. The Ciba Foundation in London and the National Academy of Sciences put forth formal efforts over 20 years ago to promote the establishment of standardized tinnitus evaluation procedures [36,37]. Both of these efforts recommended routine tinnitus assessment to include pitch-matching, loudness-matching, the maskability of tinnitus, and residual inhibition (reduction in tinnitus loudness following the presentation of specified noise). In association with the Ciba Symposium, Vernon and Meikle offered a set of procedural details for these four tests [17]. A considerable number of studies ensued, but standardized test procedures still have not been adopted universally. The procedures advocated by Vernon and Meikle have been "more or less" the standard, although special equipment is required to perform their techniques. Currently, no special tinnitus testing equipment exists that is commonly accepted for routine clinical tinnitus evaluations. Most audiologists who do tinnitus evaluations use their audiometers in some manner to obtain some or all of these measures. Apparently, interest in providing standardized techniques and procedures has diminished considerably. The question as to why these measures are important thus needs to be revisited.

In all medical disciplines, disorders are quantified through diagnostic procedures. Whenever treatment for a condition is required, the symptoms need to be thoroughly described to implement appropriate intervention. During the 1980s, the field of clinical tinnitus management seemed to be advancing toward establishing standardized methods of diagnosis and treatment. The Tinnitus Masking approach was especially popular during that time, and the tinnitus measures were important to establish an individualized masking program [38–40]. Although the Tinnitus Masking approach continues to be used, the TRT approach also became popular during the 1990s [41]. Unlike Masking, TRT places little importance on tinnitus measurements [21,38,42,43]. With TRT, tinnitus measurements are considered inconsequential in the design of appropriate intervention and are used mainly for patient counseling. The cognitive-behavioral approach to tinnitus management also places little importance on a tinnitus psychoacoustic assessment and emphasizes a description of the psychological impact of tinnitus as the critical diagnostic factor [2]. Furthermore, with the branching of tinnitus management into a variety of medical and alternative specialties, the greater emphasis regarding diagnosis seems to have become primarily the use of outcomes instruments. Presently, the main concern with describing the perceptual symptoms of tinnitus may be that practitioners generally do not know what to do with the information obtained.

Although the prevailing attitudes of some would seem to downplay the importance of tinnitus psychoacoustic assessment, we contend that it is needed now more than ever. Most generally, standardization of tinnitus measures would advance international understanding and facilitate clinical and research efforts. Consistent with all healthcare disciplines, quantifying the symptom is fundamental to understanding its mechanisms and treatments [32]. Some of the reasons why valid and reliable measurement of tinnitus symptoms is important include—

- Diagnosing the underlying cause(s) of tinnitus.
- Designing a treatment that addresses the tinnitus symptoms as well as the psychological effects.
- Determining effects of treatment on tinnitus perceptual characteristics.
- Providing quantification of a patient's subjective symptoms to counsel patient and family.
- Providing a means to evaluate the presence of tinnitus for detecting tinnitus "malingering."
- Enabling the determination of psychoacoustic "rules" of Tinnitus Masking.
- Assisting in elucidating tinnitus mechanisms.

A Proposed Interpretation of Pitch-Match Reliability

Tinnitus loudness matches have been shown to be provided reliably by most individuals with chronic tinnitus—repeated measures are usually obtained within a few decibels of each other both within and between sessions [8]. Obtaining consistent tinnitus pitch matches has proven to be more of a challenge for investigators [22]. Historically, every clinical study that has reported repeated pitch matches has shown high variability of responses. The present results corroborate those studies, and the reader might wonder how such seemingly unreliable responses can have value for either research or clinical application. The reliable responses seen with tinnitus loudness matches might suggest that the same level of reliability should be seen with tinnitus pitch matches. This does not appear to be the case, but proposing a different interpretation of pitch-match reliability may now be appropriate.

Our studies have obtained multiple pitch matches from many subjects, and clearly, very few individuals with tinnitus can identify the same tinnitus frequency consistently. Tinnitus patients typically describe their tinnitus as "tonal," but the percept may in fact be more "spectral," i.e., containing a band, or bands, of frequencies. Also, tinnitus patients often report that their tinnitus consists of "multiple sounds." It, therefore, may be the percept itself that results in the response inconsistencies that are typically observed. Of course other potential sources of error can be found, which could include frequency-resolving ability, level of musical or acoustic sophistication, and normal response variability.

The present study, combined with our previous studies, seems to be highlighting the importance of documenting the range of pitch matches for each patient rather than just a single measure. Obtaining repeated pitch matches as standard clinical protocol might show that individuals are reliable with respect to the range within which responses are given repeatedly. Reliability of tinnitus pitch matches thus may be a more complex issue than just obtaining the measures repeatedly. Future investigations should explore these complexities so that patients can be categorized with respect to the characteristics of their repeated pitch matches. Performing repeated pitch matches within a single session would require that patients could perform the testing rapidly. Such investigations may indicate that describing a patient's "tinnitus frequency" may be done most appropriately by specifying the range of tinnitus pitch matches as a single measure.

CONCLUSIONS

Tinnitus has become a major problem both within and outside of the VA healthcare system. Yet, no consensual techniques can be found to quantify the perceptual aspects of the disorder. The present study represents another step in our efforts to develop clinical methodology to accurately and reliably quantify the phantom sensation of tinnitus. This work is specifically directed toward determining a uniform method for obtaining tinnitus pitch matches routinely in the clinical environment. The findings of this study suggest that tinnitus pitch might be better viewed as a range of frequencies that can only be obtained through repeated testing.

The development and documentation of uniform pitch-matching procedures will provide clinicians and researchers the ability to quantify a patient's tinnitus perception, to identify therapeutic masking noise that is specific to the tinnitus frequency range, and to detect changes that may occur during treatment [9,22,32,44]. The ability to quantify tinnitus pitch reliably would further lead to studies that could establish "rules" for the masking of tinnitus and other psychoacoustic effects of tinnitus. Such rules currently cannot be determined because of the apparent range of pitch matches that are typically obtained. Valid and reliable pitch-match procedures are also important for clinical and basic research purposes—tinnitus must be quantifiable to observe effects of any sort of intervention. Continuation of the present work is expected to provide a reliable technique for tinnitus psychoacoustic assessment that can be used as a useful clinical tool.

ACKNOWLEDGMENT

We acknowledge the helpful assistance of Craig Dennis in the preparation of this paper.

REFERENCES

1. Axelsson A. How to evaluate treatments for tinnitus. In: Vernon JA, editor. Tinnitus treatment and relief. Needham Heights (MA): Allyn & Bacon; 1998. p. 1–7.
2. Henry JL, Wilson PH. The psychological management of chronic tinnitus. Needham Heights (MA): Allyn & Bacon; 2001.
3. Davis A, Refaie AE. Epidemiology of tinnitus. In: Tyler R, editor. Tinnitus handbook. San Diego (CA): Singular Publishing Group; 2000. p. 1–23.
4. Office of Policy and Planning: VA Central Office. Washington, DC; September 2002.
5. Erlandsson SI, Hallberg LRM, Axelsson A. Psychological and audiological correlates of perceived tinnitus severity. Audiology 1992;31:168–79.
6. Tyler RS. Tinnitus disability and handicap questionnaires. Seminars Hear 1993;14:377–83.

7. 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.
8. 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.
9. 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.
10. Vernon JA, Meikle MB. Measurement of tinnitus: an update. In: Kitahara M, editor. Tinnitus. Pathophysiology and management. Tokyo: Igaku-Shoin; 1988. p. 36–52.
11. Reed GF. An audiometric study of two hundred cases of subjective tinnitus. Arch Otolaryngol 1960;71:94–104.
12. Meikle M, Walsh ET. Characteristics of tinnitus and related observations in over 1800 tinnitus patients. J Laryngol Otol Suppl 1984;9:17–21.
13. Meikle MB, Creedon TA, Griest SE. Tinnitus archive. 2d ed. 2004 [cited 2004 Mar 13]. Available from: http://www.tinnitusArchive.org/.
14. 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: 557–66.
15. Goodwin PE, Johnson RM. The loudness of tinnitus. Acta Otolaryngol 1980;90:353–59.
16. Vernon JA, Fenwick J. Identification of tinnitus: a plea for standardization. J Laryngol Otol 1984;45 Suppl 9:45–53.
17. 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.
18. Graham JT, Newby HA. Acoustical characteristics of tinnitus: an analysis. Arch Otolaryngol 1962;75:82–87.
19. Vernon J, Johnson R, Schleuning A. The characteristics and natural history of tinnitus in Meniere’s disease. Otolaryngol Clin North Am 1980;13:611–19.
20. Evered D, Lawrenson G. Appendix II. Guidelines for recommended procedures in tinnitus testing. In: Evered D, Lawrenson G, editors. Tinnitus. Ciba Foundation Symposium 85. London: Pitman Books, Ltd.; 1981. p. 303–6.
21. Henry JA, Jastreboff MM, Jastreboff PJ, Schechter MA, Fausti SA. Assessment of patients for treatment with Tinnitus Retraining Therapy. J Am Acad Audiol 2002;13:523–44.
22. Henry JA, Meikle MB. Psychoacoustic measures of tinnitus. J Am Acad Audiol 2000;11:138–55.
23. Gulick WL, Gescheider GA, Frisina RD. Hearing: physiological acoustics, neural coding, and psychoacoustics. New York: Oxford University Press; 1989.
24. Penner MJ. Tinnitus as a source of internal noise. J Speech Hear Res 1986;29:400–406.
25. Smith PA, Parr VM, Lutman ME, Coles RRA. Comparative study of four noise spectra as potential tinnitus maskers. Br J Audiol 1991;25:25–34.
26. Penner MJ, Bilger RC. Consistent within-session measures of tinnitus. J Speech Hear Res 1992;35:694–700.
27. Burns EM. A comparison of variability among measurements of subjective tinnitus and objective stimuli. Audiology 1984;23:426–40.
28. Penner MJ. Variability in matches to subjective tinnitus. J Speech Hear Res 1983;26:263–67.
29. Tyler RS, Conrad-Armes D. Tinnitus pitch: a comparison of three measurement methods. Br J Audiol 1983;17:101–7.
30. Feldmann H. Homolateral and contralateral masking of tinnitus by noise-bands and by pure tones. Audiology 1971;10: 138–44.
31. Hazell JWP. Measurement of tinnitus in humans. In: Evered D, Lawrenson G, editors. Tinnitus. Ciba Foundation Symposium 85. London: Pitman Books, Ltd.; 1981. p. 35–48.
32. Tyler RS. The psychoacoustical measurement of tinnitus. In: Tyler RS, editor. Tinnitus Handbook. San Diego (CA): Singular Publishing Group; 2000. p. 149–79.
33. Meikle MB, Vernon JA, Johnson RM. The perceived severity of tinnitus. Otolaryngol Head Neck Surg 1984;92(6): 689–96.
34. McFadden D, Wightman FL. Audition: Some relations between normal and pathological hearing. Annu Rev Psychol 1983;34:95–128.
35. Westphal BE, Ellingson RM, Schechter MA, Fausti SA, Henry JA. Development of a new clinical system for automated tinnitus evaluation. In: Patuzzi R, editor. VII International Tinnitus Seminar Proceedings. Perth, Western Australia: Physiology Department, University of Western Australia; 2002. p. 18.
36. Evered D, Lawrenson G, editors. Tinnitus. Ciba Foundation Symposium 85. London: Pitman Books, Ltd.; 1981.
37. McFadden D. Tinnitus—Facts, theories and treatments. Washington (DC): National Academy Press; 1982.
38. Henry JA, Schechter MA, Nagler SM, Fausti SA. Comparison of Tinnitus Masking and Tinnitus Retraining Therapy. J Am Acad Audiol 2002;13:559–81.
39. Vernon JA, Meikle MB. Tinnitus Masking. In: Tyler RS, editor. Tinnitus Handbook. San Diego (CA): Singular Publishing Group; 2000. p. 313–56.
40. Vernon JAE. Tinnitus treatment and relief. Needham Heights (MA): Allyn & Bacon; 1998.
41. Jastreboff PJ. Phantom auditory perception (tinnitus): mechanisms of generation and perception. Neurosci Res 1990;8:221–54.

42. Jastreboff PJ. Usefulness of the psychoacoustical characterization of tinnitus. In: Reich GE, Vernon JA, editors. Proceedings of the Fifth International Tinnitus Seminar 1995. Portland (OR): American Tinnitus Association; 1996. p. 158–66.
43. Jastreboff PJ, Hazell JWP, Graham RL. Neurophysiological model of tinnitus: dependence of the minimal masking level on treatment outcome. Hear Res 1994;80:216–32.
44. Tyler RS, Aran J-M, Dauman R. Recent advances in tinnitus. Am J Audiol 1992:36–44.

Submitted for publication January 6, 2003. Accepted in revised form April 7, 2003

APPENDIX

Pretesting Evaluation to Determine Subjects' Understanding of "Loudness" and "Pitch"

Hearing Thresholds

1. Instruct subject for threshold testing: "You will hear soft beeping tones. Raise your hand when you hear a tone."
2. Obtain hearing thresholds at 1,000, 1,260, and 4,000 Hz (to closest 5 dB).

"Loudness"

1. Instruct subject to choose the louder of each pair of tones: "You will hear two tones, one followed by the other. After you hear both tones, tell me which tone was the louder of the two."
2. Present 1,000 Hz tone at 10 dB SL, followed by 1,000 Hz tone at 20 dB SL. Log subject's choice.
3. Instruct subject: "Listen to two tones again, and choose the louder of the two."
4. Repeat Steps 1–3, except reverse the order of presentation, i.e., 20 dB SL followed by 10 dB SL. Log subject's choice.
5. If subject chose correctly for each of the two-tone pairs, log subject as "understands loudness."
6. If subject chose incorrectly for at least one of the first two presentations, ask the subject, "Is it clear to you how to tell whether one sound is louder than another?"

(a) If the subject responds that it is clear, retest as for Steps 2–4.

(b) If subject reports that he or she does not understand "loudness," instruct, "A louder tone pushes harder on your eardrum than a softer tone. For example, a jet engine is louder than a whisper. Think about making your radio louder by turning up the volume." Then retest as for Steps 2–4.

7. If subject does not respond correctly for three total presentations (i.e., three times Steps 2–4), subject is logged as "doesn't understand loudness."

"Pitch"

1. Instruct subject to choose the "higher pitched" tone of each pair of tones: "You will hear two tones, one followed by the other. After you hear both tones, tell me which tone was the higher in pitch of the two."
2. Present 1,000 Hz tone at 10 dB SL, followed by 4,000 Hz tone at 10 dB SL. Log subject's choice.
3. Instruct subject: "Listen to two tones again, and choose the higher pitched of the two."
4. Repeat Step 2, except reverse the order of presentation, i.e., 4,000 Hz followed by 1,000 Hz. Log subject's choice.
5. If subject chose correctly for each of the two-tone pairs, repeat Steps 2–4, except use 1,260 Hz instead of 4,000 Hz.
6. If subject chose correctly for all presentations, log subject as "understands pitch."
7. If subject chose incorrectly for any presentations, ask the subject, "Is it clear to you how to tell whether one sound is higher in pitch than another?"

(a) If the subject responds that it is clear, retest as necessary for Steps 2–5.

(b) If subject reports that he or she does not understand "pitch," instruct, "The pitch of a sound refers to whether it is a low sound (such as a man's voice) or a high sound (such as a woman's voice)." Then retest as necessary for Steps 2–5.

17. If subject does not respond correctly for three total presentations (i.e., three times Steps 2–5), subject is logged as "doesn't understand pitch."