

# Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds

Bram Van Dun,<sup>1,2</sup> Lyndal Carter,<sup>1,2</sup> Harvey Dillon<sup>1,2</sup>

<sup>1</sup>The HEARing CRC, Melbourne; <sup>2</sup>The National Acoustic Laboratories, Sydney, Australia

## Abstract

Cortical auditory evoked potentials (CAEPs) are an emerging tool for hearing aid fitting evaluation in young children who cannot provide reliable behavioral feedback. It is therefore useful to determine the relationship between the sensation level of speech sounds and the detection sensitivity of CAEPs, which is the ratio between the number of detections and the sum of detections and non-detections. Twenty-five sensorineurally hearing impaired infants with an age range of 8 to 30 months were tested once, 18 aided and 7 unaided. First, behavioral thresholds of speech stimuli /m/, /g/, and /t/ were determined using visual reinforcement orientation audiometry. Afterwards, the same speech stimuli were presented at 55, 65, and 75 dB sound pressure level, and CAEPs were recorded. An automatic statistical detection paradigm was used for CAEP detection. For sensation levels above 0, 10, and 20 dB respectively, detection sensitivities were equal to 72±10, 75±10, and 78±12%. In 79% of the cases, automatic detection P-values

became smaller when the sensation level was increased by 10 dB.

The results of this study suggest that the presence or absence of CAEPs can provide some indication of the audibility of a speech sound for infants with sensorineural hearing loss. The detection of a CAEP might provide confidence, to a degree commensurate with the detection probability, that the infant is detecting that sound at the level presented. When testing infants where the audibility of speech sounds has not been established behaviorally, the lack of a cortical response indicates the possibility, but by no means a certainty, that the sensation level is 10 dB or less.

## Introduction

Since the introduction of universal hearing screening, early intervention programs have progressed the aim for optimum speech and language development in congenitally hearing impaired infants.<sup>1,2</sup> When hearing loss is detected at a very young age, hearing threshold estimations through auditory brainstem responses (ABRs) or auditory steady-state responses (ASSRs) allow audiologists to fit hearing aids relatively confidently, even with young children.<sup>3,4</sup> Unfortunately, there can be inaccuracies or even errors in either estimation of threshold or adjustment of the hearing aid, and some means of evaluating the adequacy of the fitting is required, just as with adult hearing aid wearers. Furthermore, for children with auditory neuropathy spectrum disorder, thresholds based on ABR responses bear no relationship to behavioral thresholds as absent or highly elevated ABR responses are part of the defining characteristics of this condition.<sup>5-7</sup> Cortical auditory evoked potentials (CAEPs), which have generators at a higher level on the auditory pathway than ABRs, are more indicative of whether neural signals are reaching parts of the auditory cortex and thus should be more closely related to perception of sound. Consequently, they are said to be more appropriate for speech and language development assessment<sup>8</sup> and have been shown to be related to speech perception scores and functional measures of hearing ability.<sup>5-7,9</sup>

CAEP morphology is dependent on age,<sup>10-12</sup> sleep state,<sup>13,14</sup> attention,<sup>15,16</sup> stimulus,<sup>17-21</sup> presentation parameters,<sup>22-24</sup> and electrode recording position.<sup>25,26</sup> In awake and alert children up to the age of about six years, a reliable CAEP recorded from the vertex (relative to one of the mastoids) at a rate of about one a second generally consists of a positive peak ranging from about 250 ms (at birth) to 100 ms (in childhood), followed by a low-amplitude negative deflection ranging from 450-600 ms (at birth) to 200 ms (in childhood). The latency decrease is explained by the development of the auditory system over time,<sup>11</sup> and is also dependent on the duration a person has been subjected to sound, the so-called *time-in-sound*.<sup>27,28</sup> From around the eighth year of life, the appearance of an extra negative deflection N1 separates the positive deflection into peaks P1 and P2. This transformation continues until adulthood, where the CAEP has a distinct P1-N1-P2-N2 pattern.<sup>26</sup>

Correspondence: Bram Van Dun, National Acoustic Laboratories, 126 Greville St, Chatswood NSW 2067, Australia.  
Tel: +61.294126967 – Fax: +61.294118273.  
E-mail: bram.vandun@nal.gov.au

Key words: cortical auditory evoked potential, sensitivity, hearing aid, infant.

Acknowledgements: the authors would like to thank both reviewers for their indispensable contributions and suggestions, the Australian Hearing centers, audiologists Loretta Brown, the late Christine Whitehead, Donna-Mae Schwarz, and Katrina Freeston for their assistance with data collection, Barry Clinch for his technical support, and Mark Seeto for his statistical guidance. The authors acknowledge the financial support of the HEARing CRC, established and supported under the Cooperative Research Centers Program - an initiative of the Australian Government.

Contributions: BVD, independent data collection, statistical analysis; BVD, LC, HD, conceptual idea and overall design, interpretation of results, writing of final manuscript, detailed experimental design; BVD, HD, leadership.

Conflict of interests: the authors report no conflict of interests.

Received for publication: 21 February 2012.

Revision received: 9 July 2012.

Accepted for publication: 1 August 2012.

This work is licensed under a Creative Commons Attribution NonCommercial 3.0 License (CC BY-NC 3.0).

©Copyright B. Van Dun et al., 2012  
Licensee PAGEPress, Italy  
Audiology Research 2012;2:e13  
doi:10.4081/audiore.2012.e13

CAEPs seem well suited to hearing aid fitting evaluation for several reasons. First, it is possible to use speech sounds, which are the sounds whose audibility we are most interested in knowing, rather than the brief tones or clicks that are needed for ABRs, or the modulated continuous sounds used for ASSRs. While it is possible to estimate the audibility of speech sounds from knowledge of hearing aid gain plus pure-tone behavioral thresholds, or tone-burst electrophysiological thresholds, many assumptions about auditory filter widths, detection efficiency, and temporal integration must be made. Second, the longer duration of speech sounds allows the hearing aid more time to react to the presented sound, so that the hearing aid is more likely to be in a state similar to the state it is in for a real-life speech signal. Third, CAEPs represent the detection of sound at, or near the end of, the auditory pathway, so they are affected by all parts of the auditory system as well as by the hearing aid gain-frequency response.<sup>29,30</sup>

One of the probable reasons for the lack of clinical take-up of CAEPs for hearing aid fitting evaluation is the difficulty in assessing the presence of a cortical response, given the variations in morphology that occur with age and with the state (alert, drowsy, sleeping) of the child.<sup>13,14</sup> This problem has been reduced with the availability of reliable automated methods of detecting cortical responses. Objective CAEP response detection - at least the one used in Carter *et al.* and Golding *et al.*<sup>31,32</sup> - does not rely on a *template* derived from an average wave form obtained from a large subject group. This is in contrast with the subjective interpretation by an observer, who generally does rely on similarities between a template and an individual's waveform for identification.

As one of the first investigators of CAEPs in hearing aid wearers, Rapin *et al.*<sup>33</sup> evaluated CAEPs in eight hearing-impaired infants presented with clicks and tone-bursts in the free field. The aided CAEPs of six out of eight children were larger or clearer than the unaided responses. The authors concluded that the use of CAEPs showed promise in assessing clinically whether a hearing aid was beneficial or not. In a similar experiment, Korczak *et al.*<sup>34</sup> showed that, on a group basis, the cortical responses of hearing impaired adults to /ba/ and /da/ speech stimuli displayed shorter latencies, larger amplitudes, better CAEP waveform morphologies, improved CAEP detectability, and increased behavioral performance when aided than when unaided. However, the amount of improvement between unaided and aided was variable among individual subjects.

Tremblay *et al.*<sup>35</sup> reported that the neural detection of time-varying acoustic cues in speech can be recorded in adult hearing aid users using the acoustic change complex. Two sounds *see* and *shee* were observed to evoke different neural responses. The consonant-vowel (CV) transition of *shee* still elicited an earlier negative peak with a predicted latency difference when compared with the transition from *see*, despite speech signals being altered by the hearing aid and delivered to a deprived auditory system. This CV transition preservation has been confirmed in another study with same speech sounds<sup>36</sup> for normally hearing adults, thereby separating the effect of amplification from hearing loss. The authors reported that cortical potentials could be reliably recorded, with a subtle enhancement of aided peak amplitudes when compared with unaided recordings.

The results of two studies<sup>30,37</sup> cast doubt over the validity of measuring the cortical response of a speech sound after it has been amplified by a hearing aid. They have shown that cortical responses measured with a hearing aid providing 20 dB of gain do not have significantly larger amplitudes than cortical responses measured while the participants were unaided. This result contrasts with the usual finding of increased amplitude with increased stimulus level, at least for low and moderate sensation levels.<sup>38</sup> There is one main reason why - in our opinion - this doubt is debatable. In both of these studies, the participants were normally hearing adults. Because the internal noise of a

hearing aid is audible to people with normal hearing, amplification by a hearing aid, while increasing the stimulus intensity, actually decreases the stimulus sensation level. Decreasing sensation level by adding noise reduces the magnitude of cortical responses.<sup>21</sup> Consequently, the two effects of amplification (increasing stimulus amplitude and decreasing the sensation level) have opposing effects on cortical response amplitude and it is unclear how amplification should then change the magnitude of the cortical response. These offsetting effects do not occur for hearing-impaired hearing aid wearers, provided their hearing thresholds exceed the equivalent input noise level of the hearing aid, which is the case for all but the mildest of hearing losses.

When recording cortical responses in infants in order to determine whether a hearing loss is present, or to evaluate hearing aid settings, it is important being able to assess the implications of the presence or absence of a cortical response. In addition, when an objective statistical method is used, the availability of a quantifying detection measure (like a P-value) will provide a wider spectrum of answers than the present/not-present pair. This paper aims to determine the relationship between the audibility of sounds at low sensation levels in individual infants and the detectability of the cortical responses they evoke. It differs from previous research in that the stimuli are presented at low sensation levels. In common with early studies of cortical responses and amplification, the participants are hearing-impaired infants. Unlike early studies, the detectability of a cortical response is quantified with a probability-based metric.

The paper first describes basic analyses of CAEP amplitudes and latencies for the different speech sounds presented at different sensation levels, and the electroencephalographic (EEG) noise characteristics. This provides a reference for comparison with future studies. Sensitivity, which is the ratio between the number of detections and the sum of detections and non-detections, is reported for the three speech sounds at different sensation levels. Finally, the practical use for the clinician when recording CAEPs for hearing aid fitting evaluation is discussed.

---

## Materials and Methods

### Subjects

Twenty-five infants with sensorineural hearing impairment were tested once, 18 aided and 7 unaided. Fourteen females and 11 males completed the study, with an age range of 8 to 30 months, mean 19 months (SD 8 months). Shows most recent audiograms and hearing aids used. The mean three frequency average (1, 2, and 4 kHz) of the participants was 56 dB HL (SD 11 dB), obtained from their most recent audiogram. Age of first hearing aid fitting was at 5.5 months (SD 4.7 months). Length of hearing aid use was 13.4 months (SD 6.7 months). The inclusion criteria for participation were as follows. Infants were 30 months or younger and developmentally ready for behavioral testing. The maximum three-frequency average (1, 2, and 4 kHz) hearing level (3FAHL) allowed for the better ear was 90 dB HL. The minimum sensorineural hearing loss was 25 dB HL. If a complete audiogram was not available, the two-frequency average (1 and 4 kHz) was accepted. Infants diagnosed with auditory neuropathy were excluded.

The number of infants tested in two pediatric hearing centers (which are part of the Australian Hearing network) was 12 and 13 infants respectively. The audiologists conducting the test underwent training in CAEP testing and equipment use. Their electrophysiological experience was limited to pediatric reports on ABR or CAEP results from tests conducted in other clinics. They did, however, have significant prior experience with pediatric behavioral testing, hearing aid fitting, and evaluation. The study was approved by the Australian Hearing Human Research Ethics Committee.

## Stimuli

Presented speech stimuli were /m/ (duration of 30 ms), /g/ (duration of 21 ms), and /t/ (duration of 30 ms). They have been used in other studies from the authors' research group,<sup>31,32</sup> and have also been incorporated in the HEARLab system in identical form. The stimuli were extracted from a recording of uninterrupted dialogue spoken by a female with an average Australian accent, with a sampling rate of 44.1 kHz. Very little vowel transition was included in the stimuli. Their length is sufficiently long to be processed by a hearing aid, and acceptably short to generate a proper CAEP. An additional high-pass filter of 250 Hz was applied to /t/ to remove low-frequency noise. These three essentially vowel-free speech sounds /m/, /g/, and /t/ have a spectral emphasis in the low-, mid-, and high-frequency regions, as shown in Figure 1, and thus have the potential of providing information about the audibility of speech sounds in different frequency regions. Cortical detection of these speech stimuli does not necessarily signify detection of *frequency cues*, but rather (mainly) a detection of transients or onsets.<sup>39</sup> However, if due to a hearing loss in the cochlea a speech sound has not been transmitted to the auditory cortex, an onset response also cannot be generated. As a result an onset response originating from the auditory cortex will also provide information about which frequency information has been transmitted by the cochlea.

## Procedure

Parents or guardians of participants gave informed consent. Prior to the assessment, hearing aids and ear molds were visually inspected, hearing aid batteries were replaced, and correct function of the hearing aid was confirmed by graphing using a HA2 2cc-coupler. The fit of the ear mold was checked with regards to possible feedback. Otoscopy and tympanometry were performed on both ears to exclude the presence of excessive cerumen and middle ear dysfunction respectively.

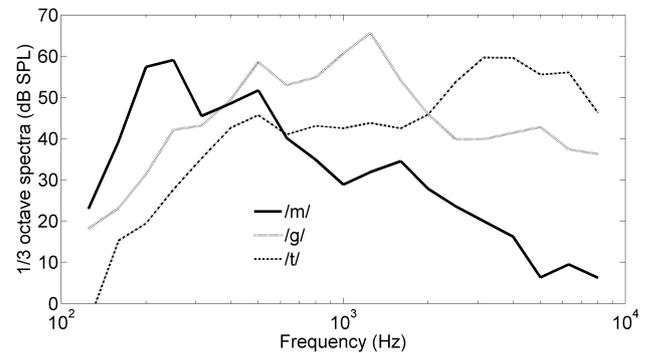
The tests were conducted either aided or unaided, depending on the degree of hearing loss of the infants. Infants with a 3FAHL (or three-frequency average) of  $\geq 50$  dB HL were assigned to the *aided condition* group. Infants with better hearing than this criterion were assigned to the *unaided condition* group. This decision was based on hearing thresholds of the better ear (given that testing was performed in free field). In the aided group, testing was performed binaurally, with the hearing aids set according to the participant's usual prescribed settings and NAL-NL1 targets. This way hearing aid settings reflected everyday use. There was an exception for the test group assignment for one participant (U4), who was originally assigned to *aided condition* group (3FAHL=62 dB HL), but as the child's hearing aids were not available at the time of testing, was evaluated unaided instead.

Given the age of the participants, the number of reliable behavioral thresholds that could reasonably be obtained in one session was likely to be limited. Therefore only two out of the three speech stimuli /m/, /g/, and /t/ were tested for each individual participant. Stimulus selection was based on a predefined counterbalanced order.

The experimental session had two assessment components occurring on the same day. Hearing aids were set identically for both assessments.

## Behavioral assessment

The behavioral thresholds for speech sounds were measured using Visual Reinforcement Orientation Audiometry (VROA) in the free field, in a sound attenuated test booth with adjacent observation room.<sup>40</sup> The speech sounds were delivered via a CD player, a power amplifier, and a Madsen OB 822 audiometer. Stimulus intensity was calibrated prior to each test. The test position was approximately 1, and 1.8 meters, from the loudspeaker in the two hearing centers respectively, with the loudspeaker positioned at an angle of 90° to the right of the test position. Stimuli were presented at a rate of 4 times per s, with presentation



**Figure 1.** The 1/3rd-octave power spectra for speech sounds /m/, /g/, and /t/ with overall levels normalized to 65 dB sound pressure level (SPL).

duration of approximately 3 s. This rate was considered sufficient to ensure that the infant's attention was maintained but not sufficiently different from the rate used for the cortical testing to affect audibility. Two clinicians were required for VROA testing. One clinician presented the stimuli from the observation room, judged the child's responses, and provided a puppet reward (in a lighted *puppet theatre* box). The second, blinded, clinician distracted the child, indicated when to apply a (non)stimulus, and also judged the child's responses. Blinding was obtained by the observer having masker noise applied through headphones.

The following procedure was used for each speech sound assigned to the child for assessment. First, a conditioned response was obtained, the initial presentation level being appropriate to the degree of hearing loss (and at the discretion of the experimenter). If the child could not be conditioned, he/she was excluded from the study. The maximum and minimum presentation levels were 85 dB sound pressure level (SPL) and 35 dB SPL respectively. After conditioning, the stimulus was presented at the minimum presentation level (35 dB SPL). If responses were observed to two out of three stimulus trials, and zero out of three responses to non-stimulus trials (which were presented randomly one out of three times), the procedure was stopped for this particular speech sound. The child's threshold was then assigned a value of 35 dB SPL for the purposes of data analysis. If a response was observed by one or both observers to a non-stimulus trial, the child's level of distraction was addressed and the full set of stimulus and non-stimulus trials were repeated. If either observer judged that no response was observed at 35 dB SPL, the stimulus intensity was increased to 65 dB SPL, as adapted from Birtles.<sup>40</sup> Then, by modifying stimulus intensity with steps of 5 or 10 dB, the child's behavioral threshold was determined as the lowest intensity that complied with the response evaluation paradigm described above. If VROA thresholds were found to be above 85 dB SPL, the child was not excluded from the study, but this resulted in subsequent CAEP recordings being performed at a negative sensation level.

## Cortical auditory evoked potential assessment

CAEP testing was performed using the HEARLab® system (Frye Electronics, Tigard, OR, USA). Stimulus calibration was performed at 75 dB SPL prior to recording, using a free field equalization and calibration method that pre-filters the stimuli with a gain-frequency response equal to the inverse of the transfer function measured from the input to the loudspeaker to the sound field at the test position, with participant absent

The session was scheduled to maximize the child's alertness. She/he was seated on the caregiver's lap, again, with his/her head approxi-

mately 1.8 or 1 meter from the loudspeaker (in the two hearing centers respectively), with the loudspeaker positioned at 0° azimuth. The child was encouraged to sit quietly in the test position using distractions such as quiet, age-appropriate toys and silent movies. The audiologist providing the distraction also observed the child's state as the test progressed, to ensure he or she remained awake and alert and that the electrodes remained in place. Stimuli were presented with a fixed inter-stimulus interval of 1125 ms (offset to onset). The two randomly assigned stimuli were interleaved automatically in blocks of 25 presentations per stimulus.

Electrode sites were prepared using a cotton applicator and electrode gel. Single use Ambu Blue Sensor N™ self-adhesive electrodes were used, placed at Cz (active), M1 (reference) and Fpz (ground).<sup>41</sup> Electrode impedance was checked, and if necessary the preparation was repeated to achieve an impedance less than 5 kohms between active and ground, and between reference and ground.

During recording, the EEG activity was amplified in two stages. Firstly at the coupling to the scalp electrodes (x121) and secondly (x10) after the signal was transported through the electrode cables. The signal was down-sampled to 1 kHz and filtered online at 0.16-30 Hz. The recording window consisted of a 200 ms pre-stimulus baseline and a further 600 ms duration. These values are limits of the recording system. Artifact rejection was set at  $\pm 110 \mu\text{V}$ . No eye blink detection channel was used, as this was not considered clinically practical.

The HEARLab system uses an automatic statistical detection procedure which does not require a subjective response interpretation from the operator.<sup>31,32</sup> This system-generated significance level (P-value) was used to determine whether or not a response was present, and to determine the end of the test run. For the calculation of the detection P-value, a Hotelling's T-squared statistic was applied to a (M x Q)-dataset, with M collected epochs in each recording, and Q bin averages. These Q=9 bin averages were taken as the means of 9 data bins ranging from 101 to 550 ms, each bin 50 ms wide. This bin width and number of bins was chosen based on earlier data, with the aim of optimizing the trade-off between bin widths being sufficiently narrow (so they do not encompass both a negative and a positive component or else their effect will cancel each other), and a reduced number of bins (as test sensitivity decreases when this number increases due to a greater opportunity for chance to affect the outcome).<sup>32</sup> Using this statistic, both waveform repeatability over all recorded epochs, and significant difference from zero could be objectively assessed.

Testing at a given intensity level was concluded immediately if HEARLab indicated that the P-value for both stimuli being tested at that level was  $P \leq 0.001$ , regardless of the number of accepted epochs. Testing was otherwise concluded after approximately 200 accepted epochs for both stimuli. A CAEP response was judged to be present if the P-value reached the level of  $P < 0.05$ . Two hundred accepted epochs were chosen as a higher number of epochs makes data collection with this age group clinically not feasible in a restricted measurement time.

Speech stimuli were initially presented at 65 dB SPL. If both CAEP responses were judged to be present, the second test run was performed at 55 dB SPL. Otherwise, the second test run was performed at 75 dB SPL. If the infant was still in suitable state for testing, a third test run was performed (either at 75 or 55 dB SPL to complete a set of three runs at different intensities).

## Data analysis

Data obtained from three children were excluded from the analysis. Given the participant age group, this was considered a reasonable success rate in obtaining both VROA thresholds and CAEP recordings. For one participant (U7) tested unaided, and one participant (A3) tested aided, the child did not remain in a suitable state for CAEP testing. For a second participant (A16) tested aided, VROA testing could not be

completed as the child was not able to be adequately conditioned, even at high intensities (85 dB SPL), which is a comparable percentage (1/25=4%) with Birtles<sup>40</sup> who reported 7% of children could not be conditioned using VROA.

Baseline correction was applied based on the first 200 ms of the EEG recording window. The averaged response was digitally filtered with a low-pass filter at 30 Hz before plotting grand averages. To smooth the waveforms for determination of the peak amplitude and latency (*peak picking*), a low pass filter at 15 Hz was applied. The infant P1 (positive) response was identified as the most positive point of the waveform in the latency range 100 to 300 ms. Similarly, the negative infant N2 response was defined in the latency range 200 to 600 ms, as the most negative point following the positive P1 response. P1 and N2 amplitudes were defined from baseline to peak. The EEG noise power was estimated as follows. For each sampled point in the epoch, the mean value of that point, across epochs, and the variance around that mean, were calculated. These variances were then averaged across all sampled points within the epoch and the square root of the average taken to produce an estimate of the root mean square (rms) noise present during that run.

## Results

Data of twenty-two infants were recorded, 6 children in the *unaided condition* group and 16 children in the *aided condition* group (10 male, 12 female). A total of 92 recordings with a single speech sound at a single intensity were available (24 unaided, 68 aided), further described as data points. For the presentation levels (55, 65, and 75 dB SPL), 29, 43, and 20 data points were collected respectively, and for /m/, /g/, and /t/ speech sounds, 30, 30, and 32 data points. Figure 2 shows a histogram of behavioral thresholds as a function of the three speech sounds. Participant A14, tested with two /g/ and two /t/ speech sounds,

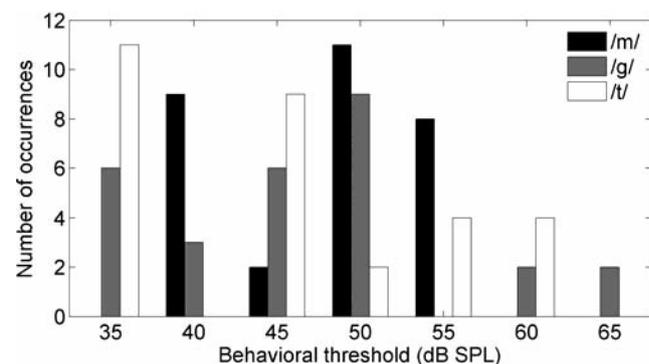


Figure 2. Distribution of 64 aided and 24 unaided behavioral thresholds (total of 88) as a function of speech sound (30 /m/, 28 /g/, 30 /t/) from 21 subjects [(one aided subject did not return behavioral thresholds equal to or less than 85 dB sound pressure level (SPL)]. It was aimed to have a range of sensation levels that varied just below to well above 0 dB sensation level (SL). The main concern was not to have too many speech sounds presented below hearing threshold (as will be the case when testing unaided severe and profound hearing losses). As speech stimuli were applied at 55, 65, 75 dB SPL, the sensation level range hence varied between -5 and 40 dB SL.

did not return behavioral thresholds equal to or less than 85 dB SPL. This was in contrast with the participant's most recent audiogram in Table 1, and the deteriorated hearing was explained by a cytomegalovirus infection during pregnancy. All presented stimuli hence were subthreshold and inaudible to this participant.

### Grand averages and regression analysis of amplitudes and latencies

The grand averages of responses to stimuli presented at four sensation level ranges are shown in Figure 3. Sensation level is defined as the difference between CAEP presentation level and the VROA threshold. Figure 3 shows that cortical response amplitudes increase with sensation level. Group means and SDs for P1 and N2 amplitudes and latencies are provided in Table 2 for each of five sensation level ranges

separately. Regression analysis on all data points found a weak, but highly significant, positive relationship between sensation level and P1 amplitude ( $r=0.41$ ) [ $r^2=0.17$ ,  $F(1.90)=18.6$ ,  $P=0.00004$ , intercept= $2.79 \mu V$ , slope= $0.09 \mu V/dB$ ]. Similarly, a weak significant negative relationship was found for sensation level versus N2 amplitude ( $r=-0.26$ ) [ $r^2=0.07$ ,  $F(1.90)=6.5$ ,  $P=0.0123$ , intercept= $-3.46 \mu V$ , slope= $-0.08 \mu V/dB$ ]. For sensation level versus P1 latency, a comparable relationship was identified ( $r=-0.32$ ) [ $r^2=0.10$ ,  $F(1.90)=10.5$ ,  $P=0.0017$ , intercept= $173 \text{ ms}$ , slope= $-1.38 \text{ ms/dB}$ ]. No significant regression ( $P>0.05$ ) was encountered between sensation level and N2 latency. When regressing CAEP latency and amplitude characteristics on the subject's age (ranging from 8 to 30 months), only P1 latencies were correlated with age ( $r=-0.50$ ) [ $r^2=0.25$ ,  $F(1.90)=29.9$ ,  $P=0.0000004$ , intercept= $216 \text{ ms}$ , slope= $-3.39 \text{ ms/month}$ ].

**Table 1. Most recent (prior to testing) hearing loss configuration estimations (in dB HL) for four audiometric frequencies (in kHz), based on free field, insert phones, or auditory brainstem response (ABR) testing. Hearing aid types are provided: Siemens Explorer 500P (SE), Phonak Una M (PUM), Siemens Prisma 2K (SP), Phonak Naida V SP BTE (PN), Phonak Una SP (PUS), Phonak Nios Micro III (PN), Phonak Una M A2 (PUM2). Both ears had hearing aids of the same type. Three participants (U7, A3, and A16) are not listed as they were excluded from further data analysis due to incomplete datasets.**

Subject	Hearing aids	Right (dB HL)				Left (dB HL)			
		0.5	1	2	4	0.5	1	2	4
U1			20	25	45	40	25	50	
U2		35	20	35	30	35	20	35	30
U3			45	55	40		50	50	50
U4		55	55	65	65	55	55	65	65
U5		30	40	55	50	30	40	55	50
U6		40	45	45	55	45	35	45	
A1	SE	50	50	60	40	50	60	60	35
A2	PUM	60	70	60	60	60	70	60	60
A4	SP	45	70	65	70				
A5	PN	55	60	75	70	45	55	60	70
A6	PUM	40	60	60	60	50	60	60	55
A7	PUS	40	50	70	90	35	55	70	80
A8	PN	65	60	65	70	65	60	65	70
A9	PN	30		55	65	30	65	75	
A10	SP		55	75		55	75		
A11	SP	40	45	50	55	40	55	50	55
A12	SE		45	65		45	65		
A13	PUM2		55	60		55	60		
A14	PUM2	45	45	45	60	55	60	45	60
A15	SP		70	70		70	70		
A17	SE	50		60	55	60	60	60	
A18	SE	45	50	55	45	50	55		

**Table 2. Group means (and SDs in brackets) for P1 and N2 amplitudes and latencies are provided for specific sensation level ranges. The infant P1 (positive) response was identified as the most positive point of the waveform in the latency range 100 to 300 ms. Similarly, the negative infant N2 response was defined in the latency range 200 to 600 ms, as the most negative point following the positive P1 response. P1 and N2 amplitudes were defined from baseline to peak. Grand averages can be found in Figure 3.**

Sensation level (dB)	P1 amplitude ( $\mu V$ )	P1 latency (ms)	N2 amplitude ( $\mu V$ )	N2 latency (ms)	No. of data points
0 and less	2.4 (2.5)	178 (63)	-3.6 (4.4)	381 (153)	11
5 and 10	3.4 (2.0)	168 (66)	-4.5 (2.7)	373 (146)	22
15 and 20	4.3 (1.9)	147 (47)	-4.1 (3.0)	336 (121)	32
25 and 30	5.2 (3.1)	135 (29)	-5.6 (4.3)	383 (118)	22
35 and more	6.7 (2.8)	119 (15)	-9.4 (6.2)	318 (98)	5

The grand averages of CAEPs in response to the three speech stimuli /m/, /g/, and /t/ are separately shown in Figure 4. Amplitudes and latencies are provided in Table 3 for each speech stimulus. No significant differences ( $P>0.05$ ) were found between speech stimuli for both amplitudes and latencies. No significant interactions ( $P>0.05$ ) were found between speech stimuli and sensation level.

### Noise levels

EEG noise amplitudes per epoch of 92 data points from 22 infants are presented in Figure 5. The mean rms noise amplitude per epoch was  $27.6 \mu\text{V}$  with a standard deviation of  $2.6 \mu\text{V}$ . Assuming that noise amplitude in the waveform average drops by the square root of the number of epochs when noise between epochs is uncorrelated and stationary (*i.e.* constant level over time), the estimated residual noise amplitude in the average would scale to  $1.95 \pm 0.18 \mu\text{V}$  after 200 collected epochs. Although EEG noise is practically uncorrelated over epochs separated in time by steps larger than a second, it is definitely non-stationary and noise levels can change significantly between epochs. Therefore the residual noise amplitude in the average needs to be interpreted as a best-case scenario, with practical noise levels generally higher than the derived levels here.

### Sensitivity of cortical auditory evoked potential detection as a function of sensation level

Figure 6 shows the P-values for all 92 CAEP recordings, versus the sensation level at which the three speech stimuli /m/, /g/, and /t/ were presented. At stimulus levels equal to or above 0 dB SL (*i.e.* the presentation level exceeded the VROA threshold for that stimulus), the presence of a CAEP would be expected if the behavioral threshold is correct. A detection probability P-value less than 0.05 should then occur whenever the magnitude of the cortical response is sufficiently large relative to the residual noise in the average waveform. Conversely, 95% of all data points below 0 dB SL would be expected to have P-values larger than 0.05, again provided that the behavioral threshold is correct.

For the sake of clarity in Figure 6, data points belonging to the same subject have not been individually identified. However, in 37/47=79% of the cases where the same speech sound was presented at two levels 10 dB apart to the same participant, the P-value decreased when the sensation level was increased by 10 dB. Of those 20 cases where a significant P-value was not obtained at the lower of two levels, in 14 of these cases (70%), the P-value became significant when the intensity was increased by 10 dB. Conversely, of those 27 cases where a significant P-value was obtained at the lower of two levels, in only two cases (7%) did the P-value become non-significant after a 10-dB increase in stimulus intensity. These observations are consistent with a significant negative correlation between the logarithm of the P-value and sensation level ( $r=-0.30$ ) [ $r^2=0.09$ ,  $F(1.90)=8.93$ , regression  $P=0.0036$ ,  $\log(P\text{-value})=-1.71-0.045*(\text{SL in dB})$ ].

Based on Figure 6, the ratios of significantly detected responses versus presented stimuli (a ratio defining detection sensitivity) can be derived per sensation level (between brackets, in dB SL): 1/4 (less than -10), 2/2 (-5), 2/5 (0), 6/10 (5), 8/12 (10), 10/15 (15), 13/17 (20), 6/10 (25), 10/12 (30), 3/3 (35), 2/2 (40). Based on these separate values, Table 4 and Figure 7<sup>42,31,43</sup> were constructed. In Table 4, the 92 available data points are divided into three overlapping groups (or ranges): sensation levels  $>0$ ,  $>10$ , and  $>20$  dB SL. This implies that the third range,  $>20$  dB SL, is entirely part of the second range,  $>10$  dB SL, and the second range entirely part of the first range,  $>0$  dB SL. Data are not shown for sensation levels  $>30$  dB SL because of the insufficient number of data points in this range. The proportion of occasions at which a cortical response was present ( $P<0.05$ , or  $P\leq 0.001$  when data collection has been truncated) when presenting above behavioral threshold, was calculated. This proportion is therefore the sensitivity of the cortical test

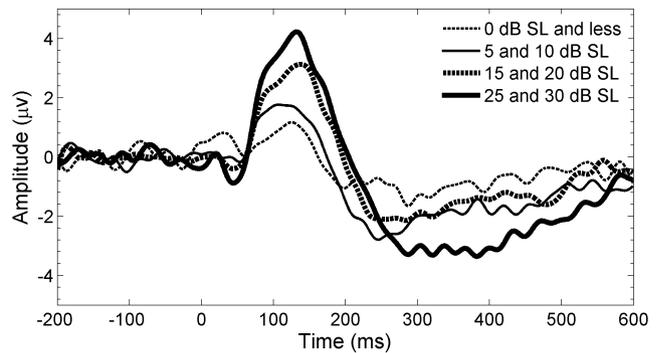


Figure 3. Grand averages for four ranges of sensation levels, based on 11, 22, 32, and 22 data points from low to high sensation levels respectively, obtained from all 22 subjects. Five data points have been omitted as they were sourced from 35 and 40 dB sensation level (SL).

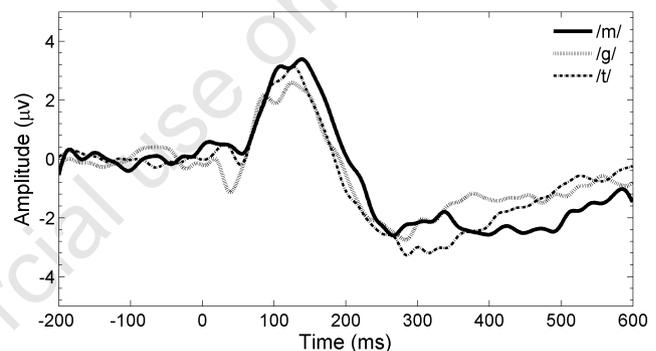


Figure 4. Grand averages for three speech sounds, based on 30, 30, and 32 data points from all 22 subjects for /m/, /g/, and /t/ respectively. All sensation levels are included (from negative up to 40 dB sensation level).

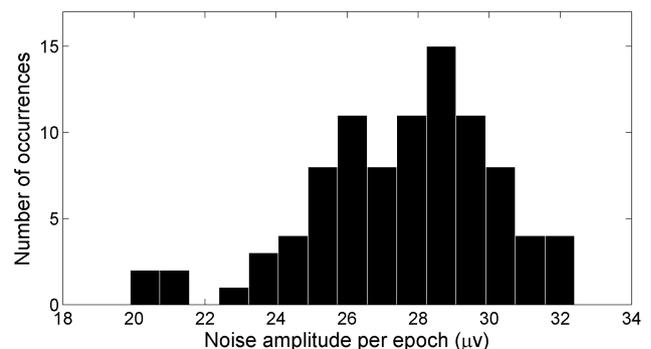


Figure 5. Histogram of rms noise amplitudes per epoch of 92 data points from 22 infants. The noise amplitude is equal to the square root of the mean of 800 point variances. The point variance is the variation around a specific data point of the grand average of the waveform. This is then calculated for each of the 800 data points available in an epoch. Noise values are inherently determined by the electroencephalographic recording filter settings, which in this study ranged from 0.16 to 30 Hz.

for sounds delivered within each of the sensation level ranges. Sensitivities all vary between 71 and 78%.

Confidence intervals for each sensitivity value are calculated by assuming that the binomial distribution approximates the normal distribution. However, confidence intervals of these proportions cannot be computed based solely on the number of available data points, as full independence of these data points cannot be assumed. Individual data points should be uncorrelated with respect to the effect of random noise voltages on the cortical waveform during each measurement. Response presence to a given stimulus across level will, however, be

correlated. If an infant does not have a response to a sound at 20 dB SL, for example, it is unlikely that a response will be detected for that infant and stimulus at a lower sensation level. To allow for this lack of independence, the calculation of confidence intervals for sensitivity are based on the number of unique participants providing the data points in a specific sensation level range. The 95% confidence intervals of the sensitivity are calculated as

$$CI_{95} = S \pm 1.96 \sqrt{\frac{S(1-S)}{N}}$$

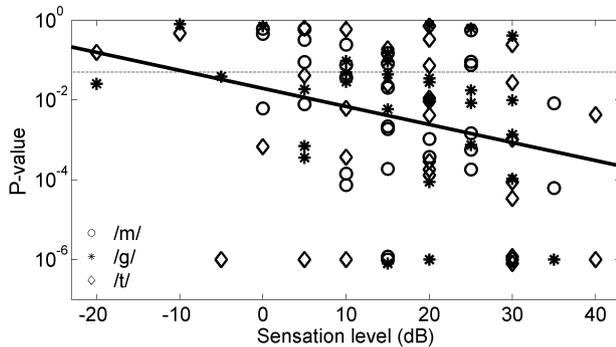


Figure 6. Scatter plot of 92 data points from 22 infants, displaying P-value detection statistic versus sensation level for speech stimuli /m/, /g/, and /t/. P-values are capped at  $10^{-6}$ , plus a small offset to avoid plotting overlapped points. The dashed horizontal line represents a  $P=0.05$  criterion used to determine whether a response could be accepted as present. A solid regression line is shown, which is significant ( $P<0.004$ ) but has a weak correlation ( $r^2=0.09$ ).

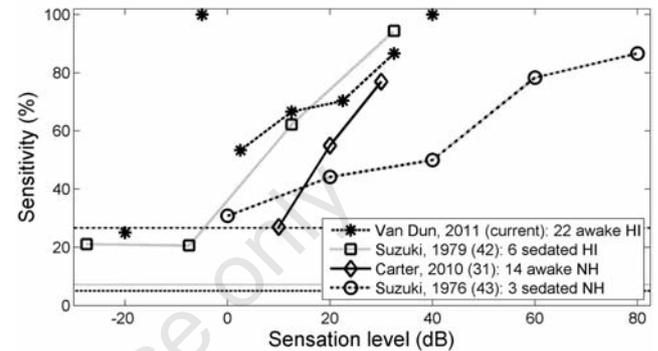


Figure 7. Detection sensitivities (the ratio between detected responses and presented stimuli at a specific sensation level) of cortical auditory evoked potentials for infants with a similar age range from four different studies. Data points (at -20, -5 and 40 dB SL) obtained from fewer than 5 recordings in the current study are not connected with dashed lines. The current study and Suzuki *et al.* report on hearing-impaired infants, while Carter *et al.* and Suzuki *et al.* investigated normally hearing infants. The horizontal lines show the false positive rates reported by each study. Carter *et al.* reported a 0% false rate, and for the current study, as there were no non-stimulus trials, the rate of 5% that results from the acceptance criterion ( $P<0.05$ ) is shown. These studies show that both false positive rates (1-specificity) and detection rates (sensitivities) can vary significantly in infants, a fact one needs to be aware of when interpreting cortical data collected in infants.

Table 3. Group means (and SDs between brackets) for P1 and N2 amplitudes and latencies are given for the three speech stimuli used in this study. The infant P1 (positive) response was identified as the most positive point of the waveform in the latency range 100 to 300 ms. Similarly, the negative infant N2 response was defined in the latency range 200 to 600 ms, as the most negative point following the positive P1 response. P1 and N2 amplitudes were defined from baseline to peak. Grand averages can be found in Figure 4.

Speech stimulus	P1 amplitude ( $\mu V$ )	P1 latency (ms)	N2 amplitude ( $\mu V$ )	N2 latency (ms)	No. of data points
/m/	4.3 (2.8)	159 (56)	-4.8 (3.8)	386 (131)	30
/g/	4.0 (2.0)	147 (48)	-3.9 (2.4)	356 (126)	30
/t/	4.2 (2.9)	148 (52)	-5.5 (4.7)	342 (129)	32

Table 4. Sensitivity versus sensation level range. Sensitivity is calculated as the ratio of the number of detections versus the number of data points (sum of detections versus non-detections) within that sensation level range. The two-sided confidence intervals are based on the number of unique participants that are available for a specific sensation level range.

Sensation level (dB)	No. of detections	No. of data points	Sensitivity (%)	No. of subjects	95% Two-sided CI (%)
>0	58	81	71.6	21	52.3–90.9
>10	44	59	74.6	19	55.0–94.2
>20	21	27	77.8	12	54.3–100.0

CI, confidence interval.

with:

CI<sub>95</sub> the 95% confidence interval lower and upper limit;

S the sensitivity (or proportion) in a sensation level range;

N the number of data points (or subjects) available in a sensation level range.

This reduced number of data points has an effect on the confidence intervals in Table 4, which are wider than when based on the total number of data points. The actual confidence interval values will lie somewhere between those based on the number of data points and those based on the number of subjects. Therefore the intervals in Table 4 are worst-case (*i.e.* widest) scenarios. Even at their widest, the reported confidence intervals exclude 50%. This indicates that the current study is statistically powerful enough to show that cortical testing is capable of making a distinction between the presence or absence of the CAEP at positive sensation levels. When responses to individual speech sounds are considered separately for sensation level ranges >0, >10, and >20 dB SL, the /m/ speech sound results in a detection sensitivity of 63, 72, 63% respectively, /g/ results in 77, 75, and 78%, and /t/ results in 75, 76, and 90%. No significant differences were found, mainly due to the low number of data points for each condition.

Similarly, when calculating sensitivities for unaided and aided conditions separately for the three sensation level ranges, the aided conditions achieve a response detection rate of 69, 76 and 80% respectively, while the unaided group had response detection rate of 80, 70, and 72%. Again, no statistical differences were found between conditions. These results should be interpreted with caution because of the low number of data points, the number of additional factors that can contribute to differences in behavioral thresholds between the different subjects, and also considering that the two groups have different hearing levels. Apart from two stimulation levels below -10 dB SL, sensation levels in this study ranged from -5 to 40 dB SL. Due to the step size of 5 dB, there was a relative small number of stimuli presented at each single sensation level. It was therefore decided to produce pairs of sensation levels starting from 0 dB SL (0 and 5, 10 and 15, 20 and 25, 30 and 35 dB SL), which are equivalent to 10 dB steps. For each pair, the combined ratio of CAEP detections versus stimulus presentations (*i.e.* detection sensitivity) is calculated. Figure 7 displays this sensitivity for each sensation level pair described above (with mean SL values for each pair corresponding to 2.5, 12.5, 22.5, and 32.5 dB SL). Sensitivities at -20, -5, and 40 dB SL contained less than 5 recordings. They are still displayed in Figure 7, but without connecting lines. When combining sensation levels in pairs starting from -5 dB SL (*i.e.* -5 and 0, 5 and 10, etc.), sensitivities are very alike and the same monotonic increasing pattern is observed. These sensitivity values were compared with three other studies.<sup>31,42,43</sup> In Suzuki *et al.*<sup>43</sup> 3 sedated normally hearing children between 1 and 4 years old were tested with 100 ms, 1 kHz pure tones with an inter-stimulus interval of about 2 s. Suzuki *et al.*<sup>42</sup> recorded CAEPs in 6 sedated hearing-impaired children between 2 and 4 years old, with the same stimuli and inter-stimulus interval. Carter *et al.*<sup>31</sup> measured CAEPs in 14 awake normally hearing infants of 12±4 months old using /m/ and /t/ sounds, stimuli identical to this study. Although the number of data points is quite low for all four studies, the curves still show an increasing trend when the sensation level rises. All graphs behave similarly, except for the data from Suzuki *et al.*,<sup>43</sup> which has a much more gradual slope.

## Discussion

This paper aims to determine the relationship between the audibility of sounds at low sensation levels in individual infants and the detectability of the cortical responses they evoke. Twenty-five sensorineurally hearing impaired infants with an age range of 8 to 30

months were tested once. Behavioral thresholds of speech stimuli /m/, /g/, and /t/ were determined and CAEP were recorded. CAEP amplitudes grew with sensation level. No morphological CAEP differences were found between speech sounds. For sensation levels above 0, 10, and 20 dB respectively, detection sensitivities were equal to 72±10, 75±10, and 78±12%.

One of the limitations in this study is the relatively low number of subjects that has been tested. Two reasons for this limitation are the scarce availability of hearing-impaired infants who were developmentally ready for behavioral testing, and the difficulty in obtaining sufficient behavioral and cortical data in one single appointment. This fact has consequences for the statistical (non-)significance and power of the results described in the previous section and any recommendations derived further.

## Grand averages and regression analysis of amplitudes and latencies

Both the relationship between CAEP amplitudes and latencies with sensation level (Table 2) on one hand, and age on the other hand, is rather weak. This is possibly caused by the presence of measurement noise, and/or the inter-subject variability being dominant in this data, given the design of only two sensation levels per stimulus and per subject. Nevertheless these relationships are briefly discussed below.

A large number of studies have investigated the amplitudes and latencies of CAEPs in a younger population (a review can be found in Wunderlich *et al.*<sup>10</sup> Among these studies, there are major differences in study design including age of participants, electrode location, inter-stimulus interval, stimulus intensity, and the type of stimulus (tonal, word, or speech sound). All these different parameters make it difficult to compare between studies.

For example, in the current study, only P1 latency was significantly correlated with age, with latencies being shorter for older participants. Kushnerenko *et al.*<sup>12</sup> confirmed this observation. In addition however, P1 and N2 amplitudes in Kushnerenko *et al.*'s<sup>12</sup> study were also significantly affected by increasing age (absolute P1 and N2 amplitudes showing increases). This might be explained by the study's different age range (0 to 12 months) and/or a shorter inter-stimulus interval (750 ms). Both differences could be responsible for dissimilar waveform morphologies with the current study. According to Sharma *et al.*,<sup>11</sup> P1 latency reduces significantly from birth to adulthood. This is valid for both normally hearing and implanted children, and has also been addressed by Wunderlich *et al.*<sup>10</sup> for children with normal hearing.

Kushnerenko *et al.*<sup>12</sup> tested mainly younger participants with normal hearing (and thus are probably longer exposed to sound than their hearing-impaired counterparts), using louder (70 dB SPL) and longer (100 ms) stimuli, with shorter inter-stimulus intervals (750 ms). Wunderlich *et al.*<sup>10</sup> reported on children with age similar to the children in this study, also using word tokens. However, these children again were normally hearing, and stimuli were louder (85 dB SPL), longer (200 ms), and less frequently presented (an ISI of 3.1 s or more). No studies have been found with similar speech stimuli, age groups, hearing conditions, and (low) stimulus presentation levels.

The trend of increasing CAEP amplitudes and decreasing latencies with increasing sensation level, as shown in Table 2 and Figure 3, has been confirmed in other studies such as Ross *et al.*<sup>38</sup> with adults, Taguchi *et al.*<sup>44</sup> with normally hearing infants, and Suzuki *et al.*<sup>42</sup> with hearing-impaired infants. This paper adds to the few available amplitude and latency data for hearing-impaired infants at low to mid sensation levels.

No significant differences in the group averages for amplitude and latency were found between the different speech sounds /m/, /g/, and /t/ (Table 3 and Figure 4). This is in contrast with Golding *et al.*<sup>23</sup> who reported that the /t/ sound evoked cortical responses were significantly larger in amplitude and earlier in latency than for the other two sounds.

This might be explained by a younger age group, by their normally hearing status, or by the greater spread in the amplitude and latency distribution due to different stimulus levels in this study when comparing with Golding *et al.*<sup>23</sup> who only use one intensity (65 dB SPL).

## Noise levels

In general, noise levels in EEG recordings have been excellently summarized in Chapter 5 of Burkard *et al.*,<sup>39</sup> together with the use of vertical (and horizontal) electro-oculogram recordings to artifact-reject eyeblinks (Chapter 23 of Burkard *et al.*<sup>39</sup>). It is a limitation of this study that no artifact rejection has been performed based on one or more ocular EEG channels, mainly because the HEARLab system is single channel only and does not have this capability. However, an artifact rejection criterion has been adopted to reject all epochs that exceed a specific value, hence excessive noise sources (including eye movements) should be handled appropriately.

More specifically however, a review of the literature suggests that little has been reported on the specifics of EEG noise in infants, toddlers or children. Of course, it has been acknowledged that infant noise variability arises because some infants are unable to remain quietly cooperative for an extended duration and the rate of artifact rejection is prohibitive to further testing.<sup>31</sup> This explains the increased required number of collected epochs for infants when comparing with adults, but a more quantitative report on infant noise figures is not readily available.

Martin *et al.*<sup>45</sup> refer to the issue of EEG noise. However, the study design was completely different to the current paper (being an evaluation of the efficiency of different stimulus strategies to elicit the acoustic change complex), which makes comparison with reported noise values (Figure 5) difficult. The impact of EEG noise, relative to the amplitude of the true response, is best considered by examining the detectability of the cortical response as this measure is related to the signal-to-noise ratio of the recording. Therefore the reported noise levels in the average (after 200 accepted epochs, with similar filter settings) should be considered as a target when one wants to obtain similar signal-to-noise ratios, and hence similar detection sensitivities.

According to previous ABR work, a signal-to-noise ratio (rms amplitude of the evoked potential versus rms amplitude of the background noise) of  $\sqrt{2.1}=1.45$  means that an ABR is present at a confidence level of 99%.<sup>46</sup> However in CAEP testing noise levels depend, more than in ABR recording, on filter settings due to the variable spectral shape of the EEG background noise at lower frequencies. Hence it is less straightforward to derive normative data for signal-to-noise ratios and detection of cortical responses. In addition, relationships between CAEP amplitudes and noise values for the Hotelling's  $T^2$  statistic have not been determined yet. But apart from these observations, in order to be detected, it seems that CAEP amplitudes indeed need to be of the same order of magnitude as the residual background noise. This is evidenced by the CAEP amplitudes visible in Figures 3 and 4 when compared with an estimated residual noise value of about 2  $\mu$ V after 200 epochs.

## Sensitivity of cortical auditory evoked potential detection in function of sensation level

All 22 children in this study showed CAEPs for at least one sound for at least one presentation level. However, based on sensitivities (the ratio of the number of detections versus the sum of detections and non-detections) reported in Table 4, as many as 25% of separate stimuli exceeding 10 dB SL did not evoke a CAEP that was detectable with  $P < 0.05$ . This shows one should take absolute care in interpreting a stand-alone result. It is therefore critical to view the larger picture and to consider test results for all stimuli and intensities obtained from the same child.

There are several possible reasons for missing CAEPs above an apparent behavioral hearing threshold in this experiment. First, the stimuli used in this study are estimated according to behavioral VROA thresholds, for which there is some degree of measurement error. For example, this uncertainty is visible in Figure 6 with a proportion of detected CAEPs for sensation levels of 0 dB and below. It can also be noted in Figure 3, which shows a low-amplitude positivity and negativity for this sensation level range. An equivalent to false-positives can be measured through non-stimulus trials. Between five and nine per cent of non-stimulus trials caused a head turn in other studies for children in the same age group.<sup>40,47</sup> In addition, a test-retest variability is reported of about 10 dB (for step sizes of 10 dB) in infants aged between 5 and 18 months using visual reinforcement audiometry, a technique which is very similar to VROA.<sup>47</sup> These studies have generally used warble tones at audiometric frequencies as the test stimulus. The current study uses a 4 Hz stimulus train of short speech sounds, which might influence both false-positive rate and test-retest variability, but to an unknown degree or direction.

Second, the infant's state of arousal is known to influence the morphology and detectability of CAEPs.<sup>16</sup> While the children in this study were observed for changes in alertness, it can be difficult to subjectively determine that young children are still in an optimal state for the CAEP to be observed, or even that they remain in a similar state throughout the assessment. Suzuki *et al.*<sup>43</sup> noted that responses from sleeping infants (1-4 years) can be unreliable due to observed theta-waves (4-7 Hz), as evidenced by the high false positive rates for their tested normally hearing infants in Figure 7. Suzuki *et al.*<sup>43</sup> employed visual examination of waveforms to determine the presence or absence of a CAEP. A different specificity may have resulted if an objective statistical CAEP detection paradigm was used. There is divided opinion about whether CAEPs are more robust in awake or sleeping children. Taguchi *et al.*<sup>44</sup> used tone-bursts as stimuli to assess CAEPs in 220 sleeping newborn infants. They reported that CAEPs during sleep were much more easily detectable than those during the waking state because of the larger amplitude (and longer latency) of the response and the greater ease in the handling of the subject. Based on this information, we could conclude that infants can be tested both awake and asleep. However, one needs to take care the child does not shift between states during the same recording as the waveform will vary during the recording, and this variation will be interpreted as noise rather than as an altered evoked response.

Third, it appears to be the case that CAEPs are not detectable in some individuals at low to medium sensation levels. For a group of normally hearing and hearing-impaired adults, assessed using tone-bursts as stimuli, between 6.4 (after bias correction) and 14.5% of the individual threshold errors are 15 dB or more (cortical thresholds being more elevated than behavioral thresholds).<sup>48,49</sup> Unpublished data from another study by the current authors (2010) showed that in 3% of the cases, the discrepancy between CAEPs and behavioral thresholds even exceeded 35 dB in some hearing impaired (older) subjects. In this unpublished study, we tested 34 hearing-impaired adults unaided with insert phones at the four audiometric frequencies (left and right ears). First, we obtained behavioral hearing thresholds, then determined the subject's hearing threshold using cortical responses while awake (5 dB steps using tone-bursts, 50 ms long, 120 repeats). Cortical detection was carried out using an automatic detection system (identical to the one in this paper). The only difference between the current infant study and this unpublished adult study is the use of free field versus insert earphones, and speech sounds versus tone-bursts. A study of both awake and sleeping children aged between 21 days and 15 years (65 with normal hearing and 93 with hearing impairment) reported a difference between cortical and behavioral thresholds (obtained through conditioned audiometry) of 10, 20, and 30 dB for 38, 36, and 9

children respectively.<sup>50</sup> It is quite remarkable that for the 20 and 30 dB differences, the cortical thresholds were almost always *better* than the behavioral thresholds. This contrasts hugely with the results described in the current study, which findings mainly point in the other direction. In a review paper, Davis *et al.*<sup>51</sup> analyzed CAEP recordings from 162 children aged 4 to 15 years, and reported no cases where a difference larger than 25 dB between behavioral and CAEP thresholds was observed. Barnett<sup>52</sup> tested 141 normally hearing, mostly sleeping infants, under 3 years of age. In 32% of the cases up to the age of nine months, a response was not detected to a presentation at 35 dB SPL. However, for 80 children nine months and older, only 4% of cases failed to show a CAEP at a presentation level of 35 dB SPL. This result corresponds with Figure 7, which seems to suggest for the three out of four displayed studies that both normally hearing (when the data of Carter *et al.*<sup>31</sup> are extrapolated) and hearing-impaired children almost always have detectable cortical responses when the sensation level is 35 dB or more. The largest difference between the current study and both Morgon *et al.*<sup>50</sup> and Barnett<sup>52</sup> studies is that these other studies used inter-stimulus intervals of 2 s or more, and included (some) sleeping infant subjects.

Fourth, the absence of a response may be inconsistent when multiple tests are performed. It is known that 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. This seems to be particularly evident in infant-generated responses.<sup>31</sup> According to Barnett,<sup>52</sup> the proportion of subjects up to nine months having no response at 35 dB SPL drops from 32 to 0% when conducting each test three times (although a correction was not made for increased false positives due to multiple testing). Nevertheless, it is interesting to consider that the number of non-detections reduces faster than the increase in false-detections. As stated, the design of the current study did not include retests, which is a limitation the authors acknowledge, and it is possible that a reduction in non-detection rates might have occurred if a retest was included. In the clinical setting it is recommended that isolated (non-)detections are interpreted with caution. Any single CAEP recording should be considered in combination with other measurements, like test runs at neighboring stimulus intensities or with speech sounds containing adjacent or overlapping frequencies. For example, when a stimulus at X dB SL does produce a significant response, it is very unlikely (7% according to this study) the recording at X+10 dB SL does not produce a significant response. Moreover, in 79% of the cases the P-value will decrease when the stimulus level increases.

Fifth, all recordings in the current study collected at least 200 accepted epochs. Although this number seems sufficient in our view, especially in the light of keeping recording time clinically acceptable, it is obvious that longer measurement times likely will result in higher sensitivities.

Sixth, due to technical limitations, the HEARLab system uses blocks of 25 stimuli. It would have been more beneficial if the system would have used an alternating stimulus presentation strategy, as Butler<sup>53</sup> indicated with tone bursts that habituation is maximum when the same neural units are activated repeatedly. Woods *et al.*<sup>54</sup> showed that short-term habituation is complete by the third stimulus and results in amplitude reductions of 50-75% for the N1 component. This effect was even more marked for speech sounds. As these studies were conducted on adults, it is difficult to derive the actual effect on infants. However, we can assume that alternating stimuli when testing infants also have a beneficial effect on their CAEP amplitudes. If this is the case, detection sensitivities might be negatively influenced by using longer blocks of stimuli, mainly because CAEP amplitudes will be smaller than using an alternating stimulus paradigm.

Finally, the issue of acoustic signal-to-noise ratios during CAEP

recordings is also an important consideration.<sup>21</sup> In the current study, acoustic noise was controlled as much as possible, by conducting the testing in a sound attenuating test booth, and by having an observer control the behavior of the child through appropriate distraction. Observers were instructed to pause measurements if the child started to vocalize too much.

Apart from these factors, several studies have questioned the applicability of CAEPs for the evaluation of hearing aid fittings, and have cautioned that more research is needed first before results can be clinically applied.<sup>21,30,35-37</sup> These studies indicate that it is possible there are unknown parameters in a hearing aid that may influence the relationship between speech sounds arriving at the hearing aid microphone and the CAEP recordings. However, their results are based largely on experimental results measured with normal-hearing adults, which makes it necessary to investigate this matter further with hearing-impaired users, the target group of hearing aids, before deriving any conclusions.

### Clinical implications of cortical responses

The population in this study was drawn from local pediatric hearing centers in Australia who see these children regularly (multiple times a year). The results obtained from this study, both cortical and behavioral, have been taken into account for further hearing aid fitting adjustment. It needs to be noted however that in Australian pediatric audiological practice, cortical recordings are the main source of information (apart from feedback from the parents) for evaluation of aided performance in young infants (or children with multiple disabilities), as the children that get cortically tested, are not (yet) behaviorally assessable. Hence, based on the results from the current study and practical experience in the clinical setting, what implications do these findings have for management of individual hearing aid fittings, where the actual behavioral thresholds are unknown or uncertain?

First, consider the situation where a cortical response is considered to be present ( $P < 0.05$ ) in response to a sound at a conversational level. This result might provide some confidence that the sound is stimulating the auditory cortex at the level tested. The smaller the P-value that is indicated by the automatic statistical detection algorithm, the less likely it is that the CAEP waveform is the result of random electrical activity on the scalp. This probability can be taken into account in deciding how much confidence to place on the finding when combining this piece of information with other information available about the child. As it is relatively common (Figure 6) to find probability levels of 0.01 or less, sometimes orders of magnitude less, it will often be the case that one could have a high degree of confidence that the infant is detecting the sound that elicited the cortical response.

Second, consider the case where no cortical response is considered to be present (*i.e.*  $P \geq 0.05$ ). If the true sensation level of the sound were to be greater than 10 dB SL, the most likely outcome (75% probability, confidence interval range 55 to 94%) is that a cortical response would have been detected with  $P < 0.05$ . If the sound truly were to be inaudible, then the statistical detection criterion adopted ( $P < 0.05$ ) will ensure no response is detected 95% of the time. If the true sensation level were to be within the range of 0 to 10 dB SL, then the probability of a significant response being detected is intermediate; results of this study indicate that there is a 16/27=59% chance of a response being detected (sensation levels 0, 5, and 10 dB SL incorporated). In summary, when the true sensation level is unknown, these experimental results seem to indicate that the lack of an apparent cortical response indicates a likelihood, but by no means a certainty, that the sensation level is 10 dB or less. One could use this likelihood to supplement other information that happens to be available about the infant (parent report of behavior, direct observation of behavior, and possibly calculated audibility based on ABR or ASSR thresholds, measured hearing aid gain, and assumed relationships between pure-tone thresholds and

audibility of speech sounds), all of which constitutes far from precise information, to guide management of the child.

Despite the possibility of audible stimuli sometimes failing to elicit a detectable CAEP, the detection of a CAEP response could provide very useful information, particularly in the context of other information available to the clinician. At the other extreme, in cases where CAEPs cannot be detected at all (particularly when professionals and parents have failed to observe behavioral responses to everyday sounds) the technique could allow an objective indication that *something* indeed might not be optimal. As hearing aid technology becomes increasingly complex it is even more necessary to continue research, with the aim of providing detailed objective measures to ensure that hearing aids are providing the audibility of speech that children need to develop.

## References

1. Joint Committee on Infant Hearing. Year 2000 position statement: principles and guidelines for early hearing detection and intervention programs. *Pediatrics* 2000;106:798-817.
2. Yoshinaga-Itano C, Sedey A, Coulter D, Mehl A. Language of early- and later-identified children with hearing loss. *Pediatrics* 1998;102:1161-71.
3. Seewald R, Scollie S. An approach for ensuring accuracy in pediatric hearing instrument fitting. *Trends Amplif* 2003;7:29-40.
4. Ching T, Dillon H. Prescribing amplification for children: adult-equivalent hearing loss, real-ear aided gain, and NAL-NL1. *Trends Amplif* 2003;7:1-9.
5. Golding M, Pearce W, Seymour J, Cooper A, Ching T, Dillon H. The relationship between obligatory cortical auditory evoked potentials (CAEPs) and functional measures in young infants. *J Am Acad Audiol* 2007;18:117-25.
6. Kurtzberg D, Stapells D, Wallace I. Event-related potential assessment of auditory system integrity: implications for language development. In: Vietze P, Vaughan H, (eds). *Early identification of infants with developmental disabilities*. New York: Grune & Stratton; 1988. pp 160–80.
7. Rance G, Cone-Wesson B, Wunderlich J, Dowell R. Speech perception and cortical event related potentials in children with auditory neuropathy. *Ear Hear* 2002;23:239-53.
8. Stapells D. Cortical event-related potentials to auditory stimuli. In: Katz J, (ed). *Handbook of clinical audiology*. 6th ed. Philadelphia, PA: Lippincott Williams & Wilkins; 2009. pp 395–430.
9. Sharma A, Cardon G, Henion K, Roland P. Cortical maturation and behavioral outcomes in children with auditory neuropathy spectrum disorder. *Int J Audiol* 2011;50:98-106.
10. Wunderlich J, Cone-Wesson B, Shepherd R. Maturation of the cortical auditory evoked potential in infants and young children. *Hear Res* 2006;212:185-202.
11. Sharma A, Dorman M, Spahr A. A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear Hear* 2002;23:532-9.
12. Kushnerenko E, Ceponiene R, Balan P, Fellman V, Huotilainen M, Näätänen R. Maturation of the auditory event-related potentials during the first year of life. *Neuroreport* 2002;13:47-51.
13. Mendel M, Hosick E, Windman T, Davis H, Hirsh S, Dinges D. Audiometric comparison of the middle and late components of the adult auditory evoked potentials awake and asleep. *Electroencephal Clin Neurophysiol* 1975;38:27-33.
14. Campbell K, Colrain I. Event-related potential measures of the inhibition of information processing: II. The sleep onset period. *Int J Psychophysiol* 2002;46:197-214.
15. Näätänen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and an analysis of the component structure. *Psychophysiology* 1987;24:375-425.
16. Picton T, Hillyard S. Human auditory evoked potentials. II: effects of attention. *Electroencephal Clin Neurophysiol* 1974;36:191-200.
17. Onishi S, Davis H. Effects of duration and rise time of tone bursts on evoked V potentials. *J Acoust Soc Am* 1968;44:582-91.
18. Picton T, Woods D, Baribeau-Braun J, Healey T. Evoked potential audiometry. *J Otolaryngol* 1977;6:90-119.
19. Sharma A, Dorman M. Cortical auditory evoked potential correlates of categorical perception of voice-onset time. *J Acoust Soc Am* 1999;106:1078-83.
20. Jacobson G, Lombardi D, Gibbens N, Ahmad B, Newman C. The effects of stimulus frequency and recording site on the amplitude and latency of multichannel cortical auditory evoked potential (CAEP) component N1. *Ear Hearing* 1992;13:300-6.
21. Billings C, Tremblay K, Stecker G, Tolin W. Human evoked cortical activity to signal-to-noise ratio and absolute signal level. *Hear Res* 2009;254:15-24.
22. Gilley P, Sharma A, Dorman M, Martin K. Developmental changes in refractoriness of the cortical auditory evoked potential. *Clin Neurophysiol* 2005;116:648-57.
23. Golding M, Purdy S, Sharma M, Dillon H. The effect of stimulus duration and inter-stimulus interval on cortical responses in infants. *Aust New Zeal J Audiol* 2006;28:122-36.
24. Picton T, Goodman W, Bryce D. Amplitude of evoked responses to tones of high intensity. *Acta Otolaryngol* 1970;70:77-82.
25. Novak G, Kurtzberg D, Kreuzer J, Vaughan Jr H. Cortical responses to speech sounds and their formants in normal infants: maturational sequence and spatiotemporal analysis. *Electroencephal Clinical Neurophysiol* 1989;73:295-305.
26. Pang E, Taylor M. Tracking the development of the N1 from age 3 to adulthood: an examination of speech and non-speech stimuli. *Clin Neurophysiol* 2000;111:388-97.
27. Ponton C, Don M, Eggermont J, Waring M, Kwong B, Masuda A. Auditory system plasticity in children after long periods of complete deafness. *Neuroreport* 1996;8:61-5.
28. Bauer P, Sharma A, Martin K, Dorman M. Central auditory development in children with bilateral cochlear implants. *Arch Otolaryngol Head Neck Surg* 2006;132:1133-6.
29. Dillon H. So, baby, how does it sound? Cortical assessment of infants with hearing aids. *Hear J* 2005;58:10-7.
30. Souza P, Tremblay K. New perspectives on assessing amplification effects. *Trends Amplif* 2006;10:119-43.
31. Carter L, Golding M, Dillon H, Seymour J. The detection of infant cortical auditory evoked potentials (CAEPs) using statistical and visual detection techniques. *J Am Acad Audiol* 2010;21:347-56.
32. Golding M, Dillon H, Seymour J, Carter L. The detection of adult cortical auditory evoked potentials (CAEPs) using an automated statistic and visual detection. *Int J Audiol* 2009;48:833-42.
33. Rapin I, Graziani L. Auditory-evoked responses in normal, brain-damaged, and deaf infants. *Neurology* 1967;17:881-94.
34. Korczak P, Kurtzberg D, Stapells D. Effects of sensorineural hearing loss and personal hearing aids on cortical event-related potential and behavioral measures of speech-sound processing. *Ear Hear* 2005;26:165-85.
35. Tremblay K, Kalstein L, Billings C, Souza P. The neural representation of consonant-vowel transitions in adults who wear hearing aids. *Trends Amplif* 2006;10:155-62.
36. Tremblay K, Billings C, Friesen L, Souza P. Neural representation of amplified speech sounds. *Ear Hear* 2006;27:93-103.
37. Billings C, Tremblay K, Souza P, Binns M. Effects of hearing aid amplification and stimulus intensity on cortical auditory evoked potentials. *Audiol Neurootol* 2007;12:234-46.
38. Ross B, Lutkenhoner B, Pantev C, Hoke M. Frequency-specific

- threshold determination with the CERAGram method: basic principle and retrospective evaluation of data. *Audiol Neurotol* 1999;4:12-27.
39. Burkard RF, Eggermont JJ, Don M. Auditory evoked potentials: Basic principles and clinical application. Philadelphia, PA: Lippincott Williams & Wilkins; 2007.
  40. Birtles G. A visual reinforcement orientation audiometry screening procedure. *Aust J Audiol* 1989;11:1-9.
  41. American Electroencephalographic Society. Guidelines for standard electrode position nomenclature. *J Clin Neurophysiol* 1991;8:200-2.
  42. Suzuki T, Taguchi K, Yoda M. Reliability of slow vertex response audiometry in young children with sensorineural hearing loss. *Audiology* 1979;18:119-24.
  43. Suzuki T, Yamamoto K, Taguchi K, Sakabe N. Reliability and validity of late vertex-evoked response audiometry. *Int J Audiol* 1976;15:357-69.
  44. Taguchi K, Picton T, Orpin J, Goodman W. Evoked response audiometry in newborn infants. *Acta Otolaryngol* 1969;67:5-17.
  45. Martin B, Boothroyd A, Ali D, Leach-Berth T. Stimulus presentation strategies for eliciting the acoustic change complex: increasing efficiency. *Ear Hear* 2010;31:356-66.
  46. Elberling C, Don M. Quality estimation of averaged auditory brainstem responses. *Scand Audiol* 1984;13:187-97.
  47. Widen J. Adding objectivity to infant behavioral audiometry. *Ear Hear* 1993;14:49-57.
  48. Lightfoot G, Kennedy V. Cortical electric response audiometry hearing threshold estimation: accuracy, speed, and the effects of stimulus presentation features. *Ear Hear* 2006;27:443-56.
  49. Tsui B, Wong L, Wong E. Accuracy of cortical evoked response audiometry in the identification of non-organic hearing loss. *Int J Audiol* 2002;41:330-3.
  50. Morgon A, Charachon D, Gerin P. Electro-encephalographic audiometry for young children. *Eur Arch Otorhinolaryngol* 1971;198:144-50.
  51. Davis H, Hirsh S, Shelnutt J, Bowers C. Further validation of evoked response audiometry (ERA). *J Speech Hear Res* 1967;10:717-32.
  52. Barnet A. EEG audiometry in children under three years of age. *Acta Otolaryngol* 1971;72:1-13.
  53. Butler RA. Effect of changes in stimulus frequency and intensity on habituation of the human vertex potential. *J Acoust Soc Am* 1968;44:945-50.
  54. Woods DL, Elmasian R. The habituation of event-related potentials to speech sounds and tones. *Electroencephal Clinical Neurophysiol* 1986;65:447-59.