Aided Cortical Auditory Evoked Potentials in Infants With Frequency-Specific Synthetic Speech Stimuli: Sensitivity, Repeatability, and Feasibility

Anisa S. Visram,[1,2] Michael A. Stone,[1,2] Suzanne C. Purdy,[3] Steven L. Bell,[4] Jo Brooks,[1,2]
Iain A. Bruce,[2] Michael A. Chesnaye,[4] Harvey Dillon,[1,5] James M. Harte,[6,7]
Caroline L. Hudson,[1,2] Søren Laugesen,[6] Rhiannon E. Morgan,[1,2] Martin O'Driscoll,[2]
Stephen A. Roberts,[1] Amber J. Roughley,[1,2] David Simpson,[4] and Kevin J. Munro[1,2]

Objectives: The cortical auditory evoked potential (CAEP) test is a candidate for supplementing clinical practice for infant hearing aid users and others who are not developmentally ready for behavioral testing. Sensitivity of the test for given sensation levels (SLs) has been reported to some degree, but further data are needed from large numbers of infants within the target age range, including repeat data where CAEPs were not detected initially. This study aims to assess sensitivity, repeatability, acceptability, and feasibility of CAEPs as a clinical measure of aided audibility in infants.

Design: One hundred and three infant hearing aid users were recruited from 53 pediatric audiology centers across the UK. Infants underwent aided CAEP testing at age 3 to 7 months to a mid-frequency (MF) and (mid-)high-frequency (HF) synthetic speech stimulus. CAEP testing was repeated within 7 days. When developmentally ready (aged 7-21 months), the infants underwent aided behavioral hearing testing using the same stimuli, to estimate the decibel (dB) SL (i.e., level above threshold) of those stimuli when presented at the CAEP test sessions. Percentage of CAEP detections for different dB SLs are reported using an objective detection method (Hotellings T[2]). Acceptability was assessed using caregiver interviews and a questionnaire, and feasibility by recording test duration and completion rate.

Results: The overall sensitivity for a single CAEP test when the stimuli were ≥0 dB SL (i.e., audible) was 70% for the MF stimulus and 54% for the HF stimulus. After repeat testing, this increased to 84% and 72%, respectively. For SL >10 dB, the respective MF and HF test sensitivities were 80% and 60% for a single test, increasing to 94% and 79% for the two tests combined. Clinical feasibility was demonstrated by an excellent >99% completion rate, and acceptable median test duration of 24 minutes, including preparation time. Caregivers reported overall positive experiences of the test.

Conclusions: By addressing the clinical need to provide data in the target age group at different SLs, we have demonstrated that aided CAEP testing can supplement existing clinical practice when infants with hearing loss are not developmentally ready for traditional behavioral assessment. Repeat testing is valuable to increase test sensitivity. For clinical application, it is important to be aware of CAEP response variability in this age group.

Key words: Cortical auditory evoked potential, Hearing aid, Infant.

Abbreviations: ABR=auditory brainstem response; ANSD=auditory neuropathy spectrum disorder; BSA=British Society of Audiology; CAEP=cortical auditory evoked potential; FPR=false positive rate; HF=(mid-)high-frequency (stimulus); HT[2]=Hotellings T[2]; ISTS=International Speech Test Signal; MF=mid-frequency (stimulus); MRL=minimum response level; rms=root-mean-square; SL=sensation level; SNR=signal-to-noise ratio; SpRefL=speech reference level; VRA=visual reinforcement audiometry.

INTRODUCTION

In the United Kingdom, the median age for prescription-fitting of hearing aids for a newly diagnosed infant is just 82 days (Wood et al. 2015). Other than subjective caregiver questionnaires (e.g., Tsiakpini et al. 2004), there is no agreed international guideline as to how to assess the benefit of hearing aids in infants who are too young to perform behavioral testing, that is, under around 7 to 9 months' developmental age. One clinical procedure that has received interest in terms of aided and unaided assessment of infants and others unable to perform behavioral testing is the cortical auditory evoked potential (CAEP). The purpose of the CAEP test is to confirm physiological detection of sound stimuli at the level of the cortex. The test is already in regular use in audiology clinics across Australia (Punch et al. 2016) but has not been taken up widely in other countries, despite the potential benefits including earlier hearing aid fitting and earlier cochlear implant referral (Mehta et al. 2017, 2020). Reasons for lack of uptake may include uncertainty over how to interpret absent responses, and how to interpret responses to relatively broadband speech stimuli. In this article, we present CAEP data from 103 infants with hearing loss using a newly-developed protocol and new test stimuli.

CAEPs are evoked responses at the level of the auditory cortex in response to a sound stimulus. The most common clinical applications have been for objectively determining auditory thresholds in adults (e.g., Lightfoot & Kennedy 2006), for

[1]Manchester Centre for Audiology and Deafness, School of Health Sciences, University of Manchester, United Kingdom; [2]Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, United Kingdom; [3]School of Psychology, University of Auckland, Auckland, New Zealand; [4]Institute of Sound and Vibration Research, Faculty of Engineering and Physical Sciences, University of Southampton, United Kingdom; [5]Department of Linguistics, Macquarie University, Sydney, Australia; [6]Interacoustics Research Unit, c/o Technical University of Denmark, Denmark; and [7]Eriksholm Research Centre, Denmark.

confirming physiological detection of suprathreshold speech sounds in infants both with and without hearing aids (Chang et al. 2012; Van Dun et al. 2012; Gardner-Berry et al. 2016; Punch et al. 2016), and for assessing hearing function in infants with auditory neuropathy spectrum disorder (ANSD) (Rance et al. 1999; Gardner-Berry et al. 2016). Audiologists that regularly use CAEPs for infant hearing assessment report they are useful for supporting earlier decisions on hearing aid fittings, fine tuning hearing aids, and earlier cochlear implant referrals (Mehta et al. 2020). However, as will be shown later, caution is advised if using CAEPs to inform hearing aid adjustments, due to uncertainties in interpreting nonresponses. CAEP assessment is a routine feature of the pediatric clinical pathway for babies with hearing loss in Australia (King et al. 2014), whereby aided responses to speech sounds are measured using the Aided Cortical Assessment module of the bespoke "HearLab" system (Frye Electronics, Tigard, OR), and the results may inform adjustments to hearing aid settings. However, reliable CAEP responses are not recorded in every infant even when the stimuli are audible. Chang et al. (2012) measured CAEPs in 18 infants with hearing loss aged 3 to 10 months, using a mix of aided and unaided conditions, and found that in 30% of cases a CAEP was not detected, although presented at >20 dB sensation level (SL). Van Dun et al. (2012) found that 22% of cases with stimulus of >20 dB SL did not produce a detectable CAEP response in infants with hearing loss aged 8 to 30 months, again with a mix of aided and unaided conditions. Gardner-Berry et al. (2016) reported similar results, with around 20 to 30% of undetected responses for SLs >20 dB in infants with ANSD and sensorineural hearing loss (SNHL) seen in a pediatric audiology clinic. Concern about the reliability of response detection suggests clinicians should be cautious about interpreting nonresponses in children with hearing loss.

Munro et al. (2019) assessed the feasibility of CAEP testing in infants with normal hearing. They recorded CAEPs using the HearLab in 104 infants (5 to 39 weeks) with no hearing concerns or significant risk factors. Detection rates were high (86 to 100% for three different consonant stimuli), with a median test duration of 27 minutes and 94% test completion. These findings met the predefined criteria for clinical feasibility (Munro et al. 2019) and led to the present study, which aimed to investigate CAEP testing in infants with hearing loss, by recording aided CAEPs at the target age of 3 to 7 months.

The HearLab system uses short duration speech sounds (extracted from running speech) with energy dominating in different frequency bands, but also containing significant energy at more distant frequencies. The somewhat broadband nature of the stimuli limits the extent to which a frequency-specific assessment of functional hearing can be achieved, at least for those with steeply-sloping hearing losses. Moreover, the stimuli are calibrated such that their level, measured with an impulse time constant, is equal to the long-term root-mean-square (rms) level of the running speech from which they were extracted, which approximates the level the stimuli had in the running speech. Statistical analysis of speech content shows considerable variability in levels of consonant bursts (Moore et al. 2008) and testing mid-to-high-frequency stimuli at the same level as running speech takes them far above the average power in those frequency bands. The present study addressed these issues by using custom-designed stimuli, which have been designed to be (1) more frequency specific than the HearLab stimuli (while

retaining some speech-like features), and (2) calibrated relative to the long-term average power present in the appropriate frequency bands in the long-term average speech spectrum (the same stimulus commonly used in fitting prescriptions). These stimuli are further described later and in Stone et al. (2019).

This study focused on infant hearing aid users aged 3 to 7 months, representing the time period classically between auditory brainstem response (ABR) testing and reliable visual reinforcement audiometry (VRA) testing, where there is a need to bridge the gap in infant hearing assessment. Infants were later seen for aided behavioral testing (using VRA) using the same stimuli as used for CAEP testing, to determine the stimulus SL. These approaches address some of the limitations of previous studies in which the infants were older than the target population (Van Dun et al. 2012) or where stimulus SL was indirectly estimated based on coupler gain, average real-ear-to-coupler differences, and unaided audiometric thresholds (Chang et al. 2012).

The primary aim of the study was to investigate CAEPs in different frequency bands in infants with hearing loss, to establish the sensitivity at different SLs of the aided CAEP test for the stimuli developed by Stone et al. (2019). A second aim was to investigate CAEP repeatability for aided infants, and to determine, for cases where a CAEP was not detected (although the stimulus was audible and recording conditions were acceptable), in what proportion of these cases the CAEP could be detected upon retest. Third, the study investigates caregiver acceptability of the test via questionnaires and interviews. Finally, the study aimed to investigate how feasible the proposed CAEP test is for application to aided assessment of infants in UK audiology clinics by recording completion rate and test duration.

MATERIALS AND METHODS

Participants and Assessment Visits

One hundred and three infants were recruited from 53 different pediatric audiology centers across the UK (Table 1). Inclusion criteria were hearing aid users with permanent bilateral hearing loss; 3 to 7 months old at start of testing; without ANSD or significant developmental delay that would affect later behavioral hearing testing. Participant hearing threshold data are summarised in Figure 1, showing a range of hearing losses, most densely represented in the moderate range. Ethical approval was obtained from the North West National Research Ethics Service Ethics Committee (reference 15/NW/0736). All testing took place in a Mobile Hearing Research Van, consisting of a single-walled sound-isolated booth and separate observation room, mounted inside a van with stiff but lightweight exterior walls. Consultation with patient-public representatives during early study development led to the suggestion of such a set-up as a necessary measure to make the study accessible to families across the UK. By prior appointment, the vehicle visited families at or close to their homes, at two time points (CAEP session and VRA session). Both the CAEP/VRA session test procedures were repeated within 7 days of each initial session, totalling two CAEP sessions and two VRA sessions. The time interval between the tests was based on caregiver and researcher availability. If the repeat session was on the same day, there was at least 1 hour break between sessions. These visits were scheduled outside of the families' regular audiology appointments. Clinical test results, including ABR, VRA,

TABLE 1. Demographics of participants at test sessions

	CAEP Session (n = 103)	VRA Session (n = 98)
Sex	59 (57%) male, 44 (43%) female	57 (58%) male, 41 (42%) female
Age (mo)	Mean = 5.2, range = 3.0–7.5	Mean = 10.8, range = 7.4–21.4
Tympanometry results	Bilateral pass = 73 (71%)	Bilateral pass = 49 (50%)
	Unilateral fail = 8 (8%)	Unilateral fail = 20 (20%)
	Bilateral fail = 22 (21%)	Bilateral fail = 29 (30%)

For tympanometry results, "Fail" includes cases where a tympanogram could not be successfully recorded. For the VRA session, in one case the infant had a bilateral tympanometry pass at the initial session and unilateral pass at the repeat session (1 day later). The VRA estimates for this child were based on the session with normal tympanograms and the infant was classed as having bilateral normal tympanograms. Infants under 6 months were tested with high frequency (1000 Hz) tympanometry, classified as a pass if a visible peak was present. Infants >6 months were tested with low frequency (226 Hz) tympanometry, classified as a pass if compliance was ≥0.2 mL and middle ear pressure ≥−200 daPa. CAEP, cortical auditory evoked potential; VRA, visual reinforcement audiometry.

tympanometry and aetiology, were acquired from the infants' local audiology centers where possible. These clinical responses are referred to as audiometric thresholds to distinguish them from aided minimum response levels (MRLs) obtained from the current study. When determining whether a progressive loss was present between the sessions, researchers made use of all clinical data, including ABR, VRA, etiology, and tympanometry, as available. Reported participant hearing thresholds (Fig. 1) are based on the estimate of hearing thresholds at the time of the latter test session (VRA session) and are typically based on VRA data obtained by the local audiology centres (i.e., clinicians who were not part of the research team).
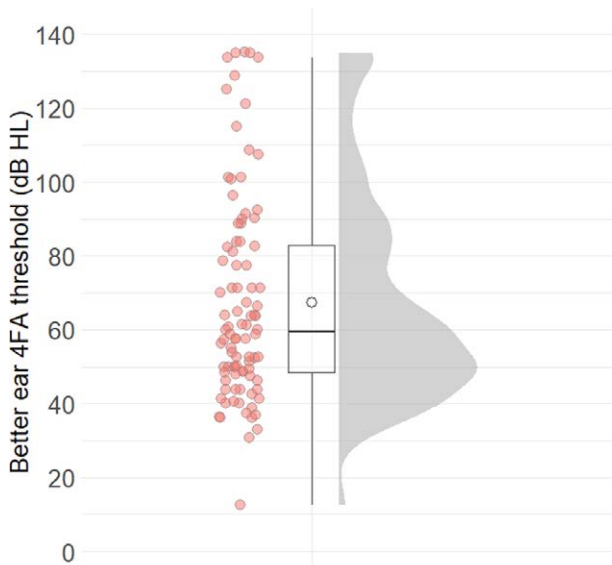

Fig. 1. Four-frequency average (4FA, 0.5, 1, 2, 4 kHz) better-ear hearing thresholds for 92 infants in whom reliable audiometric hearing thresholds (representing the time of the visual reinforcement audiometry [VRA] test session) could be estimated. These figures were estimated from clinical records (i.e., not data collected for the study), typically using clinical VRA thresholds. Jittered red dots show individual participants, and the white dot shows the mean value. The distribution curve shows the probability density. Where no hearing threshold was obtained at the maximum test level, it was entered as maximum test level +20 dB.

CAEP Set-Up and Testing

CAEP Set-Up • CAEP data were collected using the Interacoustics Eclipse system (Interacoustics A/S, Denmark). Sound stimuli were routed from the Eclipse via an RDL TX-PA40D 20W power amplifier to an Eminence Alpha-6A 8-ohm loudspeaker attached to the booth wall. A single-channel EEG recording was made between the high forehead (FPz) and right mastoid, using the left mastoid as ground. The high forehead was justified because Munro et al. (2019) compared infant CAEP responses recorded at the high forehead and vertex, finding no effect of electrode location on CAEP signal-to-noise ratio (SNR), but finding increased electrode retention difficulties and test time in some hirsute infants with placement at the vertex. Online 1 to 100 Hz bandpass filtering (via the Eclipse interface) was applied. Infants were tested wearing their own hearing aids and earmoulds without modifying any settings. The infant sat on their caregiver's lap, 1.1 m in front of the loudspeaker. The testers were two experienced pediatric audiologists. The lead tester controlled stimulus presentation from the observation room outside the booth (via window and video link) and was able to pause acquisition as needed, for example if the infant was very vocal, sleepy, or unsettled. The second tester sat in front, and slightly to the side of the infant, keeping them alert with their attention facing forwards using a selection of silent toys on a table with a minimally reflective surface.

CAEP Stimuli • Stimuli were the mid frequency (MF) and mid-high frequency (HF) stimuli described in Stone et al. (2019). For simplicity in this context, the mid-high-frequency stimulus has been labelled HF, as it represents the upper end of hearing thresholds typically obtained for infants. Note that this is distinct from the high-frequency stimulus described in Stone et al. The MF stimulus was centered on 1.36 kHz and made with a harmonic structure to resemble a voiced consonant such as /g/. The HF stimulus was centered on 3.55 kHz and had an inharmonic structure to resemble an unvoiced consonant such as /t/. Both stimuli were two-thirds of an octave wide in bandwidth and 70 msec in duration[*] including 10 msec raised cosine onset/offset ramps. Erbograms of the two stimuli can be seen in Figure 2.

Stimulus levels were calibrated at the test position with the baby and caregiver absent. During calibration, the chair where the caregiver and infant sat was covered with soft material to avoid reflections from the vinyl surface affecting the calibration. Stimuli were calibrated in what will henceforth be referred to as dB speech reference level (dB SpRefL). This means that the band-limited stimuli were leveled to have the same power as that contained in the same bandwidth of a reference speech-spectrum, here the International Speech Test Signal (ISTS, Holube et al. 2010). The power in the MF band was −14.5 dB compared to the broadband power, and the power in the HF band was −20.6 dB compared to the broadband power. A 65 dB SpRefL stimulus, therefore, had the same average power as the equivalent frequency band in the ISTS signal at 65 dB SPL. A 65 dB SpRefL stimulus was, therefore, equivalent to 50.5 dB SPL for MF and 44.4 dB SPL for HF. In the majority of subjects, the initial CAEP presentation level was either 65 dB SpRefL (n = 66) or 75 dB SpRefL (n = 29). In a small number of subjects (n = 8), it was 65 dB SPL (note: not SpRefL), equating to 79.5 dB SpRefL for MF, and 85.6 dB SpRefL for HF.

---

[*]The stimuli were 70 msec in total duration, not in half-amplitude duration as mistakenly described in Stone et al. (2019).
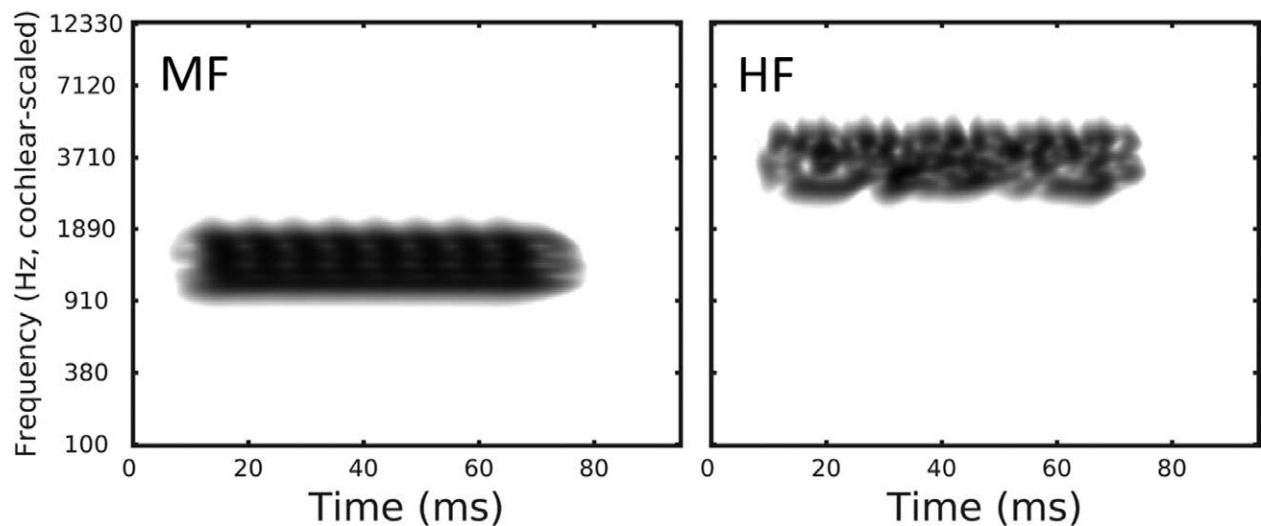
Fig. 2. Erbograms of the two test stimuli, as reported in Stone et al. (2019). The erbogram represents the time-course of the energy presented to the cochlea by the stimuli (after external and middle ear filtering).

Decisions were made as the study progressed to alter the initial level to ensure a wide range of SLs, starting with 65 dB SPL for the earliest test subjects, then 65 dB SpRefL, and then 75 dB SpRefL.

Aided CAEP Test Sessions • Otoscopy and tympanometry were performed, and hearing aid checks were carried out, including visual inspection, listening check, recording of hearing aid settings in manufacturer's software, noting stored data-logging, and recording ISTS coupler gain to inputs at 55, 65, 75, and 90 dB SPL. The MF and HF stimuli were presented in runs of 20 accepted epochs (stimulus presentations), with an online artefact rejection threshold of ±110 μV and presentation rate of 0.9 Hz. Block 1 consisted of eight runs of 20 accepted epochs, at the initial presentation level: four runs for each stimulus, with stimulus type interleaving between runs. Block 2 was a repeat of block 1, but with the stimulus order reversed. During the test session, the runs in each block were averaged yielding two averages of 80 epochs for each stimulus to give a visual estimation whether a response was likely to be present. For subsequent analysis, averages of all 160 epochs for each stimulus/test session were calculated. A summary of the CAEP test procedure can be found in Figure 3.

For each individual CAEP run (20 epochs), the tester who controlled the child's attention made a separate judgement of baby state, that is, how unsettled, sleepy, and vocal the infant had been, each of which was scored from 0 (not at all) to 3 (extremely). This was a subjective judgement by the tester. The scores were averaged over the runs to give an overall score from 0 to 3 in each of the three domains for the test session.

Completion of blocks 1 and 2 was classed as a successful test. If the baby was sufficiently settled, and the family had enough time, blocks 3 and 4 were also carried out, in which the level of
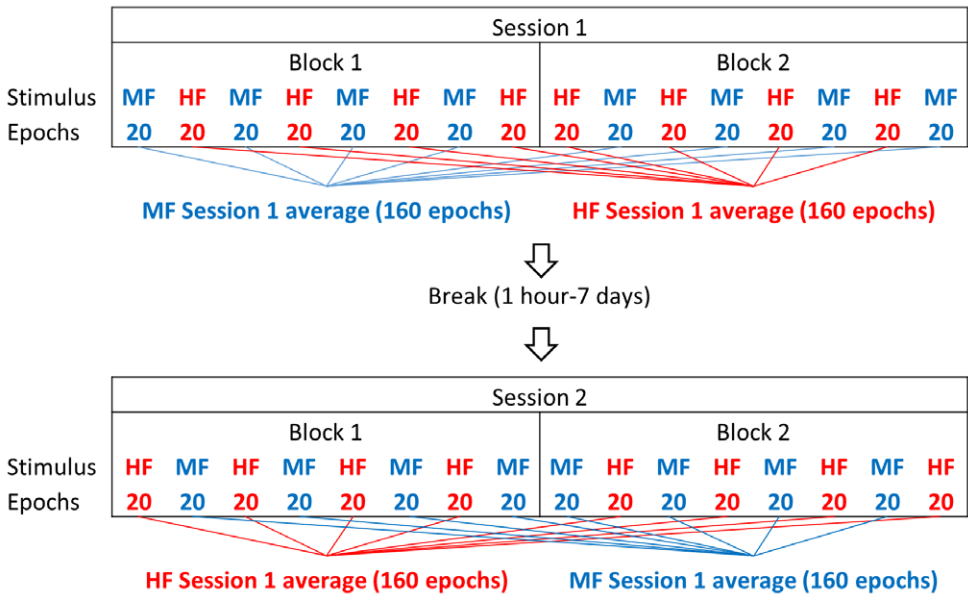

Fig. 3. A summary of the test methodology for cortical auditory evoked potential (CAEP) sessions, showing how two averages of 160 epochs were obtained for each stimulus across two sessions. This shows the process for an odd-numbered participant ID. For even-numbered participants, the stimulus order was reversed.

each stimulus was either reduced by 10 dB, if visual analysis of the waveforms showed a clear response, or increased by 10 dB if, visually, the response appeared absent or inconclusive. Test duration was recorded in terms of preparation time (time from beginning to prepare the infant for testing, i.e., preparing the skin and getting into position, to the start of the test) and testing time (time taken to complete the test protocol for the two stimuli at a single input level, i.e., blocks 1 and 2).

After the CAEP test sessions were completed, caregivers were emailed a link to a short Likert-scale-response online questionnaire asking about their experiences of the test (see Supplemental Digital Content 1, http://links.lww.com/EANDH/ B105). Caregivers who did not complete the questionnaire were invited to complete a paper version at the later, VRA, session. Some of these completed the questionnaire during the session, while others were invited to complete it and return it by post. Families were also invited to be interviewed by telephone about their experience of the test; the semistructured interview was carried out by a research audiologist not involved in the main study, and used an exploratory qualitative approach, encouraging caregivers to share their own story. Families participated in the interview based on their, and the research audiologist's availability, to the point that the research audiologist felt that data saturation was reached, in terms of themes identified. The interview guide is given in Supplemental Digital Content 2, http:// links.lww.com/EANDH/B106. Transcripts of the interviews were independently coded by authors A.S.V. and C.L.H. to identify themes emerging using NVivo 12 software. Responses were reported narratively.

VRA Set-Up and Testing

VRA Set-Up • As with the CAEP test, the MF and HF stimuli were calibrated in dB SpRefL at 1.1 m directly in front of the loudspeaker. Stimuli were presented via a Callisto audiometer (Interacoustics A/S, Denmark), routed through the same amplifier and loudspeaker as for the CAEP test. The repetition rate of the presented stimuli was 4 per second, chosen to ensure a suitably attention-grabbing stimulus (Van Dun et al. 2012).

VRA Test Sessions • Otoscopy, tympanometry, and hearing aid checks were carried out, as in the CAEP session. If the infant's hearing aid settings had changed since the CAEP test, or any fault was identified, temporary hearing aids were used for the VRA test, programmed to the infant's previous settings. The testers were two experienced pediatric audiologists, who carried out aided VRA to determine MRLs (i.e., the lowest level stimulus that elicits a reliable behavioral response) for the MF and HF stimuli. The British Society of Audiology's recommended procedure for VRA (BSA 2014) was followed as a guideline. Typically infants were conditioned at a high level, around 100 dB SpRefL, with simultaneous presentation of the visual reinforcer for two trials. Two head-turn responses (prior to the associated visual reinforcement) were then required before reducing the presentation level. Each MRL was given a reliability rating of good, satisfactory, or poor. In cases where the infant would not condition to auditory stimuli, conditioning was performed with a vibrotactile stimulus (40 dB HL 250 Hz warble tone presented by bone vibrator to a convenient position such as forehead, mastoid, hand, or foot) before returning to testing with auditory stimuli. At the end of the two separate VRA sessions, the tester noted what they thought was the overall most reliable

estimate of the MRL for each stimulus. This value was used in further analyses. At the time of VRA testing, the testers typically had some knowledge of the infant's hearing loss from the clinical history, and were able to review whether or not the CAEP response was visually detected as a clear response, but did not have access to objective CAEP detection results. In some cases, testing had to be delayed or repeated at a later date due to issues such as the infant not being developmentally ready for testing, unreliable responses recorded, persistent middle ear dysfunction/ear infections, faulty or missing hearing aids, or limited caregiver availability.

Data Analysis

To avoid bias in the data based on which cases were included/ excluded in the CAEP detection analysis, the data (audiometric thresholds, aided MRLs and associated reliability, tympanograms, etiology, and other known clinical factors) were presented blinded (i.e., not knowing the baby in question nor the CAEP results) to at least two experienced audiologists who analyzed the available information separately and then made a joint decision on whether the case should be excluded. Cases where abnormal tympanograms could affect the hearing at the CAEP or VRA sessions were excluded if this could affect whether or not the stimulus was audible. For example, an infant with abnormal tympanograms at both sessions was excluded due to the unknown impact of possible conductive overlay at each session. An infant with abnormal tympanograms at the CAEP session (i.e., a potential conductive overlay to an underlying loss) and normal tympanograms at the VRA session was excluded if the VRA MRL was less than or equal to the CAEP presentation level. This is because the VRA results suggest the CAEP stimulus would have been audible in the absence of middle ear dysfunction, but it is not known whether middle ear dysfunction at the CAEP session impacted audibility. Infants with abnormal tympanograms only at the VRA session, combined with VRA MRLs less than or equal to the CAEP presentation level were included. In such cases, the stimuli were truly audible at the CAEP session, although it is possible the obtained SL was a slight underestimation in these cases. Infants with confirmed or suspected progressive loss between the two test sessions were excluded.

CAEPs were analysed in Matlab (Mathworks, USA). For each infant, data were combined into a single ensemble of 160 epochs per session for each stimulus. Residual noise was estimated for the ensembles of 160 epochs by calculating the voltage standard error at each time point, then calculating the rms of the standard errors across time points. Waveform analysis was performed over the 0 to 500 msec time window following stimulus onset, calculating voltage means across nonoverlapping 50 msec time intervals. Objective analysis of the CAEP recordings was performed offline using the one-sample Hotellings T^2 (HT^2) test (Golding et al. 2009; Carter et al. 2010) to test whether amplitudes of the coherently averaged CAEP waveforms deviated significantly ($p < 0.05$) from zero. A 30-Hz low-pass filter was applied prior to carrying out the HT^2 analysis (in addition to the 1 to 100 Hz filtering applied by the Eclipse system). Results for alternative detection strategies, where responses from both initial and repeat sessions are considered, are presented and discussed later. This includes a Bonferroni correction for multiple comparisons, and an approach that reduces alpha (α, the

significance criterion) sequentially, as a possible method to use in clinics when repeat testing is selectively carried out. SL (or input SL) was calculated as CAEP presentation level minus MRL, both recorded in dB SPRefL in the sound field, to give the dB level above threshold of the CAEP stimulus in terms of input to the hearing aid. Percentage of CAEP detections was plotted against SL.

Test conditions were investigated for impact on CAEP detection. Mann–Whitney U tests were performed to determine whether there was a significant difference in recorded baby states (ratings of how unsettled, sleepy, or vocal they were) for cases where a CAEP was detected versus CAEP not detected, for audible stimuli. A t test was used to make the same comparison for residual noise. Participant characteristics were also investigated: t tests were used to determine whether there was a significant difference in age, hearing age (age since hearing aid fitting), or hearing aid use, for cases where a CAEP was or was not detected in at least one session (including only participants for whom the stimuli were audible).

RESULTS

CAEP Test Completion

All 103 infants successfully completed the initial CAEP test session, but one did not complete the repeat session due to crying and failing to settle. In one further case, testing was completed but data from a session was lost due to a technical error.

This gives an overall completion rate of >99%. Eighty four percent of infants underwent repeat testing on the same day, and the remainder between 1 and 7 days later.

CAEP Test Duration

Median test duration was 24 minutes for the initial test session (interquartile range 22 to 30 minutes). This included 9 minutes preparation (interquartile range 8 to 12 minutes) time and 15 minutes testing time (interquartile range 14 to 18 minutes). At the repeat session, median preparation time was slightly shorter at 7 minutes (interquartile range 5 to 8 minutes), and median testing time was unchanged.

Aided VRA Results and SL

Five of 103 babies were unavailable for follow-up behavioral testing using VRA. Of the remaining 98 infants, 91 underwent repeat VRA testing on the same day as the initial VRA test, and the remainder underwent repeat testing 1 day later. In 72% of cases, reliable MRLs were recorded at both sessions and in 17% of cases reliable MRLs were recorded in only one of the two sessions. In 6% of cases, reliable MRLs were not recorded, but audiometric thresholds (reported from the child's audiology clinic) revealed profound losses, consistent with the stimuli being inaudible, and MRLs were taken to be >105 or >110 dB SPRefL, that is, above the maximum presentation levels for the MF and HF stimuli, respectively. In the remaining 5% of cases, reliable MRLs were not obtained in the absence of profound
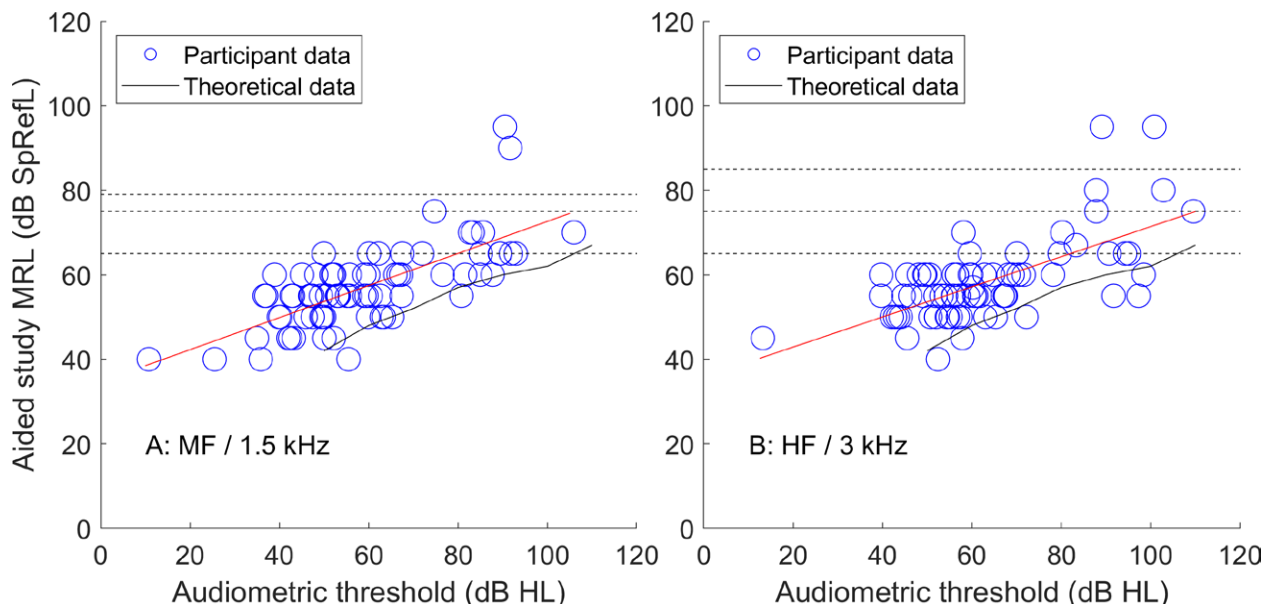

Fig. 4. Measured aided minimum response levels (MRLs) (dB speech reference level [SpRefL]) vs. clinically-determined better ear unaided (interpolated) audiometric thresholds at 1.5 kHz for the mid-frequency (MF) stimulus (n = 76) and 3 kHz for the (mid-)high-frequency (HF) stimulus (n = 75). The red lines show the linear regression plots for the patient data (MF, R^2 = 0.513; HF, R^2 = 0.428). The horizontal dashed lines show the different initial input levels used for cortical auditory evoked potential (CAEP) stimuli in the study. Of 98 cases tested using visual reinforcement audiometry (VRA), three were excluded from both MF and HF data due to inappropriate hearing aid configuration (one unaided, one bone conduction aid, one unilaterally aided with mild loss on unaided ear). Ten were excluded from the MF data and 11 from the HF data due to not having reliable responses for either audiometric thresholds or aided MRLs to the study stimuli. A further nine for the MF and nine for the HF frequency range had profound losses, and no aided response to the stimulus at maximum test levels; these data are also omitted from the figure. The audiometric threshold data is the best estimate of the infant's hearing at the time of VRA testing for the study, that is, clinical audiometric threshold data was used where possible with the same tympanometric configuration and close in age to when the study data were collected. The solid black line shows theoretical data, calculated using the DSL V child prescription for a given flat loss in dB HL (using Interacoustics Callisto software). The theoretical input level of the ISTS was varied within the software until the real-ear aided response (SPL output at the eardrum) was equal to the eardrum-referenced SPL hearing threshold at 1.5 kHz and 3 kHz, respectively. This is therefore the theoretical hearing threshold for a perfectly-fit hearing aid.

TABLE 2. Breakdown of cases excluded from the primary analysis to determine CAEP detectability

Main Exclusion Reason	Total Excluded	
	MF	HF
Tested unaided	1 (1%)	1 (1%)
Lost to follow up	4 (4%)	4 (4%)
Audibility unknown due to pattern of tympanograms at CAEP and VRA test	13 (13%)	13 (13%)
Progressive loss (positive SL)	6 (6%)	6 (6%)
Progressive loss (negative SL)	8 (8%)	9 (9%)
Bone conduction aid	1 (1%)	1 (1%)
Fluctuating hearing loss	1 (1%)	1 (1%)
Unreliable aided MRL	2 (2%)	3 (3%)
Total	36	38

Progressive loss refers to a significant, or likely significant, drop in hearing between the CAEP and VRA sessions, as judged by the panel of audiologists on reviewing the audiometric threshold data. Progressive losses have been split into positive and negative SLs (i.e., the estimated CAEP SL). This is because in cases where there was a progressive loss but positive SL, it is known that the CAEP stimulus would have been audible (though the true SL is unknown), whereas for those with a progressive loss and negative SL, the CAEP stimulus could have been audible at the time of CAEP testing when hearing may have been better.
CAEP, cortical auditory evoked potential; HF, (mid-)high-frequency (stimulus); MF, mid-frequency (stimulus); MRL, minimum response level; SL, sensation level; VRA, visual reinforcement audiometry.

loss. Median aided MRLs were 55 and 60 dB SpRefL for the MF and HF stimuli, respectively. Figure 4 shows the relationship between aided MRLs to the test stimuli and equivalent unaided audiometric threshold (reported from the child's audiology clinic and interpolated as necessary) including theoretical data based on DSL prescription targets. Of the cases where reliable MRLs were recorded in both sessions, the repeated MRLs differed from the initial MRLs by ≤5 dB for 95% of the cases and by 10 or 15 dB for the remaining 5%. Median SLs (CAEP presentation level minus MRL) were 10 dB for both stimuli.

CAEP Response Detection: Included Cases

In total, 67 MF cases and 65 HF cases were included in the primary CAEP analysis. While 5 cases were lost to follow-up before the VRA session, one of these was still included as review of the audiogram and hearing aid gain confirmed that the CAEP stimuli would have been inaudible. The reasons for excluding cases are summarized in Table 2.

CAEP Waveforms

Figures 5A–C show exemplar sets of waveforms for three participants for whom both signals were audible at the initial test level (15 to 20 dB SL) and for whom all responses were highly significant ($p < 0.003$) according to HT[2]. Significant variability is observed in the waveforms both between and, especially in Figure 5A, within participants. The lack of consistency in evoked responses is clearly evident from the very low amplitude grand average of all CAEP waveforms at the initial test level with SL ≥0 dB (Fig. 5D).

CAEP Response Detection

CAEP response detection at the starting stimulus level was analysed in two ways: First, treating sessions 1 and 2 as separate, hence doubling the effective number of CAEP recordings available for analysis. Second, data were analyzed according to

whether a response was present in at least one of the two sessions ($p < 0.05$ in either session), that is, replicating what may happen clinically if no response was seen on an initial test. Figure 6A shows the percentage of detected responses versus SL with the initial and repeat sessions treated as independent data points. The detection rate increases with increasing SL (from SL 0 to 10 to SL >10 dB) and is higher for MF than HF. The overall detection rate (across both stimuli) was 70% for >10 dB SL (80% for MF and 60% for HF), and 49% for 0 to 10 dB SL (54% for MF and 43% for HF). When combining all responses ≥0 dB SL, the detection rate across both stimuli was 62% (70% for MF and 54% for HF). An appropriately low false positive rate (FPR) of 1/44 (2.3%) was observed using the <0 dB SL recordings.

Figure 6B shows the same data but considering a significant response from the initial session or the repeat session to indicate a detection, thus showing higher overall detection rates. The overall detection rate was 87% for >10 dB SL (94% for MF and 79% for HF), and 65% for 0 to 10 dB SL (70% for MF and 60% for HF). When combining all responses ≥0 dB SL, the detection rate across both stimuli was 78% (84% for MF and 72% for HF). The FPR was 1/22 (4.5%). Figure 7 shows the SL and corresponding significance level of CAEP responses from individual sessions. See figure caption for further details.

Table 3 shows the repeatability across sessions of CAEP detection using a criterion of $p < 0.05$. Overall, in 33% of cases where the stimulus was audible (36 of 110), a different outcome was found between sessions 1 and 2, despite no obvious change in the test conditions (further discussed later). For inaudible stimuli, the CAEP outcome was repeatable (i.e., a true negative response) 95% of the time (21 of 22 cases).

Table 4 summarizes the sensitivity of the CAEP test for different analysis conditions. The measured FPR using the current dataset is given, but due to the small number of cases in which the CAEP stimuli were <0 dB SL, a theoretical (taking into account multiple comparisons) FPR is also given. The first three rows summarize the data from Figure 6. Note that the theoretical FPR for the "initial or repeat HT[2] tests, single stimulus" condition was calculated using $\alpha + \alpha(1 - \alpha)$. To clarify, an α-level of 0.05 is expected to give 5% false-positives at the initial test session, and an additional 4.75% at the repeat session (i.e., 5% of the remaining 95%). By extension, for the "Initial and repeat HT[2] tests, either stimulus" condition where a total of four tests were carried out, and the theoretical FPR equals 18.5%. Rows 4 and 5 show the performance of the test taking a reduced α level to obtain more acceptable FPRs. Row 4 maintains an overall α level of ~5% whereas row 5 uses different sequential α levels, with more stringent requirements for detection in the repeat session, which is more akin to a potential clinical approach when retesting is an option. Under no test condition did the measured FPR exceed the theoretical FPR. The benefits of repeat testing (i.e., increased test sensitivities) are largely maintained even when using the adjusted α-levels.

CAEP Response Detection at Additional Input Level

Of the nine (16%) infants showing no response to MF at either session, despite the stimulus being audible, six went on to complete testing at a higher test level (+10 dB) in at least one of the sessions. Of the 15 (28%) with no response to HF at either session, despite the stimulus being audible, five completed testing at the higher test level. The results from these cases at the
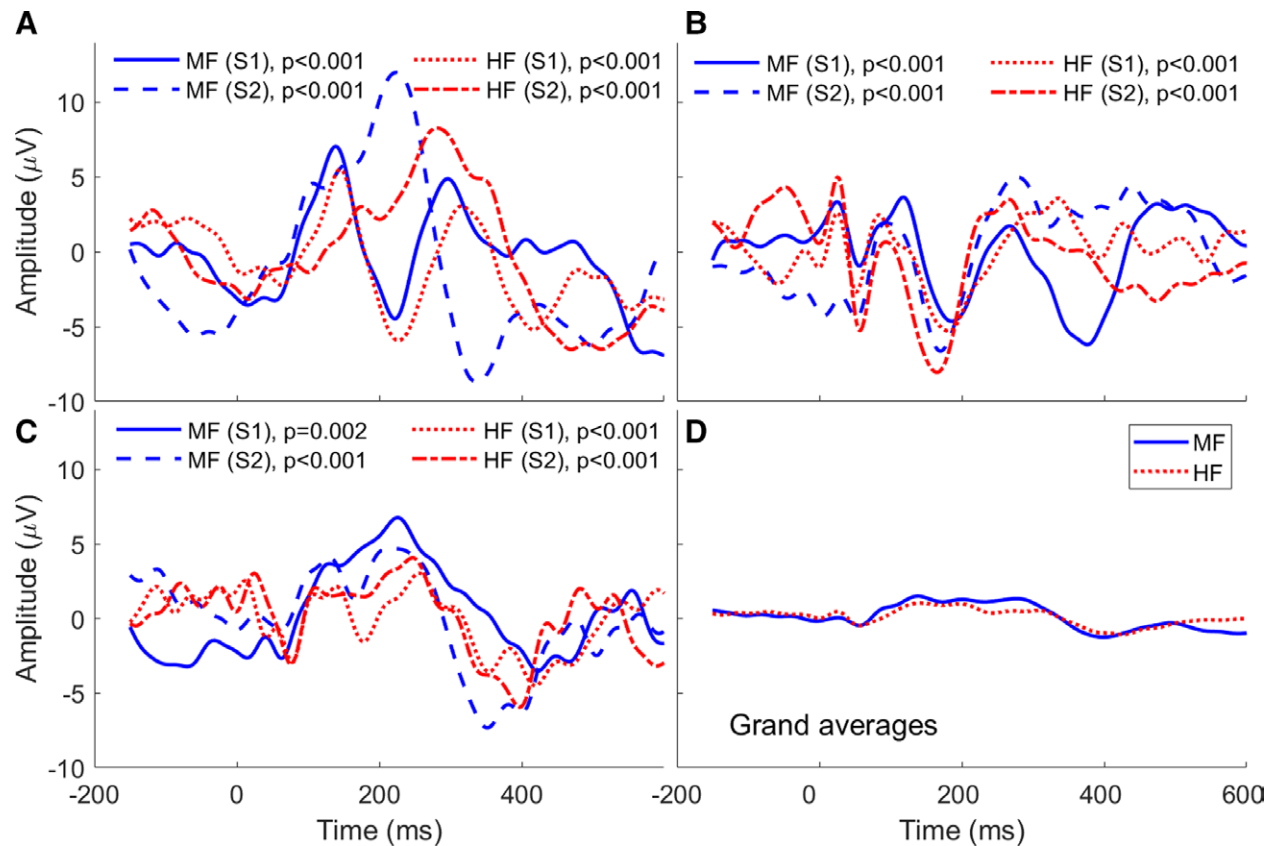
Fig. 5. Example cortical auditory evoked potential (CAEP) waveforms for three participants, at the initial test level, showing consistently significant CAEP detections (A–C), for whom stimuli were audible in the 15 to 20 dB SL range. Results are shown for session 1 (S1) and session 2 (S2), with mid-frequency (MF) responses in blue and (mid-)high-frequency (HF) response in red. The *p* values are the result of the Hotellings T² (HT²) test. D, This shows the grand average of all CAEP waveforms at the initial test level with SL ≥0 dB.

higher test level are summarized in Table 5. In two out of six cases for MF and five out of five cases for HF (denoted by **), highly significant, often repeatable, *p* values of <0.004 would give a clinician high confidence in these true-positive results, despite the impact of performing multiple tests.

Relationship Between CAEP Detections and Test Conditions or Participant Characteristics

Differences in test conditions were investigated for all audible stimuli at the initial test level (taking different sessions as unique datapoints), grouped by CAEP response
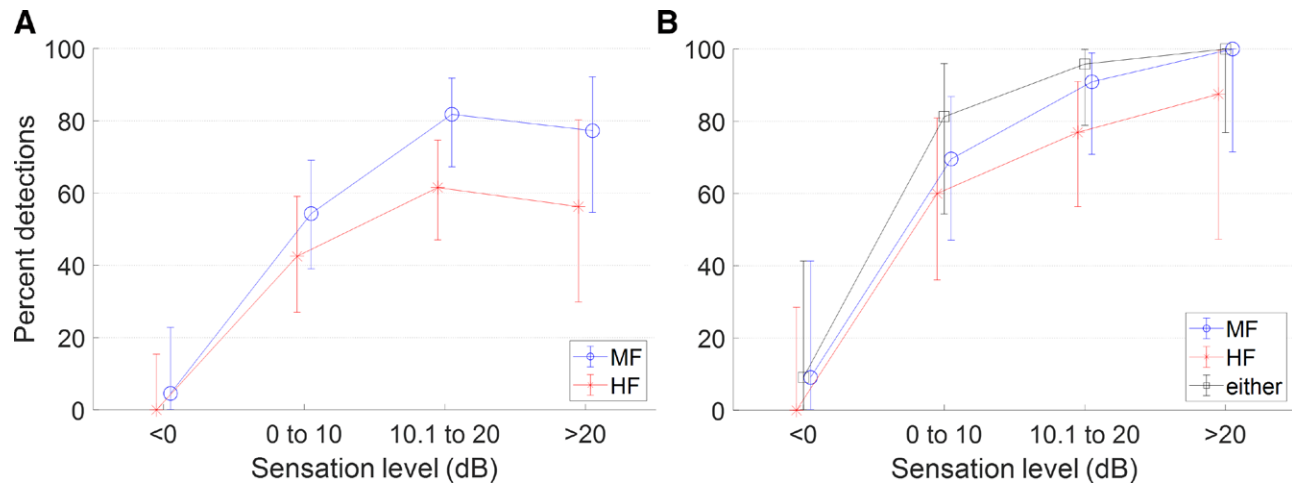

Fig. 6. Cortical auditory evoked potential (CAEP) detections vs. sensation level (SL). A, This shows detections when treating the initial and repeat sessions as separate data points. B, This shows detections that were significant in at least one of the two test sessions. "Either" in 6B shows cases where a response was present for at least one stimulus and session. The higher of the two stimulus SLs was used in this case. A significance criterion of *p* < 0.05 was used throughout. The total number of participants represented in each SL category from <0 to >20, respectively, was as follows: 11, 23, 22, 11 for mid-frequency (MF) and 11, 20, 26, 8 for (mid-)high-frequency HF (stimulus). Error bars show 95% confidence intervals calculated from the binomial distribution.
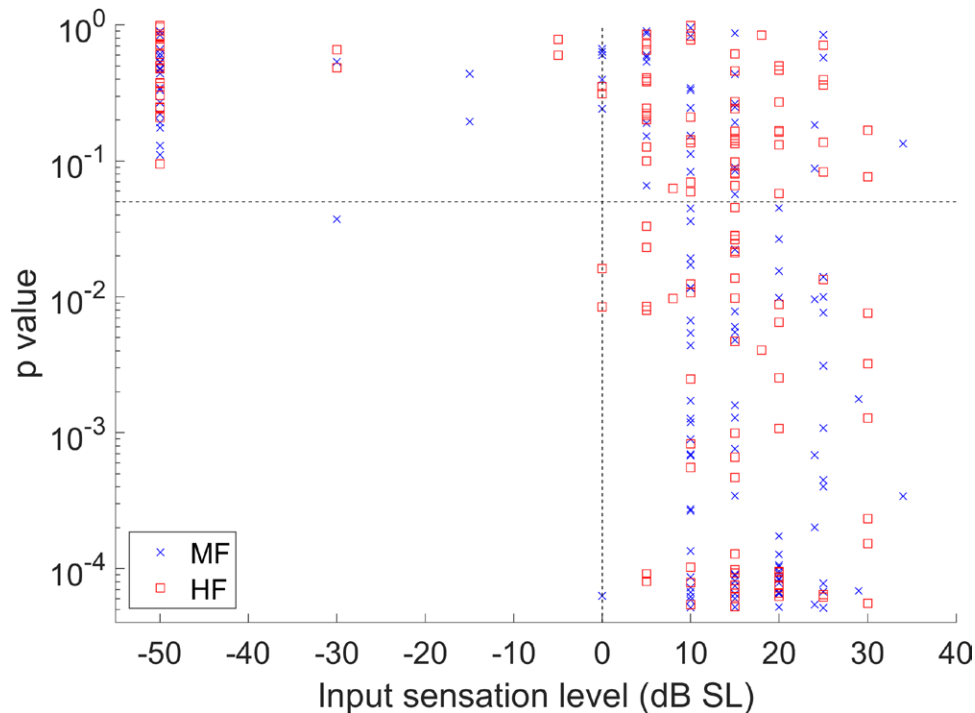
Fig. 7. The *p*-value vs. sensation level for mid-frequency (MF) (blue, n = 134) and (mid-) high-frequency (HF) (red, n = 130) stimuli. The dashed horizontal line shows the *p* = 0.05 significance criterion, and the dashed vertical line shows 0 dB sensation level (SL). The bottom left quadrant shows false positives, bottom right true positives, top left true negatives, and top right false negatives. The *p*-values were capped at a lower limit of 10^{-4}. A small amount of vertical jitter was added to the capped points for visualization purposes. Subjects with SL shown as <−40 dB had no recordable aided minimum response level (MRL): the exact point at which they appear on the figure is arbitrary.

detected (n = 136) and CAEP response not detected (n = 84). Results from the Mann–Whitney *U* test (comparing the CAEP detected cases to the CAEP not detected cases) showed no significant differences in unsettled scores (*p* = 0.494) or sleepy scores (*p* = 0.346). A significant difference was found in vocal scores (*p* = 0.007), with a lower median score for vocality for cases where a CAEP was detected (median = 0.19) than not detected (median = 0.32). There was no significant difference between residual noise recorded in cases

where the CAEP was detected and the CAEP was not detected (*p* = 0.429).

The tester was able to pause data collection as needed (i.e., if the baby was too unsettled, sleepy, vocal, or any other circumstance). Data collection was paused during the initial test level in 55% of cases where a CAEP was detected and 65% of cases where a CAEP was not detected. A χ^2 test showed no significant relationship between CAEP detection and the need to pause the recording (*p* = 0.130).

TABLE 3. Repeatability of CAEP detections between sessions 1 (S1) and 2 (S2)

		3A: Audible Stimuli (SL ≥0 dB)				
		S1: CAEP Detected?				
		MF			HF	
		Yes	No		Yes	No
S2: CAEP detected?	Yes	31 (55%)	10 (18%)	Yes	19 (35%)	12 (22%)
	No	6 (11%)	9 (16%)	No	8 (15%)	15 (28%)
		3B: Inaudible stimuli (SL <0 dB)				
		S1: CAEP detected?				
		MF			HF	
		Yes	No		Yes	No
S2: CAEP detected?	Yes	0 (0%)	1 (9%)	Yes	0 (0%)	0 (0 %)
	No	0 (0%)	10 (91%)	No	0 (0%)	11 (100%)

3A shows data for stimuli with SLs ≥0 dB, and 3B shows data for SLs <0 dB. Each table has 2 separate 2×2 sections, representing repeatability data for each stimulus. The numbers show how many participants fall in each category.
CAEP, cortical auditory evoked potential; HF, (mid-)high-frequency (stimulus); MF, mid-frequency (stimulus); SL, sensation level.

TABLE 4. Summary of sensitivity and FPR of CAEP tests for different analysis methods

		FPR			Sensitivity		
					Overall (≥ 0	0 to 10 dB	>10
Test Description	α	Theoretical	Measured	Stimulus	dB SL)	SL	dB SL
Single HT^2 test, single stimulus (Fig. 6A)	0.05	5.0%	4.5%	MF	69.6%	54.3%	80.3%
			0.0%	HF	53.7%	42.5%	60.3%
Initial or repeat HT^2 tests, single stimulus (Fig. 6B)	0.05	9.8%	9.1%	MF	83.9%	69.6%	93.9%
			0.0%	HF	72.2%	60.0%	79.4%
Initial and repeat HT^2 tests, either stimulus (Fig. 6B, black line)	0.05	18.5%	9.1%	Either	92.6%	81.3%	97.4%
Initial or repeat HT^2 tests, single stimulus (adjusted *p* value on both tests)	0.025	4.9%	0.0%	MF	82.1%	65.2%	93.9%
			0.0%	HF	66.7%	60.0%	70.6%
Initial or repeat HT^2 tests, single stimulus (adjusted *p* value on second test)	0.05 then 0.025	7.4%	0.0%	MF	82.1%	65.2%	93.9%
			0.0%	HF	70.4%	60.0%	76.5%

The theoretical FPR has been calculated based on the number of tests and α level, as described in the main text.
CAEP, cortical auditory evoked potential; FPR, false positive rate; HF, (mid-)high-frequency (stimulus); HT^2, Hotellings T^2; MF, mid-frequency (stimulus); SL, sensation level.

EEG noise was also calculated per-epoch (the rms voltage over all time points within each epoch), and CAEP detections were recalculated using the HT^2 test, but excluding the noisiest 5% of epochs in each block of 160 runs. For cases where the stimulus was audible at the initial test level, this resulted in n = 141 CAEP detections and n = 79 nondetections (where previously it had been n = 136 detections and n = 84 nondetections). Excluding the noisiest 5% of epochs resulted in n = 8 cases showing significant detections that had not previously been detected, and n = 3 cases showing no significant detection, where previously it had been significant (in all cases for dB SL ≥0). The overall impact of excluding the noisiest 5% of epochs on whether a CAEP was detected at either session was minimal; for MF one infant did show a response who did not previously while another did not show a response who did previously. For HF, two infants showed responses that did not previously. There was no change in the FPR. Therefore, overall sensitivity for SL ≥0 dB remained the same for MF, and for HF increased from 72% (39/54) to 76% (41/54).

Differences in participant characteristics were investigated for all audible stimuli at the initial test level (taking a significant response from either test session), grouped by CAEP response detected (n = 86) and CAEP response not detected (n = 24). There was no evidence that age, hearing age, nor daily hours hearing aid use differed between participants who did and did not show a CAEP response in at least one session when

the stimulus was audible (*t* test, *p* > 0.05). Data logging was not available for every participant; hence, this was evaluated on a smaller sample (n = 72 detected; n = 22 not detected).

Caregiver Acceptability

The caregiver acceptability questionnaire was completed by 85 caregivers. Table 6 shows median scores were 1 to 2 for all questions, showing good overall acceptability.

Caregiver Interviews

Eighteen caregivers took part in semistructured telephone interviews to share their experience of the CAEP test. There were no significant discouraging responses from the caregivers in relation to the CAEP test procedures. All caregivers reported positive or highly positive experiences of the test, and individual comments related to the experience of the test were overwhelmingly positive (*"I think it was brilliant, it was easy, they were great…"; "there was nothing stressful about the test whatsoever"*). Slightly negative comments were occasionally made relating to electrode placement, but these were qualified showing this was not particularly problematic overall (*"Sometimes she didn't like it, but if she didn't like it she would cry for two seconds and then it goes away," "…he didn't like them to begin with, but I just sort of settled him, and then he was absolutely fine. He forgot they were there"*).

TABLE 5. HT^2 CAEP detection *p*-values at the higher test level for infants with no response at the initial test level

		p (MF)			*p* (HF)	
Case (Infant)	MF SL (dB)	Session 1	Session 2	HF SL (dB)	Session 1	Session 2
A	25	<0.001**	<0.001**	15	0.858	0.002**
B	10	0.807	x	15	0.001**	<0.001**
C				20	0.098	<0.001**
D	15	<0.001**	0.004**			
E	15	0.818	x			
F	25	0.320	0.310			
G	15	x	0.302			
H				30	0.001**	0.716
I				30	x	0.001**

The sensation level refers to the higher test level (i.e., the initial sensation level would have been 10 dB lower). "x" refers to a case where the test was not completed at the higher level in that session. Blank cells indicate that the infant showed a significant response at the initial level.
**Indicates highly significant responses (p < 0.004).
CAEP, cortical auditory evoked potential; HF, (mid-)high-frequency (stimulus); HT^2, Hotellings T^2; MF, mid-frequency (stimulus); SL, sensation level.

TABLE 6. Scores from caregiver acceptability survey (n=85)

Question	Scale	Mean (SD)	Median (Range)
1. The information I was given about the hearing test before my baby was tested was...	1 = very good; 7 = not good at all	1.32 (0.86)	1 (1–5)
2. The test procedure made me feel*...	1 = not anxious at all; 7 = very anxious	1.53 (1.08)	1 (1–6)
3. During the hearing test, my baby appeared to be...	1 = very happy; 7 = very unhappy	2.1 (1.6)	2 (1–6)
4. Compared to other tests and procedures that my baby has experienced, tolerating the hearing test seemed*...	1 = not difficult at all; 7 = extremely difficult	1.86 (1.01)	2 (1–5)
5. Keeping my baby awake and quiet during the hearing test was*...	1 = not difficult at all; 7 = extremely difficult	2.63 (1.48)	2 (1–7)
6. Seeing the tester attach the recording sensors onto my baby's head made me...	1 = not worried at all; 7 = extremely worried	1.39 (0.87)	1 (1–5)
7. The test environment was...	1 = very pleasant; 7 = very unpleasant	1.68 (1.18)	1 (1–7)

Lower scores, on a scale of 1–7, reflect more positive experiences.
*These items were presented in the questionnaire in reverse order compared to the other items (i.e., negative responses appearing to the left and positive responses appearing to the right).
See Supplemental Digital Content 1, http://links.lww.com/EANDH/B105, for a copy of the questionnaire.

The same was true for occasional negative comments about babies becoming unsettled during the test ("*He did [enjoy it] for a while and then he started I think because it was the same thing I think getting a bit agitated but he wasn't like really whinging*"). Caregivers were very happy with the information given to them about the test ("*they just explained things really, really well*"; "*I understood what they were doing, they were very clear in what they doing*") although several expressed a desire to receive results from the test, which was not part of the protocol ("*…thought that I'd be able to get some kind of more specific results on her hearing on the different frequencies, but I understand that's not going to happen*"). Caregivers felt the test would translate well into a clinical environment, and would have value ("*I think they could benefit from testing like that in the hospital*"). Many commented on the benefits of the infant being awake during the test ("*So compared to, like, his previous hearing tests I thought this was a good idea because he didn't have to be asleep for it*") or on the possibility of tests being available that could be done with the infant asleep or awake ("*I think up at the hospital they should actually have a test that if a baby is asleep they'll do it or if a child is awake they'll do it*"). Caregivers were overwhelmingly happy with the van as a test environment in terms of space ("*It's brilliant, I didn't realise there was so much in one wee van, to be honest. I think it looked really good inside as well, it looked really professional. It wasn't just, come to our van, and there's nothing really there, it was clean and smart and professional-looking, so I did like it*") and convenience ("*Obviously, it was helpful that they came in the van because I didn't have to go up to the hospital, which can be anywhere between half an hour and an hour and a half drive*"). In terms of advice for other caregivers, comments were made related to ensuring baby had had a good sleep and feed before the test, and the option to bring baby's own toys ("*I would advise all the parents to feed the kid and let them have a little bit of a nap*"; "*Maybe take a toy that they're particularly interested in*").

CAEP Response Detection: Alternative Inclusion Parameters and Data Visualization

The criteria for inclusion in the CAEP analysis were chosen to ensure the sample was not too greatly reduced while giving confidence that the audibility of the stimulus was known, with only small potential deviations from the true SL. When more restrictive inclusion criteria were applied (accepting only participants with normal tympanograms at both CAEP and VRA sessions in at least one ear, that being the equal or better ear in terms of clinical thresholds), the overall pattern of responses remained very similar (Figure 1, Supplemental Digital Content, http://links.lww.com/EANDH/B107). Data are also visualized in terms of aided study MRL, audiometric threshold, CAEP test level, and presence/absence of CAEP response (Figure 2, Supplemental Digital Content, http://links.lww.com/EANDH/B107). The visualized cases with CAEP nondetections are discussed further in the Supplemental Material. Finally, Figure 3 in Supplemental Digital Content, http://links.lww.com/EANDH/B107, shows the calculated input level expected to achieve 20 dB SL for a well-fit hearing aid. Implications for appropriate test level are discussed in Supplemental Digital Content 3, http://links.lww.com/EANDH/B107.

DISCUSSION

This article addresses a gap-in-knowledge by measuring CAEP sensitivity in a large group of infant hearing aid users of target age 3 to 7 months, including repeat testing. The results are promising for use of aided CAEPs as part of a clinical test battery, showing that the infant aided CAEP test is feasible, acceptable, and reaches high levels of sensitivity when repeat testing is carried out, especially for the MF signal at SLs >10 dB. Both feasibility and caregiver acceptability were demonstrated in infant hearing aid users to a similar degree to that found for infants with normal hearing (Munro et al. 2019). As such, the aided CAEP may have a place for motivating caregivers to continue with hearing aid use when CAEP responses can be demonstrated, and indicating alternative intervention when responses are consistently absent.

Aided CAEP Sensitivity

In line with previous studies, the results show a large number of CAEP nondetections for audible stimuli presented in a single test session. Taking results from a single test session, for stimuli with an input SL of 0 to 10 dB, CAEPs were detected, on average, in 49% of cases (54% for MF and 43% for HF) and for stimuli with input SL >10 dB, CAEPs were detected, on average, in 70% of cases (80% for the MF stimulus and 60% for the HF stimulus). These averages are similar to summary results reported in

Gardner-Berry et al. (2016) for the same SL ranges, additionally showing poorer detection for the higher frequency stimulus. In this regard, it is worth recalling that the presentation levels in the present study were considerably lower than those used by Gardner-Berry et al. (although both report results in terms of SLs). A pattern of increasing CAEP detections for increasing SLs was shown, similar to previous reports in infants with hearing loss (Chang et al. 2012; Van Dun et al. 2012; Gardner-Berry et al. 2016) and in infants with normal hearing (Cone & Whitaker 2013).

An important aspect of the study was repetition of the CAEP test with the aim to understand in what proportion of cases a CAEP could be detected on retest where it had not been detected initially, despite acceptable recording conditions. The results from this study demonstrate the importance of retesting cases of nondetection, even when test conditions appear good. In particular, accepting detections from either initial or repeat sessions, the detection rate (for SLs >10 dB) increased from 70% to 87% average (94% MF and 79% for HF). For SLs ≥0 dB, repeat testing increased detection from 62% (70% for MF and 54% for HF) to 78% (84% for MF and 72% for HF). In other words, in 42% of the cases where the stimulus was audible, but the CAEP was not detected at the initial test, the CAEP was detected on retest. For SLs >10 dB, the figure was 57%, and for SLs of 0 to 10 dB, the figure was 31%. This benefit of retesting was largely maintained using reduced α-levels to account for multiple comparisons. The effect of repeat runs on the response detection rate suggests significant variability in responses in the individual.

The reason for the response variability is not entirely clear. Significant variability in CAEP response characteristics between infant subjects has been well-documented (e.g., Wunderlich et al. 2006), although variability within subjects has been less thoroughly reported. There was a significant effect of how vocal the infant was determined to be, with more vocal infants more likely to be associated with lack of CAEP response detection. Vocalizations could impact response detection by masking the target sound, but also potentially by instigating CAEP responses that add noise, and by increasing EMG noise due to facial movement. There was no effect of how sleepy or unsettled the baby was, but it should be noted that these judgements of baby state were subjective; hence, do not give a full and accurate indication of baby state. While it was possible to pause the CAEP data acquisition during periods when infants were not in a good state for testing, there was a delay on the pause being actioned by the equipment, which made it difficult to capture quiet, and eliminate noisy, periods for babies who were vocalizing intermittently. Testers should be aware that infant vocalizations will mean less likelihood of a significant CAEP detection; they should choose appropriate play to keep infants in a quiet state, and not over-interpret negative results, particularly where a baby was vocal during testing. A further possibility is that future EEG recording devices include a microphone near the patient to monitor noise levels and use this information to exclude trials where the signal/response would have likely been affected by acoustic noise. However, median scores for vocality were low in both CAEP present and CAEP absent groups, suggesting vocality does not fully explain absent responses. Results comparing cases where a CAEP was and was not detected showed no effect of residual noise in the data. Hence, the level of residual noise is not a sufficient criterion to identify inconclusive cases. There was also no discernible effect of how unsettled or sleepy the baby was on the rate of detection, although low scores in these domains show babies were mostly settled and alert.

The very low amplitude grand average waveforms for audible stimuli (Fig. 5D) are an indication of the variability in waveform morphology across subjects in the current data. Although the grand averages are of small amplitude, the waveform peak/trough latencies are similar to those seen in other studies such as Munro et al. (2019), though the current dominant peak is considerably broader and smaller in amplitude. Significant variability is expected in young infants, especially those with unaided hearing loss in the first weeks of life, as the auditory system is still very immature, and waveforms may not yet have developed typical morphologies. Grand average waveform amplitudes for infants with hearing loss in Chang et al. (2012) were in the region of 3 to 4 µV: a little higher than those seen here. This could in part be due to the different stimuli used, the present stimuli being narrower in bandwidth and lower in overall level than the equivalent HearLab stimuli. The mean better-ear hearing threshold in the present study (67 dB HL) was slightly worse than that in Chang et al. (62 dB HL), which may also account for some of the difference. Munro et al. (2019) reported grand averages in the region of 6 to 10 µV when using the HearLab stimuli on infants with normal hearing.

The observation in our data of fewer detections to HF than MF could be related to immaturity of the auditory system. For infants with more pronounced high frequency losses than low frequency losses, the high frequency auditory pathways may remain under-developed for longer, having had less access to sound over the lifetime. It may also reflect the tonotopic organization of the cortex. Wunderlich et al. (2006) showed CAEP response amplitudes for the dominant positive peak in newborns, children, and adults (all with normal hearing) were smaller for high than low frequency tones (presented at the same dB HL). This was attributed to early tonotopic organisation of parts of the cortex, with high frequency generators being deeper within the cortex, resulting in smaller responses at the scalp. Munro et al. (2019) showed fewer responses in normal hearing infants for the equivalent HF stimulus (/t/, 92% detections) than for the equivalent MF stimulus (/g/, 100% detections). A simple frequency-based argument may be insufficient, however, since the fewest detections were seen for the lowest frequency sound (/m/, 86% detections). This is consistent with Ponton et al.'s (1992) finding that developmental changes occurred faster and mature function was attained earlier from the mid-frequency region compared with the highest or lowest frequency conditions based on derived response ABR.

In our data analyses, an appropriately low FPR was seen, in line with the 5% value set in the HT^2 calculation. Some published data suggests that infant CAEP responses may sometimes be recordable at levels below the behavioral response (Cone & Whitaker 2013). The work by Cone and Whitaker included behavioral hearing test data from infants aged <7 months, in whom MRLs are raised compared to infants developmentally ready to perform VRA, which is a likely source of variation between the methods in both studies.

Clinical Feasibility

This study demonstrated a CAEP completion rate of >99%. It should be noted that the testers tended to have only one patient

booked in per day, and were visiting them by their homes, so this excellent completion rate is based on a higher degree of flexibility than would be realistic at a clinic appointment. For example, if the baby was asleep, the testers could wait for them to wake up, and if the baby became unsettled or sleepy during the session but before the CAEP test began, the caregiver could take them back to the house until they were in a better state for testing.

Median test duration was clinically feasible at 24 minutes in the initial test session, including patient preparation and test time. Test duration is expected to be shorter in a clinical scenario with software designed for clinical use, as opposed to the research implementation used here, which required altering settings at the start of each run. Duration, completion, and caregiver acceptability all demonstrate the potential clinical applicability of the test.

Caregiver Acceptability

The results from the caregiver acceptability questionnaire and interviews were positive. Mean scores on all aspects of the questionnaire were low, indicative of positive experiences. The highest scoring (least positive experience) question related to: "Keeping my baby awake and quiet during the hearing test," which was a challenge in some cases. This is in line with results reported on experiences of the CAEP test from caregivers of infants with normal hearing (Munro et al. 2019). This further demonstrates the potential benefit of developing an automated system for rejecting epochs where recording conditions are compromised, for example by acoustic noise. It may also be plausible and desirable to reject epochs based on EEG patterns indicative of sleep. These are potential areas for future research. The positive remarks from interviews with caregivers highlight that they felt that this test was not a stressful testing experience and could be beneficial, and hence has a place in pediatric audiology practice.

Limitations

There are several limitations to the present study, which are important in interpreting the findings. In particular, the reader should note how several study limitations impact our reporting of SL. One limitation is the significant incidence of otitis media with effusion in the age group of interest, coupled with the fact that the infants were seen at two different time points. This means that infants with any abnormal tympanograms may have had hearing thresholds affected by different degrees of conductive overlay at the two time points. To avoid having to exclude the majority of test data, cases with abnormal tympanograms were included, so long as any pattern of potential conductive overlay did not preclude categorizing the CAEP stimulus as audible or inaudible. This means that in some cases the true SL of the CAEP stimuli may have been underestimated. Figure 1B in Supplemental Digital Content 3, http://links.lww.com/EANDH/B107, shows CAEP detections only for infants with normal tympanograms in their better ear at both sessions. The overall data pattern is not significantly changed, which provides reassurance that the approach used is valid. For clinicians performing aided CAEP, testing would not need to exclude infants with flat tympanograms as in the current study (this was only done so as to make a valid comparison between the test sessions). CAEP testing may be of value for

infants with flat tympanograms: absent cortical responses may be a consequence of the conductive overlay, whereas present responses would confirm audibility. CAEPs, therefore, could still help indicate, alongside other clinical testing and observations, whether speech sounds are audible or further action may be warranted.

A second limitation is that SL is reported in terms of input to the hearing aid (input SL), that is, no adjustment was made for hearing aid compression or change in infant ear canal characteristics between the two test sessions. For SLs >0 dB, hearing aid compression will cause the SL measured at the input of the hearing aid to overestimate the SL actually available to the child, whenever the CAEP presentation level exceeds the hearing aid compression threshold. When both the CAEP presentation level and the MRL exceed the compression threshold, and the hearing aid has fast compression, the SL actually experienced by the child will equal the input SL divided by the compression ratio. That is, the true SL would be lower than the input SL. The effect of changing ear canal acoustics is likely to be only very small in the frequency regions of interest here. Predicted real-ear-to-coupler-differences according to the DSL formula at 1.5 kHz are 11 to 12 dB at 3 to 7 months and 9 to 10 dB at 8 to 24 months (read from the Callisto software [Interacoustics A/S, Denmark], for an insert tip and HA2 coupler). At 3 kHz, predicted real-ear-to-coupler-differences are 11 to 13 dB at 3 to 7 months and 9 to 11 dB at 8 to 24 months. We therefore expect the effect of changing ear canal size to be only around 2 dB. The direction of the effect would be such that we may very slightly underestimate SL. The SL was calculated using CAEP input level minus MRL. MRLs are not true thresholds, as infants require a higher threshold to elicit a behavioral response than would an adult (Parry et al. 2003). This would also have the effect of slightly underestimating the true SL. Figure 4 shows that the aided MRLs were above the theoretical thresholds using the DSL formula, which is in line with the expectation of MRLs being elevated compared to true thresholds. To summarize the limitations regarding SL, factors that could lead to underestimation of SL in our design include inclusion of some infants with abnormal tympanograms at the VRA session; changes in ear canal acoustics between the CAEP and VRA sessions, infant MRLs being typically higher than an equivalent true threshold. A factor that could lead to overestimation of SL in our design is hearing aid compression. It should also be noted that while VRA constitutes the best available method for behavioral hearing assessment in infants, the technique is inherently a subjective one and can be prone to some bias or inaccuracies (Baldwin et al. 2010). This could be a contributing factor in lack of CAEP responses when the signal is apparently audible.

Another possible limitation of the study was that no attempt was made to adjust the input test level based on the audiometric thresholds of the infant. The rationale was that the aim of a good hearing aid fitting is to optimize speech audibility for a given hearing loss, hence wanting an assessment procedure that measures audibility of conversational speech. However, the study data (Fig. 6) show only around 50% detections for SLs of 0 to 10 dB. Figure 7 shows particularly few detections at 0 to 5 dB SL. In a clinical protocol, it may be appropriate to test at a higher input level than the 65 dB SpRefL, which was the default for most of the current participants, and potentially to test at different input levels for different degrees of hearing

loss. This approach would reach higher SLs, but would ideally maintain input levels representative of common speech listening. For infants, common speech listening levels may be higher than those for adults, as caregivers will tend to direct speech toward infants, speak from a short distance, and use exaggerated speech. One approach could be to aim to test at input SLs of around ≥15 dB, to give an excellent chance of seeing a detection in at least one stimulus, with test repetition. Figure 3 in Supplemental Digital Content 3, http://links.lww.com/EANDH/B107, plots one possible derivation of this data, suggesting 65 dB SpRefL would be a suitable input level for mild losses, while 75 SpRefL dB would be more suitable for losses of 50 to 60 dB HL, and 85 dB SpRefL more suitable for losses of 70 dB HL or greater. For losses over 80 dB HL, it is likely to be difficult to achieve the desired SL: CAEP responses may still be observed, but detection rates are likely to be lower. Theoretical questions arise as to what SL would be acceptable for a given input level and hearing loss, and what speech input level gives the audiologist useful information about everyday speech audibility. In light of the current data, 75 dB SpRefL may be a more appropriate default choice for test level than 65 dB SpRefL, which often yields low SLs and hence more variable responses.

It is not known what the most efficient stimulus presentation rate is for infant CAEP detection. While many studies have used presentation rates similar to the 0.9 Hz used in this study, others have used slower presentation rates, to reduce refractory behavior and hence increase response amplitude, an effect that may be particularly salient in infants (Wunderlich et al. 2006; Cone & Whitaker 2013). Further research directly addressing this issue in infants would be of value to develop a clinically efficient test, and could lead to improved detection rates.

It is possible that the short duration (70 msec) sounds used in this study resulted in higher behavioral hearing thresholds (and hence lower SLs) than would have been seen for longer duration sounds, due to temporal integration in the auditory system (Florentine et al. 1988; Gerken et al. 1990). However, preliminary investigations on adults in our laboratories have not shown any systematic differences in threshold for the individual 70 msec stimuli, for stimulus trains at 4 Hz as used in VRA, nor for the equivalent longer duration (500 msec) stimuli. Further work is underway to define the reference equivalent threshold sound pressure levels for these stimuli in adults. In some cases, the effect of duration on threshold in infants may be larger than that in adults (Berg 1991). Any potential effect of temporal integration does not invalidate the findings. However, an increased understanding of this relationship will aid interpretation of the CAEP test. It may, for example, explain a portion of the offset between expected aided thresholds in an idealized hearing aid fitting and those observed (Fig. 4).

The rationale of the study was to use aided CAEPs to confirm physiological detection of sound stimuli at the level of the cortex. It should be noted that this does not additionally confirm that hearing aid fittings are matching prescription targets. CAEPs as described are also not suitable for demonstrating discrimination between different stimuli (Billings et al. 2012). Acoustic change complexes, derived from a cortical response, have been shown to demonstrate speech discrimination in adults (Cheek & Cone 2020) and infants (Cone 2015; Cone et al. 2022), and would hence be the more appropriate choice for investigating discrimination.

Clinical Applications

Overall, the data are promising for clinical application of the CAEP test, although clinicians must be aware of the strengths and limitations, and interpret results appropriately. In particular, the significant variability in response detection necessitates caution around interpreting negative results without a repetition. Overall, better detection rates were observed for MF than HF stimuli, so an absent response at HF may be less informative than an absent response at MF. Demonstrating aided CAEP responses may be a useful tool to encourage caregivers to persist with infant hearing aid use, which can often be challenging (e.g., Walker et al. 2013; Muñoz et al. 2015; Caballero et al. 2017; Muñoz et al. 2019; Visram et al. 2021). Using CAEP results to inform hearing aid adjustments should be just one tool, and a decision to adjust gain should be made in conjunction with the full clinical picture, understanding the possibility of false negative CAEP results, especially for HF. While the present study followed a rigid protocol, clinical protocols would have more flexibility, and this would likely increase test efficiency. For example, a criterion for an early stop could be set if an appropriately significant result were reached after a suitable number of presentations. Individual attention to each case would help clinicians to interpret results, for example, if the baby were very vocal, the clinician may choose not to pursue testing, or not to over-interpret a negative result.

When responses were collated from all tests carried out across both sessions, only two infants showed consistent nondetections when both stimuli were audible. In one case, the test input level was 65 dB SpRefL and achieved 0 and 5 dB SL for MF and HF, respectively. Testing was not fully completed at the higher level, but one of the two usual blocks of testing for the higher level was completed, and this did show a significant response (on an 80 epoch average rather than 160 epoch average as has been used elsewhere) for the MF ($p = 0.004$, SL of 10 dB). This baby had a profound bilateral hearing loss (average threshold 100 dB HL) and has now received cochlear implants. In the second case, the test level was 65 dB SpRefL and achieved 10 and 5 dB SL for MF and HF, respectively. No testing was completed at the higher level. This baby had a moderate loss (57 dB HL average threshold 1 to 4 kHz). In both cases, the clinical protocol suggested above would have led to testing at an initial higher input level for these cases. The circumstances around the cases of nondetection gives reassurance that a consistent picture of nondetections, across both stimuli, and with appropriate test levels and test repetition, does indicate lack of audibility, and clinically would support alternate intervention such as referral for cochlear implants. Earlier referral for cochlear implantation is beneficial to allow sufficient time for assessment by cochlear implant services and enable earlier implantation so as to maximize outcomes.

CONCLUSIONS

Results from a large sample of infant hearing aid users in the target age range of 3 to 7 months suggest that aided CAEPs can be a valuable tool for assessing audibility related to conversational speech, and supplement clinical testing for infants not developmentally ready to perform behavioral hearing testing. Aided CAEPs can provide additional clinical information for infants for whom there are few other reliable clinical hearing tests. The addition of a repeat test when responses are not detected is important because this reduces false negatives (for SL >10 dB) from 20%

to 6% for MF and from 40% to 21% for HF. Consistent nonde-tections across both stimuli with appropriate test levels and rep-etition does indicate lack of audibility and hence would support intervention such as cochlear implant referral. The test is clini-cally feasible and acceptable to caregivers.

REFERENCES

Baldwin, S. M., Gajewski, B. J., Widen, J. E. (2010). An evaluation of the cross-check principle using visual reinforcement audiometry, otoacoustic emissions, and tympanometry. *J Am Acad Audiol*, *21*, 187–196.

Berg, K. M. (1991). Auditory temporal summation in infants and adults: effects of stimulus bandwidth and masking noise. *Percept Psychophys*, *50*, 314–320.

Billings, C. J., Papesh, M. A., Penman, T. M., Baltzell, L. S., Gallun, F. J. (2012). Clinical use of aided cortical auditory evoked potentials as a measure of physiological detection or physiological discrimination. *Int J Otolaryngol*, *2012*, 365752.

BSA. (2014). *British Society of Audiology, Recommended Procedure, Visual reinforcement Audiometry*.

Caballero, A., Muñoz, K., White, K., Nelson, L., Domenech-Rodriguez, M., Twohig, M. (2017). Pediatric hearing aid management: challenges among hispanic families. *J Am Acad Audiol*, *28*, 718–730.

Carter, L., Golding, M., Dillon, H., Seymour, J. (2010). The detection of infant cortical auditory evoked potentials (CAEPs) using statistical and visual detection techniques. *J Am Acad Audiol*, *21*, 347–356.

Chang, H. W., Dillon, H., Carter, L., van Dun, B., Young, S. T. (2012). The relationship between cortical auditory evoked potential (CAEP) detec-tion and estimated audibility in infants with sensorineural hearing loss. *Int J Audiol*, *51*, 663–670.

Cheek, D., & Cone, B. (2020). Evidence of vowel discrimination provided by the acoustic change complex. *Ear Hear*, *41*, 855–867.

Cone, B. K., Smith, S., Cheek, S. D. E. (2022). Acoustic change complex and visually reinforced infant speech discrimination measures of vowel contrast detection. *Ear Hear*, *43*, 531–544.

Cone, B., & Whitaker, R. (2013). Dynamics of infant cortical auditory evoked potentials (CAEPs) for tone and speech tokens. *Int J Pediatr Otorhinolaryngol*, *77*, 1162–1173.

Cone, B. K. (2015). Infant cortical electrophysiology and perception of vowel contrasts. *Int J Psychophysiol*, *95*, 65–76.

Florentine, M., Fastl, H., Buus, S. (1988). Temporal integration in normal hearing, cochlear impairment, and impairment simulated by masking. *J Acoust Soc Am*, *84*, 195–203.

Gardner-Berry, K., Chang, H., Ching, T. Y. C., Hou, S. (2016). Detection rates of cortical auditory evoked potentials at different sensation levels in infants with sensory/neural hearing loss and auditory neuropathy spec-trum disorder. *Semin Hear*, *37*, 53–61.

Gerken, G. M., Bhat, V. K. H., Hutchison-Clutter, M. (1990). Auditory tem-poral integration and the power function model. *J Acoust Soc Am*, *88*, 767–778.

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*, 833–842.

Holube, I., Fredelake, S., Vlaming, M., Kollmeier, B. (2010). Development and analysis of an International Speech Test Signal (ISTS). *Int J Audiol*, *49*, 891–903.

King, A., Carter, L., Van Dun, B., Zhang, V., Pearce, W., & Ching, T. (2014). *Australian Hearing Aided Cortical Evoked Potentials Protocols. April*, 1–13.

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.

Mehta, K., Watkin, P., Baldwin, M., Marriage, J., Mahon, M., Vickers, D. (2017). Role of cortical auditory evoked potentials in reducing the age at hearing aid fitting in children with hearing loss identified by newborn hearing screening. *Trends Hear*, *21*, 2331216517744094.

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

Moore, B. C., Stone, M. A., Füllgrabe, C., Glasberg, B. R., Puria, S. (2008). Spectro-temporal characteristics of speech at high frequencies, and the potential for restoration of audibility to people with mild-to-moderate hearing loss. *Ear Hear*, *29*, 907–922.

Muñoz, K., Olson, W. A., Twohig, M. P., Preston, E., Blaiser, K., White, K. R. (2015). Pediatric hearing aid use: parent-reported challenges. *Ear Hear*, *36*, 279–287.

Munoz, K. F., Larsen, M., Nelson, L., Yoho, S. E., Twohig, M. P. (2019). Pediatric amplification management: parent experiences monitoring chil-dren's aided hearing. *J early hear detect interv*, *4*, 73–82.

Munro, K. J., Purdy, S. C., Uus, K., Visram, A., Ward, R., Bruce, I. A., Marsden, A., Stone, M. A., Van Dun, B. (2020). Recording obliga-tory cortical auditory evoked potentials in infants. *Ear Hear*, *41*, 630–639.

Parry, G., Hacking, C., Bamford, J. (2003). Day J. Minimal response lev-els for visual reinforcement audiometry in infants. *Int J Audiol*, *42*, 413–417.

Ponton, C. W., Winkelaar, R., Eggermont, J. J., Coupland, R. W. (1992). Frequency-specic maturation of the eighth nerve and brainstem auditory pathway: evidence from derived Auditory BrainStem Responses (Abrs). *J Acoust Soc Am*, *91*, 1576–1586.

Punch, S., Van Dun, B., King, A., Carter, L., Pearce, W. (2016). Clinical experience of using cortical auditory evoked potentials in the treatment of infant hearing loss in Australia. *Semin Hear*, *37*, 36–52.

Rance, G., Beer, D. E., Cone-Wesson, B., Shepherd, R. K., Dowell, R. C., King, A. M., Rickards, F. W., Clark, G. M. (1999). Clinical findings for a group of infants and young children with auditory neuropathy. *Ear Hear*, *20*, 238–252.

Stone, M. A., Visram, A., Harte, J. M., Munro, K. J. (2019). A Set of Time-and-Frequency-Localized Short-Duration Speech-Like Stimuli for Assessing Hearing-Aid Performance via Cortical Auditory-Evoked Potentials. *Trends Hear*, *23*, 2331216519885568.

Tsiakpini, L., Weichbold, V., Kuehn-Inacker, H., Coninx, F., D'Haese, P., & Almadin, S. (2004). *LittlEARS Auditory Questionnaire*. Innsbruck, Austria: MED-EL.

Van Dun, B., Carter, L., Dillon, H. (2012). Sensitivity of cortical auditory evoked potential detection for hearing-impaired infants in response to short speech sounds. *Audiol Res*, *2*, e13.

Visram, A. S., Roughley, A. J., Hudson, C. L., Purdy, S. C., Munro, K. J. (2021). Longitudinal changes in hearing aid use and hearing aid management challenges in infants. *Ear Hear*, *42*, 961–972.

Walker, E. A., Spratford, M., Moeller, M. P., Oleson, J., Ou, H., Roush, P., Jacobs, S. (2013). Predictors of hearing aid use time in children with mild-to-severe hearing loss. *Lang Speech Hear Serv Sch*, *44*, 73–88. Erratum in: *Lang Speech Hear Serv Sch*. 2015 Jan;*46*(1):64.

Wood, S. A., Sutton, G. J., Davis, A. C. (2015). Performance and characteristics of the Newborn Hearing Screening Programme in England: The first seven years. *Int J Audiol*, *54*, 353–358.

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