Recommended Procedure

Auditory Brainstem Response (ABR) Testing in Babies

Date: February 2019

Due for review: February 2024

General foreword

This document presents Practice Guidance by the British Society of Audiology (BSA). This Practice Guidance represents, to the best knowledge of the BSA, the evidence-base and consensus on good practice, given the stated methodology and scope of the document and at the time of publication.

Although care has been taken in preparing this information, the BSA does not and cannot guarantee the interpretation and application of it. The BSA cannot be held responsible for any errors or omissions, and the BSA accepts no liability whatsoever for any loss or damage howsoever arising. This document supersedes any previous recommended procedure by the BSA and stands until superseded or withdrawn by the BSA.

Comments on this document are welcomed and should be sent to:

British Society of Audiology
Blackburn House,
Redhouse Road
Seafield,
Bathgate
EH47 7AQ
Tel: +44 (0)118 9660622

bsa@thebsa.org.uk
www.thebsa.org.uk

Published by the British Society of Audiology

Authors

Produced by: The Electrophysiology Special Interest Group (EPSIG) and the Professional Guidance Group

Key Authors: Guy Lightfoot[1], Siobhan Brennan[2], John FitzGerald[3], Inga Ferm[4]

1 ERA Training & Consultancy Ltd.
2 University of Manchester; Sheffield Teaching Hospitals
3 Norfolk & Norwich University Hospitals NHS Trust
4 Croydon Health Services NHS Trust

Declarations of interests

- Declaration of interests by the authors: ERA Training & Consultancy Ltd offer training courses in ABR testing, training and accreditation in ABR peer review and offer clinical support for centres performing ABR testing.

With thanks to:

Those who made contributions to this and earlier versions, including , Constantina Georga, Rob Gardner, Graham Sutton, John Stevens, Rachel Booth, Rachel Feirn, Rachel Beeby, Rhys Meredith, Rob Low, Steve Mason, David Stapells, Clive Elliott, Dave Parker, Mike Vidler and Amanda Hall for providing the literature review presented in Table 2.

The expert reviewers for their advice and guidance and to BSA members who provided feedback in the membership consultation.

Citation

Please cite this document in the following format:

BRITISH SOCIETY OF AUDIOLOGY (2019), *Recommended Procedure Auditory Brainstem Response (ABR) Testing in Babies.* [Online]. Available from: *insert web link*. [Accessed *date*]

Shared Decision-Making

It is implied throughout this document that the service user should be involved in shared decision-making when undertaking audiological intervention, receiving subsequent information and understanding how it will impact on the personalisation of care. Individual preferences should be taken into account and the role of the clinician is to enable a person to make a meaningful and informed choice. Audiological interventions bring a variety of information for both the clinician and the patient which can be used for counselling and decision-making regarding technology and anticipated outcomes.

Contents

1. Abbreviations

ABR	Auditory brainstem response
AC	Air-conduction
ANSD	Auditory Neuropathy Spectrum Disorder
ASSR	Auditory steady state response
BC	Bone-conduction
ckABR	Click-evoked ABR
CM	Cochlear microphonic
CR	Clear Response
dBeHL	Estimated PTA from electrophysiological thresholds
dBnHL	Decibels Hearing Level (the "n" is a hangover from days before an international calibration reference was available, and the scale used was derived from "nominal" or "normal" studies. nHL has been retained by convention to distinguish it from dBHL used for long duration tonal stimuli)
ECochG	Electrocochleography
eSP	e-Screener Plus (Electronic record system for NHSP), now replaced by S4H
Inc	Inconclusive
NBchirp	Narrowband chirp
NHSP	Newborn Hearing Screening Programme (England)
RA	Response Absent
RETFL	Reference Equivalent Threshold Force Level
RETSPL	Reference Equivalent Threshold Sound Pressure Level
S4H	Smart 4 Hearing: Electronic database for recording newborn hearing test results
tpABR	Tone pip evoked ABR

2. Introduction

2.1. Development of the recommended procedure

This document has been adapted from the document 'Guidance for Auditory Brainstem Response testing in babies', version 2.1, March 2013, produced by the Clinical Group of the NHSP and an Update Document 2015 (see Appendix A for major changes incorporated into this procedure). The development of this recommended procedure was undertaken by the members of the Electrophysiology Special Interest Group (EPSIG) and has been developed in accordance with BSA Procedure for Processing Documents (2003).

2.2 Background and aims

Prior to a national standard for performing ABR testing there was a wide variation in practice within the UK, both in terms of recording parameters and test stimuli used, as well as in interpretation of waveforms. Through the introduction of guidelines by the Clinical Group of the NHSP, uniformity in

equipment set up was achieved along with improvements in test performance and waveform interpretation.

This procedure aims to set out how to ensure good quality recordings of ABR waveforms are obtained using earphones, inserts and bone-conduction transducers and how to identify an accurate estimate of hearing thresholds at different frequencies. It aims to define criteria by which to identity a 'clear response', 'response absent' or 'inconclusive' response when performing ABR testing in babies to ensure uniform standards can be achieved by those using this procedure.

2.3 Scope

These guidelines are for the use of ABR in assessing hearing in babies up to a corrected age of 12 weeks. Frequency-specific information is required[a]. The document should be read in conjunction with the former NHSP (now BSA) document 'Guidelines for early assessment and management of babies referred from the Newborn Hearing Screening Programme' which describes the whole process of assessing hearing in neonates including the use of ABR (BSA, NHSP, 2013 or later revision). A separate document, in preparation, provides guidance on ABR testing of post-newborn and adults patients. Reference should also be made to the other NHSP Audiology protocols and guidelines available on the BSA web site http://www.thebsa.org.uk/ , and in particular the BSA recommended procedure Assessment and management of Auditory Neuropathy Spectrum Disorder (ANSD) in young infants (BSA, 2019).
The document covers the technical procedure of carrying out an ABR test and the reporting of the results. It does not cover equipment for 'automatic' or screening ABR.

3. Patient preparation

3.1 Test environment

Threshold ABR tests should ideally be performed in a sound-proofed room or environment which meets the same standards as used in pure-tone audiometry. The minimum standard should be an environment in which the lowest air-conduction and bone-conduction stimulus levels that are to be used (typically 15dBnHL) can be clearly heard by a normally hearing adult. Fan noise from the equipment can cause masking of stimuli at low stimulus levels: if this is the case the equipment should be sited further away from the test subject. Also, levels of electrical interference (*e.g.* 50Hz mains) should be sufficiently low such that the signal baseline is not adversely affected. Test rooms should not be sited close to potential sources of interference such as high powered mains equipment, transformers, or plant equipment.

Where ABR testing is performed outside the designated clinic area - for example on the ward or in the operating theatre - levels of acoustical and electrical interference must be sufficiently low so as

[a] This may include tone pip ABR, narrowband chirp ABR and ASSR.

not to influence the results of the test. Careful selection of the local test area or room may be necessary in order to achieve satisfactory environmental conditions.

3.2 Precautions against cross-infection

All local procedures should be adhered to. These should cover hygiene, use of equipment and electrodes.

3.3 Choice of electrodes and application

The following procedure is recommended. The skin should be gently and carefully abraded. Appropriate options include abrasive electrode preparation paste and a clean gauze pad, a disposable abrasive pad or a cleaning stick with soft cotton material on the end. Single use disposable electrodes are recommended.

Artefact size from induced electrical interference is proportional to the difference in the electrode impedances. This difference in impedances is most easily minimised by ensuring all electrodes have low impedances. The impedance, as measured between each electrode pair should be under 5000 ohms and similar across electrodes. However in good recording conditions and in a screened room higher electrode impedances can be tolerated. High impedance would also give an unacceptably large stimulus artefact at high stimulus levels; particularly for bone-conduction ABR.

The ABR system (dedicated hardware or associated computer) must not be switched on or off (re-booted) with the patient attached. If it is necessary to re-boot the ABR system, first disconnect the patient electrodes and transducers then re-connect them once the system is fully operational. This ensures that no potentially hazardous currents pass through the electrodes and no high-level stimuli are presented to the patient.

3.4 Electrode location – AC and BC

A single channel recording is recommended for AC and BC with electrodes located as follows:

- Positive electrode: **high forehead** as near to Cz[b] as possible and midline. The fontanelle should be avoided but the electrode should be placed as close as possible to this otherwise the ABR response will be reduced in size. A mid-forehead position is not appropriate.

- Negative electrode: **ipsilateral mastoid**. Sufficient space should be allowed for a bone vibrator to be placed on the mastoid without interfering with the electrode. To allow possible recording of CM, the electrode should be no more than 1cm lower than the meatal level of the ear

- Common electrode: **contralateral mastoid**.

[b] Cz is the standard position used in adults. It is defined in the 10-20 electrode system for electroencephalography. For these purposes it can be taken as the point along the midline of scalp half way between the bridge of the nose (nasion) and the start of the skull at the rear of the head (inion).

This configuration should result in wave V being plotted upwards on the display. If this is not the case then the positive and negative electrode connections should be reversed.

Alternative montages:

- There is some evidence that the nape of the neck rather than the mastoid gives a larger wave V response although it is reported (Stevens et al. 2013) that there is little difference in test efficiency. If the nape of the neck is used for the negative electrode, the common electrode can be placed on the forehead (at least 4cm from the positive electrode) or either mastoid, whichever is the more practical in the individual case. The mastoid should be used for the negative electrode if wave I or CM is required for neuro-diagnostic purposes.

Two-channel recording of AC & BC may also be considered. This records the contralateral response in an attempt to determine which cochlea is generating the ABR. Note that this technique has limitations that should be understood if it is to be used. Details are provided in Appendix B for those interested. If this option is being considered then the common electrode can be placed on the forehead as the ipsilateral and contralateral electrodes will be used in the two channel recording.

4. Stimulus

The recommended values (or ranges) for stimulus parameters are summarised in Table 1. These ensure optimum recording of the III/V-SN_{10} complex which is crucial to paediatric threshold testing.

4.1 Stimulus and stimulus rate: Air-conduction and single-channel bone-conduction

The stimulus should be of alternating polarity to minimise the stimulus artefact.

An electrical pulse of 100µs should be used for click ABR (ckABR) and a 2 cycle rise/fall and 1 cycle plateau for tone pip ABR (tpABR), the reference stimuli described in IEC, 2007. Narrow band (pip-like) chirps (NBchirp) can also be used (Elberling & Don, 2010). The envelope for the rise, plateau and fall phases of the tone pip can be Blackman[c] or linear (Blackman is preferred). Some equipment specifies a Blackman envelope by stating the total number of cycles (this should be 5).

[c] A Blackman envelope does not strictly have a plateau. If the equipment has the option of entering the total number of cycles this is preferable for a Blackman envelope. Where it does not we recommend entering 2 cycles rise/fall and 1 cycle plateau for the Blackman envelope.

Table 1: Summary of recommended ABR parameters[d]

	Click, NBchirp & 2kHz / 4kHz tone pip	**0.5kHz / 1kHz tone pip**
Electrode location[e]	Positive : High forehead (as close to vertex as possible but avoiding fontanelle) Negative : Ipsilateral mastoid Common : Contralateral mastoid	
Stimulus type	Alternating polarity	
Stimulus timing	Click: 100µs Tone pip: 2-1-2 cycles (linear rise–plateau–fall) or 5-cycle Blackman	
Stimulus rate[f]	45.1 - 49.1/s	35.1 - 39.1/s
Calibration values for 0dBnHL	Refer to NHSP calibration data	
Amplifier reject levels	±3 to ±10µV[g] peak-to-peak. Start at ≤ ±5µV peak-to-peak	
Amplifier filters	Low frequency: 30Hz High frequency: 1500Hz	
Window length[h]	20ms	25ms
Number of sweeps averaged per replication	Typically: 2000 click & NBchirp, or 3000 for TP Minimum: 1500 click & NBchirp, or 2000 for TP	
Display scales	Within range 25-100nV ≡ 1ms See equipment specific settings	
Display	Wave V up	

The frequencies used for frequency-specific testing should be 0.5, 1, 2 or 4kHz.

The slower stimulus rates for 0.5kHz and 1kHz tpABR allow for longer window lengths to be used so that the full SN_{10} part of the waveform is recorded[i]. A range of recommended values has been given to fit with those available on commonly used equipment. Lower repetition rates will not give invalid results but will be less time efficient. Equipment-specific optimal parameters for NHSP work are

[d] Note that some equipment offer more advanced features or stimuli, not covered in this table. See the NHSP equipment-specific parameter document for details.
[e] Alternative electrode montage: Negative – nape; Common – forehead at least 4cm from positive or negative electrodes (low forehead or to one side). Note that some manufacturers label positive and negative as active and reference respectively. Referring to common as Ground or Earth is technically incorrect; indeed it is dangerous to ground a patient. For two-channel BC see Appendix B.
[f] Most equipment can provide a rate within these ranges for the suggested window length (see equipment-specific parameters). The rate must not be related to 50Hz. If chirp stimuli are used the optimum rate depends on the chirp duration.
[g] See Appendix D for note on using 10µV rejection setting.
[h] These window lengths are nominal values and should be set to the closest value available on the equipment. Chirps should be used with a window length of 20ms regardless of stimulus frequency.
[i] There is however evidence that equal test efficiency can be achieved at both lower and faster rates (Stevens et al. 2013a).

available; it is particularly important to use these if chirps are used since a compromise is required between stimulus rate, stimulus duration and "blocking" which is needed at both the start and end of the recording epoch[j]. Rates such as 35.1/s, 49.1/s etc. (with no common relationship with mains frequency) should be chosen to minimise any mains artefact.

The level of stimulus output should be checked at the start of a session (see the NHSP document on routine 'stage A' checks for ABR systems, (NHSP Clinical Group 2008) and monitored by listening to the earphone at critical points during the test, particularly if unexpected results occur.

Stimulus levels should be recorded in dBnHL. The "nHL" can be taken to imply the use of either ISO, 2007 or NHSP-recommended calibration values (NHSP Clinical Group 2012)[k]. For testing carried out under the NHSP in England, the advice in the NHSP early assessment guidelines should be adhered to (BSA, NHSP 2013).

4.2 Earphone

This should be able to deliver a stimulus up to 140dBSPL peak (about 107dBnHL for a click stimulus) without distortion. Supra-aural or insert earphones (*e.g.* type ER-3A) are suitable. The actual stimulus level is more uncertain with insert earphones due to the greater variation in the enclosed volume of a baby's ear canal and this has implications for the precision of ABR results when inserts are used in babies. However insert earphones reduce the need for masking and attenuate ambient noise more than supra-aural earphones. If insert earphones are used, take care that wax is not compacted by the probe, so blocking the sound pathway. Supra-aural earphones should be centred over the ear canal to avoid collapsing the ear canal due to excess pressure.

4.3 Warning – insert earphones

Insert phones should not be used above the maximum levels given in the NHSP guidelines for early audiological assessment[l] (BSA, NHSP 2013). This is because a baby has a much smaller ear canal which will lead to a 10-20dB higher stimulus level compared to the same insert earphone used in an adult. This uplift is thought to diminish over the early months of life as the ear canal grows (see NHSP early assessment guidance for more detail). In cases where no ABR response is recorded at the maximum recommended stimulus level using insert phones, testers should consider the use of supra-aural earphones up to the maximum recommended levels for those transducers.

---

[j] A chirp stimulus begins before (and extends to just after) the "zero" point on the recording timebase and evokes a response which does not appreciably change its latency with frequency. For this reason a 20ms window is used for all frequencies of NB chirps. In order to use a fast rate we allow the next stimulus to start just before the end of the previous timebase and since this could produce a stimulus artefact, "blocking" periods are needed at both the start and end of the timebase. These periods are affected by the chirp frequency, timebase and rate so it is vital that recommended values are adopted without modification. For more details of chirps see Elberling and Don 2010.

[k] If using nonstandard stimuli, one may require local nHL values referenced to the average psycho-acoustic threshold of a group of normally hearing young adults.

[l] The only exception would be if the equipment included a microphone to automatically adjust the stimulus level for ear canal volume - but this is not yet available.

4.4 Bone vibrator

This should be able to deliver a stimulus up to 60dBnHL (50dBnHL at 0.5kHz) without obvious waveform distortion. Stimulus levels should not exceed these values unless the bone vibrator has been passed in calibration as being able to deliver higher levels without distortion. The Radioear type B71 bone vibrator should be used as calibration data are available for this and not for other types. A Radioear type B81 bone vibrator is now available; preliminary data (Radioear Type B81 Data Sheet) suggest it is acceptable to apply the B71 calibration reference levels to the B81. A check should be made that the impedance of the bone vibrator is correct for the equipment being used. A 'Stage A' (NHSP Clinical Group 2008) listening test near threshold, and at 50dBnHL or above should be carried out at the start of each session in which a bone vibrator is used.

4.5 Placement of bone vibrator

The bone vibrator should be placed on the mastoid approximately 1 finger's width above the electrode. If possible, move hair away from where the bone vibrator is to be placed. The bone vibrator lead should be kept away from the electrode and electrode lead. This placement gives a higher stimulus level when compared to a forehead placement in young babies (Webb 1993). A mastoid location also takes full advantage of the inter-aural attenuation which is at least 20dB for clicks in babies under 12 weeks corrected age (Webb 1993).

Placement on the temporal bone slightly posterior to the upper part of the pinna may be a good alternative where the mastoid is difficult due to proximity of the electrode (Small, Hatton, and Stapells 2007).

4.6 Pressure to apply to bone vibrator

A moderate force ('finger pressure') should be applied to the bone vibrator, but the exact force is not critical - tests on an artificial mastoid have demonstrated an error of no more than 2dB over a wide range of applied forces (Webb 1993). It is not good practice to ask the parent or carer to hold the bone vibrator as this may lead to an inconsistent pressure or changes in position.

4.7 Effect of age on the bone-conduction stimulus

The effective level of the stimulus changes with age. Please refer to the NHSP Early assessment guidance (BSA, NHSP 2013).

5. Data collection and analysis

5.1 Amplifier and artefact rejection level

The key to successful testing is a relaxed and sleeping baby. To ensure that unwanted electrical activity does not contaminate the recording we recommend that the artefact rejection levels are set between ±3 and ±10µV. An initial value of no more than ±5µV should be set up in test protocols. If occasionally the background activity is above ±5µV for long periods, it is usually best to wait until the activity reduces. If this does not happen then the rejection level can be raised, but not generally

to more than the maximum of ±10µV. Of course doing so allows more noise into the recording, requiring a substantial increase in the number of sweeps. Appendix D gives more detailed advice, including advice on the use of noise-weighted averaging. Consult the equipment manuals for the value of amplifier gain/sensitivity required to achieve these rejection levels. In some equipment, for each amplifier gain/sensitivity, a range of reject levels can be set.

The use of higher rejection levels when recording conditions are difficult (*e.g.* high muscle activity) will lead to a poorer signal to noise ratio in the averaged signal.

5.2 Blocking of stimulus artefact

At high stimulus levels the stimulus artefact may exceed the artefact rejection level and so prevent recording. Some equipment will allow the artefact rejection to be ignored for a set time after the start of the stimulus to prevent this happening. This is referred to as 'blocking' by some manufacturers. It is suggested that if this facility is available it is set to a default value of 1.5ms for clicks and for the duration of the stimulus for tone pips. Where separate ABR test protocols exist for each frequency then the blocking value should be set to the duration of the stimulus (1.5ms for clicks). Refer to equipment-specific recommendations for the blocking period when using chirps.

If this 'blocking' facility is not available it may be possible to delay the start of the recording to the end of the stimulus artefact. See Appendix C for advice on the display options for the blocking period. If such a delay is introduced, then the time of the delay should be noted in the results and any latency measurements adjusted if necessary.

5.3 Recording Filters

Low frequency (high pass): a value around 30Hz is recommended. This has been found to give the best signal to noise ratio of wave V near threshold. Higher values should be avoided; although less electrical and myogenic noise is recorded the response is also attenuated, making interpretation difficult.

High frequency (low pass): a value around 1500Hz is recommended. There is little response energy above this frequency. A higher value generally adds more electronic noise from the amplifier.

5.4 Use of digital filters (*e.g.* smoothing filter on the averaged waveform)

Digital filtering in the range 50 Hz to 1000 Hz is not recommended for routine use as it is likely to result in a change in the waveform shape and response amplitude which makes interpretation and audit of the waveforms more difficult. For the same reason smoothing filters should not be used and should not be necessary as the 1500Hz filter used in the recording is sufficient to remove any unwanted high frequency noise. The exception to this is when a sloping baseline is recorded, making the placement of markers problematic for wave V and SN10 for the measurement of response amplitude. If digital (display) filtering is available it may be helpful to apply a high-pass filter at 50Hz or more to reduce the extent of the baseline slope and so obtain a more accurate estimate of response amplitude. If used, details should be included in the clinical report, peer review document and S4H entry.

5.5 Notch filter

This will not be required under normal recording conditions and with good electrode practice as 50Hz mains artefact should be absent or minimal. If mains artefact levels are high it is better to identify and remove the source of the problem rather than rely on the use the notch filter. However if there is an unusual and exceptional degree of mains interference which cannot be eliminated the temporary use of a notch filter is preferable to raising the high pass filter or abandoning the test. When a notch filter is used this must be noted in the clinical report. The available evidence (Lightfoot et al. 2014) is that notch filtering does not distort the newborn ABR, with the exception of testing at 0.5kHz where waveform distortion has been observed and could compromise waveform interpretation. At 0.5kHz therefore the notch filter must not be used.

5.6 Window length and averaging

Recommended window lengths are given in Table 1 These values ensure collection of the complete waveform including the SN_{10} component, taking into account tone pip frequency, age of baby and stimulus level.

The number of sweeps per waveform accepted will vary depending on both the size of the response and the level of background activity. The aim is to achieve a clear response or response absent rating (see later) without the need for performing additional averaging runs followed by weighted addition of waveforms, though there will be occasions where this is necessary. In order to provide flexibility, the number of sweeps set up in the equipment protocols should be 4000 or more; the tester may often terminate an average at a lower number of sweeps, as required. The number of sweeps needed will normally vary between 1500 and 3000, although it may be higher when the responses are small or the background noise is high. Typically a figure of 2000 sweeps is recommended (minimum of 1500) for ckABR & NB chirp ABR and 3000 (minimum of 2000) for tpABR.

Exceptionally there may be such a large ABR response or low background activity that fewer sweeps can be used (subject to a minimum of 1500 for tpABR and 1000 for ckABR and NB chirp ABR). Any waveform must still be judged against the full clear response (***CR***) criteria, described later.

In order to resolve inconclusive waveforms additional replications may be needed. To judge these, the waveforms should be combined in a pair-wise fashion, with interpretation based on a single pair of optimally superimposed waveforms.

More detail on the number of sweeps to use is given in Appendices D and E; a strategy of always using a pre-defined number of sweeps is no longer advised. Instead we should use a number appropriate to the prevailing test conditions and make use of objective measurement of response confidence and residual noise, if available. This will maximise test precision and efficiency. However, it is important to continue to use at least the minimum number of sweeps suggested above; objective measurements of response confidence can give spurious results if less than 1000 sweeps are used.

5.7 Display

Always adhere to the convention of plotting wave V upwards. The display should be always set at a fixed number of µV or nV per division (1µV=1000nV). The amplitude (y) and time (x) scales should be

such to ensure that small waveforms near threshold are visible. The broad range of acceptable values is 25- 100nV (0.025-0.1µV) on the response amplitude (y) axis to match 1ms on the time (x) axis. Please refer to the equipment specific settings for how to achieve these in practice.

In some equipment the display aspect ratio of the on-screen and printed waveforms are not the same. It is important to ensure the printed waveforms' aspect ratio is within the ranges recommended above. This is of particular importance in order to facilitate peer review or re-interpretation at a later date. Information on specific scales to use for different types of equipment is available on the EP Resources of the BSA website.

Do not use an automatic display gain as this may set an inappropriate display gain for assessment of the response.

5.8 Masking

As with pure-tone audiometry, masking of the contralateral ear is required in certain circumstances where the stimulus level is high enough to cross to the other cochlea and produce a response. If masking is not used a crossed shadow response may mislead as to the true threshold. A masking noise calculator spreadsheet designed by Guy Lightfoot is available to download from http://abrpeerreview.co.uk/resources.html, and this should be used to alert when masking may be needed and to calculate the level of masking noise required. Details of the background and evidence used is given in an appendix of the NHSP guidelines for early assessment (BSA, NHSP 2013). As a practical guide, the following table may be helpful. It is based on the Noise Calculator spreadsheet 2013b and assumes the non-test ear is normal. Consider the need for masking when using stimuli at or above the following levels (in dBnHL) for babies of 0 to 8 weeks corrected age (consult the spreadsheet for other ages). Note: the stimulus levels for chirps are 5dB lower than shown below.

Transducer	Click	0.5kHz	1kHz	2kHz	4kHz
Supra-aural	65	75	75	65	75
Insert	60	75	75	70	75
BC	20	15	15	25	20

Appendix B on ear-specific testing includes details of masking, 2-channel recording and identifying wave I as means of determining the ear generating the ABR.

5.9 Criteria for accepting the presence of a response

The primary method of establishing the presence and absence of a response is visual interpretation. When objective measurements are available these can be valuable in helping us to be confident in our interpretation and in deciding when to stop averaging (see Appendix E).

Replication of waveforms contributing to the reported result (as defined below) is essential if a correct visual interpretation is to be made. Replication is not needed at other stimulus levels. For example if the first stimulus level is 40dBnHL and a flat waveform is obtained then the best use of time may be not to replicate until a response is observed at higher stimulus levels and it is clear which levels need to be replicated to determine threshold.

Decision criteria for the result at each stimulus level

For each stimulus level the result should be marked in one of three ways.
CR: Clear Response present
RA: Response Absent *, or
Inc: Inconclusive.

*Note that the term ***RA*** was chosen for simplicity and clarity in clinical reporting.

The reader should make reference to Appendix C which contains more detailed advice and examples on this process. The rules for marking the results require two waveforms which are optimally superimposed (displayed on a common baseline representing the average value of all points in the averaged waveform). Where there are more than two recordings these should be combined to form two waveforms (see also later note about when a further pair of recordings have been made).

The interpretation process should be carried out according to the flow diagram in Figure 1.
Firstly, does the result meet the criterion for a clear response (***CR***)?

For ***CR*** there must be a high degree of correlation between the replications and the waveforms should show the expected characteristics in terms of amplitude, latency and morphology.

The size/amplitude of the response (as judged from the wave III/V to the following SN_{10} trough – refer to appendix for examples) should be a minimum of 40nV and at least 3 times the background noise level (the noise level can be estimated from the average difference between the waveforms).

The waveform should be judged over the whole time window excluding any stimulus artefact.

Waveforms should be compared with those at other stimulus levels (where available) to confirm that they follow the expected changes with stimulus level.
See also Appendix E Section 3.1 Modified Criteria for Single (unreplicated) waveform

This criterion ensures a high degree of confidence of the presence of an ABR response. Examples showing where this criterion has been met and where it has not been met are given in Appendix C,

Secondly, if the result does not meet the criterion for a clear response, the question should be asked 'Is this a response absent (***RA***)?'

For ***RA*** the waveforms must be appropriately flat, with no evidence of a response and the average gap (noise) between a pair of optimally superimposed waveforms should be less than or equal to 25nV (using the same method for measuring the background noise for ***CR*** described above).

This average difference (noise) criterion is designed to ensure a small response is not being obscured by noise. The waveform should ideally be flat but this is not always achieved - see Section 5.15 and Appendix F on baseline drift. As with ***CR***, the principle underpinning ***RA*** is that there must be a high degree of confidence that a response is genuinely absent.

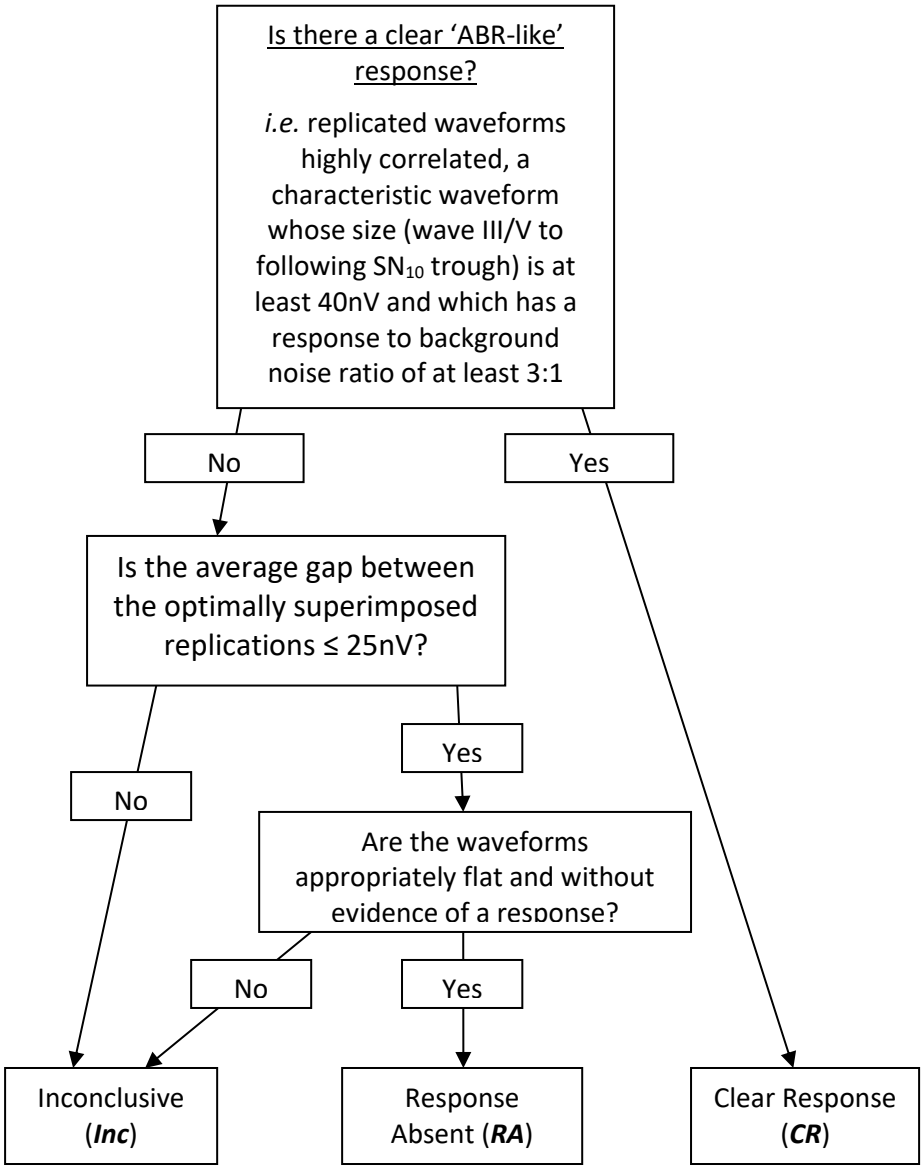
See also Appendix E Section 3.2 Modified Criteria for Single (unreplicated) waveform

Finally, if the result does not meet the criteria for either a clear response (***CR***) or response absent (***RA***), the result should be marked as inconclusive (***Inc***).

Inconclusive waveforms should not contribute to the derivation of threshold. Waveforms that appear to show a likely response or likely absence of response need not be replicated if they do not contribute to the definition of threshold as defined in section 5.11, below, but they should be labelled as ***CR*** or ***RA*** only if replicated. However see also Appendix E Section 3 for an exception to this advice.

Each threshold measurement should continue until there is a very high degree of confidence, with any inconclusive results being resolved.
The above criteria and decision-making process are summarised in Figure 1. Note that Appendix E Section 3 details an optional variation in the flow chart below when objective measurements are used to help determine ***CR*** and ***RA***.

Figure 1: Flow chart showing summary of decision making process (see text for details).



NOTE: If the decision on threshold relies on only one very low amplitude ***CR*** (between 40 and 50nV), confirmatory tests should be carried out 5 or 10dB higher, where the response amplitude should be at least 50nV.

5.10 Resolving inconclusive results

The prime strategy is to test at a given stimulus level until a decision of ***CR*** or ***RA*** can be made.

Where there is some evidence of a response but criteria for ***CR*** are not met, further replications may help resolve the result by reducing the residual noise (this is only needed where the level is around threshold). For example, two further recordings could be carried out and the waveforms could be added in pairs, using weighted addition (see Appendix C, section C5). The waveforms should also be added in a fixed order to avoid bias (we recommend that waveform 1 be added to waveform 3 and that waveform 2 be added to waveform 4). This will produce two summed waveforms, each the average of a higher number of sweeps. If the waveforms are in equally good recording conditions the effect should be to reduce the noise level, and the new pair of waveforms can be examined to see if they meet the 3:1 condition for ***CR,*** or whether they meet the criteria for ***RA***.

On some occasions one may be able to obtain only one rather than two further waveforms. In this case we recommend that waveform 1 be added to waveform 3 and superimposed with waveform 2 to test against the criteria.
When reducing the noise by performing additional runs does not resolve an inconclusive result then a blocked stimulus run (see also Section 5.15) can be helpful. This technique is particularly valuable when a likely ***RA*** or ***CR*** is recorded at the maximum available stimulus level.

In making a decision at a given stimulus level ***all*** recorded waveforms at that stimulus level should be considered. A waveform may be rejected only where there is a good technical reason or where the noise levels (*e.g.* electro-myogenic activity) were untypically high. We cannot cherry-pick waveforms simply because they demonstrate a favoured result.

5.11 Definition of ABR threshold for NHSP

ABR threshold has been defined in the NHSP early assessment guidelines (BSA, NHSP 2013) as the lowest level at which a clear response (***CR***) is present, with a response absent (***RA***) recording at a level 5 or 10dB below the threshold[m], obtained under good recording conditions.

This is the definition that should be used for entering results onto S4H. Refer to the NHSP early assessment guidelines for the use of ABR threshold in management.
Independent auditing / peer review of the results should not give thresholds more than 10dB different from those originally recorded.

5.12 Gold standard thresholds

The term 'gold standard' has been given to the combination

an ear-specific clear response (***CR***) at threshold
a clear response (***CR***) at 10dB or 5dB above threshold, and
a response absent (***RA***) at 10dB or 5dB below threshold.

[m] The reason for including a replicated response at 10dB or 5dB above threshold as well as one at threshold is that it provides a safeguard against the possibility that the response at threshold is spurious. The response at 10dB or 5 dB above threshold should be larger and clearer than that at threshold. If, on audit, there is disagreement about the clear response at threshold, then the response at the higher level will normally limit the disagreement to 10dB or 5dB and prevent any serious error. If the response morphology across levels is abnormal a blocked stimulus run may be necessary to confirm the validity of the responses.

To minimise errors in estimating ABR threshold (see note below), the gold standard should be obtained for at least one AC ABR threshold and one BC ABR threshold (where carried out) for each ear.

The minimum of one AC and one BC threshold in each ear to the gold standard will normally apply to one of the frequency-specific ABR thresholds, ideally 4kHz and be ear-specific. When the gold standard has been obtained at 4kHz, the gold standard may also be claimed (e.g. for peer review purposes) at other frequencies when it has been obtained, even though it is not strictly necessary, including in the circumstances detailed below.

The definition of the gold standard needs to be modified slightly in the following situations:

(a) Where the minimum discharge criterion of <=30dBeHL for 4kHz AC ABR is met; gold standard is met by recording a ***CR*** at 30dBeHL and at 35 or 40dBeHL[n]. An ***RA*** at 10dB or 5dB below threshold is not required.

(b) Where the minimum criterion of ≤20dBeHL for 4kHz BC ABR is met and there is evidence that the BC response is ear-specific, Gold standard is met by recording a ***CR*** at 20dBeHL and at 25 or 30dBeHL. An ***RA*** at 10dB or 5dB below threshold is not required. Whilst <=20dBeHL (ear-specific) would satisfy the minimum criteria, in practice it may be pragmatic to test to <=15dBeHL to remove the possibility of a crossed response without recourse to masking or other ear-specific techniques in babies <12 weeks old.

(c) Where it is not possible to test at 10dB or 5dB above threshold due to the maximum stimulus level having been reached. In this case at least one further recording should be carried out at threshold but displayed slightly above or below the superimposed pair. This allows estimation of the residual noise from a single gap between the first two waveforms yet still facilitates a comparison of response features. An alternative is to perform a blocked stimulus run (see section 5.15 and display as suggested above); if the response is genuine then it will disappear when the stimulus is blocked.

(d) Where there is ***RA*** at the maximum stimulus level. The ***RA*** should be of good quality but 2 recordings suffice. The gold standard status for this scenario remains but the implication this has for tests at other frequencies has been changed. If an ***RA*** is obtained at one frequency (as above) yet clear responses are recorded and the threshold defined at another frequency in the same ear for the same transducer, that result should independently meet the gold standard. Example: an ***RA*** at maximum stimulus level is recorded at 4kHz and a response is recordable at 1kHz. Even though the 4kHz ***RA*** is gold standard, gold standard recording requirements should be met when defining the 1kHz threshold. This is to avoid the situation where the gold standard associated with the ***RA*** at the maximum stimulus level removes the requirement for a gold standard where a response is seen.

[n] For the derivation of the dBeHL value from the dBnHL value see section entitled 'Prediction of the estimated hearing level threshold' in the NHSP early assessment guidelines.

5.13 Testing at other frequencies

One should usually achieve a gold standard threshold at 4kHz before moving on to other frequencies.

Once a gold standard has been achieved at one frequency (for at least one AC ABR threshold, and one BC ABR threshold where carried out) for each ear, this can be relaxed for other frequencies. There should still be a high degree of certainty for each threshold measurement and any inconclusive results resolved, but there is no requirement to also test at 10dB or 5dB above threshold (the confirmatory level).

As with the 'gold standard', waveforms should always be replicated if they are used in the definition of the ABR threshold. Some examples will help to illustrate this. If there is any doubt in the result then revert to the 'gold standard procedure'.

Example of threshold measured to the 'gold standard'
See Appendix C, Figure 5.

Example of threshold not measured to the 'gold standard'
If the AC 4kHz tpABR threshold is measured using the 'gold standard', then if testing at AC 1kHz tpABR, a ***CR*** at 70dBnHL and ***RA*** at 60dBnHL is sufficient to report the threshold as =70dBnHL

5.14 Reporting thresholds (including those which are not gold standard)

Results should be clearly marked clearly using the symbols '=', '≤' or '<=', and '>', and notes should always be made of any limitations or caveats about interpretation so this information is available to those who may carry out further tests.

For example
'=45dBnHL' means ***CR*** at 45dB (and 5-10dB above for 'gold standard'), and **RA** at 35 or 40dB.
'<=45dBnHL' means ***CR*** at 45dB but not tested (or inconclusive) below this level.
'>80dBnHL' means ***RA*** at 80dB, but not tested (or inconclusive) above this level.

Where the gold standard defined above has not been achieved, threshold should be reported as follows:

a) If no 'confirmatory' ***CR*** Is obtained at 5 or 10dB above threshold, report threshold = lowest ***CR*** obtained
e.g. ***CR*** at 70dB, ***RA*** at 60dB, threshold= 70dBnHL.

b) If no ***CR*** is obtained above an ***RA*** result, report threshold > highest ***RA***;
e.g. ***Inc*** at 70dBL, ***inc*** at 60dB, ***RA*** at 50dB, report threshold as >50dBnHL.

c) If an ***RA*** response is obtained but not within 10dB of the lowest ***CR,*** report threshold as <=lowest ***CR*** and > highest ***RA***;
For reporting in S4H (because only a single value can be entered) apply the following rules
1. 15dB or 20dB gap between lowest ***CR*** and highest ***RA*** -

e.g. ***CR*** at 70dB, ***Inc*** at 60dB, ***RA*** at 50dB. In S4H: enter =70dBnHL (in preference to <=70dBnHL), with a note in the session summary that the threshold is in the range 55 to 70dBnHL.
2. More than 20dB gap between lowest ***CR*** and highest ***RA*** -
e.g. ***CR*** at 70dB, ***Inc*** at 60dB, ***RA*** at 40dB. In S4H: enter <=70dBnHL, with a note in the session summary that the threshold lies in the range 45 to 70dBnHL.

d) If no ***RA*** is obtained below a ***CR***, report threshold as <=lowest ***CR***;
e.g. ***CR*** at 70dB, ***Inc*** at 60dB, ***inc*** at 50dB, report threshold <=70dBnHL

Note that the situations in c) and d) are best avoided, and efforts should always be made to reduce the gap between ***CR*** and ***RA*** to 10 dB or less when testing.

In all these cases all clinical factors should also be taken into account, particularly where thresholds are being used for fitting hearing aids, and it is important to avoid over-amplification.

When wave I presence or two-channel recording is used as an alternative to masking and shows that the ipsilateral ear is responsible for the response, it is recommended that the threshold is entered into S4H and the clinical report using the qualifier (M) (as though it had been masked), and adding an appropriate clinical note. If wave I presence or two-channel recording does not indicate that the ipsilateral ear is responsible for the response (because it is unclear or that the contralateral ear is responsible for the response) then the qualifier (M) should not be used and a clinical note added which warns that the result is not ear-specific.

5.15 Baseline drift and the use of blocked-stimulus runs

This term (also known as a sloping baseline) is used to describe non-flat recording baselines such as those due to large stimulus artefacts. Other types of artefact may also give rise to baselines that drift such as contamination of the recording with cardiac activity. A moderate amount of baseline drift is acceptable if it does not affect the ability to observe an ABR response and it cannot be taken for a false ABR response (*e.g.* baseline drift due to stimulus artefact should end before the first key component of any ABR response). If these rules regarding baseline drift are not met then the result should be considered as inconclusive. Where doubt exists in the possibility of genuine response presence or absence a blocked-stimulus run can sometimes help resolve the matter. Appendix F contains a series of examples to help the reader distinguish between what can be called ***RA*** and what is ***Inc*** when 'baseline drift' is present and of the role of blocked-stimulus runs.

5.16 Post-auricular myogenic (PAM) responses

Occasionally a post-auricular myogenic response which occurs around 12-15ms will be recorded when the ABR response does not meet the ***CR*** criteria. This could occur in a baby who is not asleep or relaxed. If the state of the baby results in high levels of myogenic noise, and increased numbers of sweeps fail to resolve an ABR waveform to ***CR*** or ***RA*** status, the tester should consider changing the position of the baby and/or electrodes and if this does not resolve the situation the a further test session should be considered. A PAM should not be considered as a ***CR***.

6. Calibration

A subjective 'Stage A' check should be carried out on the ABR equipment prior to use (NHSP Clinical Group 2008).

ISO 389-6 (2007) provides reference equivalent threshold sound pressure levels (RETSPL) for click and tone pip stimuli used for AC ABR for certain types of transducer. It also provides a standard for reference equivalent threshold force levels (RETFL) for use for click BC ABR for the B71 type bone vibrator. However there are no RETFLs for tone pip BC ABR. A procedure for calibration is given in IEC (2007). In July 2005 a provisional set of reference levels for AC and BC ABR was agreed for NHSP; including reference threshold values for narrow band CE-chirps, ("NHSP Recommended Stimulus Reference Levels for ABR Systems, 2012" 2012), with the ISO389-6 values used where appropriate. These values should be used to calibrate equipment used for hearing assessment.

It is important to note that the stimulus rate affects the psycho-acoustic threshold but not the ABR threshold (Lightfoot et al. 2007). When performing a 'stage A' listening check therefore, using the recommended stimulus rate of close to 50 per second will make the stimuli sound about 3dB too loud. Bear this in mind or perform the listening check at 20 per second.

The equipment should not change the physical intensity of the stimulus when the stimulus repetition rate is changed since this would introduce errors. Detecting if it does is difficult subjectively and further information on this is available in the equipment-specific ABR parameter document.

7. Artefacts

7.1 Recording system checks

The equipment should be checked at regular intervals (weekly is recommended) for system artefacts.

Using the normal protocol for testing, use a "dummy patient" resistance network, "loop back" box or, if not available, connect the electrodes together. Run the normal protocol twice and check that flat waveforms are obtained with a minimal level of residual system noise (peak-to-peak below 50nV and no significant correlation between repetitions indicating system artefacts).

At regular intervals (monthly is recommended), when testing a baby, take the opportunity to carry out an additional control recording (a blocked-stimulus run) to check that there are no artefacts in the recording system. Do this on a baby where clear ABR responses have been obtained at discharge levels. Set the stimulus to 30dBnHL and block the sound from reaching the ear (see next section on how to do this (30dBnHL is chosen as it may not be possible to completely acoustically block a high stimulus level). Obtain replicated waveforms with the stimulus still on (do not reduce the level of the stimulus as this invalidates the check). A pair of waveforms should be obtained which meet the ***RA*** criteria defined above.

7.2 Control (blocked stimulus) recordings during testing

Control recordings should be carried out whenever the ABR response is marginal and/or is of the form that could be an artefactual response - *e.g.* mains artefact that is time locked to the stimulus could result in replicated waveforms that mimic some types of ABR response.

When carrying out the control recording the stimulus should remain at the test level but prevented from stimulating the ear. If the response is artefactual, and is not an electro-physiological response to the sound stimulus, it will still be present. Turning the stimulus level right down is not appropriate.

Note on how to achieve stimulus blocking.
For AC ABR the acoustic block can be in the form of a cover for earphones, or a tubing clamp for insert earphones, and should give a substantial reduction (>30dB) in the sound level. Note that as control recordings are carried out where responses are marginal and therefore close to the ABR threshold, a 30dB reduction is normally sufficient although one may see a response from the contralateral ear if that has a much better threshold.

For BC ABR the bone conductor can be lifted a few millimetres from the scalp to prevent transmission of the sound. Note that a response may still occur by air-conduction of the stimulus and if necessary the bone conductor should be covered to reduce the airborne sound.

Touching the baby's skin may change the extent of mains-related activity recorded, but it may be very difficult to avoid this when undertaking a control recording.

Action required
If artefactual responses are observed, then it is essential to determine their source and remove them from the recording process. Advice should be sought and if necessary manufacturers contacted, so that the source of the artefacts can be eliminated.

Appendix A: Major changes in this recommended procedure

New advice has been added to Section 5.4 on the use of digital filtering on waveforms with a sloping baseline.

Section 5.6 has been changed to reflect a move away from a fixed-sweeps approach towards the use of a number of sweeps appropriate to the prevailing test conditions.

The definition of Gold Standard thresholds in Section 5.12 has been expanded; ear-specific thresholds are required. An ear-specific BC threshold of ≤20dBeHL or lower does not require an RA to qualify as Gold Standard.

Text previously in Section 5.13 (testing at other frequencies) has been moved to 5.10 (resolving inconclusive results) for clarity.

Section 5.16 on Post-auricular myogenic (PAM) responses has been changed. A PAM may no longer be used for the definition of ***CR***. Previous versions of this guidance (2010, 2013) allowed a waveform to qualify for ***CR*** on the basis of the PAM, even when the ABR element of the waveform was inconclusive (for example because of excess noise) or absent. This advice has been withdrawn, because of the dangers of misinterpreting an abnormal morphology ABR as a PAM. Current advice is to base interpretation on a recognisable ABR only.
Appendix B has been widened to include methods for ensuring ear-specific testing using the presence of wave I.

An expanded Appendix E (Objective measures for ABR interpretation in babies) includes revised criteria for the acceptance of ***CR*** & ***RA*** waveforms without replication in certain circumstances.

A new Appendix F on baseline drift has been written and some examples previously in Appendix C have been moved to Appendix F. The interpretation of some existing examples has been revised.

Appendix B: Ear-specific recording

There will be circumstances in which the ear responding to test stimuli is in doubt, for example in cases of asymmetric hearing loss, especially when testing by bone-conduction. The classical means of ensuring ear-specific results is to apply a masking technique (Section 5.8). In many instances where there is a risk that a response could be crossed in fact it is not. As an alternative to masking therefore it may be sufficient to demonstrate that a potentially crossed response is not crossed. This is particularly attractive in cases of “masking dilemma”, where cross-masking is a possibility such as in a bilateral conductive loss. Two methods are available: 2-channel recording, where the characteristics of wave V in the two channels are compared) and identifying wave I, which can be recorded in single-channel or 2-channel tests. If it is not possible to demonstrate that the response is not crossed (either because it is crossed or because the result in uncertain) then masking will be necessary.

B1. 2-channel recording of wave V

The purpose of two channel recording is to record both the ipsilateral and the contralateral response in an attempt to determine which cochlea is generating the ABR. Here, we look at the wave V asymmetries between the ipsilateral and contralateral recordings; the channel with the larger amplitude and earlier latency wave V is likely to be the side generating the response. If this is ipsilateral to the stimulated side then the response is not crossed and no masking is needed. If the larger & earlier response is in the channel contralateral to the stimulated side (or if the result is unclear) then crossover may be occurring and masking should be used to determine the test-ear threshold.
This method does not identify the correct side in 100% of cases, may falsely label some unilateral conductive losses as sensorineural and although it may be useful, masking should be regarded as the definitive method. It is important to note that the 2-channel method is helpful only in newborns (note the ages of patients in the studies shown in Table 2) and may give erroneous results in older children, where masking should be used.

Electrode positions for 2-channel recording:

- Positive electrode: high forehead as near to Cz as possible, and midline. The fontanelle should be avoided but the electrode should be placed as close as possible to this otherwise the ABR response will be reduced in size

- Negative electrodes: ipsilateral mastoid and contralateral mastoid. Common electrode: forehead (at least 4cm from the positive electrode).

The above electrode configurations should result in wave V being plotted upwards on the display. If this is not the case then the positive and negative electrode connections should be reversed.
Table 2 provides a summary review of the literature on 2-channel BC ABR.

Table 2 Literature summary: comparison of ipsilateral and contralateral Bone-Conduction ABR waveforms in infants

Study	N	Age of infants	Hearing status	ABR or ASSR	Stimulus Freq (Hz)	Amplitude	Latency	Difference in threshold	% with absent contra responses
Foxe and Stapells 1993	21	2 weeks to 13 months	Normal	ABR	0.5 & 2k	Mean contralateral response significantly smaller (size of the difference not reported)	0.5kHz Mean approx 2ms later in contra 2kHz Mean approx 0.5-1 ms later in contra*		
Stapells and Ruben 1989	48	2 weeks to 2 years	Normal and conductive hearing loss	ABR	0.5 & 2k	0.5k Hz Mean contra approx 0.1uV less in contra*. 94% had smaller contra response 2kHz 93% had smaller contra response	0.5kHz Mean approx 1ms later in contra *		0.5kHz At 40 dBnHL 0% At 30 dBnHL 5% At 20 dBnHL ~ 25% At 10 dBnHL ~50% At 0 dBnHL ~75%
Small and Stapells 2008	14	8 to 44 weeks	Normal (passed DP screen)	ASSR	0.5 to 4k	Mean contra response was 57-73% of mean ipsi (all frequencies)		4kHz Mean 14.2 dB worse in contra ear SD 5.2 Range 10-20 N=12	34% (unclear which frequency)

*No range or SD presented

Foxe JJ, Stapells DR. *Normal infant and adult auditory brainstem responses to bone-conducted tones*. Audiology. 1993;32(2):95-109.

Small SA, Stapells DR. *Normal ipsilateral /contralateral asymmetries in infant multiple auditory steady-state responses to air- and bone-conduction stimuli.* Ear Hear. 2008 Apr;29(2):185-98.
Stapells DR, Ruben RJ. *Auditory brain stem responses to bone-conducted tones in infants*. Ann OtolRhinolLaryngol. 1989 Dec;98(12 Pt 1):941-9.

Figure B1 gives an example where the BC stimulus generates the ABR via the ipsilateral ear. The contralateral waveforms have a lower amplitude and longer latency wave V so in this case it is safe to ascribe the responses to the left ear without recourse to masking.

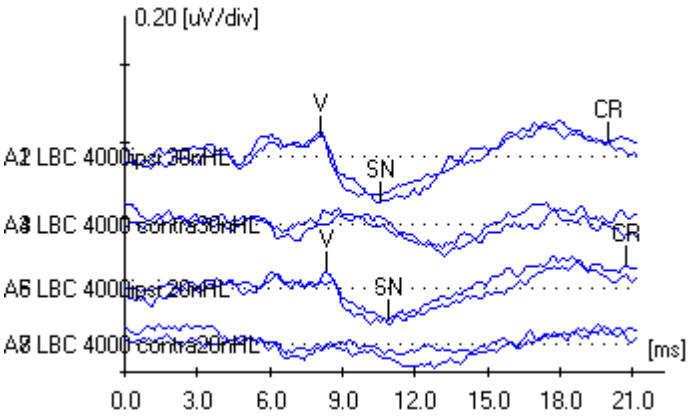


Figure B1: 4kHz BC waveforms (top to bottom): 30dB ipsi; 30dB contra; 20dB ipsi; 20dB contra.

B2. Identifying wave I

Wave V is generated in the lateral lemniscus in the upper brainstem, anatomically close to the midline and therefore it can be considered as a "far field" response, usually recordable in both ipsilateral and contralateral channels. Conversely, the generator of wave I is anatomically in the cochlea, far closer to the ipsilateral mastoid electrode than the contralateral mastoid electrode so is considered as a "near-field" ipsilateral response and cannot be seen in the contralateral channel. Identifying wave I in the ipsilateral channel is therefore confirmation that the response is not crossed. The absence of wave I in the ipsilateral channel cannot be taken as evidence of a crossed response since wave I is not always recordable (in either channel), especially close to threshold.

This method was originally described as a 2-channel technique in much the same way as the 2-channel wave V method but a recent study (Ferm and Lightfoot 2016) demonstrated the utility of a single-channel variant (using normal stimulus rates) and identified that the use of NB-Chirps increases the likelihood of identifying wave I close to threshold compared to tone pips at the same stimulus frequency. Thus, when there is a risk of cross-hearing (refer to the table in section 5.8) one should first examine the waveform and if wave I is unequivocally present then we can be confident that the response is not crossed. Wave I should be present in both waveforms and larger enough to be confident it is not residual noise.
Figure B2 is an example of single-channel recording (4kHz BC, left) in which wave I is seen. The right ear was normal so this could have been a crossed response. For the purposes of this example, masking has also been applied, confirming that the response is not crossed.

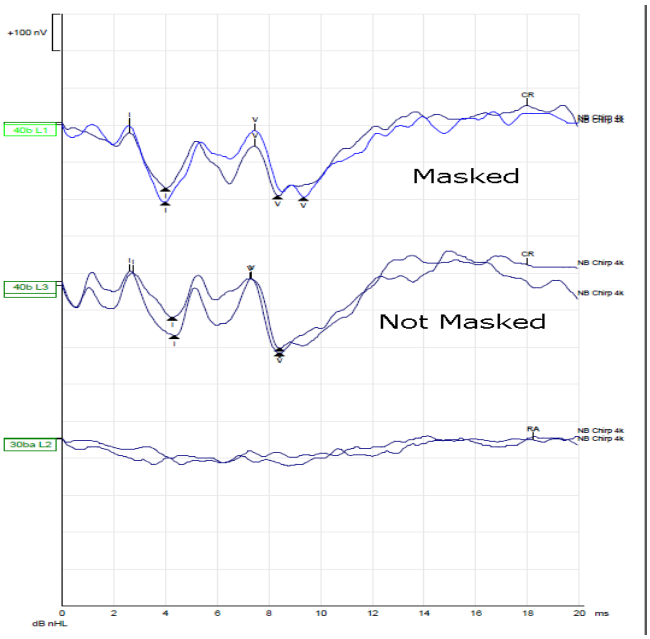


Figure B2: Single channel 4kHz BC waveforms.

B3. Reporting ear-specific results

When masking has been used to ensure an ABR threshold is ear-specific we should insert "(M)" after the result (*e.g.* =45dBnHL(M)) in the clinical report, S4H entry or peer review spreadsheet. Likewise, if 2-channel wave V or wave I tests have confirmed that a response is not crossed "(M)" should follow the result to denote that the result is ear-specific. A clinical note should be made to provide details.

Appendix C: More detailed advice and examples of ABR waveforms meeting the response criteria CR, RA and Inc

C1. Example of a clear response (*CR*), satisfying the 3 to 1 signal to noise criterion

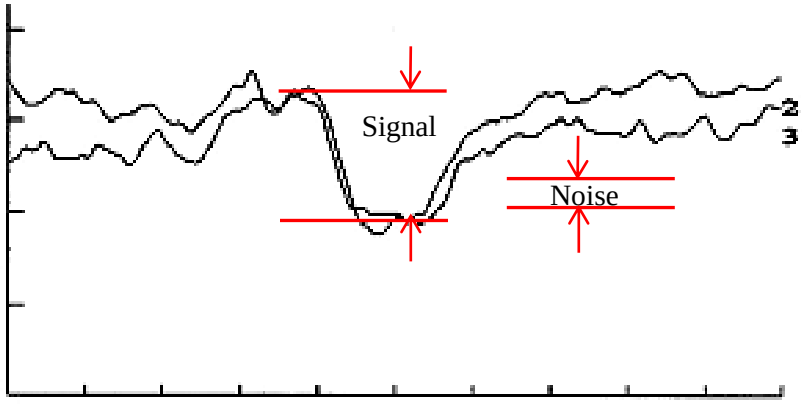


Figure C1: Example of a clear response (CR) using the wave 5 peak and SN_{10} trough to measure signal amplitude Scales: 150nV / div. 2 ms / div

It is important to remember, as detailed in the main text, that as well as meeting the 3:1 signal to noise criteria the waveforms must show the expected characteristics in terms of amplitude, latency and morphology.

In Figure C1 the two waveforms have been superimposed so as to minimise the average gap between them. Some systems have a "superimpose" function which does this optimally. If superimposing the waveforms manually do not simply align the wave V peaks. Rather, position the waveforms so as to minimise the total area between them (ignore any region containing a stimulus artefact). The "signal"[o] is taken as the average of the two waveforms' wave III/V to SN_{10} trough amplitude: in this example it is one and a third divisions (200nV or 0.2µV). The noise is the gap between the waveforms, averaged across the entire window, excluding any region of stimulus artefact, and in this example it is about a

[o] The term "signal" is used to denote the ABR response and has nothing to do with the sound stimulus. "Signal to noise ratio" can be interpreted as the "response to noise ratio".

third of a division (50nV). The signal to noise ratio in this example is therefore 200/50 (working in nV) which is about four to 1. Note that for ***CR*** there is no need for the noise to be less than a specific value – it just needs to be less than a third of the response amplitude. This, together with a characteristic waveform of more than 40nV amplitude allows us to regard this as a clear response (***CR***).

The points taken to measure the response amplitude are usually the wave V peak and the SN_{10} trough but there are a few issues that should be highlighted. Firstly, when measuring the vertical position of a peak (or trough) take the average vertical position of each replicate's peak, not the highest (or lowest). Secondly, wave III is occasionally more positive than wave V and under such circumstances it is appropriate to use wave III instead of wave V as the positive point from which amplitude is measured since it is a valid component of the ABR. Figure C2 illustrates such a case. Finally, it is important to note that SN_{10} is not necessarily the first dip or trough following wave V. In Figure C2 there is such a dip, marked X (at higher test levels this would probably be the trough between waves V & VI) but, for the purposes of amplitude measurement, we should mark SN_{10} as the lowest point in the waveform following wave V (providing the waveform is not sloping – see appendix F). Although this definition of SN_{10} may not be what is seen in textbooks, we use the term as a convenient label for the reference point when measuring response amplitude.

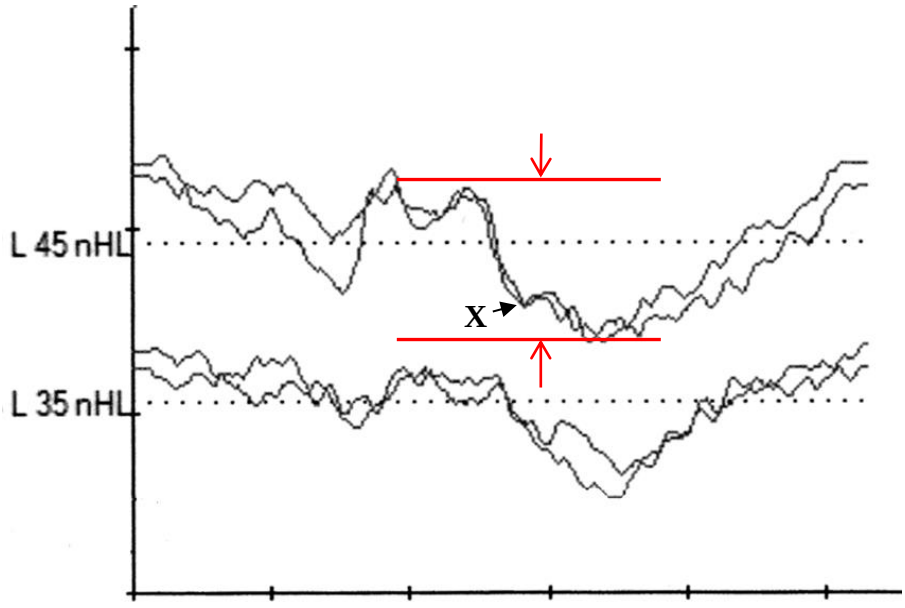


Figure C2: Example of a clear response (CR) using the wave III peak and SN_{10} trough to measure signal amplitude
Scales: 300nV / div. 4 ms / div

C2. Estimating residual noise: deciding between response absent (*RA*) and inconclusive (*Inc*)

As with the assessment of clear responses, the assessment of waveforms for ***RA*** & ***Inc*** status should be across the entire window but excluding any region of stimulus artefact. With the waveforms optimally superimposed estimate the residual noise from the average gap between the pair of waveforms. For ***RA*** status at a given test level the average gap must be no more than 25nV.

Figure C3 contains a possible response and the waveforms have been superimposed in order to evaluate the signal to noise ratio and residual noise level. The signal to noise ratio as illustrated in Fig C1 is <3:1 so it cannot be classified as ***CR***. Note that one "response" is twice the size of the other - this is quite possibly just noise. The average gap between the waveforms is about 40nV so it fails the 25nV ***RA*** noise criterion too. We must therefore categorise this as inconclusive (***Inc***).

If it were important to resolve this level into ***CR*** or ***RA*** then a further pair of averaging runs would be needed, to reduce the noise level. In order to interpret the resulting four waveforms they should be combined pair wise using a "weighted add" (see section C5 below) function and the resulting two waveforms' signal and noise should be reassessed. When combining waveforms it is important not to choose which pairs to combine on the basis of their shape. For consistency it is recommended that waveforms are always combined thus: 1 & 3; 2 & 4.

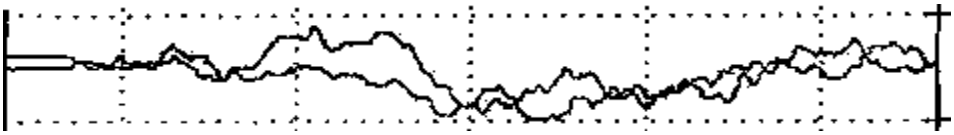


Figure C3: An inconclusive response due to the 3:1 amplitude rule for a CR not being met and the <25nV noise between waveforms not being met for a RA.
Scales: 150nV / div. 4ms / div

Figure C4 below gives two examples of replicated waveforms that just meet the requirements for ***RA*** *i.e.* the conditions for a ***CR*** have not been met and the average gap between optimally superimposed waveforms is less than 25nV. Note that it is the *average* gap we must estimate, excluding any region of stimulus artefact. There are several points at which the gap is zero (where the waveforms cross) and other points where the gap is relatively large. In both these examples the average gap is around one fifth of a vertical division: (25nV).

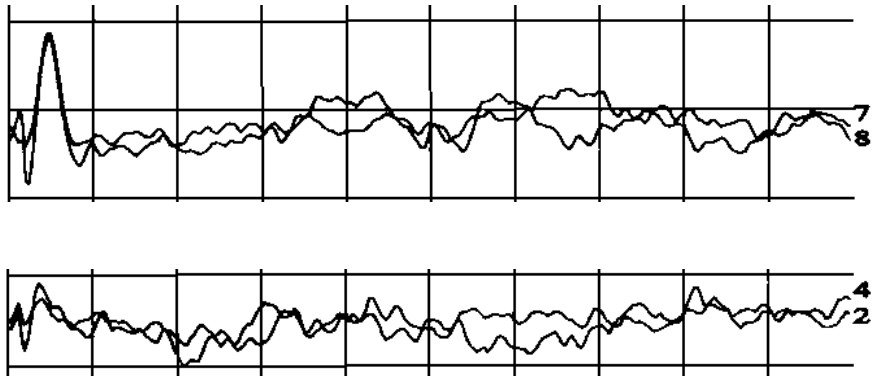


Figure C4: Waveforms that meet the criteria for RA.
Scales: 120nV / div. 2 ms / div.

C3. An example of ABR threshold being found ('gold standard')

Figure C5 shows a correctly displayed and interpreted ABR series using 1kHz tone pips (note the 2.5ms / div scale giving a 25ms window). Stimulus levels are in dBnHL. The waveforms at 60dB and 50dB are ***CR.*** At 40dB there is no candidate response and the average gap between the waveforms is about a fifth of a division or 24nV so it meets the maximum noise criterion and 40dB can be graded as ***RA***.

The ABR threshold is therefore 50dB and 'gold standard' since there are ***CR***s at 50dB and at 5 to 10dB above (in this case at 60dB) and there is an ***RA*** at 5 to 10dB below (in this case at 40dB).

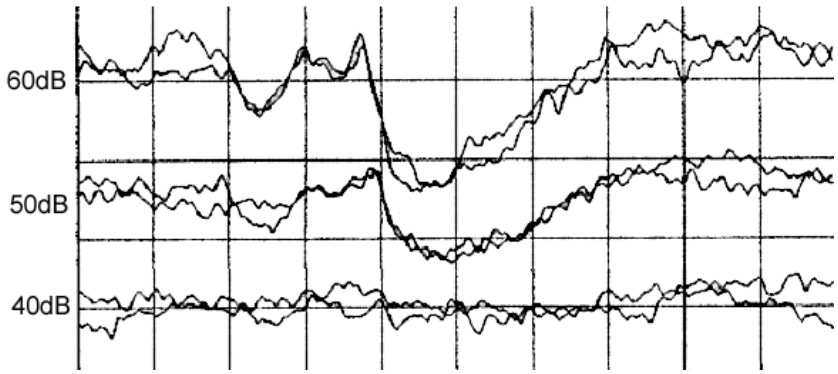


Figure C5: Waveforms meeting the 'gold Standard' of test performance for a 1kHz Tone Pip.
Scale: 120nV / div. 2.5 ms / div.

C4. The use of blocked-stimulus control runs to assist interpretation

Figure C6 gives an "intensity series" for BC clicks (note the stimulus artefact that one often sees when using a bone vibrator). Clear responses are evident at 45dB & 40dB but how do we interpret the waveforms at 35dB & 30dB? Both have a residual noise of less than 25nV but are these ***CR*** or ***RA***? For ***CR*** there needs to be a high degree of correlation between the replications, a characteristic waveform, a minimum response amplitude of 40nV and a signal to noise ratio of at least 3:1. At 35dB the replications are highly correlated, the response is a fraction under half a division: 60nV, and, given the waveforms at higher levels, an acceptable response morphology. The noise is 20nV so the signal to noise ratio is 3:1. This level can therefore be accepted as a ***CR***. The same cannot be said of the 30dB waveforms since the wave V- SN_{10} response is 40nV, the signal to noise ratio is below 3:1 and the response morphology is more questionable. The noise is about 25nV but given the possible response it is not sufficiently flat to be ***RA***. The activity before wave V is higher (perhaps the vestiges of wave III) and if we allow this to determine the response amplitude the signal to noise ratio is >3:1.

This is a genuinely difficult case. The control run (with the vibrator slightly lifted from the baby's mastoid) was performed in an attempt to resolve the 30dB waveforms. These are flatter than the result at 30dB and this supports the possible presence of physiological activity in the 30dB result. As a consequence, 30dB can be taken as ***CR***. Had the control run not been performed (or had not produced flat waveforms) 30dB would have to be interpreted at ***Inc***. [Note that in the 2010 version of this document the 30dB was accepted as ***RA****]*. The current definition of ***RA*** (requiring no evidence of a response) and the use of wave III together with the control run allows us to accept this as ***CR***.

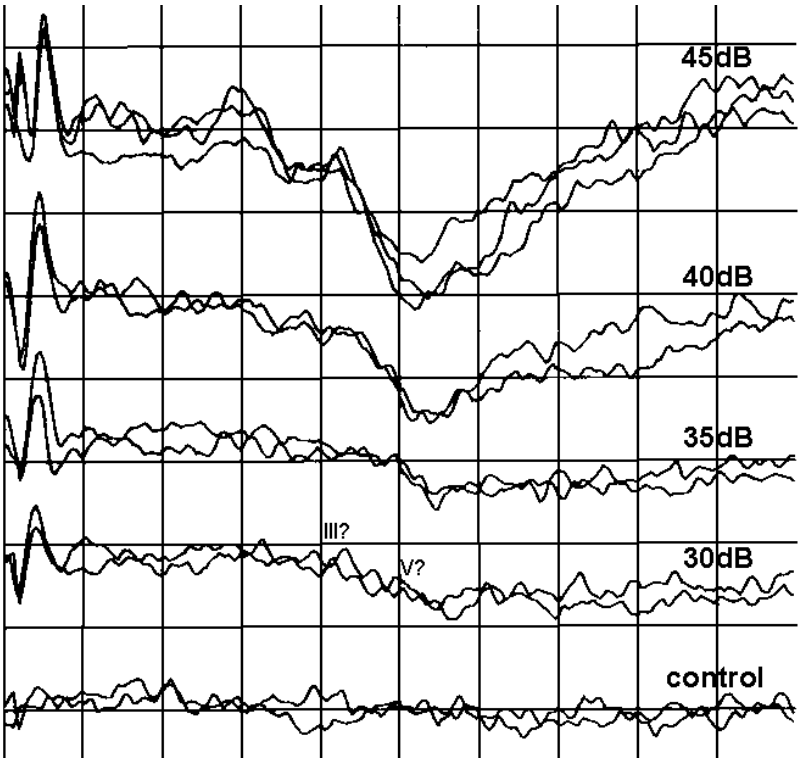


Figure C6: The "intensity series" for BC clicks showing a CR at 45dB, 40dB, 35dB (and also 30dB confirmed by the controlled run).
Scales: 120nV / div. 2 ms / div.

Figure C7 (4kHz AC) shows a ***CR*** in the upper waveforms but interpretation of the lower waveforms (at 10dB lower stimulus level) is more challenging. There is a high degree of correlation between the replications but the expected characteristics are not met in terms of latency as required under ***CR*** criteria. The next stage in the process is to ask if the criteria for ***RA*** *have* been met. The noise content (based on the average gap between replicates) is about a sixth of a division (20nV) so the *noise* criterion for ***RA*** is satisfied. However we cannot be as certain that a response is truly absent. The only safe interpretation here is inconclusive (***Inc***) even though the criteria for ***CR*** have been partially met (signal to noise ratio, response size (if considered a response) and the residual noise) and the criterion for residual noise has been met for ***RA***. In other words this is a very difficult example but the tester should be guided by the question: '**Do you have a very high degree of confidence that your interpretation is correct?'** In this case the answer has to be 'no' so ***Inc*** is the only safe conclusion. A control run may have resolved this case.

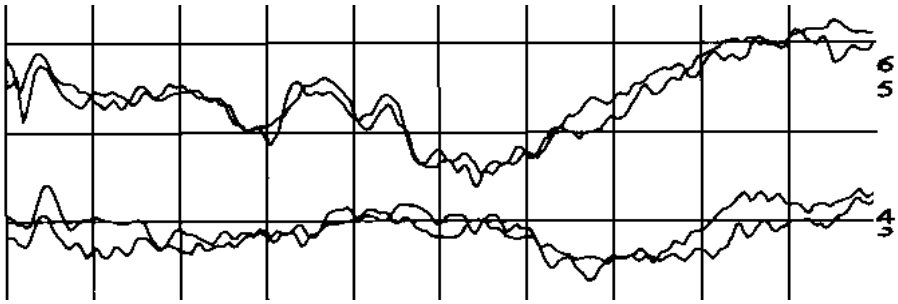


Figure C7: 4kHz AC waveform showing CR in the upper waveforms and an Inc in the lower waveforms which may have been resolved if a 'control run' had been performed.
Scales: 120nV / div. 2 ms / div.

C5. The use of A+B and A-B facilities

Most ABR systems offer the facility to *add*, *merge* or *combine* waveforms (often referred to as A+B) and *subtract* waveforms (referred to as A-B or *difference*). Both can be "weighted" or "unweighted" calculations, which lead to different results. A weighted addition/subtraction scales the contribution of each averaged waveform to the sum by the number of sweeps in each average. An unweighted addition/subtraction does not use this scaling. **It is very important that the correct type (*i.e.* weighted or unweighted) is used as specified in the following advice.**

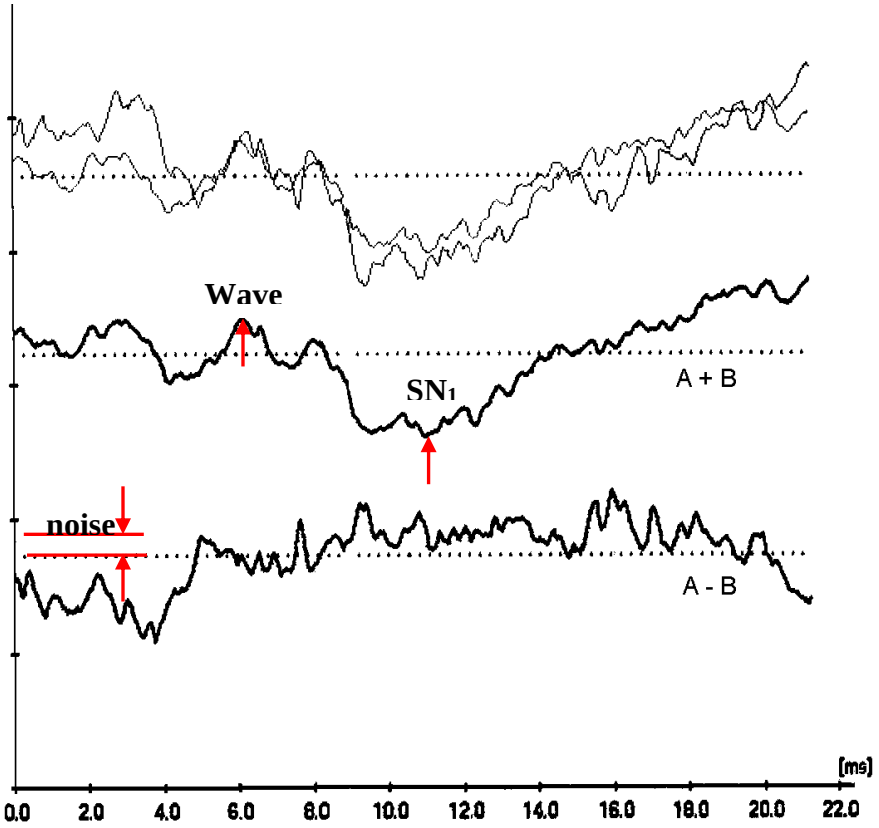
The addition of waveforms using a ***weighted*** A+B function can be useful when assessing the size of the response.

The use of ***unweighted*** A-B offers an alternative method for assessing the magnitude of residual noise as used in ***CR*** and ***RA*.** The two measurements are equivalent and the same criterion of 0.025µV (25nV) is used for ***RA***. This is not considered to be any easier than the method described in the main text on estimating the average gap between waveforms as it still requires the average of a waveform to be visually estimated and, as decisions need to be made quickly whilst testing, may use up valuable test time. However the process is now described for those who wish to use this method.

'A+B/A-B' Example 1

Figure C8 (top) illustrates a pair of waveforms recorded to 55dBnHL clicks that are to be assessed. The requirements for ***CR*** are met in terms of a characteristic waveform but we need to be sure the signal to noise ratio is at least 3:1 and that the response is >0.04µV (40nV). The middle waveform is the weighted addition (A+B) of the two top waveforms. In this case wave III is more positive (higher) than wave V so we can use the wave III – SN_{10} amplitude to represent the size of the response, which is 7/8 of a division or 175nV. The lower waveform is the unweighted difference (A-B). This represents the residual noise in the two upper waveforms and to assess the extent of the noise we must estimate the average height of the waveform above and below the baseline, averaged across the entire width of the waveform. This is shown in the figure and is about 1/6th division, or 30nV. This is the same as the average gap between the

original waveforms. The signal to noise ratio is therefore 175/30 or just under 6:1, easily exceeding the minimum 3:1 requirement - we are therefore able to classify this as ***CR***.



'A+B/A-B' Example 2

Figure C8: Click waveforms at 55dBnHL showing how the signal amplitude can be measured from a weighted add and the noise level measured from an unweighted subtraction of the replicated waveform.
Scales: 200nV / div. 2ms / div.

The waveforms in Figure C9 (top) were evoked by 4kHz tone pips at 40dBnHL and there is no question they can be classified as a ***CR***. The response (middle waveform, weighted A+B) is 1.3 divisions so is 260nV. The noise (lower waveform, unweighted A-B) is estimated as shown, at 1/8th division or 25nV. The signal to noise ratio is therefore over 10:1.

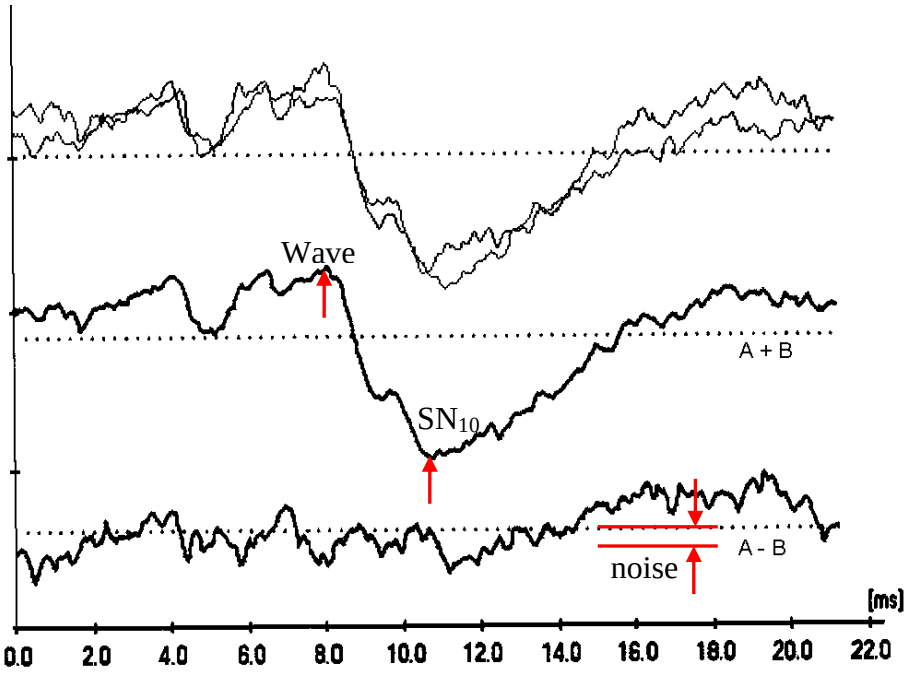


Figure C9: 4kHz waveforms at 40dBnHL showing how the signal amplitude can be measured from a weighted add and the noise level measured from an unweighted subtraction of the replicated waveform.
Scales: 200nV /div. 2 ms / div.

'A+B/A-B' Example 3

The residual noise in the upper waveforms of Figure C10 needs to be less than 25nV to qualify for ***RA*** status otherwise we cannot be certain that the noise is not obscuring a small response. The lower waveform is the unweighted A-B and its average value (and the average gap between the upper waveforms) is again about 1/8th division or 25nV. It therefore just qualifies for ***RA*** status.

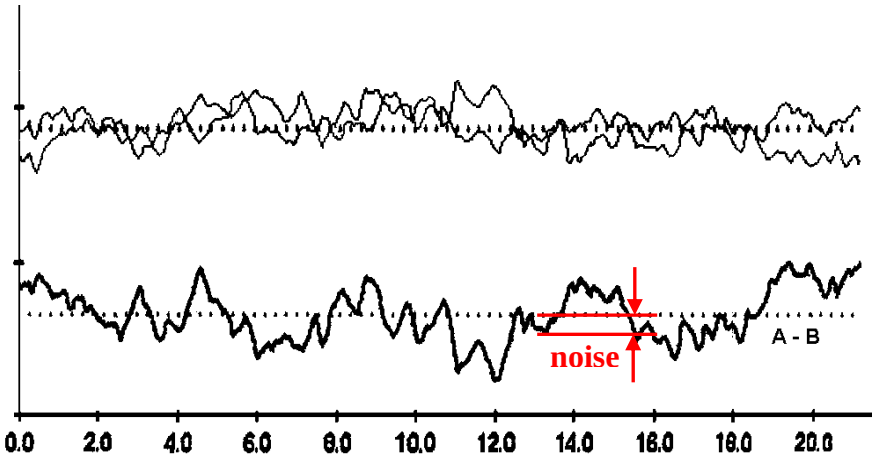


Figure C10: Example of RA calculated from an unweighted subtraction of a replicated waveform. Scales: 200nV / div. 2 ms / div.

C6. The use of rarefaction and condensation sub-averages in lieu of true replication

When alternating polarity stimuli are used, some ABR systems allow the separate display of the rarefaction and condensation waveforms. The question arises: can these separate polarity sub-averages be used as an alternative to conventional replication? This would be valid only if there is no observable difference in response morphology to these separate polarity stimuli. Provisional data from Rhys Meredith using near-threshold very low noise waveforms suggests this is indeed the case and response morphology does not appear to be influenced by stimulus polarity at near-threshold levels. The use of separate polarity sub-averages is therefore considered acceptable if they are superimposed and analysed as suggested for true replications and providing that each sub-average contains the number of sweeps suggested in section 5.6. If differences in response morphology or large stimulus artefacts are encountered then the acquisition and display of replicated alternating polarity waveforms is necessary.

C7. Advice on the use of waveform display options for the “blocking” period

Some ABR equipment offer the facility whereby the latency range over which the artefact rejection system is disabled during stimulus presentation (commonly referred to as the “blocking” period) can be set to either display the recorded waveform or display a flat line. These options do not affect the blocking function; merely the appearance of the displayed waveform. Systems should display the waveform by default. The ‘flat line’ option is appropriate only in cases where a stimulus artefact is so large as to cause waveforms to appear over several pages when printed. Its inappropriate use can hamper correct interpretation in CM testing and when using stimuli below 2kHz.

Appendix D: Artefact rejection level and number of sweeps per average

This appendix aims to provide testers with an insight into the relationship between the artefact rejection (AR) level and the number of sweeps needed in the waveforms to reach a satisfactory signal-to-noise ratio (SNR). This knowledge should guide testers' choice of AR level and sweeps - something that may need to be adjusted during the course of testing in response to changes in test conditions. This advice applies only to babies tested before 12 weeks corrected age.

In the following, reference to "signal" refers to the ABR we want to record whereas "noise" refers to everything else, which can be a combination of EEG, ECG, muscle activity and other electrical interference[p]. A large ABR can be around 0.5µV whereas near threshold it can be 0.1µV or even less (the minimum acceptable amplitude is taken as 0.04µV). The noise can be as little as 2µV in a sleeping baby (EEG) to over 20µV (muscle activity). This means that the signal we want to record is far smaller than the competing noise and in order to correctly identify the presence of the ABR we have to use an averaging technique in order to reduce the noise so that it is a fraction of the ABR.

Averaging improves the SNR by the square root of the number of sweeps. For a 10-fold improvement in SNR we need 100 sweeps; for a 100-fold improvement we need 10,000 sweeps. The number of sweeps needed depends on the relative sizes of the response and the noise. We should try to maximise the signal and minimise the noise. This involves optimal electrode positioning and avoiding sources of noise, both electrical and from the baby.

Artefact rejection simply rejects sweeps (stops them from being added to the average) if the peak-to-peak amplitude of the activity in the sweep is more than a defined level: the artefact rejection level. This ensures that data are collected only when the instantaneous test conditions are favourable.
The NHSP recommended range for AR is ±3 to ±10µV with a default starting level of no more than ±5µV (±3µV for CM tests) so these values should be used when setting up test protocols. Some equipment allows the AR level to be changed in an interactive fashion, whilst averaging; in others it is more cumbersome and time consuming. Noise-weighted (sometimes referred to as Bayesian or Kalman) averaging (see below) should be used if available.

In non-ideal recording conditions the background activity may be above ±5µV most of the time, leading to a high proportion of sweeps being rejected. This is usually caused by the baby's muscle activity. Since bursts of muscle activity are often short lived it is usually best to wait for the baby to settle. However if this does not happen then the rejection level can be raised but since this allows higher amplitude noise into the recording, we must increase the number of sweeps if we are to achieve the necessary SNR in the final recording.

[p] Separate advice is available to help testers identify the sources and minimize the effects of electrical noise.

An example will help. We will assume a satisfactory SNR and a ***CR*** is achieved using 3000 sweeps and a rejection level of ±5µV with a modest number of rejections. If the baby becomes slightly restless and we choose to raise the rejection level to ±7µV, how many sweeps are needed to achieve the same SNR? The answer can be as much as 3000 x (7µV / 5µV)^2 or 5880, *i.e.* roughly double the original 3000. This is the worst case scenario but Lightfoot & Stevens (2013) report that 4500 to 6000 will be needed to achieve the same SNR. Similarly, changing from ±5µV to ±10µV means using 6000 to 12000 sweeps. The cost of increasing the rejection level therefore carries a heavy penalty. Another study (Stevens et al. 2013) also provides evidence that use of ±5µV is more test efficient than ±10µV reject level.

Testers need to know what to do in clinic.
In non-ideal (but not impossible) test conditions which of the following is most efficient?

- Stick to ±5µV (but many sweeps will be rejected, extending test time) or
- Relax the rejection level (allowing more noise into the average and use many more sweeps to deal with this).

The recent study by Lightfoot & Stevens, (2014) tried to identify the best clinical strategy to use. The conclusion was that ±5µV should be used so long as no more than around 30%[q] of sweeps were rejected. If 30% is exceeded, pause to see if the baby will settle. If the baby does not settle but test conditions are not impossible increase to ±7µV but expect to use 4500 - 6000 sweeps. Doing this will reduce the rejection rate. If the rejection rate falls below, say, 10% because the baby has settled then return to ±5µV. If the baby is very restless and substantially more than 30% rejections occur at ±7µV stop the test and wait for the baby to settle. Exceptionally, testing may not be possible without raising the AR level up to ±10µV but it should be realised that it will be very difficult to record an acceptable ABR even with very high numbers of sweeps.

The speed or frequency at which these AR changes are made depends to a large extent on the facility to make AR level changes during averaging and whether the instantaneous reject rate can be easily estimated. Users of equipment that is unable to do this must balance any benefit from changing the AR level with time lost in making an AR change.

Appendix E deals with objective measures that can guide us in our decision of when to stop averaging but if these are not available or not used then the numbers of sweeps given above should be our guide, as well as observation of the averaged waveform.

This strategy is summarised in Figure D1.

[q] A 33% rejection rate is where the number of rejected sweeps is half the number of accepted sweeps.

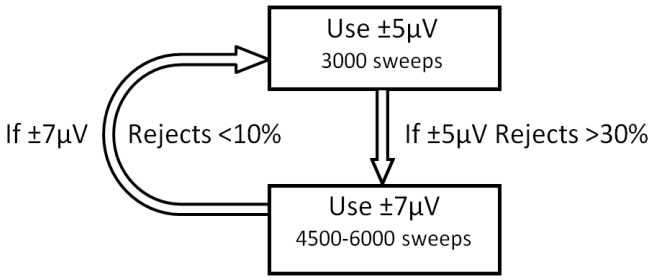


Figure D1: A proposed artefact rejection / sweeps strategy

D1. Noise-weighted averaging

This technique, referred to by various manufacturers as Bayesian or Kalman weighted averaging, is available on some ABR systems. Rather than a fixed artefact rejection level, the idea is to use a more lenient level but allow the software to place more importance on sweeps with a low noise content and less importance on noisier sweeps. This waters down the destructive effects of the noise and can be thought of as "shades of grey" rather than the "black and white" technique of simple artefact rejection. The use of noise-weighted averaging combined with an artefact rejection level of ±10µV performs well and is certainly easier than implementing the above interactive strategy.

When ECG activity is evident in the incoming signal (where the EEG waveform "jumps" about twice a second; a baby's typical heart rate), reducing the AR from ±10µV to ±7µV or even ±5µV will exclude this source of noise and be advantageous. See also appendix F.

However we need to know how many sweeps are needed to achieve an acceptable SNR. The use of an objective response presence confidence measure such as Fmp is appropriate as a guide when a response is present whilst an objective residual noise figure is appropriate as a guide when a response is absent (see appendix E). Both require accurate tester judgement to confirm that all requirements of ***CR*** and ***RA*** are satisfied.

Advice
Use ±5µV artefact rejection level and typically 3000 sweeps as the starting point and use a strategy based on Fig D1, including an increase in the number of sweeps as suggested.
If noise-weighted averaging is available together with objective measures of response confidence and residual noise then use noise-weighted averaging, with an artefact rejection level of no more than ±10µV; less if cardiac activity is recorded. The number of sweeps needed will be informed by the objective measure of response confidence or residual noise.

Appendix E: Objective measures for ABR interpretation in babies

The NHSP clinical group has considered the availability and reliability of objective measures for ABR response detection and the measurement of residual noise on current commercial ABR equipment. This appendix contains some provisional advice on the use of response confidence measures (Fsp in the Natus Biologic Nav Pro EP system; Fmp in the Interacoustics Eclipse system) and residual noise measures. These assist the user to decide whether and when sufficient sweeps have been acquired, rather than using a simple fixed number of sweeps approach.

E1. Use of response confidence measures (Fsp / Fmp) in ABR testing in babies

E1.1 Introduction

Fsp is one of a number of measures available to determine the degree of confidence in the presence of an ABR response (C Elberling and Don 1984). It compares the variance of the averaged waveform to the variance of the background noise level. The variance of the averaged waveform is a measure of the size of the ABR response (if present) plus any residual noise. The higher the Fsp value the greater the ABR response compared to the background noise and the greater the confidence of a clear response. Fmp (Don and Elberling 1994) is a slightly more sophisticated (multiple point rather than single point) version of Fsp.

In the NHSP Guidance for ABR testing in babies, the standard measure used to estimate this confidence is a visual estimate of the signal-to-noise ratio[r] (SNR) which has to reach a value of 3:1 as one of the conditions for a clear response. Note that there are other criteria, such as 'ABR-like waveform and good correlation between replications' that also have to be met before the result can be considered a clear response. The Fsp and Fmp statistics offer an alternative to measuring the SNR by inspection of a pair of replicated waveforms. An obvious use of Fsp and Fmp is to guide the tester when to stop averaging, providing the other criteria for ***CR*** have been satisfied and the minimum number of sweeps have been acquired. This has the potential to save time, especially when the response is large.

Note that Fsp & Fmp can help us decide whether a response is likely to be present; they cannot be used to tell us about response absence. For that, a measure of residual noise is needed, together with other ***RA*** criteria.

E1.2 Validation of Fsp and Fmp

In order to validate the Fsp & Fmp technique and confirm that the measurement agrees with the theoretical value, we need to obtain Fsp and Fmp values from a number of babies when no stimulus is present. Two studies were undertaken. For Fsp John Stevens & Siobhan Brennan (Sheffield) collected 42

[r] Because of the way in which this is estimated, this signal-to-noise ratio is not a true SNR; the 3:1 criterion probably corresponds to a true SNR of around 1.5

no-stimulus tests from a group of 19 babies using the Biologic Nav Pro ER system. For Fmp, Inga Ferm (Croydon) collected 47 no-stimulus tests in a group of 40 babies using the Interacoustics Eclipse system (software version 4.2). Figure E1 shows the cumulative distributions for Fsp and Fmp in these studies. The theory based on an F distribution of F(5,250) suggests that the value of these statistics should be 0.87 at the 50% point. The results of the studies gave a lower than expected Fmp value for the Eclipse system whilst those for the Biologic Nav Pro ER system were higher than expected. The origin of these differences is being investigated. The two measurements agree very well and have the closest match to F(5,250) if Fsp is multiplied by 0.5 and Fmp is multiplied by 1.4 (Figure E2).

The two distributions are very similar and a reasonable match to the F(5,250) distribution when Fsp and Fmp are scaled as suggested above.

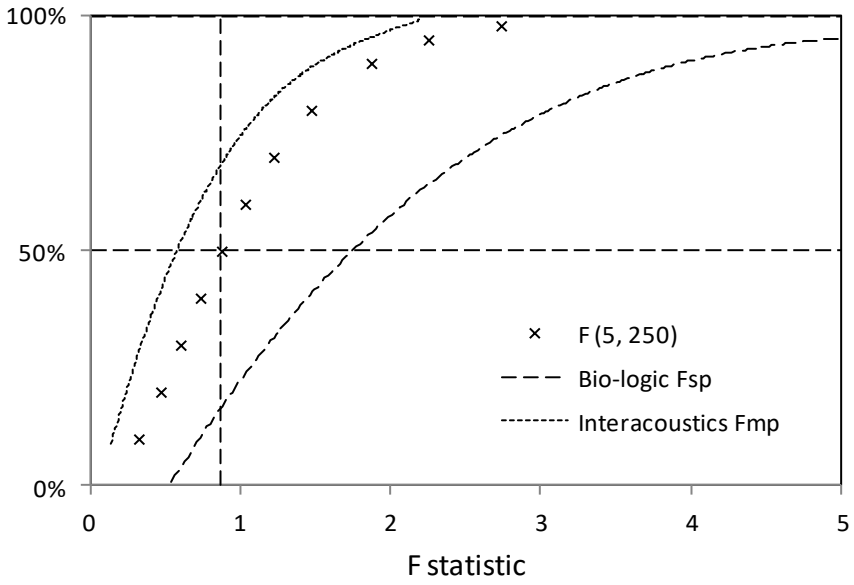


Figure E1: Cumulative distributions for Fsp and Fmp together with the cumulative distribution of F(5, 250). Axes: vertical: proportion of cases; horizontal: F statistic.

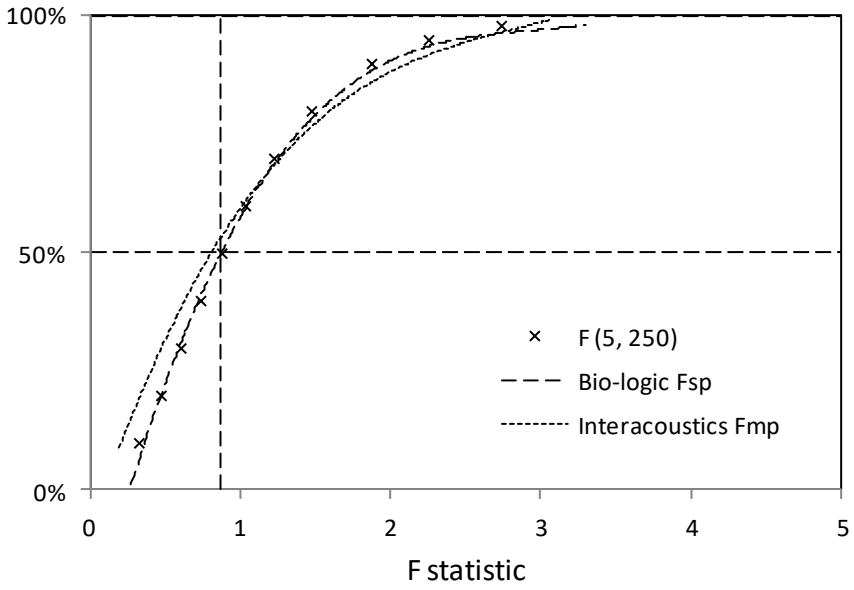


Figure E2: Cumulative distributions after Fsp values are multiplied by 0.5 and Fmp values are multiplied by 1.4.Axes: vertical: proportion of cases; horizontal: F statistic.

Based on an F(5,250) distribution, 2.25 corresponds to a 95% confidence in the presence of a response and values of 2.65 and 3.1 correspond to 97.5% & 99% respectively. Although the rescaled plots of the measured Fsp and Fmp shown in Figure E2 closely match this distribution there are limitations in the measured data that limit the direct use of these values. Two of these are (i) the uncertainty due to the relatively small sample size in the measured values and (ii) the underestimation of the low frequency residual noise in the numerator of the Fsp value caused by the relationship between the low frequency filter and the window length. This is difficult to resolve until more studies are undertaken and these are in hand. As an interim measure it is important to adopt a cautious approach and so the 95% confidence limits for the estimate of the mean of the measured Fsp and Fmp values have been used to estimate the confidence in the scaling factors derived from the two studies (1.4 & 0.5).
The results lead to scaled values for the F(5,250)statistic at the 97.5% confidence level:
Interacoustics Eclipse (v4.2):$Fmp^{97.5\%}$ = 2.37
Biologic Nav Pro:$Fsp^{97.5\%}$ = 6.8
The scaled values related to the 97.5% confidence level have been chosen, with rounding up to the nearest 0.5.

Occasionally a sloping baseline or low-frequency component is recorded; when this occurred in the no-stimulus trials a high (>10) value of Fsp was recorded. This problem is rare for Fmp since the Eclipse applies a 100Hz high-pass filter to the waveforms from which the calculations are made (not to the waveforms seen on screen) as advised by Elberling & Don (1984). The Nav Pro does not do this and the user must remain vigilant for suspiciously high values of Fsp in waveforms that do not appear to contain waveform features suggestive of a genuine response. Figures 12 & 13 in appendix C are good examples of waveforms that could have misleadingly high values of Fsp.

Following an update to the software for the Interacoustics Eclipse system (version 4.4.2.x or later), the guidance relating to Fmp and residual noise for the Eclipse has changed. Direct comparisons between Eclipse 4.2 and 4.4 systems simultaneously recording ABRs has led to the following modified advice (advice for the Biologic NavPro is unchanged):

E1.3 Advice on use of confidence measures

Fmp >2.5 for the Eclipse (software version 4.2), and Fsp & Fmp >7.0 for the Eclipse (software version 4.4.2 or later) and Nav Pro in *both* replicated waveforms using the recommended settings support the conclusion that the response probably exceeds the 3:1 condition component of the *CR* criteria.

Note that it is possible to obtain clear responses (***CR***, SNR >3:1) with Fsp and Fmp less than these values in one or both waveforms.

Important Notes:

- Low values of Fsp or Fmp cannot be used to imply response absence.
- High values of Fsp or Fmp should be accepted only if the waveforms appear to contain classic features of an ABR and not just a low-frequency or sloping baseline.
- If the stimulus artefact is close to the Fsp / Fmp window or there is any other artefact in the waveform the Fsp / Fmp result should not be used.

Check that the recommended protocols for the Biologic Nav Pro and Interacoustics Eclipse are used and that they have the settings shown below for Fsp / Fmp.

E1.4 Current settings for the Fsp & Fmp range:

NB. These values may be already entered in the test protocols available from Biosense Medical and Interacoustics UK for the diagnostic ABR testing of babies referred under the English NHSP programme.

Stimulus	Nav Pro Fsp range	Eclipse Fmp range
Click	6 to 14ms	5 to 15ms
4kHz tone pip	6 to 14ms	5 to 15ms
2kHz tone pip	8 to 16ms	7 to 17ms
1kHz tone pip	11 to 19ms	10 to 20ms
0.5kHz tone pip	13 to 21ms	10 to 20ms
Chirps (all types)	N/A	5 to 15ms

Note:
Enabling the Fsp function on the Biologic Nav Pro results in changes in the way the system acquires and displays data. The display is refreshed every 256 sweeps and upon completion of an average, the number of sweeps used is rounded down to an integer multiple of 256. When testing it is therefore

recommended that tests are stopped just after a multiple of 256 sweeps *i.e.* at 1536, 1792, 2048, 2304, 2560, 2816, 3072, 3328, 3584, 3840 or 4096 sweeps.
The figures for the number of sweeps in the current NHSP guidance for ABR testing in babies become typically 2048 (minimum of 1536) for click ABR and typically 3072 (minimum of 2048) for tone pip ABR. It is also recommended that the maximum numbers of sweeps, set in the protocols, are 4096 for tone pips and 3072 for clicks to allow for extended averaging in non-ideal conditions.

E2. Residual noise

E2.1 Background

Residual noise in ABR testing is a measure of the background electrical activity weighted by the averaging process. A baby when asleep will typically have a level of electrical activity (mainly EEG) of about 2 to 4µV peak to peak or about 0.75 to 1.5µV root-mean-square (RMS). Averaging will reduce this by the square root of the number of sweeps -*e.g.* after 2500 sweeps the value will be reduced by a factor of 50. So for a typical baby the value would be expected to be between 15 and 30nV. This assumes that the artefact rejection level was adjusted such that little or no myogenic or electrical artefacts at all were allowed into the average. Considerably greater noise levels will result if any such activity is allowed into the average.

Both the amplitude of the background electrical activity in babies and the ABR response vary considerably. In babies with a large ABR response the level of residual noise required to achieve a clear response will not be as low as for babies with a small ABR response.

The way in which ABR equipment calculates residual noise is not the same as the NHSP “average gap between superimposed replicates” method. When using the residual noise figure offered by ABR systems it is therefore necessary to estimate the value equivalent to the NHSP 25nV average gap.

E2.2 Comparison of equipment residual noise to NHSP average gap measures

Three studies were undertaken, comparing the residual noise values reported by the Nav Pro and Eclipse in order to establish equipment-specific noise values that correspond to the NHSP “average gap” ***RA*** criterion of 25nV. When combined the studies concluded that these are 18nV for the Nav Pro and for the Eclipse (software version 4.4.2 or later). However there was some variability in the relationships and the above values represent the 50% point (the most likely). If we were to use these values as a surrogate for the NHSP 25nV gap method then noise would be underestimated 50% of the time. A more cautious approach is to use slightly lower values, as given below under ‘Advice’. These values should only be used as target values for when to stop averaging: they are not meant to be true estimates of residual noise. A further study was undertaken to determine the equivalent of the NHSP 25nV ***RA*** noise criterion for the Vivosonic Integrity V500 by feeding the same raw EEG data into the Integrity and Eclipse systems with noise-weighted averaging disabled. The value for the Integrity is also included below under “Advice”.

E2.3 Advice: Use of residual noise measures in determining RA

Residual noise values (if provided by the equipment) may be used as a guide of when to stop averaging if the outcome of the test appears to be a candidate for ***RA*** status. The recommended target values for three types of equipment are as follows.

Biologic Nav Pro & Interacoustics Eclipse (software version 4.4.2 or later):	**15nV**
Vivosonic Integrity V500	**20nV**
Interacoustics Eclipse (software version 4.2):	**25nV**

Even lower values of residual noise will make achievement of the NHSP ***RA*** noise criterion more secure.

NB. Replicated waveforms will still have to meet the 25nV average gap criterion and the other ***RA*** criteria in section 5.9 and notes on baseline drift.

E3. Using objective confidence measurements as an alternative to waveform replication

The primary purpose of waveform replication is to allow the tester to judge whether a response is reliably present or absent, using the criteria developed for ***CR*** and ***RA***. However, the availability of objective response confidence and residual noise values could, theoretically, make replication unnecessary, providing other important criteria are satisfied. What follows is optional advice for those with access to objective measurements of residual noise (for ***RA***) and Fsp or Fmp (for ***CR***) in Biologic Nav Pro or Interacoustics Eclipse systems (other systems' objective measurements have not yet been validated). Fsp or Fmp and residual noise values should be available on the printout for peer review.

E3.1 Clear Response

The definition of a Clear Response in section 5.9, relying on waveform replication, remains the reference and continues to be a requirement for the ***CR*** at threshold or for screening at discharge level (or lower levels being used in lieu of discharge level).

Single (unreplicated) waveforms qualify as ***CR*** at other stimulus levels (including those levels needed to define the gold standard) if the following modified criteria are satisfied:

- The waveforms should show the expected characteristics in terms of amplitude, latency and morphology.
- The size/amplitude of the response (as judged from the wave III/V to the following SN_{10} trough) should be a minimum of 100nV and Fsp/Fmp should exceed the criterion given in section 1.3 of this appendix.
- Waveforms should be compared with those at other stimulus levels (where available) to confirm that they follow the expected changes with stimulus level.

The principle underpinning ***CR*** is that there must be a high degree of confidence that a response is genuinely present. If the only recordable ***CR*** is at the maximum stimulus level, replication is required.

Replication is essential if the waveform appears sloping or has a low-frequency component since such waveforms can have a large of Fsp even without a genuine response.
An example of the above modified criteria being applied to a baby meeting the discharge criteria is given in Figure E3. The Fsp of waveforms at 50dBnHL were well above the criterion for acceptance as ***CR***, the response were >100nV and morphology was appropriate; the waveforms did not slope.

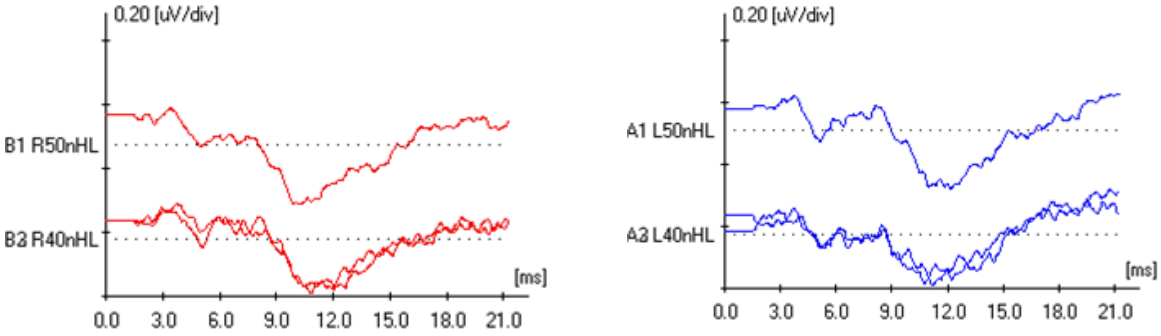


Figure E3: 4kHz Waveforms with supra-aural earphones meeting the criteria for use of an unreplicated waveform at +10dB above the discharge level of 40dBnHL (30dBeHL).

An example of the above modified criteria applied to a raised threshold is shown in Figure E4 and includes an unreplicated ***RA*** (see below).

E3.2 Response Absent

The definition of Response Absent in section 5.9, relying on waveform replication, remains the reference.

Single (unreplicated) waveforms qualify as ***RA*** at any stimulus level if the following modified criteria are satisfied:

- The waveform must be appropriately flat, with no evidence of a response and the reported residual noise value must be no more than that given in section 2.3 of this appendix.

The principle underpinning ***RA*** is that there must be a high degree of confidence that a response is genuinely absent. If there is any doubt (for example if there is a possible vestigial response), then replication is the most appropriate means of resolving the issue.

Replication is required for an ***RA*** at the maximum available stimulus level, with the residual noise meeting the criterion in section 5.9 (average gap method) or section E2.3 (objective residual noise value of both waveforms).

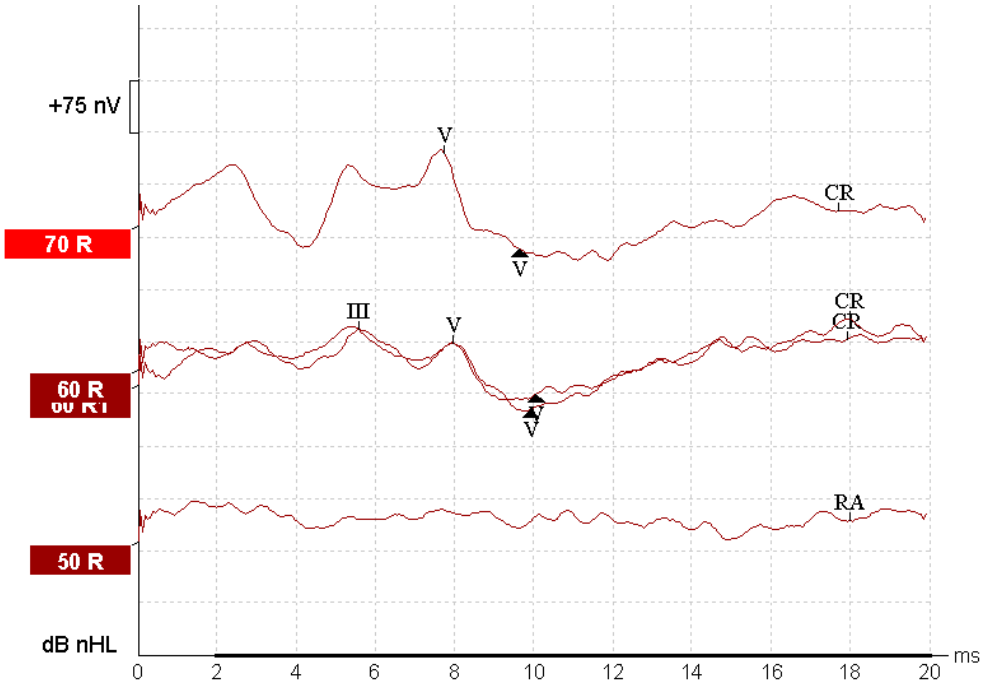


Figure E4: Tone pip 4kHz waveforms. The responses at threshold (60dB) are replicated as normal. The unreplicated 70dB waveform has an Fmp well in excess of the criterion for *CR*, the response was >100nV and the morphology agrees with the responses at 60dB. The unreplicated 50dB waveform shows no evidence of a response, is appropriately flat and the residual noise value is well below the *RA* criterion for this equipment. The threshold can be reported as =60dBnHL, gold standard. For the purposes of peer review, the Fmp value at 70dBnHL and residual noise value at 50dBnHL must be available to the reviewer in order to confirm their *CR* & *RA* status. In this example, at 60dB, wave III was marked and used when assessing the candidacy for CR since it is higher than wave V.

Appendix F: Baseline drift

This appendix expands on section 5.15.

F1. Possible mechanisms

There are a number of possible mechanisms that could result in a sloping ABR baseline and the stimulus artefact has often been sited, principally because a more likely cause had not been identified. However a recent study (Lightfoot, 2017) has suggested that a probable cause in many cases is the cardiac activity of the patient, which is unintentionally recorded by the ABR electrodes. The degree of the resulting slope is influenced by the artefact rejection level, which the tester could change to investigate whether this makes the slope steeper or flatter (it could be either).

F2. Strategies available to the tester to deal with baseline drift

If changing the artefact rejection level does lead to a change in the recorded slope or if performing a blocked stimulus run does not eliminate the slope then this would support the conclusion that the slope is not ABR-related. Simply knowing this can be helpful in waveform interpretation.

The immunity of ABR systems to reject “common mode” signals like the ECG is degraded if the recording electrode impedances are dissimilar, so an obvious first step is to check (and if necessary improve) electrode contact.

If baseline drift makes interpretation problematic then raising the high-pass filter above 30Hz (for example to 75Hz) may help. Some equipment offers digital “display” filters which can be applied to existing waveforms, avoiding the need to obtain further averaging runs. Note however that this will have the effect of reducing the amplitude of any ABR so filtered waveforms should be identified as such and interpreted with caution.

F3. Examples of ‘baseline drift’ and other waveforms with non-flat baselines that represent interpretation difficulties

Figure F1 shows an example where the response is ‘flat’ enough to accept the conclusion of RA. There are ***CR***s for AC 1kHz tone pips at 50dB & 45dB. At 40dB the average gap is less than one-eighth of a division (25nV) but does this qualify as ***RA?*** Are there response features present? The first half of the waveform appears highly correlated to the 50dB and 45dB waveforms but it is doubtful that this is part of the response; it is probably instrument or mains-related. All suggestion of wave V and SN_{10} have disappeared. A blocked-stimulus run could have been used to increase confidence if required. The threshold is therefore =45dBnHL.

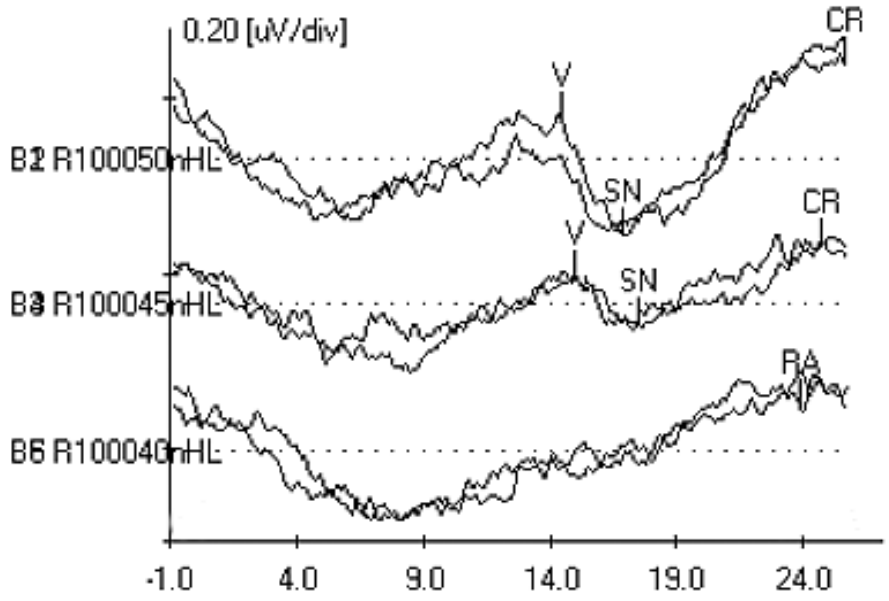


Figure F1: AC 1kHz waveforms at 50dB, 45dB and 40dB.
Scales: 200nV / div. 5 ms / div.

Figure F2 meets the requirements for ***RA*** status but has a non-flat baseline. It has no features characteristic of an ABR and because of this, ***RA*** can be accepted. Nevertheless if such a result is critical (for example if obtained at the maximum available stimulus level, suggesting a very elevated threshold) then confidence in this interpretation should be increased by attempting to remove the cause of the baseline drift and obtaining a further pair of replications. It may be helpful to run a blocked stimulus run to see if the baseline drift remains.

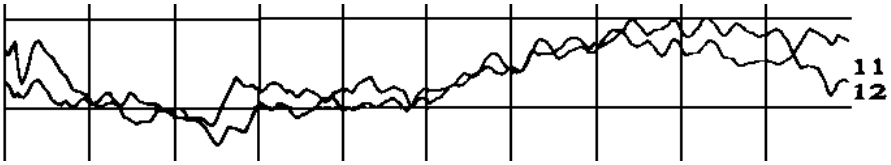


Figure F2: Non-flat baseline that meets the criteria for RA.
Scales: 120nV / div. 2 ms / div.

In Figure F3 the AC 4kHz waveforms marked at 70dB (stimulus levels are in dBnHL) have been positioned so as to superimpose the ABR peaks III & V which is tempting but not ideal in terms of assessing the residual noise. Another interesting feature is that in this example, the size of the response should be judged from wave III to SN_{10}. That being the case it is obvious that at 70dB there is a ***CR***, even though superimposition is not ideal. However our main concern in this example is how to interpret the 60dB waveforms. The noise is sufficiently low for ***RA*** yet there is a feature that some observers may interpret as a tiny wave V (if so, the signal to noise ratio is clearly below the required 3:1). Note that the baseline drift is apparent at both test levels. A previous version of this guidance stated: "*The waveform at 60dB in this example may be taken as **RA** unless the outcome was critical or was at variance with results obtained for other stimuli*." We wish to change this advice as the 60dB waveforms may contain a vestigial wave V so must be taken as Inc. A blocked stimulus run could be obtained and **RA** concluded only if the feature remained.

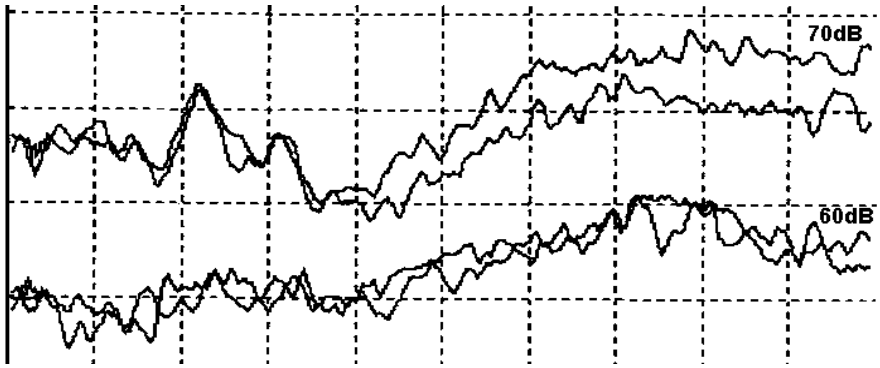


Figure F3: AC 4kHz waveforms at 70dBnHL and 60dBnHL.
Scales: 120nV / div. 2.5 ms / div.

A further example of a possible ***RA*** with 'baseline drift' is illustrated in Figure F4. This again illustrates that it is sometimes difficult to distinguish non-flat baselines from near-threshold responses particularly for low frequency stimuli. The tester must rely on their judgement of what constitutes a "characteristic waveform" and how the features of this waveform (particularly latency) relate to those obtained at higher test levels. If a similar pattern was recorded with a blocked stimulus, ***RA*** would be acceptable.

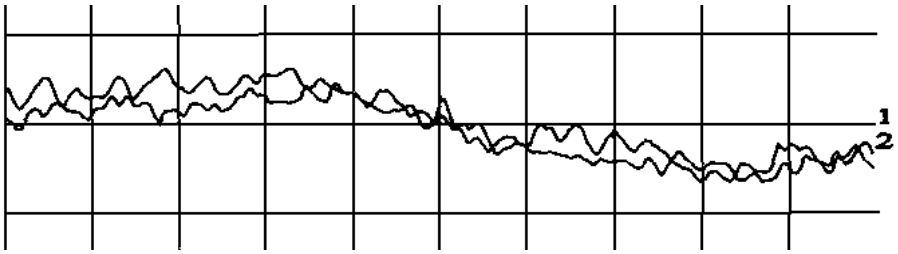


Figure F4: Non-flat baseline.
Scales: 120nV / div. 2 ms / div.

The waveforms of Figure F5 might be incorrectly accepted as clear responses because they have a signal-to-noise ratio well in excess of 3:1 and some features of an ABR waveform of the form often seen with lower frequency stimuli close to threshold. As such the case could have been discharged. However closer inspection reveals no apparent increase in amplitude above threshold. This may be 50Hz noise that just happens to be in a phase that looks very response-like or it may be the ECG artefact. Tests at a higher level and a blocked-stimulus control run (producing flat waveforms) would have helped resolve this case. Without that additional information these waveforms must be regarded as inconclusive.

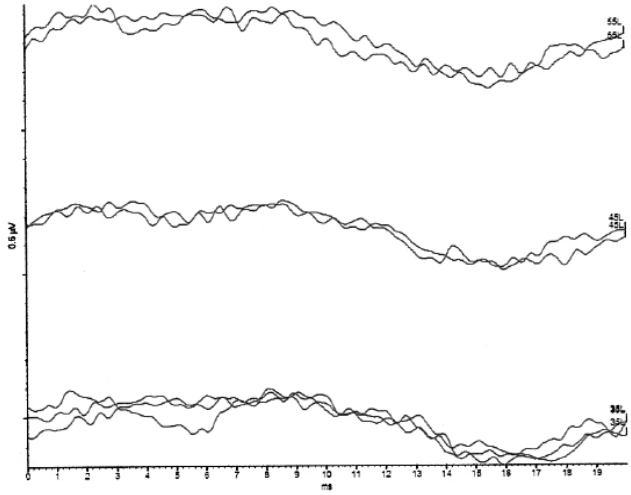


Figure F5: An inconclusive case that needs a blocked-stimulus control run to be performed to distinguish between a true response and possible artefact.
Intensity values: upper waveforms 55dBnHL, middle waveforms 45dBnHL, lower waveforms 35dBnHL.
Scale: 100nV / minor div. 1 ms / div.

BC waveforms are illustrated in Figure F6, with a ***CR*** evident at 50dB (stimulus levels in dBnHL). How do we interpret the waveforms at 40dB & 30dB? Both have residual noise below 25nV which on this scale is an eighth of a division. The "characteristic waveform" requirement is arguably just satisfied at 40dB and is probably not satisfied at 30dB but the degree of confidence in either conclusion is not high. We must try to ensure that any disagreement between independent observers is no more than 10dB. With that in mind it would reasonable to interpret 40dB as ***CR*** and 30dB as ***Inc***, reporting the result as ≤40dB. This example is very challenging but fortunately examples like this are not very common in clinical practice! The attraction of a blocked stimulus run is obvious in this case. For BC, this is performed by holding the vibrator just clear of the normal application site yet with the stimulus being delivered to the vibrator (so that any stimulus artefact is maintained).

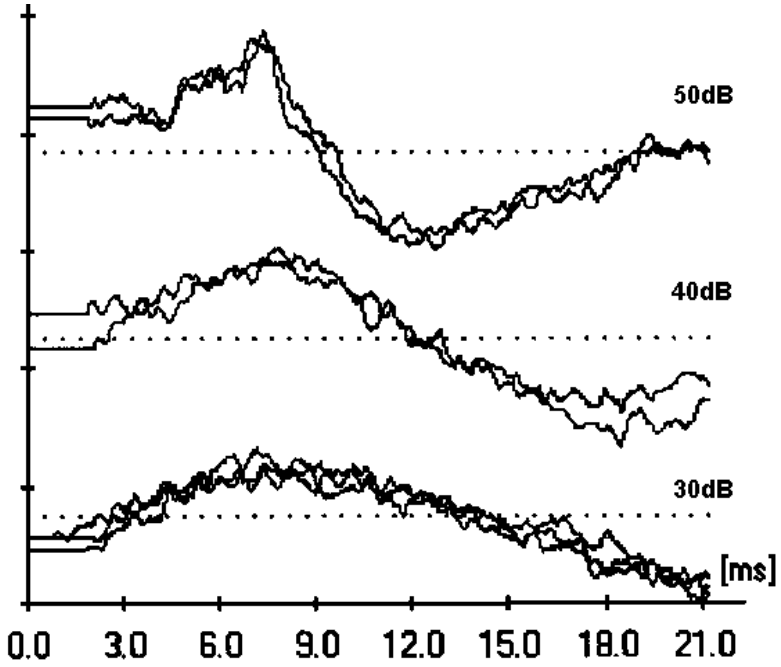


Figure F6: BC waveforms at 50dBnHL, 40dBnHL and 30dBnHL.
Scale: 200nV / div. 3 ms / div.

References

BRITISH SOCIETY OF AUDIOLOGY (NHSP) (2013). Guidelines for the early audiological assessment and management of babies referred from the Newborn Hearing Screening Programme, Version 3.1, July, 2103. [Online]. Available from: https://www.thebsa.org.uk/wp-content/uploads/2014/08/NHSP_NeonateAssess_2014.pdf [Accessed 10/11/2018].

BRITISH SOCIETY OF AUDIOLOGY (2019) Recommended Procedure Assessment and Management of Auditory Neuropathy Spectrum Disorder (ANSD) in Young Infants [Online]. Available from: insert web link. [Accessed date].

Don, M, and C Elberling. 1994. "Evaluating Residual Background Noise in Human Auditory Brain-Stem Responses." *The Journal of the Acoustical Society of America* 96 (5 Pt 1): 2746–57.

Elberling, C, and M Don. 1984. "Quality Estimation of Averaged Auditory Brainstem Responses." *Scandinavian Audiology* 13: 187–97.

Elberling, Claus, and Manuel Don. 2010. "A Direct Approach for the Design of Chirp Stimuli Used for the Recording of Auditory Brainstem Responses." *The Journal of the Acoustical Society of America* 128 (5): 2955–64.

Ferm, I, and G Lightfoot. 2016. "ABR Wave I Presence as an Alternative to Masking: Do NB CE-Chirps Offer an Advantage over Tone Pips?" In *Conference Presentation at: Hearing Across The Lifespan. Lake Como, Italy.*

IEC. 2007. "IEC 60645-3:2007. Electroacoustics - Audiometric Equipment - Part 3: Test Signals of Short Duration." International Electrotechnical Commission. https://webstore.iec.ch/publication/2773.

ISO. 2007. "Reference Zero for the Calibration of Audiometric Equipment Part 6: Reference Threshold of Hearing for Test Signals of Short Duration. ISO 389-6." *International Organisation for Standardisation*. International organisation for standardisation.

Lightfoot, G, and J Stevens. 2014. "Effects of Artefact Rejection and Bayesian Weighted Averaging on the Efficiency of Recording the Newborn ABR." *Ear and Hearing* 35 (2): 213–20.

Lightfoot, Guy. 2017. "Sloping ABR Baselines and the ECG Myogenic Artefact." *International Journal of Audiology* 56 (8): 612–16. doi:10.1080/14992027.2017.1313463.

Lightfoot, Guy, Inga Ferm, Amanda Hall, and Kathryn Evans. 2014. "The Effect of Notch Filtering on the Waveform of the Newborn Auditory Brainstem Response." *International Journal of Audiology* 53 (9): 629–32.

Lightfoot, Guy, Yvonne Sininger, Robert Burkard, and Andre Lodwig. 2007. "Stimulus Repetition Rate and the Reference Levels for Clicks and Short Tone B ..." *Am J Audiol* 16: 94–95.

NHSP Clinical Group. 2008. "Check List for Daily and Monthly Function Check of Auditory Brainstem Response Systems (Stage A Check)." http://abrpeerreview.co.uk/onewebmedia/ABR_stage_A.pdf.

———. 2012. "NHSP Recommended Stimulus Reference Levels for ABR Systems." [Online]. Available

from: http://abrpeerreview.co.uk/onewebmedia/14 NHSP Calibration Data 2012.pdf [Accessed 10/11/18].

Small, Susan a, Jennifer L Hatton, and David R Stapells. 2007. “Effects of Bone Oscillator Coupling Method, Placement Location, and Occlusion on Bone-Conduction Auditory Steady-State Responses in Infants.” *Ear and Hearing* 28 (1): 83–98. doi:10.1097/01.aud.0000249787.97957.5b.

Stevens, John, Siobhan Brennan, Denise Gratton, and Michael Campbell. 2013. “ABR in Newborns: Effects of Electrode Configuration, Stimulus Rate, and EEG Rejection Levels on Test Efficiency.” *International Journal of Audiology* 52 (10): 706–12. doi:10.3109/14992027.2013.809482.

Webb, HD. 1993. “Auditory Screening in High Risk Neonates: An Evaluation of the Bone Conduction ABR.” University of Sheffield, UK.