Pure-Tone Audiometric Threshold Test-Retest Variability in Young and Elderly Adults

Variabilité test-retest du seuil audiométrique tonal chez les jeunes adultes et les aînés

Janine A. Landry, MSc and Walter B. Green, PhD
Dalhousie University
Halifax, Nova Scotia

ABSTRACT

Audiometric threshold test-retest variability was studied in three age groups: young adults aged 22-34 years, older adults aged 50-63 years, and seniors aged 65-81 years. Audiometric thresholds were tested at six frequencies (250, 500, 1000, 2000, 4000, and 8000 Hz) utilizing supraaural and insert transducers. No statistically significant differences were found in test-retest differences at 250, 500, and 1000 Hz as a function of age group or transducer (p > .05). Statistically significant group, transducer, and group by transducer effects were found at 2000 and 8000 Hz, 2000 Hz, and 4000 Hz (p < .05), respectively. These results were speculated to be due to changes in the aging ear canal structure combined with inherent variability in the transducer coupling characteristics. The outcomes of this study suggest caution should be exercised when interpreting audiometric threshold test-retest changes in adult listeners.

KEY WORDS: pure-tone threshold test-retest variability • earphones • elderly

Pure-tone threshold audiometry is one of the foundations of audiological assessment. It is the standard procedure used in determining auditory sensitivity and plays an integral role in diagnostic interpretation and in the planning of (re)habilitative programs. Inherent assumptions underlying pure-tone audiometry are that persons provide an appropriate response when a stimulus is audible and that responses are repeatable upon retesting. Numerous studies have suggested that both behavioural and physiological changes accompany aging (Cobb, Jacobson, Newman, Kretschmer, & Donnell, 1993; Davis, Onni, & Paving, 1990; Lutman, 1990; Quaranta, Salonna, & Longo, 1990; Willott, 1996). Further, these changes may compromise an elderly person’s auditory sensitivity and their ability to provide appropriate responses in a consistent fashion over time. Schuknecht (1974) noted specific histopathologic and morphologic changes in the aging peripheral auditory system. The changes include hair cell damage, atrophic changes in the stria vascularis, mechanical alterations in the cochlear duct, and the loss of spiral ganglion cells and/or damage to cochlear neurons. Central nervous system changes may include loss of myelin, hypertrophy of the internal auditory meatus (Grimes, 1995), reduced neuron counts (Brody, 1955), circulatory problems (Hinchcliffe, 1990; Kasten & McCrosky, 1982) as well as other factors such as cerebral atrophy, reduced dendritic branching, and decreased effectiveness of neurotransmitter substances (Kaufman, 1994). Behavioural variables which may affect an elderly person’s ability to provide consistent audiometric responses include the effect of physical discomfort, antagonism toward the task (Yantis, 1994), and difficulty in attending to the required task (Green, 1972). Mauer and Rupp (1979) considered the elderly to be “difficult to test” and recommended that standard audiometric test procedures be modified to ensure valid findings. Presumably, all the above age-related histopathologic and morphologic changes cited above could contribute to the observation that hearing sensitivity decreases with age. The behavioural aspects, on the other hand, would seem to effect the reliability of audiometric findings. In addition, the repeatability of audiometric test results should also vary due to...
the unreliability of audiometric earphones. That is, irrespective of the age of the listener, there is variability associated with the coupling of the stapes transducer (i.e., supraaural < insert earphones) to the auditory mechanism (Zwislocki et al., 1988). Audiometric test-retest variability among elderly listeners should, therefore, be dependent upon changes in the aging auditory system, behavioral variables, and inherent unreliability of audiometric earphones.

To date, to the best of our knowledge, no studies have examined pure-tone audiometric threshold test-retest variability in elderly listeners. The purpose of the present study was, therefore, to examine pure-tone audiometric threshold test-retest variability in elderly listeners using both supraaural and insert earphone transducers.
Table 1. Means (and Standard Deviations) of Signed Test-Retest Audiometric Threshold Differences (dB) as a Function of Age Group (young, old, oldest), Transducer, and Test Frequency.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Supraaural Inert</th>
<th>Frequency (Hz)</th>
<th>Young</th>
<th>Old</th>
<th>Oldest</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>1.0</td>
<td>0.5</td>
<td>0.6</td>
<td>0.05</td>
<td>0.25</td>
</tr>
<tr>
<td></td>
<td>(2.0)</td>
<td>(3.0)</td>
<td>(4.0)</td>
<td>(5.0)</td>
<td>(6.0)</td>
</tr>
<tr>
<td>500</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(3.5)</td>
<td>(4.5)</td>
<td>(5.5)</td>
<td>(6.5)</td>
</tr>
<tr>
<td>1000</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
<td>1.5</td>
</tr>
<tr>
<td></td>
<td>(2.5)</td>
<td>(3.5)</td>
<td>(4.5)</td>
<td>(5.5)</td>
<td>(6.5)</td>
</tr>
</tbody>
</table>

Note: The young group was comprised of 20 participants 22-34 years old; the old group consisted of 10 participants 50-63 years old; and the oldest group included 10 participants 65-81 years old.

In order to examine the significant between-group effect at 2000 Hz (see Figure 1), separate Scheffe pair-wise comparisons were undertaken for each transducer. These results are shown in Table 3. For the supraural transducer, a significant difference was found between the young group and the oldest group (p < .05). All other pair-wise comparisons were nonsignificant (p > .05).

At 4000 Hz, the significant interaction of group by transducer (see Figure 2) was examined with single-df comparisons and Scheffe pair-wise comparisons (see Table 4). Three single-df comparisons were used to examine within-group differences in test-retest threshold differences between transducers (see Table 4). Significantly higher test-retest differences were found for the oldest group with the insert earphone (p < .05). There were no significant differences between transducers for the young and old groups (p > .05).

Between-group effects were also examined with Scheffe pair-wise comparisons at 4000 Hz for each transducer (see Table 5). A significant difference was found between the test-retest difference in thresholds with the insert earphone between the oldest group and the other young and old groups (p < .05). All other pair-wise comparisons were nonsignificant (p > .05).

The significant main effect of group at 8000 Hz (see Figure 3) was examined with Scheffe pair-wise comparisons (see Table 6). The analysis revealed that the oldest group of listeners had significantly greater test-retest variability than the other two groups.

Table 2. Two-Way ANOVA for Signed Test-Retest Audiometric Threshold Differences by Age Group and Transducer as a Function of Frequency.

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Transducer</th>
<th>Age Group</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>df</th>
<th>F</th>
<th>p</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>Supraaural</td>
<td>Young</td>
<td>15</td>
<td>1.37</td>
<td>0.20</td>
<td>30</td>
<td>37</td>
<td>0.26</td>
<td>23</td>
<td>37</td>
<td>0.43</td>
</tr>
<tr>
<td></td>
<td>Inert</td>
<td>Old</td>
<td>15</td>
<td>0.86</td>
<td>0.26</td>
<td>30</td>
<td>37</td>
<td>0.03</td>
<td>23</td>
<td>37</td>
<td>0.36</td>
</tr>
<tr>
<td>500</td>
<td>Supraaural</td>
<td>Young</td>
<td>15</td>
<td>1.37</td>
<td>0.25</td>
<td>30</td>
<td>37</td>
<td>0.25</td>
<td>23</td>
<td>37</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Inert</td>
<td>Old</td>
<td>15</td>
<td>0.86</td>
<td>0.25</td>
<td>30</td>
<td>37</td>
<td>0.03</td>
<td>23</td>
<td>37</td>
<td>0.36</td>
</tr>
<tr>
<td>1000</td>
<td>Supraaural</td>
<td>Young</td>
<td>15</td>
<td>1.37</td>
<td>0.25</td>
<td>30</td>
<td>37</td>
<td>0.25</td>
<td>23</td>
<td>37</td>
<td>0.35</td>
</tr>
<tr>
<td></td>
<td>Inert</td>
<td>Old</td>
<td>15</td>
<td>0.86</td>
<td>0.25</td>
<td>30</td>
<td>37</td>
<td>0.03</td>
<td>23</td>
<td>37</td>
<td>0.36</td>
</tr>
</tbody>
</table>

Note: Considered significant at p < .05.
groups (p < .05). There was no significant difference in test-retest audiometric threshold differences between the young and old group of listeners (p > .05).

Table 3: Scheffe Pair-Wise Comparisons of Between-Group Mean Test-Retest Auditory Threshold Differences as a Function of Transducer at 2000 Hz

<table>
<thead>
<tr>
<th>Transducer</th>
<th>Groups Comparison</th>
<th>Mean Difference</th>
<th>Critical Difference</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supranaural</td>
<td>Young vs. Old</td>
<td>2.5</td>
<td>4.4</td>
<td>.20</td>
</tr>
<tr>
<td></td>
<td>Young vs. Oldest</td>
<td>8.0</td>
<td>4.4</td>
<td>.0003*</td>
</tr>
<tr>
<td></td>
<td>Old vs. Oldest</td>
<td>5.5</td>
<td>1.1</td>
<td>.23</td>
</tr>
<tr>
<td>Insert</td>
<td>Young vs. Old</td>
<td>.05</td>
<td>5.4</td>
<td>.09</td>
</tr>
<tr>
<td></td>
<td>Young vs. Oldest</td>
<td>2.0</td>
<td>5.4</td>
<td>.43</td>
</tr>
<tr>
<td></td>
<td>Old vs. Oldest</td>
<td>3.7</td>
<td>2.2</td>
<td>.47</td>
</tr>
</tbody>
</table>

Note. *considered significant at p < .05.

Table 4: Single df Comparisons Investigating Within-Group Mean Test-Retest Auditory Threshold Differences Between Transducers at 4000 Hz

<table>
<thead>
<tr>
<th>Group</th>
<th>Transducer</th>
<th>Mean Difference</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Supranaural</td>
<td>Supranaural</td>
<td>1.0</td>
<td>1</td>
<td>1.19</td>
<td>.24</td>
</tr>
<tr>
<td></td>
<td>Earphone</td>
<td>5.5</td>
<td>1</td>
<td>1.19</td>
<td>.084</td>
</tr>
<tr>
<td>Old</td>
<td>Supranaural</td>
<td>5.5</td>
<td>1</td>
<td>1.19</td>
<td>.048*</td>
</tr>
<tr>
<td></td>
<td>Earphone</td>
<td>5.5</td>
<td>1</td>
<td>1.19</td>
<td>.048*</td>
</tr>
</tbody>
</table>

Note. *considered significant at p < .05.

Critical differences for ascertaining whether two sets of auditory thresholds are different at a 95% confidence level as a function of frequency were computed from the standard deviations of test-retest differences. In cases where the omnibus and post hoc analyses failed to reveal significant between and/or within group differences (see above), data were collapsed across age group and/or transducers. Table 7 displays the critical difference values. These critical differences suggest test-retest

Figure 2. Means of the signed test-retest differences as a function of group and transducer for the test frequency of 4000 Hz. Error bars represent plus/minus one standard deviation of the mean.

Figure 3. Means of the signed test-retest differences as a function of group and transducer for the test frequency of 8000 Hz. Error bars represent plus/minus one standard deviation of the mean.
threshold variability in the 10-14 dB range (assuming a 5 dB step size) could be due to chance alone at 250, 500, and 1000 Hz. In other words, one would need to observe a 15 dB difference in test-retest audiometric threshold to be 95% confident that the difference is real and not due to measurement variability. At frequencies of 2000, 4000, and 8000 Hz, differences between two audiometric thresholds from an individual of 5 to 20 dB would have to be observed before one could be 95% confident that the difference is real and not due to measurement variability. The actual critical difference varies as a function of group and transducer.

Discussion

The purpose of this study was to examine auditory threshold test-retest variability in elderly listeners, relative to young adults, using both supraaural and insert earphones. This was done by determining if the signed test-retest difference differed as a function of age group and transducer. The results of this study suggest that there was significant test-retest variability of audiometric pure-tone thresholds between the young (22-34 year olds), old (50-63 year olds), and oldest adult listeners (65-81 years old). These age effects were found to be significant only at frequencies higher than 1000 Hz (i.e., 2000, 4000, and 8000 Hz).

There are a number of factors which could have contributed to the significant test-retest variability found in this study. The first contributing factor could have been transducer variability resulting from the method used to couple the stimulus to the auditory system. The supraaural and insert earphones used in this study both have different coupling characteristics. Differences in the method of coupling of the transducer to the ear could have resulted in changes in the level of the stimuli which, in turn, could have caused variability in threshold responses across groups.

Zwislocki et al. noted that supraaural earphones can be unreliable at low, mid, and high frequencies due to the acoustic coupling between the sound source and the tympanic membrane. Zwislocki et al. noted that variability can occur at low frequencies due to air leaks between the earphone cushion and the pinna. Mid frequency variability can be due to unstable amounts of sound pressure enhancement in the mid frequencies. Also, variability at high frequencies can be due to listener to listener variability in earphone position and the anatomy of the pinna and cartilaginous ear canal.

Zwislocki et al. (1988) stated that insert earphones have some advantages over supraaural earphones in that inserts are less susceptible to air leaks. Insert earphones also minimize some of the wave effects and increase interaural attenuation. Yet Zwislocki et al. noted that ear canal geometry, eardrum impedance, and controlling for exact insertion depth are important considerations.

Table 7: Critical Differences at a 95% Confidence Level for Auditory Threshold Differences (dB) as a Function of Frequency, Group, and Transducer

<table>
<thead>
<tr>
<th>Frequency (Hz)</th>
<th>Group</th>
<th>Transducer</th>
<th>Critical Difference (dB)</th>
</tr>
</thead>
<tbody>
<tr>
<td>250</td>
<td>Young, Old, and Oldest</td>
<td>Supraaural and Insert</td>
<td>17.2</td>
</tr>
<tr>
<td>500</td>
<td>Young, Old, and Oldest</td>
<td>Supraaural and Insert</td>
<td>17.3</td>
</tr>
<tr>
<td>1000</td>
<td>Young, Old, and Oldest</td>
<td>Supraaural and Insert</td>
<td>17.4</td>
</tr>
<tr>
<td>2000</td>
<td>Young</td>
<td>Supraaural</td>
<td>17.5</td>
</tr>
<tr>
<td>2000</td>
<td>Old</td>
<td>Supraaural</td>
<td>17.6</td>
</tr>
<tr>
<td>2000</td>
<td>Young, Old, and Oldest</td>
<td>Insert</td>
<td>17.7</td>
</tr>
<tr>
<td>4000</td>
<td>Young, Old, and Oldest</td>
<td>Supraaural</td>
<td>17.8</td>
</tr>
<tr>
<td>4000</td>
<td>Young and Old</td>
<td>Insert</td>
<td>17.9</td>
</tr>
<tr>
<td>8000</td>
<td>Young and Old</td>
<td>Supraaural and Insert</td>
<td>18.0</td>
</tr>
<tr>
<td>8000</td>
<td>Oldest</td>
<td>Supraaural and Insert</td>
<td>18.1</td>
</tr>
</tbody>
</table>

Table 6: Scheffe Pair-Wise Comparisons of Between-Group Mean Test-Retest Auditory Threshold Differences at 6000 Hz

<table>
<thead>
<tr>
<th>Group Comparison</th>
<th>Mean Difference</th>
<th>Critical Difference (dB)</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young vs. Old</td>
<td>-2.2</td>
<td>4.3</td>
<td>0.02</td>
</tr>
<tr>
<td>Young vs. Oldest</td>
<td>-8.5</td>
<td>4.4</td>
<td>&lt;0.001*</td>
</tr>
<tr>
<td>Old vs. Young</td>
<td>-3.3</td>
<td>5.1</td>
<td>0.011*</td>
</tr>
</tbody>
</table>

Note: *p < 0.05.
discussed changes to the aging auditory system which cause
degeneration to the outer and middle ear noting structural
differences in the tympanic membrane, tympanic membrane,
degeneration of the incudostapedial and incudostapedial joints of the ossicles, and tensor tympani and
temporal muscle atrophy. These changes have been shown to
affect the higher frequencies first in aging adults as shown by
Willott's profile of hearing loss (Willott, 1991). The changes in
the aging ear canal structure combined with the inherent
variability in the transducer coupling characteristics could have
caused an increase or decrease in the sound pressure level at the
tympanic membrane. The variablesound pressure levels could,
therefore, contribute to an increase in test-retest variability in
the elderly in the higher frequencies.

Other listener variables, as discussed earlier, are behavioral
factors which suggest that the elderly population may be
difficult-to-test. These include emotional motivation, physical
discomfort, difficulty in attending, fatigue, and antagonism
about the test process. Overall, an interaction effect between
the physiological and behavioral factors could have resulted in
the observed outcome of increased test-retest variability in
the higher frequencies for the oldest group compared to the
young and old adult groups. For example, the interaction
between coupling effects and anatomical variation in the senior
group could be responsible for both the greater test-retest
difference at 8000 Hz and the lesser change at 2000 Hz. The
key consideration are the acoustic interaction between
transducer type, stimulus frequency and aging changes.

According to the outcomes found in this study, no significant
variability was found in the lower frequencies between the three
groups tested. This finding is similar to the results reported by
Stuart, Stenstrom, Tompkins, and Vandenhoff (1991) which
examined the test-retest audiometric threshold variability with
supraaural and insert earphones among children and young
adults. Stuart et al found that a 10-15 dB change in test-retest
variability would be necessary to be 95% confident the
difference did not occur due to the variability of test-retest error.
The performance of the young adult listeners is very
similar between the Stuart et al study and this study. There was,
however, greater test-retest variability in audiometric thresholds for the
oldest group of listeners in this study. That is, for thepure part critical differences for a 95% confidence level were
in the order of 10-15 dB for all listeners except at 4000
and 8000 Hz where a difference of 20 dB would have to be
observed for the oldest group for one to be confident that the
difference was not due to inherent transducer variability.
In conclusion, the findings of this study support that
test-retest audiometric threshold variability is significantly larger at
higher frequencies for geriatric listeners above 65 years of age
compared to the younger adults/older adults age groups. Based
on these findings, audiologist with adult patients should be
aware of test-retest audiometric threshold variability as a function
of transducer and patient age. Caution should be taken when
interpreting high frequency audiometric threshold variability in
the elderly population. It is suggested that more research in
geriatric audiology should be conducted in the literature to enhance
the current pure-tone air conduction assessment tools. In
particular, further research should be conducted to examine the
differences in transducer coupling procedures and physiological
differences between younger adult and older adult age groups
versus seniors. It may prove to be the case that assessment of
geriatric patients should be modified in the future to reduce test-retest variability (Ornage, MacNeil, & Stouffer, 1997).

Submitted: May, 1998
Accepted: February, 1999

References

noise levels for audiometric testing rooms. (ANSI S1.1-1991). New York:
ANSI.

audiometers. (ANSI S3.6-1996). New York: ANSI.

screening for hearing impairments and middle ear disorders.asha, 52
(Supp. 2), 17-24.

Brody, H. (1975). Organization of the central cortex: Study of
aging in human cerebral cortex. Journal of Comparative Neurology, 162,
511-556.

performance of the ER-3A tubephone and the TDH-50P earphone.
Ear and Hearing, 9, 268-274.

Cobb, F. E., Jacobson, G. P., Newman, C. W., Kretschner, L. W.,
masking task performance. Evidence of the aging temporal resolution

Davis, A. C., Oreri, B., & Paving, A. (1990). Longitudinal study

Folstein, M. F., Folstein, S. E., & McHugh, P. R. (1975). Mini-
Mental State: A practical method for grading the cognitive state


REVUE D'ORTOPHONIE ET D'AUDIOLOGIE, VOL. 23, NO. 2 JUIN 1999
Landry and Green

Béliveau, M. D., Williams & Wilkins.


