Michael R. D. Maslin and Greg A. O'Beirne
University of Canterbury

25 March 2022

Reference levels for short duration signals used in Auditory Brainstem Response testing: An overview in the New Zealand context.

Executive summary

The Auditory Brainstem Response (ABR) is an electrophysiological response used for objective hearing testing. A sound is played in the ear to stimulate nerves in the brain, and the nerve activity is then measured. The technique is particularly important for testing the hearing of infants and children as the process is relatively automatic. In New Zealand, the sounds delivered by ABR devices are calibrated to different levels than devices used in most places around the world. This is because New Zealand uses a non-standard calibration reference while other regions around the world are increasingly using a standardised calibration reference, published in ISO 389. The different references mean that hearing test results and their interpretation will also differ from those performed overseas, and are not able to be easily compared to published data.

No set of reference levels are objectively right or wrong. The purpose of standardisation is to offer a common approach in order to avoid differences in sound levels across regions and across transducers like insert earphones and bone conductors in widespread use. Only with standardisation can valid comparisons be made in routine clinical as well as research activities. However, it is worth adding that in addition to reference levels being non-standard in New Zealand, the research from which they were derived has not been fully published and we believe the values contain an unexplained anomaly. Further compounding this issue, the techniques used to calibrate ABR devices are also non-standard, which has resulted in hitherto unrecognised errors.

The New Zealand reference levels are sufficiently different from the standardised reference levels to cause real-world human and financial implications. For example, the higher levels used with insert earphones can cause an apparent reduction in the severity of a hearing loss, and thus differences in the way a hearing aid might be set up for anyone diagnosed with a hearing loss. Given that standardised reference levels are available via ISO 389, and given that following this standard would enable alignment of clinical and research activities in New Zealand and internationally (and alignment within New Zealand across transducers and stimuli), the rationale for continuing with non-standard calibration references is unclear. To resolve these problems, we recommend that New Zealand adopts the standardised reference levels in ISO 389, and the associated techniques for calibration of ABR devices.

Contents

Introduction

A topic of current interest concerns the reference levels used for the calibration of short duration audiometric signals such as clicks, tonebursts and chirps in New Zealand. By reference levels, we specifically refer to Reference Equivalent Threshold Sound Pressure Levels (RETSPLs), measured in dB re: 20 µPa. In addition to a range of research activities in New Zealand, these short duration stimuli are used in two areas of clinical practice:

- Audiologists in New Zealand use clicks and tonebursts when measuring ABRs to determine hearing status
- Newborn hearing screeners in New Zealand use an automated ABR device (the Maico MB11 Beraphone) which delivers chirp stimuli as part of our UNHSEIP programme

Around the world, reference levels used for the calibration of all audiometric signals are obtained in much the same way: by measuring hearing thresholds behaviourally from a cohort of normally hearing young adults, and basing the reference level on the median threshold from this cohort (1).

For long duration audiometric signals (such as pure tones and masking noises), the behaviourally derived reference levels are applied to behavioural tests of hearing, such as pure-tone audiometry. In contrast, for the short duration signals used for physiological assessments (such as the ABR), we have behaviourally derived reference levels applied to physiological tests of hearing. It is important to appreciate the parameters used in the behavioural derivation, since this will influence the resulting reference level and physiological test. Moreover, if these parameters vary from one piece of research to the next, then so will any associated reference levels.

Reference levels for audiometric signals and their method of derivation are standardised in the International Organization for Standardization (ISO) 389 series of standards. Standardisation of reference levels for audiometric signals offers two key benefits that are necessary for implementation of evidence-based practice and audit

i. Hearing test procedures (both clinical and research) allow quantification of hearing loss relative to normal hearing in a way that is comparable everywhere
ii. Reference levels for different stimulus types (both existing and future) delivered via different transducers (existing and future) are internally consistent i.e. one person's hearing loss causes a change in what they can hear (relative to normal) in a way that is comparable with different signals and transducers

At present, the reference levels in use for ABR devices in New Zealand differ from the ISO standard (2). The reference levels for use in ABR devices in New Zealand were largely adopted from those used by many of the Canadian screening programmes (Ontario and British Colombia in particular) in the mid-2000s. In the absence of international standards for short duration signals at that time, the Canadian programmes used reference levels based on data obtained over the preceding decades by Prof. David Stapells, who had wisely cautioned against simply using the various default calibrations provided by ABR device manufacturers that were otherwise being used. These reference levels were compiled from a variety of sources from Prof Stapells' group. For the insert earphones commonly used with infants, these were namely: i) an unpublished study from Wu and Stapells for 500 Hz and 2 kHz tonebusts, first mentioned in a 1997 paper by Stapells and Oates (6); ii) 1 kHz toneburst data from Nousak and Stapells published in 2005 (18); and iii) unpublished data for 4 kHz tonebursts from Stapells. It is understandable why, in the absence of an international standard, New Zealand adopted these values at the onset of our UNHSEIP programme. However, we argue here that the development of the ISO 389.6 standard, and its consistency with the wider audiological calibration framework, now renders these data obsolete.

A recent article in the NZAS Bulletin from Maslin et al. (2021) explores some of the real-world consequences of these reference-level differences in more detail, in relation to the fitting of hearing aids (3). The aim of the present report is to offer an overview of the reference levels themselves, their derivation, and the methods used in calibration of ABR devices in New Zealand. Although the general concepts have received extensive coverage and decades of research, with several excellent reviews available (e.g. 4, 5), the present aim is to offer an exploration of the concepts in the New Zealand context.

Some parallels with the standardisation process for long duration pure tones

The ISO standard that contains reference levels for short duration signals used with ABRs (ISO 389.6) was first introduced comparatively recently, in 2007. Prior to this time, clinicians, researchers and/or authors of ABR protocols in various regions around the world had to derive their own reference levels or choose from a variety of published (but not standardised) reference levels (6, 7). For example, one relatively early ABR protocol from the UK in 2002 provided a sample of four different sets of reference levels. The thought process was that the information could be used as a guideline against which audiology services could compare their locally derived reference levels (8). These types of situation bring about unwanted and avoidable variation in the sound levels used during hearing assessments. But since 2007, this issue has been addressed and ABR protocols worldwide have progressively been updated in favour of the standardised reference levels so that obsolete and non-standardised reference levels are no longer used. Note, this does not mean that obsolete reference levels are invalid, nor that standardised reference levels reflect the absolute "truth". Instead, the purpose of adopting standardised values is to realise the two benefits described in the Introduction. So far however, the standardised values have not been adopted everywhere and some places (such as New Zealand) continue to use non-standard references. This means that the benefits are not fully realised.

A complete historical account of standardisation of audiometric signals is beyond the present scope, but it is possible, and perhaps helpful, to briefly draw parallels between the present topic and the standardisation process for long duration pure tone signals. The first recommendation for standardised long duration pure tone reference levels was published by ISO in 1964 (ISO/R 389; 9). This document later evolved into the first edition (1975) of ISO 389.1 (10). But prior to this, one can observe many examples of different pure tone reference levels that were in use simultaneously in different places around the world (e.g. 11, 12, 13). As a result, research (and presumably clinical) protocols needed to specify which reference levels their data were based on, or derive their own in just the same way as the above-mentioned example ABR protocol (e.g. 14).

With multiple reference levels in use at the same time, pure tone audiometry data inevitably did not all align, leading to a need to correct for the differences in reference levels before any interpretations or comparisons of hearing tests could be made, or in some cases necessitating repeating the same research. A clear example of these difficulties is illustrated Lawton's 2016 technical report offering a meta-analysis of historical research involving pure tone audiometry in the assessment of noise-induced hearing loss (15). Today, however, these difficulties are largely avoided in the context of pure tone audiometry because just one set of long duration pure tone reference levels are in widespread (perhaps universal) use while obsolete reference levels are no longer used.

How do the reference levels for ABR signals in New Zealand differ from those in ISO 389.6?

The non-standard reference levels of tonebursts (all frequencies) and clicks are higher in New Zealand than those in ISO 389.6. This can make someone's hearing appear better when insert earphones or headphones are used to deliver sounds during an ABR assessment. Figure 1 shows ISO reference levels for short duration stimuli (tonebursts and clicks) and those used in New Zealand for insert earphones (the most commonly used type of transducer).

Tonebursts centred on three of the four audiometric frequencies that are routinely used when performing ABRs (0.5, 2 and 4 kHz), are calibrated **1 to 4 dB higher** in New Zealand compared with the ISO values. The fourth frequency (1 kHz) is calibrated **9 dB higher**, while click signals are calibrated **3.5 dB higher** in New Zealand. (Note, short duration signals are also measured/calibrated in New Zealand in a non-standardised way that introduces an error of +2/-4dB for the clicks, which is explored in more detail in a later section below.)

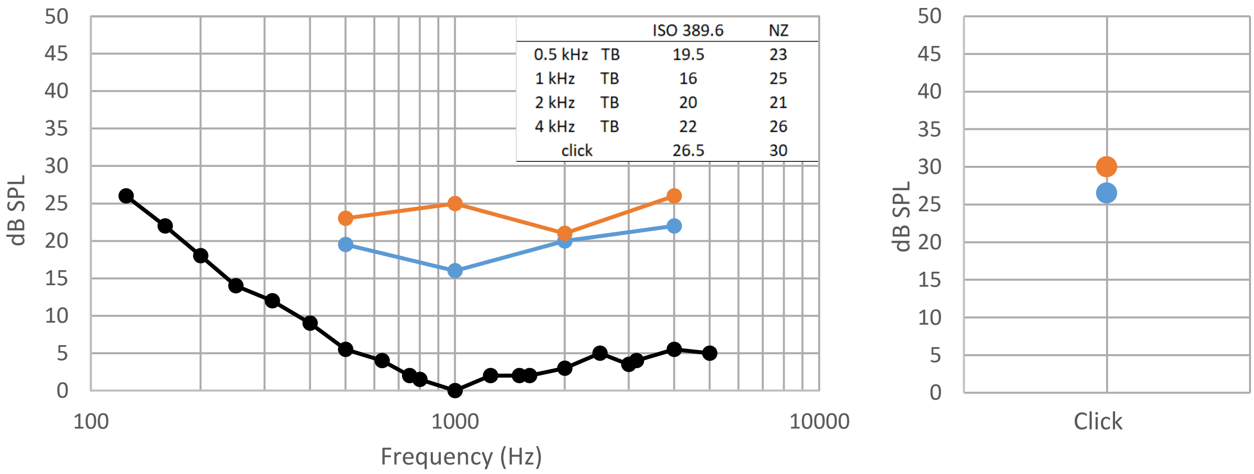


Figure 1. Left panel: Reference levels for 2-1-2 cycle tonebursts (TB) for use with ABR assessment. The values are in dB SPL (peak-to-peak equivalent) delivered via insert phone to a 2cc coupler, and standardised in ISO 389.6 (blue symbols). Non-standard values used in New Zealand are also shown (orange symbols). The black symbols show reference levels for long duration pure tones used with behavioural audiometry. The values are in dB SPL delivered via insert earphone to a 2cc coupler, standardised in ISO 389.2 (1994), and included here as a visual reference. Right panel: Reference levels for clicks used in ABR assessments. The values are in dB SPL (peak-to-peak equivalent) delivered via insert phone to a 2cc coupler. The value standardised in ISO 389.6 is the blue symbol and the non-standard value used in New Zealand is the orange symbol. Inset = tabulated values in ppeSPL.

How were the reference levels in ISO 389.6 and in New Zealand derived?

As a first step in addressing this question, it is important to consider the basic technique for deriving reference levels for short duration signals. A train of sounds (e.g. a 1 second burst of clicks or tonebursts) is delivered to the listener and they are asked to respond when they hear these sounds. This process continues as the sound level is adjusted to find the lowest level at which the listener can hear the signals – the average value from a group of listeners becomes the reference level. Various parameters can influence the result, for example:

- the background noise in the testing room,
- the task given to the listener
- the size of the increments and decrements in level
- the rate of stimulus presentation (i.e. the number of signals delivered in each burst)
- how many listeners are tested

- whether a mean or median value is used to set the reference level from the group data
- the type of stimuli (e.g. how long each short duration signal is)
- how the sound level is measured using a sound level meter

For reference levels that are described in ISO 389.6, each of these parameters is specified explicitly so that there is minimal variation across different studies e.g. when gathering reference levels for different stimulus types, or transducers. The test conditions and procedures are described in ISO 389.9 (1) and the sound level measurement technique is described in IEC 60645-3 (16). The sources of the data in the standard are also published so that others can evaluate the data transparently.

The source of the reference levels for tonebursts listed in ISO 389.6 is a 2007 study by Fedtke and Richter (17). The key features of this study are summarised in Table 1 and data are displayed in Figure 1 (blue symbols).

For comparison, Table 1 also summarises the key features of the work by Wu and Stapells, and Nousak and Stapells which are the sources of the reference levels for three of the four frequencies of toneburst used in New Zealand (also in Figure 1, orange symbols). As mentioned above, this work was performed at the well-known and respected Human Auditory Physiology laboratory (HAPlab) at the University of British Columbia, and the earliest mention of these values appears to be in a review article by Stapells and Oates in 1997 (6). In this review, published 10 years prior to the availability of standardised reference levels, the reader is offered values for two of the four frequencies of tonebursts used clinically (0.5 and 2 kHz). The origin of these two values is unpublished work by Wu and Stapells[a]. The research underpinning values for the third frequency (1 kHz) was later published by Nousak and Stapells (18), but the research underpinning the value for the fourth frequency (4 kHz) has been lost to time (Stapells, personal communication; 19).

Importantly, neither of these two pieces of research (Wu and Stapells / Nousak and Stapells) had the apriori aim of deriving reference levels for short duration stimuli. Both were studies related to measuring the ABR and similar brain activity. The inclusion of a preliminary study to derive reference levels is probably just a reflection of a lack of standardised reference levels available when both pieces of research were conducted, prompting the researchers to derive their own reference levels locally before addressing their main research aims. Thus, clinical practice in New Zealand is informed by the data from these preliminary studies.

Table 1: Summary of key parameters of research used to derive reference levels for short duration toneburst signals delivered by insert earphones used in New Zealand and ISO 389.6 (2007)

	Wu and Stapells (unpublished)	Nousak and Stapells (2005)	Fedkte and Richter (2007)
Toneburst type	2-1-2 cycle (Blackman window)	2-1-2 cycle (Blackman window)	2-1-2 cycle (linear window, standardised in IEC60645-3)
Frequencies	0.5 and 2 kHz	1 kHz	0.25, 0.5, 1, 2, 4 and 8 kHz
Transducer	ER3-A	ER3-A	ER3-A (+ five other transducer types)
Stimulus delivery period	unspecified	unspecified	1 second presentations

[a] A copy of the unpublished manuscript was gratefully received from the lead author, C-Y Wu in 2021

Rate	11.9 Hz	10.9 Hz	20 Hz
Cohort	12 otologically normal adults: one (right) ear only	10 otologically normal adults: one ear only	25 otologically normal adults: one (randomised) ear only
Threshold finding	Method of limits with 1.5 dB step size	Modified-Hughson-Westlake with 2 dB step size	Bracketing method standardised in ISO 8253-1 with 2 dB step size
Measurement device	DB0138 2cc coupler	DB0138 2cc coupler	Occluded ear simulator according to IEC 60711, as specified in ISO 389.6
Statistic	Mean ppeSPL	Mean ppeSPL	Median ppeSPL

One reflection on the work from the HAPlab is that the parameters differ not only from standardised parameters but also from each other. This is also probably just a reflection of the fact that the researchers did not have the primary aim of developing reference levels for clinical use. It is likely that if this was the aim then the researchers might have gone about things differently, similarly to their earlier work for TDH-49 supra-aural headphones (20). For example, they might have recruited more participants, tested the complete range of toneburst frequencies in a single cohort of adult listeners, specified median values because these are the most robust estimation of the central location of the group data, and used the same test parameters throughout. Meanwhile, although the details for the 4 kHz reference level are missing and a detailed appraisal has not been possible, the parameters are likely to have been similar to their stablemates. Taking all the values together the obvious outlier is at 1 kHz, where the reference level is higher than the values at other frequencies, and 9 dB higher than values in ISO 389.6. We would expect the reference level to be lowest at 1 kHz, not highest, because 1 kHz is at or near the resonant frequency of the middle ear so sounds at 1 kHz are transferred most efficiently to the inner ear. Hence this data appears to be anomalously high.

Is the ISO 389.6 standard for short duration signals in use everywhere other than New Zealand, or do other regions use non-standard reference levels?

Without having surveyed all regions around the world it is difficult to know, but it is possible to make an informed judgement. It seems the vast majority of regions around the world adhere to the reference levels in ISO 389.6 either explicitly (it is stated in open-access ABR protocols such as in Australia or the UK) or by default since ABR devices are now shipped worldwide by manufacturers calibrated (and with service manuals) according to ISO 389.6. The United States typically follows American National Standards Institute (ANSI) rather than ISO standards. In 2007, the ASHA Joint Committee on Infant Hearing called for a newborn ABR calibration standard, but in the absence of an ANSI standard they recommend manufacturers provide calibration information as per the ISO 389.6 standard (21). The overwhelming majority of US states specify that ABR equipment should be calibrated to manufacturer specifications (22). However, values similar to those in New Zealand continue to be used across Canada (23-25), and in at least one US state close to their border (26).

It may be relevant to add that since new ABR devices purchased by audiology services in New Zealand (mostly via public funds) are calibrated by default to ISO 389.6, then public money is therefore required to cover the added cost of an immediate recalibration of each new device to the non-standard New Zealand reference levels, before they can be delivered for front-line use.

The relevance of repetition rate

One of the most important factors in deriving reference levels for short duration sounds is the rate at which trains of the sounds are presented. A relatively low rate of 10 Hz to 20 Hz is optimal (4), and it is also important that for standardisation a single rate is used for all stimulus types and transducers. The value alighted on for reference levels that appear in ISO standards is 20 Hz. A slightly lower rate such as 10 Hz would tend to produce reference levels a few dB higher due to the temporal integration properties of the auditory nervous system, whereas a higher rate would tend to produce lower reference levels. It is likely that the main reason why reference levels for ABR stimuli delivered in New Zealand are higher than those in ISO 389.6 (by a few dB, excepting 1 kHz tonebursts) is because a lower repetition rate (10.9-11.9 Hz) was used in their behavioural derivation.

Does use of a different rate or other parameters make non-ISO reference levels objectively wrong?

No. None of the examples of reference data described or cited here, and others such as Sininger et al. (7), nor any locally derived values that are not cited here, are objectively wrong. These various examples have made contributions of notable importance to clinical audiology, particularly in the period before standardised reference levels were available. But, when a standardised reference for short duration signals was produced then these data became obsolete with respect to clinical use.

Much like with past examples of reference levels for long duration pure tones described earlier, most have fallen out of use accordingly. Some obsolete reference levels remain in use like in New Zealand, but there is no clear rationale for this, and their continued use only means that the benefits of standardisation continue to be unrealised.

In summary:

- **The data upon which the reference levels for short duration signals used in New Zealand are based has not all been published. Thus, a complete picture on their derivation is unavailable.**
- **The reference levels are not objectively wrong, although the values for tonebursts contain an unexplained anomaly at 1 kHz.**
- **Even if all information on the derivation of the British Columbia/New Zealand reference levels were published and there were no anomalies, this would not override the rationale, need for, or value of standardisation.**

What about short duration signals delivered by bone vibrators?

It is important to extend the discussion from signals delivered by insert earphone to bone vibration, as this is an important part of the diagnostic process undertaken by audiologists in order to differentiate conductive components of hearing losses from sensorineural components. Most concepts on the derivation of reference levels are transferable from those described already, but one difference is that signals are calibrated according to vibratory force level (giving Reference Equivalent Threshold Force Levels, or RETFLs, measured in dB re: 1 μN) rather than sound pressure level.

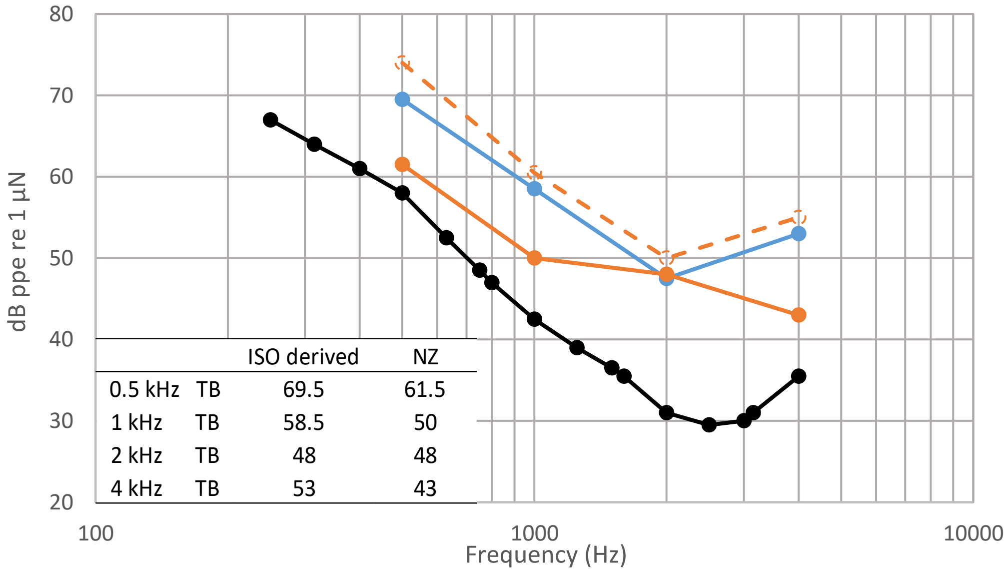
Unfortunately, the standardised references levels for audiometric signals (ISO 389.6) do not currently include values for vibratory force for short duration tonebursts. This is an important limitation, although values for long duration pure tones delivered by bone vibrators are provided (27).

Some published reference levels for short duration signals for bone vibrators do exist (e.g. 28, 29), but these studies used different acquisition parameters, and so would be misaligned with ISO389.6 air conduction values. Ideally, a set of empirically derived bone vibration reference levels analogous to those from Fedkte and Richter's 2007 research will be published in the future. In the meantime, an interim solution currently adopted by most regions around the world who otherwise follow ISO389.6, is to derive the reference levels based on a temporal integration correction factor for sounds delivered at a rate of 20 Hz, which is added to the values for long duration pure tones.

It follows that if a lower rate like 10.9 Hz or 11.9 Hz was used then the resulting bone vibration reference levels would be higher by a few dB than for a 20 Hz-based reference just as they are for insert earphones. For this reason, one would expect the bone vibration reference levels in New Zealand to be higher by a few dB just as they are for insert earphones.

However, on the contrary the bone vibration reference levels in current use are in fact *lower* in New Zealand for three of the four frequencies, as displayed in Figure 2.

Thus, while air and bone conducted tonebursts elsewhere align (because both references correspond to a 20 Hz repetition rate), those in New Zealand are misaligned because air conducted toneburst references correspond to a 10.9-11.9 Hz repetition rate and bone conducted toneburst references correspond to a different (unspecified) higher rate. The misalignment is an example of the lack of internal consistency that can result from a lack of standardisation. A real-world consequence of this misalignment is that while a hearing test performed with insert earphones would appear better in New Zealand, a hearing test performed with a bone vibrator would appear worse.



		ISO derived	NZ
0.5 kHz	TB	69.5	61.5
1 kHz	TB	58.5	50
2 kHz	TB	48	48
4 kHz	TB	53	43

Figure 2: Reference levels for tonebursts (TB) for use with ABR assessment via bone vibration. The values are in dB Force Level (peak-to-peak equivalent) delivered via a B71 (or B81) bone vibrator. The blue symbols show reference levels for short duration signals in widespread use. These are derived from ISO 389.3 by adding a temporal integration correction, corresponding to a 20 Hz repetition rate for short duration signals. The open orange symbols show approximately where the values would be expected to fall if a 11.9 Hz rate was used in their derivation. The values in current use in New Zealand are shown with the filled orange symbols. Tabulated values are inset. The black symbols shows pure tone reference levels in ISO 389.3, as a visual reference.

For anyone with a conductive hearing loss (outer and/or middle ear disorders) this misalignment is expected to make the 'air-bone gap' appear smaller (or absent) compared with a hearing test performed on the same person elsewhere. Also, the misalignment does not exist for long duration pure tone signals in New Zealand (where ISO standardised references are used). Therefore one would expect the air-bone gap to abruptly change depending on whether the hearing assessment was performed via the ABR (infancy) or via behavioural techniques (childhood and adulthood). One would hypothesise the gap appearing to widen abruptly as a patient transitions from ABR to behavioural hearing assessments.

What about chirps?

So far, the discussion about reference levels has focussed on tonebursts and clicks because these stimuli are used in the ABR protocol in New Zealand by audiologists.

However, chirps are in routine use in New Zealand during automated ABR assessment. Building on the information described above, chirp reference levels were derived behaviourally according to standardised methods, including at a rate of 20 Hz. The main rationale here is to align with the reference levels already described in ISO 389.6 for tonebursts and clicks. Reference levels for chirps are available across a range of transducer types including insert phones, and the circumaural phones used in automated ABR devices in New Zealand (30-32). Although these reference levels are not included in the current iteration of ISO 389.6, they are published and have been gathered according to the same methods and criteria in the anticipation that they could be included in a future iteration.

The alignment is important for internal consistency, as shown in Figure 3. But in New Zealand, the toneburst reference levels used by audiologists performing ABRs are non-standard, resulting in a misalignment to the chirp signal. Thus, when a patient has an automated ABR during hearing screening and a follow-up ABR with an audiologist, the relationship between results is complicated by the misalignment. As with the discussion on bone vibrators, this is an example of internal inconsistency that results from non-standardisation.

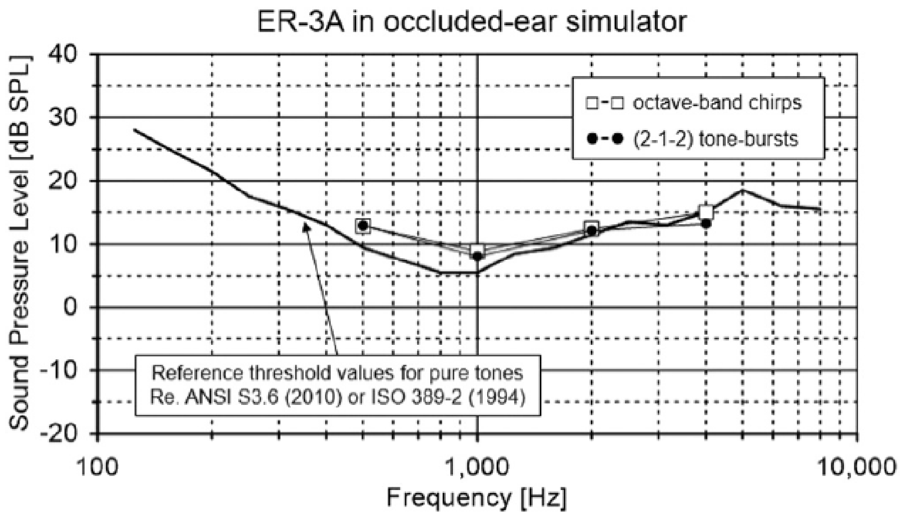


Figure 3: Reference levels in dB SPL (not ppeSPL) for long duration pure tones and short duration tonebursts and chirps under the same recording conditions, including a repetition rate of 20 Hz for the two short duration signals (30). The alignment of the chirps and tonebursts is clear, but a misalignment would be expected in New Zealand due to the non-standard reference levels for tonebursts. (The figure is taken from 30.)

Is there any issue with the standardised reference levels for short duration signals that would argue against their use?

No set of data are objectively correct but the standardised reference levels for tonebursts, clicks and chirps are peer-reviewed, and there are no anomalies such as described above for the 1 kHz toneburst reference in New Zealand. If flaws do emerge, standards committees can be petitioned for change as has happened in the past (33), rather than the standardisation process be rejected or bypassed.

One might wonder if any research conducted prior to 2007 (ISO 389.6) would require reinterpretation if non-standard reference levels were used during this research. The answer would depend on the specifics of each piece of research, but it is unlikely. From a clinical perspective, most ABR protocols are developed based on evidence that converges from multiple avenues of research and is interpreted at a meta-level. Like with Lawton (15), sometimes meta-analyses might be complicated by non-standardisation, but should not be impeded altogether. Even if reinterpretation of some historical data was required, the solution to prevent this scenario from repeating itself in the future is standardisation, thus making any such scenario a supporting argument *for* standardisation.

The calibration procedure in use in New Zealand: pSPL

When measuring short duration signals two approaches are used: peak SPL (pSPL; giving the maximum pressure of a sound) or peak-to-peak equivalent SPL (ppeSPL; giving the mean pressure). The reference levels in the ISO 389.6 standard are specified in ppeSPL, and the signals (for insert phones) are appropriately measured in a test cavity called an occluded ear simulator (711-coupler) according to a procedure standardised in IEC 60645-3 (16).

But at present the typical procedure in New Zealand is for toneburst and click signals to be calibrated by measuring pSPL in a different type of test cavity, a 2cc coupler (Jaques, personal communication; 34)[b] . A correction is then used to infer the ppeSPL (which is then compared to the reference level, also in ppeSPL)[c]. Therefore, not only are non-standard reference levels used in New Zealand, but non-standard calibration procedures are employed too.

Are there likely to be any real-world consequences of calibrating signals using pSPL instead of ppeSPL?

Fortunately there are unlikely to be any consequences for tonebursts, but unfortunately there almost certainly are with respect to click signals.

For tonebursts the relationship between pSPL and ppeSPL is fixed, therefore it should always be possible to measure pSPL and accurately infer the ppeSPL[d] via a correction.

But for clicks the relationship between pSPL and ppeSPL is not fixed. The relationship depends on the exact properties of the click[e] delivered by any given ABR device and earphone/headphone pairing, and this can vary. A correction anywhere between 3 and 9 dB is required to convert between pSPL and ppeSPL. This is important because despite this range, only a single correction value is used in practice in New Zealand (5 dB; 34) i.e. an individualised correction according to the ABR and earphone/headphone pairing is not used/available.

[b] Appropriate techniques are used for chirps.
[c] Clicks are specified in pSPL and tonebursts are specified in ppeSPL in New Zealand, for reasons unknown.
[d] pSPL is always +3dB re: ppeSPL for tonebursts
[e] The exact correction to convert between pSPL and ppeSPL for a click depends on how symmetrical the acoustic click is around the baseline. The relationship can vary over a range of 6 dB i.e. pSPL is anywhere between +3 and +9 dB re: ppeSPL for clicks.

Using a fixed (5 dB) correction to infer ppeSPL from pSPL for click signals is inappropriate. This approach means there must be errors of up to +2 dB / -4 dB in the levels of click signals used by audiologists in New Zealand.

It is impossible to know the error on any given ABR device since this can only be determined by measuring both pSPL and ppeSPL at the same time, which is not done. All that can be known is that the error will not be more than 4 dB from the values recorded on the device's calibration report(s).

The potential real-world consequences of such a degree of avoidable uncertainty in click levels has not been explored presently, but a key reason why ppeSPL measures are favoured over pSPL measures is because when the stimulus could be asymmetrical, ppeSPL is not affected by the asymmetry (4).

Concluding summary

- New Zealand uses non-standard reference levels for the calibration of toneburst and click signals used in ABR assessments
- The reference levels for insert earphones (and supra-aural headphones) are higher in New Zealand, making hearing appear better than elsewhere
- In contrast, the bone vibration reference levels for tonebursts are lower in New Zealand, making hearing appear worse than elsewhere
- Since New Zealand uses a standardised approach with reference levels for chirps there is another internal misalignment, this time between references for the signals used in the screening (chirp) and diagnostic (toneburst and click) stages of infant hearing assessments
- The methods used to calibrate tonebursts and clicks in New Zealand are also non-standard. The method involves measuring peak (not mean) SPL. This method leads to errors of up to 4 dB from the indicated level for click signals and hitherto this report, these errors have gone unrecognised
- Different stimuli and different transducers (including those in use today and any future developments) have their calibration references performed according to a standardised procedure so that data are comparable across regions, but also across different transducers and different stimuli within a region
- Non-standard reference levels are not objectively wrong and standardised reference levels are not objectively 'the truth'
- The purpose of standardisation is to avoid the complications and unintended or unrecognised consequences, by giving a common process for everyone.
- All of the complications described here that are at play in New Zealand would be resolved by switching to standardised reference levels in ISO 389 and the associated calibration procedures

References

1. ISO. 389: Acoustics - Reference zero for the calibration of audiometric equipment - Part 9: Preferred test conditions for the determination of reference hearing threshold levels. 2009.
2. Ministry of Health NZ. Universal Newborn Hearing Screening and Early Intervention Programme National Policy and Quality Standards Appendix F : Diagnostic and amplification protocols.; 2016.
3. Maslin MRD, Cross N, Welch D. A variation in standardisation: do the unique calibration reference levels used for ABR stimuli in New Zealand have a clinically significant impact on paediatric hearing aid fittings? NZAS Bulletin 2021.
4. Picton TW. Human auditory evoked potentials. San Diego: Plural Pub.; 2010. ix, 634 p. p.

5. Burkard RF, Eggermont JJ, Don M. Auditory evoked potentials : basic principles and clinical application. Philadelphia: Lippincott Williams & Wilkins; 2007. xix, 731 p., 16 p. of plates p.
6. Stapells DR, Oates P. Estimation of the pure-tone audiogram by the auditory brainstem response: a review. Audiol Neurootol. 1997;2(5):257-80.
7. Sininger YS, Abdala C, Cone-Wesson B. Auditory threshold sensitivity of the human neonate as measured by the auditory brainstem response. Hear Res. 1997;104(1-2):27-38.
8. Mason S, Elliot C, Lightfoot G, Parker D, Stapells DR, Stevens J, et al. Auditory Brainstem Response testing in babies using tone pip stimulation: A recommended test protocol. Newborn Hearing Screening and Assessment, UK: National Health Service England; 2002.
9. ISO/R. 389: Standard reference zero for the calibration of pure-tone audiometers,. 1964.
10. ISO. 389: Acoustics — Reference zero for the calibration of audiometric equipment. 1975.
11. Davis H, Kranz FW. The International Standard Reference Zero for Pure-Tone Audiometers and Its Relation to the Evaluation on Impairment of Hearing. J Speech Hear Res. 1964;7:7-16.
12. Wheeler LJ, Dickson ED. The determination of the threshold of hearing. J Laryngol Otol. 1952;66(8):379-95.
13. Dadson RS, King JH. A determination of the normal threshold of hearing and its relation to the standardization of audiometers. J Laryngol Otol. 1952;66(8):366-78.
14. Taylor W, Pearson J, Mair A, Burns W. Study of Noise and Hearing in Jute Weaving. J Acoust Soc Am. 1965;38:113-20.
15. Lawton BW. A baseline for the assessment of Noise-Induced Hearing Loss: contrasting thresholds-by-age for otologically normal persons and typical persons. ISVR Technical Report 338: University of Southampton; 2016.
16. IEC. 60645: Electroacoustics - Audiometric equipment - Part 3: Test signals of short duration. 2020.
17. Fedtke T, Richter U. Reference zero for the calibration of air-conduction audiometric equipment using 'tone bursts' as test signals. Int J Audiol. 2007;46(1):1-10.
18. Nousak JK, Stapells DR. Auditory brainstem and middle latency responses to 1 khz tones in noise-masked normally-hearing and sensorineurally hearing-impaired adults. Int J Audiol. 2005;44(6):331-44.
19. Stapells DR. Personal communication. 2021.
20. Stapells DR, Picton TW, Smith AD. Normal hearing thresholds for clicks. J Acoust Soc Am. 1982;72(1):74-9.
21. Hearing JCoI. Year 2019 position statement: Principles and guidelines for early hearing detection and intervention programs. Journal of Early Hearing Detection and Intervention. 2019;4(2):1-44.
22. Management NCfHAa. State Early Hearing Detection and Intervention Information: Utah State University; 2022 [Summary information for EHDI in the United States]. Available from: https://infanthearing.org/states/index.html.
23. Program BCEH. Audiology Assessment Protocol v4.1. 2012.
24. Program OIH. Protocol for Auditory Brainstem Response-based Audiological Assessment (ABRA) v2018.01. 2018.
25. Audiologists SAoS-LPa. Pediatric Audiology Practice Guidelines (0-5 years). 2019.
26. Minnesota NSP. Early Hearing Detection and Intervention (EHDI) Guidelines for Audiologists. 2018.
27. ISO. 389: Acoustics — Reference zero for the calibration of audiometric equipment — Part 3: Reference equivalent threshold vibratory force levels for pure tones and bone vibrators. 2016.
28. Small SA, Stapells DR. Normal brief-tone bone-conduction behavioral thresholds using the B-71 transducer: three occlusion conditions. J Am Acad Audiol. 2003;14(10):556-62.
29. Gorga MP, Kaminski JR, Beauchaine KL, Bergman BM. A comparison of auditory brain stem response thresholds and latencies elicited by air- and bone-conducted stimuli. Ear Hear. 1993;14(2):85-94.
30. Elberling C, Crone Esmann L. Calibration of brief stimuli for the recording of evoked responses from the human auditory pathway. J Acoust Soc Am. 2017;141(1):466.

31. Gotsche-Rasmussen K, Poulsen T, Elberling C. Reference hearing threshold levels for chirp signals delivered by an ER-3A insert earphone. Int J Audiol. 2012;51(11):794-9.
32. Fedtke T, Hensel J. Determination of peak-to-peak equivalent reference equivalent threshold sound pressure levels for an audiometric headphone, for ten different test signals of short duration. Report, Physikalish-Technische Bundesanstalt: Physikalish-Technische Bundesanstalt; 2008.
33. Margolis RH, Eikelboom RH, Moore BCJ, Swanepoel D. False 4-kHz air-bone gaps: a failure of standards. Audiology Today. 2019.
34. Jaques R. Acoustician with Diatec NZ. Personal communication. 2021.