

Identification of Conductive Hearing Loss Using Air Conduction Tests Alone: Reliability and Validity of an Automatic Test Battery

Elizabeth Convery,^{1,2} Gitte Keidser,^{1,2} Mark Seeto,^{1,2} Katrina Freeston,^{1,2} Dan Zhou,^{1,2}
and Harvey Dillon^{1,2}

INTRODUCTION

Objectives: The primary objective of this study was to determine whether a combination of automatically administered pure-tone audiometry and a tone-in-noise detection task, both delivered via an air conduction (AC) pathway, could reliably and validly predict the presence of a conductive component to the hearing loss. The authors hypothesized that performance on the battery of tests would vary according to hearing loss type. A secondary objective was to evaluate the reliability and validity of a novel automatic audiometry algorithm to assess its suitability for inclusion in the test battery.

Design: Participants underwent a series of hearing assessments that were conducted in a randomized order: manual pure-tone air conduction audiometry and bone conduction audiometry; automatic pure-tone air conduction audiometry; and an automatic tone-in-noise detection task. The automatic tests were each administered twice. The ability of the automatic test battery to: (a) predict the presence of an air–bone gap (ABG); and (b) accurately measure AC hearing thresholds was assessed against the results of manual audiometry. Test–retest conditions were compared to determine the reliability of each component of the automatic test battery. Data were collected on 120 ears from normal-hearing and conductive, sensorineural, and mixed hearing-loss subgroups.

Results: Performance differences between different types of hearing loss were observed. Ears with a conductive component (conductive and mixed ears) tended to have normal signal to noise ratios (SNR) despite impaired thresholds in quiet, while ears without a conductive component (normal and sensorineural ears) demonstrated, on average, an increasing relationship between their thresholds in quiet and their achieved SNR. Using the relationship between these two measures among ears with no conductive component as a benchmark, the likelihood that an ear has a conductive component can be estimated based on the deviation from this benchmark. The sensitivity and specificity of the test battery vary depending on the size of this deviation, but increase with increasing ABG size, with decreasing test frequency, and when results from multiple test frequencies are taken into account. The individual automatic tests comprising the battery were found to be reliable and valid, with strong, significant correlations between the test and retest results ($r = 0.81$ to 0.99 ; $p < 0.0001$) and between automatic and manual audiometry procedures ($r = 0.98$ to 0.99 ; $p < 0.0001$).

Conclusions: The presence of an ABG can be predicted with a reasonably high degree of accuracy using AC tests alone. Applications of such a test battery include any clinical context in which bone conduction audiometry or specialized diagnostic equipment is unavailable or impractical. Examples of these include self-fitting hearing aids, whose efficacy relies on the ability of the device to automatically administer an in situ hearing test; self-administered adult hearing screenings in both clinical and home environments; large-scale industrial hearing conservation programs; and test environments in which ambient noise levels exceed the maximum permissible levels for unoccluded ears.

(Ear and Hearing 13:XXX;00–00)

¹HEARing Cooperative Research Centre, Carlton, Victoria, Australia; and ²National Acoustic Laboratories, Macquarie Park, New South Wales, Australia.

Automatic audiometry refers to a method of measuring hearing thresholds that is driven by a predetermined algorithm, rather than by a clinician. The validity of this technique has been well-established, from early implementations in which a nonclinician is directed by a series of computer commands to select the appropriate presentation level and frequency (Sparks 1972) to modern devices that are capable of obtaining masked air conduction (AC) and bone conduction (BC) thresholds (Ho et al. 2009; Swanepoel et al. 2010) and assessing the validity of the test results (Margolis et al. 2007), including implementations in smartphones and tablet computers (Kam et al. 2012; Van Tasell & Folkeard 2013). Although it has not yet made significant inroads into everyday clinical practice, automatic audiometry is useful for efficiently testing large numbers of people, as in adult hearing screening (Sakabe et al. 1975) and industrial noise conservation programs (Sparks 1972; Picard et al. 1993). It also allows specialist personnel to focus their available clinical time on counseling (Sparks 1972), aural rehabilitation (Wood et al. 1973), and pediatric work (Margolis & Morgan 2008; Ho et al. 2009), all of which are better delivered in a face-to-face client–audiologist interaction. Automatic audiometry has particular application in self-fitting hearing aids (HAs), which are intended to be assembled, programmed, fine-tuned, and managed entirely by the user (Convery et al. 2011). As the programming step necessitates the administration of a hearing test conducted through the HA itself, self-fitting devices require the capability to accurately and automatically perform these measurements (Keidser et al. 2011).

Several studies have compared the test–retest reliability of automatic audiometry with that of manual audiometry and concluded that the automatic procedure has superior repeatability (Jervall et al. 1983; Ho et al. 2009; Margolis et al. 2010; Swanepoel et al. 2010; Van Tasell & Folkeard 2013). This is primarily thought to be due to the transfer of responsibility for complex calculations and judgments from the clinician to the computer, thus eliminating the effects of human error or bias (Campbell 1974; Swanepoel et al. 2010). Developers of early automatic audiometers chose to implement computerized versions of existing manual procedures for sensible reasons, not the least of which was the basic level of computer technology available at the time (Sakabe et al. 1975, 1978). Modern computing power, however, allows the implementation of highly sophisticated threshold-seeking paradigms with more precise control and faster decision-making procedures than those used by human clinicians. There are many factors not specified by the standards for manual audiometry that are more easily controlled

by a computer than by a human being, all of which have the potential to affect the reliability or validity of the test results. These include stimulus duration, interstimulus interval, step size, the time window in which a response is accepted, the number of reversals that constitute a threshold, and whether these characteristics should be varied or held constant within a single test (Harris 1979; Jerlvall et al. 1983).

Many of the investigations into the validity and reliability of automatic audiometry, particularly the early studies, examined AC thresholds only. However, of those studies that also obtained BC data, all found that both its validity (Wood et al. 1973; Ho et al. 2009; Margolis et al. 2010) and reliability (Ho et al. 2009; Swanepoel & Biagio 2011) were lower than that of AC, particularly for low frequencies. Reported reasons for these findings include the effect of forehead versus mastoid location of the bone vibrator, variability in the amount of static force applied by the bone vibrator, the relative admittance and impedance of the middle ear system, and the distortion by the bone vibrator of low-frequency stimuli (Margolis et al. 2010; Swanepoel & Biagio 2011). In these studies, a clinician or trained layperson, not the research participant, was responsible for placing the bone vibrator. While this likely played a role in reducing test-retest variability, it could also be argued that the involvement of a professional rendered the procedure nonautomatic. That is, to be considered *truly* automatic, the entire procedure, including placement of the transducer, would have to be completed by the test participant alone. Given the variability inherent in BC measurement even with the involvement of a skilled professional, requiring the test participant to self-place the bone vibrator would likely serve to decrease the reliability of the test results. Further, in the context of automatic in situ audiometry, in which the test stimuli are generated and delivered to the ear by an HA, the idea of BC testing with a bone vibrator is impractical, if not impossible.

Measurement and comparison of AC and BC thresholds are performed to distinguish between conductive and sensorineural sites of lesion (Silman & Silverman 1997, p. 48). Identification and quantification of the air–bone gap (ABG), an indicator of the presence and extent of conductive pathology, is a critical component of audiological assessment. The presence of a conductive component to the hearing loss, depending on etiology, may contraindicate amplification or at least dictate medical or surgical intervention instead of, or before, amplification. In the presence of situational constraints that prevent the measurement of BC thresholds, such as those described earlier, the importance of detecting a conductive component is such that alternative methods for measuring the ABG need to be explored. Existing methods for identifying the presence of a conductive component that do not rely on BC thresholds include tympanometry and such tuning fork tests as the Weber, Rinne, and Bing tests (Silman & Silverman 1997, pp. 26–27). Although valid and reliable, tympanometry and the tuning fork tests require additional test equipment. The validity of tympanometry further relies on the ability to seal the probe in the ear canal, a skill that a layperson could not reasonably be expected to possess. An ideal alternative measure of the ABG would therefore ensure that specialized equipment beyond that used for AC audiometry is not required, deliver stimuli via an AC pathway, make use of nonspeech stimuli, and have clinically acceptable sensitivity and specificity values.

The main objective of this study was to determine whether the presence and size of an ABG can be detected through the

administration of an automatic test battery comprising two AC tests, one measuring pure-tone thresholds in quiet and the other measuring the lowest signal to noise ratio (SNR) at which a pure tone can be detected in a background of spectrally and temporally modulated narrowband noise. Our hypothesis was that participants would perform differently on the combination of tests depending on their hearing loss type. The primary characteristic of hearing impairment is a loss of auditory sensitivity, a characteristic shared by losses of both sensorineural and conductive origin. On this basis, participants with sensorineural and conductive hearing loss should be distinguishable from participants with normal hearing by measuring their pure-tone thresholds in quiet. Unlike ears with conductive loss or normal hearing, which have normal cochlear function, the psychophysical tuning curves measured on ears with hearing loss of cochlear origin are significantly broader and flatter, an indication of a loss of frequency selectivity resulting from outer hair cell damage (Moore 2003, p. 117). One of the perceptual consequences of broader auditory filters and impaired frequency selectivity is a reduction in the ability to detect a signal in a background of noise, particularly at low levels. In contrast to ears with normal cochlear function, whose narrow auditory filters are able to attenuate the unwanted noise outside of a narrow band around the signal frequency, the broader filters associated with cochlear hearing loss allow more of the noise to pass through and hence interfere with detection of the desired signal (Van Tasell 1993; Moore 1996). On a test of the ability to detect a pure tone in a background of spectrally and temporally modulated narrowband noise, participants with normal cochlear filters should be able to use the temporal and spectral gaps in the noise to detect the tones at lower SNRs, which would include people with normal hearing and many people with conductive hearing loss, but would exclude most people with sensorineural hearing loss. As the success of the test battery hinges on the reliability and validity of its individual components, a secondary aim was to evaluate the suitability of a novel automatic audiometry algorithm for potential incorporation into the battery.

PARTICIPANTS AND METHODS

Participants

Data were collected from 120 ears of 64 participants, 31 male and 33 female. Eight participants were tested in one ear, as their other ear was outside the dynamic range of one or both automatic tests. Data were collected from both ears of the remaining 56 participants. Participants ranged in age from 18 to 81 years, with a median age of 62 years ($SD = 20$ years). Participants were selected for inclusion in the study if their pure-tone AC thresholds were less than or equal to 80 dB HL (hearing level) at 500, 1000, 2000, and 4000 Hz, an upper limit set due to the output limitations of the test equipment. Ears were given a single overall classification as normal hearing (NH), conductive component (CC), or sensorineural hearing loss (SNHL) on the basis of manual AC and masked BC audiometry at the four test frequencies. Forty ears met the criteria for NH, which we defined as AC thresholds less than or equal to 20 dB HL at all four test frequencies. Categorization of loss type for the ears with hearing loss was based on the ABG size at the majority of test frequencies, rather than at all the test frequencies. Ears with AC thresholds greater than or equal to 25 dB HL at two or more of the four test frequencies

and an ABG of less than 15 dB at three or more test frequencies were categorized as SNHL, a group that included 44 ears. The remaining 36 ears were classified as CC and included both pure conductive losses (6 ears) and those with mixed loss (30 ears). To be placed in the CC group, ears were required to have an ABG greater than or equal to 15 dB at two or more of the test frequencies and thresholds that had been stable (i.e., not fluctuating or progressively deteriorating) for at least the previous 6 months according to self-report. Fifteen of the bilaterally tested participants had different loss types in each ear.

Travel costs incurred by the participants were reimbursed on completion of the study appointment. The treatment of participants in this study was approved by the Australian Hearing Ethics Committee and conformed in all respects to the Australian government's National Statement on Ethical Conduct in Human Research.

Procedure

Participants underwent a series of audiometric assessments: (1) pure-tone AC and masked BC audiometry, administered manually by an audiologist with 7 years of clinical experience; (2) automatic, computerized pure-tone AC audiometry; and (3) an automatic, computerized tone-in-noise (TIN) detection test. The two automatic tests were each administered twice, for a total of five hearing assessments. The order of assessment was randomized according to a balanced Latin square. Testing was conducted in an audiometric booth that adhered to ANSI standard S3.1-1999 (R2008).

All AC stimuli were delivered to the test ear through an E-A-R 5A insert earphone (Aearo Technologies, a Division of 3M, St. Paul, MN). For manual audiometry, the tones (and masking noise, where appropriate) were generated by an Interacoustics AC40 clinical audiometer (Interacoustics AS, Assens, Denmark). Manual BC stimuli were presented via a Radioear B-71 bone conductor (Radioear, New Eagle, PA) with the ears unoccluded. The Carhart correction factor (Dillon 2012, p. 319) was applied to the BC thresholds of the four ears with confirmed otosclerosis to allow for the effect of otosclerosis-related stapes fixation on BC thresholds (Table 1). For automatic audiometry, the signals originated from .wav files stored on a laptop computer and were presented via an Edirol FireWire Audio Capture FA-101 external sound card (Roland Corporation, Shizuoka, Japan). The manual and automatic audiometers were both calibrated according to ISO standard 389-2 (ISO 1994). For the TIN test, the stimuli originated from individual tone and noise .wav files stored on the laptop computer and were presented via a chain of equipment that included the external sound card, a Technics SU-7300 stereo integrated amplifier (Panasonic Corporation, Osaka, Japan), a purpose-built remotely controlled digital attenuator, and a fader. A voltmeter was also included in the equipment chain to monitor the output of the

TABLE 1. The correction factors that were subtracted from the BC thresholds of the four ears with otosclerosis

Frequency (Hz)	Correction Factor (dB)
500	5
1000	10
2000	13
4000	6

insert earphone. The TIN test was calibrated with an artificial ear using white noise as the calibration stimulus. The levels of the narrowband test stimuli were set to match the average root-mean-square broadband level of the white noise.

The automatic audiometry implementation used a stimulus consisting of a train of three tone bursts, each of which was 290 msec in duration with 30 msec on- and offset ramps and separated by 140-msec gaps. The interstimulus intervals were of random duration and ranged from 1000 to 4600 msec. Test parameters were chosen based on existing automatic audiometry implementations and subsequent informal adjustments made during in-house experimentation. Participants were instructed to respond to the tone bursts by pressing a button on a numeric keypad: a response was considered valid if it occurred within a 1.5-second time window commencing from the onset of the stimulus. The threshold-seeking procedure included two phases. In phase 1, a 10 dB up/down step size was used to zero in on the likely threshold range. Phase 1 ended when the first nonresponse to a stimulus presentation was recorded. In phase 2, the first nonresponse to a stimulus triggered a 10 dB increase in the level of the next stimulus presentation. Subsequent non-responses resulted in an increase in stimulus level in 5 dB increments until a positive response was recorded. Positive responses to the stimuli always resulted in a 5 dB decrease in stimulus level. Phase 2 ended when a standard error of ≤ 2.5 dB was reached, or a maximum of four reversals was recorded. If the standard error criterion was met, the threshold was calculated by averaging the level of each presentation for which a positive response was obtained that was immediately before a presentation for which no response was obtained. If the maximum number of reversals criterion was met, a trimmed mean (i.e., removal of the highest and lowest values before averaging the remaining values) of the positive responses was calculated to determine the threshold. The presentation levels of the automatic audiometer ranged from 0 to 80 dB HL.

The TIN test was originally developed and validated at the National Acoustic Laboratories for use in a national telephone-based hearing screening service. Results of the initial TIN test validation indicated high intrasubject repeatability, and the measured SNRs were found to correlate significantly with pure-tone thresholds (Zhou et al. 2009). The aim of the TIN test is to determine the lowest SNR at which the listener is able to detect a train of three tone bursts in a continuous background of temporally and spectrally modulated narrowband noise. Each tone burst was 450 msec in duration with 20-msec on- and offset ramps and was separated from the other tone bursts by 136-msec gaps. The interstimulus intervals were of random duration and ranged from 1000 to 5000 msec. All noises used in this test were temporally modulated with a 20 Hz sinusoidal envelope that had a peak-to-trough ratio of 10 dB. Single .wav files were used for all noise presentations of the same frequency. The spectral characteristics

TABLE 2. The spectral characteristics of the NBN used in the tone-in-noise test

Test Frequency	Passband 1	Passband 2	Gap
500	200-450	550-800	100
1000	675-925	1075-1325	150
2000	1400-1800	2200-2600	400
4000	2800-3600	4400-5200	800

All values shown are in Hz.

varied for each test frequency and are summarized in Table 2. For the 2000 and 4000 Hz masking noises, frequencies below the lower cutoff frequency of passband 1 (1400 and 2800 Hz, respectively) were attenuated by 15 dB, rather than filtered out completely, to ensure the narrowband noise did not sound too tonal. Responses to the stimuli were effected by pressing a key on a numeric keypad; a response was considered valid if it occurred within a 1.3-second time window. The test was executed in two phases, using an SNR-seeking algorithm similar to that used by the automatic audiometer. In phase 1, a 4 dB decrement step size was used. Phase 1 ended when the first nonresponse to a stimulus presentation was recorded. Phase 2 used a 2 dB decrement step size. After a positive response, the first nonresponse resulted in a 5 dB increase for the next stimulus presentation. Subsequent nonresponses triggered 2 dB increases in stimulus level until a positive response was registered. Phase 2 ended when a fixed number of four reversals was reached. The SNR was calculated by averaging the level of each presentation for which a positive response was obtained that was immediately before a presentation for which no response was obtained. The SNR range of the TIN test was -30 to +10 dB.

Before the participant completed the TIN test at each frequency, he or she was instructed to use the remotely controlled attenuator to select a masking noise level that corresponded to the "loud, but ok" category on a seven-point categorical loudness scale. The noise was then fixed at that level for the duration of the test, while the level of the tones changed adaptively according to the participant's responses. The purpose of instructing the participant to select a relatively high level of noise was to ensure that the participant's achieved SNR on the TIN test was due to the masking effects of the noise, rather than a lack of audibility. To confirm this, a check was performed at the end of each TIN test run. The masking noise was turned off and the tone was presented alone at a level approximately 5 dB lower than the signal level of the achieved SNR. For example, if the participant's TIN result was -1 dB SNR, with the noise presented at a level of 76 dB SPL (sound pressure level) and the signal at 75 dB SPL, audibility of the tone was checked at 70 dB SPL. If the participant was able to detect the tone, the results of the TIN test were considered valid. If the participant was unable to detect the tone, the protocol dictated that the level of the noise be increased and the test repeated. However, no participant in this study required a repeat of the TIN test. The same masking noise level was used for both test and retest.

RESULTS

Pearson's product-moment correlation analysis was conducted to determine the reliability of the automatic audiometry and TIN results. Test-retest results on both measures were found to be highly and significantly correlated at all test frequencies (automatic audiometry: $r = 0.97$ to 0.99 ; $p < 0.0001$; TIN: $r = 0.81$ to 0.94 ; $p < 0.0001$). The two test runs were therefore averaged to produce a single audiometric threshold (threshold in quiet) and a single TIN result (SNR) for each participant at each test frequency. Correlation analysis of manual and automatic audiometric thresholds, performed as a measure of validity, revealed a strong, significant correlation at all test frequencies ($r = 0.98$ to 0.99 ; $p < 0.0001$). All subsequent analyses involving thresholds in quiet make use of the automatic, rather than manual, AC thresholds.

A potential effect of cognitive changes as a result of aging on TIN test performance was investigated for the participants with sensorineural hearing loss, using multilinear regression. Age was found to significantly affect participants' achieved SNRs on the TIN test at 4000 Hz ($p = 0.006$), but not at 500, 1000, or 2000 Hz ($p = 0.14$ to 0.49).

SNRs achieved on the TIN test were plotted as a function of thresholds in quiet for each frequency. As shown in Figure 1, the SNRs achieved by the SNHL ears increase with thresholds in quiet. In contrast, many CC ears have SNRs that are similar to those achieved by the NH group despite impaired thresholds in quiet. To calculate an estimate of the expected relationship between the threshold in quiet and SNR for ears without a conductive component, the following quadratic function was fitted to the 84 data points obtained from the non-CC (i.e., NH and SNHL) ears:

$$SNR = a + b(HTL + 1)^2$$

The function was constrained to be flat at a threshold in quiet of -1 dB HL, as this was the lower limit of the automatic audiometer. Values for a and b , which varied for each test frequency, were estimated using the least squares method of linear regression. We note that the function is affected by the nonindependence of some of the data points, as most participants provided data for both left and right ears. The rationale for including both NH and SNHL ears in the calculation of the function is that hearing thresholds exist on a continuum; despite nominal cutoffs that are used clinically to categorize audiograms by degree of loss, there is no discrete point at which normal hearing suddenly becomes impaired hearing.

The measured SNR was then subtracted from the fitted SNR to yield a difference value, which was plotted as a function of the measured ABG at each test frequency (Fig. 2). Perfect agreement between the estimated and observed SNR (i.e., a difference value of zero) is shown by the dotted line. As the fitted SNRs were derived from the non-CC data, we can assume that a difference value of zero is characteristic of the average non-CC ear and thus hypothesize that a sufficient deviation from zero would be characteristic of a CC ear. The graphs in Figure 2 further illustrate the fact that as the ABG grows larger, so too does the difference between the estimated and observed SNR.

To determine what constitutes a deviation sufficient to validly indicate the presence of a conductive component, sensitivity and specificity were calculated for the actual SNR difference values at each test frequency using a range of theoretical difference or "cutoff" SNR values. Ears were classified as CC if the actual SNR difference value was greater than or equal to the cutoff value. On the assumption that the difference values for each test frequency were normally distributed within each of the CC and non-CC groups, receiver operating characteristic (ROC) curves were fitted to the data using the following formula:

$$sensitivity = 1 - \Phi\left(\left(\frac{\sigma_n}{\sigma_c}\right)\Phi^{-1}(1-x) - \left(\frac{\mu_c - \mu_n}{\sigma_c}\right)\right)$$

where Φ represents the standard normal cumulative distribution function, $x = 1 - specificity$, σ_n is the standard deviation of the difference between the actual and fitted SNR values for non-CC cases, σ_c is the corresponding standard deviation for CC cases, μ_n is the mean of the difference between actual and

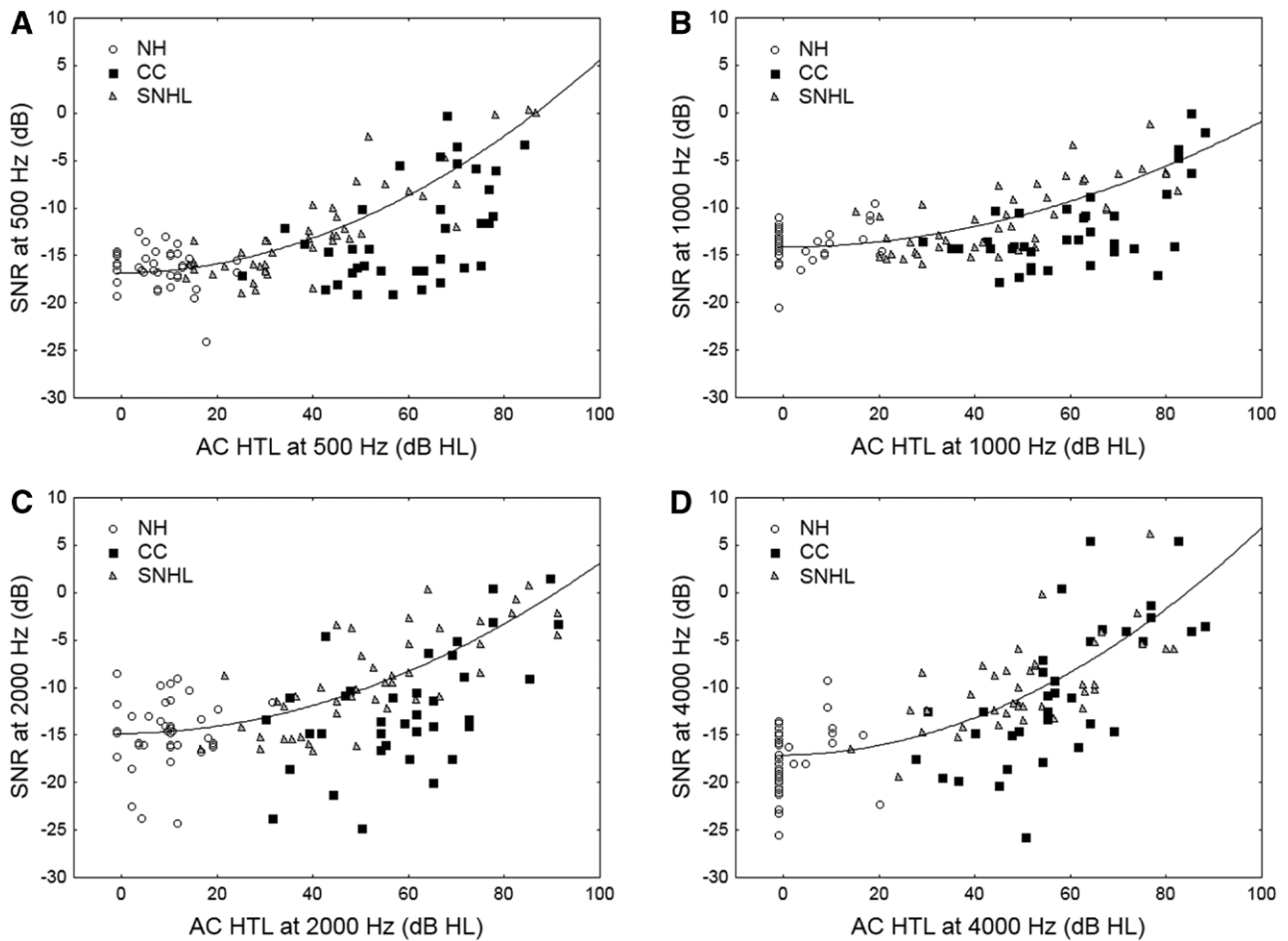


Fig. 1. The relationship between thresholds in quiet and achieved SNRs on the TIN test for the different participant groups at (A) 500, (B) 1000, (C) 2000, and (D) 4000 Hz. The lines indicate quadratic functions fitted to the data of the participants with normal hearing and sensorineural loss. AC indicates air conduction; CC, conductive component; HTL, hearing threshold level in quiet; NH, normal hearing; SNHL, sensorineural hearing loss; SNR, signal to noise ratio.

fitted SNR values for non-CC cases, and μ_c is the corresponding mean for CC cases. ROC curves show the proportion of correctly identified ABGs among CC ears as a function of incorrectly identified ABGs among non-CC ears.

Figure 3 shows the ROC curves for the difference values at 500, 1000, 2000, and 4000 Hz, as well as a condition in which the difference value at 500 or 1000 Hz, whichever is greater, was used in the calculation of sensitivity and specificity. Use of the “500 or 1000 Hz” condition yields the highest sensitivity and specificity values, with sensitivity and specificity decreasing as the test frequency increases. This finding is consistent with the fact that middle ear pathology causing a conductive component to the hearing loss is more likely to result in impairments to the low and mid, rather than high, frequencies. When the CC group is categorized by ABG size and the same formula applied to generate ROC curves, the prediction of the presence of a conductive component becomes increasingly accurate as the ABG grows larger. Figure 4 illustrates this relationship for test results at 500 Hz.

DISCUSSION

The presence of an ABG can be predicted with a reasonably high degree of accuracy across ears with a variety of hearing loss types, using a combination of automatic pure-tone audiometry

and a TIN detection task. The rationale for including ears with mixed as well as conductive loss in the study was to ensure that the test battery could detect a conductive component even in the presence of sensorineural hearing loss, rather than simply distinguish between purely conductive and purely sensorineural losses. The degree of accuracy with which this can be accomplished was found to vary according to both the size of the ABG and the test frequency for which results are available. The prediction becomes more accurate with increasing ABG size and decreasing test frequency, with accuracy further improved when multiple test frequencies are taken into account. As per the ROC curves shown in Figure 3, the test battery can yield, for example, a sensitivity of 80% and a specificity of 77% for any ABG size if the threshold in quiet and SNR are known at 500 and 1000 Hz. For ABGs larger than 35 dB, sensitivity and specificity for test results, even using the results of 500 Hz alone, can increase to 98 and 80%, respectively (Fig. 4).

Determining a clinically acceptable combination of sensitivity and specificity values is beyond the scope of this article, but will depend to a large extent on the consequences of missing ears with true conductive components (false negatives) and misidentifying as conductive those ears with true sensorineural hearing loss (false positives) for a given population. A high false-negative rate could lead to significant morbidity or mortality, as some middle ear pathologies that cause conductive

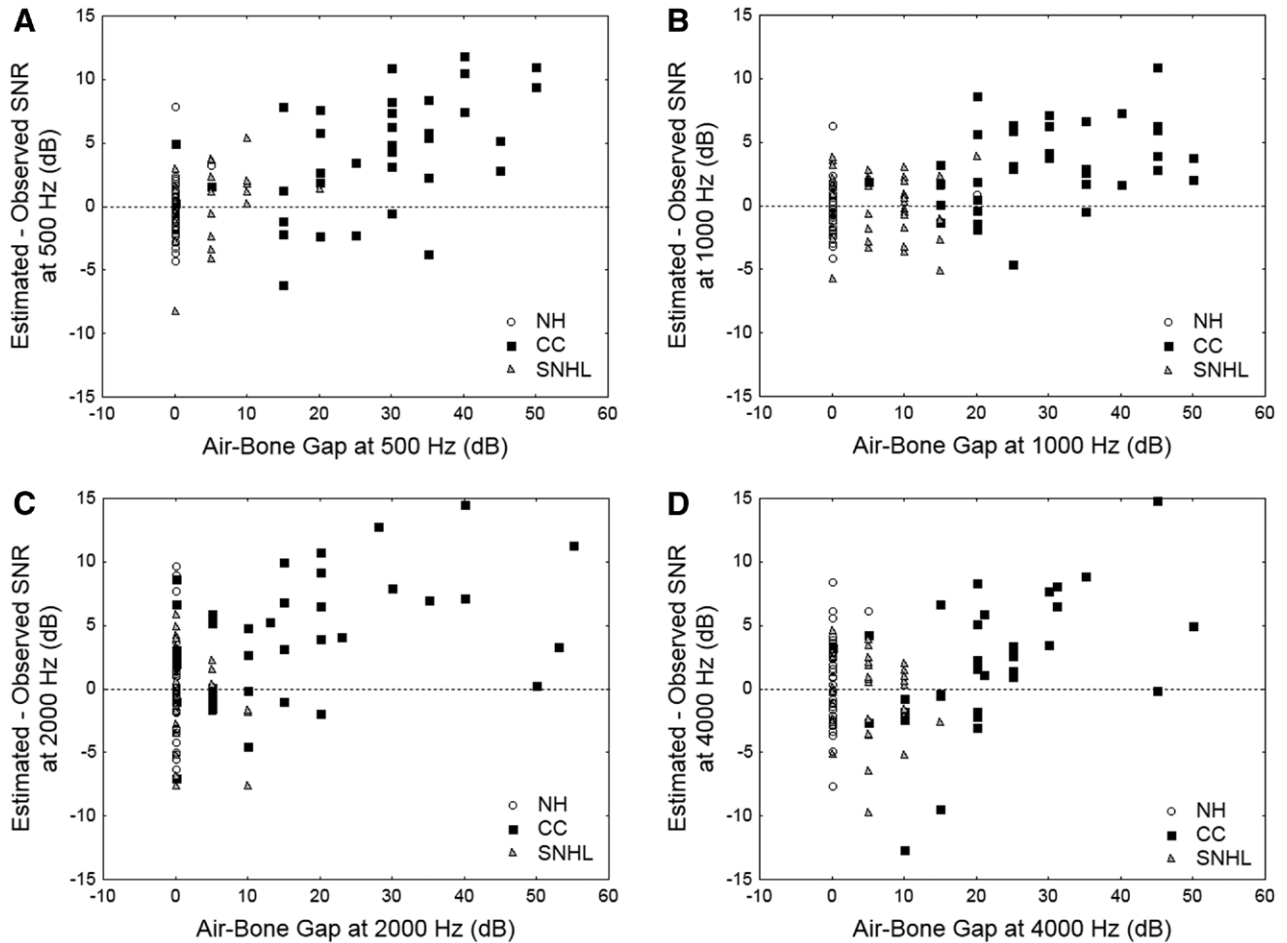


Fig. 2. The relationship between ABG size and the difference between the estimated and observed SNRs for the different participant groups at (A) 500, (B) 1000, (C) 2000, and (D) 4000 Hz. ABG indicates air-bone gap; CC, conductive component; NH, normal hearing; SNHL, sensorineural hearing loss; SNR, signal to noise ratio.

hearing loss, such as cholesteatoma and severe otitis media, can have life-threatening sequelae if left undetected and untreated (Munz et al. 1992; Spilsbury et al. 2010). However, a high false-positive rate is likely to result in needless anxiety and an unnecessary, possibly expensive, visit to a health professional. As the proportion of false negatives is reduced at the expense of false positives and vice versa, the prevalence rate of conductive hearing loss for the specific population in which the test battery will be used should be taken into account when selecting clinically appropriate sensitivity and specificity values (i.e., an acceptable difference value between predicted and measured SNR).

One of the major applications of our test battery is likely to be in a self-fitting HA, a fully self-contained personal amplification device that is designed to be assembled, programmed, and managed entirely by the user (Convery et al. 2011). In the absence of clinician involvement, a viable self-fitting HA requires, among other things, the ability to: (1) automatically obtain hearing threshold measurements such that the device's settings can be prescribed and applied; and (2) predict the presence of a conductive component, a potential contraindication to aid use. The concept of a self-fitting HA is currently being evaluated with clients in developing countries and geographically remote locations in mind, populations in which the prevalence of conductive loss is often high (e.g., Liu et al. 2001; Bowd

2005; Lasisi et al. 2007), and diagnosis and treatment are often delayed due to extremely high client-to-audiologist ratios (e.g., Klein 2000). It is likely that the high prevalence of conductive hearing loss and the potential seriousness of its attendant consequences will outweigh the costs of some users of the device being mistakenly identified as having conductive loss and thus making a needless visit to a health care professional. Similarly, conductive hearing loss tends to be more prevalent among children than adults as a consequence of higher rates of otitis media and Eustachian tube dysfunction. An implementation of our test battery in school-based screening programs, which are often carried out in environments in which the ambient noise levels preclude accurate bone-conduction testing, should likely err on the side of wrongly identifying some children in order to capture the maximum number of children who truly do have conductive hearing loss. In determining clinically acceptable sensitivity and specificity values for these populations, it is probably safer to allow for a higher false-positive rate to keep the false-negative rate low.

Further applications of our test battery include industrial hearing conservation programs and self-administered adult hearing screenings in both clinical and home environments. In these contexts, identification of conductive components would assist in separating conductive losses from sensorineural losses

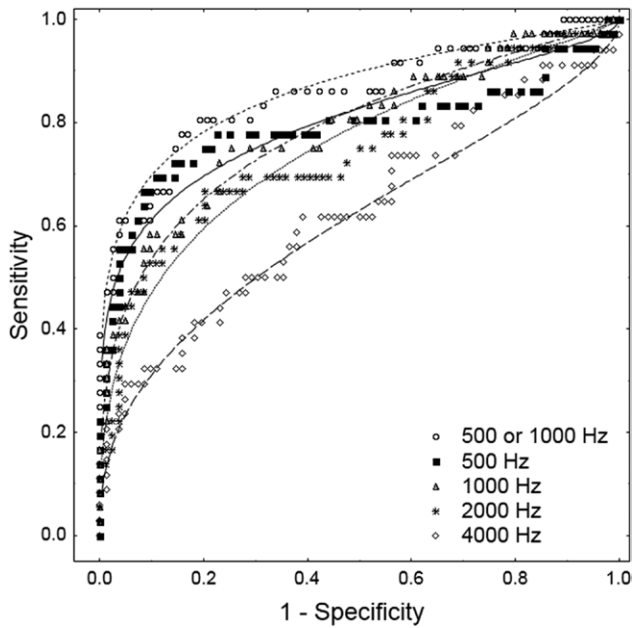


Fig. 3. ROC curves showing the combination of sensitivity and specificity values found when the difference between the estimated and observed SNRs are compared with various cutoff values. Data are shown separately for different test frequencies. ROC indicates receiver operating characteristic; SNR, signal to noise ratio.

likely to have been incurred on the job or as a result of aging. In contrast to the target population for a self-fitting HA, however, the prevalence rate of conductive hearing loss among participants in hearing conservation and self-screening programs, particularly those in developed countries, is likely to be relatively low. In these contexts, allowing for a higher false-negative rate would keep the false-positive rate acceptably low.

In any application of our test battery, it is critical that its individual components are reliable and valid, otherwise they risk compromising the reliability and validity of the battery as a whole. The reliability of our automatic audiometry algorithm compares well with recently published figures for other automatic implementations (Ho et al. 2009; Margolis et al. 2010; Swanepoel et al. 2010) and exceeds published figures for manual AC audiometry (Ho et al. 2009; Swanepoel et al. 2010), highlighting the overall advantage of computer- versus clinician-administered testing. One aspect of audiometric testing in which a human clinician can outperform a computer, however, is in making judgments about when to modify or deviate from the threshold-seeking procedure in response to the behavior of the test participant. For example, a clinician can extend the acceptable response window on the fly for a participant whose response times are slow, or reinstruct participants who appear to be guessing about the presence of the stimulus. Individual response window alteration and reinstruction of the test participant are features that could certainly be incorporated into an automatic test; however, the algorithm used in this study did not have these capabilities. Because our test battery could not take individual behavioral variations into account, the automatic audiometry algorithm used in this study was designed in an attempt to offset this limitation. The stop criterion used by our algorithm is the standard error of all positive responses; when this value drops below a particular point, the results are deemed reliable and threshold is then calculated by

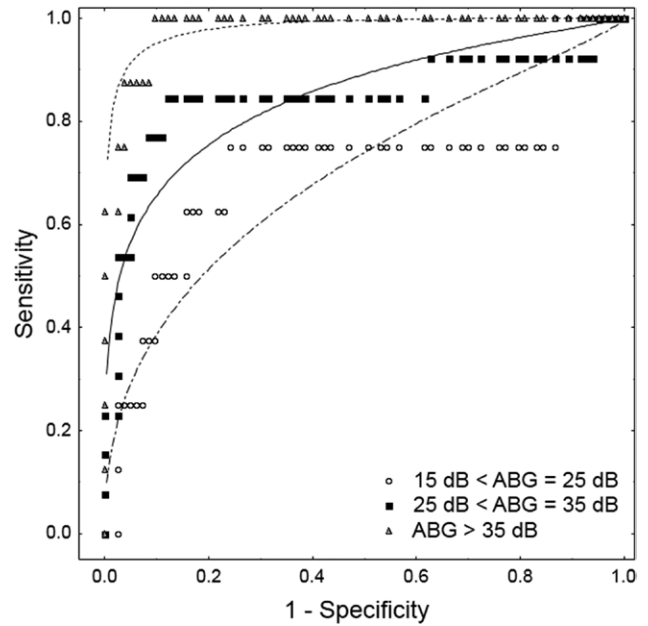


Fig. 4. ROC curves showing the combinations of sensitivity and specificity values found when the difference between the estimated and observed SNRs at 500 Hz are compared to various cutoff values. Data are shown separately for different ABG sizes. ABG indicates air-bone gap; ROC, receiver operating characteristic; SNR, signal to noise ratio.

averaging these responses across all runs. If, however, the standard error criterion is not met, testing stops after a maximum of four reversals and threshold is calculated differently: first by removing the highest and lowest response values, then by averaging the remaining values. The purpose of excluding the highest and lowest values is to minimize the effect of random responses due to guessing, fatigue, or inattention. That is, if a test participant coincidentally responds to an inaudible tone (or, conversely, responds at a suprathreshold level and then stops responding due to inattention or fatigue), these responses will not unduly influence the final threshold determination.

Before performing the TIN test at each frequency, participants were asked to select a relatively loud level at which the noise would be presented. The purpose of this was to ensure that the achieved SNR was due to the masking effects of the noise, rather than a lack of audibility. As the noise levels were not fixed at a particular SPL or sensation level, it is possible that variability in masker level across the participant group influenced the results. Multiple regression analyses demonstrate that when age, threshold in quiet, overall hearing loss type, and selected noise level are used as independent variables, threshold in quiet is the strongest predictor of the achieved SNR for all four test frequencies ($p < 0.0001$). Selected noise level contributed significantly to the achieved SNR at 4000 Hz only. However, as with the analyses performed on the other test frequencies, the selected noise level parameter displayed the lowest β value and the highest tolerance value, indicating that it does not add much unique predictive power when the other parameters are taken into consideration.

Although our findings suggest that the presence of an ABG can be predicted using AC tests alone, our fitted regression lines should be verified in a larger population. Ideally, the validity of our proposed test method should also be evaluated in a different

group of participants with pure conductive, pure sensorineural, and mixed hearing losses.

CONCLUSION

The presence of an ABG of sufficient magnitude can be predicted with a reasonably high degree of accuracy through the automatic administration of two AC tests: pure-tone audiometry and a TIN detection task. This prediction becomes more accurate with increasing ABG size, with decreasing test frequency, and when the results from multiple test frequencies are taken into account. The novel automatic audiometry algorithm used in this study compares well with other automatic procedures in terms of both reliability and validity. The test battery evaluated in this article is therefore suitable for use in self-fitting HAs; self-administered adult hearing screenings in the home or in a clinic; large-scale industrial hearing conservation programs; test environments in which ambient noise levels exceed the maximum permissible levels for unoccluded ears, such as schools; or any other context in which the traditional suite of audiological assessment tools is unavailable.

ACKNOWLEDGMENTS

The authors acknowledge the financial support of the HEARing Cooperative Research Centre, established and supported under the Cooperative Research Centres Program, an Australian government initiative; and the Commonwealth Department of Health and Ageing.

Preliminary results from this study were presented at an oral poster session at the 20th Audiology Australia National Conference, Adelaide, in July 2012; at the International Hearing Aid Research Conference, Lake Tahoe, in August 2012; and at Advances in Audiology: Tomorrow's Solutions for Today's Challenges, Las Vegas, in December 2012.

The authors declare no conflict of interest.

Address for correspondence: Elizabeth Convery, National Acoustic Laboratories, Australian Hearing Hub, 16 University Avenue, Macquarie Park, NSW 2109, Australia. E-mail: elizabeth.convery@nal.gov.au

Received February 13, 2013; accepted May 27, 2013.

REFERENCES

- American National Standards Institute. (2008). *Maximum Permissible Ambient Noise Levels for Audiometric Test Rooms (ANSI S3.1-1999, R2008)*. New York, NY: American National Standards Institute.
- Bowd, A. D. (2005). Otitis media: Health and social consequences for aboriginal youth in Canada's north. *Int J Circumpolar Health*, *64*, 5–15.
- Campbell, R. A. (1974). Computer audiometry. *J Speech Hear Res*, *17*, 134–140.
- Convery, E., Keidser, G., Dillon, H., et al. (2011). A self-fitting hearing aid: Need and concept. *Trends Amplif*, *15*, 157–166.
- Dillon, H. (2012). *Hearing Aids* (2nd ed.). Turrumurra, Australia: Boomerang Press.
- Harris, J. D. (1979). Optimum threshold crossings and time-window validation in threshold pure-tone computerized audiometry. *J Acoust Soc Am*, *66*, 1545–1547.
- Ho, A. T., Hildreth, A. J., Lindsey, L. (2009). Computer-assisted audiometry versus manual audiometry. *Otol Neurotol*, *30*, 876–883.
- International Standards Organization. (1994). Reference zero for the calibration of audiometric equipment, part 2: Reference equivalent threshold sound pressure levels for pure tones and insert earphones (ISO 389-2:1994).
- Jerlvall, L., Dryselius, H., Arlinger, S. (1983). Comparison of manual and computer-controlled audiometry using identical procedures. *Scand Audiol*, *12*, 209–213.
- Kam, A. C., Sung, J. K., Lee, T., et al. (2012). Clinical evaluation of a computerized self-administered hearing test. *Int J Audiol*, *51*, 606–610.
- Keidser, G., Dillon, H., Zhou, D., et al. (2011). Threshold measurements by self-fitting hearing aids: Feasibility and challenges. *Trends Amplif*, *15*, 167–174.
- Klein, J. O. (2000). The burden of otitis media. *Vaccine*, *19* (Suppl 1), S2–S8.
- Lasisi, A. O., Sulaiman, O. A., Afolabi, O. A. (2007). Socio-economic status and hearing loss in chronic suppurative otitis media in Nigeria. *Ann Trop Paediatr*, *27*, 291–296.
- Liu, X. Z., Xu, L. R., Hu, Y., et al. (2001). Epidemiological studies on hearing impairment with reference to genetic factors in Sichuan, China. *Ann Otol Rhinol Laryngol*, *110*, 356–363.
- Margolis, R. H., Glasberg, B. R., Creeke, S., et al. (2010). AMTAS: Automated method for testing auditory sensitivity: Validation studies. *Int J Audiol*, *49*, 185–194.
- Margolis, R. H., & Morgan, D. E. (2008). Automated pure-tone audiometry: An analysis of capacity, need, and benefit. *Am J Audiol*, *17*, 109–113.
- Margolis, R. H., Saly, G. L., Le, C., et al. (2007). Qualind: A method for assessing the accuracy of automated tests. *J Am Acad Audiol*, *18*, 78–89.
- Moore, B. C. (1996). Perceptual consequences of cochlear hearing loss and their implications for the design of hearing aids. *Ear Hear*, *17*, 133–161.
- Moore, B. C. J. (2003). *An Introduction to the Psychology of Hearing* (5th ed.). London, United Kingdom: Academic Press.
- Munz, M., Farmer, J. P., Auger, L., et al. (1992). Otitis media and CNS complications. *J Otolaryngol*, *21*, 224–226.
- Picard, M., Ilecki, H. J., Baxter, J. D. (1993). Clinical use of BOBCAT: Testing reliability and validity of computerized pure-tone audiometry with noise-exposed workers, children and the aged. *Audiology*, *32*, 55–67.
- Priede, V. M., & Coles, R. R. (1974). Interpretation of loudness recruitment tests—some new concepts and criteria. *J Laryngol Otol*, *88*, 641–662.
- Sakabe, N., Hirai, Y., Itami, E. (1978). Modification and application of the computerized automatic audiometer. *Scand Audiol*, *7*, 105–109.
- Sakabe, N., Ooki, T., Terai, E., et al. (1975). A computerized automatic audiometer. *Scand Audiol*, *4*, 75–81.
- Silman, S., & Silverman, C. A. (1997). *Auditory Diagnosis: Principles and Applications*. San Diego, CA: Singular Publishing Group.
- Sparks, D. W. (1972). The feasibility of computerized audiometry. *J Aud Res*, *12*, 62–66.
- Spilsbury, K., Miller, I., Semmens, J. B., et al. (2010). Factors associated with developing cholesteatoma: A study of 45,980 children with middle ear disease. *Laryngoscope*, *120*, 625–630.
- Swanepoel, D. W., & Biagio, L. (2011). Validity of diagnostic computer-based air and forehead bone conduction audiometry. *J Occup Environ Hyg*, *8*, 210–214.
- Swanepoel, D. W., Mngemane, S., Molemong, S., et al. (2010). Hearing assessment—reliability, accuracy, and efficiency of automated audiometry. *Telemed J E Health*, *16*, 557–563.
- Van Tasell, D. J. (1993). Hearing loss, speech, and hearing aids. *J Speech Hear Res*, *36*, 228–244.
- Van Tasell, D. J., & Folkeard, P. (2013). Reliability and accuracy of a method of adjustment for self-measurement of auditory thresholds. *Otol Neurotol*, *34*, 9–15.
- Wood, T. J., Wittich, W. W., Mahaffey, R. B. (1973). Computerized pure-tone audiometric procedures. *J Speech Hear Res*, *16*, 676–684.
- Zhou, D., Carter, L., Dillon, H. (2009). Telscreen III. *National Acoustic Laboratories Research and Development Report 2008/2009*, 31–32.