Validation of a telephone-administered screening test for hearing impairment for use in the US.

Charles S. Watson (Communication Disorders Technology, Inc., Bloomington, Indiana)
Gary R. Kidd and Larry E. Humes (Indiana University, Bloomington)
C. Smits (VU University Medical Center, Amsterdam;
Rachel McArdle, Andrea Bourne, and Richard H. Wilson
(VA Medical Centers, Bay Pines/San Francisco/Mountain Home)

presented at Aging and Speech Communication 2011:
An International and Interdisciplinary Research Conference
October 13, 2011, Indiana University, Bloomington, IN
Four seven countries now have national screening tests for hearing impairment, administered by telephone. All follow the original Netherlands model, using three-digit sequences spoken in the country’s language, presented in a background of speech-spectrum noise. Thresholds are estimated using an adaptive procedure and the tests have been validated by comparing telephone-test thresholds to other audiometric measures collected on the same listeners. A US version of this test (Watson et al. 2009; 2010) has now been administered to two clinical groups, first at the Indiana University Hearing Clinic and second with much larger samples at three clinics operated by the Department of Veterans Affairs. Validity and reliability of the US test are shown to be comparable to that reported for other national tests, and similar to those of screening tests commonly used for other health-related conditions.

### Telephone-administered screening tests currently used in seven countries

Table 1 shows the tests currently in use in seven countries, together with the year each was introduced and information that has been reported about the numbers of callers taking these tests. The manner of publicizing the tests has been different in each country and it appears that the numbers of persons taking the tests has been strongly dependent on the amount and form of publicity for them. It is clear that many thousands of people have taken advantage of the opportunity to take a quick, convenient and inexpensive hearing test.

<table>
<thead>
<tr>
<th>Country</th>
<th>Date first available to public</th>
<th>Supported by</th>
<th>No. Calls Reported over various periods</th>
<th>Cost to Caller</th>
</tr>
</thead>
<tbody>
<tr>
<td>Netherlands</td>
<td>Jan. 2003</td>
<td>Nationale Hoorstichting</td>
<td>65,000, 4 months</td>
<td>Long distance fee</td>
</tr>
<tr>
<td>UK</td>
<td>Dec. 2005</td>
<td>RNID/MRC</td>
<td>449,490, 40 months</td>
<td>Long distance fee</td>
</tr>
<tr>
<td>Australia</td>
<td>Sept. 2007</td>
<td>Hearing CRC/NAL</td>
<td>166,000, 32 months</td>
<td>Free</td>
</tr>
<tr>
<td>Germany</td>
<td>July 2008</td>
<td>Her Tech gGmbH</td>
<td>32,000, 6 months</td>
<td>Long distance fee</td>
</tr>
<tr>
<td>Poland</td>
<td>(test version)</td>
<td>University Poznan</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Switzerland</td>
<td>Feb. 2009</td>
<td>Pro Audio Schweiz</td>
<td>17,000, 4 months</td>
<td>Long distance fee</td>
</tr>
<tr>
<td>France</td>
<td>Feb. 2009</td>
<td>France Presbyacusie</td>
<td>n/a</td>
<td>Long distance fee</td>
</tr>
</tbody>
</table>
The US version of the telephone screening test (Watson et al., 2009; 2010) is closely modeled on the Netherlands test developed by Smits and colleagues (Smits, Kapteyn and Houtgast, 2004). Sixty-four naturally spoken three-digit sequences, selected from an original set of 160, were individually adjusted through preliminary tests with normal hearing listeners to yield 50% correct recognition scores at a nominal speech-to-noise ratio (SNR) of -10.5 dB. On each trial one of these sequences is drawn at random, without replacement. The initial SNR is -4.5 dB and thereafter a one-up, one-down tracking procedure determines successive SNR values, converging on the value required to yield 50% correct responses. Forty trials were presented on each test. Additional audiometric measures were collected from all listeners, including a six-frequency audiogram, and often the WIN test of words presented in noise. Total listeners tested at each site were: Indiana University: 90; Mountain Home: 293; Bay Pines: 248; San Francisco: 152. The analyses presented here are for the data collected at the VA clinics.

Figure 1a shows the distribution of ages of the subjects; Figure 1b their mean audiogram.

![Age Distribution](image)

![Audiogram](image)

Figure 2a shows the correlations between SNR thresholds, based on the final 37 SNRs visited in the tracking history, and the four-frequency (500, 1000, 2000, 4000 Hz) pure-tone average (PTA) thresholds for the current study. Figure 2b shows the correlation with WIN scores. The other panels show similar data for the original Netherlands Test (Figure 2c), the French Test (Figure 2d), and the earlier data from the IU Hearing clinic (Figure 2e). The correlations for each of these sets of data are shown on the individual panels. From these data and from reports at meetings and informally about the other tests listed in Table 1, it is evident that the US test is as strongly related as the others to the lower audiometric frequencies.

Figure 3 shows the correlations between telephone-test thresholds and pure-tone thresholds at 500, 1000, 2000, 4000, 6000, and 8000 Hz and also between the WIN test and the pure-tone thresholds. While the size of the correlations decrease as the test frequency is increased, the correlations between telephone test and pure tones remain highly statistically significant up to 8000 Hz. This is not unexpected, of course, since lower and higher frequency losses are correlated in the majority of cases. Telephone tests might be less sensitive, or even insensitive, to rarely occurring narrow regions of loss at high frequencies, and callers must be alerted to that possibility.
Results

Figure 2. Data from the current study compared with other telephone tests.

Figure 3. Correlations with audiometric frequencies

Figure 4. Correlations w. track length: WIN & PTA

Number of trials required for a reliable telephone test.

The number of trials required to obtain a reliable estimate of the SNR threshold was determined by plotting the correlation between the telephone test and two different audiometric measures, the four-frequency PTA and the WIN test, as shown in Figure 4. For either the PTA or WIN criterion measures, it is evident that thresholds based on approximately 25 trials are nearly as reliable as those based on the entire 40-trial series.
Determination of criteria for feedback to the test taker.

The large numbers of listeners tested at the facilities operated by the Department of Veterans Affairs provide a unique opportunity to determine appropriate feedback criteria. Because the test is identified to listeners as a functional test of the ability to hear speech in noise, scores on the WIN test were used as the primary validation measure, rather than PTA values, although the strong correlation between the two made this choice less meaningful. Figure 5 shows the receiver operating characteristic (ROC) curve, for functional impairment defined by a WIN score of less than 7 dB.

Figure 6 shows fitted overlapping distributions for the data shown in the ROC curve. The two vertical lines (see points on ROC) distinguish those told that they should definitely seek a further assessment of their hearing (FAIL), those whose scores suggest that they would be well advised to seek a more complete evaluation (MARGINAL), and those whose performance shows no clear evidence of impaired hearing (PASS).

Choice of pass-fail criteria on the basis of detection theory (TSD, Green and Swets, 1966) requires joint consideration of the values and costs of the decisions (hits, misses, false positives, correct rejections) and the ratio of impaired to non-impaired cases. The TSD-based maximum-gain criterion for a pass on telephone screening tests is so strict that only a few percent would pass. Expected callers, however, are not ideal decision strategists and the criteria have been chosen on the basis of consumers proven tendencies to base decisions on short-term considerations. If even modest numbers of the failures can be motivated to seek comprehensive hearing evaluations, the test would be justified.
References


