Recommended Procedure

Cortical Auditory Evoked Potential (CAEP) Testing

Date: Nov 2022
Review date: Nov 2027

General foreword

This document is a Recommended Procedure by the British Society of Audiology (BSA). This Recommended Procedure 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, with reviews by national and international experts, 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document supersedes any previous statement on cortical auditory evoked potential assessment by the BSA and stands until superseded or withdrawn by the BSA.

This document will be reviewed by the date given on the front cover. However, should any individual or organisation feel that the content requires immediate update, review or revision, they should contact the BSA using the details below. Please add 'BSA document revision request' in the title. You will be asked to complete a short form with your reasons and this will be passed to the Professional Guidance Group for assessment. Comments should be sent to:

British Society of Audiology
Blackburn House,
Redhouse Road
Seafield,
Bathgate
EH47 7AQ
bsa@thebsa.org.uk
www.thebsa.org

Published by the British Society of Audiology

Authors and acknowledgments

Produced by:
The BSA Electrophysiology Special Interest Group and the Professional Guidance Group

Key Authors:

Steven Bell (Editor)	University of Southampton
Kevin Munro (Appendix 1)	University of Manchester
Siobhan Brennan	University of Manchester
Constantina Georga	Royal Berkshire Hospital
Anisa Visram	University of Manchester

With Thanks To:

Guy Lightfoot	ERA Training Consultancy
Inga Ferm	Croydon Health Services NHS Trust
Paul Bacon	Sheffield Teaching Hospitals
Vivien Thorpe	NHS Greater Glasgow & Hyde
Ben Mann	Click Hearing, Essex
Keiran Josef	Guys and St Thomas NHS Foundation Trust

The contributions of all the above and special thanks to the suggestions of international reviewers, Dr. Michael Vidler, and Mrs Rosie Mayer who provided the reference on the work of Dr Harry Beagley are all gratefully acknowledged.

Contents

1. Introduction

This document is not intended to provide guidance on specific circumstances or on interpretation of results. It is important that the competent person carrying out, or responsible for, the test (the 'tester') uses professional judgement when deciding on the particular approach to be used with each person being tested (the 'subject'), given the specific circumstances and the purposes of the test, and the tester's level of competency.

The term 'shall' is used in this document to refer to essential practice, and 'should' to refer to desirable practice. Unless stated otherwise, this document represents the consensus of expert opinion and evidence as interpreted by the Professional Guidance Group of the BSA in consultation with its stakeholders. The document was developed in accordance with the BSA Procedures for Processing Documents (BSA).

1.1. Historical Setting

Hallowell Davis identified the auditory cortical evoked response in 1939 (Davis, Davis, Loomis, Harvey & Hobart, 1939). The response is an example of an obligatory exogenous "event-related potential" (ERP) and is most easily identified using an averaging technique, in which a stimulus is presented repeatedly and the post-stimulus EEG is averaged. The computers necessary for averaging were originally analogue but the availability of digital computers facilitated the ease and precision of ERP measurement. The term "Evoked Response Audiometry" (ERA) covers all ERPs to auditory stimuli. Historically many different terms have been used to describe auditory ERPs of cortical origin, such as the slow vertex response (SVR), auditory late response (ALR) or the auditory cortical potential (ACP). However current recommended terminology to describe auditory ERPs of cortical origin is cortical auditory evoked potentials (CAEPs). Much of the basic research on CAEPs dates from the mid-1960s to mid-1970s; pioneers from the UK included Harry Beagley and Bill Gibson.

1.2. Characteristics and uses of the CAEP

The mature supra-threshold CAEP comprises a series of peaks and troughs usually termed P1 at typically 50-70 ms, N1 at typically 100-130 ms and P2 at typically 200-250 ms. Response morphology is dependent on age, arousal state, attention, stimulus and presentation parameters. For example in infants, wave latencies can be very delayed and mature over a number of years to reach adult values. Response amplitude is generally smaller without attention to the stimuli. At higher stimulus rates, neural adaptation occurs, reducing wave amplitudes. Filtering can affect CAEP morphology. For more information see Hall (2007). A review of the neural generators, characteristics and maturation of the CAEP is given by Pratt & Lightfoot (2012). More general information on auditory evoked response methods and their clinical applications are given by Hall (2007), Burkard, Don, & Eggermont (2007) and Picton (2011).

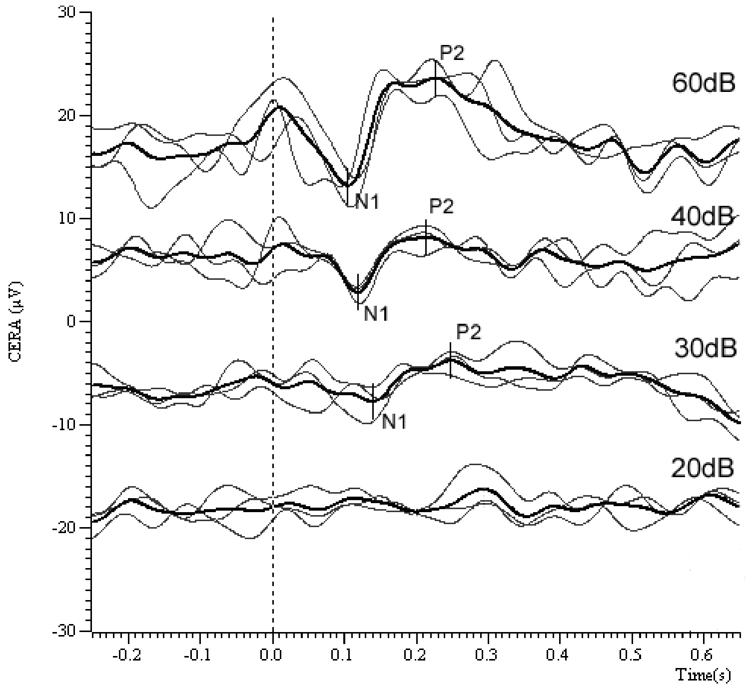
One of the most important and clinically useful aspects of the CAEP is that in adults, the response can be observed close to auditory threshold, and therefore can be used as an objective estimator of the auditory threshold (Hyde et al. 1986; Hyde 1997; Tsui et al. 2002). This has obvious applications in patients that cannot or will not produce reliable responses during pure-tone audiometry (PTA).

As with other auditory evoked responses, the size of the response diminishes and its latency increases as the stimulus level is reduced towards the patient's threshold (Picton et al. 1970). Figure 1 illustrates this for a 4 kHz tone burst stimulus in an adult whose PTA suggested a 4 kHz threshold of 20 dBHL. Responses are observed at stimulus levels of 30 dBHL and above.

The CAEP is triggered by the onset or offset of a transient stimulus or by a perceptible change in an ongoing signal such as a gap in a continuous sound, a change in level or a change in frequency of a continuous tone. The largest responses occur for louder stimuli and for stimuli with a short rise time (Onishi & Davis 1968).

Another important aspect of the CAEP is that it may not be reliably present in drowsy or sleeping patients (Ornitz et al. 1967). As a result, CAEP testing is performed in awake and alert patients.

It is generally accepted that the CAEP does not fully mature until the late teens (Sharma et al. 2002; Wunderlich et al. 2006; Sussman et al. 2008) but that the technique can be attempted in older children, although auditory brainstem response (ABR) or 80 Hz auditory steady-state response (ASSR) methods can be inaccurate because of muscle activity in awake patients. In awake newborns, infants and young children the immature response is usually recordable for stimuli well above threshold and a resurgence of interest in the last decade has revealed that whilst the response is not a reliable predictor of the hearing threshold, it may nevertheless be capable of providing clinically useful information. A**ppendix 1** provides more information of this application.



4kHz RIGHT EAR

HL dB	SWPS (n)	STAT	N1 (ms)	AMP (µV)	CC	RN	S/N	p
60	15	Y	105	10.4	0.84	2.78	3.75	0.000
40	30	Y	119	5.3	0.80	1.94	2.73	0.003
30	30	Y	139	3.9	0.52	1.56	2.52	0.012
20	45	N	-	-	0.10	1.53	1.41	0.286

Figure 1
An example of CAEPs at a number of stimulus levels in dB HL. The bold lines are the grand averages of the three sub-averages at each level. A 900 ms recording window was used, with 250 ms being pre-stimulus onset. Stimulus onset is at the vertical dotted line. In the table, CC=correlation coefficient, RN=residual noise in µV, S/N is signal (response) to noise ratio, p= p-value.

2. Scope

This document primarily describes the use of the CAEP in hearing threshold estimation in adults and older children (typically over 8 years), which may include testing for medico-legal applications.

The use of the response for supra-threshold testing of infants with and without hearing aids is attracting considerable research interest and has been widely adopted in some countries for hearing aid evaluation. **Appendix 1** gives practice guidance on using CAEP with infants (and is now expanded from the previous version of this document). At this point limited availability of clinical test systems means that we are not in a position to recommend a specific protocol for infant CAEP testing.

Other uses of CAEP testing include testing patients with learning difficulties. An example of this is given in **Appendix 3**

3. Equipment & Test Environment

3.1. Equipment

The CAEP may be recorded on most popular auditory evoked potential systems whose most common clinical application is ABR testing. Table 1 offers suitable basic stimulus and recording parameters. However, specialist CAEP systems are available that offer greater ease of use, speed of testing, narrow-band masking noise and objective response scoring. Unfortunately such systems do not additionally perform ABR tests and few centres have both types of system (Carter et al. 2010). All electrophysiological equipment shall satisfy the electrical safety requirements detailed in BS EN 60601-1 for type BF equipment and be tested by a qualified person on an annual basis.

Table 1: Summary of basic stimulus and recording parameters

Parameter	**Value**	**Comment**
Electrode Montage	Cz +ve; Mastoid –ve; Fpz Gnd	Linked mastoids may reduce noise (unverified)
High Pass Filter	1 Hz	
Low Pass Filter	15 Hz	30 Hz if 15 Hz is not available
Timebase/window	500 to 1000 ms	250 ms pre-stim is desirable
Stimulus type	Tone burst	Clicks, pips and speech tokens also work
Stimulus rise & fall time	10 - 20 ms	Linear ramp
Stimulus plateau	30 - 200 ms	Only the first 30 - 50 ms evokes the response
Stimulus modality	Air or Bone conduction	Soundfield also possible
Stimulus calibration	As for audiometers	Only if using tone bursts
Number of sweeps	5 to 30 per sub-average	Depending on response size
Number of sub-averages	2 to 3	Sum to form a grand average
Repetition Rate (adults)	0.5 to 1.0 per second	Randomise if possible
Repetition Rate (older children)	0.25 to 0.5 per second	Randomise if possible
Display aspect ratio	100 ms = 5 µV	

3.2. Test Environment

For hearing threshold estimation, the patient shall be clearly visible to the tester. The patient shall not be able to see or hear the tester adjust the equipment. When the test is controlled from outside the audiometric test room, the patient shall be monitored visually through a window or by a closed-circuit TV-system and acoustically via an intercom to ensure that any movement or patient-generated noise is identified.

Excessive ambient noise will affect the test results, and ambient noise shall not exceed the levels set out in BS EN ISO 8253-1. These are the levels appropriate for routine PTA. The problems caused by ambient noise are greater when testing by bone conduction or loudspeakers as there are no earphones in place to reduce the noise reaching the ears. In general, the ambient noise should not exceed 35 dB(A). A higher level ambient noise may be tolerated if tests down to 0 dBHL are not conducted.

For all evoked measurements the test environment should be electrically quiet, with steps taken to eliminate local sources of electrical interference. These include the use of light dimmers (minimum interference occurs when such lights are switched off or are fully on rather than dimmed), microwave ovens, lift shaft motors, X-ray equipment, mobile phones, staff bleep systems, surgical diathermy and pulse oximeters. The vulnerability of CAEP testing to electrical interference is far less than ABR testing because lower filters are used and the response is larger.

4. Staff Training

Staff undertaking CAEP tests on patients and reporting on their results shall receive specialist training and ensure they practice within the limits of their clinical competence. Some introductory instruction is a component of some British graduate or post-graduate courses in Audiology but this is unlikely to be sufficient and attendance on a specialist course or a programme of structured private study is advised. Recently trained staff should arrange to receive mentorship and support from an experienced colleague and arrange for their practice to be peer reviewed.

If tests are to be conducted on non-clinical medico-legal patients it is wise to obtain clear understanding with the employing health body that they accept responsibility for this activity. An alternative, for example in private practice, is to arrange appropriate professional indemnity and public liability insurance.

5. Preparation for testing

5.1. Preparation of test patients

The tester should adopt an effective communication strategy with the patient throughout. This shall take account of the patient's age, hearing, language skills and any other possible communication difficulties, including any suspected non-organic hearing loss (NOHL). Testing shall be preceded by otoscopic examination and tympanometry (see relevant BSA recommended procedures) and the findings recorded, including the presence of any wax. Occluding wax may be removed prior to testing but if wax is removed the procedure shall be documented and undertaken by someone who is qualified and competent to do so.

If there is a likelihood of ear canals collapsing with supra-aural earphones in position this should be recorded as it may lead to measurement of a false air-bone gap. In some cases the use of insert earphones (e.g. Etymotic ER3 and ER5) will avoid this problem.

The appointment letter should include instructions to avoid exposure to loud noise in the 24 hours prior to the test as this can cause a temporary hearing loss. The patient should be asked if they have complied with this instruction. "Loud" can be determined by having to shout or use a raised voice to communicate at a distance of 1 metre or 3 feet. If the results may have been affected by recent noise exposure then it may be necessary to re-test the patient at a time when they have had no recent exposure to noise.

The identity of the patient shall be checked according to local policy. Additionally testers should be alert to the possibility of identity fraud. If required by the instructing solicitor or clinician in medico-legal tests, the appointment letter should instruct the patient to provide appropriate documentation confirming their identity (e.g. driving licence or passport).

In the case of a patient with a learning disability and/or autism, the presence of hypertactility may cause considerable anxiety about the possibility of electrodes or transducers on the head. Pre-appointment work can be carried out involving the patient's care team introducing the individual to electrodes and, for instance, a pair of headphones in a home environment.

If applicable, inform the patient about intercom facilities.

5.2. Patient Instructions

The detail given to the patient (or their carer/parent/guardian) immediately prior to testing will depend on the test, the clinical background and age of the patient. In all cases the patient should be instructed that the test is automatic and that they are required to sit quietly, with the transducers and electrodes (“sensors” or “measurement pads” may be a less worrying term and it may be best to avoid using the term “electrodes”) in place. If the patient expresses concern, they should be reassured that the electrodes are passive and do not introduce any electricity. The likely duration of the test should be given, together with what the patient should do if they want a break. There are some patient groups for whom the test may require more than one test session to achieve any thresholds. This possibility should be discussed with both the patient and their care team prior to the commencement of the test. Consent for testing shall be obtained from the patient or their advocate. If there is to be no post-test discussion of results (as may be the case for medico-legal tests) then this should be disclosed prior to testing. For cases of suspected NOHL it may be helpful to first outline the tests to be performed (which may include a pre-CAEP PTA), highlighting the objective nature of the cortical test. It is not uncommon for the patient to then provide an accurate PTA or at least a PTA with a smaller non-organic component than previously recorded. After giving the test instructions, remove any hearing aids, headwear or ear-rings that may obstruct the correct placement of the transducers, cause discomfort or affect sound transmission. Wherever possible, hair, scarves etc., should not be allowed to sit between the ear and the transducer. Unlike conventional PTA in which any spectacles should be removed, CAEP testing requires that the patient be alert yet not physically active and in adult testing this is most conveniently achieved by asking the patient to read a magazine or watch a silent video (Lavoie et al. 2008). The patient should wear any spectacles appropriate for this, taking care they do not compromise correct and comfortable placement of transducers.

5.3. Electrodes

The following sterile procedure is recommended for skin preparation. The skin should be gently and carefully abraded using a suitable sterile abrasive electrode paste and a clean gauze or cotton bud. An alternative is the use of a disposable abrasive pad. Disposable electrodes are recommended. Electrodes with integral adhesive of the type often used in ABR testing are usually difficult to attach securely at the Cz site unless the patient is bald. Disposable EEG-type electrodes with electrode conducting paste (e.g. 10-20 EEG paste), secured by tape, are available and may be used with success.

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 should be similar across electrodes and no more than 5 kΩ. However in good recording conditions and in an electrically quiet room higher electrode impedances can be tolerated.

A single channel recording is typical, with electrodes located as follows:

- Positive electrode: vertex (Cz). A high forehead position may reduce the response amplitude and can be tolerated but a mid-forehead position is not appropriate.
- Negative electrode: low mastoid (on either side: the response may be recorded from either mastoid). Sufficient space should be allowed for a bone vibrator to be placed on the mastoid above the electrode without interfering with the electrode.
- Common electrode: other mastoid or mid-forehead.

This configuration should result in N1 being plotted downwards on the display. If this is not the case then the positive and negative electrode connections should be reversed.

An alternative to the 3-electrode montage described above is to adopt a 4-electrode arrangement in which the positive is at the vertex (Cz), the negative is a linked pair of electrodes (using "jump leads") on the mastoids and the common is placed mid-forehead (Lightfoot & Kennedy 2006). This is believed by Lightfoot & Kennedy (but not verified) to be associated with a reduction in patient-generated myogenic noise and a corresponding slight improvement in signal to noise ratio.

'As the CAEP can be recorded from both ipsilateral and contralateral sides, 2 channel recording using 4 electrodes may be used as an alternative electrode configuration. For example, channel 1 measures between the positive vertex electrode to the negative left mastoid electrode, channel 2 measures between the positive vertex electrode and the negative right mastoid electrode and the common electrode to mid forehead. This can offer benefit if myogenic noise and changes in electrode impedance (e.g., on changes of patient body position) adversely affect one channel and the better channel can be selected for analysis. However those using 2-channel recording should specify how channels have been selected when classifying responses as clear or absent. A similar 4-electrode arrangement in which the negative mastoid electrodes are linked using jump leads (Lightfoot & Kennedy 2006) can also be used. This is believed by Lightfoot & Kennedy (but not verified) to be associated with a reduction in patient-generated myogenic noise and a corresponding slight improvement in signal to noise ratio.'

6. Stimuli

6.1. Calibration

Auditory tonal stimuli shall be calibrated to the relevant part or parts of BS EN ISO 389 relating to pure tones, depending on the transducers used. BS EN ISO 389-6 which relates to the very brief stimuli used in ABR testing should not be used for the calibration of tonal stimuli used in CAEP testing.. The calibration should be checked annually by a qualified person. For the purpose of calibration, the tone should be made continuous (or near-continuous for example by selecting a rise/fall time of 1 ms, a plateau time of 997 ms and a repetition rate of 1/s). The system may then be calibrated as a pure-tone audiometer. Additionally, testers should adopt and apply the principles daily subjective "Stage A" checks given in the BSA PTA recommended procedure.

Where non-tonal stimuli such as speech tokens or white noise are used their calibration details shall be provided by the manufacturer or determined by an appropriate panel of normally hearing subjects, tested behaviourally.

6.2. Stimulus type

A linear-ramped tone burst is most commonly used for hearing threshold purposes as this allows an objective estimation of the audiogram. A rise/fall time of 10 ms at frequencies of 1 kHz and above (20 ms at lower frequencies) provides a good compromise between frequency specificity and response size. Some systems specify rise/fall time in time (ms) whilst in others it is specified in cycles. Care is needed to select the appropriate values.

Tone burst plateau time is a parameter over which there is some debate. A longer time (e.g. 200 ms) will minimise the effects of temporal integration but it is likely that it is only the first 30-60 ms of the stimulus that evokes the response (Cody & Klass 1968; Weber 1970; McCandless & Best 1966) and a longer duration is likely to decrease the magnitude of the response to the following stimulus so extending the stimulus beyond this may be pointless. A plateau time of 100 ms carries a theoretical disadvantage: the offset of a tone burst also evokes an offset cortical response and the N1 of this second response will destructively interfere with the P2 of the tone onset response. In practice the offset response is very small, so such destructive interference is probably negligible. A 40 ms plateau may be a reasonable compromise.

Any standard audiometric transducer may be used.

7. Masking

For tone burst stimuli, the normal PTA rules apply to the need for masking and when it should be applied. The plateau masking method is too time consuming to be applied. Instead the level of narrow-band masking noise should be calculated for a given stimulus level as follows:

MdB = StimdB - TTL + 10 + ABGnt where:

MdB is the narrow-band masking noise level (calibrated to normal audiometric masking standards, i.e. is calibrated in terms of effective masking);

StimdB is the stimulus level (calibrated to normal audiometric pure tone standards);

TTL is the minimum transcranial transmission loss (inter-aural attenuation) associated with the transducer (a figure of 40dB is quoted for supra-aural earphones in the BSA PTA recommended procedure but many practitioners consider 45 dB to be a more appropriate figure (Lightfoot et al. 2010)*);*

ABGnt is the air-bone gap in the non-test ear at the test frequency.

In the client groups for whom CAEP testing is most useful, we often do not know ABGnt so an educated guess is required, based on available information. A reasonable compromise would be to use a value of 30dB in cases of a non-peaked tympanogram. Even so, there may be a risk that the level of noise used for masking may, in certain cases, lead to overmasking or undermasking; the thresholds given in the CAEP report should be qualified where uncertainty exists.

The formula above relates to narrow-band masking noise. One common problem with the design of most ERA equipment primarily designed for ABR testing is that manufacturers frequently provide only wide band noise for masking purposes, for which there is no international calibration standard. If narrow band noise is available then this should be used for masking. One practical, albeit inelegant, solution is to use a conventional audiometer and associated earphone to supply the narrow band masking noise.

8. Data Collection

8.1. Stimulus repetition rate

This parameter represents a compromise between response size and speed of testing. The N1-P2 response takes typically 10 s to fully recover and if the objective was to record the largest possible response regardless of test time a repetition rate of one stimulus every 10 s would be appropriate (Appleby 1964; Davis et al. 1966). If stimuli are presented more rapidly than this then a diminished response will be recorded. The majority of this response habituation occurs following the first few stimuli (Walter 1964; Ozesmi et al. 2000).

However the purpose of averaging is to improve the response signal to noise ratio (SNR) and the more sweeps are averaged per minute the better. In adults the optimum improvement in SNR corresponds to a repetition rate of one stimulus every 1 to 2 seconds (a rate of 0.5 to 1.0 per second) (Rapin 1964; Davis & Zerlin 1966). The response recovery time is thought to be somewhat longer for the immature response of older children, requiring a slower repetition rate (0.25 to 0.5 per second). Further research is needed in this area.

Using these repetition rates the response to the first stimulus in an averaging sequence will be untypically large as it will have been preceded by a period of silence; the second will be somewhat smaller and so on.

Habituation of the response is thought to be greatest for predictable stimuli, as used in conventional averaging, with somewhat greater response amplitudes recorded in response to stimuli with less predictability. Varying the stimulus repetition rate (or inter-stimulus interval) may reduce habituation (Rapin 1964; Rothman et al. 1970) as might randomising the ear to which the stimuli are presented (Butler 1972). However standard auditory evoked potential systems do not offer this functionality. Lightfoot & Kennedy (2006) used the CED system to investigate whether a variety of stimulus manipulations increased response size but no effect was significant. However their study did not include long exposure to monotonous stimuli as reference. An alternative explanation to their negative findings could be that the novelty value of unpredictable stimuli is a temporary effect that needs to be applied sparingly.

It has been suggested that the CAEP from individuals with some pathologies, such as Down Syndrome, are not affected by habituation (Schafer & Peeke 1982).

8.2. Number of sweeps

In research, where the latency or amplitude of a response must be accurately defined, a large number of sweeps (100 or more) are often used to obtain an averaged waveform with a high SNR. In clinical hearing threshold estimation the objective is to decide whether, for a particular stimulus level, a response is present or absent and for reasons of clinical efficiency this is done in the minimum time.

For large responses (presumed to be supra-threshold) only 5 to 30 sweeps may be needed per sub-average to correctly identify a response; the smaller responses close to threshold will require more sweeps in order to achieve the SNR that is needed to identify a response with an acceptable degree of confidence (see section 9.3). Still greater numbers of sweeps are usually required to reduce the residual noise of an average waveform before it can be concluded that no response is present (section 9.4).

The number of sweeps used at a given stimulus level therefore depends on the size of any response seen and the residual noise in the waveform. The issue of large responses being seen at the start of an averaging sequence (see section 9.1) is a potential trap for testers who, on seeing an apparent response after just a few sweeps, terminate the average. To guard against this error it is vital that the number of sweeps per grand average is never less than 10, that responses are always replicated and that the methods for analysis (see section 10) are followed.

8.3. Replication

In systems that do not have objective scoring, the subjective visual analysis of waveforms requires that there are at least two sub-averages, to allow the tester to judge whether a potential response is sufficiently repeatable to be accepted as genuine. Sub-averages should be superimposed and optionally displayed with their grand average. In the absence of objective scoring facilities, an unreplicated waveform should never contribute to the definition of threshold but is acceptable in the initial phase of testing (see Section 9.8).

8.4. Artefact rejection level

This is the voltage limits above which an epoch (a single sweep) of data is rejected since it is likely to contain considerable non-response activity, often associated with muscle activity. A value of around ±50 µV is recommended. Few epochs should be rejected; it is particularly advantageous to capture (not reject) the epochs associated with the first few stimuli in an averaging run as their signal to noise ratio will be particularly favourable. In research it is usual to employ electrodes positioned to detect eye blinks for the purpose of artefact rejection since blinks are one common source of muscle interference. This is generally not needed in the clinical setting.

8.5. Filters

The spectrum of the near-threshold CAEP is greatest in the 2-10 Hz range which shifts towards the lower frequencies as threshold is approached and in order to optimise the SNR, the incoming electrical activity is filtered prior to digitisation and averaging. A low (high-pass) filter setting of 1 Hz and, where available, a high (low-pass) filter setting of 15 Hz should be used. In some systems the lowest available low-pass filter setting is 30 Hz.
A narrower bandwidth of 5 Hz to 9 Hz has been suggested (Bacon et al. 1990) which was derived from an analysis of CAEP responses at 40 dB and 60 dB above the subjective threshold but it is the responses close to threshold that are of greatest clinical interest.

In research a low-pass setting of 100 Hz is common, which provides more accurate determination of response latency but at the expense of more noisy waveforms.

8.6. Timebase/window length

This is the period over which the incoming electrical activity is recorded and averaged. For the purposes of response detection any genuine response must be clearly distinguishable from ongoing spontaneous noise so it is important for the window to include regions where no response is likely as well as the region where a response is expected. A minimum of 500 ms is needed, which starts at stimulus onset. However the inclusion of an additional 200-300 ms pre-stimulus baseline can be helpful in judging the background activity, thus aiding the process of identifying a response as a feature which is distinct from the noise. A time base of 800 ms including 250 ms pre-stimulus baseline would be ideal but not all systems allow this.

8.7. Display

The vertical (voltage) display scale shall be fixed (not automatically adjusted by the software as a result of the waveform size) and such that small responses can be seen yet several test levels can be displayed on the same chart, in order of descending stimulus level. A display aspect ratio of typically 100 ms = 5 µV is suitable. An automatic display scale shall not be used.

All sub-averages should be displayed, superimposed, unless there is a good technical reason for not doing so in which case a comment should be made in the clinical notes for the purpose of future reference. Waveforms should not be discarded simply because they do not show good correlation with other sub-averages. If a “response” in one waveform is not repeated in other waveforms and the patient’s state of arousal has not significantly altered it is likely to be noise masquerading as a response.

Grand averages may be displayed superimposed with their constituent sub-averages or displayed slightly above or below their sub-averages.

If the timebase/window includes any pre-stimulus baseline the time of stimulus onset shall be indicated.

8.8. Choice of stimulus levels

The starting level will depend to some degree on what is already reliably known for the patient but this is often very little. A stimulus at a moderate level (e.g. 60 dBHL) is usually chosen. If this initial level reveals an obvious (albeit unreplicated) response reduce the level in 20 dB steps, then use 10 dB steps to define the threshold. If the initial level fails to reveal a likely response (replication at this stage is often not a good use of time) then increase by 20 dB. If no likely response is seen continue to increase in 10 dB steps.

The use of a 5 dB step size when near threshold is at the discretion of the tester and the clinical or medico-legal requirements of the case.

Replication and the formal application of criteria for response presence and response absence (see below) are necessary only for those levels that define the threshold. However if the status of an unreplicated waveform is difficult to judge, replication is appropriate.

8.9. Maintaining patient arousal

Drowsiness and both natural and sedation-induced sleep can make the response unpredictable, so the patient's state should be monitored by the tester and action taken if appropriate. Patient eye closure should be avoided as this is associated with EEG alpha activity which may be mistaken for a response.

Protracted test sessions should be avoided as poorer responses have been noted when the test session extends beyond 30 minutes (Davis & Zerlin, 1966).

The tester should have a range of options available to them to maintain patient arousal, for example reading or, for patients who cannot read, watching a silent video.

9. Data Analysis/Interpretation

9.1. Who should interpret & report?

There are many patient and technical issues capable of influencing the quality of recorded CAEPs. Examples include patient drowsiness, eye closure, patient-generated noise or muscle activity, electrical interference and difficulties with electrode security. It is good practice for the tester to document these for future reference and, whilst it is usually possible for someone not present during testing to accurately interpret and report on a case, the tester is undoubtedly best placed to interpret and report the results of a CAEP session. The decisions required for appropriate test strategy include skilled response interpretation during the recording process so anyone sufficiently skilled to perform the test has the skills necessary for interpretation and reporting. It is nevertheless valuable for the tester to seek a second opinion and peer review, particularly in challenging cases. Where possible, peer review is recommended. The report shall not be prepared by a person that does not possess the training, skills and experience needed to perform CAEP testing.

9.2. Objective measurements

Some systems specifically designed to conduct CAEP tests offer objective response scoring and estimation of the residual noise in the waveforms. The Frye Electronics HEARLab system, developed at the Australian National Acoustics Laboratories (NAL) uses statistical analysis based on Hotelling's T^2 statistic, resulting in a p-value (VanDun et al. 2012). The Cambridge Electronic Design CERA system, developed in Liverpool UK, uses a combination of SNR and sub-average correlation to give a p-value. Both systems also estimate residual noise. A comparison of objective detection methods for the auditory brainstem response (Chesnaye et al., 2018) showed Hotelling's T^2 to be most sensitive, followed closely by the Q-samples test. Data from Chesnaye et al. in review using adult CAEP data suggests Q-samples may be best for small number of epochs and Hotelling's T^2 for larger samples. Several groups are currently exploring machine learning for evoked response detection. These objective measurements have an obvious attraction when testing medico-legal patients but in all cases, objective measurements should be used as an adjunct to the tester's interpretation described below. **Appendix 2** gives an example of medico-legal CAEP testing. The tester should override the statistical information if they feel it is likely that it is the result of an artifact.

The primary utility of these objective measurements is that they guide the tester's decision of when sufficient sweeps have been acquired to be able to categorise a stimulus level as showing a response or indicating response absence.

9.3. Criteria for response presence

For a CAEP to be categorised as present the following criteria should be met:

a) The response shall have an appropriate waveform morphology, amplitude and latency
b) The response shall be repeatable, as judged by similarity between replicates
c) The response morphology, amplitude and latency shall follow the expected trend of smaller amplitudes and longer latencies compared to responses obtained for a higher level stimulus, when available

d) The response shall have a sufficiently high SNR to satisfy the tester with a high degree of confidence that the "response" is genuine

If the equipment does not offer objective measurements[1] the tester shall estimate the SNR visually. The following method is recommended: The "signal" is the size of the response, measured as the N1-P2 amplitude. In all but marginal cases it is usually sufficient to do this by eye, with reference to the vertical scale rather than using cursor measurements. The noise may be estimated from the average gap between a pair of optimally superimposed replicates, assessed across the entire timebase. This mirrors the method advocated by the BSA Recommended Procedure for ABR testing in babies, where examples can be seen. Using this method, an SNR of 2.5 or more is usually associated with a p-value of 0.05 or less (for example see Figure 1); sufficient to conclude that the response is present, provided the other criteria have been satisfied.

There is no minimum response size requirement for response presence but in practice, responses smaller than 2.5 µV can rarely be distinguished from residual noise with confidence, even in good test conditions.

9.4 Criteria for response absence

For a CAEP to be categorised as absent the following criteria should be met:
a) There should be no likely response present; this may be reinforced by objective measurements of SNR (or p-value) failing to suggest a response. However, a possible response with an SNR (or p-value) less than that needed for response presence is not sufficient to qualify for response absence.
b) The residual noise in the grand average waveform shall be sufficiently low to be confident that a small response is not obscured by noise.

It is not sufficient to say "I can't see a response", the tester must have a high degree of confidence that a response is genuinely absent.

If the equipment does not offer an objective measurement of residual noise the tester shall estimate the residual noise visually, as above. The noise may be estimated from the average gap between a pair of optimally superimposed replicates, across the entire timebase. Using this method, residual noise of about 2 µV is usually low enough to provide the required degree of confidence that a response is genuinely absent. Systems reporting residual noise objectively may do so using a variety of algorithms; the reported value may differ across

[1] Fsp or related measurements designed for use in ABR detection are unsuited to cortical tests.

systems. The tester shall define what response absence noise criterion to use for their system.

Waveforms that do not meet the above criteria for response presence or response absence must be regarded as inconclusive and take no part in the definition of the CAEP threshold. Such responses should be labelled in a report as 'inconclusive'. Resolving inconclusive responses normally require further averaging but occasionally, small or odd-looking responses remain inconclusive even after further averaging.

9.5 Definition of the CAEP threshold

The CAEP threshold is defined as the lowest level at which a response is present, with a response absent at a level of 10 dB or less below this level and a response present at 10 dB or less above this level.

It is sometimes sufficient to obtain responses down to a certain level without the need to obtain a formal threshold, for example if responses are recorded down to 20 dBHL. Such results should be described using the format ≤20 dBHL. Conversely if no response was recorded at any stimulus level up to, say, 100 dBHL, where response absence was demonstrated, the results should be described using the format >100 dBHL.

If a step size of 10 dB has been used the tester may, if desired, adopt an "interpolation" approach (Lightfoot & Kennedy 2006) in which the threshold may be reported as 5 dB below the lowest level at which a response is seen providing that response is larger than a specified amplitude (Lightfoot & Kennedy 2006 employed 5µV at and below 1 kHz and 3µV at higher frequencies); when smaller than the specified amplitude the threshold is taken as the level of the lowest response. When this interpolation technique is employed it shall be stated in the report, together with reference to the basis of the interpolation. Alternatively the tester may choose to not use an interpolation approach and simply state that a 10 dB step size was applied.

Note that replication is necessary to satisfy the criteria for response presence and response absence but replication may not always be needed for stimulus levels well above or well below threshold. The clarity of responses will guide the tester's decision to replicate at such levels; where there is uncertainty, replication is helpful even for levels that do not contribute to the definition of threshold.

10. Accuracy & Limitations

There is an average difference between CAEP thresholds and the PTA thresholds in cooperative patients; this is known as a "bias" and is typically 5-10 dB (e.g. a figure of 6.5 dB has been reported (Lightfoot & Kennedy 2006)) with the CAEP threshold suggesting a slightly greater hearing loss than the PTA). The value of this bias will depend to some extent on the methods used for stimulus calibration, response acquisition and analysis and the presence of certain co-morbidities. It is technically valid to subtract the bias when predicting the PTA threshold but this subtraction shall be stated in the report and details given on how the bias was derived.

After subtracting any bias, there will be a spread of values in the CAEP – PTA difference in cooperative patients. After accounting for their 6.5 dB bias, Lightfoot & Kennedy (2006) found that 94% of threshold estimate differences were ≤±15 dB. Such information may allow a confidence range to be associated with CAEP results. However, the upper limit of the confidence range will not exceed the level of the CAEP threshold. Example: the above bias is rounded to 5 dB and a CAEP threshold of 50 dBHL is obtained. Subtracting the bias gives 45 dBHL as the best estimate of the PTA and there is 95% confidence that the PTA lies in the range 30 – 50 dBHL (45 - 15 and whichever is the lower of 45+15 or 50, in this case 50).

11. Reporting

A CAEP report should include the following information:

- Hospital name and department
- To whom the report is addressed (e.g. doctor, care home staff, solicitor) including their reference if any
- Test date
- The patient's name, date of birth and reference number
- Tester name
- CAEP threshold results in dBHL
- Confidence Intervals typical for the frequency/patient age (see above)
- Details of the test equipment used, including model & serial number
- The transducer types used
- The date of audiometric calibration of the equipment
- Notes on any issues that might have a bearing on the accuracy of the results

If an interpolation method has been used to arrive at the reported results this shall be stated. If any PTA-ERA bias has been subtracted from the CAEP thresholds to give a prediction of the PTA, this shall be stated including the value subtracted and its origin.

References

Appleby, S., 1964. The slow vertex maximal sound evoked response in infants. *Acta Otol (suppl)*, 206, pp.146–152.

Bacon, P. et al., 1990. Optimal filtering of the auditory cortical evoked potential. *Clinical Physics and Physiological Measurement*, 11(2), pp.135–42.

BS EN, 60601-1-11:2015 Medical electrical equipment – Part 1-11: General requirements for basic safety and essential performance.

BS EN ISO, 389-6:- Acoustics. Reference zero for the calibration of audiometric equipment. Reference threshold of hearing for test signals of short duration.

BS EN ISO, 8253-1:- Acoustics -- Audiometric test methods -- Part 1: Pure-tone air and bone conduction audiometry.

BSA, 2014. Recommended Procedure: Visual Reinforcement Audiometry.

Burkard, R., Don, M., Eggermont, J., 2007. *Auditory Evoked Potentials: Basic Principles and Clinical Application*, Lippincott Williams & Wilkins.

Butler, R.A., 1972. The influence of spatial separation of sound sources on the auditory evoked response. *Neuropsychologia*, 10(2), pp.219–25.

Carter, L. et al., 2010. The detection of infant cortical auditory evoked potentials (CAEPs) using statistical and visual detection techniques. *Journal of the American Academy of Audiology*, 21(5), pp.347–56.

Chang, H. W., Dillon, H., Carter, L., van Dun, B., & Young, S. T. 2012. The relationship between cortical auditory evoked potential (CAEP) detection and estimated audibility in infants with sensorineural hearing loss. *International Journal of Audiology*, 51(9), pp.663–70.

Chesnaye, M. A., Bell, S., Harte, J. M., & Simpson, D., 2018. Objective measures for detecting the auditory brainstem response: comparisons of specificity, sensitivity and detection time. *International Journal of Audiology*, *57*(6), 468-478.

Cody, D.T. & Klass, D.W., 1968. Cortical audiometry. Potential pitfalls in testing. *Archives of Otolaryngology (Chicago, Ill. : 1960)*, 88(4), pp.396–406.

Davis, H. et al., 1966. The slow response of the human cortex to auditory stimuli: recovery process. *Electroencephalography and Clinical Neurophysiology*, 21(2), pp.105–13.

Davis, H. & Zerlin, S., 1966. Acoustic relations of the human vertex potential. *The Journal of the Acoustical Society of America*, 39(1), pp.109–16.

Davis, H.; Davis P.A.; Loomis, A.L.; Harvey, E.N.; Hobart, G., 1939. Electrical reactions of the human brain to auditory stimulation during sleep. *Journal of Neurolophysiology*, 2, pp.500–514.

Gardner-Berry, K., Chang, H., Ching, T. Y. C., & Hou, S. (2016). Detection Rates of Cortical Auditory Evoked Potentials at Different Sensation Levels in Infants with Sensory/Neural Hearing Loss and Auditory Neuropathy Spectrum Disorder. *Seminars in Hearing*, *37*(1), 53–61.

Hall, J.W., 2007. *New Handbook of Auditory Evoked Responses* 1st ed., Pearson.

Hyde, M. et al., 1986. Auditory evoked potentials in audiometric assessment of compensation and medicolegal patients. *The Annals of Otology, Rhinology, and Laryngologyaryngology*, 95(5 Pt 1), pp.514–9.

Hyde, M., 1997. The N1 response and its applications. *Audiology & Neuro-otology*, 2(5), pp.281–307.

Lavoie, B.A., Hine, J.E. & Thornton, R.D., 2008. The choice of distracting task can affect the quality of auditory evoked potentials recorded for clinical assessment. *International Journal of Audiology*, 47(7), pp.439–44.

Lightfoot, G., Cairns, A. & Stevens, J., 2010. Noise levels required to mask stimuli used in auditory brainstem response testing. *International Journal of Audiology*, 49(10), pp.794–8.

Lightfoot, G. & Kennedy, V., 2006. Cortical electric response audiometry hearing threshold estimation: accuracy, speed, and the effects of stimulus presentation features. *Ear and Hearing*, 27(5), pp.443–56.

McCandless, G.A. & Best, L., 1966. Summed evoked responses using pure-tone stimuli. *Journal of Speech and Hearing Research*, 9(2), pp.266–72.

Mehta K, Mahon M, Van Dun B, Marriage J, Vickers D., 2020. Clinicians' views of using cortical auditory evoked potentials (CAEP) in the permanent childhood hearing impairment patient pathway. *Int J Audiol.* 59 (2) 81-89.

Munro, K. J., Purdy, S. C., Uus, K., Visram, A., Ward, R., Bruce, I. A., Marsden, A., Stone, M. A., & Van Dun, B. (2019). Recording Obligatory Cortical Auditory Evoked Potentials in Infants. *Ear and Hearing*.

Onishi, S. & Davis, H., 1968. Effects of duration and rise time of tone bursts on evoked V potentials. *The Journal of the Acoustical Society of America*, 44(2), pp.582–91.

Ornitz, E.M. et al., 1967. The variability of the auditory averaged evoked response during sleep and dreaming in children and adults. *Electroencephalography and Clinical Neurophysiology*, 22(6), pp.514–24.

Ozesmi, C. et al., 2000. Habituation of the auditory evoked potential in a short interstimulus interval paradigm. *The International Journal of Neuroscience*, 105(1-4), pp.87–95.

Pearce, W., Golding, M., & Dillon, H. (2007). Cortical auditory evoked potentials in the assessment of auditory neuropathy: two case studies. *J Am Acad Audiol*, *18*, 380–390. http://www.ncbi.nlm.nih.gov/pubmed/17715648

Picton, T.W., 2011. *Human auditory evoked potentials*, Plural Publishing Inc.

Picton, T.W., Goodman, W.S. & Bryce, D.P., 1970. Amplitude of evoked responses to tones of high intensity. *Acta oto-laryngologica*, 70(2), pp.77–82.

Pratt, H. & Lightfoot, G., 2012. Physiological mechanisms underlying MLRs and cortical EPs. In K. L. Tremblay & R. F. Burkard, eds. *Translational perspectives in auditory neuroscience; Physiological assessment of audition*. San Diego.

Punch S, Van Dun B, King A, Carter L, Pearce W., 2016. Clinical Experience of Using Cortical Auditory Evoked Potentials in the Treatment of Infant Hearing Loss in Australia. *Semin Hear*. 37(1):36-52.

Rapin, I., 1964. Practical considerations in using the evoked potential technique in audiometry. *Acta Otol (suppl)*, 206, pp.117–122.

Rothman, H.H., Davis, H. & Hay, I.S., 1970. Slow evoked cortical potentials and temporal features of stimulation. *Electroencephalography and Clinical Neurophysiology*, 29(3), pp.225–32.

Schafer, E.W. & Peeke, H. V, 1982. Down syndrome individuals fail to habituate cortical evoked potentials. *American Journal of Mental Deficiency*, 87(3), pp.332–7.

Sharma, A., Dorman, M.F. & Kral, A., 2005. The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hearing Research*, 203(1-2), pp.134–43.

Sharma, A., Dorman, M.F. & Spahr, A.J., 2002. A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and Hearing*, 23(6), pp.532–9.

Sussman, E. et al., 2008. The maturation of human evoked brain potentials to sounds presented at different stimulus rates. *Hearing Research*, 236(1-2), pp.61–79.

Stone, M. A., Visram, A., Harte, J. M., & Munro, K. J. (2019). A Set of Time-and-Frequency-Localized Short-Duration Speech-Like Stimuli for Assessing Hearing-Aid Performance via Cortical Auditory-Evoked Potentials. *Trends in Hearing*.

Tsui, B., Wong, L. & Wong, E., 2002. Accuracy of cortical evoked response audiometry in the identification of non-organic hearing loss. *Interna Journal of Audiology*, 41(6), pp.330–333.

VanDun, B., Carter, L. & Dillon, H., 2012. Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds. *Audiology Research*, 2(1), p.13.

Walter, B.A., 1964. Retrospective summary of definitive tests for hearing in young children. *Acta Otol. (suppl)*, 206, pp.162–172.

Weber, B.A., 1970. Habituation and dishabituation of the averaged auditory evoked response. *Journal of Speech and Hearing Research*, 13(2), pp.387–94.

Wunderlich, J.L., Cone-Wesson, B.K. & Shepherd, R., 2006. Maturation of the cortical auditory evoked potential in infants and young children. *Hearing Research*, 212(1-2), pp.185–202.

Appendix 1: Supra-threshold CAEP testing in infants

Clinical use of CAEP testing in infants

There is an increasing evidence base that CAEP testing can be beneficial for clinical care of infants with hearing loss and the method appears well tolerated by caregivers.

CAEP testing, using the HearLab system, is widely used in Australia to verify infant hearing aid fittings and assess hearing impairment in infants with ANSD (Mehta et al., 2020; Punch et al., 2016). Punch et al. comment that 'The results indicated that clinical CAEP testing influenced audiologists' approach to rehabilitation and was well received by parents and that they were satisfied with the technique. Three case studies were selected to illustrate how CAEP testing can be used in a clinical environment.' Mehta et al. comment that 'The main benefit from the use of CAEPs (using speech token stimuli) was for infant hearing aid fitting programmes, to facilitate earlier decisions relating to hearing aid fitting, for fine-tuning the aids and as an additional measure for cochlear implant referrals.'

Children with ANSD may show CAEP responses when ABR responses are absent. Pearce et al. (2007) report that children with ANSD who showed a CAEP response to a 55 dB SPL stimulus were likely to have no worse than a mild hearing loss. A recent study at the University of Manchester, in collaboration with Interacoustics Research Unit, has investigated infant aided CAEPs using more frequency-specific stimuli,the 'ManU-IRU' stimuli (described in Stone *et al.*, 2019), which can be integrated into the Interacoustics Eclipse system, and are available from the researchers directly, especially if variants of presentation method are used. The protocol was clinically feasible and acceptable to caregivers.

The response may also provide an indication of the maturation of the auditory pathway (Sharma et al. 2002). For example, Sharma et al. (2005) used the CAEP to demonstrate that the cortical responses of children who received cochlear implantation before the age of 3.5 years matured more quickly than the responses of children implanted after 7 years, suggesting a sensitive period of neural development.

Infant CAEP test environment

Facilities and Resources

For infant testing, a test environment suitable for visual reinforcement audiometry (VRA) is appropriate (BSA, 2014). It is also useful to be able to adjust lighting levels.

The optimum state for recording a CAEP is calm and quiet but engaged. To achieve this in infants it is important to have a broad range of engagement and distraction resources. Suitable resources might include light toys, tactile toys, tablets and video projectors.

Recording

A single channel recording is recommended, with electrodes located as follows:

- Positive electrode: high forehead or vertex (avoiding anterior fontanelle)
- Negative electrode: right or left mastoid.
- Common electrode: contralateral mastoid or forehead.

While adult head topography graphs by Sussman et al. (2008) indicate that a positive electrode positioned a little in front of Cz produces the highest signal, this is not recommended in infants because of the location of the anterior fontanelle. Munro et al. (2019) found similar CAEP SNRs for placement of the positive electrode at a high forehead (Fpz) or vertex (Cz) position, but reported that the high forehead position often allowed for easier electrode placement due to the lack of hair.

Strategy

The order in which different stimuli are presented and at which level will be dependent on the clinical question. E.g. when testing infants with a mild high frequency hearing loss, establishing audibility of a high-frequency /s/ at a soft level may be the most clinically valuable approach. Whereas, this would not be as valuable for an infant with a profound hearing loss.

CAEP morphology in infants

The morphology of the CAEP can vary considerably from one infant to the next, and is influenced by a number of factors including attention, arousal and recording parameters. In general, and in awake infants using slow stimulus repetition rates (0.5 to 1.0 /s), there is a prominent positive (upward) wave with a latency of around 150-250 ms. This will typically be preceded by a small negative component at around 80 ms and followed by a broad late negativity around 400 ms. Figure 2, panel A shows the grand average CAEP waveform obtained with the HEARLab system in a group of 74 infants with normal hearing, between 5-39 weeks of age, who were awake and alert during testing (Munro et al., 2019). This was obtained with the stimulus /g/ (duration 21 ms), extracted from a recording of uninterrupted dialogue, and presented in the sound field at a level of 65 dB SPL and at 0^0 azimuth. CAEP waveforms for infants with hearing loss tested when aided show greater variability. This is demonstrated in Figure 2 panel B, which shows the grand average of 104 waveforms from infants with hearing loss, tested wearing their hearing aid(s) with the mid-frequency ManU-IRU signal, (which has a centre frequency equivalent to that of the /g/ stimulus). The infants were tested at input sensation levels ranging from 0 to 35 dB SL, and the average includes only waveforms where a significant response was detected. The low amplitude, broad, bumpy, peak is indicative of the variability seen in the individual waveforms. For this reason, objective detection methods, such as the Hotelling's T^2, are very valuable in interpreting infant CAEP responses.

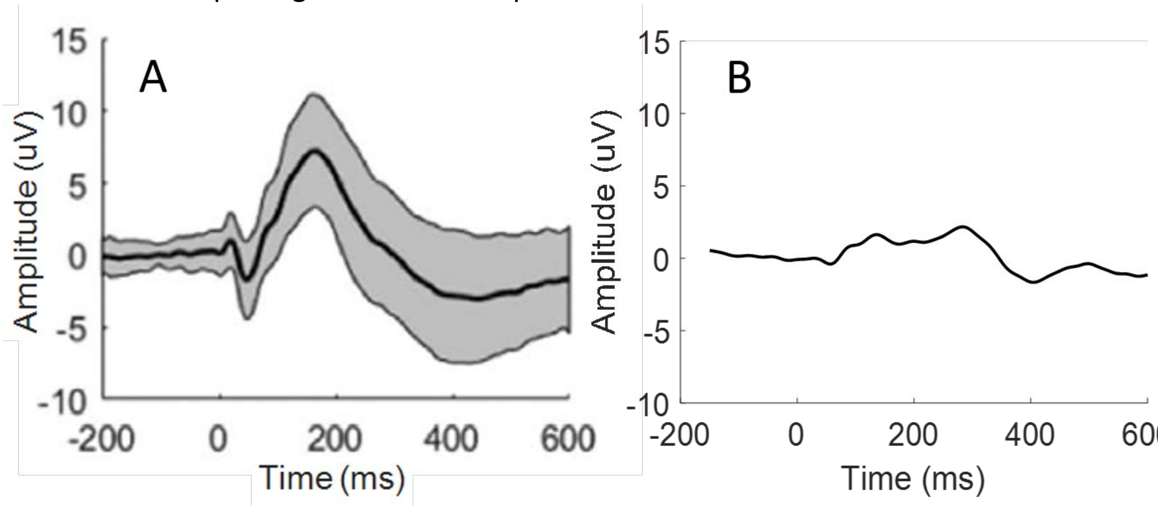


Figure 2: Panel A: (From Munro et al., 2019). CAEP grand average waveform from 74 infants with normal hearing to a 65 dB SPL /g/ stimulus presented at zero degrees azimuth using the HearLab system. The shaded area shows one standard deviation around the grand average. Individual waveforms were averages of around 150 epochs, presented in blocks of 25 Panel B: CAEP grand average of 104 waveforms for infants with hearing loss, tested when aided using the mid-frequency ManU-IRU signal (Stone et al., 2019) at input sensation levels of 0-35 dB SL, presented at zero degrees azimuth using the Interacoustics Eclipse. Individual waveforms were averages of 160 epochs presented in blocks of 20.

CAEP response detection in infants

Munro et al. (2019) showed fair CAEP response detection to speech stimuli in infants with normal hearing. Infants with bilaterally normal tympanograms all showed a response to at least one of the three speech stimuli tested, with 77% showing a response to all three stimuli, and 20% showing a response to two of the three stimuli, for averages of approximately 150 epochs per stimulus presented at 65 dB SPL.

For infants with hearing loss, tested aided or unaided, stimulus sensation levels for a given input level are lower compared to infants with normal hearing, and detection rates are poorer for a given input level. Studies using the HearLab system have found speech signals of 65 dB SPL elicited no response for around 33% of cases when the signal was known to be audible (Chang et al., 2012; Gardner-Berry et al., 2016; Van Dun et al., 2012). A study by the University of Manchester (in preparation) found similar levels of non-detections using ManU-IRU stimuli with the Interacoustics Eclipse. The Manchester study showed the benefits of repeat testing: in 50% of cases where a CAEP was not detected although it was audible, the CAEP was detected on repeat testing. This shows there is significant variability in response waveforms within as well as between subjects. The exact nature of this variability is unclear, but the Manchester study showed an association between how vocal the infant was judged to be during testing, and lack of CAEP detection (even when the recording conditions were judged to be satisfactory). Testers should thus be cautious of interpreting non-responses, especially in cases where the baby was vocal during the test.

Clinical Interpretation of infant CAEP tests

A possible concern in CAEP aided response measurements relates to the transient nature of the stimuli (tone bursts or natural/synthetic speech tokens) which, unlike normal speech, may be shorter in duration than the attack time of a hearing aid's compression circuit thus leading to an unrepresentative assessment of hearing aid performance. However, evidence has shown that the ManU-IRU stimuli are suitable for use with the range of hearing aids tested (Stone et al., 2019). Further research could confirm whether a wider range of CAEP stimuli are suitable for a wider range of hearing aids. The reader should bear these points in mind when interpreting aided CAEP results, and appropriate stimuli should be chosen for aided testing, such as speech tokens or the ManU-IRU stimuli.
Aided CAEPs can be used in infants to confirm physiological detection of sounds, and thus help to reassure families about the benefits of hearing aids, or to expedite alternative care pathways such as referral for possible cochlear implantation. Prior to performing aided CAEP testing care should be given to ensure that hearing aids are set optimally which will include: up to date audiogram and RECD measurement and good fitting earmoulds as these will all impact the ability to record responses. Caution should be exercised in interpreting absent responses, particularly if there has not been an opportunity to repeat the test. The use of CAEP results to inform hearing aid adjustments should be just one of a suite of tools, and a decision to adjust gain should be made in conjunction with the wider clinical picture.

For infants with ANSD, *unaided* CAEP testing has been recommended (Punch et al., 2016). Pearce et al. (2007) reported that, in children with ANSD, if a CAEP was detectable at a low level, such as 55 dB SPL, it would indicate no more than a mild loss. In this way CAEPs can help inform the need for amplification in infants with ANSD.

Appendix 2: An example of CAEP testing of an adult medico-legal compensation claimant.

The claimant, a male in his late 50's worked for 35 years in the textile industry. As he was a non-English speaker, his son acted as interpreter.

The claimant attended for CAEP testing, which at the hospital providing the service included tympanometry, 1 kHz ipsilateral reflex testing, conventional pure-tone audiometry in addition to 4-frequency air conduction CAEP hearing threshold estimation.

Results:
Tympanometry: normal bilaterally.
Ipsilateral 1 kHz Reflexes: equivocal on the right; ≤85 dBHL on the left.
Pure-tone audiometry is shown in Figure 3. This did not correspond to an informal clinical assessment of the claimant's hearing ability at interview.
CAEP thresholds are also shown in Figure 3 and reveal thresholds that, whilst not normal, are substantially better than the audiogram suggest, with a configuration that is not characteristic of noise damage. Note that the "interpolation" approach (Lightfoot & Kennedy 2006) has been used when estimating the objective thresholds.

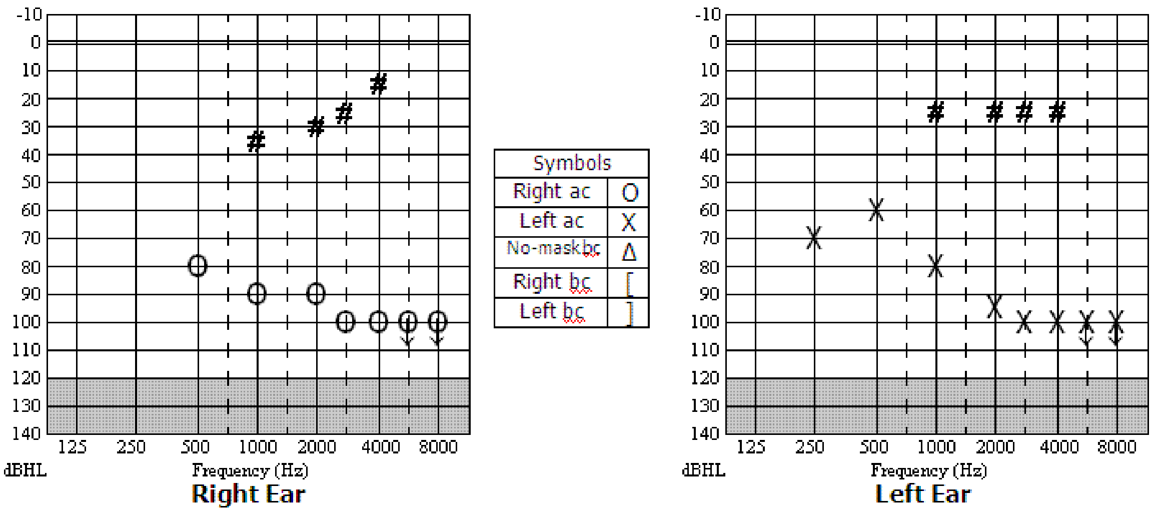
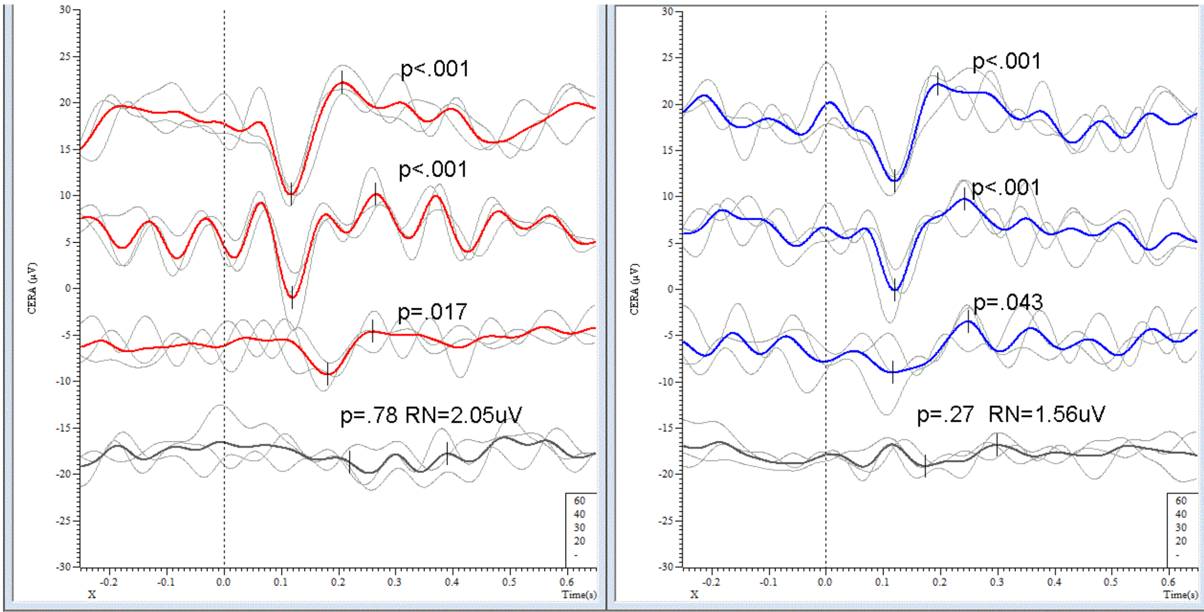


Figure 3
Pure-tone air conduction and corresponding CAEP estimated thresholds (#) of the claimant.

Figure 4 shows the CAEP averaged waveforms at 3 kHz. The pale grey waveforms are 3 sub-averages at each stimulus level whereas the bold coloured waveforms are the grand average of the 3 sub-averages. The system used offers a p-value and residual noise (RN) figure to assist interpretation. The p-value is derived from a combination of the cross correlation and signal to noise ratio of the waveforms, referenced to a data set of 1000 no-stimulus values. The testers used the response present (but using $p \leq .05$) and response absent criteria suggested in this document. This was a real clinical example in which the judgement of the tester was that the recordings were sufficient to define threshold, however it should be noted that in this example fairly large alpha activity can be seen in the R trace at 40 dB HL and some alpha activity can be seen in the L trace at 30 dB HL. The R trace at 40 dB HL was only averaged to 15 sweeps and is deliberately included to show an example of alpha activity. As there was a clear and significant response below this at 30 dB HL, and as the p value at 40 dB HL is significant, then the judgement of the clinician was not to test further at 40 dB HL R. However to obtain a higher quality trace at 40 dB HL R, then more averages could be used and the state of the patient could be checked (for example that they are not too drowsy). The recording on the L at 30 dB HL was judged to be present with a significant response indicated using the statistical approach described above and the grand average latency being in a position expected from higher stimulus levels, so was judged to be sufficient to define threshold with no response at 20 dB HL below. Again if one wanted to reduce noise further on the recording then checking the state of the patient and collecting more averages would be recommended.



Results *

RIGHT EAR FREQ (kHz)	HL dB	SWPS (n)	STAT	N1 (ms)	AMP (µV)	CC	RN	S/N	LEFT EAR HL dB	SWPS (n)	STAT	N1 (ms)	AMP (µV)	CC	RN	S/N
3	60	15	Y	117	12.1	0.82	2.30	5.24	60	15	Y	119	10.5	0.75	2.55	4.11
3	40	15	Y	119	11.1	0.86	1.94	5.73	40	15	Y	120	9.9	0.63	2.39	4.13
3	30	30	Y	181	4.6	0.58	2.01	2.31	30	30	Y	116	5.5	0.30	2.40	2.31
3	20	30	N	-	-	-0.16	2.05	0.45	20	30	N	-	-	0.09	1.56	1.44

Figure 4

CAEP waveforms for the 3 kHz tone burst stimulus. SWPS: sweeps used per grand average; N1: N1 peak latency; AMP: N1-P2 amplitude; CC: cross correlation coefficient; RN: residual noise based on the average gap between sub-averages; S/N signal to noise ratio, measured as the AMP/RN values. The vertical dotted line denotes stimulus onset. The CAEP threshold is 30 dBHL in both ears but is reported as 25 dBHL because the amplitude of the lowest present response is over 3 µV (the hospital adopts a criterion of 5 µV at and below frequencies of 1 kHz). EEG alpha activity is evident and sometimes (as in 40dB on the right) by chance the activity can be similar in sub-averages and so make response identification difficult. At 30dB, chance works in our favour and the alpha activity differs across sub-averages so cancels in the grand average.

Appendix 3: An example of CAEP testing of an adult with learning difficulties.

Case background:
The patient was a female in her mid-thirties with Down Syndrome with severe intellectual impairment and non-verbal communication. Previous behavioural assessment using VRA suggested a brief turn to 60 dBA at 1 kHz. HaLD Speech Discrimination was 60% accurate responses to single words presented at 70 dBA.
Otoscopy revealed narrow ear canals, but was otherwise satisfactory.

CAEP Logistics:
The patient showed initial anxiety about the test and the room. A pre-test visit was arranged so the patient could explore the room and see the equipment before an appointment on another day for the CAEP.
Additional time was required to explain the test.
Time was spent introducing the electrodes and headphones to the patient.

Note that the CAEP responses in Figure 5 are large; this is sometimes seen in adults with learning difficulties.

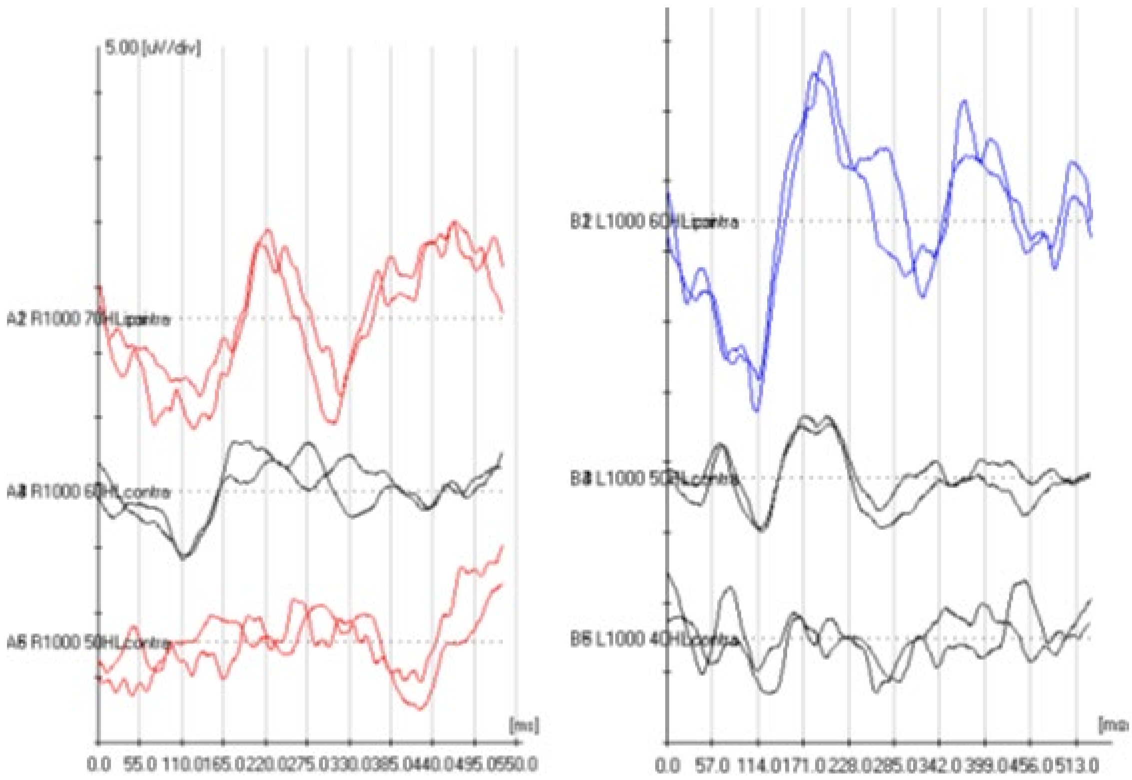


Figure 5

CAEP waveforms for the 1kHz tone burst stimulus in an adult with learning difficulties.

CERA Thresholds:	Left ear:	1 kHz ≤ 50dBnHL	3 kHz ≤ 60dBnHL
	Right ear:	1 kHz ≤ 60dBnHL	3 kHz ≤ 70dBnHL

In this case the criteria for response absence were not met for stimuli below that where a response was present, so the CAEP results were reported using the "≤" qualifier. Clinical judgement was applied when using the results for management purposes.

Management : A hearing aid was fitted to the CAEP thresholds.

Outcomes: The patient's family and care team reported greater responsiveness to sound and increased vocalisations.