

# Variation in tone presentation by pure tone audiometers: the potential for error in screening audiometry.

Christopher Barlow

Technology School, Southampton Solent University, Southampton, UK.

Lee Davison

Technology School, Southampton Solent University, Southampton, UK.

Mark Ashmore

Strategic Audiology Services, Street, Somerset, UK

## Summary

Manual pure tone audiometry has been in consistent use for a long period of time, and is considered to be the ‘gold standard’ for the assessment of hearing thresholds. Increasing legislative requirements, and a significant global cost impact of noise induced hearing loss, means that a significant amount of reliance is placed on this tool for diagnosis.

There are a number of questions regarding the degree of accuracy of pure tone audiometry when undertaken in field conditions, particularly relating to the difference in conditions between laboratory calibration and clinical or industrial screening use.

This study assessed the test-retest variability of a number of commercially available screening pure tone audiometers, all with recent calibration, using both laboratory tests and clinical tests in accordance with ISO 8253-1.

The results of both lab and clinical studies showed a high level of test-retest variability, with maximum between test variation of 21 decibels at some frequencies in the lab tests, and 35 dB in the clinical tests. Considerable variation occurred at all frequencies, with a particularly high level of variation at 6kHz for all meters. Levels of variation measured in this study suggests a high potential for diagnostic error when using screening pure tone audiometry.

**Keywords:** Audiometry, Calibration, Variance, Hearing Threshold, Audiometer

## 1. Introduction

In the field of audiology, manually operated hearing assessment using pure tones according to ISO 8253-1:2010 [1] is considered the ‘gold standard’ for the assessment of hearing thresholds by airborne conduction [2]. It is for this reason that it is vitally important to make efforts to continually assess the accuracy and practicality of the practice, in order to see that it meets modern requirements.

Exposure to high levels of noise has long been recognized as a health hazard, with the long term result of noise-induced hearing disorder in the majority of people. Several sources [3,4, 5] state that long-term exposure to sound pressure levels as low as 80 dBA poses some risk of noise induced hearing disorder, while exposure to levels

regularly above 85 dB LAeq poses a risk of ‘mild’ hearing damage to most people, with the risk of more severe damage increasing with both length and level of exposure [5, 6].

Hearing loss presents a significant global cost impact, including the costs to productivity as well as the cost of long term healthcare [7]. With an ageing population and legislative requirements to screen and protect workers from occupational hearing damage, the traditional method of audiometric screening in which tests are run on a 1:1 basis with a qualified audiometrist is expensive [8]. The reliance on pure tone audiometry means that it is important to continually assess the test procedure and equipment used for repeatability and accuracy.

## 2. Background

The fundamental methodology and equipment for audiometry have stayed very similar for a long

period of time. Pure tone audiometry was originally developed from “tuning fork” tests of the early 20th century, and has now been in use for over 90 years [9], with early audiometers by Western Electric available as early as 1923. More advanced features such as bone conduction audiometry and masking noise were already available on audiometers from Sonotone as early as 1928 [10].

The basic structure of an air conduction audiometric test involves an audiometrist, an audiometer and a patient response system. The audiometrist manipulates the audiometer to deliver pure (sine wave) tones to the patient at known amplitudes. The patient then responds to which sounds are heard through the patient response system [11]. The audiometrist uses the pattern of patient responses to determine the threshold of hearing for that patient at the various audiometric frequencies. If a person is hard of hearing in a single ear, masking noise may be used – a broad frequency spectrum sound that masks one ear from hearing loud tone presentations from a very insensitive ear.

The systems used for screening have remained fundamentally unchanged for several decades. Particular transducers (notably the Telephonics® TDH-39 and TDH-49 supra-aural headphones) have been at the core of audiometric screening since the 1960s, and despite having been conceived during WWII, [12] and being in general usage since at least the 1960s [13] these are still recommended [11, 14] and commonly used.

Although there have been studies of the reliability of different types of screening (automated, computer controlled, manual), and examination of the need for traceable calibration, there has been relatively little research into the degree of variation in performance of calibrated audiometers in clinical situations.

There are a number of aspects which suggest that calibrated audiometers may not be as reliable as generally thought. One issue is that the specifications do not currently require accreditation of the calibrating organization, as the guidelines simply state that the calibration should be performed by a ‘competent’ laboratory [1].

This leaves the standard open to interpretation, and many audiometer manufacturers recommend annual calibration to take place in their own facilities, which may or may not be accredited. This has the potential for errors in the accurate

production of tones in audiometer systems, with no centralized standardizing authority to supervise.

A second, but potentially more important issue is that of the level of uncertainty in ‘field’ testing compared to laboratory conditions. The acoustic coupler (artificial ear) defined in IEC 60318-1:2009 [15] used to assess particular headphones is a regular shape, standardized to particular dimensions [16]. The headphone is coupled to the artificial ear with a static force of 4.5 N (+/- 0.5 N) from either a mass or calibrated jig [14], rather than using the tension from the headphone band.

While this method allows for a high level of standardization in the testing of the transducer and tone generator in the system, it assumes that there is a minimal effect on the sound pressure level presented at the ear from non-standard shapes and sizes of ears and heads, as well as variations in force of coupling.

This study aimed to assess the level of variation between audiometer measurements under both laboratory and clinical conditions, using a variety of different manual audiometers, to assess whether there were any requirements to address accuracy of tone presentation, calibration methods or clinical screening methods.

### 3. Method

The study used four commercially available audiometers. These were chosen to represent the cost range of typically used screening audiometers, and ranged in cost from £995 GBP for the least expensive up to £4500 GBP for the most expensive. The method was designed to give a representative sample of the performance of typical audiometers under both simulated and real clinical conditions. Each of the audiometers had recently undergone certified traceable calibration by its recommended laboratory, meaning that the tone presentation from each should theoretically be identical.

#### 3.1. Laboratory testing

Laboratory testing took place in a hemi-anechoic laboratory, with an overall noise floor rated at NR 18, and which considerably exceeds the absolute noise criteria of ISO 3745:2012.

A calibrated Brüel and Kjær® Head and Torso Simulator (HATS) of type 4100 was used with an NTi XL2 Sound Level Meter to record tones presented by the audiometers under test. A qualified audiometrist placed the headphones on

the HATS, and the absolute sound pressure level at the ear was recorded for each ear at three presentation levels, over 6 frequencies (250 Hz, 500 Hz, 1 kHz, 2 kHz, 4 kHz, and 6 kHz).

The average sound pressure level  $L_{EQ}$  (unweighted, fast response), was measured over a 5s timed period. The headphones were then removed and replaced by the audiometrist in order to reflect 'clinical' placement, and full test was repeated 3 times per audiometer.

### 3.2. Clinical testing

For clinical testing, a cohort of 13 volunteer subjects was assembled using the respondents to a University-wide email. The cohort was not selected with any gender or age bias, other than those imposed by the demographic of the University staff and students. The cohort included subjects with normal hearing and also subjects with a registered hearing loss.

The test schedule was designed over 3 weeks, with the subjects taking the test at the same time of day each week. The reason for this was to reduce the possibility of subjects becoming better at the tests through repetition. The tests were held at a constant time of the day in order to control for varying levels of concentration experienced during the day, which might affect the test. Testing took place in the same hemi-anechoic chamber used for the laboratory testing.

Each subject was tested with each of 3 calibrated audiometers using an identical method, by the same qualified audiometrist. The 4<sup>th</sup> audiometer used in the laboratory tests was purchased at a later date so was not used for the clinical tests. Each of the tests were performed to British Society of Audiology procedural guidelines [11], which are based on, and stand alongside the ISO Standard procedure [1]. Otoscopy was performed on all patients before each test in order to ascertain that excessive cerumen or other otological issues did not affect the test.

Subjects were measured over the normal audiometric frequencies of 500 Hz, 1 kHz, 2 kHz, 3 kHz, 4 kHz, 6 kHz, and 8 kHz.

## 4. Results

### 4.1. Laboratory testing

Figure 1 shows the maximum, minimum and mean sound pressure levels for tone presentation under laboratory conditions for each presentation level and frequency for all the audiometers tested.

The range of maximum variation between presentations at theoretically identical values is between 4.9 dB and 21 dB, with a mean value of 9.8 dB and standard deviation of 5.5 dB.

Results are also presented in Table I.

Table I: Minimum, maximum and mean measured Sound Pressure Level for different tones

<i>Level</i>	<i>Frequency</i>	<i>min dB LEQ</i>	<i>max dB LEQ</i>	<i>mean dB LEQ</i>
30 dB HL	250 Hz	25.6	34.9	29.9
	500 Hz	24.8	36	29.9
	1000 Hz	34.4	41.4	39.1
	2000 Hz	35	40.9	38.8
	4000 Hz	32.3	37.2	35.5
	6000 Hz	34.1	54.9	44.4
50 dB HL	250 Hz	45.8	54.8	50.0
	500 Hz	44.8	55.8	49.8
	1000 Hz	57.3	64.4	59.6
	2000 Hz	55	60.9	58.7
	4000 Hz	52.3	57.2	55.4
	6000 Hz	53.9	74.8	64.3
80 dB HL	250 Hz	75.9	84.9	80.5
	500 Hz	74.8	86	80.3
	1000 Hz	84.4	91.2	88.9
	2000 Hz	85	91.2	88.8
	4000 Hz	82.3	87.2	86.2
	6000 Hz	83.9	104.9	94.1

### 4.2. Clinical testing

The measured thresholds of each subject were analysed, with a particular focus on variation in results between the hearing threshold for each frequency for the same subject.

Figure 2 shows the maximum variation from any test in the cohort on a single subject. This is a useful tool as it shows the largest potential error between screening tests. The same figure also shows the mean and standard deviation of variation in threshold measurements of the same ear across the whole cohort.

Results are also presented in Table II.

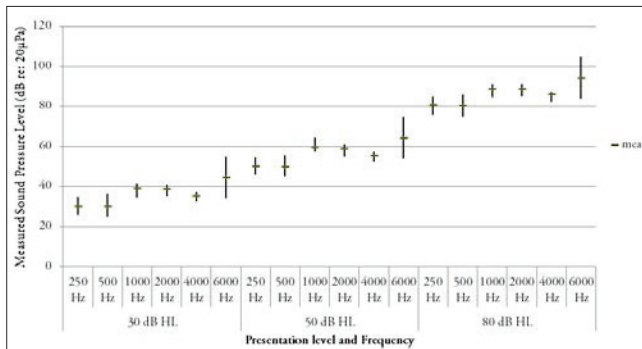


Figure 1: Minimum, maximum and mean measured Sound Pressure Level for different tones under laboratory conditions.

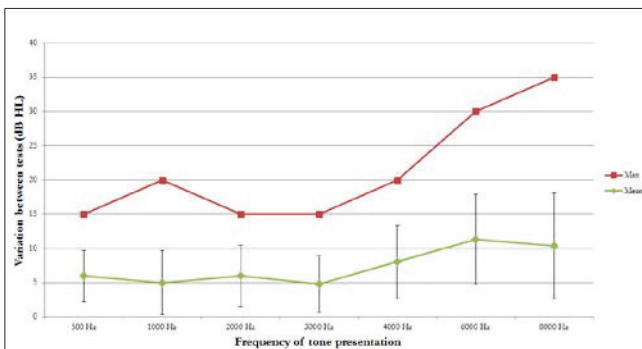


Figure 2: Maximum and mean variation with standard deviation for a series of tests of the same ear for a cohort of 13 subjects in a clinical setting.

Table II: Maximum, mean and standard deviation of variation for a series of tests of the same ear.

Frequency	Test-retest variation (dB HL)		
	Max	Mean	StDev
500 Hz	15	6.0	3.7
1000 Hz	20	5.0	4.7
2000 Hz	15	6.0	4.5
3000 Hz	15	4.8	4.1
4000 Hz	20	8.1	5.3
6000 Hz	30	11.3	6.6
8000 Hz	35	10.4	7.7

The between test variation for the same ear of the same subject ranges from a minimum of 0 dB (where the subject responded with no variation in each test at that frequency) to a maximum of 35 dB for the worst case scenario. Mean variation ranges from 4.8 dB to 11.3 dB, with a standard deviation between 3.7 and 7.7 dB.

## 5. Discussion

There was a high level of variation in the measured sound pressure levels at the ear from laboratory tests, as well as an even higher degree

of variation in the results of hearing thresholds for the cohort of subjects.

These two sets of results support each other, and while the sample size was relatively small, the audiometers used are typical of the manufacturers and types of audiometer used in the UK and Europe. Each of the audiometers was calibrated to the appropriate standards by a competent laboratory.

Under laboratory conditions the mean sound pressure levels for tone presentations between audiometers testing the same tone/level varied by 3 to 12 dB. However the maximum degree of variation across all the meters is particularly important, as this shows how much it is possible to vary in results between two theoretically identical tests.

The highest variation between identical tone presentations was encountered at 6 kHz with presentation level of 80 dB HL. This exhibited a maximum range of 21 dB between the absolute maximum and minimum values recorded in either ear across all audiometers.

As the audiometers were calibrated, the maximum variation permitted by the standard is  $\pm 3$  dB from 125 Hz to 4 kHz and  $\pm 5$  dB at higher frequencies [14]. This degree of error was considerably exceeded in the laboratory tests which were simulating a clinical setting. As each audiometer should theoretically present identical tones to the ear, any significant variation of presentation in a laboratory setting is a cause for concern, as this is likely to be exacerbated in a clinical situation, which was the case here.

Laboratory testing took place on a calibrated system meeting the requirements for a Class 1 sound level meter, with identical microphones and preamplifiers in left and right ears, so there would be minimal measurement error caused by the testing system itself. It is therefore reasonable to assume that this increased variation was due to headphone positioning and acoustic coupling of the headphone to the auditory canal on the HATS, despite being fitted by a qualified audiometrist.

The clinical tests showed an even higher degree of variation than the laboratory tests. While this would generally be expected, due to the lower degree of control of conditions in clinical situations, the degree of error shown was very high.

The highest variation between results of tests with different audiometers for the same subject/ear was 35 dB at the 6kHz tone, although



the standard deviation across the cohort for this frequency is only 6.6 dB. All frequencies had a maximum error of at least 15 dB in the clinical tests, which brings into question the accuracy of clinical pure tone testing as the primary mode of hearing screening, as this degree of error is sufficient to cause misdiagnosis.

It is worth noting that variance in results is very subject dependent. From the raw data, it was observed that there was a particularly high variation on one test in the left ear at 6 kHz, which was not reflected in the other tests on the same subject. It can be hypothesized that particularly high levels of variability could be due to different levels of patient noise exposure, or even simply down to their mood or concentration.

Subject variability is more pronounced in the high frequency areas of the audiogram. There are two main reasons that could attribute this error. One is that directionality effects in the TDH-39 headphone design can cause headphone placement to be a factor, with some audiologists potentially placing the headphones onto the patient's ears in an off on-axis orientation, in which some occlusion is caused by the tragus, and causing a reduction in the sound pressure level at the eardrum. The other main factor could be temporary threshold shift caused by exposure to loud noise [17]. Human ears are more efficient in the 6 kHz range [18], and so this frequency area is more likely to be affected by threshold changes due to loud noise. However this high frequency error appeared in both the clinical and the 'simulated clinical' testing under laboratory conditions. It is to be suggested that a high degree of the variation is caused primarily by headphone placement error. Interestingly, some authors have suggested that the hearing threshold at 6 kHz is set too high, as there is a high proportion of patients who present a threshold shift at this frequency [19]. These results indicate that this could instead be linked to variation in performance of the headphones with slight differences of placement.

Another source of error is the variation of tension in the headband which couples the transducers to the subject's ears. The calibration standards for audiometers [15] require a static force of 4.5N (+/-0.5N) in order to obtain a high quality acoustic coupling between the transducer and the ear. The different headband designs and different sizes of the heads of subjects under test will cause variation in this acoustic coupling, and poor headband tension is noted by the British

Society of Audiologists as a potential problem in testing. It is therefore reasonable to assume that some of the variability seen in this data can be attributed to differences in headband tension.

## **6. Conclusion**

Results from tests under both laboratory and clinical conditions show a wide range of variability in results from theoretically identical tests. This suggests that test results from conventional screening systems need to be carefully assessed for the possibility of error or misdiagnosis.

While the test itself is still considered 'fit for purpose', the potential for error is high, and other tests such as speech audiometry should always be used in conjunction, in order to reduce the possibility of diagnostic error.

The transducers commonly used in audiometric screening should be revisited, as developments in headphone technology over recent years has resulted in the availability of far higher quality transducers than the TDH-39, and it is suggested that a move should be made towards adopting a more contemporary design as a standard.

There is also potential for improving accuracy of the current methods of pure tone audiometry. Two particular areas are identified for further study in order to improve audiometer design. The first is the placement of the transducer, in order to minimize occlusion effects at high frequencies caused by slight misplacement, which could reduce the sound pressure level at the eardrum. The second is the relationship between headband tension and tone presentation. This may require the use of higher headband tensions or different headsets appropriate to different sizes of head. As this is potentially an important contributing factor, further research needs to be done on the exact impact of headband tension on results.

## **Acknowledgement**

This project was supported by the Technology Strategy Board, UK 2013. Grant reference: SKTP 1000821.

## **References**

- [1] International Organisation for Standardization: ISO 8253-1:2010 Acoustics-Audiometric Test Methods. Part 1: Pure Tone and Bone Conduction Audiometry. ISO: Geneva, 2010.
- [2] R.J Roeser, J. Clarke: Pure tone tests. In: RJ Roeser, M. Valente, H. Hosford-Dunn,

- (eds). *Audiology: Diagnosis*. New York: Thieme Medical, 2000. 238-260.
- [3] Action on hearing loss. Action on hearing loss; 2012. Available online from: <http://www.actiononhearingloss.org.uk/your-hearing/about-deafness-and-hearing-loss/types-and-cause-of-hearing-loss/noise.aspx>. [Last cited on 2015 Feb 25].
- [4] American Speech-Language-Hearing Association. Hearing Disorders – Noise. Available online from: <http://www.asha.org/public/hearing/disorders/noise.htm>. [Last cited on 2015 Feb 25].
- [5] Health and Safety Executive: UK Statutory Instrument 2005 no 1643 The Control of Noise at Work Regulations 2005. HMSO: Norwich, 2005.
- [6] C. Barlow, F. Castilla-Sanchez: Occupational noise exposure and regulatory adherence in music venues in the United Kingdom. *Noise Health* (14) 2012, 86-90.
- [7] A Davis, P Smith, M Ferguson, D Stephens and I Gianopoulos: Acceptability, benefit and costs of early screening for hearing disability: a study of potential screening tests and models, *Health Technology Assessment*; 11(42). 2007.
- [8] C. Barlow, L. Davison, M. Ashmore and R. Weinstein. Amplitude variation in calibrated audiometer systems in clinical situations. *Noise and Health* 16 (72). 2014, 299-305.
- [9] Feldmann H. A History of Audiology; a Comprehensive Report and Bibliography from the Earliest Beginnings to the Present. Translations of the Beltone Institute for Hearing Research. 22. (17). Chicago: The Beltone Institute 1970.
- [10] D.A. Vogel, P.A. McCarthy, GW Bratt, C. Brewer. The clinical audiogram: Its history and current use. *Commun Disord Rev* 1(2) 2007, 83-94.
- [11] British Society of Audiology. Recommended procedure. Pure-tone air-conduction and bone conduction threshold audiometry with and without masking. British Society of Audiology, Reading, 2011.
- [12] Telephonics Corporation, TDH Series Audiometric Earphones Specifications, Telephonics Corporation, New York, 1939.
- [13] W. G. Thomas, M. J. Preslar, R. R. Summers, and J. L. Stewart, Calibration and working condition of 100 audiometers, *Public health reports*, vol. 84, p. 311, 1969.
- [14] British Standards Institute. BS EN ISO 60645-1: 2001 Electroacoustics – Audiological Equipment. Part 1 – Pure Tone Audiometers. British Standards Institute, 2001.
- [15] International Electrotechnical Commission. IEC 60318-1:2009. Electroacoustics-Simulators of Human Head and Ear-Part 1: Ear Simulator for the Measurement of Supra-aural and Circumaural Earphones. IEC, Geneva, 2009.
- [16] American National Standards Institute. ANSI S3.6-2004 Specifications for Audiometers. New York: ANSI/ASA; 2004.
- [17] American Hearing Foundation. Noise induced hearing loss. Available online from <http://american-hearing.org/disorders/noise-induced-hearing-loss>. [last cited 01/03/2015]
- [18] H Takeshima, Y. Suzuki. Equal loudness level countours for pure tones. *Journal of the Acoustical Society of America*, 2004. 16(2) 918-933
- [19] D. McBride and S. Williams: Characteristics of the audiometric notch as a clinical sign of noise exposure. *Scandinavian Audiology*, 30,(2). 106-111. 2001.