Masked High-Frequency Bone-Conduction Audiometry: Test Reliability

John C. McDermott
Stephen A. Fausti
James A. Henry
Richard H. Frey

Abstract

The present study examines the reliability of masked high-frequency bone-conduction threshold measurements in 95 normal-hearing subjects. High-frequency pure-tone air- and bone-conduction thresholds were measured with a dedicated laboratory high-frequency auditory evaluation system using matched, modified Koss Pro/4X Plus earphones, and the Pracitronic KH 70/5 bone vibrator. A 400-Hz wide band masking noise centered at the frequency of the test tone was used to mask the nontest ear. Monaural masked bone-conduction threshold measurements were obtained at the ipsilateral mastoid of the ear with better high-frequency hearing. Two measurements were performed in each session, and each subject participated in two sessions. In several comparisons for test-retest consistency, high-frequency bone-conduction threshold measurements were as repeatable as air-conduction thresholds of identical frequency, or bone-conduction thresholds for frequencies of 4 kHz and less. High-frequency bone-conduction threshold measurement appears to be a sufficiently reliable tool for diagnosis of auditory disorders.

Key Words: Bone-conduction thresholds, high frequency, reliability, threshold force levels

Standard pure-tone audiometry has long remained on the frontline of auditory assessment because this procedure has a high intrasubject consistency. A measurement of this consistency, the standard error of the estimated threshold, is considered to be approximately 5 dB for both standard air- and bone-conduction threshold measurements (Carhart and Hayes, 1949; Jerger, 1962). It is expected that, in the absence of organic or nonorganic change between tests, test-retest reliability of pure-tone thresholds will be within ±5 dB.

High-frequency auditory threshold measurement has been advanced as a useful addition to standard pure-tone audiometry because of its utility in the diagnosis of auditory disorders induced by excessive noise (Fausti et al, 1981), ototoxic agents (Fausti et al, 1984a, b), recurrent middle-ear disease (Ahonen and McDermott, 1984; Ising et al, 1986; McDermott et al, 1986), and iatrogenic effects of surgery for chronic ear disease (Tos et al, 1984; Mair and Laukli, 1986; Domenech et al, 1989; Mair et al, 1989).

Thresholds for frequencies above 8 kHz (4 kHz in bone-conduction threshold tests) are not standardized. Equivalent high-frequency thresholds for both sound pressure level and force level have been examined, but neither measure is near acceptance for standardization because of unique problems that do not exist for standard pure-tone audiometry.

Research of high-frequency bone-conduction hearing threshold measurement has been limited due, in part, to difficulties in the development of stable, linear transducers for these higher frequencies. Richter and Frank (1985) compared five existing bone vibrators for capability to assess high-frequency thresholds. They observed that the Pracitronic KH 70/5 bone vibrator had the flattest frequency response for pure-tone frequencies above 6 kHz. It was among
those vibrators that required the least amount of input voltage and was most consistent across repeated calibration measurements.

Repeatability of unmasked high-frequency bone-conduction measurements made within a test session and among five test sessions was examined for 1.0 through 16 kHz (Frank and Ragland, 1987). These researchers showed that, across frequencies, mean intra-session threshold differences ranged from -1.7 to +1.9 dB; mean inter-session threshold differences ranged from 0.5 to 3.0 dB. Because interaural attenuation for bone-conduction approaches zero, we have examined the differences between contralateral masked and unmasked high-frequency bone-conduction thresholds in the same subjects. Our results demonstrated that masked high-frequency bone-conduction thresholds were 1.5 to 3.4 dB poorer than the unmasked thresholds and that these differences were statistically significant at 0.05 level of confidence, except at 12 kHz (McDermott et al., 1990). This contralateral masking procedure has been incorporated in our studies of high-frequency bone conduction. The present report describes an investigation of the reliability of masked high-frequency bone-conduction threshold measurements within and between sessions in a large group of normal-hearing subjects.

**METHOD**

**Subjects**

Ninety-five subjects, age 11 through 49 years, participated in this study. The study sample consisted of 68 female and 27 male subjects. All subjects had a negative history of otologic disorders, particularly middle-ear disease, with crossed and uncrossed acoustic reflexes present at 105 dB or less at 0.5, 1, and 2 kHz, and pure-tone thresholds of 15 dB HL (ANSI, 1989) or less for octave frequencies between 0.25 to 8 kHz. For frequencies 9 through 16 kHz, each subject had air-conduction thresholds within one standard deviation of the mean threshold observed in the age-specific data reported for subjects of a comparable age range (Schechter et al., 1986).

**Instrumentation**

Subjects were evaluated in a double-wall, sound attenuated suite (Acoustic Systems, Model 19701A) that meets or exceeds the standards for evaluation of hearing in an uncovered ear (ANSI, 1977). Conventional pure-tone air-conduction audiometry (0.25 to 8 kHz) was assessed using a clinical audiometer (Grason-Stadler [G-S], 1701) with matched earphones (Telephonics, TDH-49) in MX-41/AR ear cushions. Prior to each test session, calibration was performed in a manner consistent with ANSI specifications (ANSI, 1989) using a 6 cm³ hard wall coupler (Brüel & Kjaer [B & K], model 4153), a 1.27 cm condenser microphone (B & K, model 4134), and a precision sound level meter (B & K, model 2231).

High-frequency (8 to 12 kHz) pure-tone air-conduction thresholds were obtained with a dedicated high-frequency auditory evaluation system (Portland Auditory Research Veterans Affairs, High-Frequency [PARVA-HF]) using matched, modified Koss Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones. (Placement of foam Pro/4X Plus earphones.) Calibration of the PARVA-HF was performed prior to each test session. A description of the PARVA-HF and calibration procedure using a flat-plate coupler has been presented elsewhere (Fausti et al., 1979; Fausti et al., 1990). The modified Koss Pro/4X Plus earphones were used to present all high-frequency stimuli, including the narrow-band masking noise.

High-frequency bone-conduction thresholds were obtained with a Pracitronic KH 70/5 bone vibrator driven by the PARVA-HF. Prior to each test session, the entire stimulus generating system, including the bone vibrator, was calibrated using a mechanical coupler (B & K, model 4930) at 23°C (± 1°C) and voltmeter (Tektronix, model TM503). The B & K mechanical coupler characteristically has a markedly resonant peak at 17 kHz in the force sensitivity response. Reliable calibration measurements using the B & K mechanical coupler above approximately 12 kHz are compromised by this resonant peak and its sharply sloping response skirt. The force sensitivity response is very flat over the frequency range of 5 through 12 kHz. Thus, high-frequency threshold force levels, for this study, were restricted to a frequency range of 6 through 12 kHz.

The terminal output voltage, system output voltage, and linearity of the KH 70/5 bone vibrator output were monitored regularly. Output force levels of the KH 70/5 bone vibrator with constant voltage at terminals remained at ± 1.5 dB throughout the course of the study. Output of the stimulus generating system was limited to 10 volts rms at the input terminal to
the bone vibrator to ensure output linearity. Subharmonic force output was not observed at -80 dB for output near maximum power for all test frequencies. Second and third harmonic distortion was consistently -30 dB or lower throughout the study.

Procedure

Air-conduction thresholds were obtained bilaterally in 2-dB ascending increments for octave frequencies from 0.25 to 8 kHz and in octave frequencies 3 and 6 kHz. High-frequency air-conduction thresholds were obtained at 1.0-kHz intervals from 8 through 12 kHz (8, 9, 10, 11, 12 kHz). The ear with better high-frequency air-conduction sensitivity was selected for masked bone-conduction threshold measurements. Bone-conduction threshold measurements, using the KH 70/5 bone vibrator, were obtained at 6, 8, 9, 10, 11, and 12 kHz. Stimulus intensity increment was 2 dB. Presentation mode was a pulsed tone of either 200-msec (for the GS-1701) or 300-msec (for the PARVA-HF) duration with a 25-msec linear rise-decay time on a 50 percent duty cycle. A modified method-of-limits procedure was used to establish both air- and bone-conduction thresholds (Carhart and Jerger, 1959).

All high-frequency bone-conduction thresholds were obtained with a mastoid placement. As recommended in ANSI standards (1981), a compressible hearing protector (E-A-R) was tightly inserted into the ear canal of the test ear to reduce the effects of acoustic radiation from bone vibrators that has been observed for frequencies greater than 2 kHz (Shipton et al, 1980; Frank and Holmes, 1981). The contralateral ear was masked with high-frequency, narrow-band noise, 400-Hz wide, centered at the test frequency (Fausti et al, 1982) and presented at 30 dB SL (re: individual subject’s threshold for the masking stimulus).

Each subject participated in two bone-conduction threshold test sessions. During each test session, two separate and complete trials were performed for high-frequency bone-conduction threshold measurements. The KH 70/5 bone vibrator was placed on the bony mastoid prominence of the test ear, and the contralateral ear was masked with narrow-band masking noise. The KH 70/5 bone vibrator was removed and replaced between trials. Sessions were separated by at least 1 week and by as much as several months. The entire test battery required about 2 hours, and frequent rest periods were provided to minimize subject fatigue.

RESULTS

At each test frequency, a paired comparison was performed for high-frequency air-conduction thresholds. Table 1 displays means and standard deviations for high-frequency air-conduction thresholds for the two test sessions. At each test frequency, differences between the two high-frequency air-conduction threshold measurements were not statistically significant at the 0.05 confidence level.

Among the 95 subjects, monaural bone-conduction threshold measurements were performed on 41 left ears and 54 right ears. Subjects were grouped according to the side of the tested ear. Mean thresholds were determined at each frequency for each group. Difference between group means were calculated at each frequency and ranged from 1.59 to -2.33 dB. As these mean differences were small, for the purpose of this investigation, left and right ear threshold data were pooled for all further analyses.

Table 2 displays the means and standard deviations for high-frequency bone-conduction thresholds for all four test trials in the 95 subjects. The largest difference between mean bone-conduction thresholds among trials was 1.5 dB at 12 kHz. A single-factor ANOVA with four repeated measures was performed for high-frequency bone-conduction thresholds at each test frequency. Differences among the four high-frequency bone-conduction threshold means were not statistically significant at the 0.05 level of confidence (F[3,94 df] ranged from 0.03 to 2.28).

---

Table 1: Means and Standard Deviations for High-Frequency Air-Conduction Thresholds in dB SPL (re: 20 μPa) for 95 Normal Hearing Subjects

<table>
<thead>
<tr>
<th>Frequency (in kHz)</th>
<th>Session 1 Mean</th>
<th>SD</th>
<th>Session 2 Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>6</td>
<td>18.8</td>
<td>9.7</td>
<td>17.9</td>
<td>8.8</td>
</tr>
<tr>
<td>8</td>
<td>21.1</td>
<td>8.3</td>
<td>20.1</td>
<td>8.7</td>
</tr>
<tr>
<td>9</td>
<td>25.8</td>
<td>9.5</td>
<td>25.4</td>
<td>10.5</td>
</tr>
<tr>
<td>10</td>
<td>26.1</td>
<td>12.3</td>
<td>25.7</td>
<td>11.3</td>
</tr>
<tr>
<td>11</td>
<td>33.0</td>
<td>13.3</td>
<td>32.2</td>
<td>12.3</td>
</tr>
<tr>
<td>12</td>
<td>33.1</td>
<td>13.3</td>
<td>31.7</td>
<td>12.9</td>
</tr>
</tbody>
</table>

*No significant differences in mean air-conduction thresholds were found between two test sessions at any test frequency at the 0.05 level of confidence (N = 95).
Table 2 Means and Standard Deviations for Masked High-Frequency Bone-Conduction Thresholds in dB (re: 1 μNewton) for 95 Normal Hearing Subjects

<table>
<thead>
<tr>
<th>Frequency (in kHz)</th>
<th>Session 1</th>
<th>Session 2</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Trial 1</td>
<td>Trial 2</td>
</tr>
<tr>
<td>6</td>
<td>35.1</td>
<td>35.2</td>
</tr>
<tr>
<td></td>
<td>(6.5)</td>
<td>(6.9)</td>
</tr>
<tr>
<td>8</td>
<td>35.8</td>
<td>35.1</td>
</tr>
<tr>
<td></td>
<td>(6.7)</td>
<td>(7.0)</td>
</tr>
<tr>
<td>9</td>
<td>36.9</td>
<td>36.0</td>
</tr>
<tr>
<td></td>
<td>(7.7)</td>
<td>(8.7)</td>
</tr>
<tr>
<td>10</td>
<td>38.7</td>
<td>38.1</td>
</tr>
<tr>
<td></td>
<td>(7.3)</td>
<td>(8.1)</td>
</tr>
<tr>
<td>11</td>
<td>38.2</td>
<td>38.4</td>
</tr>
<tr>
<td></td>
<td>(9.0)</td>
<td>(9.5)</td>
</tr>
<tr>
<td>12</td>
<td>38.6</td>
<td>38.4</td>
</tr>
<tr>
<td></td>
<td>(10.1)</td>
<td>(10.5)</td>
</tr>
</tbody>
</table>

*No significant differences in bone-conduction thresholds were found among test trials at any test frequency at the 0.05 level of confidence (N = 95).

The intersession repeatability of bone-conduction thresholds was examined. Figure 1 illustrates the cumulative frequency distribution of intrasubject differences between the first trial of sessions 1 and 2 for test frequencies 6 through 12 kHz. The number of data points falling within the ±10 dB range represent 88.8 percent of all intrasubject intersession comparisons for all test frequencies. The number of comparisons falling outside the ±10 dB range increases as a function of frequency. At 6 kHz only 3 of the 95 comparisons exceed ±10 dB, while at 12 kHz, 21 comparisons of subject’s intersession thresholds exceed this range. Across frequencies, between 61 and 73 percent of all intrasubject intersession comparisons for bone-conduction thresholds were within ±5 dB. The greatest difference was noted in a subject for whom a 25-dB difference existed at 11 kHz. Similar comparisons were noted for intrasession repeatability for high-frequency, bone-conduction thresholds.

The difference and the absolute-value-of-the-difference between the first trial in the two test sessions for each subject were computed. Means and standard deviations for the actual and absolute-value-of-the-differences are displayed in Tables 3 and 4.

DISCUSSION

The reliability of masked high-frequency bone-conduction thresholds in a large group of normal-hearing subjects was investigated. From the results above, masked thresholds appeared to be as reliable as unmasked bone-conduction thresholds reported for a comparable frequency range (Frank and Ragland, 1987). The intrasession and intersession consistency for high-frequency bone-conduction thresholds satisfies a requirement of ±5 dB in 61 to 73 percent of all threshold comparisons.

The difference between test-retest thresholds can reveal a tendency among subjects for thresholds to be significantly different in either session. The absolute-value-of-the-difference reflects the magnitude of the discrepancy between test-retest measurements, regardless of the direction of the difference. Difference scores are considered another measure of agreement between test sessions and allow comparisons of...
group mean difference scores and standard deviations with previous research on reliability of pure-tone threshold measurements. As seen in Table 3, the means of difference scores for high-frequency bone-conduction thresholds were very close to zero for each test frequency. The largest mean difference score was 1.5 dB at 11 kHz. Standard deviations ranged from 3.5 to 6.1 dB. Carhart and Hayes (1949) systematically examined the reliability of standard audiometric bone-conduction thresholds and reported mean difference scores of -0.25 to +0.55 dB, with standard deviations of 7.65 to 9.05 dB. A comparison of Carhart and Hayes’ study of standard bone-conduction and the present study of high-frequency bone-conduction audiometric measurements supports the conclusion that bone-conduction threshold measurements from 6 through 12 kHz are as repeatable as measurements made for bone-conduction thresholds for frequencies of 4 kHz and less.

In the present study of reliability for high-frequency threshold measurements, the means of the absolute-value-of-differences, reported in Table 4, ranged between 2.5 and 4.4 dB for bone conduction. The means of the absolute-value-of-differences ranged between 3.1 and 5.9 dB for air conduction. The magnitudes of the test-retest differences for high-frequency air- and bone-conduction thresholds appear to be comparable.

Jerger (1962) considers absolute consistency of pure-tone threshold measurements more important in pure-tone audiometry than relative consistency. Absolute consistency, as defined by Jerger, is “the absolute variability in performance from test to test” and is expressed as the standard error of measurement. Absolute consistency is concerned with the precision with which a test instrument predicts the criterion measure across repeated applications. On the other hand, the relative consistency of the measure, as expressed in the coefficient of correlation between test-retest threshold measurements, is not relevant if the test-retest difference is inordinately large and variable. For standard pure-tone bone-conduction thresholds reported, Jerger observed a high correlation coefficient of 0.75 to 0.96, and standard error of measurement of 4.2 to 6.0 dB. In the present study, the coefficient of reliability ranged from 0.74 to 0.84, and the standard error of measurement was calculated at 2.5 to 4.3 dB. A recent study of high-frequency air-conduction thresholds (Stelmachowicz et al., 1989) revealed standard errors of 2.5 through 3.8 dB for comparable type earphones using a Bekesy tracking procedure. Both the standard error and coefficient of correlation calculated for the present study compared favorably with those values observed in earlier studies of reliability for pure-tone threshold measurements.

A final observation is made that deserves further study: the sensitivity of auditory function in dB SPL reported in Table 1 shows a positive slope of 14 dB over a range of 6 through 12 kHz, while for threshold in dB force level displayed in Table 2 the function shows a slope of only 4 dB over the same octave range. This difference in auditory sensitivity for sound pressure level and force level in the same ears exemplifies the differences between the transfer function of the air-conduction pathways and the bone-conduction pathways. Studies are underway that will explore the amplitude and phase relationships between high-frequency air- and bone-conduction threshold and suprathreshold responses using cancellation techniques developed by Békésy (1932) and revised recently by Levitt (1987).

**SUMMARY**

Utility of high-frequency bone-conduction audiometry depends upon the degree of
reliability and validity of the measurement tool. In comparisons of various measures for test consistency, the test-retest reliability of high-frequency bone-conduction threshold measurements favorably compared with test-retest reliability of both standard bone-conduction and high-frequency air-conduction audiometry. This measurement, then, is sufficiently precise to be considered for utilization as a clinical tool in auditory diagnostic assessment. Further studies are needed to validate the measurement as an assessment of only the high-frequency region of the cochlea, before examining its accuracy in the differential diagnosis of various auditory disorders, such as the effects of middle-ear disease and of surgeries for middle ear and inner ear dysfunction.

Acknowledgments. This investigation was supported by Merit Review Award #335 from the Medical Research Service, U.S. Department of Veterans Affairs. Thanks are due to Deanna Olson for her assistance in the collection of data for this study, and an anonymous editorial reviewer who provided cogent suggestions for improving this report.

REFERENCES


