P300 Versus MMN: The Clinical Potential of ERP Components to Assess Auditory Discrimination Abilities in Cochlear Implant Users

Rosanne Abrahamse[1]
Supervisors: Vitória Piai[1,4], Margreet Langereis[2,3], Anneke Vermeulen[2,3]

[1]Radboud University Nijmegen, Donders Centre for Cognition, Nijmegen, the Netherlands, [2]Radboud University Medical Centre, Donders Institute, dept. Hearing and Implants, Nijmegen, the Netherlands, [3]Radboud University Medical Centre, dept. of Otorhinolaryngology, Nijmegen, the Netherlands, [4]Radboud University Medical Centre, dept. of Medical Psychology, Nijmegen, the Netherlands

The current study was designed as a starting point in developing an electrophysiological marker of speech perception abilities in cochlear implant users. Two event-related potentials (ERPs), the Mismatch Negativity response (MMN) and the P300 response were compared in their ability to assess auditory discrimination abilities in prelingually deaf adolescent cochlear implant users (n = 8) and normal-hearing controls (n = 14). The ERPs were compared in terms of their robustness on an individual level, with an equally limited amount of stimuli in each condition. A frequency contrast (500 Hz vs. 1000 Hz tone) and a consonant contrast (/ba/ vs. /da/ syllable) were used as stimuli. The P300 response, as opposed to the MMN response, was elicited in all individuals in both contrast conditions and, therefore, was deemed the most robust ERP of the two. Further analyses on differences in amplitude and latency of the P300 response as a function of group and/or contrast condition yielded significantly longer latencies for the consonant as opposed to the frequency contrast condition. It is suggested that the absence of group differences can be ascribed to a ceiling effect in the auditory discrimination abilities of the cochlear implant group. The relations between amplitude of the P300 in response to the consonant contrast, behavioural speech perception scores and duration of deafness indicate that the amplitude of the P300 has the potential to objectively inform us about speech perception abilities of cochlear implant users, as well as about the development of the auditory cortex after implantation. Future research can focus on measuring the P300 response in younger cochlear implant users, as well as measuring the P300 response with more complex input.

Keywords: P300, MMN, cochlear implantation, discrimination, speech perception

Corresponding author: Rosanne Abrahamse, **E-mail:** abrahamseros@gmail.com

Cochlear implants (CIs) are devices used to provide profoundly deaf children and adults with better hearing function. These devices surpass the impaired cochlea and restore the hearing pathway towards the auditory nerve. Although the use of these devices considerably improves speech perception, the speech perception abilities of CI users remain limited compared to the speech perception abilities of normal-hearing children and adults (American Speech-Language-Hearing Association, ASHA, 2004). This is a challenge in itself, but the implant outcomes also vary greatly per individual. While some implant users seem to perform almost equivalently to normal-hearing peers on speech perception measures, others perform considerably below average (2004; Beynon, Snik & van den Broek, 2002; Pisoni & Geers, 2000).

Several factors have been shown to play a role in the variable speech perception outcomes of CI users, such as age at implantation (Pisoni & Geers, 2000; Ruffin, Kronenberger, Colson, Henning & Pisoni, 2013), communication mode (Geers, 2002; Pisoni & Geers, 2000; Ruffin et al., 2013) and IQ (Geers, 2002). Despite these examples, a large proportion of variability remains unexplained. Clinically, the need to detect the variability early in development is high. The earlier clinicians know that children are t risk for non-optimal speech perception outcomes, the sooner appropriate interventions can be applied.

In addition to speech perception problems, research shows non-normal and variable performance in linguistic domains among CI users (de Hoog, Langereis, Weerdenburg, Knoors & Verhoeven, 2016a; 2016b; Pisoni & Geers, 2000; Schorr, Roth & Fox, 2008; Svirsky, Robbins, Kirk, Pisoni & Miyamoto, 2000). Interestingly, these linguistic problems are not always directly related to speech perception problems (de Hoog et al., 2016a). If the origin of speech perception problems in CI users can be more properly located, it may be possible to either relate these to, or differentiate these from, the existing variability in linguistic performance. This way, appropriate interventions can be designed for each type of problem.

Differences in the central auditory processing function of the CI users provide a possible explanation for the observed variability in speech perception outcomes (Groenen, Beynon, Snik, Broek, 2001; de Hoog et al., 2016a; Kraus et al., 1993; Näätänen, Paavilainen, Rinne & Alho, 2007; Pisoni & Geers., 2000). These differences can be due to having experienced a period of auditory deprivation, or, when deafness occurs at a younger age or congenitally, to an overall immature auditory system. The central auditory processing function of the brain can be measured using electro-encephalography (EEG). This may offer a means to clinically detect the individual variation in speech perception outcomes after implantation. It provides insight into the development of the auditory cortex and it has been linked to subjective speech perception outcomes (Groenen et al., 2001; Kelly, Purdy & Thorne, 2005). Furthermore, it provides the objectiveness that is needed to assess the perception abilities of very young children.

The current study will firstly assess the clinical potential of two electrophysiological approaches to measure the central auditory processing function of CI users. To be suitable for the clinic, the measures should be robust on an individual level, their acquisition should have an appropriately short duration, and the measure should have task requirements that fit the attention span of young children. The two EEG-approaches will be tested in an adolescent population (a normal-hearing group, and a group of prelingually deaf CI users). Starting out with longer EEG-recordings (as a result of testing two methods instead of one) in an older, more flexible population, provides the opportunity to choose the best method for future research. This way, young children can in the future be exposed to less demanding tasks. As a secondary aim, the relation between the central processing function of the CI users, measured using EEG, and their speech perception scores as measured in the clinic, will be explored. This will be the first step in establishing a link between the objective measurement and the subjective scores.

Marking speech perception abilities in cochlear implant users: two approaches

The electrophysiological approaches we decided to focus on are the P300, or P3b component (Polich, 1987), and the mismatch negativity (MMN) component (Näätänen et al., 2007). These are both late event-related potentials (ERPs) that can be measured using EEG. Late ERPs are proposed to reflect the central auditory processing function of the brain. The MMN and P300 components, for example, appear when the brain performs auditory discrimination of two stimuli (Johnson, 2009). Auditory processing can be elicited by means of an auditory oddball paradigm, in which a participant hears a stream of frequent standard stimuli (for example 80% of the time) that are randomly alternated

with infrequent deviant stimuli (for example 20% of the time). If auditory discrimination is performed, it is reflected in the difference between the averaged brain responses to the standard stimuli and the averaged brain responses to the deviant stimuli.

The MMN response requires no attention from the participant and is said to reflect how accurately the auditory sensory memory substrate of the brain can perform lower-level discrimination, based on perceptual stimuli characteristics (Näätänen, 2001). Auditory MMN response is a negative deflection in amplitude around 150-250 ms, which is observed over fronto-central regions of the brain. The P300 response is said to reflect a more conscious, higher-level cognitive process. Each new stimulus is evaluated against a model of the earlier one held in working memory. If a change in stimulus is detected, the model is updated. Besides perceptual discrimination, attention to the stimuli is required for the updating of the model (Polich, 2012). The auditory P300 response is characterized by a positive deflection in amplitude around 300 ms. It is often observed over centro-parietal areas of the brain (Johnson, 2009).

Speech perception problems may be specific to certain speech contrasts, and abilities may differ for speech as opposed to non-speech stimuli. Both the MMN and the P300 components vary in amplitude and latency with respect to the input they are given. This effect has been found for intensity contrasts (e.g., 80 dB stimulus vs. 90 dB stimulus), frequency contrasts (e.g., 500 Hz pure tone vs 1000 Hz pure tone) and speech-sound contrasts (e.g., consonant contrasts /ba/ vs. /da/ or vowel contrasts /i/ vs. /a/). More difficult contrasts lead to longer latencies and altered amplitudes (MMN: see Näätänen et al., 2007 for a review; P300: Polich, 1987; see Polich, 2004 for a review). Furthermore, differences are also evident across conditions, with more complex stimuli (speech-sound contrasts) yielding longer latencies and altered amplitudes as opposed to simple stimuli (frequency contrasts; MMN: Näätänen et al., 2007; P300: Polich, 2004). The MMN and the P300 components thus give the opportunity to investigate not only auditory processing in general, but to also distinguish between responses to complex as opposed to simple stimuli. With respect to future clinical marker abilities, varying input stimuli may lead to finding out which ones are more (or less) sensitive predictors for speech perception.

Earlier findings on the P300 and MMN components in cochlear implant users

In terms of task-requirements, the MMN is optimal to measure auditory processing in critical populations like infants and severely disabled people, because it does not require attention. The P300 requires attention and is therefore less attractive for this purpose. However, the task requirements for measuring the P300 are sufficiently low (participants are instructed to count the deviant in their heads or press a button when hearing the deviant stimulus), to be suitable for children from ages 3-4 onwards (Johnson, 2009).

Both components have been shown to be measurable on a group level and on an individual level in both (pre- and postlingually deaf) children and adults with cochlear implants (MMN: Kileny, Boerst & Zwolan, 1997; Kraus et al., 1993; Ponton et al., 2000; Singh, Liasis, Rajput, To & Luxon, 2004; Turgeon, Lazzouni, Lepore & Ellemberg, 2014; Watson, Titterington, Henry & Toner, 2007; P300: Beynon et al., 2002; Beynon, Snik, Stegeman & van den Broek, 2005; Groenen et al., 2001; Jordan et al., 1997; Kileny, 1991; Micco et al., 1995). In some studies, differences in the CI user ERPs compared to the normal-hearing control ERPs appeared, such as a prolonged latency (MMN: Turgeon et al., 2014; P300: Beynon et al., 2005) or a different amplitude (MMN: Ponton et al., 2000; Watson et al., 2007; P300: Beynon et al., 2005).

Furthermore, both components are relatively sensitive in distinguishing between well-performing and poor-performing users, on a group level and on an individual level. The ERPs of well-performing CI users are similar to that of normal-hearing controls, while the ERPs of poor-performing CI users (categorised as such due to low behavioural speech-perception scores or a below average subjective discrimination of the stimuli) are often found to be absent or different (MMN: Kraus et al., 1993; Singh et al., 2004; Turgeon et al., 2014; P300: Beynon et al., 2002; Groenen et al., 2001; Jordan et al., 1997; Kileny, 1991; Micco et al., 1995).

In terms of contrast conditions, there is a trend for more different responses as conditions become more complex. For the MMN component, longer latencies were found for increasing complexity of conditions. That is, the speech-sound contrast condition yielded longer latencies than the frequency or loudness conditions (Kileny et al., 1997). For the P300 component, well-performing CI users showed a P300 only when hearing a frequency and a vowel contrast (Beynon et al., 2005; Groenen et al., 2001). When hearing a consonant contrast, the P300 was absent in a significant number of participants (Beynon et al., 2005; Groenen et al.,

2001). Furthermore, a poor-performing group of CI users in another study again showed only a P300 for the frequency contrast, albeit with a prolonged latency. The well-performing group in this study showed a P300 for the consonant contrast and performed therefore similar to the normal-hearing control group. Vowel contrasts were not addressed in this study (Beynon et al., 2002). On the basis of this research it is expected that the more complex consonant contrast condition may show more differences in robustness among CI users.

Interestingly, the simpler conditions seem to be best for predicting behavioural speech perception, although research is scarce and results are inconsistent. For the MMN, in studies with frequency and vowel contrasts, duration of the component is found to correlate with perception scores (Kelly et al., 2005; Kileny et al., 1997), while in studies with consonant contrasts, amplitude is found to correlate with perception scores (Turgeon et al., 2014). For the P300, a relation between amplitude and perception scores was found, again only in the frequency and the vowel conditions, not in the consonant condition (Groenen et al., 2001).

Direct comparisons of the two approaches are scarce. One study combined both approaches in three CI-participants and three participants without hearing problems. The ERPs were identified in both the inattentive and the attentive paradigm. Although the study does not address the clinical potential of the components in particular, it stresses that both play an important role in the comprehension of the central auditory processing function (Obuchi, Harashima & Shiroma., 2012).

The next step in finding a suitable clinical marker

All of these results are quite promising in terms of the clinical potential of the MMN and P300 response to indicate speech perception difficulties. There are, however, a few shortcomings of these studies. Firstly, most studies mentioned above are outdated. Several aspects of cochlear implantation have improved during the past 15 years. Children now are more often, as well as earlier in life, eligible for an implant. The adolescents that were chosen to participate in this study were some of the first from this 'new generation' of less strict implantation eligibility. There is a need for replicating older findings under these contemporary circumstances. Secondly, a number of studies had small sample sizes (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991; Obuchi et al., 2012), or did not compare their results to a group of normal-hearing controls (Kileny et al., 1997). The latter can hamper interpretation of the results. In the current study, sample size remains a problem. Still, eight CI users were tested, whereas the studies mentioned above drew their conclusions on only half the amount of participants.

Thirdly, analysis techniques that were used to determine the presence of the ERP in individual waveforms were not consistent over studies. The early studies all chose to manually determine the amplitude and latency of the response, making use of the difference between the response to the standard stimuli and the response to the deviant stimuli (the difference waveform). Statistical analysis was consequently performed over the manually determined amplitude and latency values in some, but not all studies (e.g., Kileny et al., 1997 did not perform statistical analysis). This manual method is subjective and prone to bias (e.g., Kilner, 2013). Only a few attempts have been made to make the analysis more objective (Ponton et al., 2000, and for healthy subject data see Bishop & Hardiman, 2010). We applied a statistical procedure (non-parametric cluster-based permutation tests, see Maris & Oosterveld, 2007) to our EEG-data to identify MMN or P300 responses. By doing this, we aimed to make ERP analyses more objective and reliable. In addition, not many studies give notice of the existence of CI artefacts. Some do and approach the problem by rejecting artefact above 100 mV (Singh et al., 2004) and 50 mV (Kelly et al., 2005). Another used a semi-automatic procedure to attenuate them (Turgeon et al., 2014). In our study, we paid specific attention to developing an efficient procedure for attenuating artefacts.

Lastly, almost no studies have compared the MMN and the P300 responses directly in one design and thus, in terms of clinical utility not much has been concluded. Comparisons are needed in order to evaluate which approach has potential to develop into a clinical marker for speech perception abilities in CI users. An advantage of the MMN compared to the P300 is that it can be measured in younger populations. However, an advantage of the P300 compared to the MMN is that the P300 response has been detected with only 12 minutes of recording (Beynon et al., 2002; 2005; Groenen et al., 2001; Kileny, 1991; Micco et al., 1995). The shorter a measurement takes to yield robust results, the more advantageous this is for the clinic. The MMN response has been detected using 25 minute EEG-recordings (Kraus et al., 1993), but there are also experiments which lasted 35 to 40 minutes (Singh et al., 2004; Turgeon et al., 2014). Interestingly, in

one study (Obuchi et al., 2012) the P300 as well as the MMN response were elicited with only four minute recordings. Here, the MMN paradigm still yielded valuable results when restrictions were imposed on duration of the measurements. If this result can be replicated, the MMN might be perfectly discernible using only a limited amount of stimuli, and perhaps a better candidate for becoming a clinical marker than the P300.

Aims and objectives

This study focused on comparing the robustness of the MMN and the P300 response to measure auditory discrimination, with an equal limited amount of stimuli data (10 min EEG-recordings), on an individual level. Robustness was defined on the basis of how many individuals showed a statistically significant amplitude difference in their neurophysiological responses to the standard as opposed to the deviant stimuli. Results were obtained for two contrast conditions: a frequency contrast (500 vs. 1000 Hz tones) and a consonant contrast (/ba/ vs. /da/ syllables). EEG was measured in 14 normal-hearing participants, and in 8 prelingually deaf young-adult CI users. In the conditions (ERP component x contrast conditions) where the individual responses were most robust, the ERPs of a matched sample (n = 8) of normal-hearing participants were compared to the CI user ERPs. For those ERPs, group differences and within-group variation was assessed using the mean amplitude and latency measure. Furthermore, the effect of contrast conditions (frequency vs. consonant contrast) was addressed for both groups to be able to evaluate differences in simple (e.g., frequency contrast) vs. complex processing (e.g., consonant contrast). Lastly, behavioural speech perception as measured in the clinic and duration of deafness were related to P300 amplitude to explore whether the P300 would be a suitable marker for speech perception abilities, and whether this suitability differs for different input contrasts.

Methods

Participants

Eight Dutch adolescent prelingually deaf CI users (M_{age} = 19.9, ranging from 16-25; 6 males) were recruited for the ERP measurements through the otolaryngology department of the Radboudumc in Nijmegen, the Netherlands. All of the adolescents had profound bilateral hearing loss. Exclusion criteria consisted of having an IQ < 85, having a developmental or neurological disorder, or having had any serious head-trauma in the past. Table 1 describes the characteristics of the CI users. All participants used the same implant processor (Cochlear™ Nucleus®), and none of them used any additional hearing aids. For the participants with bilateral implants, EEG recording was done using only one implant. These users were allowed to choose on which implant (left or right) they wanted to be tested. Participation was on a voluntary basis and the participants received a monetary reward for their participation of 20 euros in vouchers.

A control group of 14 Dutch normal-hearing participants was also tested (M_{age} = 21.4, ranging from 18-25; 6 males). All of them had no history of hearing problems or speech/language problems. Furthermore, they had no psychiatric or neurological disorders. We restricted the educational levels of the normal-hearing participants to level 6 (out of 7), according to the Dutch neuropsychological educational level coding (Hendriks, Kessels, Gorissen, Schmand & Duits, 2014). This was done to achieve a more proper matching of the two participant groups. We recruited 11 participants with an educational level of 6, three participants with an education level of 5 and one participant with an educational level of 4. The control participants were recruited via flyers and experiment databases. They received a monetary reward for their participation of 20 euro in vouchers. This research was approved by the ethical review board of the Radboud University Medical Centre.

There was no significant difference between the mean age of the normal-hearing participants and the mean age of the CI user group (W = 73.5, p = .24). This was tested using a Wilcoxon rank-sum test.

Materials

Two conditions, an inattentive and an attentive condition, were designed to elicit the MMN component and the P300 component separately. In both conditions, the auditory oddball paradigm was used. Two stimuli types were used: a frequency contrast (a 500 vs. 1000 Hz tone) and a consonant contrast (syllables /ba/ vs. /da/).

This resulted in an experiment with four separate conditions: attentive-frequency, attentive-consonant, inattentive-frequency and inattentive-consonant. The order of the four conditions was randomized

Table 1.
Demographic and clinical information of the 8 CI users.
Note. F: female, M: male, Educ. Level: education level, DD: duration of deafness before implantation, Bi: bilateral, Uni: unilateral, CI: cochlear implant, MOC: mode of communication.

ID	Sex	Age (yrs)	Educ. Level	Etiology	Age at implantation (yrs)	DD (yrs)	Bi/ Uni	CI use per day (hrs)	Main MOC
1	M	24	5	Meningitis	3	2.08	Uni	14	Speech
2	M	25	6	Meningitis	3.6	2.17	Uni	16	Speech
3	M	23	5	Congenital	2.7	2.7	Uni	14	Speech
4	M	18	4	Meningitis	1.6 and 1.6	0.08	Bi	15	Speech
5	M	16	4	Prematurity	2.1 and 5	2.08	Bi	12	Speech
6	F	16	4	Unknown	3	3	Uni	16	Half/ half
7	F	18	6	Meningitis	1	0.5	Uni	16	Speech
8	M	19	6	Congenital (LADD syndrome)	5.8 and 19	5.8	Bi	14	Speech

with breaks in between. Half of the participants started with the frequency contrast (first: inattentive-frequency, second: attentive-frequency), and did the consonant contrast after that (third: inattentive-consonant, fourth: attentive-consonant). The other half of the participants did this the other way around (first: inattentive-consonant, second: attentive-consonant, third: inattentive-frequency, fourth: attentive-frequency).

Auditory stimuli.

For the frequency contrast, a 500 Hz pure tone burst and a 1000 Hz pure tone burst of 120 ms each were generated with Praat (Boersma & Weenink, 2018) (settings: stereo channels, 20 ms linear rise and fall time, 80 ms plateau time, sampling frequency of 44100 Hz and an amplitude of 0.2). The 500 Hz tone was used as the standard stimulus, the 1000 Hz tone was used as the deviant stimulus. For the consonant contrast we used the syllable /ba/ as the standard stimulus and /da/ as the deviant stimulus. The duration of the stimuli was 170 ms. These synthesized stimuli were the same stimuli that Beynon et al., (2005) used for their ERP experiment. They, in turn, adapted these stimuli from the ones used in Groenen et al., (2001). For a detailed description of these stimuli please consult the articles mentioned above.

Behavioural assessment.

Before each set of ERP conditions, two short reaction-time tasks were performed by the participants. There was a consonant contrast version and a frequency contrast version. This was to see whether the participants could subjectively distinguish between the contrasts. The same stimuli as the stimuli used in the ERP conditions were randomly presented 20 times (50% standard, 50% deviant). The participants were asked to press the left button when they heard the standard stimulus and the right button when they heard the deviant stimulus. When a button was pressed, the next stimulus was automatically presented. If no button was pressed within 1500 ms, the next stimulus appeared. For both versions there was a familiarization phase of five trials in which the stimuli were presented and it was shown on the screen which button to press for which stimulus. The reaction times of the participants were analysed.

When the first reaction time task was done, all participants were asked to judge the loudness of the sound on a 5-point scale, with 1 = too soft, 2 = a bit soft, 3 = good, 4 = a bit loud, 5 = too loud. Subsequently, the CI participants were given the opportunity to adjust their speech processor if they wished, to avoid any discomfort while listening to the stimuli. From a total of twenty-two participants (8 CI, 14 NH), fifteen participants rated

the sound as 'good' (7 CI, 8 NH), four participants rated the sound as 'a bit soft' (1 CI, 1 NH), and three participants rated the sound as 'a bit loud' (0 CI, 3 NH). None of the CI users felt the need to adjust their processors.

Procedure

The ERP measurements were performed in a sound-proof EEG-lab. Subjects were seated in a comfortable chair. Sound was presented via speakers that were approximately 2.5m away from the participant. The sound presentation at ear-level was kept at 65 dB at all times, as measured by a measuring amplifier (Bruel & Kjaer Type 2610) and a microphone (Bruel & Kjaer Type 4192).

Stimuli were presented with Presentation® software (Version 18.0, Neurobehavioral Systems, Inc., Berkeley, CA, www.neurobs.com). For both the inattentive and attentive measurements, the standard stimuli occurred at a probability rate of 85% and the deviant stimuli occurred at a probability rate of 15%. In each of the four conditions, there were two blocks of 220 stimuli, resulting in a total of 440 stimuli per condition. In each block, first 20 standard stimuli were presented, followed by 30 deviant stimuli that were randomly embedded in 170 standard stimuli. It was made sure that between two deviant stimuli at least three standard stimuli were presented. We controlled for a list-specific effect by generating multiple stimulus lists and assigning these at random to the participants. For each of the four conditions three random stimulus lists of 440 stimuli were generated, resulting in 12 lists in total. In the control group, one-third of the participants got four A-lists (n = 5), one-third got four B-lists (n = 5) and one-third got four C-lists (n = 4). Because the CI user group turned out to be smaller than expected, five participants in the CI user group got four A-lists, and three participants in the CI user group got four B-lists. The lists were made with the program Mix (van Casteren & Davis, 2006), and adjusted by hand to remove any presentation patterns that arose even after randomization.

For the inattentive measurements, the stimuli were presented with an inter-stimulus interval (ISI) of 1000 ms with 10% jitter. In both contrast conditions (frequency and consonant), the participants were asked to watch two different 10-minute silent snippets of a film. The snippets were selected for having emotionally neutral contents. They were instructed thereafter not to pay attention to the sounds that would be presented during the movie. The two inattentive measurements (frequency and consonant) lasted eight minutes each. The video-snippet was automatically quit when the inattentive measurement had fully run. There was no break between the two blocks of stimuli (note: there were breaks between all four experiments, but not within the inattentive experiment(s), so not between blocks).

For the attentive measurements, the stimuli were presented with an ISI of 1500 ms with 10% jitter. This differed from the inattentive measurements because the P300 is a component that spreads out over a longer time-window. We did not want to risk any overlap in neurophysiological responses to the stimuli. In both contrast conditions (frequency and consonant), the participants were instructed to, for each block of 220 stimuli, count in their heads the number of deviant stimuli that occurred. At the end of each block they were asked to type in how many deviant stimuli they had heard (30 deviants in each block). Between the two blocks there was a break to enhance the participants' attention and dismiss fatigue. The participants were told that they could close their eyes during the measurement if they had difficulty not-blinking, but that they had to be careful not to fall asleep during the measurement. As a result of this, some people closed their eyes during the measurements, but most kept their eyes open. Each of the two measurements (frequency and consonant) lasted 11 minutes.

EEG data acquisition.

The EEG was continuously recorded from 24 active electrodes embedded in a 10-20 international system electrode cap (Acticap 32Ch standard-2). The reference electrode was placed at Cz for online referencing and the EEG signal was re-referenced offline using the common average method. The ground electrode was placed at AFz. Due to too much space between the EEG-cap and the scalp, where the electrodes around the processor were supposed to be placed, and due to possible CI artefacts, we did not fill electrode places around the cochlear implant(s) and its contralateral side (CP6; T8; P8; TP10; TP9; T7; CP5; P7). This configuration was kept for both control participants and CI users to enhance consistency. Electrooculography (EOG) was recorded from two horizontal electrodes, placed at the left and right temples, and two vertical electrodes, placed above and below the left eye. Electrode impedance was kept below 20 KOhm. The EEG signal was online filtered with the low cutoff of .016 Hz and the high cutoff of 125 Hz. The EEG was recorded at a sampling rate of 500 Hz.

Analysis

We analysed the EEG-data using the MATLAB-toolbox Fieldtrip (Oostenveld, Fries, Maris & Schoffelen, 2011). Data of the inattentive-frequency condition of one control participant (pp9) were missing due to an experimenter error. For the inattentive conditions, the data were cut into segments with a time-frame of -0.3 to 0.7 seconds before and after onset of the stimulus, respectively. For the attentive conditions this time-frame was -0.3 to 1 seconds. Vertical and horizontal EOG were re-referenced following a bipolar montage. The data were de-trended.

Data cleaning and CI artefact removal. Data were filtered with a low-pass filter of 80 Hz. For both the removal of eye-artefacts and the removal of CI-artefacts, we did an independent component analysis (ICA; Jung et al., 2000). We performed the ICA over all four conditions together. We visually inspected the component topographies, the component time-courses and the corresponding EEG segments. Eye-blink components were rejected.

We developed a procedure for the removal of possible cochlear implant EEG artefacts. The implant artefact is independent of brain processes or task design. It is a reaction of the implant electrode array to the presentation of a sound, that is detected by the EEG. The artefacts are described in the literature as a systematically occurring increased or decreased amplitude peak. (Gilley et al., 2006; Turgeon et al., 2014; Viola et al., 2012). The artefacts do not occur in each CI user, but only in some of them. To attenuate these CI artefacts, we performed a time-locked analysis over the ICA components. For any process to be time-locked to a stimulus, this means that in each trial during the EEG (regardless of paradigm, input stimulus or standard/deviant classification) there should be a deflection in amplitude as a consequence of a cognitive, lower-level, or external process. This deflection should occur at the exact same time in all 1600 trials. If an ICA analysis is performed and the mean of these trials per component is plotted, it shows which components from the ICA are time-locked to the stimulus. These components can then be removed. Removing activity related to the biological processes of the P300 or MMN is unlikely. These neural processes are known to occur later than at stimulus presentation. It is also unlikely that what comes out of the time-locked analysis is non-CI artefact noise. Non-artefact noise, such as eyeblinks, will cancel out because they occur at different time-points throughout the trials. Auditory presentation is the only factor that occurs roughly around the same time in all conditions. It is possible that biological processes related to this auditory presentation (such as the N1 or P2) are filtered out by the CI-artefact analysis. However, the input stimuli that were used were different. The biological processes that occur in reaction to the syllables may be different in time than the biological processes that occur in reaction to the tones. Even if it should be the case that we eliminate activity occurring from these processes, they are not the processes we focus on in this article. We did a time-locked analysis over ICA components for all participants with a CI. If, per participant, there were components that had time-locked amplitude deflections occurring all at the exact same time after stimulus presentation, they were removed from the data. Furthermore, we checked whether the spatial morphologies of these identified components matched morphologies from earlier papers (Gilley et al. 2006; Viola et al. 2012). The components could not occur later than 150 ms after stimulus presentation, otherwise they were not removed. In the end, we deleted artefact component(s) for CI users 1, 5, 6 and 7. Although this procedure was semi-automated, it still remains a subjective task to determine which components should be eliminated.

After the ICA, the data (per participant) were split into the four individual conditions. For each condition we used a semi-automatic artefact rejection approach (ft_rejectvisual in Fieldtrip) to identify and throw out any trials that were outliers. This approach shows the preprocessed data in all channels or trials and allows the user to select noisy data (trial and/or channel) and delete it. Furthermore, it is possible to compute the variance in each channel and/or trial and delete outliers based on this. For each participant on average 17 (for the CI users) or 18 (for the controls) out of 440 trials were deleted per condition. Channels that were noisy were noted down (not deleted) for each experiment. Later on, this information was taken in consideration when selecting the channels to perform statistics on.

ERP calculations and statistics

Group-level analysis.

The artefact-free data were used to compute group ERPs. The ERPs were computed by averaging waveforms across trials per stimulus condition (standard vs. deviant), per group (normal-hearing and CI user), per task by contrast

condition (attentive-frequency, attentive-consonant, inattentive-frequency, inattentive-consonant). The data were filtered with a low-pass filter of 50 Hz and down-sampled to 512 Hz. We used cluster-based permutation tests (Maris & Oostenveld, 2007) to statistically evaluate the presence of the ERPs in all four conditions per group. This was done using a within-subjects design in which the grand average response to the standard trials was compared to the grand average response to the deviant trials. Statistics were performed as follows: first a dependent samples t-test was calculated for every electrode/time-point. The comparison was based on all time-points from 150 ms to 800 ms post-stimulus onset for the attentive task-condition and 50 ms to 350 ms for the inattentive task condition. Statistical tests were based on channels 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task-conditions and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task conditions. Decisions for these time-points and channels were based on previous literature describing the location of effects (Johnson, 2009) and on the exclusion of channels that were deemed excessively noisy during data acquisition. The electrodes/time points were clustered based on spatial and temporal adjacency at an alpha level of 0.05. Channels had on average 3.3 neighbours. Cluster-level statistics were calculated by taking the sum of the t-values within every cluster. The largest cluster-level statistic was taken for evaluation under a permutation distribution. This distribution under the null hypothesis of exchangeability between trial conditions was constructed by randomly re-assigning the standard trial and the deviant trial labels to the original individual ERP waveforms, followed by the construction of spatiotemporal clusters, in the same way as for the observed data. 1000 permutations were used to make the permutation distribution. The p-value was determined as the proportion of random permutations that yielded a more extreme cluster statistic than the cluster in the original data. The alpha-level was set to 0.05 (two-sided test). If the p-value was smaller than alpha, the difference between the standard and the deviant trials was deemed significant.

Individual-level analysis.

We also performed individual ERP analyses per stimulus-condition (standard vs. deviant) per task by contrast condition (attentive-frequency, attentive-consonant, inattentive-frequency and inattentive-consonant). To test for ERP presence, we used the same cluster-based permutation procedure as described for the group analysis. That is, we tested the difference between the standard and the deviant waveforms per individual. However, a between-trials design was used. This means that an independent samples t-test was performed.

Amplitude and latency analyses.

The attentive task-condition yielded robust results for all participants for both contrast conditions, while the inattentive task condition did not (see Results below). Therefore, we performed the amplitude and latency analyses only on the attentive task-condition data. To avoid the different sample sizes of the two groups, we took a sub-sample of normal-hearing participants (n = 8) to match the sample of CI user participants (n = 8). This sub-sample firstly was matched to the CI user group on the order in which the contrast conditions appeared during data collection. Then, each CI user was matched to a control on at least one of the following three criteria: education level, age, or sex. This was done because it was not possible to match the two groups on one criterion only and because all criteria were deemed equally important to match on. This semi-objective way of matching shielded the matching from a selection bias. In the end, the sub-sample consisted of normal-hearing controls 1, 3, 4, 5, 6, 8, 11 and 12. Their significant clusters and time-windows from the individual ERP statistics are reported in Appendix A. We performed the amplitude and latency analysis over one electrode: 'Pz'. This decision was based on earlier studies that also performed their analyses over one or two electrodes. (Beynon et al. 2002, 2005; Groenen et al., 2001; Kileny, 1991; Micco et al., 1995; Obuchi et al., 2012). This way, we would be able to more accurately relate our findings back to earlier ones. We did not pursue this decision for the ERP presence analysis (see above), because the cluster-based permutation approach is more conservative when using more channels.

Amplitude analysis.

We assessed amplitude differences in the attentive task condition difference waveforms between groups, per contrast condition. Because analysing amplitude differences in ERP designs has been shown to be prone to bias and may lead to incorrect conclusions (Luck, 2012; Woodman, 2010), we explored the outcomes of two different analyses. We calculated the mean amplitude (MA) and the peak amplitude (PA). As the MA, we calculated the mean voltage over a pre-specified time-window of the difference

wave (standard deviant condition). We chose 220 ms to 705 ms for the frequency contrast condition and 315 ms to 760 ms for the consonant contrast condition as time-windows. These time-points were the minimum and maximum time-points of the time-windows in which the individual ERPs were significant, as present in the cluster-based statistics. As the PA, we took the peak amplitude of the difference waveform (standard-deviant condition). For the consonant contrast condition, we chose the same time-windows as in the MA approach. For the frequency contrast condition, however, we chose a time-window of 300 ms to 705 ms. This was done because CI user 3 showed two peaks: one around 250 ms and one around 400 ms. As the first peak was very early in comparison to all other CI users, it was suspected to be a CI artefact. We therefore wanted to make sure we got the PA of the second peak. The peaks of all other participants did not start until 300 ms so there was no danger of missing other peaks.

We performed statistics over the means and standard deviations of both the MA and PA outcome values. Because of a relatively small *n*, we used a non-parametric Wilcoxon rank-sum test to test for significant mean differences between groups and a non-parametric Fligner-Killeen test to test for homogeneity of variances between-groups. more, we used the Wilcoxon signed-rank test to test for significant mean differences between contrast conditions. Lastly, we used the Wilcoxon rank-sum test to explore an interaction effect between contrast condition and group.

By comparing the results of these two analyses, this paper may contribute to the discussion on the reliability of calculating the peak amplitude versus the mean amplitude of the difference wave. On the one hand, it is debated whether the peak of an ERP component has any meaningful value in itself. Also, the peak of an ERP is very sensitive to high-frequency noise (Luck, 2012; Woodman, 2010). On the other hand, the mean amplitude is not free of disadvantage either. A lower-to-noise level may cause the ERP waveform of an individual to be fluctuating at several places in the waveform (see for example the alpha noise in Figure 5, CI user 2), causing the mean amplitude (calculated over the entire window) to decrease.

Latency analysis.

We also assessed latency differences between groups and contrast conditions in the attentive task condition. Measuring ERP latency differences on a single-subject level is deemed a cautious undertaking. Firstly, the relationship between the underlying component and the local shape of a component is not obvious (Luck, 2005). Secondly, the signal-to-noise level is low due to averaging over a small amount of trials. Therefore, we measured latency differences with the jackknife-based approach (Kiesel et al., 2008). In this approach, latencies are scored for each of *n* grand average waveforms in a group, with each grand average waveform computed from a subsample of *n-1* individual waveforms. Each participant in a group is omitted from the analysis once, and each latency score is calculated not from a single-subject waveform, but from a grand average.

Using the peak latency of the component as a scoring method has been deemed misleading and arbitrary in the ERP literature (Luck, 2005; Woodman, 2010). Therefore, we again relied on Kiesel et al. (2008) for our scoring method. The scoring was done as follows: First, we determined a latency onset criterion of 300 ms to 700 ms, based on our set time-windows for determining the peak amplitude of the difference wave. The P300 ERP was determined as the first positive going peak from the set onset criterion (300 ms in our case). Then, for each subsample, ERP latency was calculated using a relative criterion technique: "the time-point at which the amplitude reaches a constant, pre-specified percentage of the peak value" (Kiesel et al., 2008). We chose to take 50% as the pre-specified value, one of the percentage values that came out to be most reliable to use when measuring latencies of the P300 in an oddball paradigm (Kiesel et al., 2008). We submitted the latency outcome values of this jackknife-based approach to a 2x2 repeated measures analysis of variance (ANOVA; group x condition). The F-value of this ANOVA was adjusted according to the following formula: Ulrich & Miller (2001). It was not possible to assess individual differences using the Jack-knife based approach. We therefore also calculated the latency values per single-subject waveform (with the same scoring method and settings) and used the Fligner-Killeen test to test for homogeneity of variance between-groups. Although we are aware of the pitfalls of the single-subject waveform method, we wanted to be as thorough as possible in exploring differences in variation per group.

Correlation with behavioural speech perception scores and duration of deafness

Lastly, we correlated MA with behavioural

speech perception scores of the CI users as measured in the clinic and with duration of deafness. We used a non-parametric correlation method: the Spearman's rho test. Peak amplitude was not included. This decision was based on the fact that both amplitude measures showed a similar pattern of results and mean amplitude has been argued to be more reliable (Luck, 2012; Woodman, 2010). For consistency, we correlated single-subject waveform ERP latency with behavioural speech perception scores and duration of deafness.

The behavioural speech perception scores were obtained in the clinic using a word intelligibility task (NVA lijsten, Bosman Wouters & Dumman, 1995). In this task, isolated one-syllable words are presented at 70 dB in an audio-booth and the CI user is asked to repeat the word they are presented with. The percentage correctly repeated words is used as an outcome measure. Almost all scores were obtained within the same half year as the EEG was conducted. For CI user 1 and user 5, there were no up-to-date perception scores so their lastly available data were used. For CI user 1 this was obtained five years ago, for CI user 5 this was obtained one year ago. The decision was made to not exclude these values in the analysis because the CI users were all implanted at a young age and their perception scores are assumed to have relatively stabilized over the years. The bilaterally implanted CI users were all tested with only their right CI on, the same CI as they chose to have on during the EEG-measurements.

Duration of deafness was obtained from the demographic information of our participants. It is displayed in years in table 1.

Results

Behavioural assessments

Reaction time experiment.

Results for the behavioural reaction-time task - per group per contrast condition - are displayed in Figure 1. A two-way ANOVA revealed a main effect of group (F (1,40) = 15.56, p < .001). The normal-hearing group pressed significantly faster (M = 457.61, SD = 108.77) than the CI user group (M = 588.35, SD = 162.29) in both contrast conditions. There was also a main effect of contrast condition (F (1,40) = 21.17, p < .001). Both groups pressed significantly faster in the frequency contrast condition (M = 431.80, SD = 124.83) than in the consonant contrast condition (M = 431.81, SD = 124.31). Descriptively, the plot shows that the CI user group tends to be slower than the normal-hearing group in the consonant contrast condition. This interaction effect was, however, not significant (F (1,40) = 2.72, p = .100).

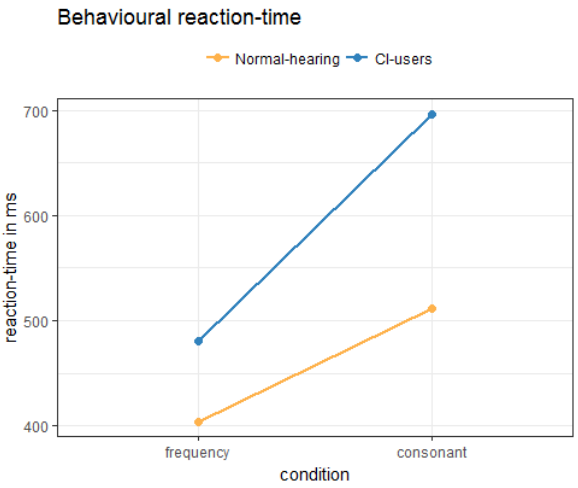

Fig. 1. Behavioural reaction time task results. Mean reaction time and standard deviations in milliseconds as a function of group (Normal-hearing (n = 14), CI users (n = 8)) and contrast condition (frequency, consonant)

Counting deviants during the attentive task-condition

Scatterplots with standard scores (score per participant in both groups minus the normal-hearing group's mean divided by the normal-hearing group's SD) of the amount of deviants counted per block during the attentive task-condition are displayed in Figure 2, for each contrast condition separately. In total, 30 deviants could be counted in each block. The normal-hearing group had a mean of 29.60 (SD = 2.47) over both blocks in the frequency contrast condition, and a mean of 29.82 (SD = 1.70) over both blocks in the consonant contrast condition. The CI user group had a mean of 30.50 (SD = 1.32) over both blocks in the frequency contrast condition and a mean of 29.38 (SD = 2.00) over both blocks in the consonant contrast condition.

Group ERP results

Group averaged ERP results for both groups collapsed over the electrodes 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task-condition

(P300) and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task-condition (MMN) are displayed respectively in Figure 3.1 A and B and 3.2 A and B. In the normal-hearing group, time-locked EEG-activity for the deviant trials was found to be significantly more positive in amplitude than activity for the standard trials in the attentive task condition. This was found for both the frequency (270-690 ms, p = .002) and the consonant contrast (390-690 ms, p = .002). The CI user group yeilded similar results. A significant positive deflection was found for the frequency (270-660 ms, p = .002) and the consonant contrast (350-660 ms, p = .002). The time-windows all correspond roughly to the P300 component as described in the literature (usually present from 350-500 ms). For the inattentive task-condition, time-locked EEG activity for the deviant trials was found to be significantly more negative in amplitude than activity in the standard trials, but only for the frequency contrast. This was found for the normal-hearing group (90-160 ms, p = .020) as well as for the CI user group (120-200 ms, p = .020). The time-windows correspond roughly to the MMN component as described in the literature (usually present from 150-250 ms). For the consonant contrast, significant negative deflections were found for neither the normal-hearing group (p = .090), nor the CI user group (p = 1.000). Scalp topographies of the difference wave (standard deviant) of the group-averaged ERPs are displayed respectively in Figure 3.1 C and D and 3.2 C and D. For the attentive task-conditions the clusters were detected over centro-parietal regions. For the frequency contrast in the inattentive task-condition, the clusters were detected over fronto-central regions.

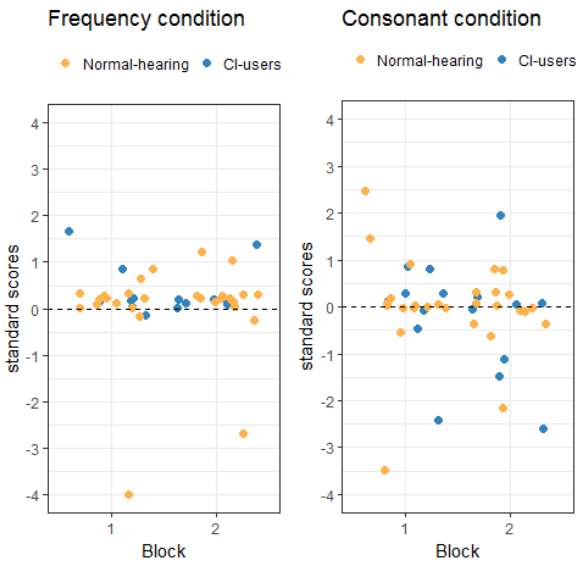

Fig. 2. Results of deviant counting. Scatterplots showing the standard scores in both groups of the amount of deviants counted during the attentive task-condition. On the left the results for the frequency contrast condition are displayed, on the right the results for the consonant contrast condition are displayed

Individual ERP results

Individual ERP results per task condition per contrast condition collapsed over the electrodes 'CP1', 'CP2', 'P3', 'P4', 'Pz', 'C3', 'C4' for the attentive task condition (P300) and 'Fz', FCz', 'F3', 'F4', 'FC1' and 'FC2' for the inattentive task condition (MMN) are displayed respectively in Figures 4 (attentive-frequency), 5 (attentive-consonant), 6 (inattentive-frequency) and 7 (inattentive-tone). In each of these figures the standard trials and the deviant trials are plotted per individual, as well as the time-window(s) in which the difference between these trials was significant (if there were significant differences). This significant time-window is indicated by the dashed lines in the figures. In each of the figures two randomly picked individual ERP plots from the normal-hearing control group are displayed. In the other eight plots of each of the figures, all individual ERPs of the CI users are displayed. Figure 8 shows the number of individuals for whom the ERP effects were statistically present per group and per contrast condition. An ERP waveform was deemed present if there was at least one positive (only for the attentive task condition) or one negative (only for the inattentive task condition) significant cluster in the pre-specified time-windows. A table with all p-values of the difference between standard and deviant trials and corresponding time-windows per participant (n = 22) is displayed in Appendix A.

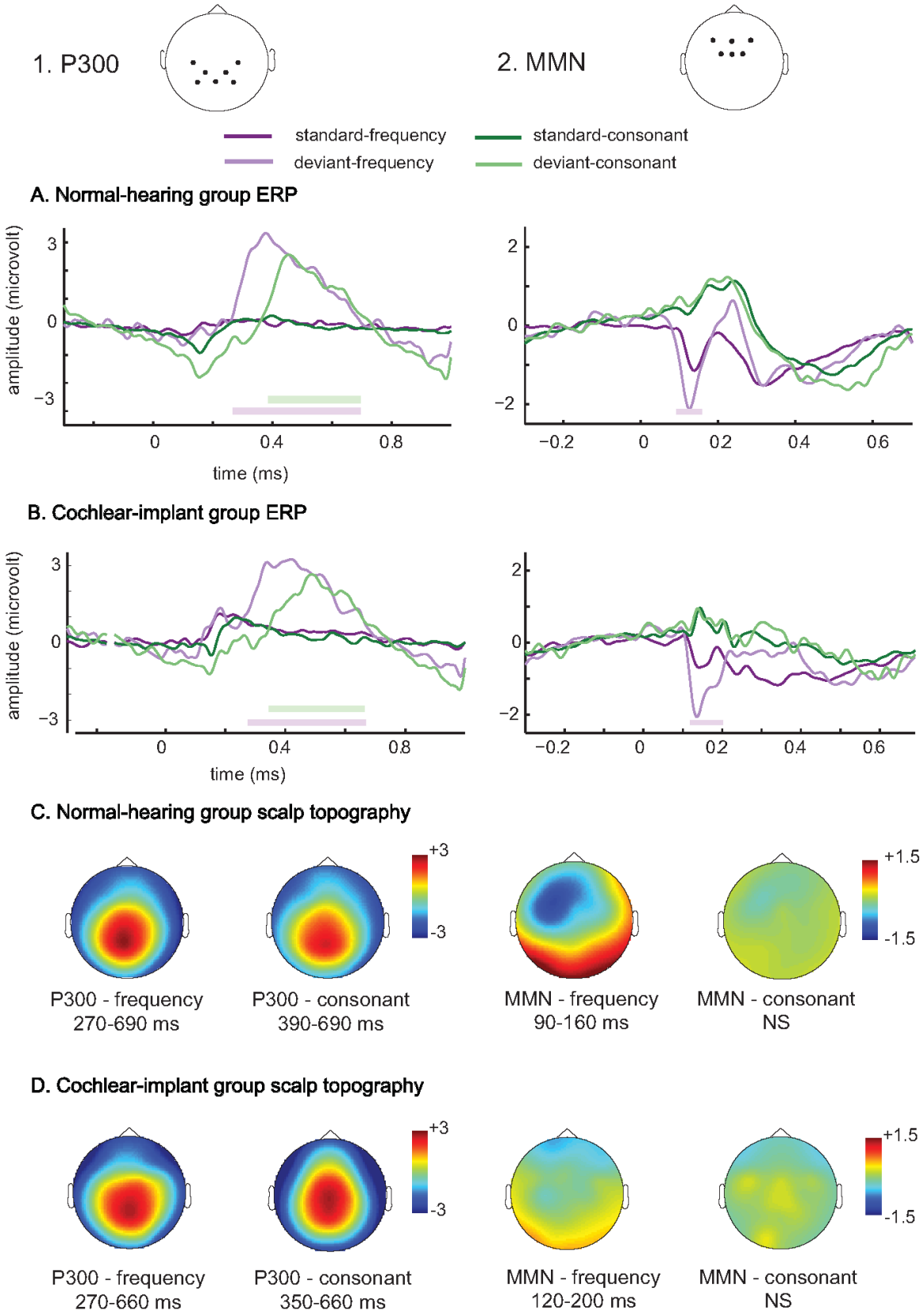

Fig. 3. Group ERP results. Group averaged ERP waveforms (**A, B**) for the standard-frequency, deviant frequency, standard-consonant and deviant-consonant contrast trials and corresponding difference-wave (standard-deviant) scalp topographies (**μV**) (**C, D**) per group (normal-hearing and CI user group) per task-condition (1. P300 [attentive], 2. MMN [inattentive]). EEG-cap configurations are also shown per task-condition. The time-windows in which the standard trials were significantly different from the deviant trials are indicated by the light purple and light green bars in the ERP plots

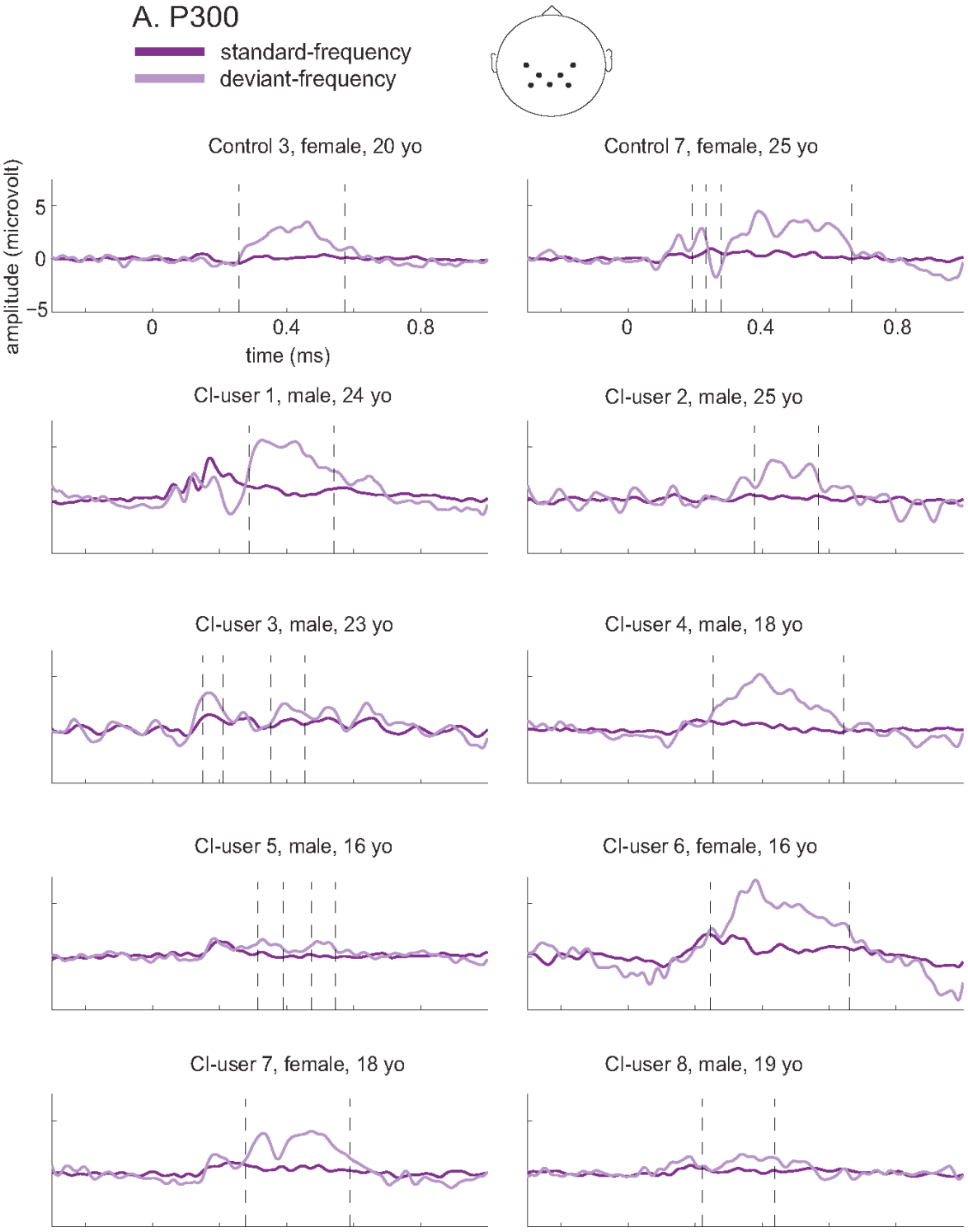

Fig. 4. Individual ERP results attentive-frequency. Individual ERP waveforms for the standard-frequency and deviant frequency trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

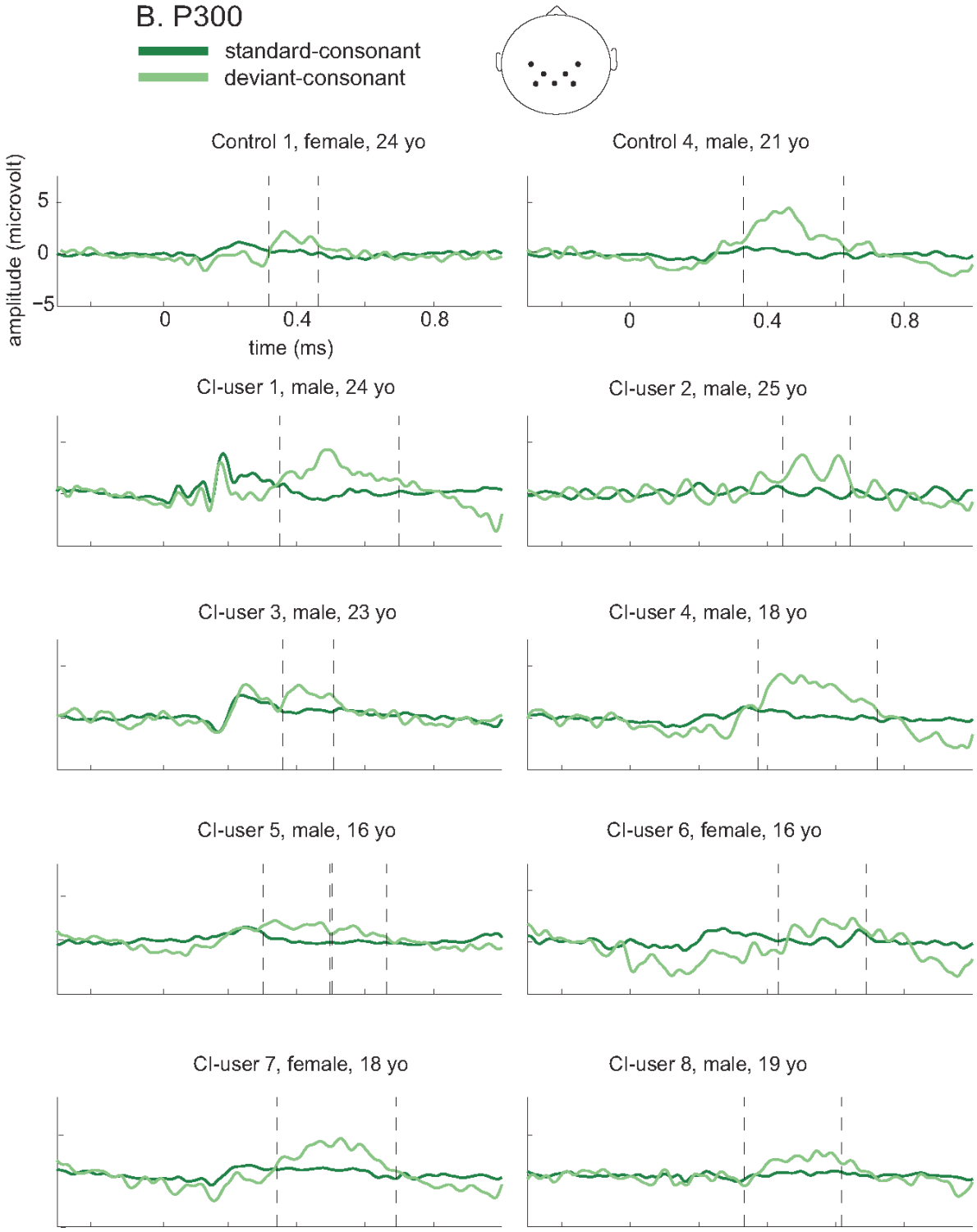

Fig. 5. Individual ERP results attentive-consonant. Individual ERP waveforms for the standard-consonant and deviant consonant trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

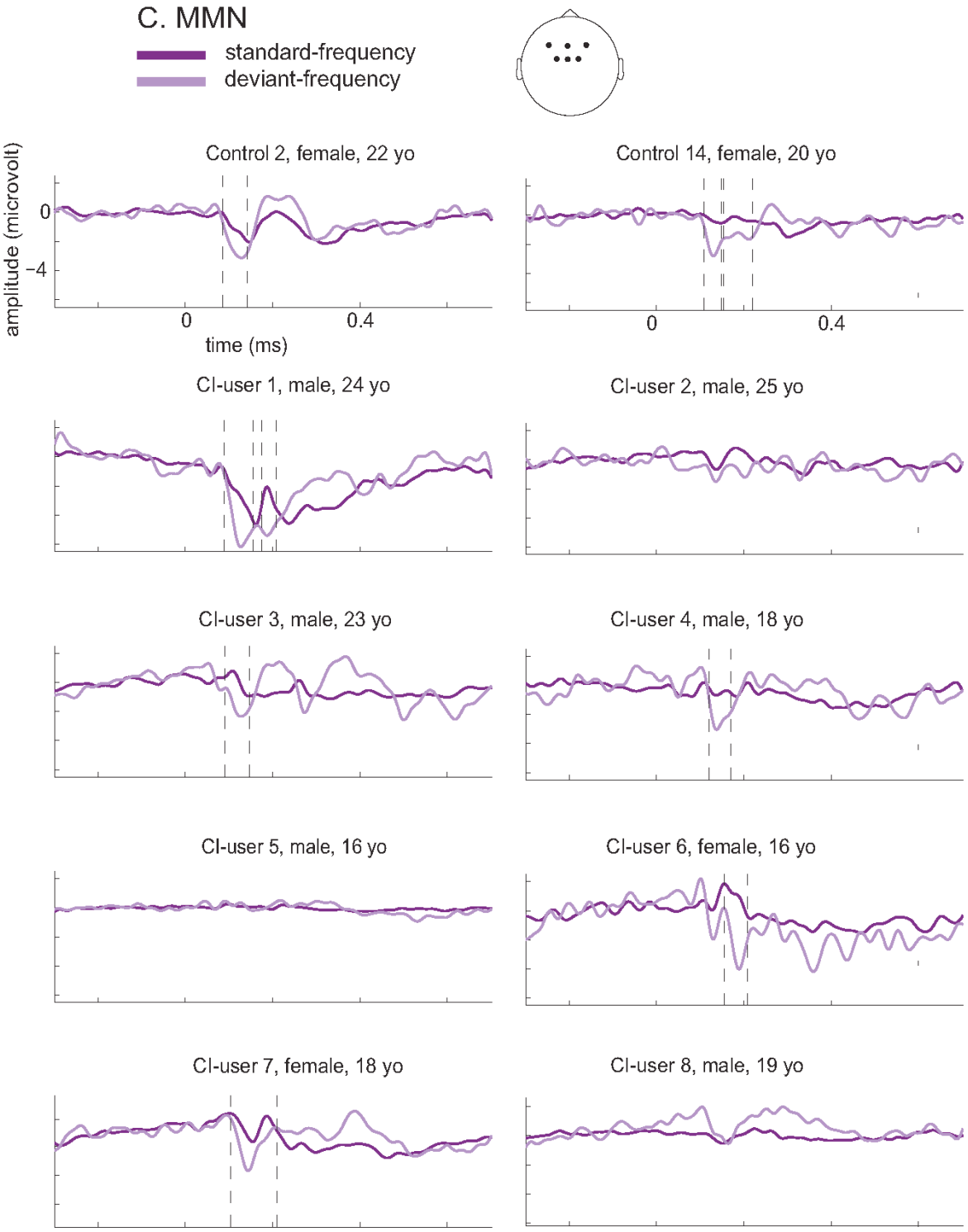

Fig. 6. Individual ERP results inattentive-frequency. Individual ERP waveforms for the standard-frequency and deviant frequency trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

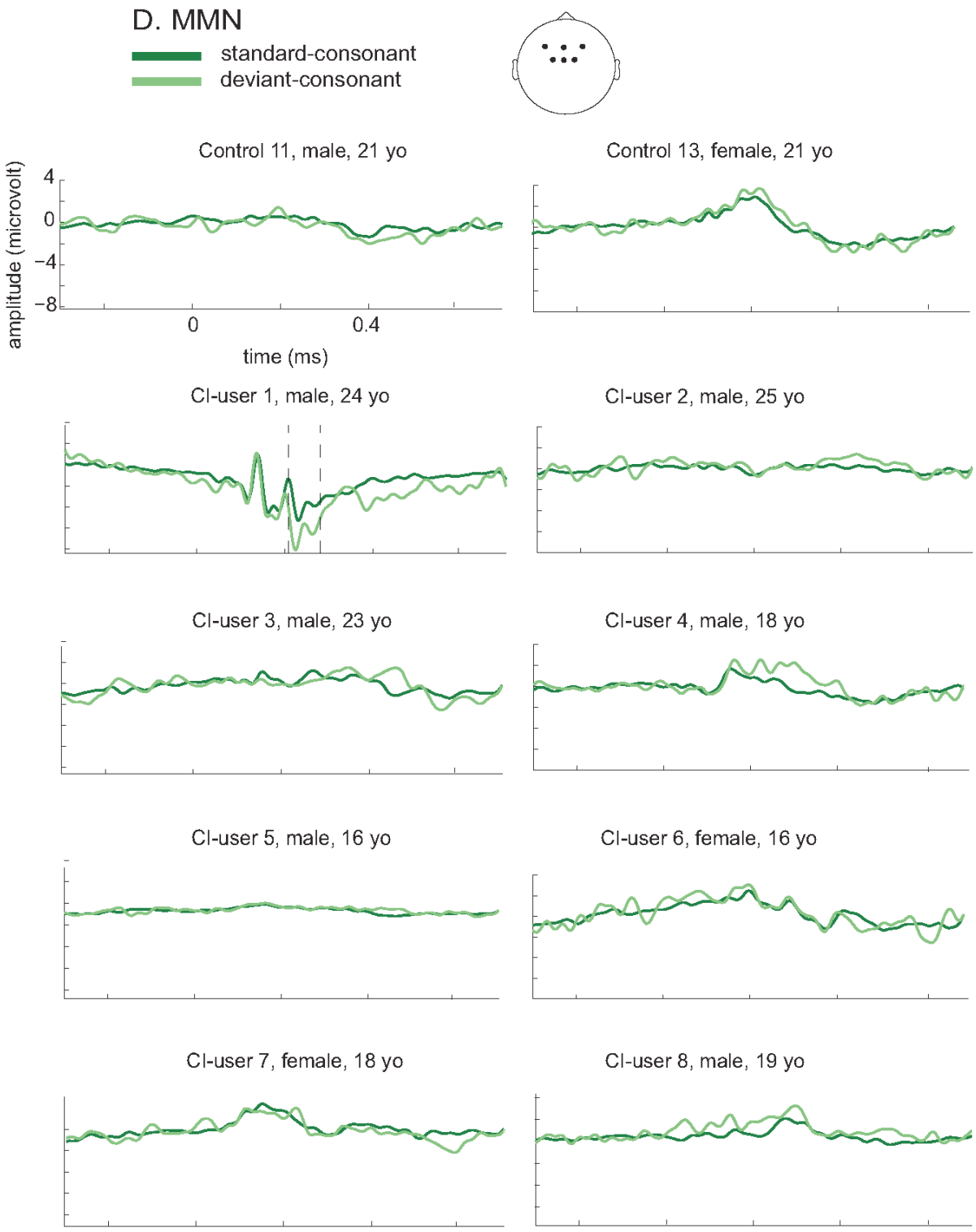

Fig. 7. Individual ERP results inattentive-consonant. Individual ERP waveforms for the standard consonant and deviant consonant trials. At the top of the graph, two randomly picked ERPs of normal-hearing controls are displayed as reference. The next eight ERPs correspond to each individual CI user (n = 8). The EEG-cap configuration is shown. Dashed lines indicate significance

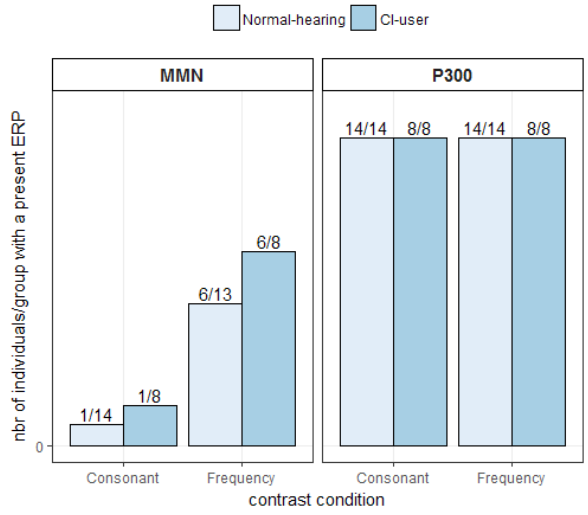

Fig. 8. Number of individuals with a significant individual ERP waveform. The number of individuals per group that had a significant ERP waveform, split by contrast condition

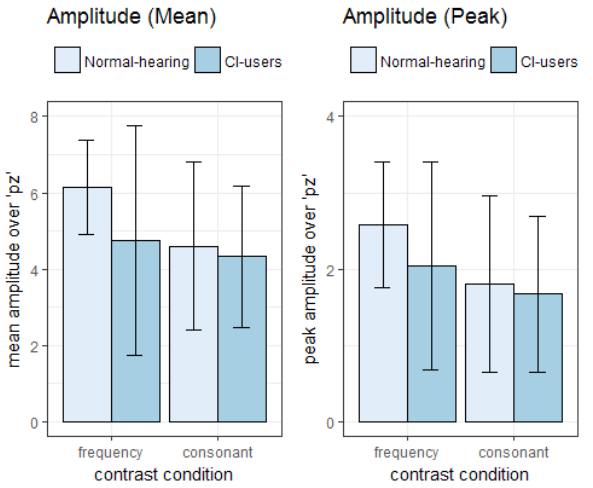

Fig. 9. Amplitude results. Plots showing the mean and standard deviations of the mean amplitude (left) and peak amplitude (right) in microvolt of the difference waveforms per group per contrast condition, as measured in the attentive condition. Amplitude was calculated over electrode 'Pz'

Amplitude results

Means and standard deviations of our two amplitude measures, the mean amplitude and the peak amplitude - per group per contrast condition - are displayed in Figure 9. The Wilcoxon rank-sum test (for group) and signed-rank test (for condition) did not show significant differences between groups and contrast conditions, neither for the mean amplitude ($W = 107$, $p = .45$ for the group comparison, $V = 38$, $p = .12$ for the contrast condition comparison), nor the peak amplitude ($W = 101$, $p = .32$ for the group comparison, $V = 40$, $p = .16$ for the contrast condition comparison). To test for the interaction effect of group x condition we calculated the difference of the frequency minus the contrast condition for each individual. Consequently, we tested the difference as a function of group, again using the Wilcoxon rank-sum test. We did not find a significant interaction effect for either measure (Mean amplitude: $W = 20$, $p = .23$, Peak amplitude: $W = 20$, $p = .23$). Furthermore, the Fligner-Killeen test showed no main effect of group or condition on the variance within groups on either the mean amplitude measure ($\chi^2(1) = 0.42$, $p = .51$ for group, $\chi^2(1) = 0.02$, $p = .90$ for condition), or the peak amplitude measure ($\chi^2(1) = 1.26$, $p = .26$ for group, $\chi^2(1) = 0.57$, $p = .44$ for condition). There was a significant interaction between group and condition in variance for the peak amplitude ($\chi^2(3) = 9.22$, $p = .03$), but not for the mean amplitude measure ($\chi^2(3) = 4.55$, $p = .20$). This means that when using peak amplitude as a measure, the variance within the CI user group was significantly greater than the variance within the normal-hearing group, and that this is true only for the frequency condition, not for the consonant condition.

Latency results

Results for the latency analysis per group per contrast condition are displayed in Figure 10. A two-way ANOVA revealed a main effect of condition ($F(1,28) = 23$, $p < .001$). The latency was significantly later in the consonant ($M = 0.402$, $SD = 0.01$) than in the frequency contrast condition ($M = 0.308$, $SD = 0.01$) for both groups. There was no main effect of group ($F(1,28) = 0.18$, $p = .67$), nor an interaction effect of group x condition ($F(1,28) = 0.004$, $p = .95$). The Fligner-Killeen test we performed on the latencies calculated from the single-subject ERP waveforms was not significant. There was homogeneity of variance between groups ($\chi^2(1) = 0.55$, $p = .46$) and between contrast conditions ($\chi^2(1) = 0.65$, $p = .41$), and there was no significant interaction effect between group x contrast condition ($\chi^2(3) = 2.8$, $p = .41$).

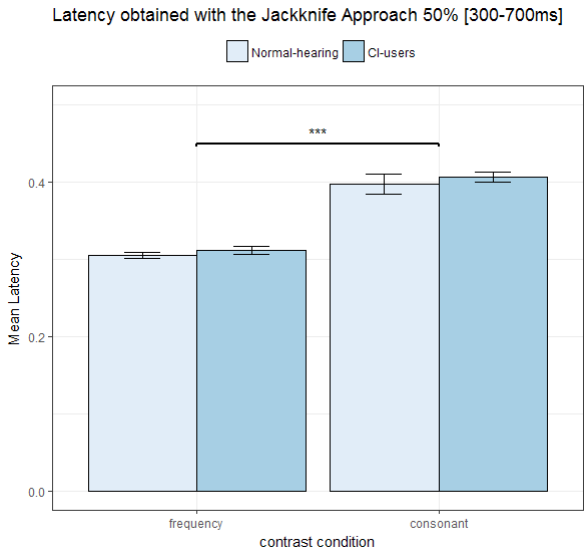

Fig. 10. Latency results. Plots showing the mean and standard deviations of the latency outcome values in seconds, as measured in the attentive task-condition. Results are shown per group per contrast condition. Results were obtained using a Jack-knife based approach and a 50% relative criterion scoring technique with a time-window of 300-700 ms. Latency was calculated over electrode 'Pz'. The asterisks indicate significant differences

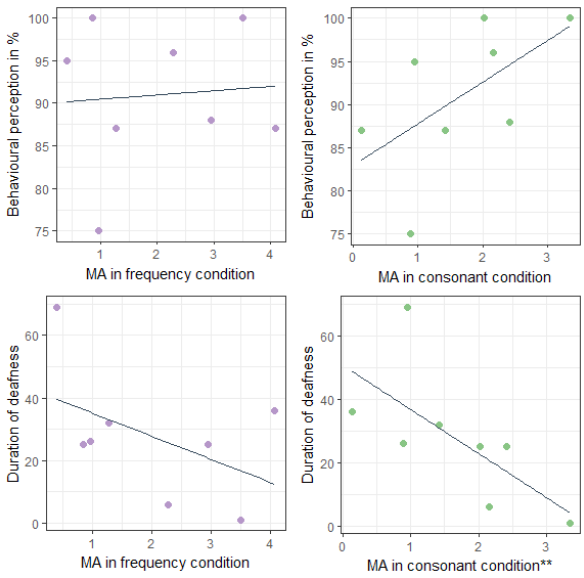

Fig. 11. Correlations between behavioural speech perception and duration of deafness. Scatterplots showing the correlations between the mean amplitude (MA) in microvolt and the behavioural speech perception scores in percentage correct (top) and duration of deafness in months (bottom), for each contrast condition (left: frequency, right: consonant)

Relation between amplitude and latency, behavioural speech perception scores, and duration of deafness

Results of the non-parametric correlation between the behavioural speech perception scores and mean amplitude and duration of deafness and mean amplitude are displayed in Figure 11. As for speech perception, there is no significant relation between mean amplitude of the P300 components and the behavioural speech perception scores in the frequency condition (r_s = -.26, p = .53). For the consonant condition, however, there was a strong correlation between the mean amplitude of the P300 components and the behavioural speech perception scores. This showed a trend towards significance (r_s = .70, p = .05). The assumed relation is positive: the higher the behavioural speech perception score of the individual, the greater the amplitude of the P300 component. As for duration of deafness, there is no significant relation between mean amplitude of the P300 components and duration of deafness in the frequency condition (r_s = -.33, p = .41). There is, however, a significant and strong correlation between the mean amplitude of the P300 components and duration of deafness in the consonant condition (r_s = -.83, p = .009). The relation is negative, which means that the shorter a CI user has been deaf, the greater the amplitude of the P300 component.

Latency as measured using the single-subject waveform approach was not correlated with behavioural speech perception scores or duration of deafness, in either condition (duration of deafness and latency for the frequency contrast: r_s = .19, p = .64 and the consonant contrast: r_s = .06, p = .88; behavioural speech perception & latency for the frequency contrast: r_s = -.15, p = .75 and the consonant contrast: r_s = -.41, p = .30).

Discussion

Robustness of the MMN and the P300 and their suitability for the clinic

The primary aim of this study was to compare two ERP components, the P300 component and the MMN component, in terms of their ability to robustly measure auditory discrimination in normal-hearing adolescents and adolescents with a CI on an individual level. The group-averaged ERP results show that, despite a difference in sample size, the CI user group performs similar to

the normal-hearing control group. The similarity of results between the two groups is in line with earlier research (MMN: Kraus et al., 1993; Ponton et al., 2000; P300: Groenen et al., 2001; Micco et al., 1995 [although not similar for the consonant contrast]), with the exception of poor-performing CI users (Beynon et al., 2002; Turgeon et al., 2014).

Furthermore, the individual results are clear-cut in indicating which ERP paradigm is most robust in measuring auditory discrimination under the set circumstances. For the attentive task condition, eliciting the P300 response, all participants showed a significant difference between the standard and the deviant waveforms, regardless of group or contrast condition. For the inattentive task condition, eliciting the MMN response, results were less robust. While in the frequency contrast condition half of the CI users and normal-hearing participants showed ERP presence, in the consonant contrast only one participant in each group showed ERP presence.

Although all individual waveforms showed significance in the attentive task-condition, the components seem to vary considerably in robustness. Although this variation is evidently present in CI users, it is also present in some controls (e.g., see control 1 in Figure 5). Therefore, on the basis of the figures, we should be careful in interpreting differences in the robustness of the waveforms as non-normal differences in auditory discrimination abilities. They may be differences that also appear in the normal-hearing population. Furthermore, it is equally likely that differences in signal-to-noise ratio may be underlying the variation in robustness.

The absence of robustness in the inattentive task condition corroborates the hypothesis that the set duration of the experiment is too short to robustly elicit the MMN. The difference in robustness between the frequency and the consonant contrast condition may be explained by a combination of a lack of power due to duration and a more complex contrast condition. This is emphasized by the findings from two earlier studies. These used almost four times as long EEG-recordings as were used in the current study to elicit the MMN with a consonant contrast (Singh et al., 2004; Turgeon et al., 2014). This finding has important implications for setting up ERP experiments in the future. When the aim is to elicit a robust MMN response on an individual level, a longer measurement than 10 minutes is needed. Interestingly, this finding contradicts the finding of an earlier study. This study found a significant MMN and P300 response in three CI users and three normal-hearing controls with only three-minute EEG-recordings, elicited using frequency contrasts (Obuchi et al., 2012). There is no straightforward explanation for this discrepancy. It may, however, be due to differences in the amount of electrodes that were used to perform the analysis over.

P300 only: differentiating between groups, within groups, and between contrasts

Amplitude and latency between- and within groups.

A second aim of this study was to evaluate whether the amplitude and latency of the individual waveforms in the attentive task condition could distinguish the CI users as a group from the normal-hearing participants as a group. Also, it was assessed whether variance on these measures was greater in the CI user group than in the normal-hearing group. This would indicate that differences in the P300 response of CI users cannot be ascribed to regular P300 variation as present in the normal-hearing population. Greater variation in the P300 response of CI users was expected based on studies that found individual differences in the behavioural speech perception abilities of prelingually deaf CI users (Pisoni et al., 2000; ASHA, 2004), and in the P300 results of prelingually deaf CI users (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991).

Results of both mean amplitude and peak amplitude showed that it was not possible to distinguish the CI user group from the normal-hearing group in either contrast condition. Latency results showed a similar pattern. This finding was not unexpected based on the results of Beynon et al. (2002), although in Beynon et al. (2005), postlinguals could be distinguished from normal-hearing controls.

A first explanation for this null result may be that differences in auditory discrimination abilities are present, but that our measurements are not able to reflect these. Firstly, because we used individual waveforms for our analysis, the signal-to-noise ratio may have been low. For the latency analysis however, this is less of an issue, because we used the Jackknife approach. Secondly, even if not only noise was measured, conclusions on the nature of an effect should be drawn with caution. It may well be that other (task-independent) cognitive or biological mechanisms such as general intelligence, attention, or the arousal state of subjects were underlying P300 components (Polich & Kok, 1994). These processes are not expected to differ between groups, which might explain why no differences were found.

A second explanation may be that the differences between the auditory discrimination abilities of CI users and normal-hearing participants are too subtle to be elicited by the chosen measurements and design in this study. Our contrast conditions are fairly simple compared to the level of difficulty in auditory perception CI users encounter on a daily basis. This assumption would contradict the findings of older studies that this contrast does not elicit a P300 for some CI users. However, in older studies, more poor-performers were included in the analysis (Beynon et al., 2002; Jordan et al., 1997; Kileny, 1991). Poor performance in those studies was defined as no behavioural discrimination ability (Kileny, 1991; Jordan et al., 1997) or a low behavioural speech perception score (e.g. <65% on an open-set speech recognition task in Beynon et al., 2002). All CI users in the current study could discriminate between the stimuli (although as a group they took somewhat longer). Their behavioural speech perception scores were high (mean 90%, ranging from 75-100%).

Another explanation for such a ceiling performance may be that differences in perception abilities (at least for speech in isolation) between (prelingually deaf) CI users and normal-hearing participants have diminished over the years. It is possible nowadays to be implanted from a very early age. Compare for example the age of implantation in older studies on prelingually deaf CI users (range 5-33 yo in Beynon et al., 2002 and Jordan et al., 1997 combined), to the age of implantation in our study (range 1-5 yo). For prelingually deaf users in general, this early implantation means that their period of auditory deprivation (duration of deafness) diminishes considerably and, furthermore, that their auditory cortex can start developing while it is still flexible.

Considering the variance within groups, greater variance in peak amplitude was found within the CI user group as opposed to the normal-hearing group. This was found only for the frequency contrast. While this implies to corroborate the expectation of non-normal individual differences in the auditory perception abilities of CI users, it does not rhyme with our behavioural results. For the deviant counting, the spread in the consonant condition was greater for both groups. For the reaction time task, although only descriptively, the normal-hearing group showed greater variance in the frequency condition, while the CI users showed greater variance in the consonant contrast condition. Moreover, because this result was not found for the mean amplitude measure, even though this measure has been argued to be more reliable (Luck, 2012; Woodman, 2010), this result should be interpreted with caution.

Latency variance as obtained with the Jackknife approach was not used in our analysis of the variance within groups. It was not possible to correct for the reduced variance as a result of this approach (Miller & Ullrich, 2001). When using the latency of the single-subject ERP waveforms, there were no differences in latency variation between the two groups. The absence of greater variance in auditory processing in the CI user group as opposed to the normal-hearing group again confirms the hypothesized ceiling effect for our CI users.

Amplitude and latency between contrast conditions.

We replicate earlier studies in finding that, using the P300 as a measure for auditory discrimination, it is likely that the discrimination of stimuli by the brain lies on a continuum of complexity, with more complex stimuli being more difficult to discriminate (see Polich, 2004 for a review on healthy subjects).

Although we found a longer latency for the consonant contrast as opposed to the frequency contrast, we did not find the same result for the amplitude. Descriptively, however, there was a trend towards the more complex condition yielding a lower amplitude. The greater variance in the frequency condition for the CI users may explain why the difference is not significant. Furthermore, the difference in robustness of the effects for amplitude as opposed to latency may be again due to the fact that we obtained the latency from eight times an n-1 group sample, while we used the individual waveforms to obtain the amplitude.

These results are fruitful for the long-term goal of this study to develop a neurophysiological predictor for speech perception abilities. If input conditions are made more complex in the future, latency and amplitude differences may increase. That way, it may be possible to highlight the more subtle differences in speech perception that were not picked up by the current design.

P300 only: relation between neurophysiological results, behavioural results, and duration of deafness

In the light of our long-term aim to develop a marker for speech perception abilities it is important to link neurophysiological results to behavioural results. Despite the small sample size of this study, it was found that lower behavioural speech perception resulted in a lower amplitude of the P300 response.

This was found only for the consonant contrast. This result implies that the lower the amplitude of the P300 component, the harder it is to perceptually discriminate between phonemes. Phoneme discrimination is a very important aspect of the speech perception process. Important to keep in mind is that perceptual discrimination is necessary, but not sufficient for the P300 to appear. The P300 amplitude is also influenced by task-dependent cognitive processes such as immediate working memory of the stimulus and attention allocation (Polich, 2004, 2010). Differences in the development of the auditory processing function may have resulted in differences in these memory and attention processes used for speech perception, resulting in more difficulty for some users as opposed to others. Conclusions on the exact contribution of processes underlying the link between the P300 amplitude and speech perception should be drawn with caution. Differences in working memory and attention allocation processes may appear independently of CI or deafness. This also has implications for the specificity of the P3 amplitude as a marker for speech perception ability. If a CI user presents with an absent P3, we cannot be sure whether this is the result of a CI and/or deafness related auditory discrimination deficit, the result of a (general) deficit in working memory updating, or simply lack of attention from the CI user during the task. However, in terms of sensitivity of the marker, it is unlikely that a P3 is present while perceptual discrimination is not.

The relation between behavioural speech perception and P300 amplitude in the consonant condition furthermore shows that tone discrimination (frequency contrast) says less about speech perception abilities than speech discrimination (consonant contrast). As was already laid out in the introduction, the relation between the consonant contrast /ba/ vs. /da/ and behavioural speech perception has not been investigated much in CI populations. Earlier articles on the P300 (Groenen et al., 2001) and the MMN (Kelly et al., 2005) found a relation between speech perception scores and frequency and vowel contrasts as opposed to consonant contrasts. However, it is hard to compare our findings to theirs for several reasons. Firstly, they tested postlingually deaf adults. Secondly, ERPs for consonant contrasts were not measured (Kelly et al., 2005) or much less robust (Groenen et al., 2001). Thirdly, it was rather the duration of the MMN, not the amplitude, that was significantly correlated with speech perception scores in Kelly et al. (2005).

While the former relation was on the verge of significance, the relation between duration of deafness and amplitude of the P300 was robust. This was again found for the consonant contrast condition only. This implies that the longer a CI user has been deaf before implantation, the lower their P300 amplitude in response to speech stimuli. This finding is evidence for the fact that the auditory cortex is flexible enough to adapt to speech input after implantation, when performed early in development as is nowadays more and more the case with prelingually deaf CI users. Adaptation success seems to decrease as a function of the duration of speech deprivation (other factors that may play a role in this process put aside). This finding has been robustly established in the literature on obligatory cortical auditory evoked potentials (CAEP; Sharma, Dorman & Spahr, 2002a; Sharma, Spahr & Dorman, 2002b) but has not often been confirmed using the discriminative CAEPs MMN and P300. Our relatively short P300 experiment was able to capture this, and the findings correspond well to the correlation between behavioural speech perception and P300 amplitude in response to speech. This is evidence that the P300 response, as we measured it, is meaningful, despite individual waveform noise.

The results of the correlation measures are in line with our other behavioural results. That is, for the deviant counting during the P300 experiments, CI users (and this is also true for some normal-hearing participants) showed more deviations from the normal-hearing mean in the consonant condition than in the frequency condition. This implies that the stimuli in the consonant condition were somewhat more difficult to discriminate. However, results should be interpreted with caution. Someone with a low score on deviant counting does not have to have a lower discrimination ability. Deviant counting also measures processes other than perceptual discrimination, such as context-updating and attention allocation. The main function of the deviant-counting task was to make sure our participants paid attention to the stimulus. The context updating required for the deviant counting task is not expected to present a confound to our results, because it is expected that perceptual difficulties occur prior to context updating and also influence it linearly. The results of the behavioural reaction time task also showed a trend towards the CI users being slower than the normal-hearing controls, but only for the consonant condition. It may be due to a small sample size that this trend was not significant. This greater spread and difficulty for CI users in the contrast condition may have led

to this condition being a more sensitive measure for differences in auditory processing of speech.

A relation between latency and duration of deafness or behavioural speech perception was not found. An explanation for this may be that the single-subject waveform method is low in power. Rho-values were found in the right direction for both conditions, but they were very small. Latency may not be the most suitable measure to elicit relations between individual ERPs and behavioural outcomes, if any present.

Recommendations and limitations

The current study shows that as a clinical predictor, the P300 is more robust on an individual level with a limited amount of duration. For future research it is important to focus on this P300 response and to extend the current findings to a population of young children. For this population, it is even more important than for adolescents to obtain objective and all-round information on their auditory abilities.

Furthermore, the current study replicates, under contemporary implantation circumstances, the earlier found quality of the P300 response as a possible marker of perceptional challenges for the cochlear implant population. It paves the way for further research into predicting the auditory processing of speech in more difficult conditions. For example, recent research has focused on measuring the P300 in noise in postlingually deaf elderly CI users. It was found that for the CI users as a group, the P300 response was absent in the measurements with white noise as opposed to the measurements in quiet (Soshi et al., 2014). Measuring the P300 in noise may be the first step towards highlighting more subtle differences in processing. Furthermore, on a behavioural level, testing speech perception abilities in different conditions of noise also yields promising opportunities to investigate in more detail the challenges for CI users. Differences have already been found for speech perception ability masked by two-talker babble as opposed to steady-state noise (Hillcock-Dunn, Taylor, Buss & Leibold, 2015 on hearing-aid users) or white noise (Soshi et al., 2014 on CI users)

The assumed ceiling effect in our study showed that speech in isolation for early implanted, prelingually deaf adolescent CI users may be peer-like. This implies that the results of our study are in favour of the hypothesis of de Hoog et al. (2016a), who proposed a discrepancy between linguistic problems and speech perception problems for some, but not all, CI users. The linguistic problems of our sample were not taken as a variable in this study, but it was observed in the clinic that linguistic ability varied considerably and that it did not show a one-to-one mapping with behavioural speech perception scores. This indicates that this discrepancy should gain attention in future research. It should, furthermore, be investigated whether the discrepancy persists when measuring speech perception in more ecologically valid conditions.

This study is limited by its small sample size. It should also be taken into account that the signal-to-noise ratio of individual waveforms remains low (both found for CI users and for controls) and this may have distorted interpretations. We, however, have tried our best to diminish this risk by using new, well-argued-for analysis methods. The trade-off between more robust results for the group level (latency analysis) or more information on individual performance (amplitude analysis) remains a problem.

Conclusions

Our study shows that when taking into account suitability for children and clinical utility in general, an attentive task paradigm (eliciting the P300 component), as opposed to an inattentive task paradigm (eliciting the MMN component), is the most robust in highlighting auditory discrimination abilities, at least when using frequency and consonant contrasts. The P300 component is robust on an individual level. When latency of the component is taken as a measure the component can distinguish between complex vs. simple conditions, even with a small sample size. The P300 component, however, cannot (yet) distinguish between CI users and normal-hearing controls, or between individual CI users. This may be due to a ceiling effect in the performance of the CI users: their speech perception is relatively good, and the input conditions may have made the auditory discrimination too easy. The suitability of the P300 as a marker for speech perception ability is backed up by the findings that the amplitude of the P300 is related to behavioural speech perception as measured in the clinic and duration of deafness. Future research should focus on investigating the P300 in relation to behavioural speech perception in younger CI users, as well as using more challenging, ecologically valid experimental conditions.

Acknowledgements

A special thanks to all the cochlear implant users

that were willing to participate in this study. I furthermore would like to thank all the normal-hearing students that participated, Milou van Helvert for helping me with the data collection, and Randi Goertz and Marlijn ter Bekke for proof-reading this thesis. I thank Dr. Margreet Langereis, Dr. Anneke Vermeulen and Dr. Andy Beynon for providing me with the creative foundations for this thesis, thorough theoretical support and their clinical expertise. Lastly, I thank my supervisor Vitória Piai for her excellent supervision during the past year, for sharing with me her ideas and her in-depth expertise, and for educating and supporting me where and whenever possible.

References

American Speech-Language-Hearing Association. (2004). *Cochlear implants*. [Technical Report]. Retrieved April 14, 2018, https://www.asha.org/policy/TR2004-00041/

Boersma, P., & Weenink, D. (2018). Praat: doing phonetics by computer [Computer program]. Version 6.0.39, retrieved 12 April 2018 from http://www.praat.org/

Bosman, A. J., Wouters, J., & Damman, W. (1995). Realisatie van een cd voor spraakaudiometrie in Vlaanderen. *Logopedie en foniatrie: maandblad van de Nederlandse vereniging voor logopedie en foniatrie*, *67*(9), 218-225.

Beynon, A. J., Snik, A. F., & van den Broek, P. (2002). Evaluation of cochlear implant benefit with auditory cortical evoked potentials. *International journal of audiology*, *41*(7), 429-435.

Beynon, A. J., Snik, A. F. M., Stegeman, D. F., & Van den Broek, P. (2005). Discrimination of speech sound contrasts determined with behavioral tests and event-related potentials in cochlear implant recipients. *Journal of the American Academy of Audiology*, *16*(1), 42-53.

Bishop, D. V. M., & Hardiman, M. J. (2010). Measurement of mismatch negativity in individuals: A study using single-trial analysis. *Psychophysiology*, *47*(4), 697-705.

van Casteren, M., & Davis, M. H. (2006). Mix, a program for pseudorandomization. *Behavior research methods*, *38*(4), 584-589.

Geers, A. E. (2002). Factors affecting the development of speech, language, and literacy in children with early cochlear implantation. *Language, Speech, and Hearing Services in Schools*, *33*(3), 172-183.

Gilley, P. M., Sharma, A., Dorman, M., Finley, C. C., Panch, A. S., & Martin, K. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clinical Neurophysiology*, *117*(8), 1772-1782.

Groenen, P. A., Beynon, A. J., Snik, A. F., & Broek, P. V. D. (2001). Speech-evoked cortical potentials recognition in cochlear implant users and speech. *Scandinavian audiology*, *30*(1), 31-40

Hendriks, M. P., Kessels, R. P. C., Gorissen, M. E. E., Schmand, B. A., & Duits, A. A. (2014). *Neuropsychologische diagnostiek. De klinische praktijk*. Amsterdam: Boom.

Hillock-Dunn, A., Taylor, C., Buss, E., & Leibold, L. J. (2015). Assessing speech perception in children with hearing loss: What conventional clinical tools may miss. *Ear and hearing*, *36*(2), e57-60.

de Hoog, B. E., Langereis, M. C., Weerdenburg, M., Knoors, H. E., & Verhoeven, L. (2016a). Linguistic profiles of children with CI as compared with children with hearing or specific language impairment. *International journal of language & communication disorders*, *51*(5), 518-530.

de Hoog, B. E., Langereis, M. C., van Weerdenburg, M., Keuning, J., Knoors, H., & Verhoeven, L. (2016b). Auditory and verbal memory predictors of spoken language skills in children with cochlear implants. *Research in developmental disabilities*, *57*, 112-124.

Johnson, J. M. (2009). Late auditory event-related potentials in children with cochlear implants: a review. *Developmental neuropsychology*, *34*(6), 701-720.

Jordan, K., Schmidt, A., Plotz, K., Von Specht, H., Begall, K., Roth, N., & Scheich, H. (1997). Auditory event-related potentials in post-and prelingually deaf cochlear implant recipients. *The American journal of otology*, *18*(6 Suppl), S116-117.

Jung, T. P., Makeig, S., Humphries, C., Lee, T. W., Mckeown, M. J., Iragui, V., & Sejnowski, T. J. (2000). Removing electroencephalographic artifacts by blind source separation. *Psychophysiology*, *37*(2), 163-178.

Kelly, A. S., Purdy, S. C., & Thorne, P. R. (2005). Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. *Clinical Neurophysiology*, *116*(6), 1235-1246.

Kiesel, A., Miller, J., Jolicœur, P., & Brisson, B. (2008). Measurement of ERP latency differences: A comparison of single-participant and jackknife-based scoring methods. *Psychophysiology*, *45*(2), 250-274.

Kileny, P. R. (1991). Use of electrophysiologic measures in the management of children with cochlear implants:

brainstem, middle latency, and cognitive (P300) responses. *The American journal of otology*, *12*(Suppl), 37-42.

Kileny, P. R., Boerst, A., & Zwolan, T. (1997). Cognitive evoked potentials to speech and tonal stimuli in children with implants. *Otolaryngology-head and neck surgery*, *117*(3), 161-169.

Kilner, J. M. (2013). Bias in a common EEG and MEG statistical analysis and how to avoid it. *Clinical Neurophysiology*, *124*(10), 2062-2063.

Kraus, N., Micco, A. G., Koch, D. B., McGee, T., Carrell, T., Sharma, A., Wiet, R. J. & Weingarten, C. Z. (1993). The mismatch negativity cortical evoked potential elicited by speech in cochlear-implant users. *Hearing research*, *65*(1-2), 118-124.

Luck, S. J. (2005). Ten simple rules for designing ERP experiments. In T. C. Hardy (Ed.), *Event-related potentials: A methods handbook* (pp. 17-32). Cambridge, Massachussetts: The MIT Press

Luck, S. J. (2012). Event-related potentials. In H. Cooper, P. M. Camic, D. L. Long, A. T. Panter, D. Rindskopf, & K. J. Sher (Eds.), *APA handbook of research methods in psychology, Vol. 1. Foundations, planning, measures, and psychometrics* (pp. 523-546). Washington, DC, US: American Psychological Association.

Maris, E., & Oostenveld, R. (2007). Nonparametric statistical testing of EEG-and MEG-data. *Journal of neuroscience methods*, *164*(1), 177-190.

Micco, A. G., Kraus, N., Koch, D. B., McGee, T. J., Carrell, T. D., Sharma, A., Nicol, T. & Wiet, R. J. (1995). Speech-evoked cognitive P300 potentials in cochlear implant recipients. *American Journal of Otology*, *16*(4), 514-520.

Näätänen, R. (2001). The perception of speech sounds by the human brain as reflected by the mismatch negativity (MMN) and its magnetic equivalent (MMNm). *Psychophysiology, 38*(1), 1-21.

Näätänen, R., Paavilainen, P., Rinne, T., & Alho, K. (2007). The mismatch negativity (MMN) in basic research of central auditory processing: a review. *Clinical neurophysiology*, *118*(12), 2544-2590.

Obuchi, C., Harashima, T., & Shiroma, M. (2012). Auditory evoked potentials under active and passive hearing conditions in adult cochlear implant users. *Clinical and experimental otorhinolaryngology*, *5*(Suppl 1), 6-9.

Oostenveld, R., Fries, P., Maris, E., & Schoffelen, J. M. (2011). FieldTrip: open source software for advanced analysis of MEG, EEG, and invasive electrophysiological data. *Computational intelligence and neuroscience*, *2011*, 1.

Pisoni, D. B., Cleary, M., Geers, A. E., & Tobey, E. A. (1999). Individual differences in effectiveness of cochlear implants in children who are prelingually deaf: New process measures of performance. *The Volta Review*, *101*(3), 111-164.

Pisoni, D. D., & Geers, A. E. (2000). Working memory in deaf children with cochlear implants: Correlations between digit span and measures of spoken language processing. *The Annals of otology, rhinology & laryngology. 109*(Suppl 12) , *92-93*.

Polich, J. (1987). Task difficulty, probability, and inter-stimulus interval as determinants of P300 from auditory stimuli. *Electroencephalography and Clinical Neurophysiology/Evoked Potentials Section*, *68*(4), 311-320.

Polich, J., & Kok, A. (1995). Cognitive and biological determinants of P300: an integrative review. *Biological psychology*, *41*(2), 103-146.

Polich, J. (2004). Neuropsychology of P3a and P3b: A theoretical overview. *Brainwaves and mind: Recent developments*, 15-29.

Polich, J. (2012). Neuropsychology of P300. In S.J. Luck, E.S. Kappenman (Eds.), *Oxford handbook of event-related potential components*,159-188. New York: Oxford University Press Inc.

Ponton, C. W., Eggermont, J. J., Don, M., Waring, M. D., Kwong, B., Cunningham, J., & Trautwein, P. (2000). Maturation of the mismatch negativity: effects of profound deafness and cochlear implant use. *Audiology and Neurotology*, *5*(3-4), 167-185.

Ruffin, C. V., Kronenberger, W. G., Colson, B. G., Henning, S. C., & Pisoni, D. B. (2013). Long-term speech and language outcomes in prelingually deaf children, adolescents and young adults who received cochlear implants in childhood. *Audiology and Neurotology*, *18*(5), 289-296.

Schorr, E. A., Roth, F. P., & Fox, N. A. (2008). A comparison of the speech and language skills of children with cochlear implants and children with normal hearing. *Communication Disorders Quarterly*, *29*(4), 195.

Sharma, A., Dorman, M. F., & Spahr, A. J. (2002a). A sensitive period for the development of the central auditory system in children with cochlear implants: implications for age of implantation. *Ear and hearing*, *23*(6), 532-539.

Sharma, A., Spahr, A., Dorman, M., & Todd, N. W. (2002b). Early cochlear implantation in children allows normal development of central auditory pathways. *Annals of Otology, Rhinology & Laryngology*, *111*(Suppl 5), 38-41.

Singh, S., Liasis, A., Rajput, K., Towell, A., & Luxon, L. (2004). Event-related potentials in pediatric cochlear implant patients. *Ear and hearing*, *25*(6), 598-610.

Soshi, T., Hisanaga, S., Kodama, N., Kanekama, Y., Samejima, Y., Yumoto, E., & Sekiyama, K. (2014). Event-related potentials for better speech perception in noise by cochlear implant users. *Hearing research*, *316*(C), 110-121.

Svirsky, M. A., Robbins, A. M., Kirk, K. I., Pisoni, D. B., & Miyamoto, R. T. (2000). Language development in profoundly deaf children with cochlear implants. *Psychological science*, *11*(2), 153-158.

Turgeon, C., Lazzouni, L., Lepore, F., & Ellemberg, D. (2014). An objective auditory measure to assess speech recognition in adult cochlear implant users. *Clinical Neurophysiology*, *125*(4), 827-835.

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring LRP onset effects in factorial designs. *Psychophysiology*, *38*(5), 816-827.

Viola, F. C., De Vos, M., Hine, J., Sandmann, P., Bleeck, S., Eyles, J., & Debener, S. (2012). Semi-automatic attenuation of cochlear implant artifacts for the evaluation of late auditory evoked potentials. *Hearing research*, *284*(1-2), 6-15.

Watson, D. R., Titterington, J., Henry, A., & Toner, J. G. (2007). Auditory sensory memory and working memory processes in children with normal hearing and cochlear implants. *Audiology and Neurotology*, *12*(2), 65-76.

Woodman, G. F. (2010). A brief introduction to the use of event-related potentials in studies of perception and attention. *Attention, Perception, & Psychophysics*, *72*(8), 2031-2046.

Appendix A

Table A1.
Time-windows per group where there was a difference between standard and deviant waveforms and their p-values, for the attentive and inattentive x the frequency and consonant conditions. In bold are the normal-hearing controls that were matched to the CI users to be included in the amplitude and latency analyses performed over the attentive-task conditions.

	Attentive (P300)		**Inattentive (MMN)**	
Group	Frequency	Consonant	Frequency	Consonant
Patients (n=8)				
1	0.288-0.542(**)	0.350-0.700(**)	0.089-0.155(*) 0.175-0.209(*)	0.208-0.280(*)
2	0.378-0.567(**)	0.444-0.641(**)	NS	NS
3	0.149-0.212(*) 0.352-0.454(*)	0.358-0.509(**)	0.091-0.147(**)	NS
4	0.255-0.643(**)	0.374-0.720(**)	0.120-0.170(**)	NS
5	0.475-0.546(*) 0.315-0.389(*)	0.302-0.497(**) 0.501-0.663(**)	NS	NS
6	0.244-0.660(**)	0.432-0.688(**)	0.155-0.208(**)	NS
7	0.278-0.589(**)	0.343-0.690(**)	0.104-0.210(**)	NS
8	0.221-0.438(**)	0.334-0.616(**)	NS	NS
Controls (n=14)				
1	0.223-0.401(**) 0.470-0.538(**) 0.413-0.454(*)	0.319-0.464(**)	NS	NS
2	0.221-0.577(**)	0.333-0.755(**)	0.085-0.141(*)	NS
3	0.259-0.575(**)	0.411-0.645(**)	NS	NS
4	0.274-0.653(**)	0.331-0.624(**)	0.302-0.350(**)	NS
5	0.300-0.575(**)	0.427-0.523(**) 0.545-0.630(**)	NS	0.270-0.311(*)
6	0.309-0.704(**)	0.244-0.315(**) 0.421-0.741(**)	0.085-0.136(**)	NS
7	0.278-0.667(**) 0.192-0.233(*)	0.401-0.741(**)	0.079-0.202(**)	NS
8	0.286-0.602(**)	0.346-0.712(**)	0.096-0.138(*)	NS
9	0.298-0.501(**)	0.372-0.565(**)	Missing	NS
10	0.288-0.554(**)	0.386-0.651(**)	NS	NS
11	0.270-0.532(**)	0.427-0.561(**)	NS	NS
12	0.263-0.471(**)	0.339-0.579(**) 0.587-0.671(*)	NS	NS
13	0.296-0.440(**) 0.546-0.632(*)	0.401-0.528(**) 0.532-0.598(**) 0.610-0.749(**)	NS	NS
14	0.366-0.452(**) 0.497-0.563(**)	0.487-0.575(**) 0.594-0.659(**)	0.108-0.149(*) 0.153-0.220(*)	NS

Note. * p < .05, ** p < .01, NS = not significant