The Detection of Infant Cortical Auditory Evoked Potentials (CAEPs) Using Statistical and Visual Detection Techniques

DOI: 10.3766/jaaa.21.5.6

Lyndal Carter*†
Maryanne Golding*†
Harvey Dillon*†
John Seymour*†

Abstract

Background: With the advent of newborn hearing screening programs, the need to verify the fit of hearing aids in young infants has increased. The recording of cortical auditory evoked potentials (CAEPs) for this purpose is quite feasible, but rapid developmental changes that affect response morphology and the presence of electrophysiological noise can make subjective response detection challenging.

Purpose: The purpose of this study was to investigate the effectiveness of an automated statistic versus experienced examiners in detecting the presence of infant CAEPs when stimuli were present and reporting the absence of CAEPs when no stimuli were present.

Research Design: A repeated-measures design was used where infant-generated CAEPs were interpreted by examiners and an automated statistic.

Study Sample: There were nine male and five female infants (mean age, 12 mo; SD, 3.4) who completed behavioral and electrophysiological testing using speech-based stimuli.

Data Collection and Analysis: In total, 87 infant CAEPs were recorded to three sensation levels, 10, 20 and 30 dB relative to the behavioral thresholds and to nonstimulus trials. Three examiners were presented with these responses: (1) "in series," where waveforms were presented in order of decreasing stimulus presentation levels, and (2) "nonseries," where waveforms were randomized completely and presented as independent waveforms. The examiners were given no information about the stimulus levels and were asked to determine whether responses to auditory stimulation could be observed and their degree of certainty in making their decision. Data from the CAEP responses were also converted to multiple dependent variables and analyzed using Hotelling's T^2. Results from both methods of response detection were analyzed using a repeated measures ANOVA (analysis of variance) and parameters of signal detection theory known as d-prime (d′) and the area under the receiver operating characteristic (ROC) curve.

Results: Results showed that as the stimulus level increased, the sensitivity index, d′, increased for both methods of response detection, but neither reached the maximum possible d′ value with a sensation level of 30 dB. The examiners with the greatest experience and Hotelling's T^2 were equally sensitive in differentiating the CAEP from noise.

Conclusions: Hotelling's T^2 appears to detect CAEPs from normal hearing infants at a rate equal to that of an experienced examiner. A clinical instrument that applies Hotelling's T^2 on-line, so that the likelihood of response detection can be assessed objectively, should be of particular benefit to the novice or less experienced examiner.

*National Acoustic Laboratories, Chatswood, NSW, Australia; †Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation, East Melbourne, Victoria, Australia

Lyndal Carter, National Acoustic Laboratories, 126 Greville St., Chatswood, NSW, Australia; Phone: +61 9412 6828; Fax: +61 9411 8273; E-mail: Lyndal.Carter@nal.gov.au

Selected results from this paper were presented at the Audiology Australia conference, May 2008, Canberra, and at the XX IERASG Biennial Symposium, June 2007, Slovenia.

This project was partially funded by the Cooperative Research Centre for Cochlear Implant and Hearing Aid Innovation.

Key Words: Cortical auditory evoked potentials, infants, normal hearing

Abbreviations: ABR = auditory brainstem response; ASSR = auditory steady state response; CAEPs = cortical auditory evoked potentials; CI = confidence interval; d′ = d-prime sensitivity index; ROC curve = receiver operating characteristic curve; SNR = signal-to-noise ratio; VROA = visual reinforcement orientation audiometry

The fitting of hearing aids in young infants, using prescriptive methods, is reliant on threshold estimates that are derived from results using electrophysiological techniques such as auditory brainstem response (ABR) or auditory steady state responses (ASSR) (Dillon, 2001; Bagatto et al, 2005). Given the stimulus output limitations of many clinic-based electrophysiological systems (Golding et al, 2007) and the uncertainties in determining electrophysiological threshold, particularly when abnormal response morphology is evident (Rance et al, 1999; Sininger, 2002), there are risks in using these estimates to fit an infant with hearing aids.

It is not until an infant reaches a developmental age of around 6 mo that threshold detection using behavioral measures becomes possible (Bess and Humes, 2003; Martin and Clark, 2003). With the advent of early identification of hearing loss programs in many parts of the world, a significant reduction in the age of infants at their first fitting has occurred (Ching and Hill, 2007). Given that reliable behavioral testing is generally impossible until after the sixth month of life, there is likely to be a substantial lapse of time between the fitting of hearing aids, based on estimated thresholds, and the confirmation of hearing threshold levels using behavioral test techniques. In addition, even if the initial estimates of hearing threshold are later proved to be correct, it does not always follow that the hearing aid will have been optimally, or even adequately, fitted (Dillon, 2001).

When adults are fitted with hearing aids, it is important to ensure that speech signals are audible, comfortable, and tolerable, and that the goals for rehabilitation that were set prior to the fitting of hearing aids have been achieved (Weinstein, 2000). This verification process is just as important when fitting hearing aids for children, but there are very few ways to evaluate parameters such as audibility and comfort in young infants (Stelmachowicz and Seewald, 1991; Sirimanna, 2001). The observations of parents and professionals are an important starting point. These observations can be more formally documented by using a functional measure of performance, such as the *Parent's Evaluation of Aural/Oral Performance of Children (PEACH) Scale,* which was developed to evaluate the effectiveness of amplification in infants by the systematic use of parental observations (Ching and Hill, 2007). The range of observable behaviors in very young infants, however, is limited (Bess and Humes, 2003).

Electrophysiological methods are also applicable to infant hearing aid evaluation. While it is feasible to use ABR and ASSR to assess hearing aid fitting when behavioral measures cannot be employed (Kileny, 1982; Picton et al, 1998; Brown et al, 1999), recording cortical auditory evoked potentials (CAEPs) for this purpose has several benefits over recording the early latency electrophysiological measures. First, the acoustic features that are relevant for speech detection and perception can be presented as stimuli (Kurtzberg et al, 1988). Second, the integrity of the response pathway through to the cortex can be assessed (Kraus et al, 1998). There may also be additional advantage in some cases of auditory neuropathy/dys-synchrony, as a cortical response may be observed (Rance et al, 2002; Cone-Wesson and Wunderlich, 2003; Pearce et al 2007) while the ABR cannot, and hence ABR results alone may be misleading. CAEPs can be reliably recorded in young awake infants with normal hearing when stimuli are presented at conversational level (Kurtzberg, 1989; Steinschneider et al, 1992; Cone-Wesson and Wunderlich, 2003). They have also been recorded to verify the audibility of stimuli presented at conversational level in infants fitted with hearing aids or in infants who are under evaluation for hearing aid fitting (Rapin and Granziani, 1967; Gravel et al, 1989; Purdy and Kelly, 2001; Cone-Wesson and Wunderlich, 2003).

The CAEP changes significantly with respect to the shape and latency of the major components over the first 14–16 yr of life (Rotteveel et al, 1986; Hyde, 1997; Pasman et al, 1999). These morphological changes are not well understood (Kushnerenko et al, 2002; Wunderlich and Cone-Wesson, 2006) but are likely to reflect underlying developmental changes in the response generators such as improved synaptic efficiency arising from increased axon myelination and maturation of intra- and interhemispheric connections throughout the cortex (Cunningham et al, 2000; Eggermont and Ponton, 2003). The average newborn infant CAEP in response to speech stimuli is dominated by a prominent peak at 200 to 300 msec when recorded at the midline (Kurtzberg, 1989; Stapells and Kurtzberg, 1991; Sharma et al, 2002) but the response amplitude, latency, and wave morphology varies substantially between and within subjects and with varying levels of alertness (Hyde, 1997; Wunderlich and Cone-Wesson, 2006). CAEPs in young children may show even more variability than those of adults because of increased electrophysiological noise brought about by movement of the electrode-skin

interface when physical movement occurs or sudden alterations of psychological state (Hyde, 1994). As a result, the common method for response detection, which is visual observation of (1) response replication, (2) a plausible latency for key response components, and (3) response tracking (i.e., increased latency and decreased amplitude of the response with decreasing stimulus presentation levels) (Elberling and Don, 2007), may be inadequate especially in infant populations. The need for an objective statistical detection technique is therefore apparent in CAEP testing, and yet automated and machine scoring methods, which are not new in ABR testing (Hall, 1992), are not commonly applied in CAEP testing.

We aimed to investigate the effectiveness of an automated statistic versus expert examiners in (1) detecting the presence of infant-generated CAEPs when stimuli of various intensities were present, and (2) reliably reporting the absence of infant-generated CAEPs when no stimuli were present. The results were compared using signal detection theory, which provides a means of calculating the sensitivity and specificity of the objective and subjective detection methods as the stimulus condition is varied. A rating scale, rather than simple forced-choice "presence/absence" of response, was used to determine the d-prime (d′) sensitivity index of stimulus versus nonstimulus trials. This measure was derived by calculating the mean of the signal distribution (i.e., the stimulus-present conditions) and the mean of the noise distribution (i.e., no stimulus-present condition) (McNicol, 1972; Oates et al, 2002; Korczak et al, 2005). Therefore, d′ provides a summary of the CAEP detection hit-rate (i.e., sensitivity) and false-alarm rate (i.e., specificity) in a single metric.

METHOD

Stimuli

The speech stimuli /m/ and /t/ were extracted from continuous discourse spoken by a female with an average Australian accent. The recording was sampled at a rate of 44.1 kHz and filtered to closely match the International Long-Term Average Speech Spectrum (ILTASS). The stimuli include very little of the vowel transition. An additional high-pass filter of 250 Hz was applied to /t/ to remove unwanted low-frequency noise. These essentially vowel-free stimuli have a spectral emphasis in the low- and high-frequency regions, as shown in Figure 1, and thus have the potential to provide information about the audibility of speech sounds in different frequency regions.

Participants

Infants recruited for this study were not "at risk" for hearing loss, having passed automated ABR screening at birth. All were considered to be in good health by their parents, and they had no known developmental disabilities. Of the 17 infants who participated, three were excluded either because of unreliable results to behavioral testing and/or noncooperation for CAEP testing. This was not unexpected, given the age group of the participants, and the stringent protocol for consistency of behavioral test results (a test/retest reliability of ±10 dB SPL was required for inclusion in the study). Nine male and five female infants (mean age, 12 mo; SD, 3.4 mo) completed behavioral and CAEP testing.

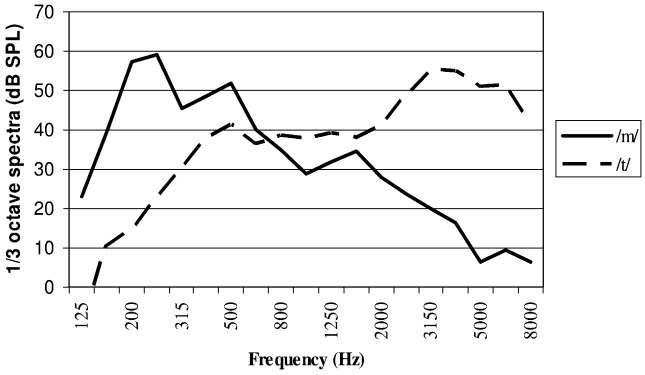

Figure 1. The spectral analysis of the two stimuli /m/ and /t/ are shown with the primary energy being below 500 Hz for /m/ and above 3000 Hz for /t/.

Procedure

The experiment and its procedures were approved by the Australian Hearing Human Research Ethics Committee.

Behavioral Assessment

All infants were assessed using visual reinforcement orientation audiometry (VROA). The /m/ and /t/ speech stimuli were delivered via a CD player, power amplifier, and Madsen OB822, two-channel clinical audiometer into the free field, which was calibrated on a daily basis. The test loudspeaker was positioned at 90° azimuth 1.2 m to the right of the test position, and stimuli were presented at a pulsed rate of 2/sec. This rate was considered sufficient to ensure that the infant's attention was maintained but not sufficiently different from that used for the cortical testing to be of any consequence. A skilled distracter was employed on all occasions to ensure quiet cooperation of the infant during testing, and a visual reward was provided via a lit puppet theater in an adjoining observation room. Response conditioning was performed at 65 dB SPL using one of the selected speech stimuli, with the order of stimulus presentation balanced among participants. Behavioral thresholds, defined as the minimum level at which two out of three responses were observed by tester

and observer, were determined. At least three nonstimulus trials were also presented for each test stimulus, to monitor for the possibility of false positive responses. A repeat session was conducted on a separate day to confirm the behavioral threshold. The degree of acceptable variability (test/retest reliability of ±10 dB SPL) was consistent with reported acceptable test-retest variations for infants aged 5 to 18 mo using operant conditioning (Widen, 1993).

CAEP Testing

CAEP testing was performed on the day of the first or second behavioral test session. The speech stimuli were presented via a loudspeaker positioned at 0° azimuth, approximately 1.8 m from the test position. These stimuli were presented with an alternating onset polarity and an interstimulus interval of 1125 msec. A daily calibration check, identical to that performed for the behavioral assessment, was undertaken to ensure consistency in the stimulus delivery across both forms of assessment. Stimuli for which reliable behavioral thresholds had been obtained were presented at 10, 20, and 30 dB relative to the behavioral threshold obtained on that day, and a nonstimulus condition was also added. There were therefore potentially four stimulus conditions and two stimuli presented to each infant participant.

The infant was seated on the caregiver's lap, or next to the caregiver during testing. A distracter, experienced in pediatric assessment, ensured quiet play and monitored the infant's state of alertness during the session. A video/audio monitor was installed to enable the examiner to pause CAEP testing when the infant became restless or vocalized. The average duration of the cortical recording session was 43 min. The cortical recording was discontinued when the distracter judged that the infant was no longer in a state suitable for reliable testing. The distracter used a silent children's DVD recording, age-appropriate toys, and/or feeding to keep the child in a relaxed and alert state for as long as possible.

During CAEP testing, brain electrical activity was recorded using the Neuroscan™ system with the active electrode positioned at Cz, referenced to the left mastoid, and forehead as ground. Electrodes were held in place by the addition of a headband or surgical tape to reduce slippage of the electrode during sudden movement. Individual sweeps of the electroencephalographic (EEG) activity were amplified and analog filtered online at 0.1–100 Hz. The recording window consisted of a 100 msec prestimulus baseline and a further 600 msec, and artifact reject was set at −150 and +150 μV. Each stimulus and stimulus condition was presented in blocks until 100 artifact-free EEG samples were acquired. If the infant remained cooperative, each block of stimuli was presented again, using a predetermined order, to form a sample consisting of two sets of 100 epochs. A review of infant cortical data from other experiments obtained at this laboratory (Golding et al, 2007) showed that an online rejection rate of <30% could be expected in most recording sessions, and this criterion was used to exclude any file sets from analysis that were overly noisy. There were 87 response samples included in the analysis; eight participants providing eight sample sets each (i.e., four stimulus conditions and two stimuli) and five participants providing four sample sets each (i.e., four stimulus conditions and one stimulus). A remaining single participant provided responses to three stimulus conditions (i.e., 10, 20, 30 dB SL) for one stimulus only. Of these 87 sample sets, 37 (i.e., 43%) contained 100 accepted epochs, and the remainder contained 200 accepted epochs.

These raw EEG files were baseline corrected using the averaged prestimulus data points, low-pass filtered at 30 Hz using a 24 dB/octave slope zero-phase filter, averaged, and then prepared for visual examination by three clinicians who had between 6 mo and 3 yr of experience in identifying infant CAEPs. Given that samples contained either 100 or 200 accepted epochs, each sample was divided so that responses derived from "odd" numbered stimuli were averaged together and those derived from "even" numbered stimuli were averaged together. Examiners were therefore presented with "odd" and "even" paired waveforms for visual detection, but this separation was not necessary for statistical analysis.

Response Detection by the Examiners

The paired waveforms were prepared for the examiners to view in series and as nonseries data sets, which are illustrated in Figures 2 and 3 respectively. The

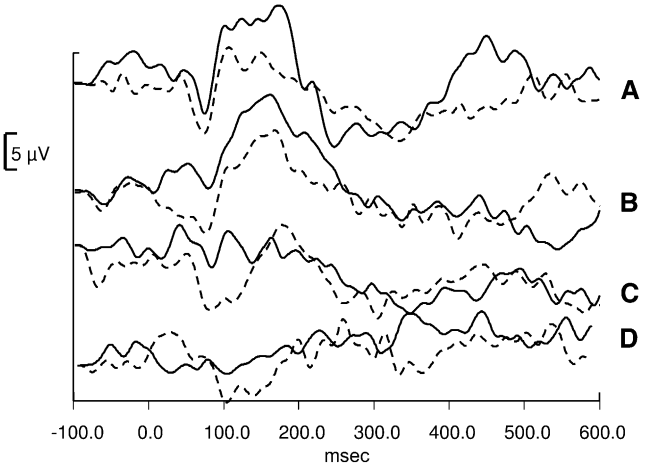

Figure 2. An in-series data set (i.e., four stimulus conditions) from one participant is shown, which was generated using one stimulus. The solid line shows the average of the even responses, and the dotted line shows the average of the odd responses.

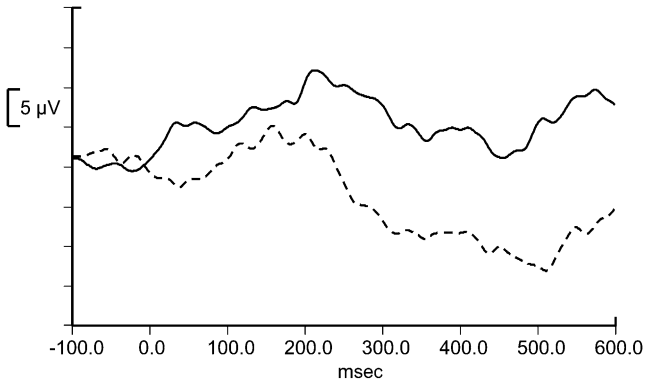

Figure 3. A nonseries paired waveform that was generated for one of four stimulus conditions is shown. The solid line shows the average of the even responses, and the dotted line shows the average of the odd responses.

examiners were informed that the in-series presentations consisted of four paired waveforms and simulated typical clinical practice where responses were presented in order of decreasing stimulus presentation level, but they were not informed of the existence of nonstimulus trials. The nonseries presentations, however, consisted of a complete randomization of paired waveforms with each pair being viewed independent of the remainder. The total set of 87 paired waveforms was presented to the examiners as a number of tasks, and each task was completed before another was issued. The tasks were: (1) nonseries, made up of 44 paired waveforms; (2) nonseries, made up of the remaining 43 paired waveforms; (3) in series, waveforms from the first seven participants only; and (4) in series, waveforms from the remaining seven participants.

Examiners were asked to study each paired waveform and determine if a cortical response to auditory stimulation could be observed or not. They were not provided with any specific criteria to use in making their judgment, but they were asked to rate their degree of certainty in making their decision using a five-point scale:

1: You are certain that a cortical response is absent.
2: It is more likely that a response is absent. but you are not certain.
3: It is equally likely that a response is present or absent.
4: It is likely that a response is present. but you are not certain.
5: You are certain that there is a cortical response present.

Statistical Method

The method used to calculate cortical detection statistically has been described previously (Golding et al, 2009). In brief, the raw EEG files that were used to create the paired waveforms for visual detection were also transferred to MATLAB™, baseline corrected and low-pass filtered at 30 Hz before undergoing statistical analysis using Hotelling's T^2. This statistic is the multidimensional equivalent of the (squared) univariate t-statistic (Flury and Riedwyl, 1988) and is well suited to situations where multiple dependent variables exist and are likely to be correlated (Tabachnick and Fidell, 2001). For statistical analysis to occur, each accepted epoched EEG file was divided into nine data-bins that cover an analysis period of 450 msec from 50 to 500 msec after stimulus onset. Within each 50 msec bin the multiple sampling points, which reflect the amplitude of the response, were reduced by averaging to form one variable per bin. Hotelling's T^2 was then applied to calculate the probability that the mean value of any linear combination of the nine created dependent variables was significantly different from zero.

RESULTS

Calculations of d′ Sensitivity Index for Examiners

For individual examiners, four cutoff criteria (>1, >2, >3, >4) were determined based on the five-point degree of certainty scale. The proportion of responses meeting each of the cutoffs was calculated for all stimulus conditions and converted to z scores. To calculate d′, the z score for the noise condition (i.e., nonstimulus condition) was subtracted from the z score for each stimulus-present condition to form a set of difference z scores that were then averaged across all cutoff criteria to form a single d′ value for each stimulus-present condition (i.e., 10, 20, 30 dB SL). By averaging the three examiners' ratings for each of the 87 paired waveforms presented in series and nonseries, d′ was found for a "composite" examiner. The calculation of d′ then followed the procedure that was described previously.

Calculation of d′ Sensitivity Index for Hotelling's T^2

For Hotelling's T^2, four cutoff criteria were selected for the p values (<0.005, <0.01, <0.02, <0.05), and the proportion of responses meeting each of these criteria was calculated for all stimulus conditions before being converted to z scores and combined to form a single d′ value for each stimulus-present condition as described above. Although a number of different cutoff criteria for the Hotelling's T^2 data could have been used, these four were consistent with commonly applied alpha levels, and they were consistent with those reported previously (Golding et al, 2009).

In calculating d′ for either examiners and/or Hotelling's T^2, some of the proportional data equaled "0"

(i.e., the proportion of responses meeting any single cutoff criteria was nil), or less commonly it equaled "1" (i.e., the proportion of responses meeting any single cutoff criteria was "1"). These extreme values cannot be converted to z scores, and therefore approximations were devised, where "0" was replaced by 0.375/R and 1 was replaced by 1–0.375/R (where R is the number of units or replications in the data). This approximation provided a workable substitute, albeit a likely underestimate of the magnitude of the true z score value (Dillon, 1984).

In Series and Nonseries

To evaluate whether the examiners' sensitivity index differed by simply presenting the data in series or without order (i.e., nonseries), at any of the stimulus presentation levels, a repeated-measures factorial ANOVA (analysis of variance) was performed using the d′ values that were generated for each examiner, with series and level as factors. Results showed that there was no significant difference in the detection sensitivity index between the mode of presentation ($F(1,2) = 0.59$, $p = 0.52$) and no interaction between mode of presentation and stimulus level ($F(2,4) = 0.707$, $p = 0.55$). A main effect of stimulus presentation level was, however, found ($F(2,4) = 35.5$, $p = 0.003$). As we predicted that a 10 dB increase in stimulus level should lead to an increase in response detectability, planned comparisons were performed and were significant for an increase in SL from 10 to 20 dB ($F(1,2) = 36.32$, $p = 0.03$) and from 20 to 30 dB ($F(1,2) = 34.13$, $p = 0.03$). Further discussion of the examiners' results is restricted to in-series data as examiners were likely to receive the greatest benefit from viewing responses in this form (although this benefit was insufficient to reach statistical significance).

Performance Outcomes for the Automated Statistic and the Examiners

Figure 4 shows d′ for Hotelling's T^2 and the "composite" examiner, for each SL. These results are shown against the possible d′ value that would be obtained for 100% accuracy in determining the presence of a response when a stimulus was present and 100% accuracy in determining the absence of a response when no stimulus was present for every cutoff criteria. Results show an increase in the sensitivity index as the SL increased for both the "composite" examiner and for Hotelling's T^2, but both fall well short of the estimated maximum d′ value, even when the stimuli were presented at 30 dB SL.

The automated statistic and individual examiner's detection sensitivity index was examined using a one-statistic summary of the receiver operating characteristic (ROC) curve. This value, known as the area under the curve, indicates the probability that a randomly chosen grade for a stimulus-present condition will exceed that of a randomly chosen grade for a nonstimulus condition (Zweig and Campbell, 1993). The area under the curve was calculated for each examiner using the five-point grading scale, and it was also calculated for Hotelling's T^2 z scores converted to five categories. To do this, the range of Hotelling's T^2 z scores was divided equally in five such that each division represented a z score range of 1.76. Category 5 on the scale was consistent with highly significant p values of <0.00001, and category 1 represented highly nonsignificant p values of >0.5. Figure 5 shows the areas under the curve for each examiner and for Hotelling's T^2, by SL. The Wilcoxon statistic was applied to test the hypothesis that examiners/Hotelling's T^2 had successfully distinguished between stimulus-present and stimulus-absent conditions (i.e., their determination was above chance) and $p < 0.05$ was achieved for all examiners/Hotelling's T^2 and all SLs. To investigate whether a performance difference between examiners/Hotelling's T^2 was likely to exist, the SE was also estimated using a nonparametric distribution model for both the stimulus-present and stimulus-absent conditions (Hanley and McNeil, 1982), and 95% confidence intervals (CIs) were calculated as shown in Figure 5.

Results show substantial overlap in the CIs associated with the different SL within and between examiners/Hotelling's T^2. With the increase in SL from 10 to 20 dB, the three examiners showed very little difference in their detection sensitivity indexes with increases in the areas under the curve of 2–5% only, whereas Hotelling's T^2 showed an area under the curve increase of 14%. At 30 dB SL, examiners 1 and 2 as well as Hotelling's T^2, showed high detection sensitivity indexes with areas under the curve greater than 90% while examiner 3 showed an improved detection sensitivity index, but the area under the curve was 82% only.

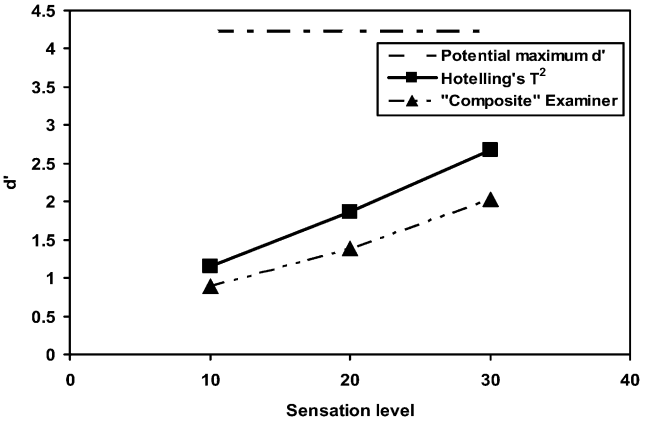

Figure 4. For the "composite" examiner and Hotelling's T^2, d′ as a function of sensation level (SL) is shown.

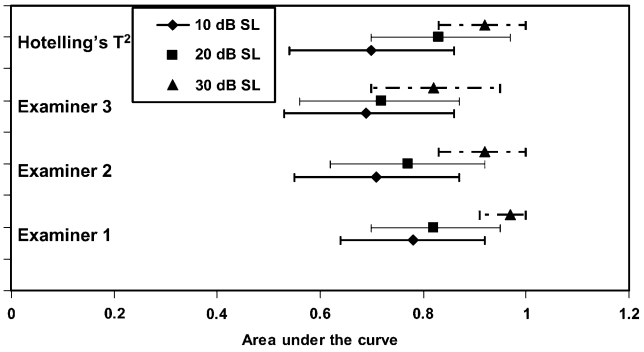

Figure 5. The area under the receiver operating characteristic (ROC) curves and 95% CI for all examiners and Hotelling's T^2 are shown.

DISCUSSION

Response replication, where two or more averaged responses are superimposed and examined for reliability, is a common aid to visual response detection (Elberling and Don, 2007). An alternative technique known as response tracking, where the response at a high stimulus level becomes a template for detection at lower stimulus levels, is also common (Elberling and Don, 2007). We hypothesized that if examiners had access to both these methods concurrently, then their detection sensitivity indexes should increase beyond that achieved using response replication alone. We therefore presented infant-generated CAEPs in two formats. First, where examiners could take advantage of response replication alone (i.e., nonseries) and, second, where examiners could use response replication combined with response tracking (i.e., in series) to make their decisions. Our results showed that examiners gained no significant benefit by the addition of response tracking, which was consistent with our findings reported previously using adult-generated cortical responses (Golding et al, 2009). This finding highlights the sensitivity of CAEPs to moment-by-moment changes to psychological and physiological state, leading to substantial individual variation in the response (Davis, 1965; Kurtzberg, 1989). It is clear then that the commonly held principle in electrophysiological testing, that responses decrease in amplitude and increase in latency as the stimulus presentation level decreases, is not reliable enough on an individual basis in CAEP testing to be a useful aid in response detection.

As the signal-to-noise ratio (SNR) increased (with increased stimulus presentation levels), the detection sensitivity index increased for the examiners as well as Hotelling's T^2. The statistical response detection technique demonstrated a slightly higher sensitivity index than the "composite" examiner for all SLs, but both fell short of the potential maximum d′ value, even with SLs of 30 dB SL. Our previous research, using adult CAEP responses (Golding et al, 2009), showed that near-perfect response detection sensitivity index could be achieved when an SL of 20 dB was applied. It is therefore plain that CAEP detection is more difficult in infant than adult populations. Our infants were well controlled in an effort to reduce electrophysiological noise but the recommended signal enhancement strategies of adequate averaging, the application of appropriate artifact reject levels, and filters (Garinis and Cone-Wesson, 2007; Picton, 1990; Thornton, 2007; Tomlin et al, 2006) appear to be inadequate for detection sensitivity that equals that achieved using adult-generated CAEPs.

Residual noise (i.e., averaged background electrophysiological noise) is typically a combination of low-level stationary as well as intermittent high-level electrophysiological activity that degrades the SNR and leads to a reduced ability to detect the evoked potential (Elberling and Don, 2007). In early latency responses, the evoked potential is relatively stable, but in late latency evoked potentials, detection can be impaired by the instability of the true evoked potential as well as residual noise. These two factors, which combine to reduce the SNR, seem to be particularly evident in infant-generated responses, and therefore methods that reduce residual noise are needed.

Our infant-averaged responses consisted of between 100 and 200 epochs. This variability arose because some infants were unable to remain quietly cooperative for an extended duration and the rate of artifact rejection was prohibitive to further testing. An adult CAEP should be detectable with 20 to 50 epochs and replication (Tomlin et al, 2006; Garinis and Cone-Wesson, 2007; Thornton, 2007), but our previous research using adult-generated data showed that the detection sensitivity index for the examiners as well as for Hotelling's T^2 increased when the number of epochs increased from 60 to 200 (Golding et al, 2009). It has been suggested that the number of epochs to detect an infant-generated CAEP is at least 50 to 100 with at least one repeat (Purdy et al, 2005). It is therefore possible that the detection sensitivity observed in this study may have been higher if all samples had consisted of 200 epochs. Theoretically it would be possible to divide the datasets into two groups (i.e., 100-epoch samples and 200-epoch samples) and recalculate d′ for smaller sample sets. However, Hotelling's T^2 and the "composite" examiner accessed the same data, and therefore performance comparisons between the two methods of detection are valid. It is also worth noting that the results obtained in this study are likely to reflect the clinical reality that collecting 200 artifact-free epochs on every infant for a range of stimuli, at a range of presentation levels, is unlikely to be feasible.

We examined the area under the ROC curve for independent examiners as a means of comparing each examiner's detection sensitivity index with the remainder and with Hotelling's T^2. Results showed very little difference

between examiners 1 and 2 in their detection sensitivity index while the third, who was the least experienced, demonstrated a lower sensitivity index at stimulus SLs of 30 dB. As all three examiners had the same level of professional qualification and none were given specific instructions in how to interpret these responses, it is reasonable to conclude that prior cumulative experience is important in interpreting infant CAEP responses without the aid of an automated statistic. Based on the results shown in Figure 5 where the Hotelling's T^2 sensitivity index was at least equal to the better performing examiners, it is also reasonable to suggest that Hotelling's T^2 can be reasonably relied upon to detect CAEPs from normal-hearing infants at SLs of 10 to 30 dB, at a rate equal to that of an examiner with prior experience.

As shown in Figure 4, Hotelling's T^2 had a sensitivity index of 2.7 when the SL was 30 dB. Our previous study, using adult data (Golding et al, 2009), showed that the same sensitivity index was achieved when the stimulus SL was approximately 14 dB (60 epoch data sets) or 5 dB (200 epoch data sets). This suggests infant CAEPs (at least, for infants without significant hearing loss), may be more difficult to detect than adult CAEPs at low SLs. It is possible however that detection sensitivity for infants was affected by measurement error in our estimates of infant SL. It has been reported that in infants aged 8 to 11 mo, intratest VROA threshold variability for individual participants is between 10 and 12 dB SPL (Primus, 1991). It is feasible, then, that the SLs used in this present study may have under- or overestimated the true SL by around 10 dB. This would naturally make detection of the CAEP to stimulation more or less difficult respectively. It is no more likely that we underestimated than overestimated the SL on average, but given the typical pattern of rate of change in amplitude with SL, reduced SLs will impact more than increased SLs on response detectability. The performance comparisons between the examiners and Hotelling's T^2 will, however, be unchanged.

Although the infants tested in this experiment did not have sensorineural hearing loss, our primary interest in recording CAEP is to evaluate the perception of speech by infants with hearing loss. Unpublished data that we have collected indicated that a stimulus of 10 dB SL evoked larger amplitude CAEP in adults with sensorineural hearing loss than does a 30 dB SL stimulus in adults with normal hearing. These results are consistent with the expected loudness growth curves for people with sensorineural hearing loss (Hellman and Meiselman, 1990). Consequently the variation in amplitude, and hence detectability of the CAEP, when hearing loss is present, is likely to be caused by recruitment that originates in the loss of the outer hair cell function in the cochlea (Lightfoot and Kennedy, 2006). If so, this phenomenon would apply equally to infants with sensorineural hearing loss, and hence the amplitude and detectability of the CAEP would be considerably greater than we have reported in this paper, where the infants did not have sensorineural hearing loss.

The group of infants tested is relatively small. Each infant, however, effectively acted as his or her own control, because the same set of waveforms was submitted to the automated and human examiners. The main question of interest was the relative sensitivity of response detection by an automated statistic versus detection by expert examiners, rather than an attempt to define the characteristics of CAEP waveforms in the infant population.

To date, it has been common practice to detect CAEPs using visual examination that relies on response replication and response tracking. Our results suggest that (1) the examiner's cumulative experience is likely to be an important determinant of detection sensitivity, and (2) the Hotelling's T^2 statistic is able to detect normal-hearing infant CAEPs at stimulus SLs of 10 to 30 dB with a detection sensitivity index equal to that of the more experienced examiners. It remains to be seen how much benefit the less and more experienced examiners might derive from combined access to visual and statistical CAEP detection techniques.

CONCLUSIONS

Automated statistical detection of cortical responses from infants, based on the Hotelling's T^2 statistic, was at least as accurate as detection based on the average of three expert examiners, two of whom were very experienced in interpreting cortical responses from infants, and one of whom had some experience.

For the infants tested, none of whom had sensorineural hearing loss, a sensation level of 30 dB led to a d' of 2.7. For a d' of 2.7, a detection power (i.e., ability to detect a cortical response when the speech sound was audible) of 0.85 can be achieved with an alpha level (i.e., false alarm rate) of 0.05.

To date, there has been limited clinical application of CAEP recording for the infant population. This may in part be due to the inherent difficulties, particularly for novice observers, in interpreting CAEP waveforms. Providing an automatic evaluation of cortical responses appears to have promise for increasing the clinical utility of CAEP testing. The inclusion of automatic response detection in clinical equipment may increase clinicians' confidence in using the CAEP technique, save time in interpreting results, and thus make this form of hearing aid fitting evaluation more accessible to the pediatric population.

REFERENCES

Bagatto M, Moodie S, Scollie S, Seewald R, Pumford J, Liu KPR. (2005) Clinical protocols for hearing instrument fitting in the desired sensation level method. *Trends Amplif* 9:199–226.

Bess FH, Humes LE. (2003) *Audiology: The Fundamentals*. 3rd ed. Philadelphia: Lippincott Williams and Wilkins.

Brown E, Klein AJ, Snydee KA. (1999) Hearing aid processed tone pips: electroacoustic and ABR characteristics. *J Am Acad Audiol* 10:190–197.

Ching TYC, Hill M. (2007) The Parent's Evaluation of Aural/Oral Performance of Children (PEACH) Scale: normative data. *J Am Acad Audiol* 18:220–235.

Cone-Wesson B, Wunderlich J. (2003) Auditory evoked potentials from the cortex: audiology applications. *Curr Opin Otolaryngol Head Neck Surg* 11:372–377.

Cunningham J, Nicol T, Zecker S, Kraus N. (2000) Speech-evoked neurophysiological responses in children with learning problems: development and behavioural correlates of perception. *Ear Hear* 21:554–568.

Davis H. (1965) Slow cortical responses evoked by acoustic stimuli. *Acta Otolaryngol* 59:179–185.

Dillon H. (1984) *A Procedure for Subjective Quality Rating of Hearing Aids. Report No. 100*. Canberra, Australia: Australian Government Publishing Service.

Dillon H. (2001) *Hearing Aids*. New York: Thieme.

Eggermont JJ, Ponton CW. (2003) Auditory-evoked potential studies of cortical maturation in normal hearing and implanted children: correlations with changes in structure and speech perception. *Acta Otolaryngol* 123:249–252.

Elberling C, Don M. (2007) Detecting and assessing synchronous neural activity in the temporal domain (SNR, response detection). In: Burkard RF, Don M, Eggermont JJ, eds. *Auditory Evoked Potentials*. Baltimore, MD: Lippincott, Williams and Wilkins, 102–123.

Flury B, Riedwyl H. (1988) *Multivariate Statistics: A Practical Approach*. London: Chapman and Hall.

Garinis AC, Cone-Wesson BK. (2007) Effects of stimulus level on cortical auditory event-related potentials evoked by speech. *J Am Acad Audiol* 18:107–116.

Golding M, Dillon H, Seymour J, Carter L. (2009) The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. *Int J Audiol* 48(12):833–842.

Golding M, Pearce W, Seymour J, Cooper A, Ching TYC, Dillon H. (2007) The relationship between obligatory cortical auditory evoked potentials (CAEPs) and functional measures in young infants. *J Am Acad Audiol* 18:117–125.

Gravel JS, Kurtsberg D, Stapells DR, Vaughan HG, Wallace IF. (1989) Case studies. *Semin Hear* 10:272–287.

Hall JW. (1992) *Handbook of Auditory Evoked Responses*. Boston: Allyn and Bacon.

Hanley JA, McNeil BJ. (1982) The meaning and use of the area under a receiver operating characteristic curve. *Radiology* 143: 29–36.

Hellman RP, Meiselman CH. (1990) Loudness relations for individuals and groups in normal and impaired hearing. *J Acoust Soc Am* 88:2596–2606.

Hyde M. (1994) The slow vertex potential: properties and clinical applications. In: Jacobson JT, ed. *Principles and Applications in Auditory Evoked Potentials*. Needham Heights: Allyn and Bacon, 179–218.

Hyde M. (1997) The N1 response and its applications. *Audiol Neurootol* 2:281–307.

Kileny P. (1982) Auditory brainstem responses as indicators of hearing aid performance. *Ann Otol Rhinol Laryngol* 91:61–64.

Korczak PA, Kurtsberg D, Stapells DR. (2005) Effects of sensorineural hearing loss and personal hearing aids on cortical event-related potential and behavioural measures of speech-sound processing. *Ear Hear* 26:165–185.

Kraus N, McGee TJ, Koch DB. (1998) Speech sound representation, perception, and plasticity: a neurophysiologic perspective. *Audiol Neurootol* 3:168–182.

Kurtzberg D. (1989) Cortical event-related potential assessment of auditory system function. *Semin Hear* 10:252–262.

Kurtzberg D, Stapells DR, Wallace IF. (1988) Event-related potential assessment of auditory system integrity: implications for language development. In: Vietze P, Vaughan HGJ, eds. *Early Identification of Infants with Developmental Disabilities*. Philadelphia: Grune and Stratton, 160–180.

Kushnerenko E, Ceponiene R, Balan P, Fellman V, Huotilainen M, Naatanen R. (2002) Maturation of the auditory event-related potentials during the first year of life. *Cog Neurosci Neuropsychol* 13:47–51.

Lightfoot G, Kennedy V. (2006) Cortical electric response audiometry hearing threshold estimation: accuracy, speed, and the effects of stimulus presentation features. *Ear Hear* 27:443–456.

Martin FN, Clark JG. (2003) *Introduction to Audiology*. 8th ed. Boston: Allyn and Bacon.

McNicol D. (1972) *A Primer of Signal Detection Theory*. Sydney: Australasian Publishing Company.

Oates PA, Kurtzberg D, Stapells DR. (2002) Effects of sensorineural hearing loss on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear* 23:399–415.

Pasman JW, Rotteveel JJ, Maassen B, Visco YM. (1999) The maturation of auditory cortical evoked responses between (preterm) birth and 14 years of age. *Eur J Paediatr Neurol* 3:79–82.

Pearce W, Golding M, Dillon H. (2007) Cortical auditory evoked potentials in the assessment of auditory neuropathy. *J Am Acad Audiol* 18(5):380–389.

Picton T. (1990) Auditory evoked potentials. In: Daly D, Pedley T, eds. *Current Practice of Clinical Electroencephalography*. 2nd ed. Philadelphia: Lippincott Williams and Wilkins, 625–628.

Picton T, Durieux-Smith A, Champagne SC, et al. (1998) Objective evaluation of aided thresholds using auditory steady-state responses. *J Am Acad Audiol* 9:315–331.

Primus MA. (1991) Repeated infant thresholds in operant and non operant audiometric procedures. *Ear Hear* 12:119–122.

Purdy SC, Kelly AS. (2001) Cortical auditory evoked potential testing in infants and young children. *N Z Audiol Soc Bull* 11(3):16–24.

Purdy SC, Katsch R, Dillon H, Storey L, Sharma M, Agung K. (2005) Aided cortical auditory evoked potentials for hearing instrument evaluation in infants. In: Seewald R, ed. *A Sound Foundation through Early Amplification*. Warrenville, IL: Phonak.

Rance G, Beer DE, Cone-Wesson B, et al. (1999) Clinical findings for a group of infants and young children with auditory neuropathy. *Ear Hear* 20:238–252.

Rance G, Cone-Wesson B, Wunderlich J, Dowell R. (2002) Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear Hear* 23:239–253.

Rapin I, Granziani LJ. (1967) Auditory-evoked responses in normal, brain-damaged, and deaf infants. *Neurol (Tokyo)* 17:881–894.

Rotteveel JJ, Colon EJ, Notermans LH, Stoelinga GBA, de Graaf R, Visco YM. (1986) The central auditory conduction at term date and three months after birth. IV. Auditory cortical responses. *Scand Audiol* 15:85–95.

Sharma A, Dorman MF, Spahr AJ. (2002) A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear Hear* 23:532–539.

Sininger YS. (2002) Identification of auditory neuropathy in infants and children. *Semin Hear* 23:193–200.

Sirimanna KS. (2001) Management of the hearing impaired infant. *Semin Neonatol* 6:511–519.

Stapells DR, Kurtzberg D. (1991) Evoked potential assessment of auditory system integrity in infants. *Clin Perinatol* 18:497–518.

Steinschneider M, Kurtzberg D, Vaughan HG. (1992) Event-related potentials in developmental neuropsychology. In: Rapin I, Segalowitz SJ, eds. *Child Neuropsychology.* Volume 6 of *Handbook of Neuropsychology*. Amsterdam: Elsevier Science Publishers, 239–299.

Stelmachowicz PG, Seewald RC. (1991) Probe-tube microphone measures in children. *Semin Hear* 12:62–72.

Tabachnick BG, Fidell LS. (2001) *Using Multivariate Statistics*. 4th ed. Boston: Allyn and Bacon.

Thornton ARD. (2007) Instrumentation and recording parameters. In: Burkard RF, Don M, Eggermont JJ, eds. *Auditory Evoked Potentials*. Baltimore: Lippincott Williams and Wilkins, 73–101.

Tomlin D, Rance G, Graydon K, Tsialios I. (2006) A comparison of 40 Hz auditory steady-state response (ASSR) and cortical auditory evoked potential (CAEP) thresholds in awake adult subjects. *Int J Audiol* 45:580–588.

Weinstein B. (2000) *Geriatric Audiology*. 1st ed. New York: Thieme.

Widen JE. (1993) Adding objectivity to infant behavioural audiometry. *Ear Hear* 14:49–57.

Wunderlich J, Cone-Wesson B. (2006) Maturation of CAEP in infants and children: a review. *Hear Res* 212:212–223.

Zweig MH, Campbell G. (1993) Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine. *Clin Chem* 39:561–577.