Telephone Screening Tests for Functionally Impaired Hearing: Current Use in Seven Countries and Development of a US Version

DOI: 10.3766/jaaa.23.10.2

Charles S. Watson*
Gary R. Kidd†
James D. Miller*
Cas Smits‡
Larry E. Humes†

Abstract

Background: An estimated 36 million US citizens have impaired hearing, but nearly half of them have never had a hearing test. As noted by a recent National Institutes of Health/National Institute on Deafness and Other Communication Disorders (NIH/NIDCD) Working Group, “In the United States (in contrast to many other nations) there are no readily accessible low cost hearing screening programs…” (Donahue et al, 2010, p. 2). Since 2004, telephone administered screening tests utilizing three-digit sequences presented in noise have been developed, validated, and implemented in seven countries. Each of these tests has been based on a test protocol conceived by Smits and colleagues in The Netherlands.

Purpose: Investigators from Communication Disorders Technology, Inc., Indiana University, and VU University Medical Center of Amsterdam agreed to collaborate in the development and validation of a screening test for hearing impairment suitable for delivery over the telephone, for use in the United States. This test, utilizing spoken three-digit sequences (triplets), was to be based on the design of Smits and his colleagues.

Research Design: A version of the digits-in-noise test was developed utilizing digit triplets spoken in Middle American dialect. The stimuli were individually adjusted to speech-to-noise ratio (SNR) values yielding 50% correct identification, on the basis of data collected from a group of 10 young adult listeners with normal hearing. A final set of 64 homogeneous stimuli were selected from an original 160 recorded triplets. Each test consisted of a series of 40 triplets drawn at random, presented in a noise background. The SNR threshold for 50% correct identification of the triplets was determined by a one-down, one-up adaptive procedure. The test was implemented by telephone, and administered to listeners with varying levels of hearing impairment. The listeners were then evaluated with pure-tone tests and other audiometric measures as clinically appropriate.

Study Sample: Ninety participants included 72 who were volunteers from the regular client population at the Indiana University Hearing Clinic, and 18 who were recruited with a newspaper ad offering a free hearing test. Of the 90 participants, 49 were later determined to have mean pure-tone thresholds greater than 20 dB hearing level (HL).

*Communication Disorders Technology, Inc., Indiana University Research Park, Bloomington, IN; †Department of Speech and Hearing Sciences, Indiana University, Bloomington, IN; ‡VU University Medical Center, Amsterdam, The Netherlands

Charles S. Watson, Communication Disorders Technology, Inc., Indiana University Research Park, 501 N. Morton Street, Bloomington, IN 47404; Phone: 812-336-1766; E-mail: watson@comdistec.com

Portions of this research were presented at the Aging and Speech Communication Conferences, October 2009 and 2011, Bloomington, IN and at the International Conference on Adult Hearing Screening, June 2010, Cernobbio, Italy.

Drs. Charles S. Watson and James D. Miller are employees and shareholders in Communication Disorders Technology (CDT), Inc., a for-profit company that collaborated with Indiana University and VU University Medical Center (Amsterdam) in the development of the screening test described in this article. CDT, Inc. may profit from the sale or licensing of this test if it is made available to the public. Both of the collaborating universities may receive portions of any revenues from the test, for use in research on communication disorders.

Support for this research was provided by National Institutes of Health/National Institute on Deafness and Other Communication Disorders (NIH/NIDCD) Grant R43 DC009719, by Indiana University, and by Communication Disorders Technology, Inc.
INTRODUCTION

The majority of men and women in the United States who consider themselves to have a significant hearing disability have had no experience with hearing aids, and roughly half of them have never had a hearing test (Kochkin, 2007). There are ample data showing that many of these people would have a greatly improved quality of life if their hearing impairment were treated, primarily by their becoming successful users of hearing aids. Yueh et al (2003) surveyed 1595 articles dealing with screening and management of hearing impairment, published between 1985 and 2001. Their findings were consistent with survey data from Kochkin (2007), noting that, “Hearing loss is the third most prevalent chronic condition in older adults and has important effects on their physical and mental health. Despite these effects, most older patients are not assessed or treated for hearing loss” (Yueh et al, 2003, p. 1976).

Outside of the United States, in response to these same problems of prevalence and the need for evaluation, there have been recent initiatives to offer self-administered tests of hearing to the general public, utilizing a digits-in-noise procedure originally developed in The Netherlands. Prior to describing the development of a US version of this test, the history and current status of modern telephone screening tests is reviewed.

Background of Telephone Screening Tests

During the past 8 yr, a National Hearing Test administered by telephone has been implemented in The Netherlands. More recently, versions of that test have been introduced in the United Kingdom, Australia, Germany, and France and are expected soon in several other countries (see Table 1). All of these tests have been based on the procedure developed by the Dutch investigators, Smits et al (2004), in which the listener identifies spoken three-digit sequences presented in a background of speech-shaped noise. An adaptive-tracking method is used to estimate the value of the speech reception threshold in noise (SRTn). Feedback in these tests informs the caller that their hearing is either within normal limits or not. If it is not, callers are advised to consult a hearing specialist for a full evaluation. The Dutch version differs from the others, in that it provides one of three messages to the caller on the basis of the SRTn: (1) “Good” (SRTn < −4.1 dB); (2) “Insufficient” or “Marginal” (−4.1 dB < SRTn < −1.4 dB); (3) “Poor” (SRTn > −1.4 dB). Those with either “Poor” or “Insufficient/Marginal” scores are advised to visit a hearing specialist for further evaluation and treatment.

These telephone-administered screening tests have been developed to help adults decide whether they should seek an evaluation by a hearing professional, generally meaning an audiologist or otolaryngologist. Those who elect to take such tests should be advised to seek a professional evaluation if they have even the slightest suspicion that they may have a hearing problem, whether or not they pass a screening test. Since the screening tests discussed here involve the recognition of speech sounds in noise, the property of hearing for which they are valid screening instruments requires some clarification, particularly in relation to the terms hearing impairment, hearing disability, and hearing loss. Current usage generally identifies hearing loss with measures of pure-tone sensitivity, impairment with both pure-tone loss and below-average recognition of speech, and disability with the effects of impaired hearing on
the capacity to cope with the demands of everyday life. Performance on a screening test that involves the identification of speech in a background of noise is thus an indirect index of hearing loss to the degree that it correlates with pure-tone sensitivity, and a measure of impairment to the degree that it correlates with both pure-tone loss and reduced performance on other speech-recognition measures. The authors consider the screening tests to be a valid basis for advising those who fail them to seek professional evaluation, but do not regard them as the equivalent of clinical measures of loss or impairment.

The Dutch, French, and Australians have published follow-up results for their tests (Smits and Houtgast, 2005; Jansen et al, 2010; Meyer et al, 2011). Those from other countries have been presented at conferences and generally are in good agreement with the published results. In The Netherlands, a total of 881 persons returned questionnaires (Smits et al, 2006) which were mailed to them several months after they took the telephone test. About 95% rated the test “easy” to take. More important are the respondents’ reports of what they did as a consequence of being told that their hearing was either “good”, “insufficient”, or “poor”. For those who had not seen a hearing specialist in the past, about 57% in the “poor” category reported that they did so, as did 45% in the “insufficient” category and, interestingly, 14% in the “good” category. Answers to other questions indicated that many of those in the “good” category elected to take the test because they suspected some problem with their hearing. Thus, one goal of the test, a significant increase in the number of persons with hearing impairment who scheduled visits with hearing specialists, was apparently achieved. These data suggest that the telephone test resulted in a meaningful reduction in the number of persons with unidentified and untreated hearing impairment, although the number that obtained hearing aids could not be determined. More recently, Meyer et al (2011) reported a small increase in the number of hearing-impaired persons who failed the Telscreen (Australian) version of the telephone test and who then sought services, compared to the numbers who do so in the absence of a screening test. It is possible that the impact of failing a screening test might be significantly increased if the caller were contacted within a few weeks by someone who could advise them about what to expect in a further evaluation and about the help that might be expected from a hearing aid if that seemed advisable. Such follow-up might be most effective if provided by an agency or organization that would not benefit directly from the caller obtaining clinical services or from hearing-aid sales.

Currently Available Digit-Based Telephone Screening Tests

Telephone screening tests currently available in five foreign countries and under development in two others are shown in Table 1. At least two other countries, Sweden and Turkey, are developing similar tests, but only limited data are currently available from those countries. Each of these tests utilizes three-digit sequences spoken in the native language of the country, by a native speaker. Adaptive tracking methods are modeled to various degrees after the procedure developed by Smits and colleagues in The Netherlands (Smits et al, 2004, Smits and Houtgast, 2005). The national tests described in Table 1 were all modeled after the Dutch test, with some variations, as a part of the European Union (EU)-funded project HearCom (Vlamming et al, 2011). The Australian test (Meyer et al, 2011), which was not part of the HearCom project, differs from the others in several ways, including the spectral and temporal properties of the masking noise, the scoring system used to identify a correct response and an algorithm that varies the number of trials required to estimate the threshold as a function of the response variability. Despite these differences, the validity of the Australian test appears to be quite similar to that of the tests using the original method developed by Smits and colleagues.

<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>January 2003</td>
<td>Nationale Hoorstichting</td>
<td>65,000, 4 mo</td>
<td>Long-distance fee</td>
</tr>
<tr>
<td>United Kingdom</td>
<td>December 2005</td>
<td>RNID/MRC</td>
<td>449,490, 40 mo</td>
<td>Long-distance fee</td>
</tr>
<tr>
<td>Australia</td>
<td>September 2007</td>
<td>Hearing CRC/NAL</td>
<td>166,000, 32 mo</td>
<td>Free</td>
</tr>
<tr>
<td>Germany</td>
<td>July 2008</td>
<td>Her Tech GmbH</td>
<td>32,000, 6 mo</td>
<td>Long-distance fee</td>
</tr>
<tr>
<td>Poland (Test Version)</td>
<td>February 2009</td>
<td>University Poznan</td>
<td>n/a</td>
<td>n/a</td>
</tr>
<tr>
<td>Switzerland</td>
<td>February 2009</td>
<td>Pro Audio Schweiz</td>
<td>17,000, 4 mo</td>
<td>Long-distance fee</td>
</tr>
<tr>
<td>France</td>
<td>February 2009</td>
<td>France Presbyacousie</td>
<td>n/a</td>
<td>Long-distance fee</td>
</tr>
</tbody>
</table>

Note: Adapted from Buschermohle (2009). All but the Dutch and Australian tests are part of the EC HearCom Project. RNID/MRC = Royal National Institute for Deaf People/Medical Research Council; CRC/NAL = Cooperative Research Centre Association/National Acoustic Laboratories; n/a = not applicable; EC HearCom Project is the European Community Hearing Communication Project.
Development of a US Version of the Telephone Screening Test

In 2008, the authors elected to develop a version of the telephone test, to address the needs of the large untested number of hearing-impaired persons in the United States. That work was undertaken in a three-way collaboration between investigators from Communication Disorders Technology, Inc. (Watson, Miller), Indiana University (Kidd, Humes), and the test’s original developers at VU Medical Center, Amsterdam (Smits, Houtgast). Development and validation of the US version of the test has been reported at three international meetings (Watson et al, 2009, 2010, 2011).

An additional reason for the development of a standardized digit test for speakers of American English was that digit-based tests could be a very useful tool for both clinical assessment and for research on speech recognition by normal-hearing and hearing-impaired listeners. It has long been known that the SRT_n for the recognition of digits is as much as 5 dB below that for words in sentences and 15 dB below that for nonsense syllables (Miller et al, 1951). This small set of the most overlearned words in the lexicon (the digits) yields a measure of auditory speech processing that is less influenced by individual differences in cognitive abilities than measures based on any category of less familiar speech sounds (Ramkissoon et al, 2002).

The strategy for the development of the US version of the telephone screening test was similar to that used by Smits et al (2004). The goal was to create a large catalog of spoken three-digit sequences (triplets), and to adjust the speech-to-noise ratios (SNRs) at which each triplet was presented to achieve equally recognizable stimuli. In order to decrease the variation among the stimuli, the set was limited to the eight English one-syllable digits (1, 2, 3, 4, 5, 6, 8, and 9). Three digit sequences were recorded, rather than concatenating sequences of individual recorded digits because concatenated recordings are sometimes judged as unnatural and also because the greater variation within the larger set of stimuli (64 triplets rather than eight digits) might yield a more valid estimate of general speech recognition.

METHOD

Recording of the Three-Digit Series

Recordings were made of a selection of 160 randomly generated three-digit sequences, utilizing the eight one-syllable digits, 1, 2, 3, 4, 5, 6, 8, and 9. A female talker was selected from a group of 16 who had previously recorded speech samples used in the Speech Perception Assessment and Training System (SPATS) (Miller et al, 2008), and whose speech was informally identified in an earlier project as one of the most easily understood when masked by noise. The talker’s dialect was Middle American, sometimes referred to as General American. The recorded stimulus files were end-pointed to include 200 ms of silence before and after each spoken triplet. Those files were then combined with speech-shaped noise for their total duration (including the silent periods before and after the speech sample). The noise spectrum was matched to the mean spectrum of the spoken triplets. A unique speech-shaped noise burst was used as the masker for each three-digit series, to avoid the learning that has been shown to occur when a single noise sample is repeated (Coble and Robinson, 1992; Lyzenga and Smits, 2011).

Development of Equally Recognizable Stimuli

Using a method of constant stimuli, each of the 160 recorded sounds were presented once at each of 10 SNRs, from −20 to −2 dB, in 2 dB steps to 10 young adult listeners with normal pure-tone audiograms. The noise level was 70 dBA. All stimuli were bandpass filtered (∼300–4000 Hz) to simulate a telephone handset frequency response. Listeners used Sennheiser circumaural earphones (HD250) and responded by entering the digits, in the order heard, on a simulated telephone keypad, displayed on a computer monitor (the layout of a telephone keypad is different from that of the digital keypad on a computer keyboard).

A set of 80 triplets was selected from among the original set of 160 on the basis of uniform slopes, giving preference to steeper functions. Figure 1A shows fitted functions for all 160 sequences and Figure 1B shows the functions for the selected set of stimuli. The SNRs of the individual triplets were then adjusted to produce a set of stimuli that are equally recognizable (for P(C) = 0.5), as shown in Figure 1C. A set of 11 nominal SNR levels using 2 dB steps was created for each digit sequence, with the 50% point defined as level 8 (mean SNR = −10.5 dB), the highest SNR as level 1 (mean = 3.5 dB), and the lowest as level 11 (mean = −16.5 dB).

Because of concerns about the differences between the earphones (with filtering to simulate a telephone receiver) used in the initial effort to equate the stimuli and actual telephone receivers, an additional evaluation of the 80 selected triplets was conducted with a second set of 10 normal-hearing listeners. The digit sequences were presented (without the simulated telephone filtering) over a landline corded telephone (Model 29369GE1-A; Thomson, Inc.) using the same hosting and delivery system to be used for the actual telephone test (provided by Basis Audionet). The triplets were presented at only three SNR levels to check recognition of the sequences at a low (level 8), middle (level 5), and high (level 2) SNR level. On the basis of this second evaluation, a set of 64 stimuli was selected for use in the telephone test. Sixteen “outliers” (i.e., stimuli for which...
the number of correct responses was substantially greater or less than other sequences at the same nominal SNR) were eliminated from the set of 80. No additional SNR correction was made.

Using the catalog of 64 equated triplets, the telephone screening test was implemented on an interactive voice response (IVR) platform. As in the other tests shown in Table 1, the threshold SNR was estimated using an adaptive tracking algorithm. Forty triplets were presented, using a self-paced procedure in which a one-down, one-up algorithm converged on the SNR required to achieve 50% correct recognition. The listener was instructed to respond by pressing the numbers that were heard on the telephone keypad, in the order that they were presented. They were encouraged to respond to each sequence, even if they were very uncertain about the numbers that were presented. They were told that if they did not respond within 10 s that the next sequence would be presented. The responses were considered correct only if all three of the digits were identified, in the order in which they were presented. The probability of a correct response by chance was approximately 0.002. In contrast with typical two- or three-choice psychophysical procedures, this low probability of success by chance makes the correct recognition of stimuli at low SNR values quite meaningful. The Dutch version of the test uses 23 three-digit sequences, and computes the threshold as the average of trials 5–24 (using the SNR that would have been presented on trial 24, based on the response to the 23rd digit sequence). In the initial version of the US test, as noted above, 40 sequences were presented, with the goal of determining the minimum number necessary for a reliable test.

Figure 1. Psychometric functions for the full set of 160 stimuli, with the selected subset of 80 stimuli highlighted (A), and for the selected subset before (B) and after (C) normalization.
Subjects

A group of 72 listeners were tested from the regular client population at the Indiana University Hearing Clinic. In addition, 18 subjects were recruited with a newspaper ad, which offered a free hearing evaluation to persons who would first take the telephone screening test. Thus, the audiometric characteristics of the subjects were unknown until they had taken the telephone test, after which their hearing was evaluated using a standard audiological protocol. A pure-tone audiogram was administered immediately after the telephone test.

The final study population included 49 participants with impaired hearing and 41 with normal sensitivity, using pure-tone average (PTA) (0.5, 1.0, 2.0 kHz) > 20 dB hearing level (HL) as the criterion measure. The mean age in this sample was 53.5, and the standard deviation (SD) was 23.0 (range = 18–90).

Procedures

All subjects took the telephone test in a small office using the corded landline telephone described above. Frequency response and distortion measures for this telephone were typical of a sample of corded telephones evaluated as part of the test-development research.

The listeners were instructed to pick up the telephone handset, dial the test number, and enter their individually assigned identification number. Following the procedure used in other countries where telephone screening tests are currently in use, no instruction was given regarding which ear to test. All further instructions were provided by recordings heard over the telephone. The listeners were told to listen to three digit sequences and then enter them using the telephone keypad. After each three-digit response was entered they would hear another sequence, for a total of 40 sequences. At the end of the stimulus series, the caller was told that their threshold SNR was some value in dB, ranging from −8 to +12. No interpretation of these numbers was provided, as an institutional review board (IRB)-approved informed-consent procedure had stressed that this test was under development and that information about their hearing would be given to them after further audiometric measurements were made.

Immediately following the telephone test, all listeners were given pure-tone threshold tests (250–8000 Hz) in each ear, otoscopic examinations, and the Hearing-in-Noise test (HINT; Nilsson et al., 1994), a test of sentence recognition in noise. Additional audiometric measures were obtained as required for the subset of listeners who were part of the regular clinic population and for those recruited through newspaper ads who requested further evaluation.

RESULTS

Thresholds were defined as the average SNR on trials 5–41 (using the value for the 41st trial that was determined by the tracking algorithm, although only 40 triplets were actually presented). These thresholds ranged from −8.6 to 3.0 dB SNR, with a median of −3.9. The correlation between thresholds derived from the telephone test and pure-tone thresholds (PTA: average of 500, 1000, and 2000 Hz) is shown in Figure 2C (r = 0.74).

Scores on the HINT speech test did not correlate as well with the telephone test thresholds (r = 0.66) as did the Plomp and Mimpen (1979) sentences-in-noise test for the Dutch version of the telephone test (r = 0.86) (Smits et al., 2004). After completing the tests described above, the authors became aware of data showing that the HINT is a less reliable instrument than some other speech-in-noise tests, especially the Words in Noise test (WIN; Wilson et al., 2007).

An additional validation study is currently under way in cooperation with the US Department of Veterans Affairs (USDVA). Testing protocols in the USDVA project, including IRB informed-consent and collection of audiometric validation measures, were similar in all significant details to the procedures used with the group tested in the Indiana University Hearing Clinic, except that the WIN test was administered to the USDVA subjects, rather than the HINT. Preliminary data from that project are presented in Figure 2D (116 listeners with a mean age of 64.8 and an SD of 13.1; range = 24.0–90.1). The correlation between the SRTn and PTA values for these data (r = 0.76) is similar to that found in the current study. The correlation between the telephone test SRTn and scores on the WIN is 0.79; considerably higher than found between the telephone test and the HINT in the current study.

DISCUSSION

The associations between thresholds estimated with the telephone test and pure-tone thresholds obtained in the Indiana University Hearing Clinic and in the additional study being conducted in collaboration with the USDVA are thus quite similar to those obtained in The Netherlands (Smits et al., 2004), shown in Figure 2A, and by the French investigators (Jansen et al., 2010) shown in Figure 2B. Although no results in this format have been published for the tests currently in use in other countries, listed in Table 1, reports at scientific meetings describing the validation of those tests and other informal reports suggest that they all are quite similar.

One of the questions to which this study was addressed was the number of trials required to obtain a valid estimate of the SRTn. Examination of the correlations between PTAs and the telephone test thresholds
for different track lengths, as shown in Figure 3, revealed that the correlation changes very little after roughly 15 trials. An incidental observation that may prove of some practical value was that thresholds were also very similar whether computed as the average of all trials (excluding the first four) or as the average of just the lowest four or five SNR values visited. This observation confirms the earlier suggestion that even a single correct response can be meaningful in this task because of the very low probability of correct responses by chance.

Using the mean of SNR values from trials 5 to 25 as the threshold estimate, sensitivity and specificity values—based on 20 dB HL PTA as the criterion for hearing loss and a SRT of 5.7 dB—were 0.80 and 0.83, respectively. Use of the three-frequency PTA as the criterion does not mean that the authors conceive the test to be a screening instrument for hearing loss, since it is only a direct measure of functional hearing for speech. However, the comparison of PTA values to SNR thresholds obtained with digits in noise is the only form of test evaluation by which the US version of the test could be compared to tests administered in other languages.

CONCLUSIONS

This US version of a spoken-digits hearing screening test administered by telephone has been developed and validated with a group of 90 listeners with a range of PTA thresholds from about 0 to 75 dB HL. This test is similar to the original Dutch version developed by Smits et al (2004) and it has comparable validity. The correlation with mean three-frequency (500, 1000, 2000 Hz) pure-tone thresholds (r = 0.74) is quite similar to those for the telephone tests in Dutch, French, and
Australian English, for which validation data have been published. The correlation is also similar to that found obtained with preliminary data from a larger study currently being conducted in collaboration with the USDVA using the test described here. Similar screening tests are currently being used in seven countries, as shown in Table 1.

**Potential Value of a Telephone Screening Test**

A valid telephone screening test could be of considerable value in the United States, where roughly half of the persons with hearing impairment have never had their hearing tested, if two further issues can be satisfactorily addressed. The first is whether the test can be justified in terms of the relative values and costs associated with the possible test outcomes (correct detection of impaired hearing, failure to detect impaired hearing, erroneous identification of normal hearing as impaired, or correct identification of normal hearing). The second issue is whether significantly more persons with impaired hearing who fail such a screening test are likely to seek treatment for their impairment than would do so in the absence of a screening test.

An analysis of the validation data was made in which persons with mean pure-tone thresholds in excess of 20 dB HL were considered to be hearing impaired. Thresholds obtained with the telephone test were examined to determine sensitivity (hit rate), miss rate, false-positive rate, and specificity (1 – the false-positive rate) for those data, as shown in Table 2. These values are shown for two criteria: one for the “need evaluation” boundary and one for the “within normal range” boundary (i.e., either including or excluding the “marginal” group from the passing category). The column labeled “feedback to caller” is merely an example of the way screening data can be used to provide a categorical form of feedback, so that those who are most likely to benefit will be strongly advised to seek a full-scale hearing evaluation. Very similar values have also been obtained in the ongoing

![Figure 3. The correlation between the telephone test threshold SNR and PTA_{0.5, 1, 2 kHz}, as a function of the number of track values used to compute the threshold (starting at the fifth track value).](image)

**Table 2. Illustration of Different Levels of Performance on the Telephone Test (SNR Thresholds) Selected as Feedback Criteria**

<table>
<thead>
<tr>
<th>X = SNR On Tel Test in dB</th>
<th>Feedback to Caller</th>
<th>Sensitivity</th>
<th>Miss Rate</th>
<th>False-Positive</th>
<th>Specificity</th>
</tr>
</thead>
<tbody>
<tr>
<td>X &gt; -5.7</td>
<td>Need evaluation</td>
<td>0.80</td>
<td>0.20</td>
<td>0.17</td>
<td>0.83</td>
</tr>
<tr>
<td>-7.4 &gt; X &lt; 5.7</td>
<td>Marginal</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X &lt; -7.4</td>
<td>Within normal range</td>
<td>0.94</td>
<td>0.06</td>
<td>0.63</td>
<td>0.37</td>
</tr>
</tbody>
</table>

Note: SNR = speech-to-noise ratio.
study with the USDVA, using the WIN, as the criterion measure, rather than a PTA "fence."

Selection of an optimal decision criterion (or two criteria, if a "marginal" category is desired) to be used as the reference point(s) in providing feedback to callers would be relatively simple, if values and costs could be computed and the entire process were regarded as a game in which the goal was to maximize gain. Unfortunately, that is not the case. In statistical decision theory, the decision criterion that is required to achieve maximum gain is determined by summing the values and costs of the four possible decision outcomes (correct-positive, false-positive, correct-negative, incorrect-negative), each weighted by their probabilities of occurrence in the population that is tested. Gain is then evaluated for all possible criteria and the criterion value that achieves the maximum gain is identified (Green and Swets, 1966). When the differences between values and costs are large, as in the present case (particularly the small cost of a clinical evaluation for a person whose hearing is within normal limits, relative to a much larger value of becoming a successful user of a hearing aid), the maximum gain criterion is often so extreme (high or low) as to make the test irrelevant. In other words, a maximum-gain strategist concerned that he or she may have impaired hearing should seek a full evaluation, rather than take a screening test. Thus, taking a screening test cannot be described as an optimal decision strategy. Most people, however, are not optimal decision makers. Human decisions are more strongly influenced by short-term gain, simple curiosity, and wishful thinking than are those of an optimal strategist (see, for example, Kahneman, 2007). Many who should seek a full evaluation will not do so unless motivated by something that is personally meaningful, such as a family member advising them to do so or, perhaps, by taking and failing a screening test. Nevertheless, some of the basic principles of decision theory are worth considering when choosing criteria for a screening test.

In selecting the decision criterion, it is important to recognize that among those taking the screening test, the a priori probability of at least a mild hearing problem is significantly greater than that the callers have normal hearing. Given the relatively low cost of professional assessment compared to the value of needed intervention, a very strict criterion for a clear pass should be used with the telephone screening test, or any other voluntary screening test for hearing impairment.

The criterion for a "pass" shown in the bottom row of Table 2 is quite strict and results in a very high hit rate (0.94) and a very small miss rate (0.06), but a high rate of false-positives (0.63). The also means that specificity (1 – the false positive rate) is relatively low (0.37). The practical consequences of such a strict criterion for a

**Is the Telephone Test Worth Implementing?**

As mentioned earlier, two studies of the impact of telephone screening tests on callers' subsequent behavior (Smits et al, 2006; Meyer et al, 2011) suggest at least a modest increase in the likelihood that those who fail the test will seek services. It is possible that aggressive follow-up explaining the potential benefits of further evaluation and appropriate treatment could result in higher rates of service seeking than were reported in those studies, particularly if the follow-up was conducted by an agency or organization without financial interests in clinical services or devices, or ties to those with such interests. In addition to the value of a valid, convenient, and inexpensive screening test to the large population of persons with age-associated hearing loss, there are other at-risk groups for which such a test might be a first step toward seeking needed services. Those include, for example, persons who fail the test and who work in noisy environments without access to industrial hearing conservation programs or who routinely expose themselves to music amplified at levels exceeding damage risk criteria. Although such persons should seek a comprehensive clinical evaluation, large numbers of them do not, and failing the telephone test might provide an incentive to do so.
How to Implement such a Test in the United States?

The test described in this report could be made available in the United States at an estimated cost of $3–4 per call, including the expenses associated with hosting the calls, testing both left and right ears of all callers, and providing the sort of vigorous follow-up mentioned in the preceding discussion. Alternative ways to pay for such a service include the government (more likely in countries with national health services), funds from commercial interests that might benefit from increasing numbers of persons seeking services, or a direct-pay option by an organization without any commercial interest in clinical services or hearing aid revenues. The authors have been advised by relevant federal agencies that the first option is highly unlikely in the United States. The second option is feasible, although it is possible that persons told that they need to be evaluated for a hearing aid by agents who profit from hearing aid sales may decrease confidence in the test. The authors are convinced that, in the United States, the only approach that would avoid these problems would be for an independent organization, one without direct ties to clinical service providers or to the manufacturers of hearing aids, to provide the test for a charge sufficient to cover the costs of the test itself and of follow-up to explain to those who fail the potential benefits of a further assessment.

Uses Other Than as a Telephone Test

The simplicity of digit-in-noise tests and their proven validity suggest that such tests might also be useful for functional hearing screening in settings where persons could be tested with earphones in a quiet place, rather than by telephone. While many schools and businesses have equipment available to conduct hearing screening tests, many more do not and those might benefit from a version of this test that is administered by a computer. Such an implementation conducted in a quiet place with good-quality earphones has been shown to yield higher rates of correct identification of hearing impairment than those achieved with the telephonic versions of the test (Smits et al, 2004).

Limitations and Directions for Future Research

As suggested above, the major limitation of a telephone screening test is its cost. While the authors favor offering the test for a small fee, it is possible that the many free tests available from hearing aid dealers and from online sites sponsored directly or indirectly by the hearing aid industry may make the public unwilling to pay even a few dollars for such a test. The behavior of those who might benefit from such a test might, however, be influenced by aggressive publicity stressing the value of a convenient, inexpensive, and valid screening test that can be administered in a quiet room in the caller’s home. Whether the test can be successfully marketed can only be determined by a vigorous effort to publicize it, initially in selected regions and then nationwide if revenues permit. Success of a screening program for hearing impairment would not be measured strictly in terms of the numbers of persons taking the tests. Even though response rates were similar to those shown in Table 1 for telephone tests in The Netherlands, the United Kingdom, and Australia, the ultimate goal is to influence those who would benefit from clinical services to seek them. The authors plan to combine the administration of the test with intensive follow-up to encourage those who fail the test to seek assessment by a hearing professional (an audiologist or an otolaryngologist). Various methods of follow-up will be investigated to determine the most effective means of increasing the numbers of successful hearing aid users.

Acknowledgments. The support and counsel of a Clinical Advisory Committee for this project is greatly appreciated. Members of that committee are Judy Dubno, CCC-A, Nancy Nelson-Barlow, CCC-A, Larry E. Humes, FAAA, Jill Preminger, CCC-A, David Wark, CCC-A, and Gail Whitelaw, CCC-A.

NOTES

1. The test was implemented by a firm specializing in IVR applications (Basis Audionet). The catalog of recorded messages consisted of the 64 three-digit sequences, each premixed at 10 different SNRs, plus recorded instructions and feedback messages.
2. Each of the national telephone tests now in use (Table 1) tests the single ear that is selected by the caller. For those with asymmetric loss, this might be their less-sensitive ear because the caller is concerned about that ear, or it might be their better ear since that may be the one that is normally used when talking on the telephone. Overall hearing impairment is probably best indicated by the status of the better ear. However, that issue is not raised in the instructions to callers. It is the intention of the developers of the US version of the test to routinely test both ears for all callers.
3. These preliminary data are shown with the kind consent of Andrea Bourne (San Francisco), Rachel McArdle (Bay Pines), and Richard H. Wilson (Mountain Home) from the US Department of Veterans Affairs. A report of an extensive validation study conducted in collaboration with those investigators will be made available at the completion of that study.

REFERENCES


