

Cortical Electric Response Audiometry Hearing Threshold Estimation: Accuracy, Speed, and the Effects of Stimulus Presentation Features

Guy Lightfoot* and Vicki Kennedy

Objectives: A number of stimulus presentation features of the tone burst-evoked N1-P2 cortical response were investigated to identify any advantage over simple stimulation when the test is used for hearing threshold estimation. The speed of establishing objective thresholds at 1, 3, and 8 kHz in both ears was also measured in what was designed to be an efficient test protocol, together with the precision of the threshold estimates with reference to subjects' conventional audiograms.

Design: Twenty-four volunteer subjects were recruited and tested by both behavioral and electrophysiological methods. A low-intensity, 3-kHz stimulus was used when the stimulus features were studied. The parameter was the N1-P2 amplitude.

Results: Changing the side of presentation (randomly or by alternating ears), varying the inter-stimulus interval and inserting a 10-second recovery period midway through an averaging run had no demonstrable effect on response amplitude, both individually or in combination, contrary to earlier reports. Establishing the 6 threshold estimates took an average 20.6 minutes. The mean error in the N1-P2 threshold estimate was 6.5 dB, with no significant effect of frequency. After correcting for this bias, 94% of individual threshold estimates were within 15 dB of the behavioral threshold and 80% were within 10 dB.

Conclusions: This study suggests that cortical electric response audiometry has a performance that is as good as or better than the auditory brain stem response for threshold estimation in adults and that sophisticated stimulation techniques do not appear to be required. An efficient test protocol that automates many laborious tasks reduces the test time to less than half that previously reported in the literature for this response.

(*Ear & Hearing* 2006;27:443-456)

The N1-P2 response to auditory stimuli can form the basis of an objective hearing threshold estimation test that can be used in adults and older children (Hyde et al., 1986; Stapells, 2002). It can estimate the

threshold to within 10 dB of the behavioral value in most cases (Hyde, 1997; Prasher et al., 1993; Tsui et al., 2002), although in a small percentage of individuals, threshold errors can exceed 20 dB (Albera et al., 1991). It is variously known as the slow vertex response, the auditory late response, cortical electric response audiometry (CERA), and the acoustic change complex. Here, it will be referred to as CERA1.

Research into the clinical utility of CERA became popular in the 1960s after the pioneering work of Hallowell Davis and colleagues, but the advent of the auditory brain stem response (ABR) in the 1970s saw the decline in CERA research. Although CERA offers similar threshold prediction accuracy to tonal ABR (Stapells, 2002), it has superior frequency specificity by virtue of a longer tone burst stimulus with a correspondingly narrower spectrum, allowing an objective audiogram to be constructed across the full frequency range, if required. Issues relating to neural synchrony that limits the accuracy of short latency (ABR) tests at low frequencies do not apply to the longer latency CERA tests. This excellent frequency specificity has particular advantages in noise-related compensation cases (Coles & Mason, 1984) in which audiometric notches, sometimes obscured by nonorganic overlay, may be resolved, thus addressing issues relating to diagnosis and causation (Alberti et al., 1987; Hone et al., 2003). In the United Kingdom and Ireland, CERA is the test of choice and is accepted by the courts when objective hearing assessment is required in adult medico-legal cases (King et al., 1992). Although it has gained wide acceptance in most countries, this test is surprisingly unrecognized and underused in the United States (Stapells, 2002).

One inherent disadvantage of CERA is the diminution of response amplitude over time (Prosser et al., 1981). Using the basic test techniques available on most ERA systems, test sessions involving 3 or 4 frequencies usually exceed 1 hr (Hyde et al., 1986) and a dwindling response reduces the test accuracy (Roeser & Price, 1969). An effective way to reduce test time is to automate all predictable manual tasks that are normally undertaken by the operator. These include automatic acquisition of multiple sub-averages, creation of grand averages from subaverages, super-

Department of Clinical Engineering, Royal Liverpool University Hospital, Liverpool, United Kingdom.

*The first author has developed a noncommercial educational resource web site giving further details of this response and its clinical uses: <http://www.CorticalERA.com>.

position of replicates, on-screen intensity sorting and cursor placement. Reducing the extent of response degradation itself is another approach: There are a number of findings from early research on this response that might allow us to optimize the stimulus parameters to minimize the problem of response degradation over time. These include:

- Choice of interstimulus interval (ISI). Although an ISI of more than 10 sec is required to avoid any rate-related response decrement effects (Appleby, 1964; McCandless & Best, 1964), the optimum ISI for response detection is 1 to 2 sec (Rapin, 1964; Davis & Zerlin, 1966). This is a compromise between faster rates allowing more averaging to be completed (and thus signal to noise ratio improvement gained) and slower rates allowing greater response amplitude recovery between stimuli.
- Choice of the number of stimuli per average. Averages that comprise fewer stimuli yield larger responses (Henry & Teas, 1968; Nelson et al., 1969). Walter (1964) also warned against overaveraging because of diminishing signal-to-noise improvement. This is because the responses to stimuli at the start of an averaging run are preceded by several seconds of silence and are therefore un-degraded and correspondingly large. As averaging continues at a rate faster than the response recovery time, response amplitude declines (Ozesmi et al., 2000) to the extent that protracted averaging is counterproductive. In other words, the start of an averaging run has a favorable signal to noise ratio and this reduces as the averaging process continues. However, the time course of this degradation appears to be very rapid. Several studies (Bourbon et al., 1987; Budd et al., 1998; Prosser et al., 1981), using an ISI of around 1 second, have shown that the amplitude decrement in N1 is complete by the second stimulus of an averaging sequence and that this represents a process of refractoriness (a rate effect) rather than habituation. Woods & Elmasian (1986) found that the process is complete by the third stimulus in a train. These findings would appear to suggest that protracted averaging runs should not lead to a diminished response.
- Combining the two points above suggests that when additional averaging is needed (for example, when a questionable response is identified at an intensity near threshold), it may be advantageous to insert a 10-sec stimulus-free period into the averaging process in order to allow the response to recover before averaging continues.
- Response amplitude is increased (degradation is reduced) if the side of presentation is alternated (Butler, 1972; Lammertmann et al., Reference Note 3).

- Variation of the interstimulus interval may increase response amplitude. Some studies have reported an effect (Rapin, 1964; Rothman et al., 1970), whereas others have not (Davis & Zerlin, 1966; Nelson et al., 1969). Any effect may be simply a consequence of the subject's attention to the stimulus: A nonrhythmical sound is inherently more "attention-grabbing" than a predictable monotonous sound. The effect of attention is well established, especially for the longer latency components of the auditory evoked response (Stapells, 2002).

We have tried to exploit the above findings by developing a stimulus presentation paradigm that aims to reduce response degradation and so enhance clinical utility in terms of test time and precision of threshold estimation. Features include a variable ISI, randomized ear presentation, a low number of stimuli per average and a stimulus-free recovery period when further averaging is needed. We have already reported that its use in obtaining CERA thresholds at four frequencies in both ears takes typically 30 minutes (Lightfoot & Horseman, Reference Note 4). However, the effectiveness, if any, of the many aspects of stimulus variation on response amplitude have not been evaluated within the context of our stimulus paradigm at near-threshold intensities.

We suspect that the sensation level of the stimuli in one ear may affect the amplitude of responses to stimuli in the other ear. We have often observed that a near-threshold stimulus is less likely to generate a clear N1-P2 response if the stimulus delivered to the other ear has a higher sensation level. In other words, we suspected that there may be an ear interaction, which may limit any advantage of randomized ear stimulation.

The present study was designed to evaluate the individual and combined effects of the above aspects of stimulus presentation on the N1-P2 response. A second aim was to establish the accuracy and speed of CERA threshold estimation in tests employing these stimulus features. A third aim was to establish to what extent the loudness of the stimuli in one ear influences the response to stimuli in the other ear when our randomized ear stimulus paradigm is used.

METHODS

Subjects

Twenty-four adult volunteers were recruited to participate as subjects, drawn from the staff of the authors' or allied departments at the Royal Liverpool University Hospital (15 men, 9 women; mean age, 39 yr; range, 22 to 59). No restrictions were applied with regard to the subjects' hearing status, though most were found to have normal hearing for

their age. Our choice of a subject group with predominantly normal hearing was intentional. The accuracy of evoked potential based hearing threshold assessment is known to be enhanced in cases of significant loss (Hyde, 1985; Johnson & Brown, 2006), probably through a process related to recruitment, and worst in cases of normal hearing. Our examination of the accuracy of CERA is therefore the most stringent one, likely to be bettered in clinical practice. No subject was excluded on the basis of any test result or because of excessive artifacts.

Procedure

All subjects underwent conventional pure-tone audiometry (PTA), with the second author using a standard audiometric technique based on a 10-dB down, 5-dB up bracketing procedure (British Society of Audiology, Reference Note 1). Equipment was a Kamplex (Interacoustics) AC30 audiometer calibrated to ISO 389-1 (similar to ANSI S3.6-1996) with Telephonics TDH-39 earphones in an audiological test room (ambient noise, 19 dBA). The first author then conducted the CERA tests at 1, 3, and 8 kHz, blinded to the PTA thresholds until after the subject's CERA thresholds had been established. The three test frequencies were chosen for the following reasons. In many countries, CERA is used to establish the true hearing thresholds in compensation or war pension (veteran) cases where there is a suspicion of nonorganic hearing loss. Disability is calculated by averaging the hearing thresholds at a number of frequencies, typically 1, 2, and 3 kHz. Establishing the precision of the CERA threshold estimate at 1 and 3 kHz will provide a measure of its utility in this context. The N1-P2 response is thought to be smaller at high frequencies (Antinor & Skinner, 1968; Gibson, 1978; Picton et al., 1978). No study could be found that evaluates the precision of threshold estimation at 8 kHz. This frequency was therefore included in this study because high frequency thresholds can be useful in the diagnosis of noise-induced hearing loss.

CERA tests used Ag/AgCl scalp electrodes at the forehead (ground), Cz (noninverting) and linked mastoids (inverting) with interelectrode impedances (measured at 20 Hz) in the range of 3 to 5 k Ω . Subjects sat in the same test environment used for their PTA and were asked to remain quiet but awake and read magazines during testing. They were informed that the procedure was automatic and no instructions were given with regard to active listening to the stimuli. This mirrors our clinical practice. A CERA system developed by the first author was used, based on a

desktop PC running CED* Signal 2 software, a CED Power 1401 processor, a CED 1902 medical grade isolated amplifier and a second Kamplex (Interacoustics) AC30 audiometer operating under PC control. A suite of software scripts were written to control stimulus generation and response processing (Lightfoot & Horseman, Reference Note 4), modified to meet the requirements of the experiments. Stimuli were tone bursts having 10-msec linear rise/fall times and a 60-msec plateau time generated by the CED Power 1401 and delivered to external inputs of the audiometer. The audiometer was used to attenuate and route the stimuli to the desired ears. Again Telephonics TDH-39 earphones were used and stimulus intensities were calibrated to the same reference equivalent threshold sound pressure levels as used for the PTA. Each average resulted from 15 or 30 stimuli, delivered at an ISI of 1.4 sec, which was either fixed or allowed to vary randomly in the range $\pm 15\%$. The recording epoch (time base) was 900 msec, 250 msec of which was prestimulus onset and 650 msec was poststimulus onset. The single-channel EEG was filtered between 1 Hz and 15 Hz (12 dB per octave), digitized (16-bit resolution) at a rate of 1 kHz, and individual sweeps were further subject to a real-time 15-Hz low-pass digital filter before being added to the average. Although artifact rejection was used ($\pm 50 \mu\text{V}$), rejects were a rare event.

Experiment 1

After their PTA, subjects underwent electrode attachment and CERA threshold estimation tests at the three test frequencies. Each subject was allotted one of the six possible orders of the three frequencies to eliminate order effects. At each frequency, tests began at 60 dB HL and a 20 dB down, 10 dB up bracketing procedure was used. Unlike the PTA procedure therefore, the CERA procedure employed a 10-dB minimum step size. The CERA threshold was established from visual inspection of the waveforms. Two values for the CERA threshold were derived. In the first (the "absolute" value) the threshold was simply the minimum intensity at which a definite response was seen. In the second (the "interpolated" value), if the N1-P2 response was larger than an empirically derived criterion (5 μV for 1 kHz; 3 μV for the higher frequencies) at that intensity, the threshold was taken as 5 dB below that intensity. If the response was smaller, that intensity was taken as the threshold. The interpolated CERA threshold therefore provides statement of a CERA threshold in 5 dB steps by exploitation of the

*Cambridge Electronic Design, Cambridge, UK <http://www.ced.co.uk>.

anticipated intensity input-output function of the response. A comparison of the absolute and interpolated CERA thresholds should indicate which approach provides the more accurate estimate of the hearing threshold.

Both ears were tested in a pseudo-simultaneous fashion, in which individual stimuli were directed to one or other ear on a random basis. The software algorithm ensured that blocks of 10 stimuli always contained equal numbers of right and left ear stimuli. Thus, the maximum possible number of stimuli delivered consecutively to a given ear was five. A variable ISI was used ($1.4 \text{ sec} \pm 15\%$, with a rectangular probability distribution). Thirty stimuli were delivered (15 to each ear, 5 contributing to each of 3 sub-averages for each ear, updated on a cyclical basis), after which a grand average was automatically created for each ear. The grand averages and sub-averages were then automatically superimposed for inspection. The cyclical averaging into the three sub-averages was designed to minimize the extent to which changes in the EEG or arousal level lead to changes in the sub-averages. Figure 1 illustrates the basis of this averaging procedure as a flow

diagram. Figure 2 shows a typical intensity series of responses to tone bursts at 40, 20, 10, and 0 dB HL. Grand averages are bold, superimposed on the 3 sub-averages.

If a clear response was seen, with good repeatability across the sub-averages, the process was halted and cursors applied for measurement of N1 and P2. If a poor or absent response was apparent, a second sequence of 30 stimuli were delivered but only after a 10 s stimulus-free recovery period had expired. Cursor placement was automatic but could be adjusted by the user. Our test protocol was also designed to save time by automating all predictable manual tasks that on a conventional ERA system would be normally undertaken by the operator. Automated activities included automatic acquisition of multiple sub-averages, creation of grand averages from sub-averages, superposition of replicates, and on-screen intensity sorting of waveforms. The overall time taken to establish the 6 CERA thresholds was noted. This was the "headphones on" time, which did not include time taken for electrode site preparation/attachment but did include response analysis/storing/printing.

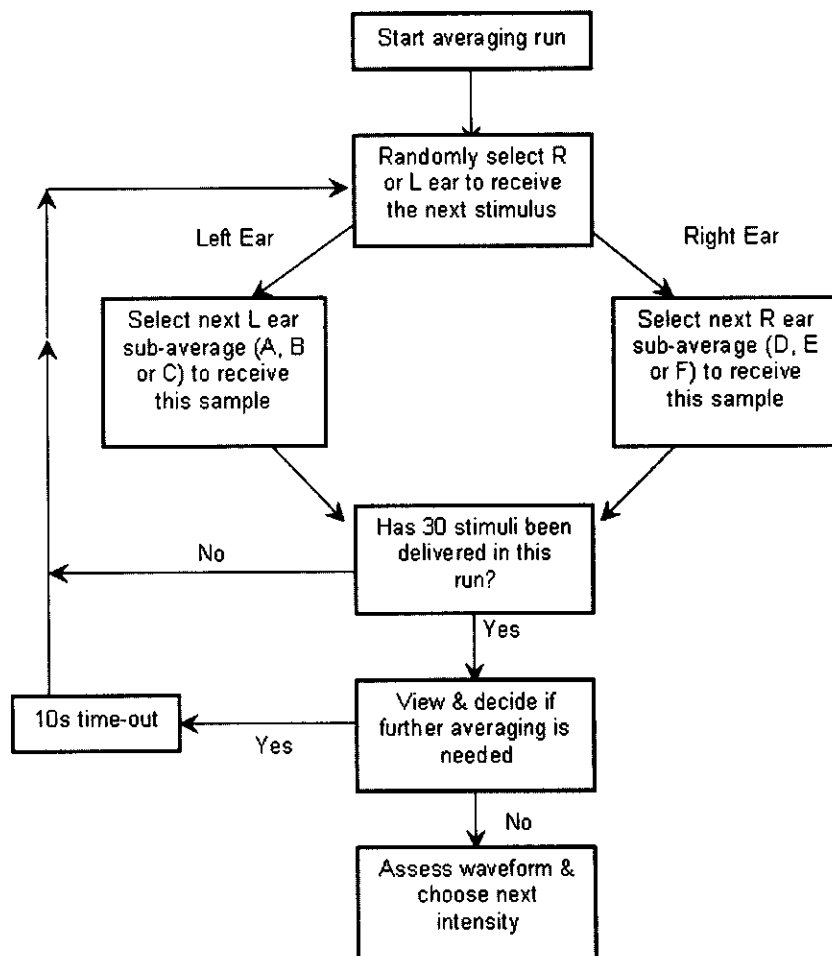


Fig. 1. Simplified flow diagram of the pseudo-simultaneous stimulus and averaging process used in Experiment 1.

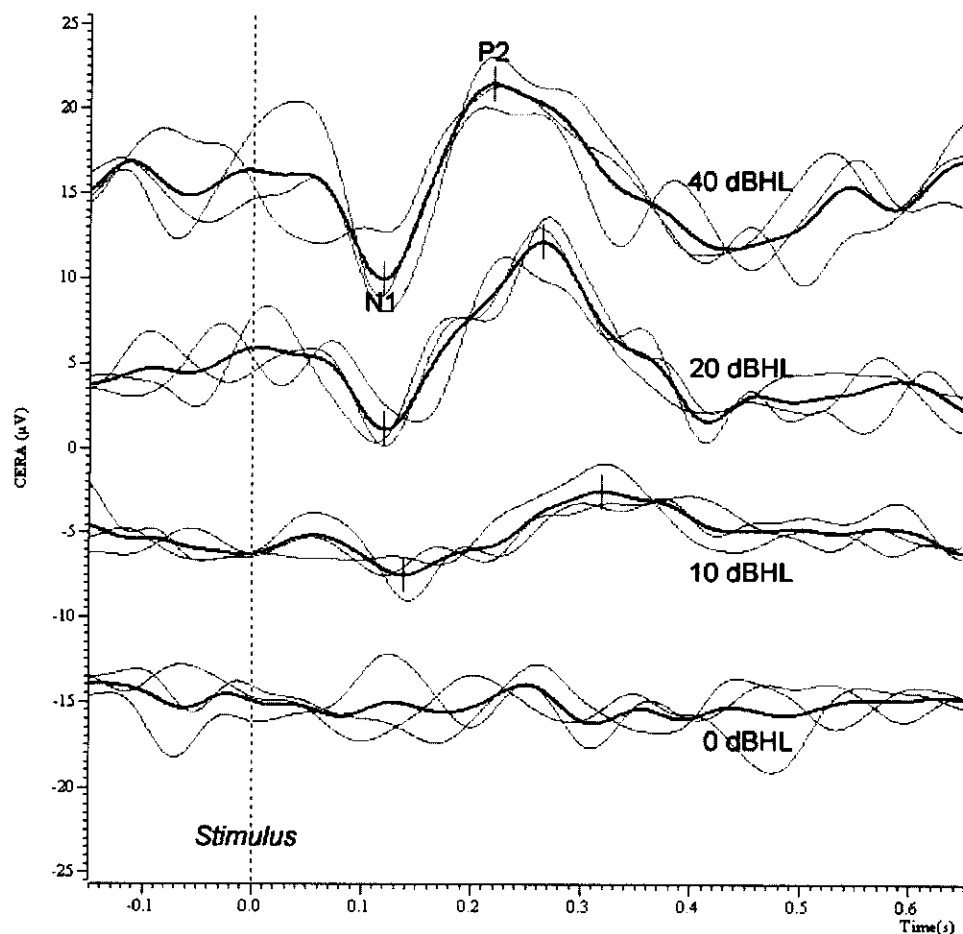


Fig. 2. Example of an N1-P2 intensity series. Waveforms are displayed vertex positive up. At each intensity, 3 sub-averages and their grand average (*bold*) are superimposed. Only 15 stimuli were presented at 40 dB HL and 20 dB HL (5 in each of the 3 sub-averages), whereas at 10 dB HL and 0 dB HL, double that number was used. The stimulus frequency was 3 kHz. No response is evident at 0 dB HL. The response amplitude at 10 dB HL was 5.2 μ V, so the threshold was estimated as 5 dB HL, using our interpolation rule.

Experiment 2

Here, any effect of selected stimulus presentation features on the latency and amplitude of the N1-P2 response was assessed. For reasons of economy, a single frequency (3 kHz was chosen, simply because it was the intermediate frequency of those used in Experiment 1) was used, at a fixed intensity: 25 dB above the threshold established from the subject's PTA (the PTA now being known to both testers). This intensity was chosen because we wanted to assess possible effects at a near-threshold intensity while maintaining waveform clarity. Experiment 2 was conducted on only one of each subject's ears, chosen randomly by tossing a coin. Four stimulus conditions were used.

- Condition A comprised 30 stimuli delivered to the test ear at a fixed ISI (1.4 s) without a break.
- Condition B was the same as condition A but with a 10-second stimulus-free silent interval interposed after the first 15 stimuli.
- Condition C was the same as condition A but with an ISI that varied (1.4 sec \pm 15%).
- Condition D was the random pseudosimultaneous binaural presentation paradigm used in experiment 1 except that the nontest ear received no

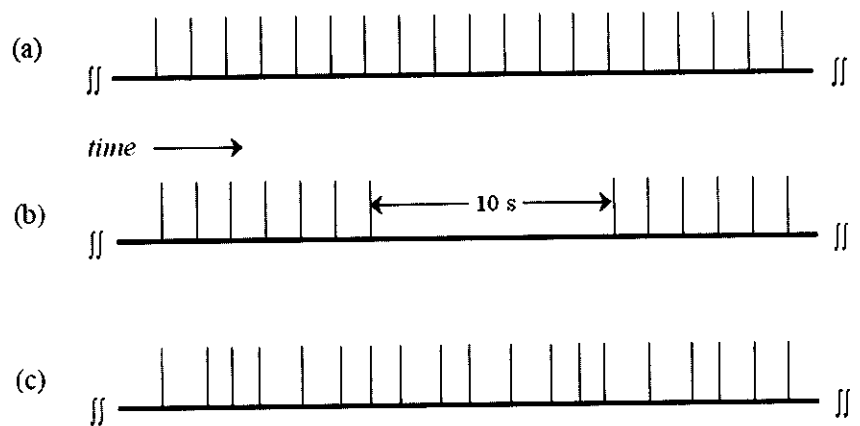
audible stimuli (i.e., the nontest ear stimulus intensity was adjusted to be below the subject's threshold). The test ear received 30 stimuli, with the variable ISI and a 10-second break after 15 stimuli. This will be referred to as $D_{25/0}$.

Conditions A, B, and C are illustrated in Figure 3; Figure 4d relates to Condition D. There are 24 possible orders of presenting these 4 conditions. Each subject was assigned one of the 24 orders to eliminate order effects. In each condition, the latency and amplitude of the N1 and P2 response was measured. Each condition was applied twice in succession (the intervening stimulus-free break being 10 to 15 sec) to establish any short-term adaptation of the response.

Experiment 3

The effect of stimulus intensity in the nontest ear on the response to stimuli in the test ear was investigated using the randomized ear presentation paradigm used in Experiments 1 and 2D. Immediately after Experiment 2, condition $D_{25/0}$ was repeated but with equal sensation level stimuli in the two ears (25 dB above the subject's 3 kHz PTA). Only one averaging run was completed, comprising

Fig. 3. Schematic representation of a segment of the monaural stimulus sequence used to acquire an average in Experiment 2, conditions A, B, and C. Each vertical bar denotes one stimulus, delivered to the test ear. (a), Thirty stimuli at a fixed ISI; (b), 30 stimuli at a fixed ISI with a 10-second recovery period at the half-way point; (c), 30 stimuli at a variable ISI (note the variable spacing between vertical bars).



30 stimuli per ear (10 stimuli into each of the three sub-averages). This will be referred to as $D_{25/25}$ and is illustrated in Figure 4c. The procedure was repeated but with the stimulus intensity in the non-test ear raised to 40 dB sensation level. This will be referred to as $D_{25/40}$. There was no attempt to correct for order effects in these two treatments. Again, in each condition, the latency and amplitude of the N1 and P2 response was measured.

Experiment 4

Finally, the effect of right/left randomized stimulus was investigated. Condition $D_{25/25}$ was undertaken again but with a fixed rather than varying ISI (Fig. 4b). This was then repeated using stimuli that simply alternated from one ear to the other (Fig. 4a). The order of these two treatments was alternated from one subject to the next to control for any order effect.

These four experiments were conducted in the order described here, in one test session. The total

time taken for all (PTA and CERA) measurements was typically 75 minutes.

RESULTS

Except where stated, paired *t*-tests were used to identify the significance of the feature under scrutiny, with $p = 0.05$ being the criterion for significance.

Accuracy and Speed of CERA Threshold Estimates

Table 1 shows the mean hearing thresholds of subjects at the three test frequencies as measured by PTA and both absolute and interpolation CERA threshold methods, together with the mean of the PTA-CERA differences (bias). Table 1 shows that all subjects had near-normal hearing at 1 kHz, but a number had elevated thresholds at the higher frequencies. Inspection of the standard deviations of the CERA thresholds reveals that the interpolation method of defining the CERA threshold yields

Fig. 4. Schematic representation of a segment of the stimulus sequence used to acquire right and left ear averages. Each solid vertical bar denotes one stimulus, delivered to the right ear (in this example), whereas a dashed bar denotes one stimulus, delivered to the left ear. In all cases, a 10-second recovery period (not shown) was inserted at the half-way point in the averaging process. (a), Thirty stimuli for each ear at a fixed ISI, simply alternating right and left; (b), as a, but with the side of presentation being randomized; (c), as b, but with a variable ISI (note the variable spacing between vertical bars); (d), as c, but with the left ear set to an intensity below threshold (denoted by the shorter dashed bars). This is condition $D_{25/0}$ used in Experiment 2.

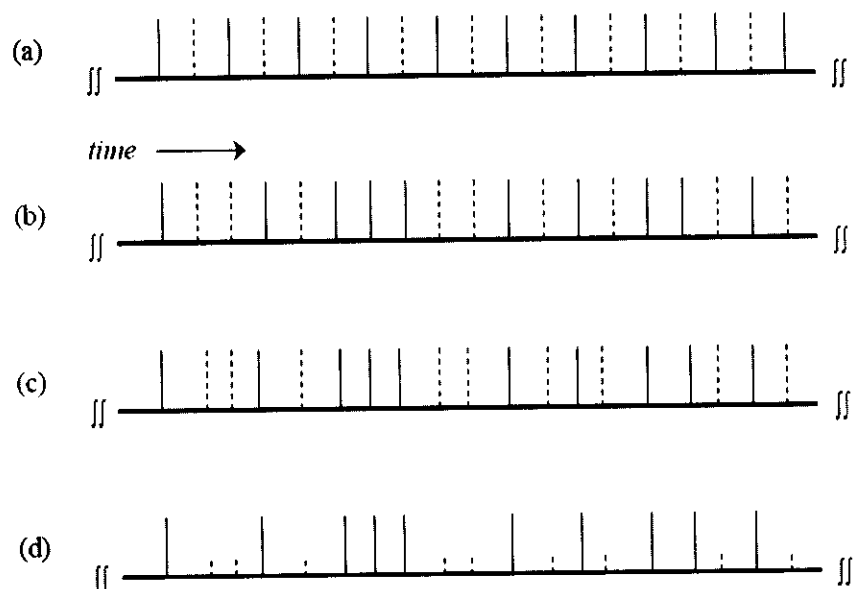


TABLE 1. Subjects' mean (and standard deviation) hearing thresholds as measured by pure-tone audiometry (PTA), cortical ERA (CERA), and subjects' PTA-CERA differences

	1 kHz	3 kHz	8 kHz	1, 3, 8 kHz average
PTA mean	3.9 (5.3)	8.4 (13.5)	15.8 (17.7)	9.4 (11.0)
PTA range	-5 to 15	-5 to 55	-10 to 65	-5 to 43
Absolute				
CERA mean	14.7 (6.2)	18.1 (13.0)	26.2 (18.6)	19.6 (10.6)
CERA range	0 to 30	0 to 60	0 to 80	3 to 50
Interpolated				
CERA mean	10.6 (5.9)	15.0 (12.7)	22.0 (18.4)	15.9 (10.5)
CERA range	0 to 25	0 to 55	- 5 to 75	0 to 47
PTA-CERA	10.7 (6.0)	9.7 (10.0)	10.4 (8.1)	10.3 (5.1)
Interpolated				
PTA-CERA	-6.8 (5.7)	-6.6 (9.8)	-6.2 (8.0)	-6.5 (5.0)

CERA thresholds are shown separately for absolute and interpolation rules (see text). Units are dBHL re: ISO 389-1:1998.

slightly less variability and a smaller PTA-CERA bias at all three frequencies. The CERA thresholds obtained by using the interpolation method will therefore be used in the remainder of this paper. Paired *t*-tests failed to reveal any effect of frequency on the PTA-CERA differences. After accounting for the average 6.5 dB bias, 80% of individual threshold estimate errors were <10 dB, 14% were in the range 10 to 15 dB; 5% were 15 to 20 dB; and 1.4% were 20 to 25 dB. Figure 5 illustrates the relation between PTA and CERA threshold results. Taking the three-frequency average for a given ear (as is common in disability assessment, though using frequencies different from those used here), the CERA threshold estimate error exceeded 10 dB in only 6% of ears.

The frequency specificity of this response and its clinical utility in quantifying hearing loss are already well established and not the focus of this study. Nevertheless, inspection of Figure 5 (b and c) shows a characteristic we have observed many times in both CERA and ABR testing: The precision of the objective threshold is greater for hearing losses and poorer for thresholds near normal. We believe that this is a byproduct of loudness recruitment, which has the effect of sharpening the transition between hearing and not hearing. The 11 PTA thresholds greater than 30 dB HL have associated CERA thresholds whose average bias is only 4.5 dB (SD = 5.8 dB).

The mean time taken to derive subjects' six CERA thresholds (not including electrode attachment) was 20.6 minutes (SD, 3 minutes).

Characteristics of the N1-P2 Response

The latency and amplitude of the responses of all subjects, collapsed across frequency, was analyzed, and the mean and standard deviations are given in Table 2. The stimulus intensity is referenced to individuals' PTA threshold. Note that P2 latency is

more variable than N1. These data are illustrated, together with the functions for individual frequencies, in Figure 6. The same general trend is evident for all three frequencies: Latencies increase and amplitude decreases as intensities reduce towards an individual's threshold. Readers more familiar with the ABR will recognize these general characteristics. Knowledge of these functions is helpful in the interpretation of responses and can be seen in Figure 2.

Effect of a 10-Second Recovery Period

In Experiment 2, test conditions A and B differed only in whether a 10-second, silent interval was introduced mid-way through the averaging process. No significant effect on response amplitude was evident, therefore there appeared to be no advantage in providing this recovery period.

Effect of Varying the ISI

Test conditions A and C of Experiment 2 differed only in their fixed or variable ISI. Again, no effect on varying ISI was seen in this experiment.

Effect of Randomizing the Ear Being Stimulated

In condition D_{25/0}, the amplitude of the N1-P2 response (7.8 μ V) was significantly larger (by 46%, $p = 0.00005$), and the N1 latency (120.3 msec) was longer (by 5.8%, $p = 0.017$) than in conditions A, B, or C. A larger amplitude would provide an advantage in randomizing the ear to which the stimulus is directed. However, in condition D_{25/0}, the stimulus intensity of the nontest ear was set to be below threshold, and because it was only those stimuli directed to the test ear that were audible, the average effective ISI was therefore doubled in comparison to the other test conditions. A more representative test of clinical practice is that of condition

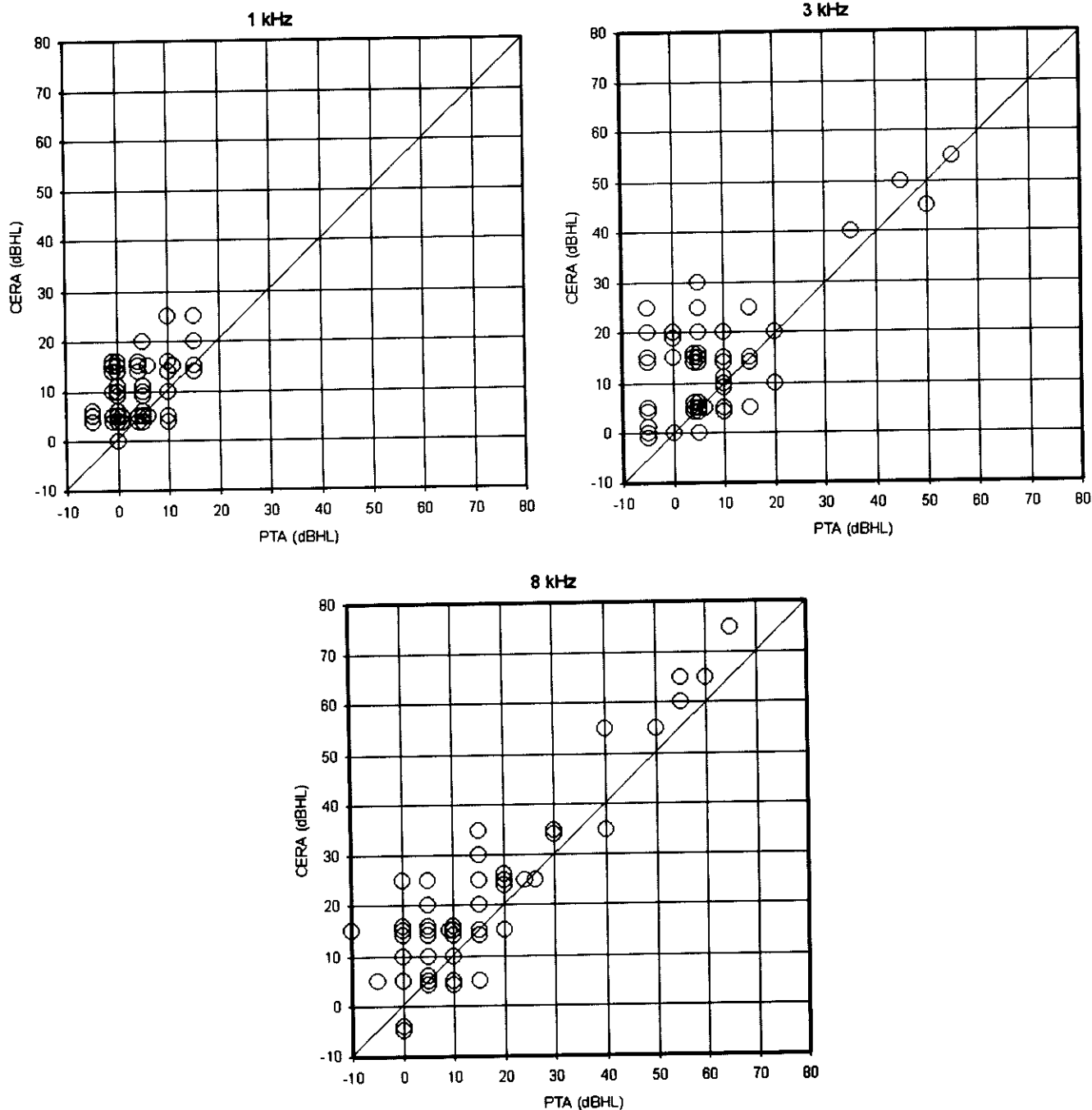


Fig. 5. Scatterplot of subjects' PTA and their corresponding CERA thresholds. Top panel: 1 kHz; center panel: 3 kHz; bottom panel: 8 kHz. Multiple results at the same point have been displaced slightly for clarity.

TABLE 2. Input-output characteristics of the N1 and P2 response: Mean (and standard deviation) values of all subjects and frequencies versus stimulus intensity expressed relative to subjects' PTA hearing threshold (sensation level)

Sensation level (dB)	-5 to +5	10 and 15	20 and 25	30 and 35	40 and 45	>45
N1 latency (msec) (SD)	149.5 (31.5)	128.5 (19.5)	122.1 (15.8)	115.3 (12.7)	115.8 (14.0)	115.5 (16.1)
P2 latency (msec) (SD)	266.6 (50.8)	237.7 (42.1)	229.1 (39.3)	227.5 (45.3)	224.2 (30.3)	222.1 (43.9)
N1-P2 Amp (μ V) (SD)	5.03 (2.12)	5.88 (2.17)	6.70 (2.87)	8.58 (3.81)	8.88 (3.73)	9.56 (3.04)

Latency in msec; amplitude in microvolts.

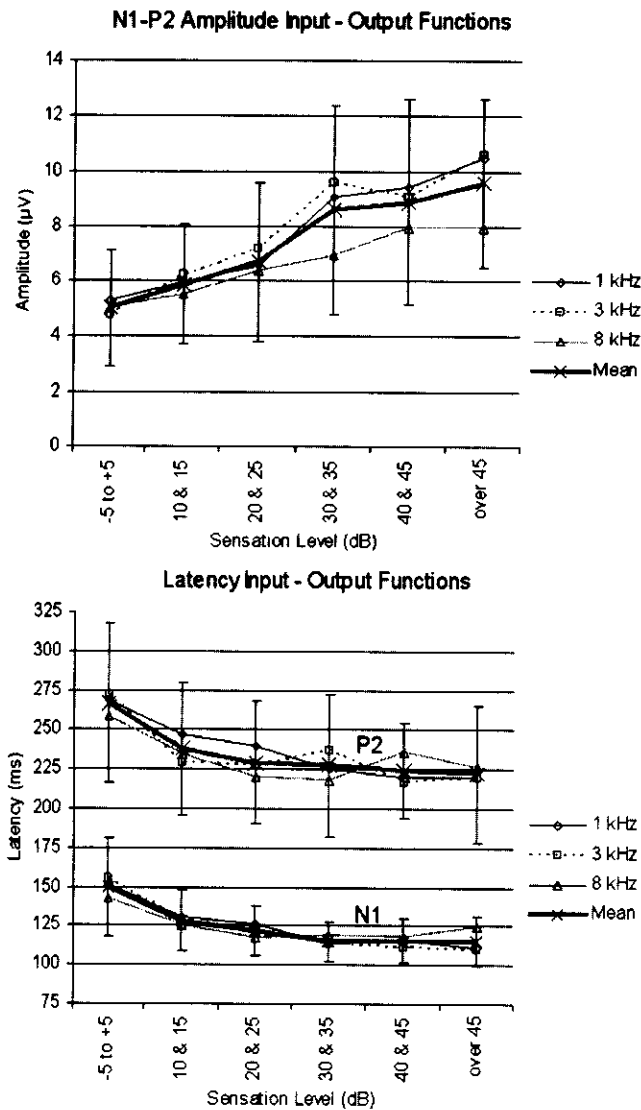


Fig. 6. Input-output functions of the N1-P2 response. *Top panel:* Latencies of N1 and P2; *bottom panel:* N1-P2 amplitude. Mean of all subjects for individual frequencies and the mean function across all frequencies are shown. Error bars show 1 standard deviation for the mean function across frequencies. Stimulus intensity is expressed relative to subjects' PTA hearing threshold (sensation level).

$D_{25/25}$, in which both ears receive randomly delivered stimuli of equal loudness (sensation level). Here, no advantage was seen over the monaurally delivered response amplitudes of conditions A, B, and C. The significant effect seen in condition $D_{25/0}$ is therefore attributable to a rate effect.

Amplitude Decrement of Consecutive Averaging Runs

The four conditions A to D comprised two runs each. Comparison of the amplitudes of the first and second runs (separated by a stimulus-free period of 10 to 15 sec) allowed any adaptation with a time

course of approximately one minute to be identified. Paired *t*-tests of the response amplitudes of the first and second runs of each condition and all conditions combined failed to reveal any significant degradation of the response over this time course.

Effect of Contralateral Ear Intensity

Experiment 3 allowed this to be investigated by obtaining the N1-P2 response to the 3 kHz 25 dB SL stimulus, whereas the other ear received 3 kHz stimuli (a) also at 25 dB SL (condition $D_{25/25}$) and (b) at 40 dB SL (condition $D_{25/40}$), using the random pseudo-simultaneous binaural presentation paradigm. Contrary to expectation, response amplitude was not significantly influenced by the intensity of the stimulus in the opposite ear in this experiment.

Random Ear Versus Alternating Ear Stimulation

Experiment 4 examined this. Paired *t*-tests on response amplitude in these two stimulus conditions failed to identify any significant effect of stimulation format using 25 dB SL stimuli.

DISCUSSION

Our decision to use the standard audiometric calibration reference zero for our tone burst stimuli rather than use locally derived behavioral values was intentional and well considered. It ignores the small degree of temporal integration associated with stimuli lasting only 80 msec and therefore part of our observed PTA-CERA threshold bias is likely to be attributable to this. However, the use of an internationally endorsed calibration reference promotes the comparison and pooling of experimental data and we believe places CERA on a more secure medicolegal footing to ABR, for whose stimuli there are still no agreed reference threshold levels, although an International Standard (ISO 389-6, Reference Note 2) covering these is in preparation.

The accuracy of hearing threshold prediction with the CERA method used in this study agrees well with that of other studies and is similar to that available using ABR. However, if one spends time averaging a greater number of stimuli, a superior signal to noise ratio will be obtained, affording greater result precision. For example, Nousak & Stapells (2005) report that 83% of subjects' 1 kHz ABR thresholds were within only 5 dB of their behavioral thresholds. To obtain this impressive precision, a total of 8000 stimuli were presented at each intensity. Although many would rightly argue that it is important to invest time in obtaining low noise recordings, we suspect that it is rarely done to

this degree in clinical practice. The practicalities imposed by limited test time are important but often overlooked in some studies. Johnson & Brown (2006) recently compared the relative accuracy of ABR and auditory steady state response threshold prediction, but each single threshold estimate for either technique took longer, on average, than our average time to obtain all six thresholds.

Determination of the objective threshold by using an interpolation rule slightly improves precision, but it is hardly a new technique (Cole & Mason, 1984). The mean PTA-CERA bias of 6.5 dB (4.5 dB for thresholds >30 dB HL) is likely to be influenced by loudness recruitment associated with cochlear hearing loss. Our predominantly normal-hearing subjects therefore probably represent a worst-case condition for this test, with somewhat better performance likely in clinical populations. Whether practitioners subtract the mean bias from the CERA threshold (it would be scientifically valid to do so, providing this is stated) is up to the practitioner; we prefer to simply quote the electrophysiological threshold, though we do adopt the interpolation method we have outlined.

The ability of this test to provide accurate and frequency specific estimates of the hearing threshold in clinical populations is already well established (for a review, see Stapells, 2002) but is limited to older children and adults. None of the stimulus manipulations we undertook influenced the spectrum of the stimulus nor introduced any other characteristic that invalidates the applicability of the literature that has already demonstrated the clinical utility of this response. Two case studies demonstrate this. Case 1 is that of a middle-aged man whose symptom profile was consistent with right-sided Ménière's disease. Figure 7 shows his audiogram. Bone conduction symbols are omitted for simplicity but demonstrated a cochlear loss. Narrow band masking noise was applied to the left ear at appropriate intensities in both PTA and CERA tests and because masking was necessary, a monaural CERA stimulus was used. His CERA air conduction thresholds at 1 kHz and 2 kHz were found to be 95 dB HL and 45 dB HL, respectively (using the interpolation rule but without subtracting any PTA-CERA bias), in excellent agreement with the PTA. Figure 8 shows the CERA waveforms. In case 2, a 65-yr-old man with intermittent BPPV was seen in clinic. His hearing loss, shown as Figure 9, was long standing, sensorineural in nature, and, despite the configuration, there was no history of noise exposure or other obvious cause. CERA waveforms (Figure 10) obtained at 2 kHz and 4 kHz were within 10 dB of the PTA in both ears. These two cases illustrate both the frequency specificity and precision that CERA typically provides, both in upward and downward-sloping audiometric configurations.

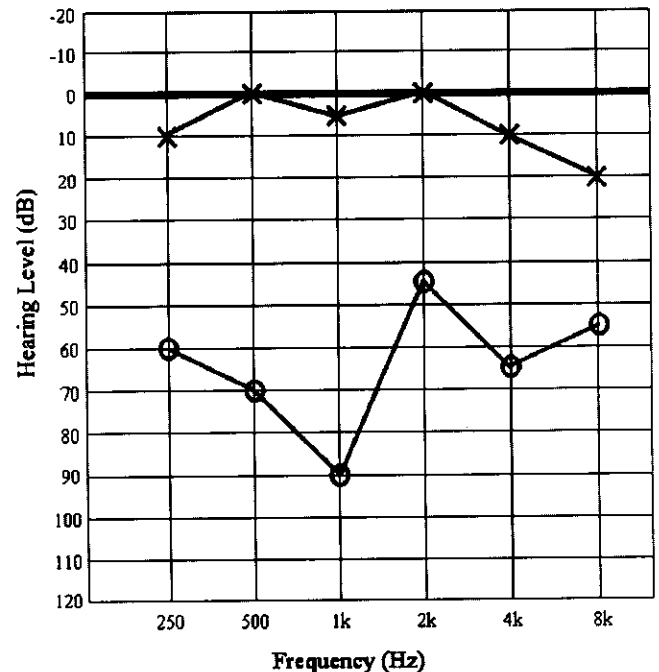


Fig. 7. Pure-tone audiogram in case study 1. Bone conduction thresholds (not shown) indicated that the loss on the right was sensorineural in nature. Symbols: O, right ear air conduction; X, left ear air conduction.

An unexpected but welcome finding in this study is that the threshold estimate at 8 kHz appears to be no less accurate than that at lower frequencies. The capacity to resolve the audiometric contour with good precision regardless of test frequency suggests that CERA may be used to provide medicolegal information concerning both the degree of disability and causation in cases of noise-induced hearing loss. Previous reports of poor precision at high frequencies may have its origin in the order in which tests were performed and in the time taken. CERA tests are often performed at midfrequencies first because it is those that are used when calculating disability. Additional (e.g., 8 kHz or bone conduction) tests are left to the end of the test session, and, with conventional ERA equipment, this may well be over 1 hr from the start. By this time, many subjects' responses will be small, making response identification more problematic. An efficient test protocol therefore has a secondary benefit of allowing a wider range of testing in a single session without compromising accuracy.

Our motivation to speed up the test by automation and minimize any habituation process led us to develop our "attention-grabbing" stimulus protocol. It is somewhat surprising, given the existing literature, that we have been unable to demonstrate any apparent advantage of our stimulus protocol over

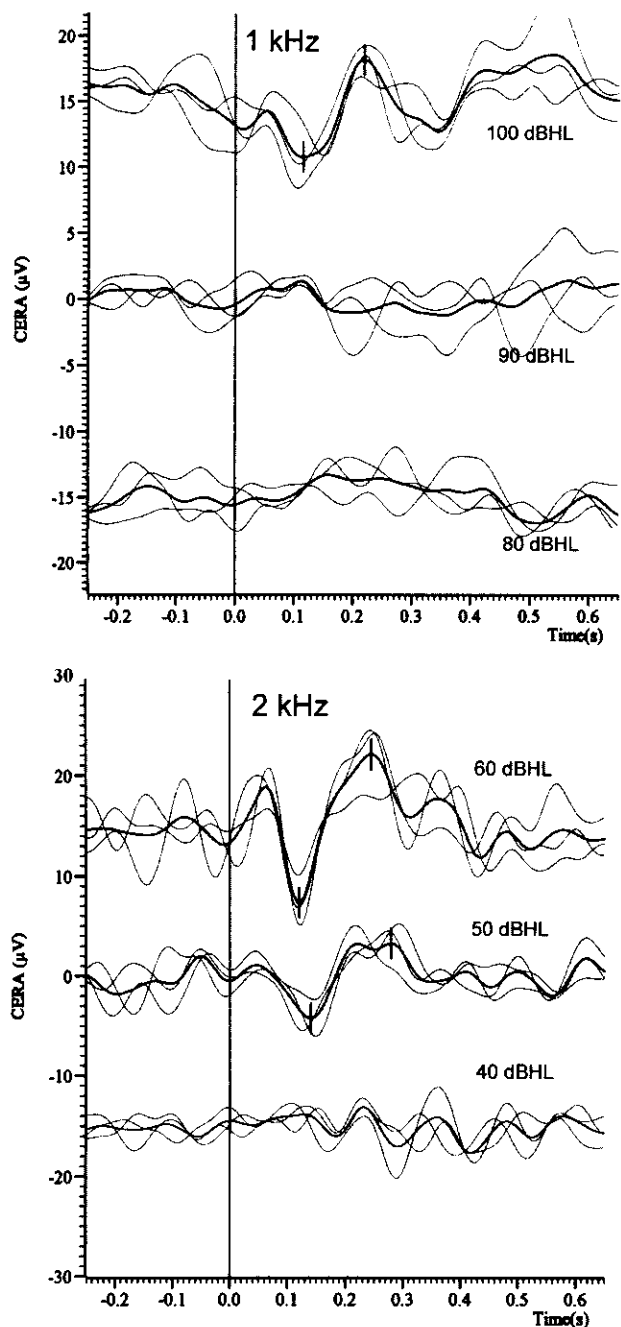


Fig. 8. CERA waveforms of the right ear in case study 1. Top panel: 1 kHz, suggesting a threshold of 95 dB HL; bottom panel: 2 kHz, suggesting a threshold of 45 dB HL.

the standard single ear rhythmical presentation method. On a positive note, our findings suggest that no special stimulus manipulation is required to achieve good test accuracy. Previous studies on the effect of varying ISI have painted a confusing picture but the effect of alternating or randomizing the side of stimulus presentation seems to be well established. Why, therefore, was this effect not seen in the present study? A number of possibilities exist. The signal-to-noise ratio of the waveforms of this

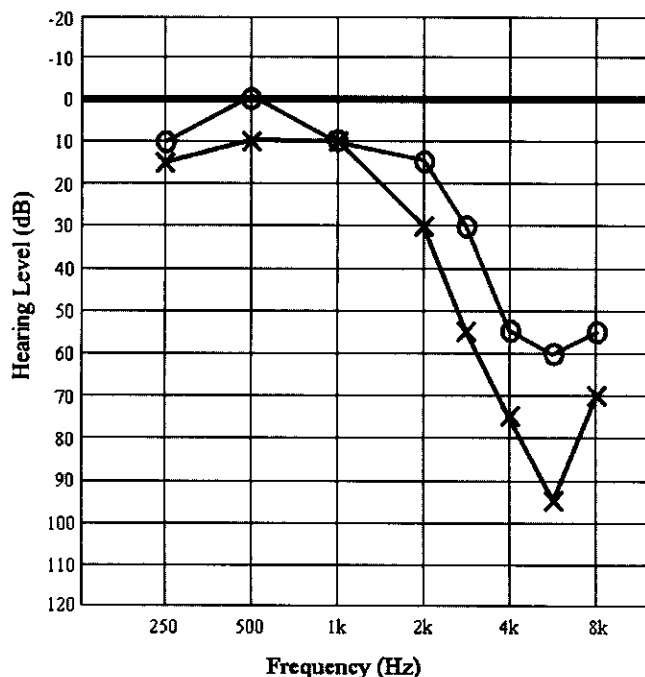


Fig. 9. Pure-tone audiogram in case study 2. Bone conduction thresholds (not shown) showed that the hearing loss was sensorineural in nature bilaterally. Symbols: O, right ear air conduction; X, left ear air conduction.

study reflect that typically obtained in clinical practice, whereas previous experimental studies have used greater averaging and therefore achieved lower noise levels, possibly allowing a small effect to reach a level of significance. A further difference is that of stimulus intensity: No other study of this effect has used a near-threshold intensity, and it is possible that there is actually little or no effect at intensities close to threshold. Finally, our experimental design could not access longer-term habituation effects. The various test conditions presented our subjects with a stimulus whose characteristics changed every few minutes. Even the standard and monotonous stimuli used in condition A may have been interesting simply because its characteristics differed from the stimuli of the previously presented condition. It is possible that this effect alone was sufficient to limit any long-term response degradation that would otherwise have been present in a more conventional test procedure. We are therefore unable to state with certainty that there is no advantage in the use of our variable stimulus.

One worry that we had concerning our random ear stimulus paradigm was that the accuracy of the hearing threshold estimate in one ear may be degraded by adaptation effects induced by stimuli of greater loudness presented to the opposite ear. It is encouraging, therefore, that a reduction in

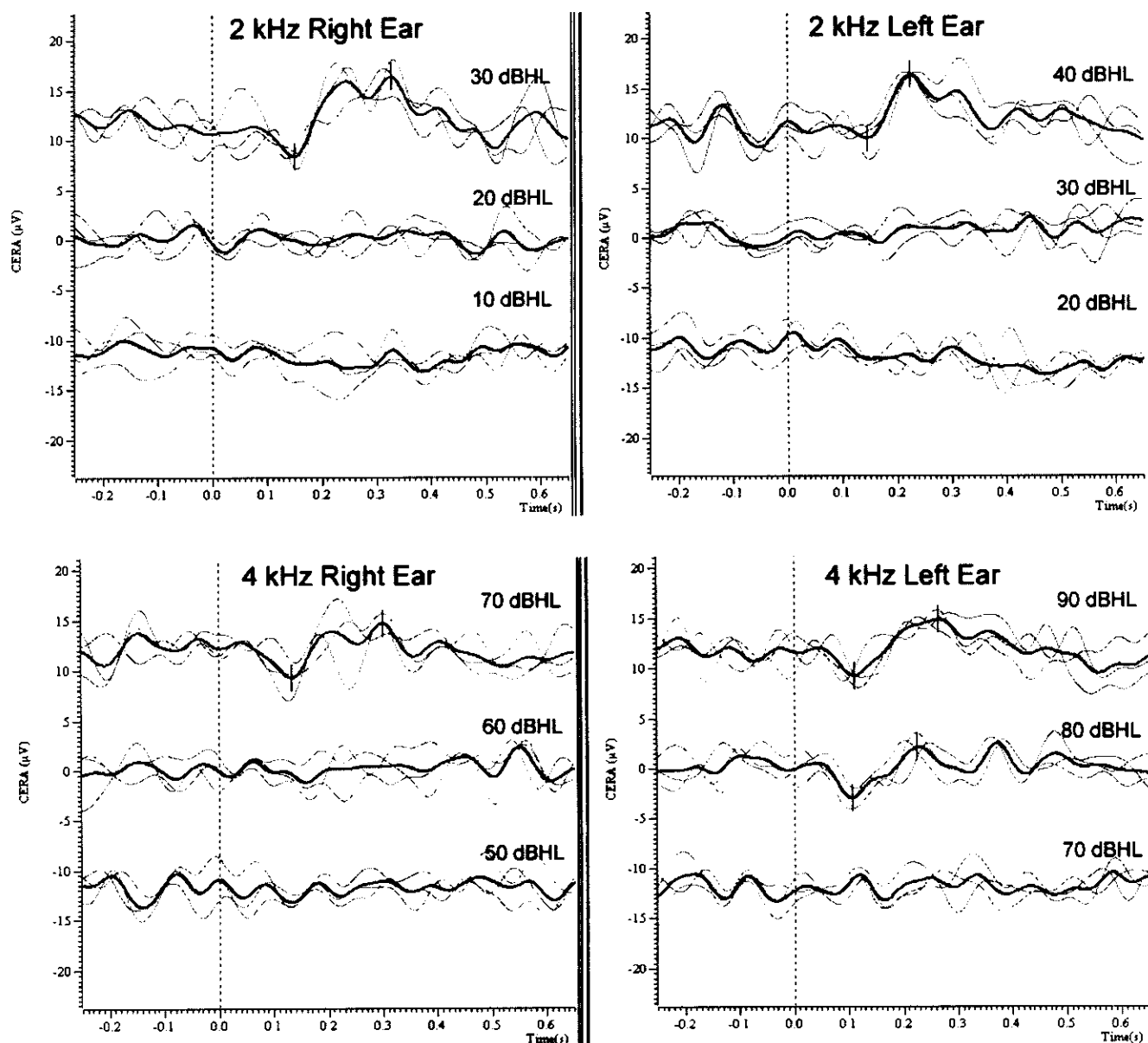


Fig. 10. CERA waveforms in case study 2. Top panel: 2 kHz, suggesting thresholds of 25 dB HL (right ear) and 35 dB HL (left ear); bottom panel: 4 kHz, suggesting thresholds of 65 dB HL (right ear) and 75 dB HL (left ear).

response amplitude was not observed when the intensity of stimuli in the other ear was increased by 15 dB. This suggests that at near-threshold intensities, such interaction is not an issue.

One result of the present study sets our test protocol apart from the standard clinical CERA procedure. The 20.6 minutes taken to establish six CERA thresholds is a half to a third of that reported elsewhere (Hyde, 1997) and is achieved largely by automating the tedious and time-consuming manual manipulation of waveforms and other operator intervention tasks. Van Maanen & Stapells (2005) recently reported that four CERA thresholds in one ear took them an average of 14.9 minutes, equating to 22 minutes for three frequen-

cies in both ears. This was not the actual time it took but rather was calculated from the number of stimuli delivered multiplied by the ISI. In other words it was the theoretical averaging time and an additional 15 to 20% was needed to manipulate and identify waveforms. Our 20.6 minutes (the actual headphones-on time) demonstrate just how efficient our protocol is. This aspect of test performance alone is sufficient to justify a reevaluation of the relative merits of CERA and ABR methods in the objective hearing threshold assessment of adults and older children and is a powerful argument for the development and availability of protocols that include an element of time-saving automation.

ACKNOWLEDGMENTS

The authors thank the subjects that participated in this study and Dr. Steve Mason for his constructive comments on the experimental design and draft manuscript.

Address for correspondence: Dr. Guy R. Lightfoot, Department of Clinical Engineering, Royal Liverpool University Hospital, Prescot Street, Liverpool L7 8XP, United Kingdom. E-mail: g.lightfoot@liverpool.ac.uk.

Received August 31, 2005; accepted April 3, 2006.

Presented in part at the XIXth Biennial Symposium of the International ERA Study Group, Havana, Cuba, June 2005, and at the joint British Society of Audiology and Irish Society of Audiology Annual Conference, Cork, Ireland, September 2005.

REFERENCES

- Albera, R., Canale, G., Magnano, M., Lacilla, M., Morra, B., Rugiu, M. G., & Cortesina, G. (1991). Relations between pure-tone audiometry and cortical evoked auditory potentials. *ACTA Otorhinolaryngologica Italica*, *11*, 551-562.
- Alberti, P. W., Hyde, M. L., & Riko, K. (1987). Exaggerated hearing loss in compensation claimants. *Journal of Otolaryngology*, *16*, 362-366.
- Antinoro, F., & Skinner, P. H. (1968). The effects of frequency on the auditory evoked response. *Journal of Auditory Research*, *8*, 119-123.
- Appleby, S. V. (1964). The slow vertex maximal sound evoked response in infants. *Acta Oto-laryngologica (suppl)*, *206*, 146-152.
- Bourbon, W. T., Will, K. W., Gary, H. E., & Papanicolaou, A. C. (1987). Habituation of auditory event-related potentials: a comparison of self-initiated and automated stimulus trains. *Electroencephalography and Clinical Neurophysiology*, *66*, 160-166.
- Budd, T. W., Barry, R. J., Gordon, E., Rennie, C., & Michie, P. T. (1998). Decrement of the N1 auditory event-related potential with stimulus repetition: habituation vs refractoriness. *International Journal of Psychophysiology*, *31*, 51-68.
- Butler, R. (1972). The influence of spatial separation of sound sources on the auditory evoked response. *Neuropsychologia*, *10*, 219-226.
- Coles, R. R. A., & Mason, S. M. (1984). The results of cortical electric response auditory in medico-legal investigations. *British Journal of Audiology*, *18*, 71-78.
- Davis, H., & Zerlin, S. (1966). Acoustic relations of the human vertex potential. *Journal of the Acoustical Society of America*, *39*, 109-116.
- Gibson, W.P.R. (1978). *Essentials of Clinical Electric Response Audiometry*. Edinburgh: Churchill Livingstone; ISBN 0 443: 01322-5.
- Henry, G., & Teas, D. (1968). Averaged evoked responses and loudness: analysis of response estimates. *Journal of Speech & Hearing Research*, *11*, 334-342.
- Hone, S. W., Norman, G, Keogh, I., & Kelly, V. (2003). The use of cortical evoked response audiometry in the assessment of noise-induced hearing loss. *Archives of Otolaryngology-Head & Neck Surgery*, *128*, 257-262.
- Hyde, M. L. (1985). The effect of cochlear lesions on the A. B. R. In: Jacobson, J.T. *The Auditory Brainstem Response*. 1st Edition. London: Taylor & Francis; ISBN 0-85066-504-3.
- Hyde, M. (1997). The N1 response and its applications. *Audiology & Neuro-otology*, *2*, 281-307.
- Hyde, M., Alberti, P, Matsumoto, N., & Li, Y. L. (1986). Auditory evoked potentials in audiometric assessment of compensation and medicolegal patients. *Annals of Otolaryngology, Rhinology, and Laryngology*, *96*, 514-519.
- Johnson, T. A., Brown, C. J. (2006). Threshold prediction using the auditory steady-state response and the tone burst auditory brain stem response: a within-subject comparison. *Ear and Hearing*, *26*, 559-576.
- King, P.F., Coles, R.R.A.C., Lutman, M.E., & Robinson, D.W. (1992). *Assessment of Hearing Disability. Guidelines for Medico-legal Practice*. London: Whurr Publishers; ISBN 1-870332-04-0.
- McCandless, G. A., & Best, L. (1964). Evoked responses to auditory stimuli in man using a summing computer. *Journal of Speech and Hearing Research*, *7*, 193-202.
- Nelson, D. A., Lassman, F. M., & Hoel, R. L. (1969). The effects of variable-interval and fixed-interval signal presentation schedules on the auditory evoked response. *Journal of Speech and Hearing Research*, *12*, 199-209.
- Nousak, J. K., & Stapells, D. R. (2005). Auditory brainstem and middle latency responses to 1 kHz tones in noise-masked normally-hearing and sensorineurally hearing-impaired adults. *International Journal of Audiology*, *44*, 331-344.
- Ozesmi, C., Dolu, N., Suer, C., Golgeli, A., & Ascioğlu, M. (2000). Habituation of the auditory evoked potential in a short inter-stimulus interval paradigm. *International Journal of Neuroscience* *105*:87-95.
- Picton, T. W., Woods, D. L., & Proulx, G. B. (1978). Human auditory sustained potentials, II: stimulus relationships. *Electroencephalography and Clinical Neurophysiology*, *45*, 198-210.
- Prasher, D, Mula, M., & Luxon, L. (1993). Cortical evoked potential criteria in the objective assessment of auditory threshold: a comparison of noise induced hearing loss with Meniere's disease. *Journal of Laryngology and Otology*, *107*, 780-786.
- Prosser, S, Arslan, W., & Michelini, S. (1981). Habituation and rate effect in the auditory cortical potentials evoked by trains of stimuli. *Archives of Otorhinolaryngology*, *233*, 179-187.
- Rapin, I. (1964). Practical considerations in using the evoked potential technique in audiometry. *Acta Oto-laryngologica (suppl)*, *206*, 117-122.
- Roeser, R., & Price, L. (1969). Effects of habituation on the auditory evoked response. *Journal of Auditory Research*, *9*, 306-313.
- Rothman, H., Davis, H., & Hay, I. (1970). Slow evoked cortical potentials and temporal features of stimulation. *Electroencephalography and Clinical Neurophysiology* *29*, 225-232.
- Stapells, D. (2002). Cortical event-related potentials to auditory stimuli. In: Katz J. *Handbook of Clinical Audiology*. 5th Edition. Philadelphia: Lippincott Williams & Wilkins; ISBN 0-683-30765-7.
- Tsui, B, Wong, L. L., & Wong, E. C. (2002). Accuracy of cortical evoked response audiometry in the identification of non-organic hearing loss. *International Journal of Audiology*, *41*, 330-333.
- Van Maanen, A., & Stapells, D. R. (2005). Comparison of multiple auditory steady-state responses (80 versus 40 Hz) and slow cortical potentials for threshold estimation in hearing-impaired adults. *International Journal of Audiology*, *44*, 613-624.
- Walter, W. G. (1964). Retrospective summary of definitive tests for hearing in young children. *Acta Oto-laryngologica (suppl)*, *206*, 162-172.
- Woods, D. L., Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. *Electroencephalography and Clinical Neurophysiology*, *65*, 447-459.

REFERENCE NOTES

- 1 British Society of Audiology. Recommended procedure. (2004). Pure tone air and bone conduction threshold audiometry with and without masking and determination of uncomfortable loudness levels.
- 2 ISO 389-1:1998 Acoustics. Reference zero for the calibration of audiometric equipment - Part 1: Reference equivalent threshold sound pressure levels for pure tones and supra-aural earphones.
- 3 Lammertmann, C., Fujiki, B., Lütkenhöner, B., & Hari, R. (2000). Short-term decrement of the auditory N1m response. In: *Biomag 2000, Proceedings of the 12th International Conference on Biomagnetism*. Nenonen J., Ilmoniemi R. J., & Katila T., Editors. pp. 50-53. Espoo, Finland: Helsinki University of Technology.
- 4 Lightfoot G., & Horseman G. (2003). Optimising the N1-P2 adult threshold estimation test. XVIIIth Biennial Symposium of the International ERA Study Group, Puerto de la Cruz, Tenerife, June 2003.