Intrasubject Reliability of High-Frequency (9–14 kHz) Thresholds: Tested Separately vs. Following Conventional-Frequency Testing

Stephen A. Fausti*
James A. Henry†
Donna Hayden†
David S. Phillips§
Richard H. Frey†

Abstract
Retrospective analysis of hearing-threshold data from a multisite ototoxicity monitoring study identified an individualized range of predominantly high frequencies (>8 kHz) that appeared to be highly sensitive to early threshold changes caused by ototoxicity. This suggested the potential for a limited-frequency monitoring protocol that could be conducted rapidly without compromising sensitivity to ototoxicity. Such testing would require high-frequency thresholds to be obtained independently, that is, without prior testing at conventional frequencies (0.25–8 kHz). This study was conducted to determine the test–retest reliability of isolated threshold testing in a "target" frequency range of high frequencies (9, 10, 11.2, 12.5, and 14 kHz) that represented a shortened ototoxicity monitoring test. Twenty normal-hearing subjects were evaluated over five sessions. During each session, subjects were tested in each of two conditions: (1) conventional frequencies (0.25–8 kHz) tested first, followed by target frequencies; and (2) target frequencies tested alone (isolation condition). Depending on test frequency, reliability of high-frequency thresholds was either unchanged or improved in the isolation condition. Although these results cannot be generalized to ill hospitalized patients, who may also have pre-existing hearing loss, they lay the groundwork for development of a time-saving limited-frequency test to monitor for ototoxicity in these patients.

Key Words: Hearing thresholds, high frequency, intrasubject reliability, ototoxicity monitoring

Therapeutic treatment with pharmacologic agents having ototoxic potential is common, and the incidence of hearing loss from such treatment is higher than previously realized (Fausti et al, 1992, 1993, 1994). The effects of ototoxicity can be reduced if ototoxicity is detected early in its course and if the information is used to modify the treatment regimen for the purpose of preventing disabling hearing loss (Gandara et al, 1991; Schweitzer, 1993).

Audiometric monitoring of high-frequency (>8 kHz) hearing sensitivity during treatment with ototoxic agents has been shown to be a sensitive method for early detection of ototoxicity (Kopelman et al, 1988; Dreschler et al, 1989; Fausti et al, 1984, 1992, 1993, 1994). As suggested in the national guidelines (ASHA, 1994), ototoxicity monitoring should include both the conventional- (0.25–8 kHz) and high-frequency ranges for optimal sensitivity to ototoxicity. The time involved to conduct such testing, however, can be overly strenuous for many ill patients, and the validity of their responses can be compromised from fatigue. Furthermore, hospitals may be reluctant to adopt a program that is so labor intensive and, therefore, costly. To enable ototoxicity monitoring for all patients who are treated with ototoxic drugs, this laboratory is working toward development of auditory monitoring techniques that can be conducted rapidly.
without compromising sensitivity to early ototoxicity.

Analysis of threshold changes in the Fausti et al. (1992, 1993, 1994) studies revealed that about 90 percent of initial ototoxicity changes occurred in a range of five test frequencies. These five frequencies—generally in the range above 8 kHz—consisted of the five adjacent highest frequencies at which each patient's hearing thresholds were 100 dB SPL or less. These findings suggested that monitoring hearing thresholds only in such a circumscribed high-frequency region would provide a short, effective method for early identification of ototoxicity. If documented as an effective procedure, most of the current functional barriers preventing ototoxicity monitoring from being accepted for routine monitoring should be removed. This documentation first requires that the technique be demonstrated to provide reliable threshold responses in the absence of hearing change, so as to be reasonably certain that any changes seen are the result of a true shift in hearing acuity rather than normal variation of repeated threshold testing.

Several studies have documented the test-retest reliability of high-frequency thresholds (Laukli and Mair, 1985; Fausti et al., 1990; Frank, 1990; Frank and Dreisbach, 1991). In each of these studies, however, high-frequency thresholds were always obtained following testing in the conventional-frequency range. Intuitively, it might seem that repeated threshold testing restricted to the high-frequency range would provide equivalent reliability. However, certain factors could be argued as potentially affecting the reliability, and these factors must be ruled out for such testing to be documented for physicians, medical personnel, and even third-party reimbursers. First, high-frequency pure tones are unfamiliar sounds to most listeners, and testing first at more familiar frequencies might cause a practice effect that biases thresholds at subsequently tested high frequencies. This type of frequency-dependent reliability might also occur within the conventional-frequency range; early studies demonstrated that test-retest reliability of thresholds tends to be best at 1000 Hz (Harris, 1945; Dadson and King, 1952). This established 1000 Hz as the initial frequency to test for an audiogram, and audiologists continue to be trained to begin threshold testing at that frequency. Second, fatigue could be a factor that could cause high-frequency thresholds to be less reliable when obtained following conventional-frequency testing.

Since effective ototoxicity monitoring is critically dependent upon good response reliability, the reliability of testing high-frequency thresholds in isolation, that is, when not preceded by conventional-frequency threshold testing, was evaluated. The present study was conducted in normal-hearing subjects to establish normative reliability of responses obtained in this manner.

**METHOD**

**Subjects**

Subjects were selected on the basis of (1) no known history of otologic disorders; (2) hearing thresholds ≤15 dB HL (ANSI, 1989) for conventional frequencies (0.25–8 kHz); (3) hearing thresholds for high frequencies (9–20 kHz) within 1 standard deviation (SD) of the age-specific means reported in Schechter et al. (1986); (4) middle-ear pressure and compliance within normal limits bilaterally as determined by tympanometry at 226 Hz, 630 Hz, and 1 kHz; and (5) ipsilateral and contralateral acoustic reflexes no greater than 95 dB HL at 0.5, 1, and 2 kHz.

Twenty subjects, ages 21–26, met all inclusion criteria and participated in this study. Four subjects completed all five test sessions in 1 week. For the remaining 16 subjects, all test sessions were completed within 2 weeks.

**Instrumentation**

Subjects were evaluated in a double-walled, sound-attenuated chamber (Model #19701A, Acoustic Systems Inc., Austin, TX). Pure-tone air-conduction thresholds were obtained using a Model 320 audiometer (Virtual Corp., Portland, OR), TDH 50P earphone transducers in MX 41-AR ear cushions were used with conventional frequencies and modified Koss Pro/4X Plus earphone transducers (Fausti et al., 1990) were used with high frequencies. Conventional-frequency earphones were calibrated in compliance with ANSI standards (ANSI, 1989) using a Bruel & Kjaer 2231 sound level meter with a 1/3-octave band filter and a Bruel & Kjaer 4153 artificial ear. The Pro/4X Plus earphones used for high-frequency testing were calibrated on a flat-plate coupler as previously described (Fausti et al., 1979, 1990). Tympanometry and acoustic reflex testing were conducted using a Model 310 (Virtual Corp.) aural acoustic immittance system.
Procedures

Test procedures were conducted over five test sessions, with each session occurring on a separate day. All hearing thresholds were obtained using the modified Hughson-Westlake audiometric test technique (Carhart and Jerger, 1959).

Five test frequencies (9, 10, 11.2, 12.5, and 14 kHz, hereafter referred to as the “target” frequencies) were chosen to represent a typical range of five high frequencies that would be used in a shortened ototoxicity monitoring test. Test procedures were designed to evaluate the test-retest reliability of thresholds at the target frequencies in two conditions: (1) conventional frequencies (0.25-8 kHz) tested first, followed by the target frequencies (“long” test); and (2) target frequencies tested alone (“short” test).

Thresholds in three frequency ranges were evaluated for each subject: conventional frequencies, target frequencies, and frequencies above the target frequencies (16, 18, and 20 kHz). Within each frequency range, testing was done separately in each ear prior to testing in the next frequency range, and frequencies were selected in ascending order. For the first session, threshold testing began in the perceived better ear (or right ear if there was no difference), and for subsequent tests, order of testing between ears was randomized.

For the “long” test, threshold testing was first done as for a conventional audiogram. That is, thresholds were obtained at 1 kHz followed by 0.25, 0.5, 1, 2, 3, 4, 6, and 8 kHz. Testing was then done at the five target frequencies, followed by the frequencies above the target range. For the “short” test, subjects were first tested at the five target frequencies, followed by the higher frequencies, then the conventional frequencies. Although the entire frequency range was tested twice during each session, the testing order of frequency ranges established the effect of evaluating the target frequencies either in isolation or following conventional frequencies.

For session 1, subjects first received an otoscopic examination, immittance testing, and speech audiometry. These evaluations were conducted only to confirm normal auditory functioning and were not repeated. Subjects were then tested using the “long” test paradigm. This was followed by an approximately 15-minute break and then testing with the “short” procedure. For the next four sessions, the order of “long” and “short” tests was alternated (Table 1). Each subject was randomly assigned to one of two groups. Group 1 was tested during session 2 as at the baseline session (i.e., “long” test followed by a break, then the “short” test). For group 2, the sequencing of tests was reversed at session 2 (“short” test followed by a break, then the “long” test). For the remaining three test sessions, testing order was alternated between the two testing procedures so that each subject started two of the last four sessions with each procedure.

RESULTS

Testing for Order Effect

Within each of the 5 days of testing, hearing thresholds were obtained twice. There was thus the potential for an effect of testing order to occur that could bias the results. To test for this effect, each hearing threshold was grouped according to “test 1” and “test 2,” and t-tests were conducted at each of the five test frequencies. None of these t-tests were significant (all p’s > .05), indicating that an effect of testing order did not occur.

Mean Thresholds across Subjects

In Table 2, mean thresholds for all ears are shown separated by the variables frequency (9, 10, 11.2, 12.5, and 14 kHz), group (1 and 2), day (1, 2, 3, 4, and 5), and test (“long” and “short”). Variability between frequencies was due to the normal progressive decrease in auditory sensitivity at higher frequencies. The remaining variables were tested using a 1 × 2 design ANOVA,
Table 2  Mean (across Subjects) of Hearing Thresholds (in dB SPL) at the Five Test Frequencies, Divided by Group* (1, 2), Frequency (9, 10, 11.2, 12.5, 14 kHz), Day (D1, D2, D3, D4, D5), and Test ("Long," "Short")

<table>
<thead>
<tr>
<th>kHz</th>
<th>Group</th>
<th>D1</th>
<th>D2</th>
<th>D3</th>
<th>D4</th>
<th>D5</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>1</td>
<td>22.5</td>
<td>21.8</td>
<td>20.9</td>
<td>21.6</td>
<td>20.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>20.3</td>
<td>20.8</td>
<td>20.3</td>
<td>20.0</td>
<td>21.7</td>
</tr>
<tr>
<td>10</td>
<td>1</td>
<td>19.8</td>
<td>19.1</td>
<td>18.6</td>
<td>20.0</td>
<td>18.6</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>19.4</td>
<td>21.1</td>
<td>20.0</td>
<td>19.4</td>
<td>20.8</td>
</tr>
<tr>
<td>11.2</td>
<td>1</td>
<td>24.3</td>
<td>24.1</td>
<td>22.7</td>
<td>23.9</td>
<td>23.0</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>22.2</td>
<td>23.1</td>
<td>22.2</td>
<td>21.1</td>
<td>22.2</td>
</tr>
<tr>
<td>12.5</td>
<td>1</td>
<td>28.0</td>
<td>27.7</td>
<td>27.3</td>
<td>27.3</td>
<td>29.1</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>30.6</td>
<td>31.4</td>
<td>31.9</td>
<td>32.8</td>
<td>31.7</td>
</tr>
<tr>
<td>14</td>
<td>1</td>
<td>40.0</td>
<td>39.1</td>
<td>39.1</td>
<td>40.0</td>
<td>40.5</td>
</tr>
<tr>
<td></td>
<td>2</td>
<td>37.2</td>
<td>38.9</td>
<td>37.5</td>
<td>37.2</td>
<td>35.8</td>
</tr>
</tbody>
</table>

*N = 22 ears for group 1; N = 18 ears for group 2.

with one independent factor (group) and two repeated factors (day and test). The 1 x 2 ANOVA was done at each of the five test frequencies. Results of these analyses revealed that none of the main effects or interactions was significant (p > .05), indicating that there were no differences between groups, tests, or test days.

Intrasubject and Intertest Variability of Thresholds

The group means shown in Table 2 reflect the overall consistency of the thresholds averaged across subjects. Group means, however, do not specifically address within-subject variability of measures, which was the primary focus of this study. Each subject was tested with both the "short" and the "long" tests during each of five testing sessions. To evaluate within-subject variability of the thresholds, SDs were calculated across the five test sessions. These were calculated for each of the two testing protocols ("long" and "short") and at each of the five test frequencies. The means, across subjects, of these SDs are shown in Table 3. It can be seen that, comparing the mean SDs between "short" and "long" conditions, the "short" mean SDs were smaller than the "long" mean SDs for all of the five comparisons. Thus, the mean threshold variability was empirically less for the short tests than for the long tests. To evaluate this trend statistically, t-tests were done to compare the mean SDs between the two test types. The p values resulting from the t-tests are also shown in Table 3. Three of the five p values were significant (p < .05), indicating that the threshold reliability was significantly better with the "short" test than with the "long" test at three of the five target frequencies.

DISCUSSION

The purpose of this study was to document the test–retest reliability of testing in the high-frequency range in isolation, that is, without preceding high-frequency testing with conventional-frequency testing. Although this particular topic has not previously been explored, the reliability of testing high-frequency thresholds has been investigated. Both Fletcher (1965) and Rosen et al (1964) concluded that, although there were difficulties in testing thresholds above 8 kHz due to inadequate equipment available at that time, the thresholds obtained were repeatable. Stelmachowicz et al (1989a, b), Frank (1990), and Frank and Dreisbach (1991) found thresholds for frequencies above 8 kHz to be within the clinically acceptable range of ±10 dB in 94 percent to 98 percent of the repeated thresholds. The present study also found that

Table 3  Means (across Subjects) of Standard Deviations at Target High Frequencies, and p Values for t-tests at Each Frequency

<table>
<thead>
<tr>
<th>Frequency (kHz)</th>
<th>&quot;Long&quot;</th>
<th>&quot;Short&quot;</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>9</td>
<td>3.206</td>
<td>2.739</td>
<td>.0071*</td>
</tr>
<tr>
<td>10</td>
<td>3.127</td>
<td>2.690</td>
<td>.0019</td>
</tr>
<tr>
<td>11.2</td>
<td>3.242</td>
<td>2.511</td>
<td>.0087*</td>
</tr>
<tr>
<td>12.5</td>
<td>3.333</td>
<td>2.822</td>
<td>.0545</td>
</tr>
<tr>
<td>14</td>
<td>4.048</td>
<td>3.365</td>
<td>.0431*</td>
</tr>
</tbody>
</table>

*Indicates significant difference (p < .05).
98 percent of the intertest thresholds varied by no more than ±10 dB.

It is important to distinguish between intra-subject and intersubject reliability of high-frequency threshold data. Good intrasubject reliability of high-frequency thresholds has been documented in this and other studies, including those mentioned above. Intersubject variability has consistently been shown to be large (Osterhammel and Osterhammel, 1979; Schechter et al, 1986; Green et al, 1987; Stelmachowicz et al, 1988, 1989a, b). Schechter et al (1986) studied high-frequency threshold sensitivity in subjects ranging from 6 to 30 years of age. Most of the younger subjects in this study had sensitive high-frequency hearing, but there was variable degeneration of sensitivity into young adulthood, as indicated by increased SDs of mean thresholds at the higher frequencies. This high intersubject variability increases to a maximum between 12 and 16 kHz, with a marked decrease at the highest frequencies (Schechter et al, 1986; Stelmachowicz et al, 1989a). Stelmachowicz et al (1989b) explained this pattern by showing that SDs of thresholds increased as a function of the absolute threshold. That is, as mean thresholds increased to about 70 to 80 dB SPL (as with the older age groups), SDs increased to a maximum of about 30 dB. At mean thresholds above and below 70 to 80 dB SPL, SDs decreased proportionally at all frequencies. This large intersubject variability of high-frequency thresholds, even in young adults, has precluded the emergence of normative reference values for high-frequency thresholds. However, the value of high-frequency threshold testing lies in the ability to obtain thresholds with good intra-subject reliability, which is the basis for monitoring high-frequency sensitivity to detect the ototoxic process.

The results of this study should be highly relevant to the auditory evaluation of patients receiving ototoxic agents. Many of these patients are often too ill, unable, or unwilling to tolerate a complete-frequency hearing evaluation in both ears (up to 17 frequencies per ear when the frequency range of 0.25–20 kHz is tested), which can require 20 to 30 minutes per ear (Fausti et al, 1994). In studies assessing changes in hearing across all frequencies during therapy with ototoxic drugs, approximately 90 percent of all patients revealed initial change within a range of five frequencies (Fausti et al, 1992, 1993, 1994). For each individual, this range includes the highest frequency with a threshold no greater than 100 dB SPL and the next four consecutively tested lower frequencies. If the independent testing of a small range of high frequencies can be conducted reliably, it is expected that initial hearing loss due to exposure to ototoxic medication would be effectively detected in at least 90 percent of ears showing change.

The present results demonstrate that testing only in a narrow range of high frequencies in a normal-hearing adult population produces reliable threshold data. Unexpectedly, reliability was actually improved relative to testing high frequencies following conventional frequencies. This could potentially be due to improved alertness/attention during an initial period of testing. Patients often have difficulty staying awake and/or maintaining focus during extended threshold testing, and this may be a particular issue when high frequencies are tested following a complete conventional audiogram.

It is unknown whether these results, obtained using normal-hearing individuals, can be generalized to the population for which this shortened methodology is intended. Hospitalized patients receiving therapeutic treatment with ototoxic drugs are generally ill and fatigued, are usually older than the subjects in the present study, and often present with reduced hearing sensitivity. Research in this laboratory is therefore currently being conducted to determine whether testing in this manner with hospitalized patients will also produce reliable results.

With the normal-hearing, healthy subjects for this study, single-ear pure-tone testing required approximately 5 minutes to test the five target frequencies between 9 and 14 kHz. One would expect testing time to be slightly greater with ill patients who may have difficulty with the performance of this task. The time of testing for five target frequencies is thus about one-third of the time that would be required for complete-frequency testing (0.25–20 kHz). Testing a shortened range of frequencies would also reduce the time spent in a patient’s room when ward testing is required. This would be a significant advantage because bedside testing time is inevitably lengthened due to hospital staff entering and leaving the room, especially a room that contains two to four patients.

A shortened testing method would facilitate clinical monitoring of ototoxicity. Audiologists who administer ototoxicity monitoring programs could more easily integrate these patients into their daily schedules, since less testing time would be required. This, in turn, would increase the accessibility of monitoring to these patients.
patients, allowing a greater percentage of them to be followed. The results of this added accessibility could reduce the number of patients suffering debilitating hearing loss from ototoxicity.

Acknowledgment. Funding for this study was provided by the Department of Veterans Affairs, Rehabilitation Research and Development (RR&D 0227).

REFERENCES


