Background information on Cortical ERA

**Electric Response Audiometry**

ERA is actually an umbrella term for a collection of techniques in which electrical potentials are recorded, usually from the scalp of the subject, evoked by a sound stimulus. The presence of the response or the response characteristics allow us to infer conclusions about the subject's hearing ability or the performance of their auditory pathways. The original term was *Evoked* Response Audiometry until one bright spark pointed out that a behavioural response such as pressing a button was an "evoked response". The term Electric Response Audiometry has therefore been used. However, the International ERA Study Group have re-adopted the term "evoked" so as to embrace OAEs (which are evoked, but not electrical).

Historical Setting and other auditory evoked responses

The earliest report of relevance was that of Davis who identified the auditory cortical evoked response in 1939 although changes in the EEG evoked by a loud sound had been observed by Berger a decade earlier. Because Cortical ERA (CERA) was the first of the ERA techniques to find widespread clinical use (in the 1970s), the term ERA is sometimes used to refer to this particular technique. Confusingly, CERA is also know by a number of other terms: the *N1-P2 response, slow vertex response* (SVR) and the *auditory cortical response* (ACR). What is more, there are a number of other auditory-evoked responses that arise from the cortex, each having their own characteristics and clinical uses. They include CNV, MMN and P300. This web site makes no attempt to cover these other cortical responses. Also included under the umbrella of ERA are ECochG, ABR and MLR.

What is the "N1-P2" response?
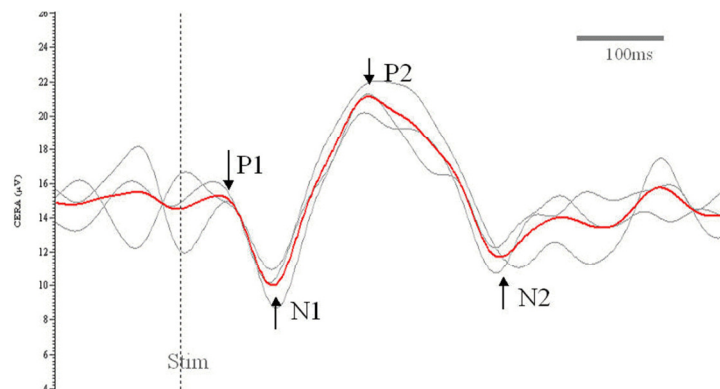
The N1-P2 response is one element of a larger series of events and arises in response to a *change* in auditory environment - it is also referred to as the *acoustic change complex.* In hearing threshold tests it is usually evoked by the onset of a tone, but it may be triggered by any abrupt change - in intensity, frequency etc. or even by the offset of a long tone. The N and P refer to the sign of the potential (negative and positive) at the vertex compared to the potential at the reference electrode.

Waveforms on this site are displayed "vertex positive up". For stimulus intensities well above threshold, N1 has a latency of about 100ms and P2 of about 200ms (you may see them referred to as N_{100} and P_{200}). As intensity is reduced towards threshold, the latencies increase to almost double these figures. The amplitude of the N1-P2 response may be up to about 25µV for moderate to high intensity stimuli, decreasing in size to zero at threshold. These relationships are referred to as input-output functions and knowledge of their characteristics helps us in evaluating an individual's hearing threshold. The generator of N1 is probably the primary auditory cortex but P2 probably has multiple generators, perhaps within the polysensory frontal areas.

Uses of the N1-P2 response

The main clinical application of this response is the objective estimation of the auditory hearing threshold. It may be most conveniently considered as the electrophysiological equivalent of the pure tone audiogram (PTA). The advantages, problems, acoustical constraints and audiological considerations of the PTA are equally applicable to CERA with one important exception: the patient is not asked to play an active part in deciding whether to report that a stimulus has been heard. As such, CERA is most useful when the accuracy of PTA results are in doubt or are clearly erroneous, for example in cases of psychogenic or non-organic hearing loss. Patients with senility or learning difficulties also often yield inaccurate PTA results, yet are willing to offer the passive cooperation required for CERA. However, probably the largest client group is that with military, industrial or occupational hearing loss for whom any pension or compensation for their disability is linked or contingent on their hearing status. Even when the PTA results are accurate, CERA serves to remove all doubt over their validity and as such, can strengthen a claimant's case. The utility of Cortical ERA in the above contexts is well established (Beagley, 1973; Coles & Mason, 1984; Hyde et al, 1986; Alberti et al, 1987; Prasher et al, 1993; Hyde, 1997; Tsu et al, 2002; Cone-Wesson & Wunderlich, 2003; Hone et al, 2003). The Cortical ERA service in Liverpool has undertaken tests on over 9,000 patients / medico-legal claimants since it's introduction 20 years ago and the technique is accepted by the British legal system as the definitive test of hearing status.

Cortical ERA has a major limitation of application: it is based on the N1-P2 response which does not mature fully until the patient's late teens (Stapells, 2002). It is therefore widely regarded as an adult threshold estimation test although it is still a viable test for children as young as about 8 years old. In these older children, the immature response has a different morphology, with N2 & P3 often being more dominant and perhaps because of this, a longer inter-stimulus interval (slower repetition rate) is necessary to record a satisfactory response. However, some audiologists claim to have used CERA with success in 2-3 year olds, though there is insufficient data in the literature to substantiate this.

Although it has no direct neurological application, CERA may be used as an adjunct to other assessment tools to assist in the diagnosis of retro-cochlear pathology. For example, the combination of clear OAEs, normal ABR and an absent CERA can occur in cortical deafness. An absent ABR and recordable CERA responses can be seen in many cases of auditory neuropathy or desynchrony. There is good evidence (Hyde, 1997; Martin & Boothroyd, 1999; Cone-Wesson & Wunderlich 2003) that the N1-P2 response can also be used to access features of auditory discrimination and central auditory processing.

Accuracy of threshold estimation

If the test parameters and protocol are chosen with care (see later), the N1-P2 response is capable of estimating the true hearing threshold of adults with a degree of accuracy at least as good as that of the ABR - within 10dB in most cases (Hyde, 1986; Tsu, 2002). A study in Liverpool using the author's original system suggested a mean Cortical ERA - PTA difference of 4dB. There have been reports that the accuracy of this technique is poor, but it is possible that inappropriate parameters or methodology are responsible. Subject factors are known to influence accuracy. The morphology and amplitude of the N1-P2 complex is degraded with drowsiness and in particular, in the different stages of sleep and although N1 is larger if the subject actively attends to the stimulus, it is sufficient that the patient remains generally alert. Requiring them to quietly read a magazine is ideal. Drugs known to induce drowsiness are to be avoided (sedatives, alcohol etc).   Nevertheless, there is a small percentage of individuals in whom, for no apparent reason, error in the threshold estimate exceeds 20dB (Albera et al, 1991). Ironically, and to our advantage, the quality and size of the N1-P2 response is often better in cases of non-organic hearing loss than in honest subjects. This author believes that this is an unintended attention effect: the stimuli may be of less interest to the honest subject than the malingerer, whose attention is irresistibly drawn to the sounds, particularly those at an intensity below their volunteered threshold yet still audible. Indeed, in

some individuals, a larger response is seen at, say, 10dBSL (sensation level) than at 40dBSL, the higher intensity posing less of a "threat" since it is above their volunteered hearing threshold. As with other ERA techniques (e.g. the ABR), CERA accuracy is better in cases of cochlear hearing loss than in normal subjects: the loudness recruitment associated with cochlear loss compresses the transition between hearing and not hearing into a narrower intensity range, thus making the input-output function steeper.

Methodology
A table below summarises the test parameters.

The electrode montage used for the N1-P2 cortical response is a Cz (vertex) /mastoid electrode pair. Some loss of response amplitude occurs if a high forehead site is chosen instead of Cz (Vaughan & Ritter, 1970). Either mastoid can be used as the reference site, regardless of test ear and indeed, a slight ($\sqrt{2}$) reduction in myogenic activity can be achieved by using a linked mastoid arrangement. By convention, a forehead ground is used.

The filter settings (recording bandwidth) depend, of course, on the spectral peak of the N1-P2 response which lies in the range 2 to 5 Hz. Since we are interested in response detection (rather than analysis), a narrow filter bandwidth helps achieve good signal to noise ratio and is optimally 1 Hz to about 15 Hz (30 Hz can be used if this is the lowest available low-pass setting).

The analysis epoch (time base or window) can be in the range 500 to 1000 ms. It is useful to include a pre-stimulus epoch of about 250 ms to assist in the assessment of background activity. As with other ERA tests, it is important to duplicate or triplicate the response, particularly when the response is small, close to threshold.

Although a click or tone pip may be used, the stimulus of choice is a tone burst of the desired audiometric frequency. The response can be detected at all audiometric frequencies although at frequencies above 2kHz, a smaller response is recorded and so the precision of the threshold estimate is probably poorer. The frequency specificity of this stimulus, and of the response it evokes, is almost ideal and far better than that afforded by tone pips used in ABR tests. This is simply a by-product of the number of cycles in the stimulus. The rise time of the tone burst is an important parameter. If this was very short (if we were to abruptly present the tone burst without a gradual rise time) then we would suffer from a loss of frequency specificity which may be important in steeply sloping or notched audiograms. However, the amplitude of the cortical response diminishes if long rise and fall times are used. A good compromise is to have a linear rise time of 10 to 20 cycles (e.g. 10 ms at 1 kHz). The "plateau" of the tone burst also needs to be defined. Very brief plateaus (<25ms) would compromise frequency specificity and also affect the loudness of the stimulus through the process of temporal integration and hence diminish the response (Davis & Zerlin, 1966; Skinner & Jones, 1968). After the first 30-50ms of the stimulus, the response has been evoked, so there is little merit in extending a plateau for much longer than this. Interestingly, many centres use tone bursts of 100 ms or more. Very long tone bursts should be avoided, since the end of the tone burst will also evoke a cortical "off response" as well as slightly and unnecessarily extending the test time. Those centres using long plateau times will argue that they do so in order to intentionally separate the on and off responses. A plateau of around 100 ms (often advocated) should be avoided since in theory, this can cause the destructive overlapping of the onset P2 and offset N1 responses. In practice, these arguments are rather academic and a plateau of either about 50 ms or 200ms is acceptable. A stimulus of this duration allows us to use the calibration reference values available for pure tone audiometry since the extent of temporal integration is small enough to ignore. Until the recent availability of ISO 389-6 (2007) giving reference values for ABR stimuli, this was a great practical advantage over ABR tests for which there was no agreed calibration values - of particular importance in the medico-legal context.

The choice of stimulus repetition rate is critical and represents a compromise between two opposing considerations. On the one hand, we would like to make the rate fast to shorten the test time, especially if we have several frequencies to test. On the other hand, we do not want to degrade the response and so make its identification difficult. A reasonable question to ask is "what is the maximum rate that does not degrade (reduce the amplitude) of the response?". To record a response unaffected by rate effects, we need to keep the rate down to about one stimulus every **ten seconds**, i.e. 0.1Hz (Appleby, 1964; Davis *et al*, 1966). Using a rate this slow would make the test very time consuming. Although rates above 0.1Hz diminish the response, the rate that yields the best *signal to noise ratio improvement per unit test time* is chosen. For cortical responses in adults it is normal to have a repetition rate between 0.5 and 1.0 stimuli per second (1 - 2 seconds between stimuli) (Rapin, 1964; Davis & Zerlin, 1966). In older children 0.25 to 0.5 Hz (2 – 4s between stimuli) is required. At these rates we record a partially adapted response but we do so in a reasonable time. Of course the very first stimulus in an averaging run is un-adapted because it is preceded by silence and is therefore large. The second is somewhat adapted and the third is more so. The amplitude continues to diminish slightly during the average, though the biggest change is at the start of the averaging run (Walter, 1964; Ozesmi *et al*, 2000).

The above feature plays a part in our choice of the number of sweeps in an average. A very common mistake is to over-average. Averages containing more than 50 sweeps (used to further improve the signal to noise ratio) are often counter-productive, and merely serve to further adapt the response (Henry & Teas, 1968). The number of stimuli required to produce an acceptable response depends upon the size of the response. Stimuli above about 20 dBSL usually produce a clear response after 20 or so stimuli whereas closer to threshold, 30 to 50 stimuli may be required. Replication is essential and for greatest efficiency, the above numbers of sweeps should be distributed across several sub-averages and then combined to form a grand average (e.g. 30 sweeps in total, 10 sweeps in each of 3 sub-averages).

Another way of enhancing response detection is to use a non-rhythmical stimulus and some systems provide the facility for a pseudo-random stimulus rate. This facility used to be common on systems 20 years ago but few systems offer it now - so much for progress! This is also useful in prolonged testing sessions where the response amplitude diminishes due to habituation - a process which can be in part reduced by making the stimulus less predictable (Rapin, 1964; Rothman *et al*, 1970). Other tactics may involve randomising other aspects of the stimulus, for example the ear under test (Butler, 1972), test frequency or test intensity. Giving the patient a brief break or making them more alert in some other (devious?) way can rejuvenate a flagging response.

Summary of recommended test parameters

Parameter	Value	Comment
Electrode Montage	Cz +ve; Mastoid –ve; Fpz Gnd	Linked mastoids may be used
High Pass Filter	1 Hz	
Low Pass Filter	15 Hz	30 Hz if 15 Hz unavailable
Epoch / time base	500 to 1000 ms	250 ms of pre-stim preferred
Stimulus type	Tone burst	Clicks & pips also work fine
Stimulus rise & fall time	10-20 ms	Linear ramp
Stimulus plateau	About 50 ms or 200 ms	Avoid 100 ms

Stimulus modality	Air or Bone conduction	
Stimulus calibration	As for audiometers	Only if using tone bursts
Number of sweeps/trials	5 to 20 per sub-average	Depending on response size
Number of sub-averages	2 to 3	Sum to form grand average
Repetition Rate (adults)	0.5 to 1.0 /s (ISI = 1 – 2 s)	Randomise if possible
Repetition Rate (older children)	0.25 to 0.5 /s (ISI = 2 – 4 s)	Randomise if possible

Below is a summary of the main early papers upon which the choices for test parameters are based. Many parameters are a compromise between conflicting requirements.

Low Pass Filter: Optimum frequency: 15Hz, ideally using a digital filter (*Beagley, 1973*)
Stimulus rise time: A shorter rise time produces a larger response. *(Skinner & Jones 1968).*
Too short a rise time makes the stimulus less frequency specific. Optimum rise time ~10 - 20 ms.
Stimulus duration (plateau): Maximum response seen with durations between 25-50ms. *(Davis & Zerlin 1966), (Skinner & Jones 1968).* A duration of about 100ms can cause onset & offset responses to destructively interfere. Duration > 200ms induces unnecessarily habituation and prolongs test time.
Number of sweeps (stimuli) per average: Response amplitude declines during each averaging run. Use as few as possible because of diminishing return of signal-to-noise improvement *(Walter 1964).*
Fewer stimuli reduces adaptation. *(Henry & Teas 1968).*
Stimulus repetition rate: Over 10s between stimuli required to avoid any adaptation effects. *(Appleby 1964), (Davis et al 1966).* Optimum rate for test is 0.5 – 1.0 Hz *(Rapin 1964), (Davis & Zerlin 1966).*
Stimulus randomisation: Randomisation increases amplitude and reduces adaptation *(Rapin 1964), (Rothman, Davis & Hay 1970).* Amplitude increased if presentation side is randomised *(Butler 1972).*
Test session duration: Poorer responses are recorded after 30 minutes *(Roeser & Price 1969).*
Electrode site: Vertex (Cz) gives optimal amplitude *(Davis & Zerlin 1966).* Amplitude at high forehead is only 60% of vertex amplitude *(Vaughan & Ritter 1970).*

Procedure

With most candidates considered for CERA testing where non-organic hearing loss is suspected, it is worth explaining first what tests will be conducted: the Author's routine is to include tympanometry with acoustic reflexes, pure tone audiometry then CERA (described as the automatic version of the PTA). One then often finds that an accurate PTA is provided, especially if the PTA method is adapted to minimise non-organic overlay (see Cooper & Lightfoot, for example).

For CERA, the patient is required to give their passive co-operation and comply with normal electrode attachment procedures. As with conventional pure tone audiometry, the patient is seated in a standard audiometric room, wearing earphones and is asked to remain quiet and awake. They should be encouraged to read a magazine or book for the duration of the test. The patient should be monitored (close circuit TV & intercom) and re-instructed if they become drowsy, close their eyes or attempt to disrupt the test. Physical relaxation (as required for ABR & steady-state tests) is not necessary and could be counter-productive.

The procedure for the estimation of the hearing threshold at a given frequency is essentially the same as

that used in conventional audiometry - obtain a definite, supra-threshold response and repeat trials at progressively lower intensities until the threshold has been established, using a bracketing technique. To minimise test time however, a 20dB down, 10dB up procedure is advantageous (steps that are twice as coarse as in behavioural audiometry), similar to the procedure often adopted in threshold ABR tests. The chosen threshold is the result of an analysis of the size and latency of the lowest intensity positive response. An interpolation to the nearest 5dB is possible even though a minimum step size is 10dB, hence the term threshold *estimation*. An agreed interpolation rule is necessary. The author uses a 5 µV amplitude criterion (3 µV at 3 kHz and above): if the response is less than this, *that* is the threshold intensity; if greater, the threshold is 5 dB lower.
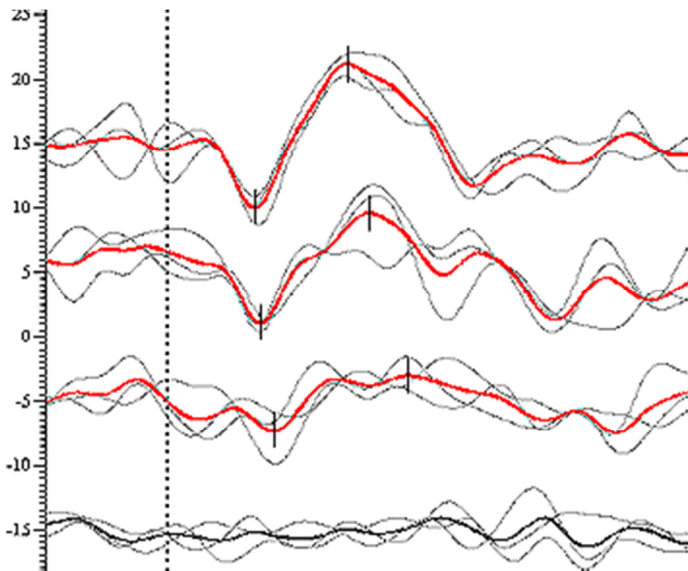
Here is an example of 500 Hz responses obtained at 40, 20, 10 (taken as threshold) and 0 dBHL. The time base extends from 250 ms prior to the stimulus (dashed line) to 650 after the stimulus onset. Three sub-averages are shown superimposed with their grand average (in red, since this is a right ear test). The N1 trough and P2 peak (displayed "vertex positive up") are marked. Collecting the data into sub-averages helps in response identification. Note how there is considerable residual "noise" in the sub-averages. They contain only five sweeps each. Traditionally evaluating the threshold is by subjective assessment but the same objective scoring techniques used in the ABR can be applied to this response.

The choice of the initial test intensity should be made without reference to any existing results from the patient's previous behavioral tests, in order to ensure tester objectivity. A fixed intensity (e.g. 60dBHL) is most appropriate. In cases where a protracted test session is envisaged (as in some medico-legal tests where results at four or more frequencies have been requested), the first threshold to be obtained will give us an approximate idea of the accuracy of a previously obtained audiogram. From this it *may* be possible to start each new frequency at, say, 20 to 30dB above the predicted *true* threshold, thus saving test time by avoiding unnecessary supra-threshold trials. However, some users prefer to retain full scientific objectivity by performing CERA tests blind to any other results.

Masking considerations

As with all audiological tests, we need to consider masking, and the basis of masking in these tests is the same as that used in conventional pure tone audiometry. We do not have the luxury of being able to find the plateau of the masking function so we must calculate the desired masking intensity:

$$Im = Is - TTL + 10 + ABGnt \quad \text{where:}$$

- Im is the masking intensity (calibrated to normal audiometric masking standards),
- Is is the stimulus intensity (calibrated to normal audiometric pure tone standards),
- TTL is the minimum transcranial transmission loss (inter-aural attenuation) associated with the transducer (e.g. 40dB for TDH series earphones)
- ABGnt is the air-bone gap in the non-test ear at the test frequency.

One common problem with the design of most ERA equipment is that manufacturers frequently provide only wide band noise for masking purposes, but if narrow band noise is available then obviously this should be used.