

Distribution Characteristics of Air-Bone Gaps – Evidence of Bias in Manual Audiometry

Robert H. Margolis¹, Richard H. Wilson², Gerald R. Popelka³, Robert H. Eikelboom^{4,5,6}, De Wet Swanepoel^{4,5,6}, and George L. Saly¹

¹Audiology Incorporated, Arden Hills, Minnesota, USA

² James H. Quillen VA Medical Center, Mountain Home, TN, USA

³ Department of Otolaryngology, Stanford University, Stanford, California, USA

⁴ Ear Science Institute Australia, Subiaco, Australia

⁵ Department of Speech-Language Pathology and Audiology, University of Pretoria, Pretoria, South Africa

⁶ Ear Sciences Centre, School of Surgery, The University of Western Australia, Nedlands, Australia

Conflicts of Interest and Source of Funding

Portions of this work were supported by grant RC3DC010986 from the National Institutes of Deafness and Other Communication Disorders and by contract No. VA118-12-C-0029 from the U.S. Department of Veterans Affairs. The Rehabilitation Research and Development Service of the U.S. Department of Veterans Affairs supported this work through the Auditory and Vestibular Dysfunction Research Enhancement Award Program (REAP) and a Senior Research Career Scientist award to the second author.

AMTAS intellectual property is owned by Audiology Incorporated (AI) in which the first and last author have commercial interests. That intellectual property may be incorporated into commercial products. The other authors report no conflicts of interest.

Correspondence:

Robert H. Margolis, Ph.D., Audiology Incorporated. 4410 Dellwood St., Arden Hills, MN 55112

Email: rhmargo001@gmail.com

(651) 639-1985 (Tel & Fax)

ABSTRACT

Objective:

Five databases were mined to examine distributions of air-bone gaps obtained by automated and manual audiometry. Differences in distribution characteristics were examined for evidence of influences unrelated to the audibility of test signals.

Design: The databases provided air- and bone-conduction thresholds that permitted examination of air-bone gap distributions that were free of ceiling and floor effects. Cases with conductive hearing loss were eliminated based on air-bone gaps, tympanometry, and otoscopy, when available. The analysis is based on 2,378,921 threshold determinations from 721,831 subjects from five databases.

Results: Automated audiometry produced air-bone gaps that were normally distributed suggesting that air- and bone-conduction thresholds are normally distributed. Manual audiometry produced air-bone gaps that were not normally distributed and show evidence of (a) inclusion of cases with conductive loss, and (b) biasing effects of assumptions of expected results.

Conclusions: Thresholds obtained by manual audiometry show bias effects from assumptions of the patient's hearing loss characteristics. Tester bias artificially reduces the variance of bone-conduction thresholds and the resulting air-bone gaps. Because the automated method is free of bias from assumptions of expected results, these distributions are hypothesized to reflect the true variability of air- and bone-conduction thresholds and the resulting air-bone gaps.

The air-bone gap (air-conduction threshold minus bone conduction threshold at a specific test frequency) plays a crucial role in the interpretation of pure-tone audiograms. The air-bone gap (ABG) can be 0 dB, positive, or negative and is used to classify audiograms as indicating conductive, sensorineural, or mixed hearing losses. In cases of normal hearing and sensorineural hearing loss, the mean ABG is expected to be 0 dB. In cases of conductive and mixed hearing losses, the ABG is expected to be positive (greater than 0 dB). Negative ABGs are usually interpreted as resulting from measurement error or from the variability inherent in air-conduction (AC) and bone-conduction (BC) thresholds.

Studebaker (1967), responding to concerns that the ABG in patients with normal middle-ear function are not always the expected value of 0 dB, explained that the inherent variability of AC and BC thresholds dictates that the ABG is a distributed variable with a variance that is dependent on the variability of the two threshold measurements that combine to form the difference. Studebaker offered that the ABG is a normally-distributed variable with a mean of 0 dB and a standard deviation of 5 dB. Based on these assumptions and a 5-dB step size, he pointed out, the ABG is expected to be 0 dB 38% of the time. Other cases scatter symmetrically around 0 dB as determined by the standard deviation and the form of the normal distribution.

Margolis (2008) pointed out that manual BC audiometry may be inherently biased. Because the audiologist often has a prior assumption of what the BC threshold should be, the measurements may be biased toward the expected threshold. That is, audiologists have an *a priori* target for the BC threshold and intentionally or unintentionally try to accommodate that target. The expected threshold may be derived from information about the patient including previous audiograms, patient history, other test results such as tympanometry and otoscopy, and AC and BC thresholds at other frequencies.

Sackett (1979) defined *bias* in analytic research as “any process at any stage of inference which tends to produce results or conclusions that differ systematically from the truth” (p. 60). He classified and described 35 forms of bias in research, two of which are particularly apropos to bias in BC testing. *Diagnostic suspicion bias* occurs when knowledge of the subject’s past history influences the outcome of a diagnostic process. The source of the knowledge could be a previous diagnosis (e.g., conductive hearing loss, otitis media) or an impression from a previous chart note. *Previous opinion bias* occurs when a previous diagnostic procedure (e.g., otoscopy,

tympanometry, a previous audiogram, or earlier threshold results within the same audiometric test) influences the administration and result of a subsequent threshold measurement. Because these biasing effects are related to the behavior of the person conducting the test they are referred to here as *tester bias*. Data obtained by automated psychophysical procedures are not subject to this form of bias.

These forms of bias may change the statistical characteristics of AC and BC thresholds, and therefore, change the distribution of ABGs. Because the automated testing method used in this investigation does not use any of these sources of information, it is not subject to the same biasing effects as manual audiometry. A comparison of the statistical characteristics of the ABG obtained with manual and automated audiometry may reveal effects of bias and explain the apparent difference in ABG variability in manual and automated testing. The availability of large databases of audiograms obtained by manual and automated testing provides the opportunity to examine ABG distribution characteristics obtained with automated and manual audiometry.

AUDIOMETRIC METHODS

Five databases were analyzed in this study, two that employed automatic audiometry and three that employed manual audiometry. Those employing manual audiometry were obtained by experienced audiologists in clinical settings using well-accepted clinical protocols. Most audiograms were obtained by supra-aural earphones. A small proportion were obtained with insert earphones. In all cases audiometers were calibrated to international and American audiometer standards. Contralateral masking was employed according to standard clinical protocols. Bone conduction was tested with the bone vibrator (Radioear B-71) on the mastoid process of the test ear with the non-test ear covered with an earphone for masking.

The databases that employed automated audiometry used AMTAS (Automated Method for Testing Auditory Sensitivity), a single-interval, forced-choice, adaptive psychophysical method with feedback. Like manual audiometry, signal levels are varied in a bracketing manner, initially using 10-dB descending steps and then ascending 5-dB steps. The use of a single observation interval allows the use of “catch trials” and the quantification of false alarms that are used to estimate the accuracy of the test. A series of validation studies demonstrated that the method produces thresholds with average values that are equivalent to those obtained by

experienced audiologists (Margolis et al. 2010; Margolis & Moore 2011; Margolis et al. 2011). An independent assessment of AMTAS (Eikelboom et al. 2013) replicated the finding of equivalence between AMTAS and manual audiometry and a meta-analysis showed good agreement between automated and manual test procedures (Mahomed et al. 2013). AMTAS employs masking based on rules that are consistent with those used in manual audiometry.

BC thresholds obtained with AMTAS are measured with the bone vibrator (Radioear B71) placed on the forehead and both ears covered by circumaural earphones (Sennheiser HDA200). Occlusion effect measurements for this arrangement indicate no effect for frequencies 500 – 4000 Hz (Madsen & Margolis, 2014). Because the circumaural earphone does not produce an occlusion effect, the procedure is equivalent to a non-occluded condition. Audiometers used for AMTAS testing were calibrated using the bone conduction RETFLs for forehead placement in the audiometer standards. Audiometers used for manual testing were calibrated using the RETFLs for mastoid placement. Anecdotal reports suggest that some users of AMTAS and other automated audiometry systems have observed that ABGs are more variable than ABGs obtained with manual audiometry.

THE DATABASES

All data were de-identified at the time of analysis, proper approvals to use the databases for research were obtained, and care was taken to abide by international standards for human subjects research (e.g. International Research Code of Ethics (1990). Bulletin of the Pan American Health Organization, 24, 604-621). The databases were mined for cases meeting certain criteria. In order to derive the distribution of the ABG without floor or ceiling effects, cases were limited to specific ranges of AC thresholds. The goal was to select cases that allowed the range of the ABG to be at least –20 to 20 dB. To accomplish this range, cases had to be excluded that had AC thresholds too low to allow a 20 dB ABG (floor effect) or too high to allow a –20 dB ABG due to the upper limit of BC stimulus levels (ceiling effect). The resulting ranges for AC thresholds for each database are given in Table 1.

Because the intent of this study is to compare ABG distributions in the absence of middle-ear dysfunction and its resulting conductive hearing loss, an attempt was made to discard

cases in which conductive components were present. To make this exclusion based on the ABG risks distortion of the natural variability. To minimize the inclusion of true conductive hearing losses without eliminating ABGs that resulted from the variability of the measurements, cases in which the ABG ≥ 10 dB at 500, 1000, and 2000 Hz were excluded. That is, the ABG had to exceed the criterion at all three frequencies for the case to be excluded. Although there is a high level of certainty that these excluded cases represent true conductive (or mixed) hearing losses, it is likely that some audiograms with milder conductive components remained in the databases. Only the Stanford database included tympanometry and otoscopy results that were used to further exclude cases with a high likelihood of conductive hearing loss. The sample sizes (number of ears) in the analysis after the exclusion criteria were applied are shown in Table 1 (parentheses).

The databases are described below. The section headings indicate automated (A) or manual (M) audiometry.

The Busselton Healthy Ageing Study (A)

AMTAS audiograms were obtained in the course of the Busselton Healthy Ageing Study, a detailed survey of the health of up to 4000 residents in the Shire of Busselton, Western Australia. (see Swanepoel et al. 2013; James et al. 2013, for descriptions of the project). All non-institutionalized residents (born between 1946 and 1964) listed on the electoral roll ($n = 6690$) and residing in the Shire are eligible to participate. Enrollment into the study is randomized, with 10% of the target sample drawn and recruited at a time. Data from the first 2023 participants were included in this study before exclusion criteria were applied. Subjects ranged in age from 45 to 66 years (mean = 56 years) at the time of testing. The audiograms were analyzed with Qualind, a method for estimating accuracy (Margolis et al., 2007). Qualind outcomes were Good in 94% of cases, Fair in 5%, and Poor in 1%.

The VAi2 Database (A)

The VAi2 database was created by a clinical trial of an automated audiometry system conducted at five community-based outpatient clinics in the Tennessee Valley Healthcare System, which is anchored at the VA Medical Center in Nashville, Tennessee. The project was funded by

Table 1. Ranges of AC thresholds in dB HL (ANSI 3.6 2010) and sample sizes (in parentheses). The first column indicates the database with (A) indicating automated audiometry and (M) indicating manual audiometry. Cases with AC thresholds outside of the indicated ranges were excluded from the analysis resulting in the indicated sample sizes (number of ears).

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz
Busselton (A)	10-20 (1,318)	10-30 (1,804)	10-40 (1,657)	10-40 (1,906)	10-40 (2,044)
V Ai2 (A)	NT*	15-25 (232)	15-40 (313)	15-40 (294)	15-45 (192)
UMH (M)	10-25 (10,614)	10-40 (14,928)	10-50 (15,151)	10-50 (17,628)	10-45 (13,305)
VA (M)	10-25 (205,544)	10-45 (741,427)	10-45 (693,036)	10-45 (474,445)	10-45 (146,781)
Stanford (M)	10-26 (4,185)	10-40 (8,810)	10-50 (8,566)	10-50 (8,471)	10-45 (6,270)

*Not Tested

Table 2. Standard deviations of air-bone gaps. The first column indicates the database with (A) indicating automated audiometry and (M) indicating manual audiometry. Values for the Busselton and V Ai2 databases are the standard deviations from the best-fit normal distributions. Values for the UMH, VA, and Stanford databases are calculated from the standard formula for standard deviation.

	250 Hz	500 Hz	1000 Hz	2000 Hz	4000 Hz	All
Busselton (A)	7.6	7.9	8.6	7.8	8.6	8.1
V Ai2 (A)	NT*	11.7	10.4	10.2	10.5	10.7
UMH (M)	12.1	14.9	15.0	10.5	10.7	12.6
VA (M)	6.3	6.5	6.1	5.9	6.3	6.2
Stanford (M)	6.3	6.0	6.4	5.5	6.0	6.0

*Not Tested

the Veterans Affairs Innovation Initiative (VAi2), now the VA Center for Innovation. Automated hearing test systems were used to perform pure-tone and speech audiometry for veterans who requested hearing evaluations. The test algorithm was identical to the automated AMTAS system used in the Busselton study except for a few minor improvements related to identifying and retesting suspicious thresholds. The VAI2 database is much smaller than the others (397 cases before exclusion criteria were applied) and provides a comparison to the Busselton data for automated testing with a different population.

Patients who were in treatment for middle-ear disease are not likely to be included in this database. The patients are predominantly male, aged 22-86 years (mean = 57 years). Qualind outcomes for all cases before exclusion criteria were applied were Good in 68% of cases, Fair in 19% and Poor in 13%. Because the Poor outcomes could significantly affect the distributions, these cases were deleted from the analysis.

The University of Minnesota Hospital Database (UMH) (M)

A database of audiograms obtained in the University of Minnesota Hospital Audiology Clinic was described by Margolis and Saly (2008). Audiograms were obtained by licensed audiologists in the normal course of clinical evaluations. A small subset were obtained by supervised graduate students. The largest source of referrals was the University of Minnesota Hospital Ear, Nose, and Throat Clinic. Other patients were referred from other clinical units, within and outside the University of Minnesota Hospital and some were self-referred. Many of these patients were seen in conjunction with a hearing-aid dispensing program.

The age distribution of the patient population was nearly constant from the first through the eighth decade of life with fewer patients in earlier and later decades (See Margolis & Saly 2008, Table 2, p. 526). About 25% of patients had normal hearing in both ears. Hearing loss severities determined by AMCLASS (Margolis and Saly, 2007) were 45% mild, 41% moderate, 8% severe, and 5% profound. The distribution of site of lesion categories was 15% conductive, 54% sensorineural, 24% mixed, and 7% sensorineural or mixed. The latter category was used when one or more AC thresholds were greater than the maximum BC stimulation level so that the ABG was undefined. These were predominantly sensorineural hearing losses. The combined prevalence of conductive and mixed hearing losses (39%) resulted in a greater representation of patients with

middle-ear dysfunction compared to the other databases, a result of the close association with the ENT clinic and probably accounts for some of the differences in ABG distributions described below. Although cases with ABGs ≥ 10 dB at 500, 1000, and 2000 Hz were excluded as described earlier, it is likely that a substantial number of patients with smaller conductive components remain in the analysis sample.

The VA Database (M)

A database of audiograms from veterans was created by the Quality Audiology and Speech Analysis and Reporting (QUASAR) Audiogram Module that is used by VA facilities across the U.S. The database is archived at the Denver Acquisition and Logistics Center (DALC), which is the unit of the VA through which hearing aids and associated devices are procured and dispensed to veterans. Audiometric data were submitted to the database at the discretion of individual clinics. Cases were unselected with regard to hearing loss characteristics and demographics but were mostly male with predominantly sensorineural hearing losses. At the time the database was accessed for this study it contained 1,000,001 audiograms. The mean age of subjects selected for the study was 66.7 years ($SD = 14.7$ years). The majority of patients were at some stage of the hearing-aid fitting process. This database has been mined in conjunction with other research projects (Wilson & McArdle 2013, 2014).

The Stanford University Database (M)

The Stanford University Medical Center audiology clinical service is part of a comprehensive tertiary care hospital outpatient clinic with referrals from the Neurotology Division, other Otolaryngology divisions, other specialty services including Oncology, Internal Medicine, and Occupational Health, external sources including independent otolaryngology and audiology practices, and self-referrals by patients with predominantly age-related hearing loss. At the time the database was accessed for this study it included audiometric records from 13,180 sequential patient visits between April 20, 2010, and November 8, 2013. The data were collected from patients seen in the normal course of diagnostic evaluations provided by Stanford licensed audiologists. Multiple audiograms from the same patient were included as well as some pure-tone audiometric data from sources outside of Stanford (<12% of the database) but entered into

the electronic medical record system for purposes of audiologic management. The database includes tympanometry and otoscopy results that provided valuable information for excluding conductive hearing losses. In addition to ABG at 500, 1000, and 2000 \geq 10 dB, exclusion criteria included low static admittance at 226 Hz (<0.4 mmho) and abnormal otoscopic findings.

Equipment and Calibration

The data analyzed for this report were collected with audiometric equipment that complies with international and American audiometer standards (ANSI S3.6-2010; IEC 389.1-1998; IEC 389.3-1994). All of the facilities from which the data were obtained maintain a regular schedule of calibration and maintenance to ensure compliance with the standards.

RESULTS

Frequency distributions (frequency of occurrence in % versus ABG) are shown in Figures 1-6. The graphs have a common aspect ratio (30% on the y-axis = 60 dB on the x-axis) to facilitate a visual comparison of distribution shapes. Frequency distributions are shown for data obtained with automated audiometry in Figures 1-3 and for data obtained by manual audiometry in Figures 4-6. The best-fit distributions in Figure 1 are replotted on each panel of Figures 4-6 to compare the shapes of the distributions obtained with manual audiometry with the best-fit normal distribution for results obtained with automated audiometry.

The Busselton Database (A)

The ABG frequency distributions from the Busselton Study are shown in Figure 1 along with best-fit normal distributions (solid curves). The best-fit normal distributions were obtained by determining the standard deviation of the normal distribution that minimized the residuals between the observed values and those from the resulting normal distribution. Those standard deviations are shown in Table 2. The peaks of the distributions are within 5 dB of the expected mean value of 0 dB except at 4000 Hz where there is a positive shift in the Busselton distribution reflecting the ABG discussed by Margolis et al. (2013). The 4000-Hz ABG is considered further in the discussion section.

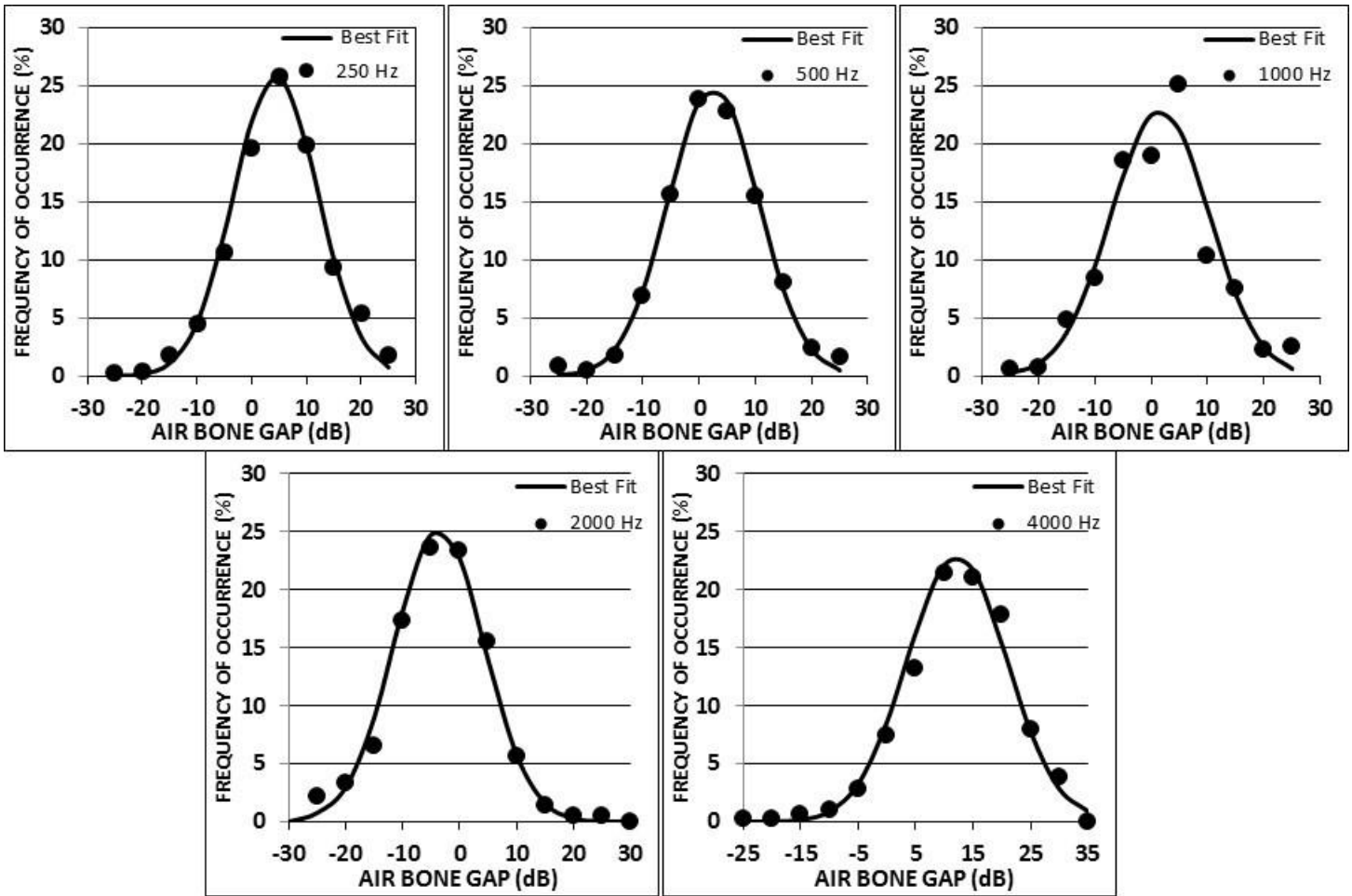


Fig. 1. Distribution of air-bone gaps at five frequencies from the Busselton study. The data were obtained by automated audiometry. Solid curves are best-fit normal distributions.

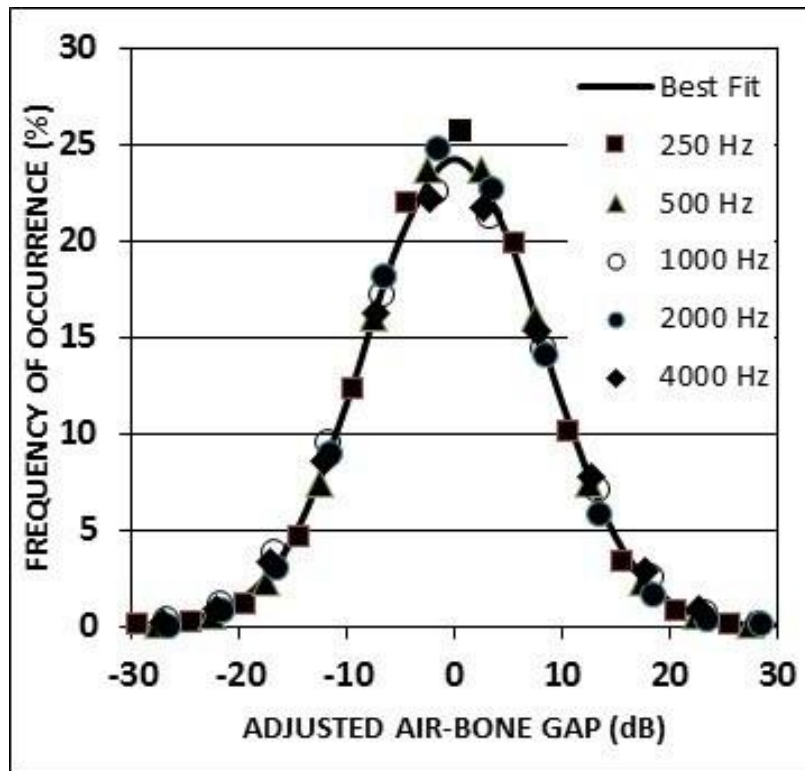


Fig. 2. Distribution of air-bone gaps at five frequencies combined from the Busselton study. The data were obtained by automated audiometry. The individual distributions from Figure 1 are normalized to a mean of 0 dB. The solid curve is best-fit normal distribution

We explored the possibility that the data for all frequencies could be combined to produce a single best-fit normal distribution. Because the means of the distributions were unequal, it was necessary to adjust the means of the best-fit distributions at individual frequencies to 0 dB. The solid curve in Figure 2 is a normal distribution with a mean of 0 dB and a standard deviation equal to the average standard deviation of the best-fit distributions in Figure 1. The single normal distribution fits the combined data quite well when the differences between means are removed (Figure 2). The fact that the composite data are well fit by a single normal distribution, suggests that the ABG has the same variance at all test frequencies.

The VAI2 Database (A)

The VAI2 database provides a second data set obtained with automated audiometry for comparison to the Busselton data. The VAI2 data were obtained with nearly identical testing software at a different site with a different population of subjects. Because the dataset is much smaller than those from the other databases (see Table 1), it was necessary to combine the data across frequencies to characterize the ABG distribution. The distributions of ABGs from the VAI2 database are shown in Figure 3 for all frequencies combined after normalizing for the differences between means. The best-fit normal distribution (solid curve) has a standard deviation of 10.7 dB in contrast to the 8.1 dB standard deviation for the best-fit normal distribution for the Busselton data in Figure 2. The greater variability in the VAI2 data compared to the Busselton data may result in part from the smaller sample size but probably reflects a true difference in the variability for the different populations from which the samples were drawn. The higher rates of Fair and Poor Qualind outcomes relative to the Busselton study supports the hypothesis that differences in the distributions are related to population characteristics. Population characteristics that could contribute to unequal variance include hearing loss characteristics, age, general health, otologic health, and many other possible sources. The results in Figure 3 suggest that the ABGs obtained with an unbiased method are normally distributed but may differ in variance for different populations.

The University of Minnesota Hospital (UMH) Database (M)

The distributions of ABGs from the University of Minnesota Hospital database, shown in

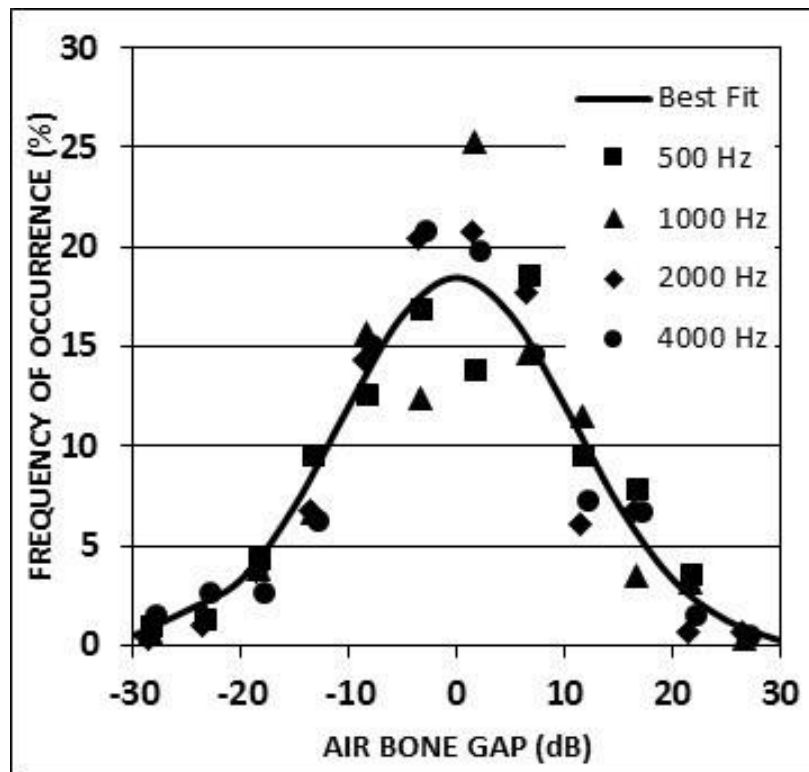


Fig. 3. Distribution of air-bone gaps at four frequencies combined from the VAI2 study. The data were obtained by automated audiometry. The individual distributions are normalized to a mean of 0 dB. The solid curve is the best-fit normal distribution

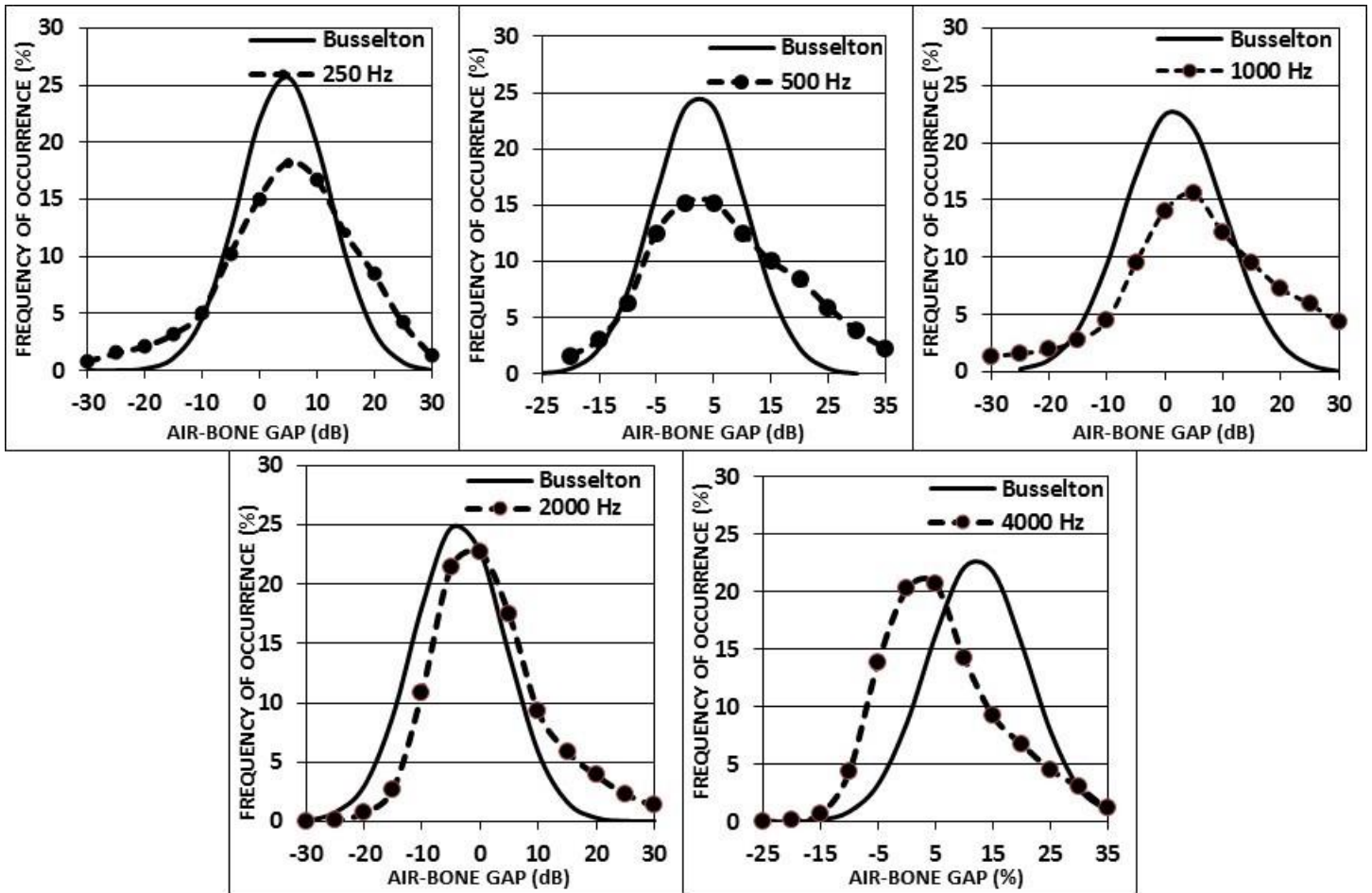


Fig. 4. Distribution of air-bone gaps at five frequencies from the University of Minnesota Hospital database (circles). The data were obtained by manual clinical audiometry. The solid curves are the best-fit normal distributions from Figure 1

Figure 4, differ from the Busselton data in two ways. First, at all five frequencies, the shapes of the UMH distributions are broader than the Busselton distributions. Second, whereas the Busselton distributions are normal in form the UMH distributions are positively skewed. The broader distributions and positive skew result in the greater standard deviations for the UMH distributions in comparison to the Busselton data (Table 2). The positive skew indicates a relative under-representation of negative ABGs and an over-representation of positive ABGs relative to the normally-distributed Busselton distributions (solid curves in Figure 4).

The VA Database (M)

The distributions of ABGs from the VA database are shown in Figure 5 along with the best-fit Busselton distributions (solid curves). The VA distributions are substantially narrower (leptokurtic) than the Busselton distributions and have a slight positive skew indicating a greater representation of positive ABGs than negative ABGs. The peak locations of the two distributions closely approximate one another at 250-2000 Hz, but at 4000 Hz the peaks of the two distributions differ due to the positive shift in the Busselton distribution. The standard deviations of the VA distributions are systematically smaller than those from the Busselton data (Table 2), possibly due to the leptokurtic form, which is further considered in the discussion section.

Stanford University Database (M)

The distributions of ABGs from the Stanford University database are shown in Figure 6 along with the best-fit normal distributions from the Busselton study (solid curves). The Stanford distributions closely resemble the VA distributions, with both sets of distributions being substantially narrower (leptokurtic) than the Busselton distributions and positively skewed. Like the VA distributions, the Stanford distributions have systematically smaller standard deviations than the Busselton distributions (Table 2).

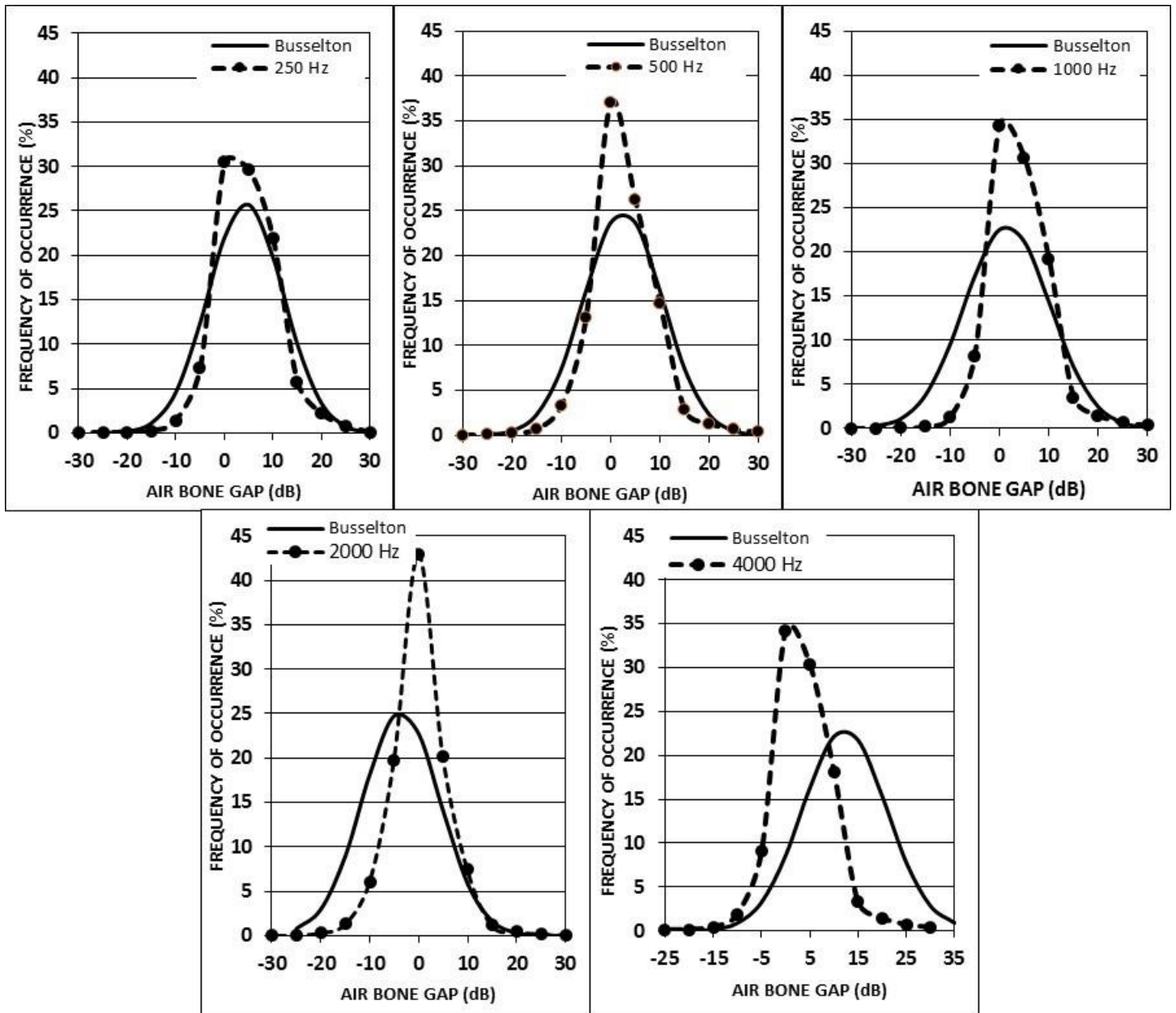


Fig. 5. Distribution of air-bone gaps at five frequencies from the VA database (circles). The data were obtained by manual clinical audiometry. The solid curves are the best-fit normal distributions from Figure 1

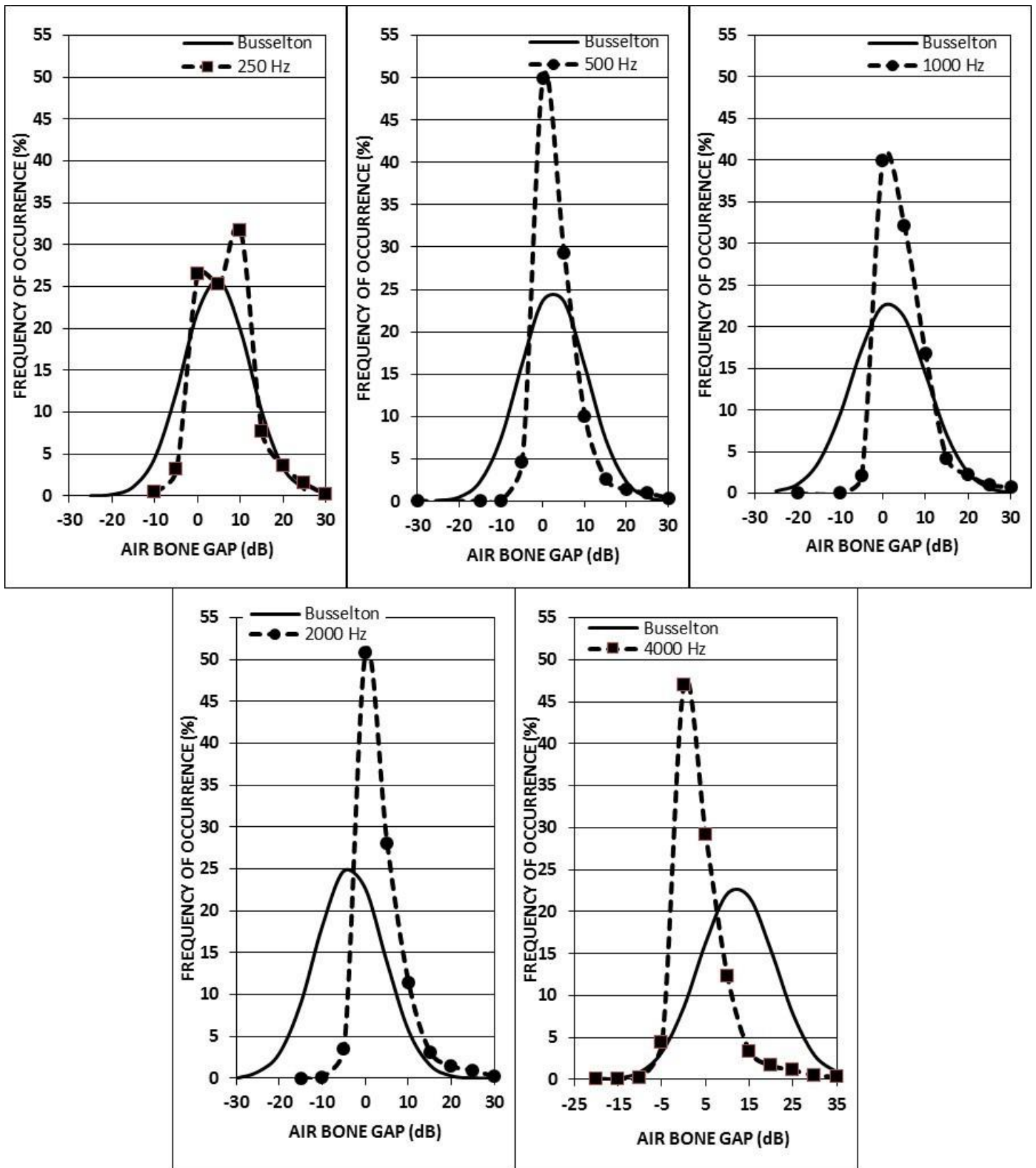


Fig. 6. Distribution of air-bone gaps at five frequencies from the Stanford University database (circles). The data were obtained by manual clinical audiometry. The solid curves are the best-fit normal distributions from Figure 1.

DISCUSSION

Distributed Nature of Air-Bone Gaps

Studebaker's analysis presciently hypothesized that the distribution of ABGs is a normal distribution with the combined variance of AC and BC thresholds (Studebaker 1967). If the AC and BC thresholds are independent, the ABG variance is the simple sum of the AC and BC threshold variances. AC and BC threshold variability have components that are independent (e.g., the variance of the different transmission pathways to the cochlea) and shared components (e.g., the variance of the sensorineural pathways). The ABG variance may not be the simple sum of the component variances but is nevertheless a distributed variable that depends on the AC and BC threshold variances. A difference in ABG distributions for automated and manual testing is expected if one method is affected by sources of bias that do not affect the other method. A prior assumption of what the BC threshold should be would affect both BC and ABG distributions in manual testing.

The ABG SD that Studebaker assumed (5 dB) is close to the 6 dB value obtained in the VA and Stanford databases (Table 2). Studebaker pointed out that a 5 dB SD results in 38% of ABGs being equal to 0 dB (assuming a 5-dB step size). Based on the 6 dB value from the VA and Stanford databases, the value drops to 32%. Because the VA and Stanford distributions are subject to tester bias as described below, the proportion of subjects with 0 dB ABGs is probably inflated relative to the unbiased variability of AC and BC thresholds. When the tester bias is removed through automated testing that does not make use of prior information, the SD of ABG rises to 8.1 dB and the proportion with 0 dB ABGs drops to 24%. The VAI2 data show a SD for ABG of 10.7 dB resulting in only 19% of ABGs equal to 0 dB. The implication is that users of automated systems will have to reset their expectations to reflect the true variability of threshold measurements unaffected by tester bias.

The normal shape of the ABG distributions suggests that AC and BC threshold distributions are normal but this has not been proven in the current study. A separate analysis of AC distributions from normal-hearing subjects in the Busselton database indicated that AC distributions obtained with automated audiometry were normal in form and those obtained with manual audiometry showed a small but significant positive skew (Margolis et al. 2015).

Departure of a distribution from the normal form can be quantified by measures of skewness and kurtosis. The skewness of a distribution is given by

$$\text{Skewness} = \frac{n}{(n-1)(n-2)} \sum \left(\frac{x_j - \bar{x}}{s} \right)^3 \quad (1)$$

where n is the number of cases, x is the value for each case, \bar{x} is the mean, and s is the standard deviation. (see Doane & Seward 2011.) A skewness value of 0 indicates a symmetrical distribution. Negative values indicate negative skew, i.e., over-representation of negative values. Positive values indicate over-representation of positive values. Skewness values can be compared by calculating standard error of skewness and converting the standard error to a confidence interval. The standard error of skewness (SE_s) is given by

$$SE_s = \sqrt{6/n} \quad (2)$$

where n is the number of cases (Tabachnick & Fidell, 1996).

The 90% confidence interval (95th %ile – 5th %ile) is given by

$$90\% \text{ C.I.} = \bar{x} \pm 1.96 SE_s. \quad (3)$$

The top panel of Figure 7 shows mean skewness values, averaged over all test frequencies, for the five databases with their 90% ranges (vertical bars). Skewness values for the databases obtained by automated audiometry (Busselton and VAi2) are near zero with slight negative skews. The near zero values are consistent with the normal distributions shown in Figures 1-3. The overlapping 90% ranges indicate that there is no significant difference in skewness between the Busselton and VAi2 distributions. The manual audiometry databases (UMH, VA, and Stanford) show significant levels of positive skew, indicating an under-representation of negative ABGs. The non-overlapping 90% ranges indicate significant differences in the degree of skewness between the three databases.

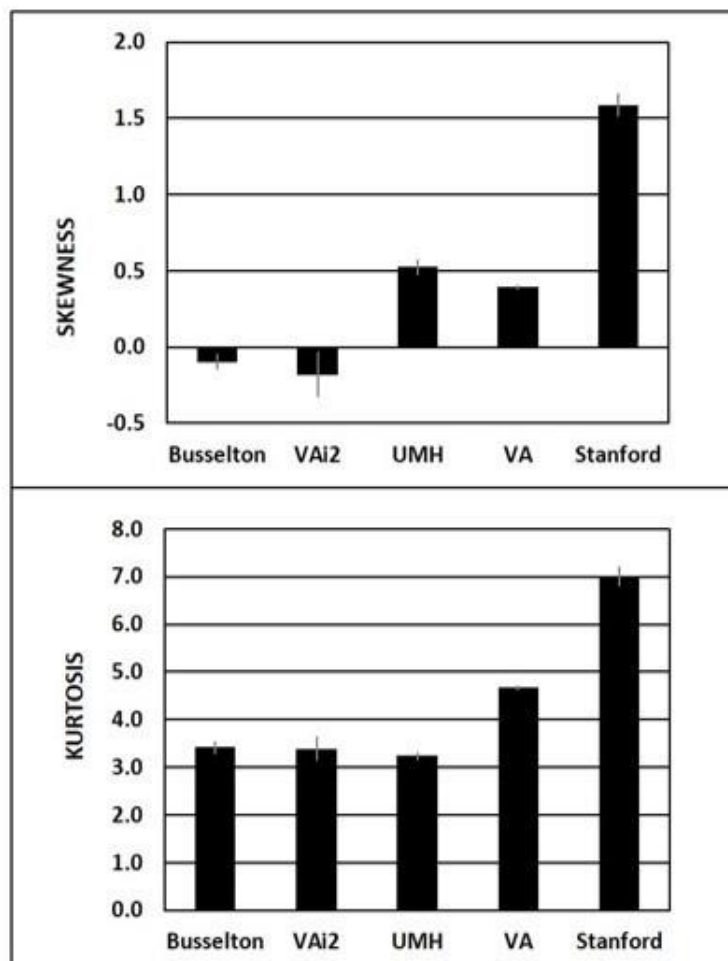


Fig. 7. Skewness and Kurtosis statistics for each database. The black vertical bars are 90% ranges (95th to 5th percentiles)

Kurtosis is the degree to which a distribution is peaked or flat relative to a normal distribution. The kurtosis statistic is given by

$$\text{Kurtosis} = \frac{\sum (x_i - \bar{x})^4}{(n-1) s^4} \quad (4)$$

where n is the number of cases, x is the value for each case, \bar{x} is the mean, and s is the standard deviation (see National Institute of Standards and Technology 2013). A normal distribution has a Kurtosis value of 3.0. Values greater than 3.0 indicate that the distribution is more peaked (leptokurtic) relative to a normal distribution. Values less than 3.0 indicate that the distribution is less peaked (platykurtic) relative to a normal distribution. Kurtosis values can be compared by calculating standard error of kurtosis and converting the standard error to a confidence interval. The standard error of kurtosis (SE_k) is given by

$$SE_k = \sqrt{(24/n)} \quad (5)$$

where n is the number of cases (Tabachnick & Fidell, 1996).

The 90% confidence interval (95th %ile – 5th %ile) is given by

$$90\% \text{ C.I.} = \bar{x} \pm 1.96 SE_k. \quad (6)$$

Kurtosis values for the five databases are shown in the lower panel of Figure 7. The values for the Busselton, VAi2, and UMH databases are close to the value of 3.0 that is expected for normal distributions. The overlapping 90% ranges indicate that there are no significant differences among the three databases. The larger values for the VA and Stanford manual audiometry databases indicate leptokurtic distributions for those databases. The non-overlapping 90% ranges indicate that the distributions from the Stanford database are more leptokurtic than those from the VA database.

If AC and BC threshold variances are normally distributed and independent, then the ABG distribution is expected to be normally distributed. AC threshold obtained with automated

audiometry for normal-hearing subjects were shown to be normally distributed by Margolis et al. (2015). Because AC and BC signal pathways are partially independent (the different pathways to the cochlear) and partially shared (the sensorineural pathways), a departure from the normal form of the ABG distributions would not be surprising. However, the ABG distributions from the Busselton and VAI2 databases are strikingly normal in form (Figures 1-3) and do not show significant levels of skewness and kurtosis (Figure 7). This suggests that the departure from the assumption of independent variance is not sufficient to affect the form of the ABG distribution in any substantial way. The Studebaker (1967) formulation of the dependence of ABG variance on AC and BC threshold distributions, and the normal form of all three distributions are supported by these results for thresholds obtained by a bias-free automated method.

Influence of Conductive Hearing Losses

Studebaker (1967) predicted that in the absence of conductive transmission loss the form of the ABG dispersion should be a normal distribution. The presence of conductive hearing loss would positively skew the distributions toward positive ABGs. Although the Busselton database did not include other tests of middle-ear function, the population (45-65 year-old adults from a community sample) probably has a very low representation of conductive components. This explains the symmetrical, normal distribution patterns in Figures 1 and 2. The VAI2 database was composed of veterans whose complaints were primarily related to hearing loss and many expressed interest in hearing aids. They were not referred by otolaryngology clinics for assessment of hearing loss related to middle-ear disease. Like the Busselton database, the VAI2 database probably had a low representation of conductive components, producing the symmetrical distribution in Figure 3 which is reasonably well described by the best-fit normal distribution.

In the samples analyzed from all databases, cases were excluded that had $ABG \geq 10$ dB at 500, 1000, and 2000 Hz. This exclusion probably eliminated the most significant middle-ear involvements but may not have excluded patients with milder conductive components. The Stanford database included tympanometry and otoscopy data for every patient. Patients with abnormal tympanograms and abnormal clinician-reported otoscopy were excluded. Thus the Stanford database was probably the least influenced by conductive components.

The UMH database probably had the highest proportion of conductive hearing losses. Although the largest conductive components were eliminated ($ABG \geq 10$ dB at 500, 1000, and 2000 Hz), it is likely that many mild conductive hearing losses remained. Unlike the Busselton, VAI2, and VA databases, the patient population included substantial numbers of patients who were being treated and followed for middle-ear disease. Because the database included repeat audiograms, these patients were probably well-represented in the sample. A report on the distributions of hearing loss characteristics from this database indicated that 39% of the audiograms in the database were categorized as conductive or mixed by AMCLASS. The higher proportion of conductive components in the UMH database probably accounts for the differences in distribution characteristics between that database and the VA and Stanford databases. The greater representation of positive ABGs relative to the Busselton distributions probably results from the inclusion of subjects with conductive components. The VA database includes the test results of veterans who are primarily in the hearing-aid assessment and fitting process. This population is not likely to include a high proportion of conductive components. The Stanford database provides the most controls for eliminating conductive hearing losses. The similarity between the Stanford and VA distributions suggests that both had either a virtual absence or a similar representation of conductive components.

Evidence of Bias

Bias is an unavoidable consequence of prior knowledge of the results of a diagnostic or treatment procedure (Sackett, 1979). Clinical trials are conducted in a double-blind fashion because prior knowledge of results by the investigator or research subject necessarily influences the results. The issue has been discussed in optometry which has adopted automated perimetry to eliminate tester bias (Pineles et al. 2006).

Hipskind and Rintelmann (1969) reviewed the evidence for tester bias in a variety of disciplines and concluded that “most of the experiments reported to date have revealed that these phenomena do in fact occur and their occurrence is seemingly beyond the conscious control of the experimenter” (p. 298). Audiologists who consciously or unconsciously “adjust” the results to convey the “correct” message are doing so in the interest of the patient. But the expectation that the ABG should be 0 dB influences the audiologist’s procedure in a way that biases the

result toward that expectation. Forcing the results toward the expected value results in an artificial decrease in the variance of the measurements, a process referred to as the Bias-Variance Tradeoff by Fortmann-Roe (2012). The smaller SDs for the VA and Stanford databases relative to the Busselton distributions (Table 2) probably result from the effect of bias on variance.

The strikingly normal form of the distributions from the Busselton and VAI2 databases are consistent with the lack of bias inherent in the automated method in which there is no prior knowledge that can influence either AC or BC thresholds. We believe that the normal ABG distributions in Figures 1-3 reflect the true behavioral variability of the populations that were sampled, unaffected by diagnostic suspicion bias and previous opinion bias. The VAI2 distributions in Figure 3 are similar to the Busselton distributions when the greater variability of the population that was sampled is taken into account. The normal form of these distributions was hypothesized by Studebaker (1967) five decades ago.

The UMH data show a tendency toward a bias against negative ABGs, particularly at 500, 1000, and 4000 Hz (Figure 4). The greater variability of the ABG distributions relative to the Busselton distributions (Table 2) probably results from the inclusion of cases with conductive and mixed hearing losses.

The ABG distributions from the VA database (Figure 5) are substantially different in form and in overall variance from the Busselton distributions. Both negative and positive ABGs are less prevalent in the VA data. Because these hearing losses are predominantly sensorineural, and in most cases the testing audiologist knows that, a bias against both negative and positive ABGs is expected. As a result there is a significant decrease in the overall ABG variance (Table 2) due to the bias-variance tradeoff.

The ABG distributions from the Stanford database (Figure 6) are very similar to those from the VA database and the overall variances are very similar for the two data sets (Table 2). The reduced variance relative to the Busselton distributions is another manifestation of the bias-variance tradeoff.

If prior knowledge of test results has a biasing effect on measurements, then it is likely that the bias can be differentially influenced by the details of that knowledge. We explored the possibility that the biasing effect could bear a relationship to the magnitude of the hearing loss indicated by the AC threshold. Figure 8 shows distributions of 1000 Hz ABGs from the VA

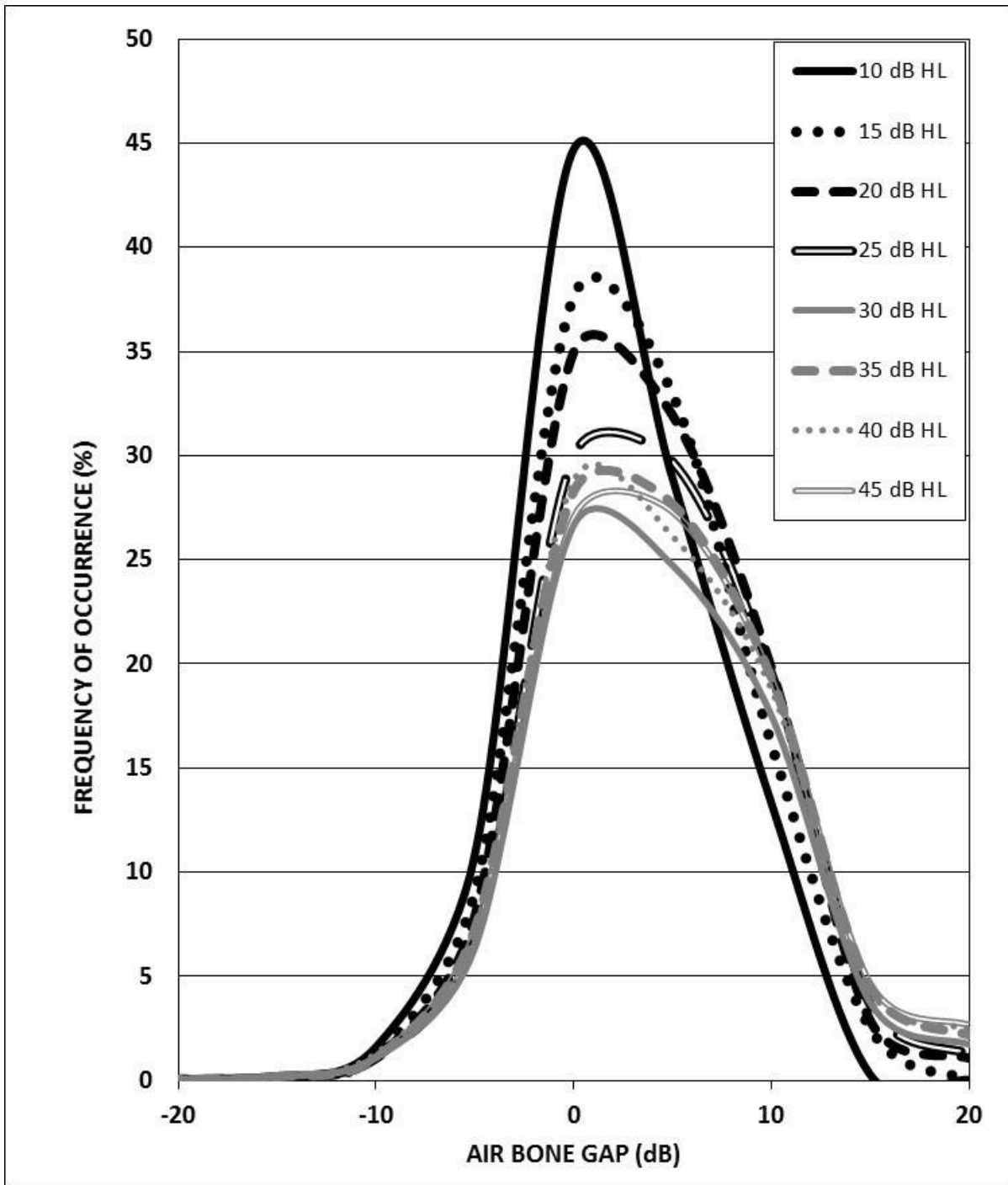


Fig. 8. Distributions of 1000-Hz air-bone gaps from the VA database stratified by the AC threshold. The data were obtained by manual audiometry. AC indicates air conduction

database stratified by the AC threshold. There is a monotonic relationship between the peak of the ABG distribution and AC thresholds over a range of 10-30 dB. Beyond that level, the distributions are unrelated to the AC threshold. These results suggest that as the AC threshold increases from normal hearing toward greater hearing loss, the biasing effect on BC thresholds diminishes. This may reflect an expectation that the ABG should be 0 dB when AC thresholds are in the normal range and are more likely to differ from zero when the AC threshold is outside the normal range.

The effect of tester bias on BC thresholds is to decrease the independence of AC and BC variability. Because the tester may target the BC threshold toward the AC threshold, the variances become highly interdependent. This can be manifested in a change in the form of the distributions and may explain the skewed distributions in Figures 5, 6, and 8.

Hipskind and Rintelmann (1969) attempted to examine the effect of tester bias by providing various types of prior information to testers who measured pure-tone thresholds, speech recognition thresholds, and word recognition scores in subjects with normal hearing, sensorineural hearing loss, and conductive hearing loss. They found no significant differences related to prior information or the experience level of the testers and concluded that they saw no evidence of tester bias, in contrast to the conclusions of this study. The lack of effect in the Hipskind and Rintelmann study may be related to (1) the fact that the testers were aware that they were participating in a study and therefore were more likely to follow the prescribed protocol, and (2) the sample size (56 audiograms compared to 721,831 audiograms analyzed in the present study). In addition, Hipskind and Rintelmann analyzed only the differences between means in their various conditions. It is the shapes of the distributions that are most revealing in the present study.

The 4000 Hz Air-Bone Gap

Margolis et al. (2013) discussed the 4000-Hz ABG that is ubiquitous in clinical practice and present in some clinical studies. That study reported results from several laboratories in three countries documenting the occurrence of the 4000-Hz ABG when the audiometers were calibrated to international (ISO 389.3-1994) and American (ANSI S3.6-2010) standards. The erroneous ABGs appear to be related to an incorrect BC Reference Equivalent Threshold Force

Level (RETFL) at 4000 Hz and seems to have its origin in early studies that employed the Brüel & Kjær Type 4930 Artificial Mastoid for calibration. Some clinics and research studies have employed correction factors for the 4000-Hz bone-conduction RETFL to eliminate the erroneous ABG (e.g., Popelka et al. 2010) and others have excluded 4000 Hz bone conduction from clinical test protocols.

The 4000-Hz ABG error is evident in the Busselton data in the shift of the distribution toward larger ABGs (Figure 1). The peaks of the UMH, VA, and Stanford distributions were near 0 dB and did not show the expected 4000-Hz ABG error. These are puzzling results.

To investigate why the UMH database did not show the 4000-Hz ABG, the calibration records were requested from the contractor that provided calibration services during the period that the database was acquired. Surprisingly, the calibration records did not include the Reference Equivalent Threshold Sound Pressure Levels (RETSPLs) to which the audiometers were calibrated or the measured sound pressure levels of the earphone outputs after the instruments were calibrated but only the deviations from the target values. We subsequently learned that it is industry policy to withhold this information from calibration records (Anon., 2014).

The absence of the 4000-Hz ABG error in the UMH, VA, and Stanford databases may have the following explanations.

- 1) The Margolis et al. (2013) result is incorrect and the standard BC 4000-Hz RETFL is correct.
- 2) The audiometers used for collection of the UMH, VA, and Stanford databases were calibrated off standard to correct for the erroneous 4000-Hz ABG. If this is the case, then it appears that the users were not always informed of this off-standard calibration.
- 3) Because the predominant majority of these cases are sensorineural hearing losses and in most cases the testing audiologist knew that, and because 4000 Hz is usually tested after lower frequencies where the average ABG is 0 dB, tester bias had the effect of eliminating the 4000-Hz ABG that was assumed to be erroneous.

Risk of Clinical Errors due to Bias

The influence of prior expectations on test results has the potential, and indeed has

probably had the effect, of causing significant errors in interpretation of audiograms. There are a number of conditions that produce ABGs that are not consistent with the usual expectations of BC thresholds in certain pathologies. Three examples are provide here.

Probably the first observation of this sort was the “Carhart notch” that occurs in patients with otosclerosis (Carhart, 1950). The lower BC threshold at 2000 Hz was not expected based on the widely held assumption that middle ear pathology does not affect BC thresholds.

The aberrant 4000-Hz ABG (discussed in detail by Margolis et al. 2013) has been with us for at least four decades. It appears in some datasets and not others. It is likely that tester bias has contributed to the lack of understanding of this problem.

More recently, the ABGs in patients with inner ear pathology have been described. Merchant et al., (2007a) reported ABGs in patients with large vestibular aqueduct syndrome and offered an explanation of the underlying mechanism. Rosowski et al. (2004) and Merchant et al. (2007b) reported ABGs in patients with superior semicircular canal dehiscence and explained the underlying mechanism. ABGs are usually not expected in these patients and prior expectations that bias bone conduction threshold measurement could lead to misdiagnosis.

It is likely that BC thresholds have been erroneously reported in cases like these where the BC threshold value is unexpected. It is dangerous, unscientific, and unprofessional to allow our prior expectations to affect clinical measurements of auditory sensitivity. In manual audiometry, care should be taken to guard against the inaccuracies that result from measurement bias.

Implications for Communication with Patients and Other Professionals

AC and BC threshold variability contribute to the variability of ABGs. The variability of BC thresholds is greater than that of AC thresholds (Dirks & Swindeman 1967; Margolis et al. 2010, 2011; Margolis & Moore 2011), resulting in a greater influence of BC variability on the variability of ABGs than AC thresholds.

Because AC thresholds are usually measured before BC thresholds, the tester has more knowledge of the expected results during BC testing probably resulting in greater bias during BC testing. Although the tester may have information from previous audiograms, case history, medical records, and other sources during AC testing, the knowledge of AC thresholds probably

has the most powerful biasing effect. If the Busselton distributions (Figures 1 and 2) represent the true variability of ABGs and the VA and Stanford distributions (Figure 5 and 6) represent the variability of the ABG as influenced by the audiologists' prior knowledge of the results, then the long-term effect is to project an erroneous impression of the variability of our threshold measurements, the bias-variance tradeoff (Fortmann-Roe 2012). Audiologists frequently experience questions about both positive and negative ABGs that are thought to be "wrong" because they are non-zero in cases that are believed to be sensorineural. The biased data has led to an expectation that ABGs should not occur as a result of the inherent variability of the measurements. Studebaker (1967) tried to disabuse us of that notion but his message has not taken root.

Manual audiometry procedures can be modified to reduce but not eliminate the biasing effect of prior knowledge. If bias due to prior knowledge could be eliminated completely, then clinical trials would not be conducted using double-blind protocols.

Limitations of the Study

Because this is a retrospective study of datasets that were not collected with our purposes in mind, there are likely to be shortcomings that would not have been present in a prospective study. Limitations include the following.

1. The databases analyzed for this report are the ones that were available to the investigators. Just as differences were observed among the datasets reported here, other databases may show different characteristics. We are hopeful that this work will inspire other studies to test our interpretations.
2. Because of differences in the methods and populations associated with the databases, different inclusion and exclusion criteria were used. This was done to control variables for the purposes of making the databases more comparable but a study that could use identical inclusion and exclusion criteria would have been preferable.
3. The use of automated and manual datasets obtained on different subject groups in different centers is a limitation. The preferred method would be to test the two methods prospectively on the same subjects.
4. Although we attempted to eliminate as many cases with conductive components as possible,

the attempt was imperfect due to the occurrence of ABGs that result from the expected variability of air and bone conduction thresholds.

5. BC thresholds were obtained with forehead placement of the bone vibrator when the testing was done with automated audiometry and with mastoid placement during manual testing.

Although the RETFLs for the two placement locations have been in the standards for decades, the different placements could contribute to differences seen with the two methods.

SUMMARY AND CONCLUSIONS

Five databases of audiometric data were analyzed to study the frequency distributions of the ABG. Two of the databases were acquired with an automated audiometry procedure (AMTAS). Three were acquired with manual audiometry procedures as used in routine clinical assessment. The results suggest the following conclusions:

1. The bias-free automated threshold procedure produces ABG distributions that are normally distributed.
2. The automated method produced ABG distributions in a smaller independent database that sampled a different population of subjects produced ABG distributions that were normal in form with a greater variability suggesting that the variance of the normal distribution is dependent on the characteristics of the populations that are sampled.
3. The presence of conductive components in audiograms included in a database may have the effect of obscuring the inherent variability of the ABG.
4. When conductive components are effectively removed from databases acquired with manual audiometry, the resulting ABG distributions are characterized by a lower variability relative to automated testing, most likely resulting from tester bias associated with prior knowledge of the patient's hearing loss characteristics (the bias-variance tradeoff).
5. The 4000-Hz ABG reported in a previous study (Margolis et al. 2013) was not evident in the databases acquired with manual audiometry. Potential explanations for the absence of that finding include off-standard calibration of bone conduction at that frequency and

tester bias.

6. Audiology clinics that use contractors for calibration should require records that include the target values to which audiometers are calibrated and the absolute measurements of signal properties after calibration.

ACKNOWLEDGMENTS

Kevin Quitmeyer, Pam Urrutia, and Mary Ann Blumenthal from the Denver Acquisition and Logistics Center who extracted the VA archived data provided important assistance. Drs. Gene Bratt, Rachel Tomasek, and Mia Rosenfeld were enormously helpful in building the VAI2 database. Dr. Michael Hunter and his team at the Busselton Health Study provided similar assistance in collecting the Busselton data. Three anonymous reviewers provided very helpful suggestions. The contents of this report do not represent the views of the Department of Veterans Affairs or the United States Government.

REFERENCES

- Anon. (2014). Understanding Your Clinical Audiometer Calibration Certificate (www.trilogyaudiology.com) p. 6. Retrieved July 2, 2014.
- ANSI S3.6-2010. (2010). Specification for audiometers. American National Standards Institute, New York.
- Carhart, R. (1950). Clinical application of bone conduction audiometry. *Arch Otolar*, 51, 798-808.
- Dirks, D., Swindeman, J. G. (1967). The variability of occluded and unoccluded bone-conduction thresholds. *J. Speech Hear Res* 10, 232-249.
- Doane, D. P., Seward, L. E. (2011). Measuring Skewness: A forgotten statistic. *J Stat Educ*, 19, 1-18.
- Eikelboom, R. H., Swanepoel, D. W., Motakef, S., et al. (2013). Clinical validation of the AMTAS automated audiometer. *Int J Audiol*. 52, 342-349.
- Fortmann-Roe, S. (2012). Understanding the Bias-Variance Tradeoff. [www.http://scott.fortmann-roe.com/docs/BiasVariance.html](http://scott.fortmann-roe.com/docs/BiasVariance.html), retrieved 2 March 2015.
- Hipskind, N. M., Rintelmann, W. F. (1969). Effects of experimenter bias upon pure-tone and speech audiometry. *J Aud Res*, 9, 298-305.
- ISO 389.1. (1998). Acoustics – Reference zero for the calibration of audiometric equipment – Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones. Geneva: International Organization for Standardization.
- ISO 389.3. (1994). Acoustics – Reference zero for the calibration of audiometric equipment – Part 3: Reference equivalent threshold force levels for pure tones and bone vibrators. Geneva: International Organization for Standardization.
- James, A., Hunter, J. A., Straker, L., et al. (2013). Rationale, Design and Methods for a Community-Based Study of Clustering and Interactions of Chronic Disease Processes and their Effects on Ageing: The Busselton Healthy Ageing Study. *BMC Public Health*. 13,936. Published online Oct 8, 2013. doi: 10.1186/1471-2458-13-936.
- Madsen, B., Margolis, R.H. Audiometric properties of three circumaural earphones. Presented to the American Auditory Society, Scottsdale, Arizona, March 7, 2014. Available at

- <http://audiologyincorporated.com/wp-content/uploads/AAS-Poster-2014-Audiometric-Properties-of-Three-Circumaural-Earphones.pdf>.
- Mahomed, F., Swanepoel, D. W., Eikelboom, R. H., et al. (2013). Validity of automated threshold audiometry: a systematic review and meta-analysis. *Ear Hear*, 34, 745-752.
- Margolis, R. H. (2008). The vanishing air-bone gap – audiology’s dirty little secret. Available at: <http://www.audiologyonline.com/articles/vanishing-air-bone-gap-audiology-901>.
- Margolis, R. H., Eikelboom, R. H., Johnson, C., et al. (2013). False Air-Bone Gaps at 4 kHz in Listeners with Normal Hearing and Sensorineural Hearing Loss. *Int J Audiology* 52, 526-532.
- Margolis, R. H., Frisina, R., Walton, J. P. (2011). AMTAS[®] - Automated method for testing auditory sensitivity: II. Air conduction audiograms in children and adults. *Int J Audiol*, 50, 434-439.
- 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., Moore, B. C. J. (2011). Automated method for testing auditory sensitivity: III. Sensorineural hearing loss and air-bone gaps. *Int J Audiol*, 50, 440-447.
- Margolis, R. H., Saly, G. L. (2007). Toward a standard description of hearing loss. *Int J Audiol*, 46, 746-758.
- Margolis, R. H., Saly, G. L. (2008). Distribution of Hearing Loss Characteristics in a Clinical Population. *Ear Hear*, 29, 524-532.
- 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.
- Margolis, R. H., Wilson, R. H., Popelka, G. R., et al. (2015). Distribution characteristics of normal pure-tone thresholds. *Int J Audiol*, early online:1-10.
- Merchant, S. N., Nakajima, H. H., Halpin, C., et al. (2007a). Clinical investigation and mechanism of air-bone gaps in large vestibular aqueduct syndrome. *Ann Otol Rhinol Laryngol*, 116, 532-541.
- Merchant, S.N., Rosowski, J.J., McKenna, M.J. (2007b). Superior semicircular canal dehiscence mimicking otosclerotic hearing loss. *Adv Otorhinolaryngol* 65, 137-145.
- National Institute of Standards and Technology (2013). Engineering Statistics Handbook, <http://www.itl.nist.gov/div898/handbook/>, retrieved October 12, 2014.

- Pineles, S. L., Volpe, N. J., Miller-Ellis, E., et al. (2006). Automated combined kinetic and static perimetry. *Arch Ophthalmol*, 124, 363-369.
- Popelka, G. R., Telukuntla, G., Puria, S. (2010). Middle-ear function at high frequencies quantified with advanced bone-conduction measures. *Hear Res*, 263, 85-92.
- Rosowski, J. J., Songer, J. E., Nakajima, H. H., et al. (2004). Clinical, experimental, and theoretical investigations of the effect of superior semicircular canal dehiscence on hearing mechanisms. *Otol Neurotol*, 25, 323-332.
- Sackett, D. L. (1979). Bias in analytic research. *J Chron Dis*, 32, 51-63.
- Studebaker, G. A. (1967). Intertest variability in the air-bone gap. *J Speech Hear Disord*, 32, 82-86.
- Swanepoel D., Eikelboom R. H., Friedland P. F., et al. (2013). Self-reported hearing loss in Baby Boomers from the Busselton Health Study – Audiometric correspondence and predictive value. *J Am Acad Audiol*, 24, 514-521.
- Tabachnick, B. G., Fidell, L. S. (1996). *Using Multivariate Statistics (3rd ed.)*. New York: Harper Collins.
- Wilson, R. H., 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, 111-132.
- Wilson, R. H., McArdle, R. (2014). A treatise on the thresholds of interoctave frequencies: 1500, 3000, and 6000 Hz. *J Am Acad Audiol*, 25, 171-186.