Hearing Research 316 (2014) 110–121

Contents lists available at ScienceDirect

Hearing Research

journal homepage: www.elsevier.com/locate/heares

Research paper

Event-related potentials for better speech perception in noise by cochlear implant users

Takahiro Soshi [a], Satoko Hisanaga [a], Narihiro Kodama [b], Yori Kanekama [a], Yasuhiro Samejima [b], Eiji Yumoto [b], Kaoru Sekiyama [a, *]

[a] *Division of Cognitive Psychology, Faculty of Letters, Kumamoto University, 2-40-1, Kurokami, Chuo-ku, Kumamoto 860-8555, Japan*
[b] *Department of Otolaryngology-Head and Neck Surgery, Graduate School of Medicine, Kumamoto University, 1-1-1, Honjo, Chuo-ku, Kumamoto 860-0811, Japan*

ARTICLE INFO

Article history:
Received 2 December 2013
Received in revised form
25 July 2014
Accepted 5 August 2014
Available online 23 August 2014

ABSTRACT

Speech perception in noise is still difficult for cochlear implant (CI) users even with many years of CI use. This study aimed to investigate neurophysiological and behavioral foundations for CI-dependent speech perception in noise. Seventeen post-lingual CI users and twelve age-matched normal hearing adults participated in two experiments. In Experiment 1, CI users' auditory-only word perception in noise (white noise, two-talker babble; at 10 dB SNR) degraded by about 15%, compared to that in quiet (48% accuracy). CI users' auditory-visual word perception was generally better than auditory-only perception. Auditory-visual word perception was degraded under information masking by the two-talker noise (69% accuracy), compared to that in quiet (77%). Such degradation was not observed for white noise (77%), suggesting that the overcoming of information masking is an important issue for CI users' speech perception improvement. In Experiment 2, event-related cortical potentials were recorded in an auditory oddball task in quiet and noise (white noise only). Similarly to the normal hearing participants, the CI users showed the mismatch negative response (MNR) to deviant speech in quiet, indicating automatic speech detection. In noise, the MNR disappeared in the CI users, and only the good CI performers (above 66% accuracy) showed P300 (P3) like the normal hearing participants. P3 amplitude in the CI users was positively correlated with speech perception scores. These results suggest that CI users' difficulty in speech perception in noise is associated with the lack of automatic speech detection indicated by the MNR. Successful performance in noise may begin with attended auditory processing indicated by P3.

1. Introduction

Nowadays, a CI is the most effective neural prosthesis for delivering auditory information to patients with profound deafness by bypassing the damaged inner ear and directly stimulating the auditory nerves (Zeng, 2004). With the use of a CI, post-lingual deaf patients rapidly improve speech perception within the first year of surgery (Hamzavi et al., 2003; Rouger et al., 2007; Ruffin et al., 2007). On the other hand, speech perception in noise is still difficult for CI users even after several years of device use (Tyler et al., 1995; Nelson et al., 2003; Nelson and Jin, 2004; Fu and Nogaki, 2005; Davidson et al., 2010). It is an immediate issue to be clarified as to what behavioral and neural foundations are responsible for speech perception in noise with CI use.

Neurophysiological studies have investigated the neural foundations for CI-dependent auditory performance in quiet, mainly using two event-related potentials (ERPs), that is, mismatch negativity (MMN) and P300 (P3) (Kaga et al., 1991; Kraus et al., 1993; Ponton and Don, 1995; Groenen et al., 2001).

The MMN is a negative ERP, appearing around 200 ms after stimulus onset, observed for deviant auditory stimuli compared with standard frequent stimuli (Näätänen et al., 1978; Kraus et al., 1992). The MMN may originate mainly from the superior and

---

Abbreviations: 2T, two-talker; AEP, auditory evoked potential; ANOVA, analysis of variance; AO, auditory-only; AV, auditory-visual; CI, cochlear implant; DF, deafness; EEG, electroencephalogram; EOG, electro-oculogram; ERP, event-related potential; LSD, least significant difference; MMN, mismatch negativity; MNR, mismatch negative response; N, noise; N1, N100; N2, N200; NH, normal hearing; P2, P200; P3, P300; RT, response time; SD, standard deviation; SNHL, sensorineural hearing loss; SNR, signal-to-noise ratio; SPL, sound pressure level; Q, quiet; WN, white noise

* Corresponding author. Tel.: +81 96 342 2845.
E-mail address: sekiyama@kumamoto-u.ac.jp (K. Sekiyama).

http://dx.doi.org/10.1016/j.heares.2014.08.001
0378-5955/

middle temporal areas (Marco-Pallarés et al., 2005; Näätänen et al., 2007) and reflects automatic auditory detection of deviant stimuli (Näätänen and Gaillard, 1983; Näätänen et al., 2007). Under attended conditions, MMN is overlapped by an attention-related posterior negativity (N2b) that peaks at around 250 ms (Näätänen and Gaillard, 1983; Novak et al., 1992; Cowan et al., 1993; Näätänen et al., 2007).

The MMN has been observed for good CI performers, but not for poor CI performers. Kraus et al. (1993) recorded the MMN response from good CI performers, using a passive auditory oddball task with speech. Similar findings about MMN elicitation for good CI performers have been reported in several studies (adult/speech: Groenen et al., 1996b; children/speech: Singh et al., 2004; adult/tone: Kelly et al., 2005; Zhang et al., 2011; Lonka et al., 2013).

P3 is another ERP component used in CI-related ERP studies. The P3 is the third positive component typically observed for attended rare targets in an active oddball task (Squires et al., 1975; Picton, 1992). Because P3 does not appear for an undetected change of stimulus properties, the elicitation is associated with an attentional evaluation of stimulus change (Donchin et al., 1978). The latency has a wide range from about 300 ms to over 600 ms after stimulus onset. The scalp distribution has a centro-posterior maximum.

P3 is also observed for good CI performers, but not for poor CI performers (Kaga et al., 1991; Oviatt and Kileny, 1991; Micco et al., 1995; Groenen et al., 1996a, 2001). Oviatt and Kileny (1991) observed that one poor CI performer could not detect stimulus change in an active oddball task, not showing the P3 to the deviant tone, while the other nine CI users could detect stimulus change, eliciting the P3.

In contrast to speech perception in quiet, very little is known about CI users' neurophysiological foundations for auditory speech perception in noise. The current study investigates the neurophysiological responses of CI users to auditory speech in noise. Participants were post-lingual adult CI users having at least 2 years of CI use, with NHs as controls. As with previous studies, we also used an auditory oddball paradigm with consonant-vowel syllables (/ba/and/ga/), comparing neurophysiological responses between deviant and non-deviant stimuli.

The main predictions of ERP results are as follows: present CI users having already used a CI device for more than 2 years, likely show good syllable detection in quiet (Hamzavi et al., 2003; Rouger et al., 2007; Ruffin et al., 2007). Accordingly, they will elicit the MMN and the N2b ('N2 deflection' noted together hereafter as 'mismatch negative response: MNR') (Näätänen and Gaillard, 1983) to deviant stimuli in quiet, similar to the NH controls (Groenen et al., 1996b). The P3 to deviant stimuli may not appear, because syllable detection in quiet may be easy for both groups; thus, the selective evaluation of deviant stimuli as a task-relevant rare target may be attenuated (Picton, 1992).

In noise, the CI users with good syllable detection performance and the NH controls may also show MNRs to deviant stimuli. They may also elicit the P3, because speech in noise probably promotes attentional stimulus evaluation (Wong et al., 2008), enhancing evaluation of deviant stimuli as a rare target. On the other hand, poor CI performers may elicit neither MNR nor P3, because degraded speech perception at a poor SNR did not elicit either MNR or P3 even for NH people (Martin et al., 1997; Whiting et al., 1998; Kaplan-Neeman et al., 2006).

We also behaviorally tested auditory-only (AO) and auditory-visual (AV) word perception in quiet and noise for the purpose of delineating an overview of noise effects on CI-dependent speech perception (Experiment 1). Experiment 1 used two types of noise (white noise (WN) and two-talker babble (2T)). Talker noise is suitable to examine noise interference effects to CI users' speech perception in ordinary communicative situations. A two-talker babble may work as not only an energetic masker such as white noise, but also as an information masker of the target speech (Brungart et al., 2001; Freyman et al., 2004; Nelson and Jin, 2004; Cooke et al., 2008; Mattys et al., 2009). As a result, the talker noise may more severely affect CI-dependent speech perception, providing the significant information that CI users are vulnerable in speech perception at two levels of noise masking. The present CI users may be weak in AO word perception in noise, in general (Nelson et al., 2003; Fu and Nogaki, 2005). In addition, the CI users' AV word perception is likely to be more degraded in the 2T noise condition than in the WN condition (Carhart et al., 1969; Brungart et al., 2001 for review of NHs' AO performances in two types of noise) because differences in AO noise interference may be enhanced in AV word perception in multiplicative ways, as suggested by a previous study (Sumby and Pollack, 1954). Therefore, Experiment 1 included not only AO, but also AV conditions. The results of Experiment 1 will be reported first.

2. Methods

2.1. Experiment 1: behavioral measure of word perception

2.1.1. Participants

Seventeen CI and twelve NH participants took part in the experiment. The CI users were post-lingually deafened (>90 dB hearing level at all test frequencies), and were monaurally implanted. Mean age of the CI users was 63.2 ± 10.6 years old (41–80 years old). Mean duration of CI use was 8.0 ± 5.5 years (2.4–19.7 years). Mean duration of deafness (DF) was 6.3 ± 7.1 years (0.3–24 years). The etiology included sudden sensorineural hearing loss (SNHL), idiopathic progressive SNHL, mitochondrial disease, sequelae of chronic otitis media, and mumps. Their primary communication method was oral, and none of them used a hearing aid on a non-implanted ear. The CI users used their individual standard comfortable device settings throughout the experiments. Table 1 summarizes the main demographical and clinical properties of the CI users.

The NH participants matched to the CI users in age (NH: 62.3 ± 9.0 years old, range from 43 to 76 years old; $t_{(27)} = 0.262$, $p = 0.796$), and male-to-female ratio (CI: female:male, 12:5; NH: 8:4; $\chi^2_{(1)} = 0.051$, $p = 0.822$). The NH participants possessed normal hearing ability (<25 dB of average hearing level at 500, 1000, 2000, and 4000 Hz as defined by the World Health Organization: left, 14.2 ± 5.0 dB; right: 15.0 ± 5.7 dB). All of the CI and NH participants had normal or corrected-to-normal visual acuity, and were right-handed. They reported no cortical and psychiatric deficits. All of the participants provided written informed consent prior to participation. All of the procedures were approved by the Human Subjects Ethics Committee of Kumamoto University.

2.1.2. Stimuli

The stimulus set consisted of 8 lists of 25 Japanese words (each word contained about 3 morae, e.g.,/ha-shi-ra/, "pilar"; /shi-ro-i/, "white"). These lists were from the CI 2004 list set (Technical Committee on Cochlear Implants in Japan, 2004).

The experimental conditions consisted of AO and AV conditions. In the AO condition, stimuli consisted of auditory speech and a visual fixation point (+). The visual cross was presented 900 ms before the onset of auditory stimuli and provided the cue for the duration of the auditory speech. In the AV condition, stimuli contained both auditory and visual speech. The visual speech used actual facial articulatory movements. These two conditions had sub-conditions of quiet and noise: in the quiet (Q) conditions (AO-Q, AV-Q), auditory words were presented without background noise. In the noise conditions, two types of noise were added to

Table 1
Demographic and clinical profiles of the 17 CI users.

ID	Age	Sex	CI use duration (year)	DF duration (year)	Speech processor	Coding	Implanted ear	Etiology
1	41	M	10.6	0.8	Esprit	ACE	L	Mitochondrial disease
2	49	F	8.8	0.4	Esprit	ACE	R	Idiopathic progressive SNHL
3	52	F	3.5	3.2	Freedom	ACE	R	Idiopathic progressive SNHL
4	53	F	3.3	10.0	Freedom	ACE	R	Idiopathic progressive SNHL
5	57	M	9.0	0.6	Sprint	ACE	R	Idiopathic progressive SNHL
6	58	F	11.0	2.0	Freedom	ACE	L	Idiopathic progressive SNHL
7	62	F	6.4	1.0	Sprint	ACE	L	Sudden SNHL
8	64	F	6.1	1.0	COMBI 40+	FSP	L	Sudden SNHL
9	65	M	19.7	10.0	Freedom	ACE	R	Idiopathic progressive SNHL
10	65	F	19.7	16.0	Esprit-3G	SPEAK	R	Mumps
11	66	F	2.4	0.3	Freedom	ACE	R	Sudden SNHL
12	66	F	2.8	15.0	Freedom	ACE	L	Sudden SNHL
13	71	F	2.4	15.0	Freedom	ACE	R	Idiopathic progressive SNHL
14	73	F	8.1	1.0	Esprit-3G	ACE	R	Idiopathic progressive SNHL
15	76	F	3.7	5.0	Freedom	ACE	R	Idiopathic progressive SNHL
16	77	M	3.5	1.7	Freedom	ACE	R	Sequelae of chronic otitis media
17	80	M	14.8	24.0	Clarion S-series	CIE	L	Idiopathic progressive SNHL

F: female; M: male; ACE: advanced combination encoder; L: left; R: right; SNHL: sensorineural hearing loss.

auditory words. One was a white noise (AO-WN, AV-WN) and the other was a two-talker babble (AO-2T, AV-2T). Multiple talker babbles have often been used to effectively mask target speech (Carhart et al., 1969; Brungart et al., 2001; Freyman et al., 2004; Tyler et al., 2006). The two types of noise were expected to mask target speech differently (Brungart et al., 2001). White noise energetically masks the spectral and temporal information of speech signals (energetic masking) whereas, a two-talker babble likely yields not only energetic masking effects, but also information masking effects such as misperception of target's phonological information, higher cognitive load of divided attention, and/or word-knowledge competition (Freyman et al., 2004; Cooke et al., 2008; Mattys et al., 2009).

Word stimuli were recorded with a digital video camera and an audio recorder. A male speech-language pathologist articulated the words. The word stimuli were edited using a movie editing software. Video components of the movies were digitized at 29.97 frames per second at 720 × 480 pixels. The auditory components of the stimuli were digitized with 16-bit, 44100 Hz resolution, and were stored in stereo. The auditory speech was presented at a sound pressure level (SPL) of 65 dB.

To create the two-talker babble noise, continuous speech was recorded from two male speakers. One read an article from a Japanese newspaper, and the other read a story (story title: “The north wind and the sun”). Each of the speeches was recorded using a sound recorder, and was digitized with 16-bit, 44100 Hz resolution. The two-talker babble was created through combining the two stories. These noises were calibrated at 55 dB SPL (the SNR +10 dB relative to the 65 dB task-relevant speech stimuli). The output levels of the speech and noises were calibrated at the position of the listener's ear. While a ceiling effect had occurred for NH participants with the SNR +10 dB in the preliminary study, this SNR was used to avoid a floor effect in CI users.

2.1.3. Procedure

The participants were seated 0.9 m in front of a loudspeaker at ear level, and 0.6 m in front of a 17-inch monitor. Speech and noise sounds were presented through a two-channel audio mixer to a loudspeaker. Participants were instructed to listen carefully to the speech stimuli, and to repeat each word aloud correctly. The examiner scored the test during the trial. The participants were tested using random orders of the six conditions. Lists of the CI 2004 and conditions within each of the AO and AV conditions were counterbalanced across the participants. If they were not sure what they heard, they were encouraged to guess and produce a response. The CI2004 test was actually conducted after the ERP measurement (Experiment 2). One CI user could not come back for the CI2004 test due to a scheduling conflict.

2.1.4. Statistical analysis

Response accuracy for the 16 CI users was tested with a two-way analysis of variance (ANOVA) with factors of modality (AO, AV) and condition (Q, WN, 2T). Because the interaction between modality and condition was almost significant, planned post-hoc analyses were conducted for each modality and condition. Multiple comparisons were done by Fisher's least significant difference (LSD) method. Almost all of the NH participants were perfect in all of the conditions, and therefore, were not statistically tested. A Greenhouse-Geisser correction was performed when the sphericity assumption about the variance of differences was violated, and will be reported with unmodified degrees of freedom and epsilon.

2.2. Experiment 2: neurophysiological and behavioral measures of syllable detection

2.2.1. Participants

The same CI and NH participants participated in Experiment 2. Among the 17 CI users, ERP analyses included 12 CI users (62.9 ± 10.9 years old) for the noise (N) condition and 7 CI users (63.9 ± 8.2 years old) for the quiet (Q) condition. The inclusion conditions will be described later. Throughout the analyses, the CI users matched to the NH participants (62.3 ± 9.0 years old) in age (12 CI users: $t_{(22)} = 0.163$, $p = 0.872$; 7 CI users: $t_{(17)} = 0.388$, $p = 0.703$) and male-to-female ratio (12 CI users: female:male, 8:4; NH: 8:4; 7 CI users: 6:1, $\chi^2_{(1)} = 0.083$, $p = 0.363$).

2.2.2. Stimuli

We prepared/ga/and/ba/auditory stimuli. These syllables differ only in place of articulation, and are difficult to discriminate in noise (Miller and Nicely, 1955). The stimuli were monaurally recorded with a sampling depth of 16 bits and a sampling rate of 32000 Hz, while a female talker was articulating/ba/and/ga/. To minimize vowel differences, the two stimuli were modified by TANDEM-STRAIGHT software (Kawahara et al., 2009): we extracted F0 traces of the two stimuli, equating mean F0 frequencies of the vowel parts (about 200 Hz). Durations of both stimuli were 330 ms.

Use of natural speech with relatively long stimulus duration may be optimal for observing clear ERP effects for speech perception (Henkin et al., 2009), although stimuli with a short duration can prevent contamination of large CI device-related artifacts. A square pulse trigger was added to the second channel of each sound file, and the trigger was recorded synchronously with event-related brain response by an EEG amplifier.

2.2.3. Experimental conditions

In the electroencephalogram (EEG) recording, each of the two groups experienced four experimental conditions with two within-participants factors: one was the frequency of the target stimuli (30% in the deviant condition and 100% in the control condition), and the other was the noise factor (absent or present). The four experimental conditions were conducted in separate blocks.

In the behavioral measurement, only the deviant condition was conducted to examine auditory discrimination between/ba/and/ga/. Each of the two groups underwent the Q and N conditions in separate blocks.

2.2.4. Experimental procedure

We used a two-stimulus auditory oddball paradigm to evaluate syllable detection performance and relevant neurophysiological responses. High (base) and low (deviant) probability of syllables was presented pseudo-randomly. Participants were instructed to count the numbers of all stimuli without physical movement in order to avoid contamination of motor-related artifacts into the EEG as well as to confirm whether or not the participants possessed normal cognitive function in counting numbers in their mind. Participants reported the number of stimuli at the end of each block.

In the deviant condition, we presented the deviant stimulus/ga/(30%) and base stimulus/ba/(70%) in a pseudo-randomized order. The condition contained a total of 180 stimuli (base/ba/: 126 times, 70%; deviant/ga/: 54 times, 30%) and consisted of 5 blocks. Each block included pseudo-randomized numbers of stimuli (base: 33 ± 5 times; deviant: 14 ± 2 times; the ratio of deviant stimuli: 30 ± 0.03%) and always repeated the base/ba/five times at the beginning.

In the control condition, the same syllable/ga/was always presented as the standard stimuli. ERPs for the standard stimuli served as the baseline for examining ERPs for the deviant stimuli presented in the deviant condition. The control condition included the 54/ga/stimuli (the standard/ga/) and consisted of 2 blocks.

The deviant and control conditions were conducted in quiet and white noise. In the Q condition, the syllable stimuli were presented without white noise. In the N condition, the syllable stimuli were presented with white noise. To prevent the participants from anticipating the appearance of deviant stimuli during the control condition, the participants first performed the control conditions in quiet and noise (4 blocks), and then secondly, the deviant conditions in quiet and noise (10 blocks).

Participants listened to the syllable stimuli from a loudspeaker located 1 m in front of their heads, while facing a 17-inch monitor at a distance of 0.9 m. The loudspeaker was located on top of the monitor. The CI users underwent free-field stimulation through the device's microphone. Each test session started with the participant's button press at the presentation of the trial instruction. After the instruction disappeared, a gray fixation mark (+) continued to appear in the center of a black screen while auditory stimuli were presented. Syllable stimuli were presented at a mean interval of about 1050 ms (stimulus onset asynchrony ranging from 1000 to 1090 ms). The syllable stimuli were presented together with white noise at the SNR +10 dB. The SPL of syllables and white noise were set to 65 dB and 55 dB, respectively. The output levels of the syllable and noise were calibrated at the position of the listeners' ears. The syllable and noise sounds were routed through a two-channel audio mixer to a loudspeaker. In both the deviant and control conditions, the participants were instructed to maintain their gaze on the fixation symbol.

After the EEG recording had finished, the participants performed behavioral trials to evaluate their syllable detection performance (percentage correct and response times (RT) for button response). Forty stimuli were presented in the Q and N conditions (base/ba/: 28 times, 70%; deviant/ga/: 12 times, 30%). Both conditions repeated the base/ba/five times at the beginning. The stimulus parameters were the same as those of the EEG recording, while the participants responded to each stimulus by a button press to discriminate/ba/and/ga/.

2.2.5. Electroencephalogram recording and analysis

The EEG was continuously recorded from three Ag/AgCl midline scalp electrodes (anterior: Fz; central: Cz; posterior: Pz). Three additional electrodes were placed around the eyes for recording horizontal (left-upper minus right-upper) and vertical (left-upper minus left-lower) electro-oculograms (EOG). All electrodes were referenced to the tip of the nose. The ground electrode was situated on the participant's forehead (Fpz). The EEG was recorded at a sampling frequency of 1000 Hz with a band-pass frequency ranging from DC to 300 Hz (thus no filtering for lower frequencies). The impedance was set below 5000 Ω throughout the recording.

It has been reported that EEG waveforms are contaminated by CI device pulse artifacts (Singh et al., 2004; Gilley et al., 2006). In fact, our individual EEG data for the Q condition were often strongly contaminated by CI artifacts. On the other hand, the EEG data for the N condition did not show CI artifact contamination for almost all of the CI users; therefore, we did not perform a CI pulse reduction algorithm (Cf. Singh et al., 2004; Zhang et al., 2011). Our approach is not exceptional, and has been taken in some studies previously (Kraus et al., 1993; Kelly et al., 2005; Henkin et al., 2009).

The continuous EEG data of the individual participants (12 CI and 12 NH participants for the N condition, 7 CI and 12 NH for the Q condition) were first filtered with a low-cutoff frequency of 0.5 Hz (24 dB/octave, zero-phase shift) and a high-cutoff frequency of 40 Hz (24 dB/octave; zero-phase shift). The EEG was segmented into each epoch from 200 ms before to 1000 ms after the onset of the standard/ga/(54 epochs) and the deviant/ga/(54 epochs) in each condition. Individual-averaged waveforms were calculated after baseline correction (mean potentials during the baseline interval from −200 to 0 ms), vertical EOG reduction (spatial singular value decomposition and spatial filtering of blink components, implemented by Scan 4.3, Compumedics Neuroscan, Inc., Charlotte, NC), and artifact rejection for residual artifacts (peak-to-peak amplitudes of ±75 μV). Mean rejection rates of the 4 conditions were 16.0 ± 2.3% for the CI users, and 14.2 ± 3.1% for the NH participants. Grand averaged waveforms were smoothed using the moving average method (21 data points, that is, 21 ms interval) only for ease of visual inspection.

2.2.6. Statistical analysis

Response accuracy in behavioral oddball performance for the 17 CI users was compared between the Q and N conditions with a paired *t*-test. Almost all of the NH participants performed the task perfectly with a ceiling effect. Thus, the accuracy data for the NH participants were not statistically tested. RTs for the 17 CI users and the 12 NH participants were tested by the ANOVA with the factors of condition (Q, N) and group (CI, NH). When significant interaction effects were obtained, pair-wise comparisons were conducted between conditions and/or groups. Response accuracies in counting the total number of stimuli during the EEG recordings were perfect for almost all of the CI users and the NH participants, and therefore, were not tested in statistical analysis.

For neurophysiological data, five CI users were excluded from the statistical tests, because (i) two users (No.8, 17) used CI devices supplied by different device manufacturers; (ii) two users (No. 7, 10) showed ERP waveforms contaminated by CI pulse artifacts in both the Q and N conditions; (iii) one user (No.2) showed an overall drift artifact for EEG waveforms even after high-pass filtering. Accordingly, a total of 12 CI users were introduced into the statistical analysis for the N condition. For the Q condition, 7 of the 12 CI users were included in the statistical analysis, because the other CI users showed CI artifacts contaminating into waveforms. All of the NH participants were statistically tested.

To explore neurophysiological effects, we examined peak latencies, durations, or amplitudes of neurophysiological components. We specified individual peak latencies of auditory evoked potentials (AEPs: N100 (N1), P200 (P2)) and ERPs (MNR, P3) in the Cz electrode temporally adjacent to peaks in grand average waveforms. Because an MNR peak was not found in the Cz electrode for one user among the seven CI users in the Q condition, peak latency in the Pz was alternatively used. The P3 latency could be specified only for the good CI performers and the NH controls in the N condition. Mean durations of the MNRs were specified by (i) intervals between the onset (the first 0 μV point) and offset (the last 0 μV point) of negative deflection of difference waveforms after appearance of the N1 (Singh et al., 2004), or (ii) intervals between the onset and offset of negative voltage increase of difference waveforms, if onset and/or offset 0 μV points were not detected. When no negative deflection was observed during the interval from 100 to 350 ms, the MNR duration was specified as 0 ms (Singh et al., 2004). Mean peak latencies and durations were compared between the CI and NH participants with unpaired *t*-tests, and/or between the Q and N conditions with paired *t*-tests. For amplitude, averaged waveforms for the standard/ga/and deviant/ga/were segmented into each 50 ms from speech onset (0 ms) and 1000 ms after stimulus onset. Mean amplitudes for the standard and deviant conditions in each electrode were compared with paired *t*-tests for the CI and NH groups. Consecutive significant time windows were combined in reporting results.

Based on response accuracy in the behavioral oddball performance in noise (the mean percent correct was 66%), the 12 CI users were grouped into good (>mean; $n = 5$) and poor (<mean; $n = 7$) performers to examine performance-related neurophysiological change. Both groups were not significantly different in age (good: 61 ± 14 years old; poor: 64 ± 10 years old; $t_{(10)} = 0.548$, $p = 0.596$), male-to-female ratio (good: female:male, 4:1; poor: 4:3; $\chi^2_{(1)} = 0.686$, $p = 0.408$), CI use duration (good: 5.4 ± 3.8 years; poor: 7.6 ± 6.2 years; $t_{(10)} = 0.712$, $p = 0.493$), and DF duration (good: 5.4 ± 6.7 years; poor: 5.4 ± 5.3 years; $t_{(10)} = 0.018$, $p = 0.986$). Although each group included small numbers of participants, the N1 and P2 latencies were compared between the groups with unpaired *t*-tests. Amplitudes for the standard and deviant stimuli were compared using paired *t*-tests in each 50 ms time window for both groups.

Correlation analyses were conducted between neurophysiological properties (peak latency, duration, or difference amplitude: deviant − standard) and response accuracy for both the oddball task and the CI2004 test in each 50 ms time window. Because chronological age, CI use duration, and DF duration may affect neurophysiological, as well as behavioral responses, partial correlation analyses were also performed using these factors as control variables. These variables were not significantly correlated with each other, keeping independency (age-CI: $r = -0.245$, $p = 0.443$; age-DF: $r = 0.183$, $p = 0.183$; CI–DF: $r = -0.131$, $p = 0.686$). A significant α level was set to $p < 0.05$ throughout the statistical analyses.

3. Results

3.1. Word perception performance in Experiment 1

In the two-way ANOVA with factors of modality (AO, AV) and condition (Q, WN, 2T) for the CI users ($n = 16$), the main effects of modality and condition were significant (modality: $F_{(1,15)} = 253.570$, $p < 0.0001$; condition: $F_{(2,30)} = 7.080$, $p = 0.003$). The interaction between modality and condition was almost significant ($F_{(2,30)} = 3.152$, $p = 0.057$), and then, follow-up ANOVAs for AO and AV modality were conducted. Both modalities yielded the significant main effect of condition (AO: $F_{(2,30)} = 7.059$, $p = 0.003$; AV: $F_{(2,30)} = 3.374$, $p = 0.048$). When the CI users did not use lipreading (AO condition), there was reduced response accuracy for the noise conditions (AO-WN, AO-2T), compared with the Q condition (AO-Q: mean ± standard deviation (SD), 47.8 ± 22.7%; AO-WN: 35.8 ± 15.6%; AO-2T: 31.3 ± 18.3%; LSD: AO-Q vs. AO-WN, $p = 0.009$; AO-Q vs. AO-2T: $p = 0.011$; AO-WN vs. AO-2T: $p = 0.241$) (Fig. 1A). For the AV condition, the difference in

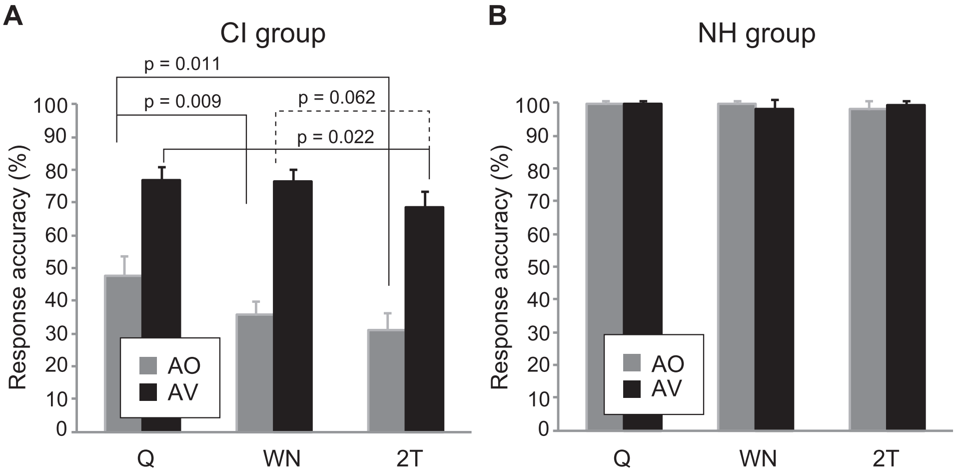

Fig. 1. Word perception performances for the cochlear implant (CI) users and normal hearing (NH) controls. (A): CI users' response accuracy in auditory-only (AO) and audio-visual (AV) word perception in quiet (Q), white noise (WN), and two-talker noise (2T). (B): NH's response accuracy in AO and AV word perception in Q, WN, and 2T. For the noise conditions, the signal-to-noise ratio was +10 dB.

response accuracy between the AV-Q and AV-WN conditions was not significant at all. The AV-2T condition, on the other hand, showed lower response accuracy than the AV-Q condition (AV-Q: 76.8 ± 16.5%; AV-WN: 76.5 ± 14.0%; AV-2T: 68.8 ± 18.3%; LSD: AV-Q vs. AV-2T: $p = 0.022$; AV-Q vs. AV-WN, $p = 0.944$; AV-WN vs. AV-2T: $p = 0.062$). These results suggest that the two-talker noise interfered more with AV word perception than white noise. As reported in previous studies, the CI users generally performed better when they used lip-reading (AO vs. AV for Q, WN, and 2T conditions: LSD, all of $p < 0.0001$). Almost all of the NH participants were perfect in all conditions (AO-Q: 99.7 ± 1.0%; AO-WN: 99.7 ± 1.0%; AO-2T: 98.0 ± 1.0%; AV-Q: 99.7 ± 1.0%; AV-WN: 98.3 ± 3.0%; AV-2T: 99.3 ± 2.0%) (Fig. 1B).

3.2. *Syllable detection performance in Experiment 2*

The CI users ($n = 17$) showed lower response accuracy in the N condition, compared to the Q condition (Q: 83.5 ± 15.5%; N: 67.4 ± 20.3%; $t_{(16)} = 3.087$, $p = 0.007$) (Fig. 2A). This indicates that CI users are vulnerable to syllable-level speech perception in noise even at above 2 years of CI use. Almost all of the NH participants were perfect in both the conditions (Q: 98.4 ± 3.3%; N: 97.7 ± 4.3%). The CI users generally responded more slowly than the NH participants (CI: 719 ± 128 ms; NH: 586 ± 70 ms; group: $F_{(1,27)} = 14.718$, $p = 0.001$) (Fig. 2B). RTs for the N and Q conditions were not significantly different in both the CI (Q: 709 ± 115 ms; N: 729 ± 142 ms; $t_{(16)} = 0.549$, $p = 0.591$) and NH participants (Q: 574 ± 59 ms; N: 598 ± 81 ms; $t_{(11)} = 1.459$, $p = 0.173$).

3.3. *Neurophysiological results in Experiment 2*

Five of our CI users showed contamination of CI pulse artifacts in the Q condition. The CI users excluded ($n = 5$) showed square-form pulse artifacts from the stimulus onset to about 400 ms for the standard and/or deviant conditions, while the CI pulse amplitudes varied among participants. However, almost all of the individuals' averaged waveforms in the noise condition ($n = 12$) did not show contamination of CI pulse artifacts. Accordingly, we calculated averaged ERP waveforms without reduction of the artifacts.

3.4. *Neurophysiological results for the normal hearing participants*

After about 200 ms from stimulus onset, the NH participants elicited MNRs to the deviant stimuli for both the Q and N conditions. In the Q condition (Fig. 3A), the MNR started from the P2 time window, peaking at 263 ± 31 ms and continuing for 233 ± 107 ms in the central-posterior sites (standard vs. deviant: Cz, 200–350 ms, $t_{(11)} = 2.707$, $p = 0.020$; Pz: 200–350 ms, $t_{(11)} = 3.588$, $p = 0.004$). The P3 did not appear in response to deviant stimuli in the Q condition.

In the N condition (Fig. 3B), the MNR appeared at relatively early time windows in the central-posterior sites (Cz: 150–350 ms, $t_{(11)} = 3.156$, $p = 0.009$; Pz: 150–350 ms, $t_{(11)} = 2.420$, $p = 0.034$). Actually, the MNR for the N condition peaked earlier than it did for the Q condition (N: 205 ± 29 ms; Q: 263 ± 31 ms; $t_{(11)} = 5.062$, $p < 0.0001$), indicating that noise facilitates early auditory processing. The NH participants also showed the P3 effect to deviant stimuli peaking at 757 ± 62 ms in the central electrode (Cz: 650–800 ms, $t_{(11)} = 4.434$, $p = 0.001$).

The NH controls generally showed about 100 ms shorter N1 and P2 peak latencies than the CI users in both the Q (N1: $t_{(17)} = 8.369$, $p < 0.0001$; P2: $t_{(17)} = 9.396$, $p < 0.0001$) and N conditions (N1: $t_{(22)} = 6.880$, $p < 0.0001$; P2: $t_{(22)} = 4.492$, $p = 0.0002$) (Table 2).

3.5. *Neurophysiological results for the CI users*

The CI users ($n = 7$) yielded an MNR to the deviant stimuli from 350 to 700 ms after stimulus onset in the Q condition (Cz: $t_{(6)} = 3.457$, $p = 0.014$) (Fig. 4A). These CI users, as well as the five CI users excluded from the analysis, showed high response accuracy in the behavioral oddball task in quiet (7 CI: 88.6 ± 11.7%; 5 CI: 88.7 ± 16.1%). This supports the previous finding that the MNR is associated with good speech perception in the CI users. On the other hand, the MNR for the CI users peaked about 100 ms later than that of the NH participants (CI: 350 ± 55 ms; NH: 263 ± 31 ms; $t_{(17)} = 4.412$, $p < 0.0001$), as well as the N1 and P2 components for auditory sensory processing. The MNR durations were not significantly different between the CI and NH participants (CI: 233 ± 107 ms; NH: 299 ± 153 ms; $t_{(17)} = 1.103$, $p = 0.286$).

In the N condition (Fig. 4B), the CI users ($n = 12$), in general, did not show MNR or P3 effects to the deviant stimuli. In order to examine the ERP effects in more detail, the CI users were separated into two performance groups with a threshold of mean response accuracy (66%) in behavioral oddball performance. Five CI users were in the good performer group (83 ± 11%). The remaining seven CI users were included in the poor performer group (53 ± 11%).

The two groups did not significantly differ in peak latencies of N1 (good: 184 ± 36 ms; poor: 196 ± 39 ms; $t_{(10)} = 0.562$, $p = 0.586$) and P2 (good: 315 ± 14 ms; poor: 299 ± 42 ms; $t_{(10)} = 0.932$, $p = 0.379$). Notably, the good CI performers showed a significant P3

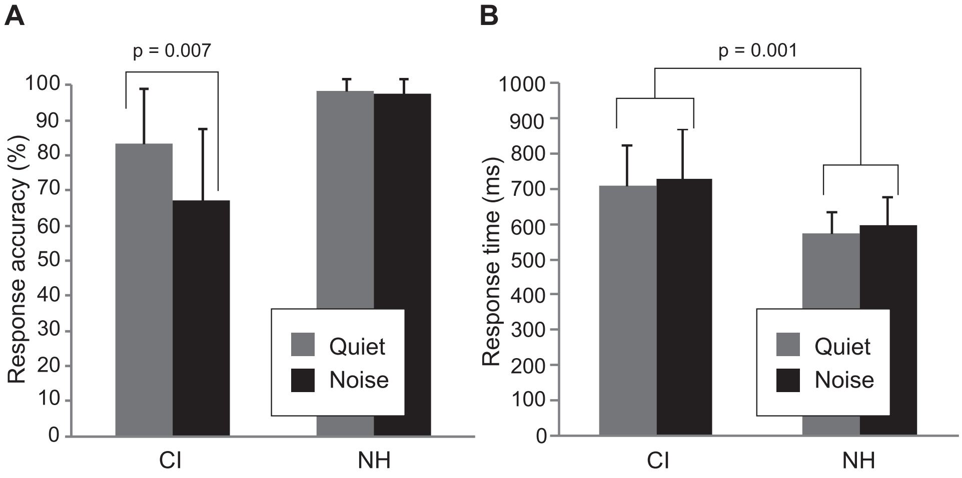

Fig. 2. Syllable detection performances of/ba/and/ga/for the cochlear implant (CI) users and normal hearing (NH) controls. (A): Response accuracy, (B): Response time. Only white noise was used at +10 dB SNR.

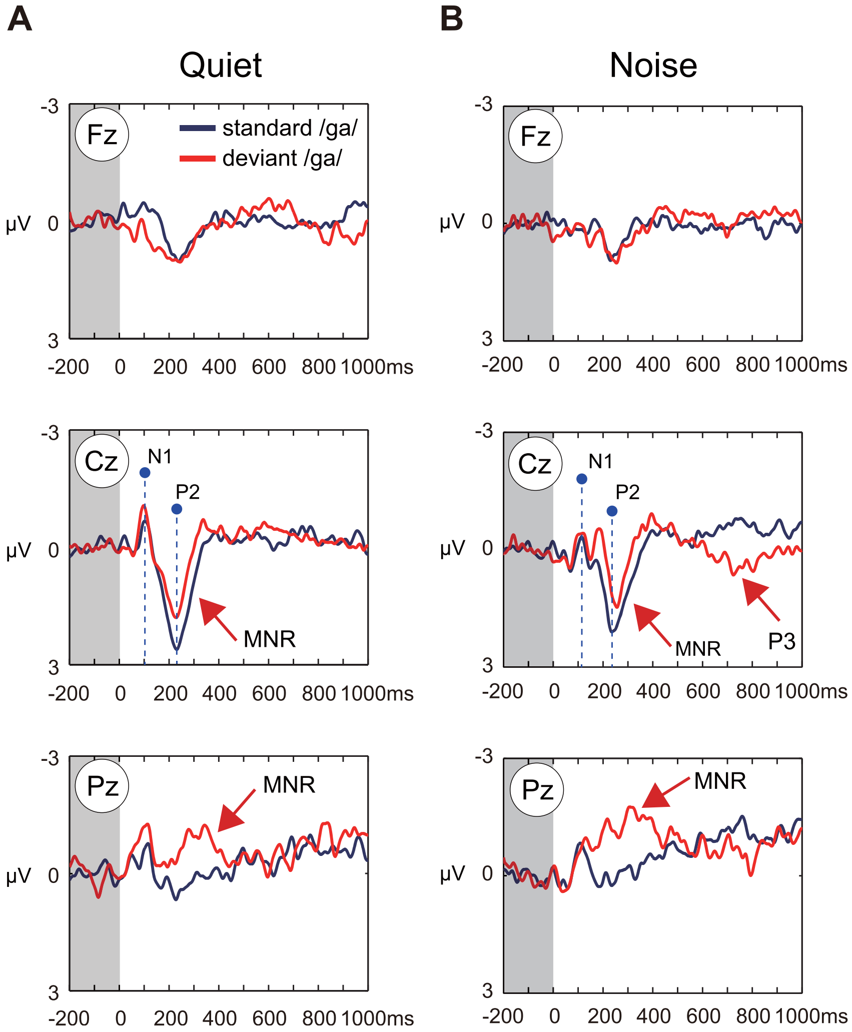

Fig. 3. EEG grand average for the normal hearing controls ($n = 12$) for deviant (red line) and standard (blue line) syllables presented in quiet (A) and noise (B) at the three scalp sites. N1 and P2 peaks (blue dotted lines) were clearly observed at the Cz position for both the quiet and noise conditions. A mismatch negative response (MNR) to the deviant syllable (deviant − standard) was mainly observed in central-posterior sites (red arrows for MNRs in A and B) from about 200 ms post speech-onset for both the quiet and noise conditions. P3 to the deviant syllable in noise was also observed in the central site (red arrow for P3 in B). Gray areas indicate the baseline for waveform comparison. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

effect to the deviant stimuli at the central-posterior site (Cz: 500–700 ms, $t_{(4)} = 2.985$, $p = 0.041$; Pz: 650–900 ms, $t_{(4)} = 7.398$, $p = 0.002$) (Fig. 5A). In fact, the P3 was observed for all of the good CI performers after about 600 ms (Fig. 6). The peak latency of the P3 was similar to that of the NH participants (CI: 762 ± 77 ms; NH: 757 ± 62 ms; $t_{(15)} = 0.143$, $p = 0.888$). The poor CI performers did not show P3 to the deviant stimuli (Cz: 500–700 ms, $t_{(6)} = 1.590$, $p = 0.163$; Pz: 650–900 ms, $t_{(6)} = 1.616$, $p = 0.157$) (Fig. 5B).

Contrary to our prediction, even the good CI performers did not show MNR to the deviant stimuli in noise. This contrasted sharply with the results for the NH controls.

3.6. Neurophysiological correlates of CI-dependent speech perception performance

Finally, we conducted correlation analyses to investigate functional/neural correlations between speech perception performance and neurophysiological properties for the CI users. For the Q condition ($n = 7$), significant correlation relationships were not observed between the response accuracy and neurophysiological properties (N1, P2, MNR).

For the N condition ($n = 12$), N1 and P2 peak latencies did not show significant correlation with response accuracy.

P3 amplitude (deviant − standard) was positively correlated with syllable detection performance in noise in the central (Cz: 500–600 ms, $r = 0.650$, $p = 0.022$), and posterior electrodes (Pz: 650–1000 ms, $r = 0.866$, $p < 0.0001$) (Fig. 7A). The positive correlation also remained significant using the controls of age, CI use duration, and DF duration (Cz: $\rho_{XY \cdot Z} = 0.804$, $p = 0.016$; Pz: $\rho_{XY \cdot Z} = 0.914$, $p < 0.0001$).

Significant positive correlation was also found between P3 amplitude and word perception performance in white noise (AO-

Table 2
Mean peak latency (ms) of auditory evoked potential components (N1, P2) for standard syllables in the quiet and noise conditions for the normal hearing (NH) and cochlear implant (CI) groups.

Group	N1		P2	
	Quiet	Noise	Quiet	Noise
NH	105.1 ± 16	114.1 ± 13	232.4 ± 19	249.5 ± 27
CI	195.7 ± 32	190.8 ± 36	323.6 ± 13	305.3 ± 33

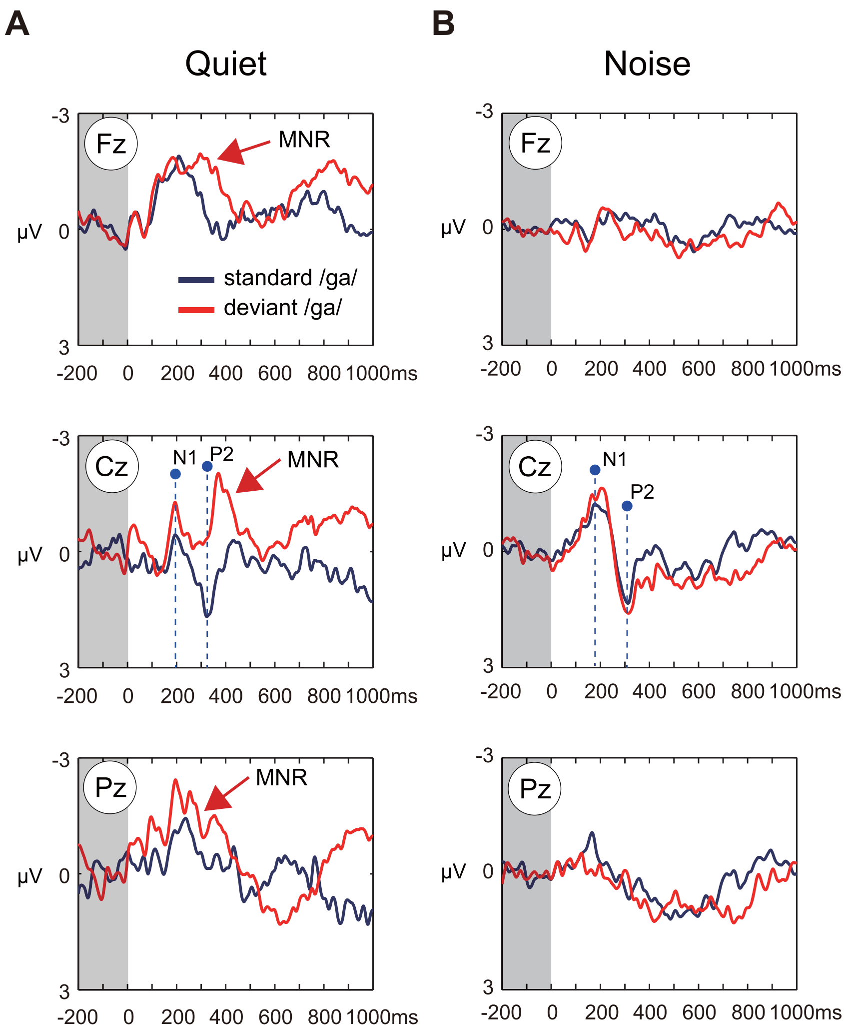

Fig. 4. EEG grand average for the cochlear implant users for deviant (red line) and standard (blue line) syllables presented in quiet (A: $n = 7$) and noise (B: $n = 12$) at the three scalp sites. N1 and P2 peaks (blue dotted lines) were clearly observed at the Cz position for both the quiet and noise conditions. A mismatch negative response (MNR) to the deviant syllable (deviant − standard) was mainly observed in the central site (red arrow in A) from about 200 ms post speech-onset only for the quiet condition. The noise condition did now show either an MNR or a P3 to the deviant syllable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

WN in Experiment 1) (Pz: 350–600 ms, $r = 0.739$, $p = 0.009$; Pz: 850–1000 ms, $r = 0.729$, $p = 0.011$). The correlation again remained significant using control of age, CI use duration, and DF duration (Pz: 350–600 ms, $\rho_{XY \cdot Z} = 0.776$, $p = 0.024$) (Fig. 7B). A similar positive correlation was observed between P3 amplitude and word perception performance in the two-talker noise condition (AO-2T in Experiment 1) (Cz: 250–400 ms, $r = 0.669$, $p = 0.024$; 800–1000 ms: $r = 0.640$, $p = 0.034$), although the correlation did not remain significant using the control of age, CI use duration, and DF duration (250–400 ms, $\rho_{XY \cdot Z} = 0.197$, $p = 0.639$; 800–1000 ms: $\rho_{XY \cdot Z} = 0.410$, $p = 0.313$). To summarize, larger P3 amplitude was associated with better speech perception performance in noise at both the syllable and word levels.

4. Discussion

4.1. Experiment 1

The CI users generally showed degraded word perception in noise at a mild SNR, which did not affect NH participants. These results are consistent with well-known findings that CI users are vulnerable to noisy surroundings (Tyler et al., 1995; Nelson et al., 2003; Nelson and Jin, 2004; Fu and Nogaki, 2005; Davidson et al., 2010). The present CI users had already used a CI device for more than 2 years, however response accuracies in the AO-WN and AO-2T conditions were still about 15% lower than those in the AO-Q condition. This presents a clear contrast to the results of the NH participants who did not show any accuracy decrease in noise at this SNR. Thus, the present results indicate again that CI users are vulnerable to AO word perception in noise even after several years of device use.

Combined with the results of the AV condition, the CI users' vulnerability to noise may be related to the two levels of energetic and information masking. In the AV condition, the masking effect of noise (accuracy in AV-noise versus in AV-Q) was significant in the two-talker but not in the white noise condition. The two-talker noise probably prevents the CI users from detecting the phonological information of the target words, or from focusing on target words because of the same sex voice, even with the use of lip-reading (Brungart et al., 2001; Freyman et al., 2004; Cooke et al., 2008; Mattys et al., 2009). In addition to such information masking, CI users' word perception may also be affected by fluctuations in the two-talker noise, as the two-talker noise reduces the availability of the envelope information of the target words (Nelson and

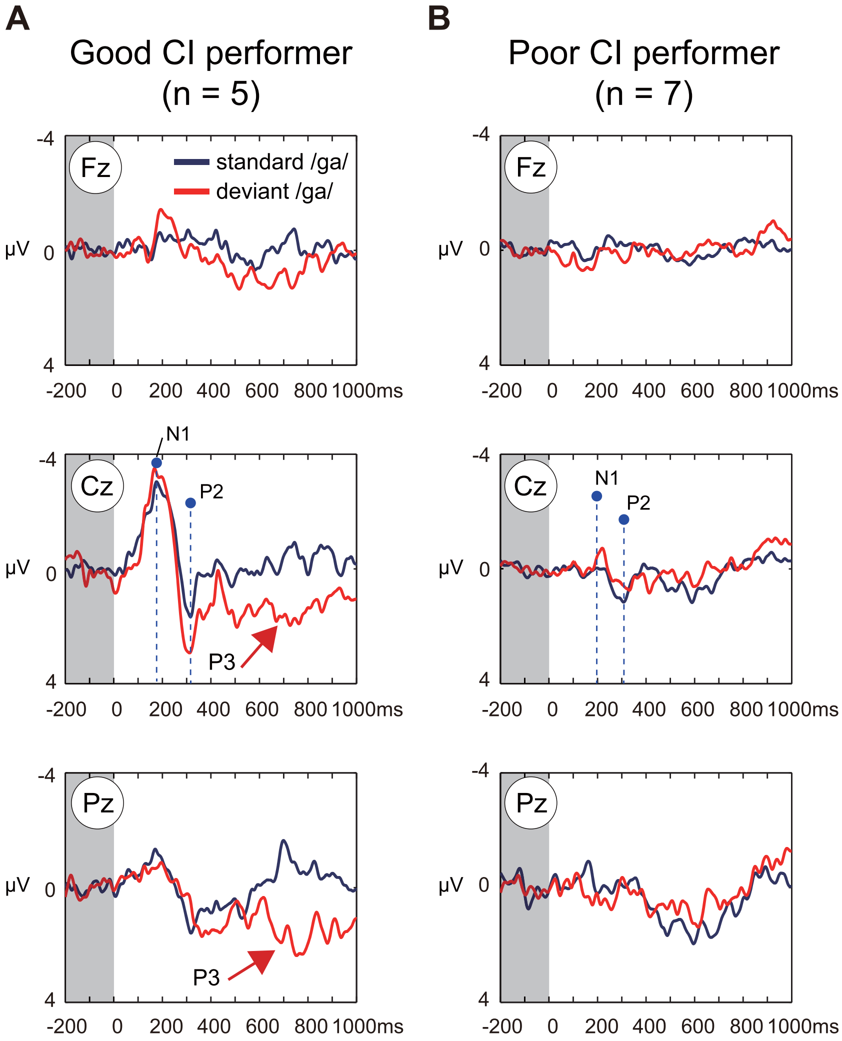

Fig. 5. EEG grand average for the good cochlear implant (CI) performers (A: $n = 5$) and the poor CI performers (B: $n = 7$) for deviant (red line) and standard (blue line) syllables presented in white noise at the three scalp sites. N1 and P2 peaks (blue dotted lines) were observed at the central (Cz) electrode for, in particular, the good CI performers. The N1 and P2 peak latencies were similar between the two CI groups. For the good CI performers, the P3 to the deviant syllable was clearly observed in central-posterior sites (red arrows in A) in later time windows. The poor CI performers did not show a P3 to the deviant syllable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Jin, 2004). On the other hand, the results of the AO word perception did not show the difference in masking between the two noises (accuracy in AO-noise versus AO-Q). This may be due to a kind of floor effect: that is, AO speech perception in noise is generally difficult for CI users (Nelson et al., 2003; Fu and Nogaki, 2005), while slight differences in masking between the two noises in the AO condition might be enhanced in the AV condition (Sumby and Pollack, 1954).

The present results confirmed the previous findings that CI users compensatively use lip-reading to improve speech perception (Kaiser et al., 2003; Rouger et al., 2007, 2008; Desai et al., 2008). The present CI users showed significantly improved word perception performances in all of the quiet and noise conditions in comparisons between the AO and AV conditions. As has been argued in previous studies (Kaiser et al., 2003; Desai et al., 2008), the present CI users used lip-reading despite having more than 2 years of CI use.

4.2. *Experiment 2*

The CI users showed the MNR to the deviant syllable stimuli in quiet. In noise, contrary to our prediction, even the good CI performers did not show the MNR to the deviant stimuli. This was in contrast with the clear MNR in the NH participants in both quiet and noise. On the other hand, the good CI performers showed the P3 to the deviant stimuli in noise similarly to the NH controls. P3 amplitude was significantly correlated with speech perception of syllables and words in noise. That is, the P3 is a marker for CI users' improved auditory speech perception in noise.

4.2.1. *ERP results in the quiet condition*

Both the CI users and the NH participants showed MNRs to deviant stimuli in quiet. The present CI users got about 90% response accuracy for syllable detection in quiet, then eliciting the MNR, as in previous studies (adult/speech: Kraus et al., 1993; Groenen et al., 1996b; children/speech: Singh et al., 2004; adult/tone: Kelly et al., 2005; Roman et al., 2005; Zhang et al., 2011; Lonka et al., 2013).

The MNR peak latency for the CI users was about 100 ms longer than that of the NH controls in quiet, as well as peak latencies for the AEPs. Because a temporal gap between acoustic and device stimulation onsets may be rather narrower than 100 ms (Sandmann et al., 2009), such a delayed latency of the MNR and AEP may be related to a longer RT for behavioral response in the CI users (Ritter et al., 1979). The present study used a difficult speech

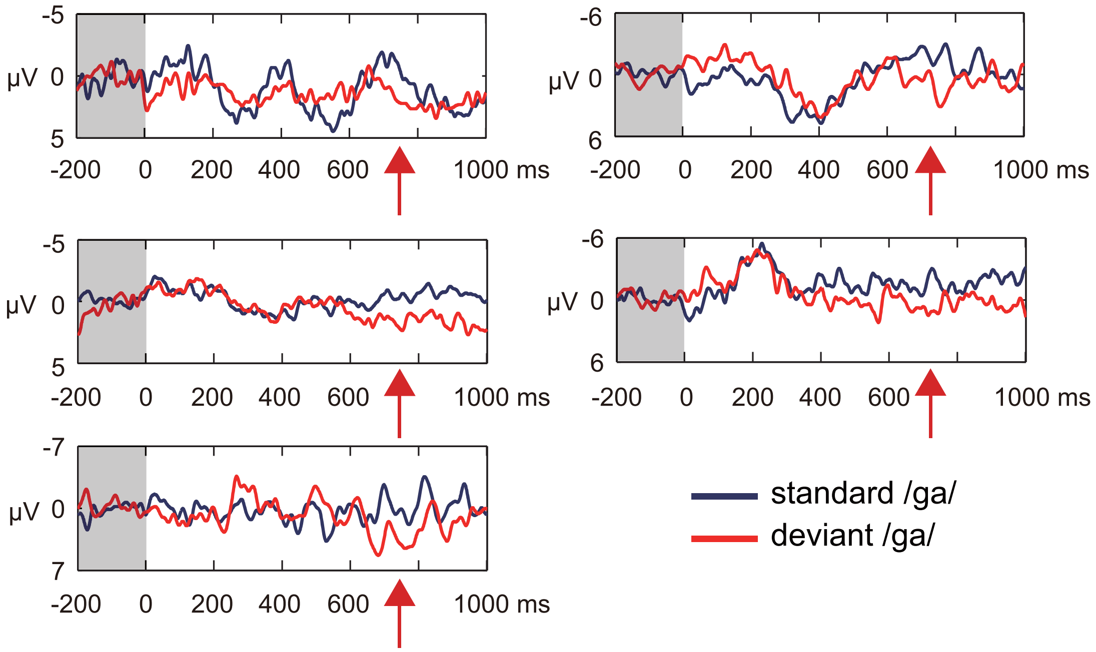

Fig. 6. Individual EEG average at the posterior (Pz) electrode for deviant (red line) and standard (blue line) syllables presented in noise for the good cochlear implant (CI) performers. All of the good CI performers showed P3 to the deviant syllable after about 600 ms post stimulus onset (red arrows). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

contrast with different places of articulation (Groenen et al., 2001; Henkin et al., 2009), likely showing similar results to the studies using a difficult tone contrast (Roman et al., 2005; Obuchi et al., 2012), but not those using an easy contrast (Zhang et al., 2011).

A significant correlation was not observed between speech perception performance and MNR properties, in contrast to previous studies (amplitude: Lonka et al., 2013; latency: Roman et al., 2005; duration: Singh et al., 2004; Kelly et al., 2005). The absence of correlation may be related to the small data range due to the high response accuracy (72–100%). The small number of the CI users may also be related to the low power of the statistical tests in finding significant correlations.

4.2.2. ERP results in the noise condition

The MNR disappeared in the CI users in noise, indicating the difference in automatic speech detection compared to the NH participants. On the other hand, the good CI performers elicited the P3 to deviant speech in noise, while the poor CI users did not. Correlation analyses showed that speech perception performance in noise was positively correlated with P3 amplitude. The good CI performers were not different from the NH controls in P3 latency, as in a previous study in quiet (Kubo et al., 2001). Because the P3 is related to an attentional stimulus evaluation rather than response decision per se (Ritter et al., 1979; Picton, 1992), the better CI users may more attentively evaluate deviant stimuli, similarly to the NH controls, to yield better syllable detection in noise.

The attentional influence to P3 elicitation was also clearly observed in the NH controls. The NH controls showed not only MNR, but also P3 under the same task requirement. This indicates that the NH participants more attentively performed the task in noise, and consciously specified deviant stimuli as a rare target. Wong et al. (2008) examined NH adults' speech perception in noise

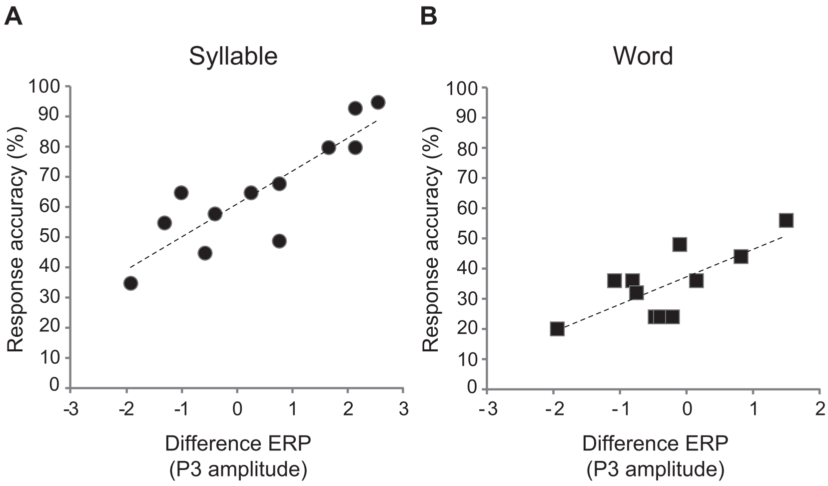

Fig. 7. Linear correlation relationships between CI users' P3 amplitude (deviant − standard) and response accuracy in speech perception of syllables (A) and words (B) in white noise. Larger P3 amplitudes were related to higher response accuracy in CI-dependent speech perception of syllables and words.

with functional magnetic resonance imaging, observing enhanced activation in cortical areas such as the auditory cortex and prefrontal cortex, as well as attentional areas including the anterior cingulate and medial frontal gyri, in a comparison between noise and quiet. This finding suggests that white noise may evoke a more attended attitude in performing the task, thereby in consequence, eliciting the P3 to deviant speech. This attentional effect by noise was also supported by a change in MNR peak latency for the NH controls: the MNR in the noise condition peaked about 50 ms earlier than in the quiet condition (N: 205 ± 29 ms; Q: 263 ± 31 ms). These results suggest that attention related training is effective for improvement of CI users' speech perception in noise (Oba et al., 2011).

Contrary to our prediction, the good CI performers did not elicit MNR in noise. The appearance of MNR probably requires further improved speech perception in noise. Three CI users among the good CI performers showed below 90% response accuracy (68%, 80% and 80%), which was not strictly comparable to that of the NH controls (97.6 ± 4.4%). The question remains for future study, whether or not easier auditory stimuli (more steady contrasts such as tones or vowels) can elicit MNR in CI users. On the other hand, a previous study indicates the possibility that improved speech perception first promotes P3 elicitation. Kelly et al. (2005) reported that in quiet, even poor CI performers could elicit P3 (P3a) to deviant tone, but not MNR. This suggests analogically that along with speech perception improvement in noise, neurophysiological change may begin from restoration of P3 in relation with an attentional stimulus evaluation. Another possible reason may be that P3 has a phase canceling effect to MNR, and as a consequence, attenuates manifestation of MNR (Wunderlich and Cone-Wessen, 2001; Kelly et al., 2005). Future studies are needed to elucidate whether or not good CI performers also elicit MNR after they have further improved speech perception in noise.

5. Conclusions

Experiment 1 replicated the previous finding that CI users can improve speech perception in noise with support by lip-reading. Because ordinary life includes a lot of multi-modal communicative situations, lip-reading benefits may be crucial for CI users' quality of life. On the other hand, the present study revealed that the CI users' auditory-only speech perception is vulnerable to noise even at a mild signal-to-noise ratio at which NH people are not affected at all. There was a tendency for information masking (two-talker noise) to affect CI users' speech perception more than energetic masking (white noise) under a multi-modal condition. The overcoming of information masking in speech perception may be an important issue for CI users to refine overall speech perception.

Experiment 2 found that the good CI performers elicit P3, an indicator of attended cognitive processing, in speech perception in noise. Larger P3 amplitude was related to higher speech perception. P3, therefore, is an objective marker for evaluating CI users' speech perception improvement in noise, and may be a gateway for their further improvement such as in unattended processing reflected by a mismatch negative cortical potential.

Acknowledgments

This study was supported by a Grant-in-Aid for Scientific Research (21243040) to K. Sekiyama from the Japan Society for the Promotion of Science (JSPS). We would like to express our gratitude to Hideki Kawahara (Wakayama University) for use of the TANDEM-STRAIGHT software; Seiko Hayashida (Association of Cochlear Implant Transmitted Audition; ACITA) for recruitment of CI users; Takao Yamada, Toshikazu Kawagoe, Saki Shikita, and Naomi Nakamura for their support in preparation and delivery of the current study. The current affiliation of the first author (T.S) is National Center of Neurology and Psychiatry (NCNP), Kodaira, Tokyo, Japan.

References

Brungart, D.S., Simpson, B.D., Ericson, M.A., Scott, K.R., 2001. Informational and energetic masking effects in the perception of multiple simultaneous talkers. J. Acoust. Soc. Am. 110, 2527–2538.

Carhart, R., Tillman, T.W., Greetis, E.S., 1969. Perceptual masking in multiple sound backgrounds. J. Acoust. Soc. Am. 45, 694–703.

Cooke, M., Garcia Lecumberri, M.L., Barker, J., 2008. The foreign language cocktail party problem: energetic and informational masking effects in non-native speech perception. J. Acoust. Soc. Am. 123, 414–427.

Cowan, N., Winkler, I., Teder, W., Näätänen, R., 1993. Memory prerequisites of the mismatch negativity in the auditory event-related potential (ERP). J. Exp. Psychol. Learn. Mem. Cogn. 19, 909–921.

Davidson, L.S., Geers, A.E., Brenner, C., 2010. Cochlear implant characteristics and speech perception skills of adolescents with long-term device use. Otol. Neurotol. 31, 1310–1314.

Desai, S., Stickney, G., Zeng, F.G., 2008. Auditory-visual speech perception in normal-hearing and cochlear-implant listeners. J. Acoust. Soc. Am. 123, 428–440.

Donchin, E., Ritter, W., McCallum, C., 1978. Cognitive psychophysiology: the endogenous components of the ERP. In: Callaway, E., Tueting, P., Koslow, S. (Eds.), Brain Event-related Potentials in Man. Academic Press, New York, NY, USA, pp. 349–441.

Freyman, R.L., Balakrishnan, U., Helfer, K.S., 2004. Effect of number of masking talkers and auditory priming on informational masking in speech recognition. J. Acoust. Soc. Am. 115, 2246–2256.

Fu, Q.J., Nogaki, G., 2005. Noise susceptibility of cochlear implant users: the role of spectral resolution and smearing. J. Assoc. Res. Otolaryngol. 6, 19–27.

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. Clin. Neurophysiol. 117, 1772–1782.

Groenen, P.A., Beynon, A.J., Snik, A.F., Van Den Broek, P., 2001. Speech-evoked cortical potentials and speech recognition in cochlear implant users. Scand. Audiol. 30, 31–40.

Groenen, P.A., Makhdoum, M., Van Den Brink, J.L., Stollman, M.H., Snik, A.F., Van Den Broek, P., 1996a. The relation between electric auditory brain stem and cognitive responses and speech perception in cochlear implant users. Acta Otolaryngol. Stockh. 116, 785–790.

Groenen, P.A., Snik, A.F., Van Den Broek, P., 1996b. On the clinical relevance of mismatch negativity: results from subjects with normal hearing and cochlear implant users. Audiol. Neurootol. 1, 112–124.

Hamzavi, J., Baumgartner, W.D., Pok, S.M., Franz, P., Gstoettner, W., 2003. Variables affecting speech perception in postlingually deaf adults following cochlear implantation. Acta Otolaryngol. 123, 493–498.

Henkin, Y., Tetin-Schneider, S., Hildesheimer, M., Kishon-Rabin, L., 2009. Cortical neural activity underlying speech perception in postlingual adult cochlear implant recipients. Audiol. Neurootol. 14, 39–53.

Kaga, K., Kodera, K., Hirota, E., Tsuzuku, T., 1991. P300 response to tones and speech sounds after cochlear implant: a case report. Laryngoscope 101, 905–907.

Kaiser, A.R., Kirk, K.I., Lachs, L., Pisoni, D.B., 2003. Talker and lexical effects on audiovisual word recognition by adults with cochlear implants. J. Speech Lang. Hear. Res. 46, 390–404.

Kaplan-Neeman, R., Kishon-Rabin, L., Henkin, Y., Muchnik, C., 2006. Identification of syllables in noise: electrophysiological and behavioral correlates. J. Acoust. Soc. Am. 120, 926–933.

Kawahara, H., Takahashi, T., Morise, M., Banno, H., 2009. Development of exploratory research tools based on TANDEM-STRAIGHT. In: Proc. APSIPA, pp. 111–120.

Kelly, A.S., Purdy, S.C., Thorne, P.R., 2005. Electrophysiological and speech perception measures of auditory processing in experienced adult cochlear implant users. Clin. Neurophysiol. 116, 1235–1246.

Kraus, N., McGee, T., Sharma, A., Carrell, T., Nicol, T., 1992. Mismatch negativity event-related potential elicited by speech stimuli. Ear Hear 13, 158–164.

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. Hear. Res. 65, 118–124.

Kubo, T., Yamamoto, K., Iwaki, T., Matsukawa, M., Doi, K., Tamura, M., 2001. Significance of auditory evoked responses (EABR and P300) in cochlear implant subjects. Acta Otolaryngol. 121, 257–261.

Lonka, E., Relander-Syrjänen, K., Johansson, R., Näätänen, R., Alho, K., Kujala, T., 2013. The mismatch negativity (MMN) brain response to sound frequency changes in adult cochlear implant recipients: a follow-up study. Acta Otolaryngol. 133, 853–857.

Marco-Pallarés, J., Grau, C., Ruffini, G., 2005. Combined ICA-LORETA analysis of mismatch negativity. NeuroImage 25, 471–477.

Martin, B.A., Sigal, A., Kurtzberg, D., Stapells, D.R., 1997. The effects of decreased audibility produced by high-pass noise masking on cortical event-related potentials to speech sounds/ba/and/da/. J. Acoust. Soc. Am. 101, 1585–1599.

Mattys, S.L., Brooks, J., Cooke, M., 2009. Recognizing speech under a processing load: dissociating energetic from informational factors. Cogn. Psychol. 59, 203–243.

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. Am. J. Otol. 16, 514–520.

Miller, G.A., Nicely, P.E., 1955. An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27, 328–352.

Näätänen, R., Gaillard, A.W.K., 1983. The orienting reflex and the N2 deflection of the event-related potential (ERP). In: Gaillard, A.W.K., Ritter, W. (Eds.), Tutorials in Event Related Potential Research: Endogenous Components. North Holland, Amsterdam, pp. 119–141.

Näätänen, R., Gaillard, A.W.K., Mantysalo, S., 1978. Early selective-attention effect on evoked potential reinterpreted. Acta Psychol. (Amst) 42, 313–329.

Näätänen, R., Paavilainen, P., Rinne, T., Alho, K., 2007. The mismatch negativity (MMN) in basic research of central auditory processing: a review. Clin. Neurophysiol. 118, 2544–2590.

Nelson, P.B., Jin, S.H., 2004. Factors affecting speech understanding in gated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 115, 2286–2294.

Nelson, P.B., Jin, S.H., Carney, A.E., Nelson, D.A., 2003. Understanding speech in modulated interference: cochlear implant users and normal-hearing listeners. J. Acoust. Soc. Am. 113, 961–968.

Novak, G., Ritter, W., Vauhhan Jr., H.G., 1992. The chronometry of attention-modulated processing and automatic mismatch detection. Psychophysiology 29, 412–430.

Oba, S.I., Fu, Q.J., Galvin 3rd, J.J., 2011. Digit training in noise can improve cochlear implant users' speech understanding in noise. Ear Hear 32, 573–581.

Obuchi, C., Harashima, T., Shiroma, M., 2012. Auditory evoked potentials under active and passive hearing conditions in adult cochlear implant users. Clin. Exp. Otorhinolaryngol. 5 (Suppl. 1), 6–9.

Oviatt, D.L., Kileny, P.R., 1991. Auditory event-related potentials elicited from cochlear implant recipients and hearing subjects. Am. J. Audiol. 1, 48–55.

Picton, T.W., 1992. The P300 wave of the human event-related potential. J. Clin. Neurophysiol. 9, 456–479.

Ponton, C.W., Don, M., 1995. The mismatch negativity in cochlear implant users. Ear Hear 16, 131–146.

Ritter, W., Simpson, R., Friedman, D., 1979. A brain event related to the making of a sensory discrimination. Science 203, 1358–1361.

Roman, S., Canévet, G., Marquis, P., Triglia, J.M., Liégeois-Chauvel, C., 2005. Relationship between auditory perception skills and mismatch negativity recorded in free field in cochlear implant users. Hear. Res. 201, 10–20.

Rouger, J., Lagleyre, S., Fraysse, B., Deneve, S., Deguine, O., Barone, P., 2007. Evidence that cochlear-implanted deaf patients are better multisensory integrators. Proc. Natl. Acad. Sci. U. S. A. 104, 7295–7300.

Rouger, J., Fraysse, B., Deguine, O., Barone, P., 2008. McGurk effects in cochlear-implanted deaf subjects. Brain Res. 1188, 87–99.

Ruffin, C.V., Tyler, R.S., Witt, S.A., Dunn, C.C., Gantz, B.J., Rubinstein, J.T., 2007. Long-term performance of Clarion 1.0 cochlear implant users. Laryngoscope 117, 1183–1190.

Sandmann, P., Tom Eichele, T., Buechler, M., Debener, S., Jäncke, L., Dillier, N., Hugdahl, K., Meyer, M., 2009. Evaluation of evoked potentials to dyadic tones after cochlear implantation. Brain 132, 1967–1979.

Singh, S., Liasis, A., Rajput, K., Towell, A., Luxon, L., 2004. Event-related potentials in pediatric cochlear implant patients. Ear Hear 25, 598–610.

Squires, N., Squires, K.C., Hillyard, S.A., 1975. Two varieties of long-latency positive waves evoked by unpredictable auditory stimuli in man. Electroencephalogr. Clin. Neurophysiol. 38, 387–401.

Sumby, W.H., Pollack, I., 1954. Visual contribution to speech intelligibility in noise. J. Acoust. Soc. Am. 26, 212–215.

Tyler, R.S., Lowder, M.W., Parkinson, A.J., Woodworth, G.G., Gantz, B.J., 1995. Performance of adult ineraid and nucleus cochlear implant patients after 3.5 years of use. Audiology 34, 135–144.

Tyler, R.S., Noble, W., Dunn, C., Witt, S., 2006. Some benefits and limitations of binaural cochlear implants and our ability to measure them. Int. J. Audiol. 45 (Suppl. 1), 113–119.

Whiting, K.A., Martin, B.A., Stapells, D.R., 1998. The effects of broadband noise masking on cortical event-related potentials to speech sounds/ba/and/da/. Ear Hear 19, 218–231.

Wong, P.C., Uppunda, A.K., Parrish, T.B., Dhar, S., 2008. Cortical mechanisms of speech perception in noise. J. Speech Lang. Hear. Res. 51, 1026–1041.

Wunderlich, J.L., Cone-Wesson, B.K., 2001. Effects of stimulus frequency and complexity on the mismatch negativity and other components of the cortical auditory-evoked potential. J. Acoust. Soc. Am. 109, 1526–1537.

Zeng, F.G., 2004. Trends in cochlear implants. Trends Amplif. 8, 1–34.

Zhang, F., Hammer, T., Banks, H.L., Benson, C., Xiang, J., Fu, Q.J., 2011. Mismatch negativity and adaptation measures of the late auditory evoked potential in cochlear implant users. Hear. Res. 275, 17–29.