J Am Acad Audiol 25:937–951 (2014)

Validation of a Screening Test of Auditory Function Using the Telephone

DOI: 10.3766/jaaa.25.10.3

Victoria Williams-Sanchez*†
Rachel A. McArdle*†
Richard H. Wilson‡§
Gary R. Kidd**††
Charles S. Watson**††
Andrea L. Bourne‡‡

Abstract

Background: Several European countries have demonstrated successful use of telephone screening tests for auditory function. The screening test consists of spoken three-digit sequences presented in a noise background. The speech-to-noise ratios of the stimuli are determined by an adaptive tracking method that converges on the level required to achieve 50% correct recognition.

Purpose: A version of the three-digit telephone screening protocol for the United States was developed: the US National Hearing Test (NHT). The objective of the current study was to determine the sensitivity and specificity as well as the feasibility of the NHT for use within the Department of Veterans Affairs (VA).

Research Design and Study Sample: Using a multisite study design with convenience sampling, we used the NHT to collect data from 693 participants (1379 ears) from three geographical areas of the United States (Florida, Tennessee, and California).

Data Collection and Analysis: The NHT procedures were as follows: the participants (1) called a toll-free telephone number, (2) entered their assigned ear-specific identification code, (3) listened to 40-sets of digit triplets presented in speech-spectrum background noise, and (4) entered in the numbers that they heard on the telephone key pad. The NHT was performed on each ear, either at home or in a VA clinic. In addition to collecting data from the experimental task, we gathered demographic data and the data from other standard-of-care tests (i.e., audiometric thresholds and speech recognition tests in quiet and in noise).

Results: A total of 505 participants completed the NHT at a VA clinic, whereas 188 completed the test at home. Although the ear-specific NHT and mean pure-tone threshold all correlated significantly ($p < 0.001$), there were more modest correlations in the low- and high-frequency ranges with the highest correlation seen with the 2000 Hz mean pure-tone threshold. When the NHT 50% point or threshold was compared with the three-frequency PTA at 500, 1000, and 2000 Hz, the sensitivity was 0.87 and specificity was 0.54. When comparing the NHT with the four-frequency PTA at 500, 1000, 2000, and 4000 Hz, the sensitivity was 0.81 and specificity increased to 0.65. The NHT also correlated strongly with other speech-in-noise measures.

*Bay Pines VA HealthCare System, Bay Pines, FL; †Department of Communication Sciences and Disorders, University of South Florida, Tampa, FL; ‡James H. Quillen VA Medical Center, Mountain Home, TN; §Department of Audiology and Speech-Language Pathology, East Tennessee State University, Johnson City, TN; **Communication Disorders Technology, Inc., Bloomington, IN; ††Indiana University, Bloomington, IN; ‡‡San Francisco VA Medical Center, San Francisco, CA

Victoria Williams-Sanchez, Bay Pines VA HealthCare System, Audiology (126), P.O. Box 5005, Bay Pines, FL 37710; E-mail: vawilli2@mail.usf.edu

Portions of this work were presented at the International Aging and Speech Communication Conference in Bloomington, IN, October 2011; the American Speech-Language-Hearing Association Conference in San Diego, CA, November 2011; and the Joint VA Department of Defense Audiology Conference in Dallas, TX, March 2012.

The screening test used in this study was developed with funding provided by National Institutes of Health/National Institute on Deafness and Other Communication Disorders grant R43DC009719 to Communication Disorders Technology, Inc. (CDT), Bloomington, IN. Charles Watson is a stockholder of CDT and might benefit financially from the licensing, sale, or other commercial use of that test. The Rehabilitation Research and Development Service, VA, also supported this work through a Career Development Award to Rachel A. McArdle and a Merit Review, the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP), and a Senior Research Career Scientist award to Richard H. Wilson.

Disclaimer: The contents of this manuscript do not represent the views of the Department of Veterans Affairs or the United States government.

Conclusions: The NHT was found to correlate with other audiometric measures, including pure-tone thresholds and speech recognition tests in noise, at sufficiently high correlation values to support its use as a screening test of auditory function.

Key Words: Diagnostic techniques and procedures, hearing loss, hearing test, screening, sensitivity and specificity, telephone, validation studies

Abbreviations: AUROC = area under the receiver operating characteristic curve; BP = Bay Pines VA Healthcare System; CDT = Communication Disorders Technology, Inc.; 4-Freq PTA = four-frequency PTA at 500, 1000, 2000, and 4000 Hz; LE = left ear; MH = James H. Quillen VA Medical Center, Mountain Home; NHT = National Hearing Test, NU-6 = Northwestern University Auditory Test No. 6; PTA = pure-tone average; QuickSIN = Quick Speech-In-Noise Test; RE = right ear; ROC = receiver operating characteristic curve; SF = San Francisco VA Medical Center; SNR = signal-to-noise ratio; 3-Freq PTA = Three-frequency PTA at 500, 1000, and 2000 Hz; VA = Veterans Affairs; WIN = Words-In-Noise Test

INTRODUCTION

Although hearing loss is reported as the third most common chronic health condition among adult Americans, hearing loss is underdetected and undertreated (Cruickshanks et al, 1998; Bogardus et al, 2003; Chia et al, 2007). Hearing loss may be underrecognized because of its gradual onset and lack of visible symptoms (Yueh et al, 2010) and may be dismissed as being either unimportant or an inevitable aspect of aging by both older listeners and healthcare providers (Weinstein, 1994; Cacciatore et al, 1999; Dalton et al, 2003). Despite the subtle symptoms and lack of urgency with which hearing health is sometimes viewed, hearing loss can be debilitating and is strongly associated with functional and cognitive decline, depression, and reduced quality of life (Cacciatore et al, 1999; Bogardus et al, 2003; Dalton et al, 2003; Chia et al, 2007; Chisolm et al, 2007; Fellinger et al, 2007; Lin, 2011; Tun et al, 2012).

The aforementioned burden of hearing loss provides justification for the development and validation of simple and convenient methods to screen for auditory impairment as a first step in encouraging persons to seek evaluation and treatment of hearing loss (Bogardus et al, 2003; Dalton et al, 2003; Chia et al, 2007; Yueh et al, 2010; Chou et al, 2011). In general, health screenings can be completed in many environments such as health fairs, open houses, and senior living centers, but most are conducted during primary care visits (Johnson et al, 2008; Medwetsky and Scherer, 2011). The US Preventive Service Task Force (1996) recommended routine screening of hearing loss in adults, and other investigators have advocated for the inclusion of hearing screenings as a routine part of physical examinations for adults 65 yr and older (Gates et al, 2003; Johnson et al, 2008; Johnson et al, 2009; Yueh et al, 2010). Hearing screenings are beneficial and cost effective in terms of time and resource management because those who pass the examination do not require extensive assessment, thus allowing resources to be allocated to those who require referrals or more comprehensive evaluations, leading to access to treatment and rehabilitative options (Strong et al, 2005; Yueh et al, 2010; Demorest et al, 2011). Unfortunately, most patients who see their healthcare practitioners do not receive routine screenings for hearing loss (Bogardus et al, 2003). Despite the evidence of benefit from hearing screenings, there is a paucity of hearing screening instruments. This deficiency in screening has been attributed to a lack of organization and innovativeness (McBride et al, 1994; Bogardus et al, 2003; Davis et al, 2007) that could be improved if screening tests were a feasible part of routine care and/or if novel and structured options existed. Effective screening instruments must be accurate, practical, and administered quickly without specialized training (Strong et al, 2005; Yueh et al, 2010; Demorest et al, 2011).

A valid screening test identifies individuals who need a diagnostic evaluation, but does not predict the diagnosis (Demorest et al, 2011). With that said, there are many diverse forms of screening tests, and some have proven to be more adequate than others because different screening protocols can evaluate different aspects of hearing function (Chou et al, 2011). Hearing loss can be explained in many domains of auditory function, but most audiologists think of hearing loss in terms of decreased sensitivity to pure tones. An individual, however, can have a decreased sensitivity to pure tones but limited or no loss in functional hearing, such as understanding speech in quiet or in noise. Or, more commonly, an individual can present with normal or near-normal pure-tone thresholds but have a self-perceived hearing disability with complaints of difficulty hearing in everyday situations. Thus, hearing screening protocol can differ in respect to what it measures, how it measures or quantifies impairment, and what aspect of hearing or functionality the screening test is designed to detect. As a valid screening test attempts to identify individuals who need a diagnostic test, several factors influence what aspect of hearing the screening instrument is assessing and the quality of that assessment.

As mentioned previously, screening tests are available in many forms. Possibly the simplest form of a screening protocol is a self-report questionnaire. Self-reported questionnaires can be standardized, inexpensive, easy, and quick to administer. Such questionnaires are designed to assess self-perceived handicap associated with hearing loss but do not provide direct measures of actual hearing abilities. Although questionnaires have demonstrated usefulness, they have their limitations and have been reported to lead to more false-positive responses and are considered less efficient because they generate numerous clinical visits without positive rehabilitative outcomes (Yueh et al, 2010). Screening instruments that emit sounds are reported to provide an aura of legitimacy that the questionnaires lack (Yueh et al, 2010). The use of sound to screen for hearing impairment can come in many forms, such as the antiquated whispered-voice test that involves whispering numbers or words typically without visual cues or using a noisemaker such as rubbing of fingers, tuning forks, or other noncalibrated noise tools and asking for a response from the listener. Many of these older tests lack standardization, validation, and reliability (Bogardus et al, 2003). However, there are several standardized sound emitting screening tools, such as tone-emitting otoscopes that are reported to be efficient screening instruments (Yueh et al, 2010) or the latest innovation of screening applications on electronic tablets (e.g., iPads) or online (Albrecht et al, 2005; Wolfe et al, 2012; Leensen and Dreschler, 2013). The option used by most hearing healthcare professionals is the pure-tone screening audiometer. From kindergarten up, pure tones are used to screen hearing in the school system, at health fairs, and in occupational settings (Steinberg et al, 1940; Meinke and Dice, 2007).

Although pure tones may be the most popular screening stimulus, the frequencies and presentation levels tested affect the sensitivity and specificity of the instrument. Although specific results about the range and configuration of hearing loss can be obtained by pure-tone screening, a lack of ecological validity exists in the results because pure-tone sensitivity does not sufficiently predict performance in everyday listening situations, such as understanding speech in the presence of background noise. Therefore, the use of pure tones as a screening protocol may not be the optimal stimuli for assessing hearing functionality. On the other hand, screening procedures are available that use speech stimuli that do not evaluate hearing sensitivity directly but, rather, indirectly by providing evidence about the auditory functionality most important for everyday communication (Paglialonga et al, 2013). Thus, the use of speech stimuli increases the ecological validity of a screening test. The selected speech stimuli, or the actual words presented, can influence the validity of a screening test. The relationship among different speech stimuli has been studied extensively, and it is well known that certain words are easier to recognize than others; in addition, words in isolation are more difficult than words presented with context. In the case of words presented in isolation, monosyllabic digits, which are basically a closed-set stimulus, are easily recognized because of their high familiarity (Miller et al, 1951; Wilson et al, 2008). The use of prerecorded digits with the Western Electric 4A (later 4C) audiometer was one of the first widely used auditory tests (Fowler and Fletcher, 1926; Fletcher and Steinberg, 1929). Digits have continued to be used as auditory stimuli more recently in dichotic listening paradigms, as well as for screening tests such as the telephone-administered screening test being used in Europe (e.g., Smits et al, 2004).

The use of speech presented in noise requires less strict acoustical requirements than pure tones, because of suprathreshold presentation of the speech signal in controlled background noise (Leensen and Dreschler, 2013). Screening protocol using speech in noise is not concerned with the amplitude or volume of the signal necessarily, as long as the listeners have sufficient audibility. Thus, the use of digits-in-noise delivered over the telephone as an auditory screening technique overcomes several of the problems associated with speech in quiet and nonspeech screening stimuli, as well as improving the convenience and privacy of the screening test. These tests determine the signal-to-noise ratio (SNR) required for 50% correct recognition of three-digit sequences spoken in a noise background (Smits et al, 2004; Smits and Houtgast, 2005; Smits et al, 2006). Whereas differences among telephone receivers leading to calibration issues may limit the validity of pure-tone data obtained with telephones, the measured reliability of SNR thresholds is reported to be sufficiently high for hearing screening (Smits et al, 2006; Watson et al, 2012; Zokoll et al, 2012). A digits-in-noise test can be completed at the listener's home, at a booth located in public area, in a primary care office, or in any relatively quiet environment, and it is easy to administer and score using the telephone. During the past eight yr, at least seven countries have developed some form of a national telephone hearing screening test, most related to the European HearCom project (Buschermohle, 2009; Zokoll et al, 2012; Watson et al, 2012). The digit stimuli for these tests are presented in the language or dialect appropriate to each country, and efforts have generally been made to ensure that the recorded sequences are equally identifiable. Validation has been completed by comparing telephone data with other audiometric data on the same listeners. Additionally, the developers of those tests report moderately high sensitivity and specificity for the prediction of performance on other speech tests or of average pure-tone thresholds. As Watson and colleagues suggest, the results of these screening tests should not be regarded as a substitute for a clinical hearing evaluation, but poor performance on the screening protocol is a reasonable basis on which to advise individuals to seek an audiological evaluation. The countries with established digit-based telephone-administered screening tests report that many thousands of people have taken advantage of the opportunity to take a quick, convenient, and inexpensive hearing screening test (Buschermohle, 2009; Zokoll et al, 2012; Watson et al, 2012).

A US version of the telephone hearing screening test was developed in a collaborative effort between Communication Disorders Technology, Inc. (CDT), Indiana University, and VU University, Amsterdam (Watson et al, 2009; 2012). Similar to the Dutch National Hearing Test, the US National Hearing Test (NHT) uses three-digit sequences presented in speech-shaped noise that is matched to the mean spectrum of the spoken digits (Watson et al, 2012). As an overview, the NHT uses the eight monosyllabic digits, which excludes the bisyllabic *seven*. The three-digit sequences are naturally spoken by an adult speaker with a General American accent. A unique sample of speech-shaped noise was added to each triplet with 200 msec of leading and trailing noise. Thus, each three-digit sequence is a unique utterance of those digits and is presented in a unique noise burst. With the NHT, the initial SNR is −4.5 dB and thereafter a one-up, one-down tracking procedure, in 2-dB steps, determines successive SNR values, converging on the value required to yield 50% correct recognition (threshold). A total of 40 triplets are presented, and the threshold is the average SNR defined by the final 37 presentations. The NHT is implemented on an interactive voice response platform accessible through a toll-free telephone number, which is hosted by a commercial telephone service company (Basis Audionet). The initial validation was completed at the Indiana University Hearing Clinic (n = 90) and showed a strong positive relationship between three-frequency pure-tone averages (3-Freq PTAs; 500, 1000, and 2000 Hz) and the NHT thresholds [$r = 0.74$, (Watson et al, 2012)]. The only speech test used in this initial study was the Hearing-In-Noise Test (HINT; Nilsson et al, 1994). The correlation between the HINT and NHT thresholds was not as strong ($r = 0.66$), as with the 3-Freq PTA, nor was it as strong as the correlation obtained by Smits et al, (2004) between the Dutch NHT thresholds and a Dutch sentences-in-noise test [$r = 0.86$, (Plomp and Mimpen, 1979)].

The aim of the present study was to evaluate the performance of the NHT with a large nation-wide veteran population and a wide range of audiological measures of hearing function. This study also included testing in the home as well as in a clinical setting using both traditional land-line and cable (VoIP-based) telephones. The sensitivity and specificity as well as the feasibility of the NHT are discussed.

METHODS

A descriptive study using convenience sampling was completed to evaluate the NHT. The veteran participants were recruited between April 2010 and October 2011 from the Audiology Clinics at the Bay Pines VA Healthcare System (BP) in the Tampa Bay area, the Mountain Home VA Medical Center in Upper East Tennessee (MH), and at the San Francisco VA Medical Center (SF). All of the participants were enrolled for VA healthcare benefits and received all audiological services free of charge. The Institutional Review Boards at each VA site approved all recruitment and study procedures.

The procedures were uniform across the three VA locations. An intake sheet was completed for each participant to gather demographic information (e.g., age, gender), collect information about their telephone service provider (e.g., cable or traditional telephone company), and attributes of the telephone used to complete the NHT (e.g., cord, cordless, amplified). Pure-tone thresholds and word-recognition performance in quiet were collected as part of routine clinical evaluations, but if not obtained during a scheduled clinical visit, a research assistant completed those tests.

Word recognition in quiet was assessed with the Northwestern University Auditory Test No. 6 (NU-6; Tillman and Carhart, 1966) presented at a level intended to produce maximum performance (Department of Veterans Affairs [VA], 2006). Monaural speech-in-noise data, measured by the Words-In-Noise (WIN) test (Wilson, 2003; Wilson et al, 2003; Wilson and Burks, 2005; Wilson and McArdle, 2007) or the Quick Speech-In-Noise (QuickSIN) test (Etymōtic Research, 2001; Killion et al, 2004), were collected from each participant. The WIN and QuickSIN provide essentially the same estimates of recognition performance in background noise (Wilson et al, 2007). The WIN test, which involves the presentation of monosyllabic words in multitalker babble at 7 SNRs from 24–0 dB in 4 dB decrements, was presented at 70 dB HL (ANSI, 2004) to participants with a 3-Freq PTA of less than 40 dB HL and at 80 dB HL for participants with higher 3-Freq PTAs (Wilson, 2003). The QuickSIN test was presented at 75 dB HL and is composed of IEEE sentences (Institute of Electrical and Electronics Engineers, 1969) embedded in multitalker babble from 25–0 dB SNR in 5 dB decrements. For both speech-in-noise tests, the participants verbally repeated what they heard after every presentation, and performance was evaluated using the 50% recognition point on the listeners' psychometric function, calculated with the Spearman-Kärber equation (Finney, 1952). All audiometric testing took place in a double-walled sound booth using a compact disc player to reproduce the speech materials and a standard audiometer (Grason-Stadler, Model 61) with insert earphones (Etymōtic, Model ER-3A).

After the audiometric data were collected, the participants were randomly assigned to complete the NHT either at home or at the VA. Regardless of where they completed the NHT, each participant was provided with an instruction sheet with the telephone number and identification codes to complete the NHT on each ear (see Appendix). The initial test ear was counterbalanced to avoid order effects. First, the participants dialed the toll-free number, listened to the instructions, keyed in the identification code using the telephone keypad, and completed the NHT on the ear indicated by keying in the three-digit response after each triplet presentation. After completing the NHT on the first ear, the participants hung up and repeated the task on the opposite ear using the second identification code provided. De-identified NHT data were stored on a server maintained by CDT. The participants who were assigned to complete the NHT at home were contacted by telephone if they had not completed the task within 7–14 days. The follow-up phone call helped ensure completion of the at-home NHT or offered the participants the opportunity to withdraw from the study if they no longer wished to complete the NHT.

RESULTS AND DISCUSSION

The aim of the present study was to evaluate the US NHT in a large VA population and validate the NHT by comparing performance on this measure with other audiological measures. Demographic and audiometric data are reported as well as speech-in-quiet and speech-in-noise performances. The NHT results are compared with the other audiological measures using correlational analyses. The sensitivity and specificity of the NHT are calculated and the feasibility of this screening test, based on the results obtained, is discussed.

Demographics and Audiometric Performance

A total of 693 veterans (20 females, 673 males) participated in the study, with 248 participants at BP, 293 at MH, and 152 at SF. The second column of Table 1 lists the mean age and SDs for the participants by site and collectively with the distribution of ages illustrated in Figure 1. The obvious age difference among sites in Table 1 was significant [analysis of variance, $F_{(2,690)} = 13.18$, $p < 0.001$], with the BP participants averaging approximately 5 yr older than the MH and SF participants ($p < 0.05$, post hoc evaluation with Bonferroni correction). The mean 3-Freq PTAs (500, 1000, and 2000 Hz) and four-frequency pure-tone averages (4-Freq PTAs; 500, 1000, 2000, and 4000 Hz) for the left ear (LE) and right ear (RE) also are listed in Table 1. When site differences in pure-tone threshold averages were evaluated with multiple one-way analyses of variance for each ear (see Table 1), the differences were statistically significant [$F_{(2,690)}$ = 16.10 (LE 3-Freq PTA), 13.34 (RE 3-Freq PTA), 16.95 (LE 4-Freq PTA), and 15.27 (RE 4-Freq PTA), $p < 0.00$]. Post hoc analysis with Bonferroni corrections revealed that SF participants had all four PTAs that were significantly better (lower) than the PTAs from each of the other two sites, and MH was significantly better than BP only for the RE 4-Freq PTA. Thus, the participants from BP were older and generally had more hearing loss. The differences among sites are a consequence of the age and audiological function of the participants and are not associated with geographical location. Site differences were not considered in any of the subsequent analyses. The descriptive statistics for the LE and RE audiograms from the 693 participants are listed in Table 2. At each of the lower frequencies (250–1000 Hz), mean thresholds for the two ears were within roughly 1 dB or less. In the higher frequencies (≥2000 Hz), the RE thresholds were 2–3 dB better (lower) than the LE thresholds. These two relationships between ears for pure-tone thresholds are a consistent finding in studies involving large numbers of participants (e.g., Wilson and McArdle, 2013). Because none of the between-ear threshold differences were significant, the pure-tone thresholds were combined by frequency into the overall mean audiogram (n = 1379 ears) that is

Table 1. Mean Ages, 3-Freq PTAs, and 4-Freq PTAs for LE and RE Listed by Study Site, Along with SDs

		3-Freq PTA (dB HL)†		4-Freq PTA (dB HL)‡	
Site	Age (yr)	LE	RE	LE	RE
Means					
BP	68.6*	36.3	34.3	43.5	41.4
MH	63.5	33.0	31.0	40.4	37.9*
SF	63.5	26.7*	25.9*	33.5*	32.1*
All sites	65.2	32.9	31.1	40.0	37.9
SDs					
BP	11.8	18.1	17.5	17.9	17.4
MH	12.1	15.8	14.5	16.2	15.3
SF	13.9	14.9	15.2	15.8	16.1
All sites	13.1	16.8	16.1	17.1	17.0

*Denotes significant differences across study sites for each measure ($p < 0.05$).

†500, 1000, and 2000 Hz.

‡500, 1000, 2000, and 4000 Hz.

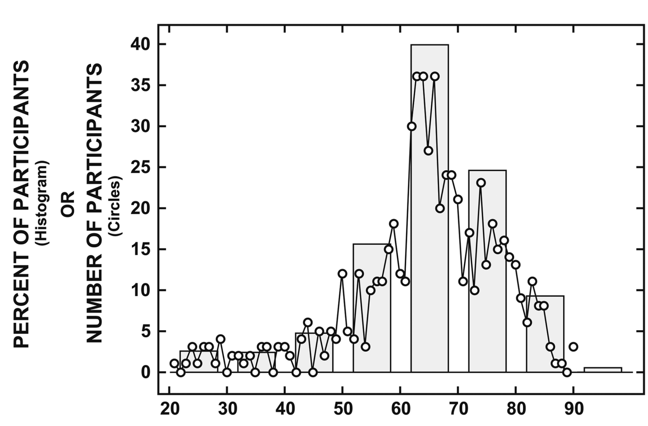

Figure 1. Distribution of ages for participants.

Table 2. Mean LE and RE Audiograms (dB HL, ANSI, 2004) for the 693 participants, Along with SDs and Mean LE Minus RE Threshold Differences (in dB)

	Frequency (Hz)					
	250	500	1000	2000	4000	80000
LE						
Mean	23.5	25.7	29.6	43.0	61.6	65.7
SD	14.4	15.0	17.7	22.8	24.1	27.2
RE						
Mean	23.3	24.6	28.4	40.1	58.3	63.4
SD	14.3	14.3	16.7	22.1	24.4	27.7
Difference						
LE − RE	0.2	1.1	1.2	2.9	3.3	2.2

illustrated in Figure 2, along with the mean audiograms from the three sites.

Speech Recognition Performance

Speech recognition performance in both quiet and noise were determined for each participant, and the number of ears that were evaluated by each test is shown in the third row of Table 3. The mean word recognition performances in quiet on the NU-6, which ranged from 0–100% on the participants, are listed in the left columns of Table 3. The mean performance in the LE was 2.7% lower than the mean performance in the RE. Also in Table 3, the ear-specific speech-in-noise mean performances (and SDs) on the WIN, QuickSIN, and NHT are listed. Performances on the WIN ranged from 0.4–26 dB SNR with mean performances of 13.1 dB SNR (LE) and 12.7 dB SNR (RE). Performance on the QuickSIN ranged from 0.5–27.5 dB SNR with mean performances of 10.9 dB SNR (LE) and 11.5 dB SNR (RE). The 1–2 dB better performance on the QuickSIN versus the WIN by the two participant groups shown in Table 3 is almost identical to the difference in a previous report from listeners who completed both speech-in-noise tests (Wilson et al, 2007). In the right columns of Table 3 are the NHT performances that ranged from −9.8 to 3.6 dB SNR with mean performances of −3.5 dB SNR (LE) and −3.7 dB SNR (RE). As a reminder, all speech-in-noise measures identified the SNR at which 50% recognition occurred, but the tests differed in regard to the target-speech stimuli, type of background noise, and administration mode, which contributed to the differences in mean performances observed among the tests.

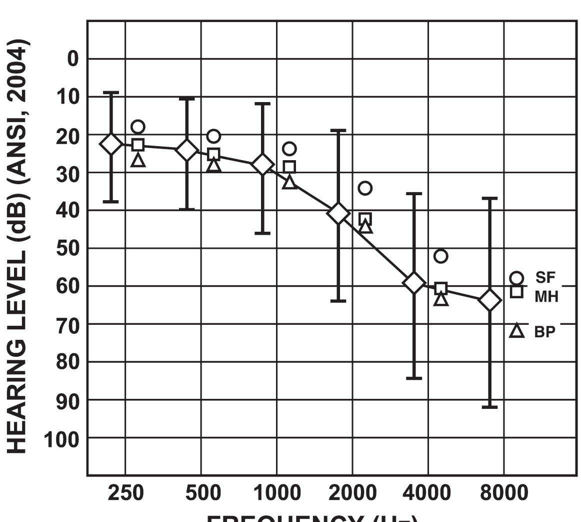

Figure 2. Mean audiometric data collapsed across ears for all participants (diamond) and separated by study sites. The SDs for the overall mean audiometric performance are shown.

To compare the performance on the NHT with another speech-in-noise measure, we generated psychometric functions. For this analysis, the listeners were divided into 10 groups (deciles) based on the NHT performance (134 or 135 ears per decile). First, performances on the NHT were rank-ordered and were then grouped into 10 contiguous segments in which those who performed best on the NHT were assigned to Group 1 and those who performed poorest were assigned to Group 10. Data points (percent correct for each presentation level visited) were taken from the tracking histories of all participants in each decile, excluding the first five trials. Logistic functions were fit to the data, omitting the data points based on fewer than 10 observations. Figure 3 is a 10-panel graph that shows the 10 sets of NHT and WIN data. Table 4 lists the 50% points and the slopes of the functions at the 50% points for the decile data. Figure 4 depicts the family of NHT and WIN functions. These comparisons show that, from a subset of the same listeners, although the NHT and WIN use different stimuli and were administered under telephone and earphone conditions, respectively, performance deficits were similar with the deficits running in the same direction. It is noteworthy that the WIN data do not appear to be as systematic as the NHT data, which is understandable because the data were organized into deciles on the basis of NHT performances. In general, although the functions for the two materials are displaced, the auditory impression gained from each is the same and the performance for both the NHT and the WIN clearly improves in a similarly systematic and orderly manner as the signal-to-noise level is increased.

NHT Validation

The main objective of this study was to validate the NHT results by comparing the NHT thresholds with other audiometric measures obtained from the same participants. The following section provides several plots comparing the NHT thresholds with the PTAs, and the WIN and QuickSIN performances (Figs. 5, 8, and 9). In addition,

Table 3. Mean LE and RE Performances and SDs for the NU-6 in Quiet, WIN, QuickSIN, and NHT Tests

	NU-6 (%)		WIN (dB SNR)		QuickSIN (dB SNR)		NHT (dB SNR)	
	LE	RE	LE	RE	LE	RE	LE	RE
Mean	79.4	82.1	13.1	12.7	10.9	11.5	−3.5	−3.7
SD	21.6	19.9	5.6	5.3	6.3	6.9	3.3	3.2
n	690	689	521	528	151	150	690	689

Note: The number of participants for each procedure is included (n).

correlational analyses were completed between the performances on the NHT and the other audiometric variables, as shown in Figure 6.

Figure 5 contains bivariate plots with the ordinate containing the NHT thresholds and the abscissae displaying the 3-Freq PTAs (top panels) and 4-Freq PTAs (bottom panels) for the LEs (left panels) and REs (right panels). The slopes of the linear regressions used to describe the data in each panel, all ~0.13%/dB, indicate a positive relationship between the two variables (i.e., the participants with higher PTAs also had higher NHT thresholds). Indeed, the two variables correlated strongly for the LE [$r_{(688)} = 0.62$, $p < 0.000$] and the RE [$r_{(687)} = 0.63$, $p < 0.000$]. The correlations between the 4-Freq PTA and the NHT were slightly higher for the LE [$r_{(688)} = 0.64$, $p < 0.000$] and RE [$r_{(687)} = 0.66$, $p < 0.000$]. Because both the 3-Freq and 4-Freq PTAs correlated strongly with the NHT performance, it is not surprising that there were significant correlations among the NHT threshold and the pure-tone thresholds across the frequency range (250–8000 Hz). Interestingly, as seen in Figure 6, the strength of the correlation was frequency dependent. Although the ear-specific NHT performance and pure-tone thresholds all correlated significantly ($p < 0.001$), there were more modest correlations in the low- and high-frequency ranges with the highest correlation observed with the 2000 Hz threshold.

An additional way of presenting the systematic relationship between the NHT and PTA is depicted in Figure 7. The fitted NHT psychometric functions for each decile from Figures 3 and 4 are shown, along with the mean 4-Freq PTA values for each decile. This family of functions consists of an orderly progression across the deciles, with a systematic increase in the PTAs across the deciles. These data reinforce the conclusion that the NHT systematically discriminates among listeners with different degrees of hearing sensitivity as measured by the PTA.

Beyond pure-tone threshold correlations, the NHT correlates strongly with other speech-in-noise measures as shown in Figures 6, 8, and 9. To minimize superimposed datum points in Figures 8 and 9, the WIN and QuickSIN 50% points were jittered randomly with an additive algorithm from −0.36 and 0.36 in 0.04 steps. As seen in Figure 8, WIN performance and the NHT had a strong positive relationship for the LE [$r_{(519)} = 0.75$, $p < 0.000$] and RE [$r_{(6526)} = 0.74$, $p < 0.000$] with regression slopes of 0.42 and 0.44dB/dB, respectively. Likewise, as illustrated in Figure 9, performance on the QuickSIN also showed a positive relationship with the NHT for the LE [$r_{(49149)} = 0.50$, $p < 0.000$] and RE [$r_{(48148)} = 0.60$, $p < 0.000$] with slopes of 0.27 and 0.28dB/dB, respectively. The correlations between the NHT and the speech-in-noise measures are high, which is not surprising because the NHT, WIN, and QuickSIN are all measures of speech recognition in background noise.

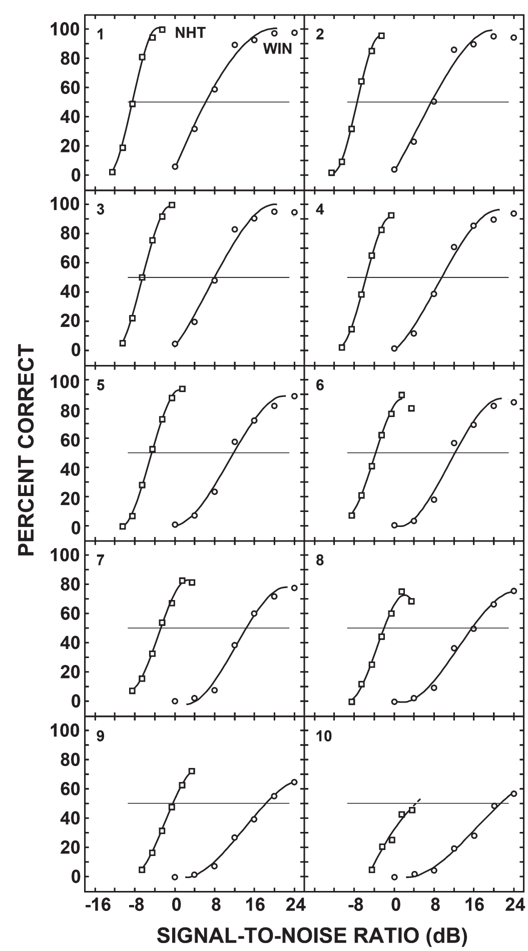

Figure 3. Psychometric functions for NHT and WIN data showing listeners performing at different levels and examining the correspondence among the functions. For this analysis, the listeners were divided into 10 groups (deciles) based on their performance on the NHT.

Table 4. Points at 50% (dB SNR) Calculated from Polynomial Equations Used to Fit Mean Data in Figure 3 Listed for NHT and WIN Protocols, Along with Slopes of Mean Functions at the 50% Points

	50% Points			Slopes at 50% Points		
DeciRange	NHT	WIN	Diff.	NHT	WIN	Diff.
1	−8.4	6.2	14.6	15.1	6.5	8.6
2	−7.4	7.4	14.8	14.9	6.4	8.5
3	−6.5	8.0	14.5	13.3	6.4	6.9
4	−5.6	9.5	15.1	12.8	6.2	6.6
5	−4.6	11.8	16.4	11.8	5.9	5.9
6	−3.7	12.3	16.0	10.7	6.3	4.4
7	−2.7	12.8	15.5	9.7	5.9	3.8
8	−2.0	15.7	17.7	9.1	4.7	4.4
9	−0.3	18.4	18.7	7.8	3.6	4.2
10	4.4	21.3	16.9	3.0	3.2	−0.2

The associations between the NHT and audiometric performance were also examined with respect to the location at which the NHT was completed (i.e., VA clinic or home). Across all study sites, 505 participants completed the NHT at a VA clinic, whereas 188 participants completed the test at home. Although an equal number of participants were randomly assigned to complete the NHT at home and at the clinic, several participants did not have landline telephones but, instead, only had cell phones, which were not included in the present study. These participants were asked to complete that task at the clinic. In addition, many other participants left the clinic and failed to take the NHT at home for unknown reasons, despite attempts to contact them to encourage completion. As seen in Table 5, the NHT thresholds were better (lower) when the task was completed in a VA clinic compared with tasks completed at the homes of the participants. The differences in performances between the two groups only were significant for the RE [$t_{(691,2)} = 1.2$, $p < 0.05$]. Factors such as a noisy room, lack of assistance with the instructions, and lower quality of telephones at home may have contributed to the performance difference. Despite these potential poorer test conditions at home, the data in Table 5 indicate that the at-home group was significantly older, had more hearing loss for pure tones, and poorer word recognition scores than the participants who completed the NHT in a clinic. Thus, the differences in NHT performance are likely related to age and hearing loss of the

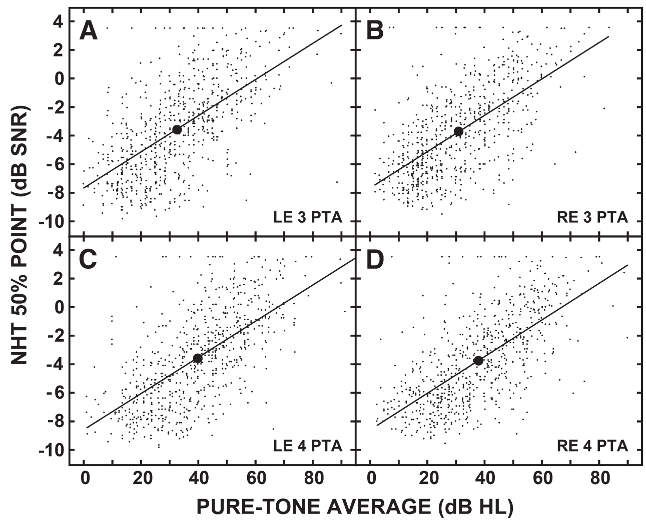

Figure 5. Ear-specific bivariate plots with NHT score shown as a function of the 3-Freq PTA (panels A and B) and 4-Freq PTA (panels C and D). The large filled circle represents the mean data. To minimize superimposed datum points, the PTA data were jittered randomly using an additive algorithm from –0.3 to 0.3 in 0.05 steps.

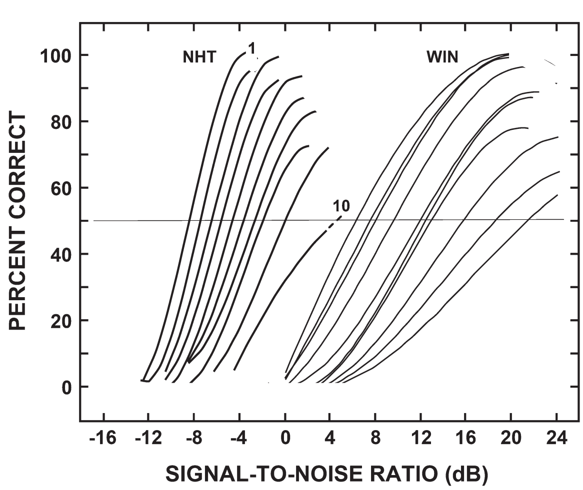

Figure 4. Family of NHT and WIN functions listed starting from decile group 1–10.

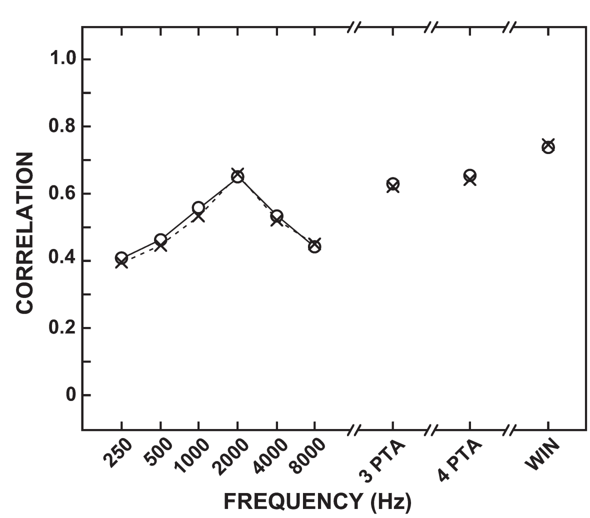

Figure 6. The correlations between the SNR required for 50% correct recognition on the NHT and other measures are shown for individual audiometric frequencies, for 3-Freq PTA, 4-Freq PTA, and for the WIN for the LE (x) and RE (o) collapsed across sites.

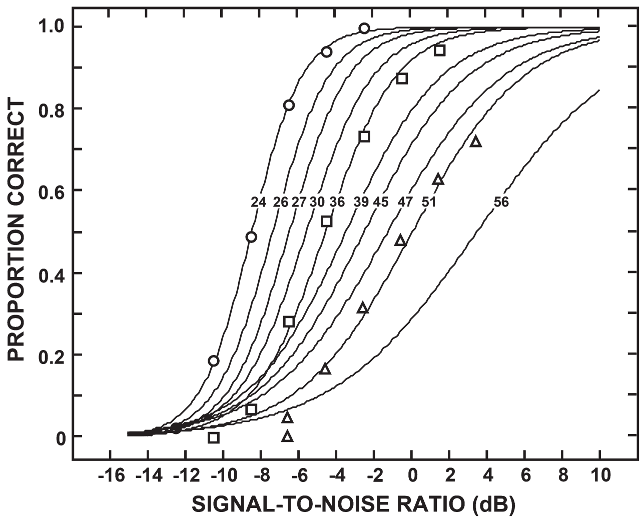

Figure 7. Group psychometric functions for each decile, based on performance on the NHT. Data points (mean percent correct for each SNR visited in the tracking procedure) are shown for the first, fifth, and ninth deciles. Also shown are the mean 4-Freq PTAs (rounded to the nearest dB HL) for the ears within each decile, reflecting the strong association between PTA and SNR on the screening test.

at-home group rather than actual setting in which the test was completed. For evaluation of this explanation, the at-home group participants were age matched (±2 mo) to a subgroup of 188 clinic participants (see Age-Matched Data in Table 5). When matched, the observed difference of poorer NHT performance by the at-home group compared with the VA-group disappeared; thus,

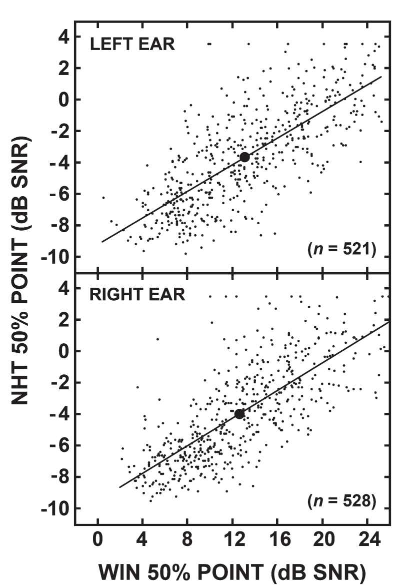

Figure 8. Ear-specific bivariate plots showing the NHT threshold (dB SNR) as a function of the WIN threshold (dB SNR). To minimize superimposed datum points, the WIN 50% points were jittered randomly with an additive algorithm from –0.36 and 0.36 in 0.04 steps. The telephone data were not jittered.

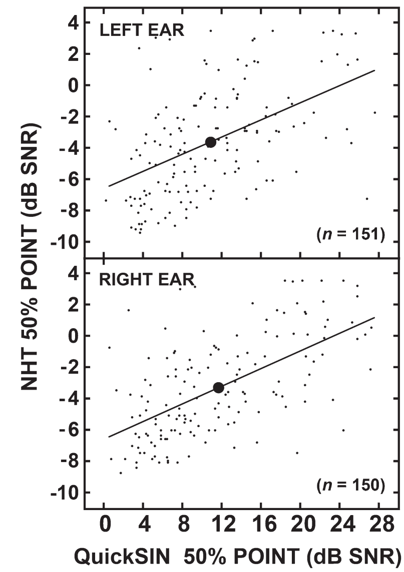

Figure 9. Ear-specific bivariate plots showing the NHT threshold (dB SNR) as a function of the QuickSIN threshold (dB SNR). To minimize superimposed datum points, the QuickSIN 50% points were jittered randomly with an additive algorithm from –0.36 and 0.36 in 0.04 steps. The NHT data were not jittered.

it is reasonable to attribute the differences associated with location of NHT completion to age and hearing loss. This finding also suggests that the broad variety of telephone receivers that must have been used by the at-home group had little or no influence on test performance.

To investigate further the possible contribution of the telephones used by the participants, we examined the performances obtained within the at-home group based on differences related to telephone attributes [i.e., cord (n = 108) versus cordless (n = 79)] and telephone service [i.e., cable company (n = 67) versus traditional telephone company (n = 118)]. No significant age or hearing-abilities differences were found between the groups that used the different telephone attributes or service providers. The mean NHT thresholds for both ears on the cord and cordless phones ranged from −2.5 to −3.0 dB SNR (with 3.0–3.5 dB SDs) and were not significantly different [$t_{(185,2)} = 1.02$ $p > 0.05$, LE; $t_{(183,2)} = 0.4$, $p > 0.05$, RE]. These results were somewhat surprising considering the acknowledged increase in interference and the decrease in the frequency response of cordless telephones (bandwidth of 3600 Hz) compared with the larger, but also limited, bandwidth of corded phones (Church and Taylor, 2007). The NHT thresholds for the participants with a cable company phone service were −3.6 dB SNR (SD = 3.4 dB) and −3.3 dB SNR (SD = 3.0 dB) for the LE and RE, respectively, whereas participants with traditional phone services had a mean NHT threshold of −2.6 dB SNR (SD = 3.2 dB) for the LE and −2.3 dB SNR (SD = 3.1 dB) for the RE. The NHT performance difference was only significant for the LE [$t_{(183,2)} = 2.7$ $p < 0.05$].

Table 5. Mean Ages, 3-Freq PTAs, 4-Freq PTAs, and Performance on the WIN and NHT Tests for LEs and REs for Participants Who Took the NHT at the VA (n = 505) or at HOME (n = 188)

		3-Freq PTA (dB HL)		4-Freq PTA (dB HL)		WIN (dB SNR)		NHT (dB SNR)	
Location	Age (yr)	LE	RE	LE	RE	LE	RE	LE	RE
Means									
VA	63.8	32.3	30.4	39.3	37.1	12.7	12.4	−3.8	−4.1
HOME	69.3*	34.1*	32.8*	41.9*	40.1*	14.3*	13.6*	−2.7	−2.8*
SDs									
VA	13.2	17.3	15.8	17.9	16.7	5.7	5.4	3.3	3.2
HOME	10.0	15.3	16.7	14.8	16.1	5.2	4.8	3.3	3.2
Age-Matched Data (n = 188)									
VA	69.2	33.2	31.2	39.7	38.2	13.1	12.9	−3.0	−3.2
HOME	69.3	34.1	32.8	41.9	40.1	14.3	13.6	−2.7	−2.8

Notes: SDs are shown in the middle panel. In the bottom panel, the age-matched data are shown. *Denotes significant differences ($p < 0.05$) across groups for each measure.

Sensitivity and Specificity

The validity of a diagnostic test is often described in terms of the ability of the test to identify which individuals have a disease of interest and which individuals do not have the disease. This is commonly expressed by measures of *sensitivity* and *specificity*. Sensitivity refers to the ability of the test to identify positive results, whereas specificity relates to the ability of the test to identify negative results. For a test of hearing function, sensitivity is the proportion of hearing-impaired individuals who are correctly identified as such, and specificity is the proportion of normal-hearing individuals who are correctly identified. Sensitivity is calculated by determining the number of *true-positives,* in our case, the number of individuals correctly identified by the NHT as having a hearing impairment, divided by the sum of the number of *true-positives* and *false-negatives* (i.e., the number of individuals with a hearing impairment who are incorrectly identified as passing the NHT). Specificity is calculated by determining the number of *true-negatives,* which is the number of individuals with normal hearing correctly identified by the NHT as passing, divided by the sum of the *true-negatives* and *false-positives* (i.e., the number of individuals with hearing within normal limits who are incorrectly identified as failing the NHT). There is always a tradeoff between sensitivity and specificity measures, and the calculated values depend on the cutoff values of the screening instrument and the gold standard diagnostic test with which the screening result is compared.

The traditional standard in diagnostic audiology for identifying individuals with hearing loss is based on pure-tone thresholds. A pure-tone, air-conduction threshold average of 25 dB HL or less is generally considered within normal limits in the adult population. For determination of the accuracy of the NHT, a criterion threshold of −6 dB SNR was used to distinguish between pass or fail (Watson et al, 2012). Using this threshold criterion, if a participant had a NHT threshold of −5.9 dB SNR or worse (higher), then that participant would be classified as failing, whereas if a participant had a threshold of −6 dB SNR or better (lower), then that participant would be classified as passing. On comparison of the 3-Freq and 4-Freq PTA results with a cutoff value of 25 dB HL and the NHT thresholds of −6 dB SNR, the ability of the NHT to identify *true-positives, false-positives, true-negatives, and false-negatives* was determined and are provided in the first two columns of Table 6. Using the above-listed classifications, sensitivity and specificity measures were determined and are shown in the bottom rows of Table 6. When the NHT threshold was compared with the 3-Freq PTA, the sensitivity was 0.87 and specificity was 0.54. On comparison of the NHT with the 4-Freq PTA, the sensitivity was 0.81 and specificity increased to 0.65. The 4-Freq comparison yielded better specificity values that were most likely owing to the increase in the number of participants with hearing impairment when the 4000 Hz threshold contributed to the calculations. Furthermore, because of convenient sampling of the study from audiology clinics, very few participants with normal hearing were recruited to participate in the study. This shortcoming has most likely influenced the

Table 6. True-Positives, True-Negatives, False-Negatives, False-Positives, and Calculated Sensitivity/Specificity for the NHT and WIN Tests

	NHT (n = 1379)		WIN (n = 1049)	
	3-Freq PTA	4-Freq PTA	3-Freq PTA	4-Freq PTA
True-Positives	405	855	588	813
True-Negatives	308	213	109	98
False-Negatives	101	196	11	22
False-Positives	265	115	341	116
Sensitivity	0.87	0.81	0.98	0.97
Specificity	0.54	0.65	0.24	0.46

Note: The tests are evaluated against the 3-Freq and 4-Freq PTAs.

specificity values obtained. However, the sensitivity and specificity values are sufficiently high to indicate that elevated NHT thresholds can be used to identify individuals who may benefit from further evaluation.

The accuracy of the screening test also can be described independent of the cutoff criterion by determining the area under the receiver operating characteristic (ROC) curve [the curve relating the hit rate (sensitivity) and false alarm rate (1−specificity) for a test] (Wikipedia, 2013; Creelman and Macmillan, 2005). Comparing the 3-Freq and 4-Freq PTA results with a 25 dB HL cutoff value defining normal hearing for pure tones and a −6 dB SNR cutoff point for the WIN, we computed the area under the ROC curve (AUROC) in relationship to both PTAs. The ROC curves are shown in Figure 10, and as listed in Table 7, the obtained AUROC values were 0.83 and 0.82 for the 3- and 4-Freq PTA, respectively. Values of sensitivity and specificity can be obtained from any point on the ROC curve, but for the purposes of comparison, pairs of values with sensitivity fixed at 0.80 were used. Specificity values of 0.70 and 0.67 were obtained for 3-Freq PTA and 4-Freq PTA (using NHT SNR cutoff values of −4.9 dB and −5.9 dB), respectively.

Although the PTA is commonly used as the primary measure of hearing ability, it is also informative to evaluate the NHT against a measure of speech-in-noise recognition. Outside of the United States, speech-in-noise measures are more commonly used by hearing professionals to assess hearing function. One such measure, the Plomp and Mimpen (1979) sentence test, was used as the criterion for the evaluation of the sensitivity and specificity of the Dutch version of the NHT (see Smits et al, 2004). In the present study, the WIN was used as the criterion measure for the assessment of the NHT. For determination of the sensitivity and specificity of the WIN, the cutoff threshold of 6 dB SNR, which defines the 90th percentile of the normal performance range, was used to distinguish between normal and abnormal speech recognition in noise (Wilson et al, 2003). Once again, on comparison of the 3- and 4-Freq PTA results with a cutoff level of 25 dB HL, the ability of the WIN test to identify *true-positives, false-positives, true-negatives, and false-negatives* was determined and are provided in the last two columns of Table 6 for the 1049 ears that underwent WIN testing. Using the above-listed classifications, the following sensitivity and specificity measures were determined and are also shown in the bottom rows of Table 6. When the WIN test was compared with the 3-Freq PTA, the sensitivity was 0.98 and specificity was 0.24. When the WIN test was compared with the 4-Freq PTA, the sensitivity was 0.97 and the specificity increased to 0.46.

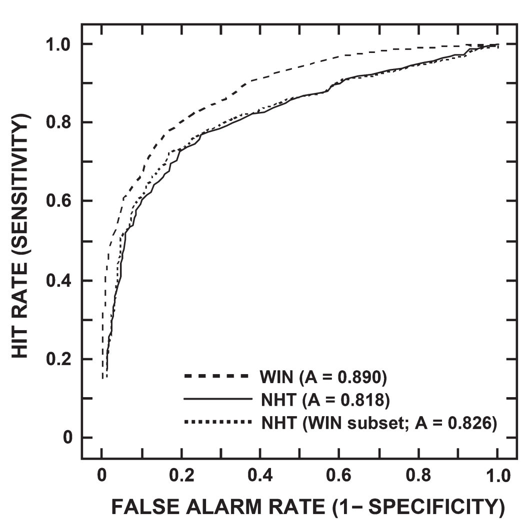

Figure 10. ROC curves for the WIN and NHT (full dataset and WIN subset) evaluated against the 4-Freq PTA (25 dB HL criterion).

Furthermore, the AUROC and specificity for a constant sensitivity value (0.80) were also calculated for the NHT and WIN. Using a threshold of 6 dB SNR on the WIN as the cutoff level for normal hearing function, we computed the AUROC and specificity (for sensitivity = 0.80) values for the NHT. As might be expected, these values were somewhat higher than those using PTA as the criterion, with AUROC = 0.86 and specificity = 0.76. The AUROC values for the WIN evaluated against the 3- and 4-Freq PTAs (≤25 dB HL criterion) were 0.86 and 0.89, respectively. The WIN specificity values corresponding to a sensitivity of 0.80 were 0.73 and 0.81 for the 3- and 4-Freq PTA. The AUROC and specificity values for a constant sensitivity of 0.80 are summarized in Table 7.

The sensitivity and specificity of the NHT and WIN were similar. Although these values are somewhat higher for the WIN than for the NHT, the differences are not great, given that the WIN was administered in a clinical setting with an audiometer, rather than over the telephone. When the same, somewhat smaller population was used to evaluate the NHT (data for only 1027 ears were available for the WIN), the NHT fared about the same as with the larger population, with AUROC = 0.83 and specificity = 0.71 (for sensitivity = 0.80) for the 3-Freq PTA, and AUROC = 0.83 and specificity = 0.70 for the 4-Freq PTA. A comparison of the ROC curves for the NHT and WIN, evaluated against the 4-Freq PTA (≤25 dB HL criterion), is provided in Figure 10. When evaluating both the WIN and NHT against PTA values, the WIN performed somewhat better relative to the 4-Freq PTA than to the 3-Freq PTA, and the NHT performed similarly for both PTA measures. This finding indicates that although the higher frequencies are important for the intelligibility of speech material, the limited frequency response of the telephone has little or no impact on the accuracy of the NHT. As mentioned earlier, the differences in performance and validity between the NHT and WIN tests may be attributed to the differences in the test stimuli and administration of the material.

Table 7. AUROC, Sensitivity, and Specificity for the NHT and WIN Tests

	NHT Against			WIN Against		NHT Subset Against	
	3-Freq PTA	4-Freq PTA	WIN	3-Freq PTA	4-Freq PTA	3-Freq PTA	4-Freq PTA
AUROC	0.83	0.82	0.86	0.86	0.89	0.83	0.83
Sensitivity	0.80	0.80	0.80	0.80	0.80	0.80	0.80
Specificity	0.70	0.67	0.76	0.73	0.81	0.71	0.70
n	1349	1349	1027	1027	1027	1027	1027

Note: The tests are evaluated against the 3-Freq and 4-Freq PTAs; the NHT is also evaluated against the WIN. The larger "n" is for the full set of usable NHT data, and the smaller "n" is for the subset of the NHT data for which WIN scores were available.

Although the background noise used in each speech-in-noise measure differed, with the WIN presented in the presence of multitalker babble and the NHT embedded in speech-shaped noise, this difference has not been shown to contribute much to performance differences on the WIN (Wilson et al, 2012). On the other hand, the NHT requires the recognition of a small set of spoken monosyllabic digits, compared with the multiple unpredictable monosyllabic words presented in the WIN test, which does contribute significantly to performance differences (e.g., Miller et al, 1951). Additionally, the mode of administration differed between the two tests; the WIN test was presented through earphones in a sound booth, whereas the NHT was presented in an open room (e.g., clinical office setting or somewhere in the home of the participants) and with the bandwidth of telephones that had a narrower frequency response than the earphones.

Feasibility

We discussed the NHT in terms of validity and have reported that the screening test provides sufficient sensitivity and specificity for use as a functional auditory screening protocol but have yet to discuss the feasibility of the NHT. The test is convenient and is cost and time efficient. Also, it is possible to modify the NHT to be even more time efficient by reducing the number of trials required to converge onto the 50% threshold. In the present study, the NHT took an average of 6.4 min (SD = 1.31 min) per ear. As seen in Figure 11, the number of trials required for a reliable estimate of the SNR threshold was determined by plotting the correlation between the NHT and the 3- and 4-Freq PTAs and WIN performance as a function of the number of trials used in the NHT threshold estimate. It is evident that thresholds based on approximately 25 trials are nearly as reliable as those based on a 40-trial series. The evidence that fewer trials are sufficient is similar to the data presented by Watson et al (2012), further supporting the reduction of the duration of the test. The NHT may be shortened to 25 trials and still maintain similar high sensitivity and specificity values. Computations based on a 25-trial track confirmed that AUROC and sensitivity/specificity values are nearly identical to those shown in Table 6 for 40 trials, as would be expected based on the similar correlations.

CONCLUSIONS

The NHT has been proposed for use in the United States for the same purposes that similar telephone-administered hearing screening tests are now in use in many European countries and in Australia (Watson et al, 2012). The NHT provides a convenient, easy-to-take, and valid screening test of functional hearing to anyone who wishes to take the test. There is no question that individuals concerned about the state of their hearing are best advised to seek a standard audiological assessment. It is well established, however, that a large proportion of individuals with impaired hearing in the United States and in most other countries have not had their hearing tested for a variety of reasons (Bogardus et al, 2003; Dalton et al, 2003). Reports of the response to telephone screening tests in other countries suggest that some of those who fail those tests comply with the recommendations that they seek further assessment and some of those are fitted with hearing aids (Yueh et al, 2010; Meyer et al, 2011). Another possible use of this brief test is as a quick screening device for use in clinics or physicians' offices that do not have in-house audiological services.

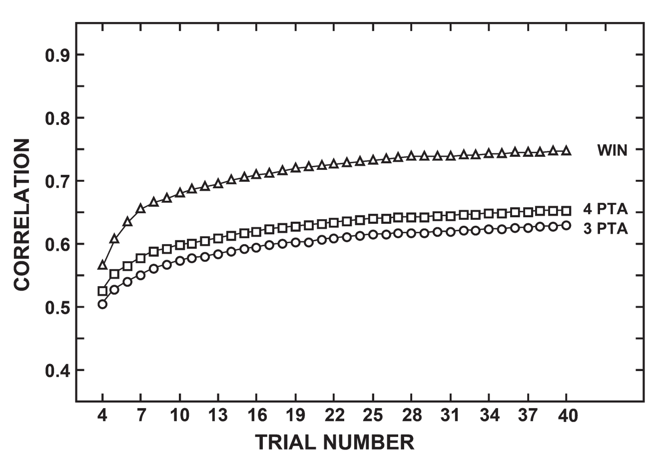

Figure 11. The correlation between the NHT and three hearing measures (3-Freq PTA, 4-Freq PTA, and WIN) as a function of the number of trials on which the threshold estimate is based.

Either for the general population or for use in clinics, value-cost analyses clearly show that the pass-fail criterion point for such a brief screening test should be very strict in the sense that only excellent performance should be considered a definite negative test outcome (meaning that the person's hearing is considered within normal limits). Requiring a high level of performance has two consequences: one desirable and one undesirable. The desirable consequence is that the "hit rate" for correctly identifying persons who should be advised to seek hearing assessment can be quite high, even 90–95%. The undesirable consequence of such a strict passing criterion is that the rate of false-positives will also be high, possibly 50–60%. The theoretical argument for this recommendation is a simple application of statistical decision theory with the goal of maximizing overall value and minimizing costs. The value of correctly advising a person with impaired hearing to seek a full-scale assessment of his or her hearing and receive appropriate treatment is vastly greater than the cost of having a person evaluated whose hearing turns out to be within normal limits. It is important to note that the false alarm rate is likely to be operating on a relatively small number of individuals with normal hearing, if taking the test is voluntary. As noted by Watson et al (2012), it is very likely that only a small number of persons who have hearing within normal limits will elect to take the test. Another value of hearing screening tests, where they have been widely publicized, is that the general public has become more aware of the importance of healthy hearing and of the value of treatment for impairments.

The most important difference between a screening test and a diagnostic test is the very different consequences of the two types of measurement. A positive result of a screening test most commonly means that the listener is referred for a complete audiological evaluation. A positive result on a diagnostic test, in contrast, most often is the basis for some form of treatment, although marginal findings can call for either repeating the test or selecting some other diagnostic measure. Another distinction between screening and diagnostic tests is that screening tests are sometimes given to entire populations, independent of any basis for believing that the tested individuals are at risk. Examples include routine mammograms, mandatory premarital blood tests, and universal vision and hearing screening in schools. Less common are elective screening tests that are offered to the general public, often in the form of kits purchased in a drugstore or, in the case of mental health, as tests that can be taken on the internet or in the popular press. Many of these elective screening tests are considered to be of dubious validity, and professional associations often advocate against their use. Blanket rejection of elective screening tests, however, should not be applied to tests of demonstrated validity, especially for clinical conditions such as hearing impairment that are known to be greatly under-reported.

Telephone-administered screening tests for functional hearing problems appear to meet most of the criteria for useful screening instruments that are accepted by the clinical community. At this time, no standards have been established by the US Food and Drug Administration for such tests, although that agency has suggested that such tests fall in the domain of medical instrumentation and therefore will eventually require approval. For such approval to be given, it is necessary for tests to be clinically validated, through studies such as the present work. Thus, the goal of the present study was to evaluate the US NHT for a large, mainly male, VA sample. It was demonstrated that the NHT provides sensitivity and specificity that are comparable to telephone screening tests evaluated with nonveteran populations in the United States and in other countries. The results indicate that the NHT is a quick, convenient, and valid screening test of functional hearing.

Acknowledgments. The authors thank Dr. Kelly Watts, along with the audiology staff and audiology graduate students who collected, submitted, and compiled the data from the clinics.

REFERENCES

Albrecht J, Elewout L, Verhage L, Vervweij C. (2005) Oorcheck: de validering van een interactief screening instrument. Internal report, LUMC Leiden.

American National Standards Institute (ANSI). (2004) *Specification for Audiometers* (ANSI S3.6-2004) New York: American National Standards Institute.

Bogardus ST, Jr, Yueh B, Shekelle PG. (2003) Screening and management of adult hearing loss in primary care: clinical applications. *JAMA* 289(15):1986–1990.

Buschermohle M. (2009) Hearing screening by telephone and internet, HearCom Implementation in Several Countries. Workshop on Hearing Screening and Technology, Brussels, January 28, 2009.

Cacciatore F, Napoli C, Abete P, Marciano E, Triassi M, Rengo F. (1999) Quality of life determinants and hearing function in an elderly population: Osservatorio Geriatrico Campano Study Group. *Gerontology* 45(6):323–328.

Chia EM, Wang JJ, Rochtchina E, Cumming RR, Newall P, Mitchell P. (2007) Hearing impairment and health-related quality of life: the Blue Mountains Hearing Study. *Ear Hear* 28(2):187–195.

Chisolm TH, Johnson CE, Danhauer JL, et al. (2007) A systematic review of health-related quality of life and hearing aids: final report of the American Academy of Audiology Task Force On the Health-Related Quality of Life Benefits of Amplification in Adults. *J Am Acad Audiol* 18(2):151–183.

Chou R, Dana T, Bougatsos C, Fleming C, Beil T. (2011) Screening adults aged 50 years or older for hearing loss: a review of the evidence for the U.S. preventive services task force. *Ann Intern Med* 154(5):347–355.

Church S, Taylor R. (2007) Telephone Network Interfacing. NAB Engineering Handbook, 609-644.

Creelman CD, Macmillan NA. (2005) *Detection Theory: A user's Guide, New York*. 2nd ed. Mahwah, N.J.: Erlbaum.

Cruickshanks KJ, Wiley TL, Tweed TS, et al.; The Epidemiology of Hearing Loss Study. (1998) Prevalence of hearing loss in older adults in Beaver Dam, Wisconsin. *Am J Epidemiol* 148(9):879–886.

Dalton DS, Cruickshanks KJ, Klein BE, Klein R, Wiley TL, Nondahl DM. (2003) The impact of hearing loss on quality of life in older adults. *Gerontologist* 43(5):661–668.

Davis A, Smith P, Ferguson M, Stephens D, Gianopoulos I. (2007) Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models. *Health Technol Assess* 11(42):1–294.

Demorest ME, Wark DJ, Erdman SA. (2011) Development of the screening test for hearing problems. *Am J Audiol* 20(2):100–110.

Etymōtic Research. (2001) QuickSIN™ (Compact Disc) 61 Martin Lane. Elk Grove Village IL: 60007.

Fellinger J, Holzinger D, Gerich J, Goldberg D. (2007) Mental distress and quality of life in the hard of hearing. *Acta Psychiatr Scand* 115(3):243–245.

Finney DJ. (1952) *Statistical Method in Biological Assay*. London, UK: C. Griffen.

Fletcher H, Steinberg JC. (1929) Articulation testing methods. Bell Telephone Systems Technical Publications, 8:806–854.

Fowler EP, Fletcher H. (1926) Three million deafened school children, their detection and treatment. *JAMA* 87(23):1877–1882.

Gates GA, Murphy M, Rees TS, Fraher A. (2003) Screening for handicapping hearing loss in the elderly. *J Fam Pract* 52(1):56–62.

Institute of Electrical and Electronics Engineers. (1969) IEEE recommended practice for speech quality measurements. *IEEE Trans Audio Electroacoust* 17:227–246.

Johnson CE, Danhauer JL, Bennett M, Harrison J. (2009) Systematic review of physicians' knowledge of, participation in, and attitudes toward hearing and balance screening in the elderly population. *Semin Hear* 30:193–206.

Johnson CE, Danhauer JL, Koch LL, Celani KE, Lopez IP, Williams VA. (2008) Hearing and balance screening and referrals for Medicare patients: a national survey of primary care physicians. *J Am Acad Audiol* 19(2):171–190.

Killion MC, Niquette PA, Gudmundsen GI, Revit LJ, Banerjee S. (2004) Development of a quick speech-in-noise test for measuring signal-to-noise ratio loss in normal-hearing and hearing-impaired listeners. *J Acoust Soc Am* 116(4 Pt 1):2395–2405.

Leensen MCJ, Dreschler WA. (2013) Speech-in-noise screening tests by internet, part 3: test sensitivity for uncontrolled parameters in domestic usage. *Int J Audiol* 52(10):658–669.

Lin FR. (2011) Hearing loss and cognition among older adults in the United States. *J Gerontol A Biol Sci Med Sci* 66(10):1131–1136.

McBride WS, Mulrow CD, Aguilar C, Tuley MR. (1994) Methods for screening for hearing loss in older adults. *Am J Med Sci* 307(1):40–42.

Medwetsky L, Scherer MJ. (2011) Factors influencing individuals' decisions to access hearing care services. *Hear Rev* 18(5):24–32.

Meinke DK, Dice N. (2007) Comparison of audiometric screening criteria for the identification of noise-induced hearing loss in adolescents. *Am J Audiol* 16(2):S190–S202.

Meyer C, Hickson L, Khan A, Hartley D, Dillon H, Seymour J. (2011) Investigation of the actions taken by adults who failed a telephone-based hearing screen. *Ear Hear* 32(6):720—731.

Miller GA, Heise GA, Lichten W. (1951) The intelligibility of speech as a function of the context of the test materials. *J Exp Psychol* 41(5):329–335.

Nilsson M, Soli SD, Sullivan JA. (1994) Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *J Acoust Soc Am* 95(2):1085–1099.

Paglialonga A, Grandori F, Tognola G. (2013) Using the speech understanding in noise (SUN) test for adult hearing screening1. *Am J Audiol* 22(1):171–174.

Plomp R, Mimpen AM. (1979) Improving the reliability of testing the speech reception threshold for sentences. *Audiology* 18(1):43–52.

QuickSIN Speech in Noise Test. (2001) Elk Grove Village, IL: Etymotic Research.

Smits C, Houtgast T. (2005) Results from the Dutch speech-in-noise screening test by telephone. *Ear Hear* 26(1):89–95.

Smits C, Kapteyn TS, Houtgast T. (2004) Development and validation of an automatic speech-in-noise screening test by telephone. *Int J Audiol* 43(1):15–28.

Smits C, Merkus P, Houtgast T. (2006) How we do it: The Dutch functional hearing-screening tests by telephone and internet. *Clin Otolaryngol* 31(5):436–440.

Steinberg JC, Montgomery HC, Gardner MB. (1940) Results of the World's Fair Hearing Tests. *J Acoust Soc Am* 12(2):291–301.

Strong K, Wald N, Miller A, Alwan A; WHO Consultation Group. Current concepts in screening for noncommunicable disease: World Health Organization Consultation Group Report on methodology of noncommunicable disease screening. *J Med Screen* 12(1):12–19.

Tillman TW, Carhart R. (1966) An expended test for speech discrimination utilizing CNC monosyllabic words. Northwestern University Auditory Test No. 6. SAM-TR-66-55. Tech Rep. SAM-TR. 1-2.

Tun PA, Williams VA, Small BJ, Hafter ER. (2012) The effects of aging on auditory processing and cognition. *Am J Audiol* 21(2):344–350.

U.S. Preventive Service Task Force. (1996) *Guide to Clinical Preventive Services*. 2nd ed. Baltimore, MD: Williams & Wilkins.

Watson CS, Kidd GR, Miller JD, Humes LE, Smits JCM, Festin JM. (2009) Development of a telephone-based hearing screening test for the United States. Poster presented at the International Aging and Speech Communication Conference, Bloomington, IN.

Watson CS, Kidd GR, Miller JD, Smits C, Humes LE. (2012) Telephone screening tests for functionally impaired hearing: current use in seven countries and development of a US version. *J Am Acad Audiol* 23(10):757–767.

Weinstein BE. (1994) Age-related hearing loss: how to screen for it, and when to intervene. *Geriatrics* 49(8):40–45, quiz 46–47.

Wikipedia. "Receiver operating characteristic," last modified December 10, 2003. http://en.wikipedia.org/wiki/Receiver_operating_characteristic_curve

Wilson RH. (2003) Development of a speech-in-multitalker-babble paradigm to assess word-recognition performance. *J Am Acad Audiol* 14(9):453–470.

Wilson RH, Abrams HB, Pillion AL. (2003) A word-recognition task in multitalker babble using a descending presentation mode from 24 dB to 0 dB signal to babble. *J Rehabil Res Dev* 40(4): 321–327.

Wilson RH, Burks CA. (2005) Use of 35 words for evaluation of hearing loss in signal-to-babble ratio: A clinic protocol. *J Rehabil Res Dev* 42(6):839–852.

Wilson RH, McArdle R. (2007) Intra- and inter-session test, retest reliability of the Words-in-Noise (WIN) test. *J Am Acad Audiol* 18(10):813–825.

Wilson RH, McArdle R. (2013) Characteristics of the audiometric 4,000 Hz notch (744,553 veterans) and the 3,000, 4,000, and 6,000 Hz notches (539,932 veterans) *J Rehabil Res Dev* 50(1):111–132.

Wilson RH, McArdle R, Roberts H. (2008) A comparison of recognition performances in speech-spectrum noise by listeners with normal hearing on PB-50, CID W-22, NU-6, W-1 spondaic words, and monosyllabic digits spoken by the same speaker. *J Am Acad Audiol* 19(6):496–506.

Wilson RH, McArdle RA, Smith SL. (2007) An Evaluation of the BKB-SIN, HINT, QuickSIN, and WIN Materials on Listeners With Normal Hearing and Listeners With Hearing Loss. *J Speech Lang Hear Res* 50(4):844–856.

Wilson RH, Trivette CP, Williams DA, Watts KL. (2012) The effects of energetic and informational masking on The Words-in-Noise Test (WIN). *J Am Acad Audiol* 23(7):522–533.

Wolfe A, Galster E, Thomas B. (2012) Sound check: Passing the test? Starkey Innovations Magazine [newsletter from Starkey Hearing Technologies]. 2(4):38–43.

Yueh B, Collins MP, Souza PE, et al. (2010) Long-term effectiveness of screening for hearing loss: the screening for auditory impairment—which hearing assessment test (SAI-WHAT) randomized trial. *J Am Geriatr Soc* 58(3):427–434.

Zokoll MA, Wagener KC, Brand T, Buschermöhle M, Kollmeier B. (2012) Internationally comparable screening tests for listening in noise in several European languages: the German digit triplet test as an optimization prototype. *Int J Audiol* 51(9):697–707.

Appendix

Subject Number: XX

Telephone Hearing Screening Study

Thank you for your participation. Please follow the directions created specifically for you below.
If you have any questions, please contact Audiology Research at (727) 398-XXXX ext. 4862.
You have been randomized to complete this study at: VA
You will hear 40 sets of 3 numbers (triplets) in the presence of varying background noise via the telephone. After each presentation of triplets, you must key in 3 numbers into your telephone keypad. If you did not hear the numbers, then please guess. Please continue the screening until you have completed all 40 triplets, and then complete the screening on your other ear.

(1) From a landline telephone, call the following number: 1-888-XXX-4240.
(2) You will be prompted to enter your IDENTIFICATION NUMBER.

Please key in the following IDENTIFICATION NUMBER: 10001707

(3) Press 2 to hear the screening instructions.
(4) To begin the screening, place the telephone to your LEFT ear and press 1.

*It is very important that you place the telephone to the correct ear indicated above.
(5) Upon completing the screening, using your LEFT ear, please hang up and redial 1-888-XXX-4240 and complete the screening using your RIGHT ear this time.

Please key in the following IDENTIFICATION NUMBER: 10001708

***IMPORTANT ***
★ Please perform the telephone hearing screening from a landline telephone only. DO NOT use a CELL PHONE or any other MOBILE telephone device.
★ If you are a hearing-aid user, DO NOT wear your hearing aids while completing the telephone hearing screening.
★ If you have any difficulties, please call Audiology Research. Or if you no longer wish to participate in the telephone hearing screening study, please call to inform us that you are longer able to participate. Audiology Research at (727) 398-XXXX ext. 4862