Objective Identification of Simulated Cochlear Implant Settings in Normal-Hearing Listeners Via Auditory Cortical Evoked Potentials

Sungmin Lee,[1] and Gavin M. Bidelman[1,2,3]

Objectives: Providing cochlear implant (CI) patients the optimal signal processing settings during mapping sessions is critical for facilitating their speech perception. Here, we aimed to evaluate whether auditory cortical event-related potentials (ERPs) could be used to objectively determine optimal CI parameters.

Design: While recording neuroelectric potentials, we presented a set of acoustically vocoded consonants (aKa, aSHa, and aNa) to normal-hearing listeners (n = 12) that simulated speech tokens processed through four different combinations of CI stimulation rate and number of spectral maxima. Parameter settings were selected to feature relatively fast/slow stimulation rates and high/low number of maxima; 1800 pps/20 maxima, 1800/8, 500/20 and 500/8.

Results: Speech identification and reaction times did not differ with changes in either the number of maxima or stimulation rate indicating ceiling behavioral performance. Similarly, we found that conventional univariate analysis (analysis of variance) of N1 and P2 amplitude/latency failed to reveal strong modulations across CI-processed speech conditions. In contrast, multivariate discriminant analysis based on a combination of neural measures was used to create "neural confusion matrices" and identified a unique parameter set (1800/8) that maximally differentiated speech tokens at the neural level. This finding was corroborated by information transfer analysis which confirmed these settings optimally transmitted information in listeners' neural and perceptual responses.

Conclusions: Translated to actual implant patients, our findings suggest that scalp-recorded ERPs might be useful in determining optimal signal processing settings from among a closed set of parameter options and aid in the objective fitting of CI devices.

Key words: Auditory evoked potentials, Cochlear implants, EEG classification, Information transfer, Speech confusion matrix, Vocoded speech.

INTRODUCTION

For the past 30 years, cochlear implants (CIs) have served as an effective means to provide access to sound for people with severe to profound hearing loss. However, music appreciation, fine speech perception, and noise-degraded listening still remain unresolved (Cullington & Zeng 2011). In terms of signal processing, a CI device decodes input acoustic signals and converts them into biphasic electrical pulses to directly stimulate auditory nerve fibers. This neuronal stimulation can be changed by manipulating the signal processing strategies (i.e., parameter settings) of the CI processor to yield measurable changes in speech perception. Various CI mapping parameters are available to the clinical audiologist for adjustment during device fitting. Thus, it is important to understand which parameters and how specific combinations of device settings affect auditory perception and address the aforementioned issues experienced by CI patients. Current approaches to CI fitting rely on the behavioral (subjective) evaluation of the patient. The present study aimed to demonstrate initial efficacy of a new, more objective approach to CI parameter selection based on listeners' neurophysiological responses.

Several studies have attempted to identify optimal parameter settings that maximize speech perception outcomes in CI patients (Fishman et al. 1997; Loizou et al. 2000; Kiefer et al. 2001; Friesen 2005; Nie et al. 2006; Arora et al. 2009; Buechner et al. 2009; Shannon et al. 2011). Among contemporary CI speech processing strategies, the n-of-m strategy (e.g., Spectral Peak [SPEAK], Advanced Combination Encoder [ACE]) has proven to be highly effective (Skinner et al. 2002). ACE is currently used as a default in Cochlear Corporation's Nucleus system. In the n-of-m strategy, the device activates only a designated number of channels which have the highest spectral energy in each cycle of stimulation. This functional reduction in channels is beneficial to achieve faster stimulation rates (SRs), a reduction in channel interaction, and an increase in battery life. Maxima is a term used in the n-of-m speech coding strategy that indicates the number of activated channels (n) among those available (m) along an electrode array. Hence, the number of channels involved (maxima) determines the amount of spectral information for decoding the input speech signal. Studies have found that while the optimal number of maxima varies across individuals, at least eight maxima are typically required for robust speech perception (Dorman et al. 2002; Plant et al. 2002).

Another relevant parameter to consider in CI processing is SR. CIs deliver acoustic information with a train of electrical biphasic pulses through multiple electrodes. SR determines the stimulation rate (pps) that temporal information is transmitted to individual electrode contacts. SR limits how precisely temporal acoustic information like envelope cues are transferred to the auditory nervous system and are ultimately available to perception. Theoretically, rapid SRs might be advantageous as they (1) simulate the normal stochastic firing of the auditory nerve, (2) improve temporal sampling and fine structure coding, and (3) expand the dynamic range of both frequency and intensity, providing access to higher pitch percepts and lower thresholds (Shannon et al. 2011). However, some studies have shown that higher SRs may not always improve speech recognition (Vandali et al. 2000; Friesen et al. 2005; Arora et al. 2009; Shannon et al. 2011). Equivocal effects of SR may stem from variations in signal processing

[1]School of Communication Sciences & Disorders, University of Memphis, Memphis, Tennessee, USA; [2]Institute for Intelligent Systems, University of Memphis, Memphis, Tennessee, USA; and [3]University of Tennessee Health Sciences Center, Department of Anatomy and Neurobiology, Memphis, Tennessee, USA.

strategies, electrode interactions, or a negative influence of additional temporal information at higher rates (Nie et al. 2006; Shannon et al. 2011). What is clear, however, is that more research is needed to understand the conditions under which SR produces improvements in speech perception. Moreover, it would be useful to develop objective techniques to determine such parameters and map implants using a "hands off" approach.

Event-related brain potentials (ERPs) are a well-established objective measure of cortical auditory processing. ERPs have been used to assess the neurophysiological detection and discrimination of sound in normal and hearing-impaired listeners (Tremblay et al. 2003; Alain et al. 2014; Bidelman et al. 2014a). The cortical ERPs consist of several prominent deflections (i.e., "waves") that develop over the first few hundred milliseconds after sound initiation. In particular, the P1–N1–P2 complex is composed of three major waves occurring at latencies of ~50 msec (P1), ~100 msec (N1), and ~180 to 200 msec (P2) after stimulus onset. They are presumed to originate from thalamus, primary, and secondary auditory cortex (Näätänen & Picton 1987; Picton et al. 1999) and thus, provide a neural index of early auditory coding in cerebral structures. To date, the utility of the obligatory P1–N1–P2 components for deployment in clinical applications has not been firmly established. Intersubject variability and normal maturational changes notwithstanding (Novak et al. 1989; Luck 2005), the cortical ERPs are remarkably stable and show clinically-acceptable ($r > 0.8$; Cicchetti 1994) test–retest repeatability across multiple sessions (Sandman & Patterson). This suggests that ERPs could be used in the objective assessment of auditory function and optimization of hearing assistive devices (Billings et al. 2011b).

Several studies have explored use of the P1–N1–P2 in assessing auditory physiological function in CI patients. For example, P1 has been used as an index of the maturational changes in auditory cortex following implantation (Sharma et al. 2002, 2005; Sharma & Campbell 2011). Following the technical challenge to minimize electromagnetic artifacts from CIs, (Gilley et al. 2006; Martin 2007), studies have been able to assess how implant parameters influence the neural encoding of complex stimuli (Firszt et al. 2002; Kim et al. 2009), the recovery function of neural responses (Zhang et al. 2009, 2010), and the neural correlates of speech perception in CI patients (Groenen et al. 2001; Gordon et al. 2005; Kelly et al. 2005; Samira et al. 2010). For example, Kelly et al. (2005) measured N1 and P2 responses from 12 CI users by presenting three types of tonal stimuli (0.25, 1, and 4 kHz pure tones). They reported that N1 amplitudes were smaller in the CI group than a control group. They further indicated that earlier P2 latencies were associated with shorter duration of deafness and greater speech perception performance. Similarly, Groenen et al. (2001) described characteristics of the N1–P2 response in 9 post lingual CI patients using an extensive range of acoustic stimuli including tones, consonants, and vowel pairs. Their findings showed that N1 and P2 latencies were prolonged in the CI group compared with control listeners. In addition, amplitudes of the N1 and P2 waves in response to consonant tokens were weaker in CI users compared with normal-hearing controls. Collectively, these studies indicate that the cortical ERPs can be used to objectively assess aspects of auditory function in CI patients and, more critically, are predictive of behavioral outcomes.

To date, few studies have explored the potential of using cortical ERPs (P1–N1–P2) to optimize the selection of CI device settings. One study by Friesen et al. (2009) found a potential use of ERPs for CI mapping. Using consonant–vowel–consonant syllables processed with a CI simulation (acoustic) vocoder, they observed a decrease in peak latencies and an increase in amplitudes with increasing number of channels from 2 to 16, indicative of improved speech encoding. Systematic changes in the ERPs with channel number implies that the characteristics of CI processing are represented in neural responses from auditory cortex. However, changes in P1–N1–P2 responses did not correlate with behavioral identification scores. Problematically, this suggests that while evoked potentials might be used to evaluate the neural encoding of CI speech, isolated waves may not provide a direct predictor of behavioral outcomes. Finding new ways to leverage the ERPs in selecting processor settings is needed to provide an objective alternative to CI fittings. Such objective fitting techniques are particularly warranted in cases where patients cannot provide a confirmatory behavioral response (e.g., young infants, difficult-to-test patients) to validate the settings chosen by the clinician.

In the present study, we extend previous ERP studies (Friesen et al. 2009) to examine the effects of parametrically varying CI processor settings on the neural encoding of speech. Our specific aim was to determine whether or not listeners' cortical ERPs could be used to identify an optimal combination of two important CI settings: SR and number of spectral maxima. To avoid complications produced by the electromagnetic artifact of CIs and demonstrate initial efficacy of the proposed technique, we recorded ERPs in response to acoustic simulations (vocoded speech) of various vowel–consonant–vowel (VCV) speech tokens in normal-hearing individuals. Simulated CI speech enabled us to more precisely control stimulation schemes while minimizing confounding factors inherent in actual CI patients (e.g., duration of device use, electrode insertion depth, etc.; Dorman et al. 1997; Friesen et al. 2009; Casserly 2015). Listeners performed a speech identification task during electrophysiological recording to assess their behavioral performance for speech processed with different combinations of SR and maxima settings.

We hypothesized that CI settings (SR/maxima combinations) yielding better behavioral speech identification would elicit cortical responses with more robust amplitudes—consistent with prior studies (e.g., Friesen et al. 2009). In addition to typical univariate peak analyses, we adopted multivariate classification techniques that have been previously employed in recent ERP studies (Chang et al. 2010; Bidelman et al. 2013). The notion of this latter approach is that sounds that are perceptually identifiable are represented by unique patterns of neural activity; more divergent neural representations between tokens reflect different speech "templates" and thus support more robust identification (cf. Bidelman et al. 2013; Bidelman & Lee 2015). For this latter approach, we conducted a discriminant analysis (DA) to classify different neural responses to CI-processed speech into mutually exclusive groups. We then constructed "neural confusion matrices" based on the accuracy we could predict the speech stimulus input from listeners' ERP responses. Neural confusion matrices were then compared with perceptual confusion matrices to evaluate which CI settings yielded superior behavioral identification of the stimulus set. Results show that multivariate analyses of the ERPs can be used to objectively determine CI processing settings from among a closed set of parameter combinations.

MATERIALS AND METHODS

Subjects

Twelve adults (4 males, 8 females) with self-reported normal hearing and no history of neurological or cognitive deficits participated in the study. They were speakers of American English aged from 22 to 50 years old (M: 28.7, SD: 7.4 years). Hearing was confirmed normal (<20 dB HL) in participants over 30 years of age via formal audiometric evaluation. Participants completed a written consent form in compliance with a protocol approved by the Institutional Review Board at the University of Memphis. All were compensated monetarily for their time.

Cochlear Implant Simulated Speech

Twelve CI-simulated speech stimuli were created from three American English VCVs /aKa/, /aNa/, and /aSHa/ (Fig. 1A). The three original VCVs were taken from the consonant set created by Shannon et al. (1999). Speech tokens were produced by a female talker to minimize possible ceiling effects (Loizou et al. 1998), and preprocessed to normalize duration (400 msec) using the change tempo function in Audacity (v. 2.1.2). This function enables changing the tempo without changing the pitch for sounds. Sample rate (48,828 Hz), bit depth (16 bits), and RMS intensity normalization were then completed using Praat (Boersma & Weenink 2013). The application "Cochlear Implant Simulation (v2.0) (CIS)" (de la Torre Vega et al. 2004) was then used to create four CI-vocoded simulations for each VCV that varied in two parameters: SR and the number of maxima (described below).

As this study is motivated in part by clinical considerations, an attempt was made to create stimuli in CIS that were similar to those processed by the default ACE strategy in the Nucleus CI system: Input frequency range, defined as fMin and fMax, was set from 150 Hz to 8 kHz. The spectral ranges of input sounds were separated into band-pass filters whose bandwidths were dependent on the filter's center frequency; filters with lower center frequencies had narrower bandwidths than those with higher center frequencies. The number of filters was determined by the designated value 22 of the parameter "n-inserted-ci," which represents the number of CI electrodes inserted into the cochlear partition. The bandwidths of the filters were designed to be allocated along a logarithmic scale, having narrower bandwidths at lower center frequencies and broader bandwidths at higher center frequencies.

With 22 designated channels in the defined frequency range (0.15 to 8 kHz), the center frequency and bandwidth for the lowest frequency channel were 171 and 42 Hz, and the highest channel were 7668 and 672 Hz, respectively. In this study, the filter bank consisted of infinite impulse response filters followed by envelope detection based on rectification and low-pass filtering (Rect-LP + infinite impulse response). Note that this differs from the Fast Fourier Transform employed in the commercial ACE CI strategy. However, previous studies indicate there is no significant difference in speech recognition between these two types of filter configurations in certain conditions (Ghrissi & Cherif 2012). Under the assumption of using a slim, straight electrode array, the CI length parameter in the CIS program was set to 20 mm and the number of inserted electrodes was set to 22. Because the objective of this study was to focus on the effect of the number of maxima and SR, other irrelevant parameters such as channel interaction and synchronization were not manipulated.

Each of the three VCVs was acoustically processed using four different parameter combinations simulating various SRs and the number of maxima of a CI signal processing strategy. These parameter settings were selected to feature relatively fast/slow SRs and high/low number of maxima; 1800/20, 1800/8, 500/20, and 500/8. They are clinically applicable parameter settings used routinely in CI mapping sessions. Most importantly, these settings and VCVs showed the greatest differences in identification performance from our extensive pilot testing of 20 consonant alternatives, thus maximizing the possibility of observing differences in the ERPs. Likewise, we found that perceptual differences between the three VCVs was greatest between 500/20 and 1800/8 than any other pairs among 12 parameter configurations. In this regard, we adopted these tokens in the present study under the assumption that they would allow us to measure more explicit neural responses in relation to behavioral performance. The rate/maxima parameter combinations used in this study are shown in Table 1.

Stimulus Presentation and Behavioral Task

Recording of ERPs and speech identification tasks were conducted simultaneously. The tasks were implemented in a double-walled electroacoustically shielded sound-attenuating chamber (Industrial Acoustics Company, Inc., North Aurora, IL). Behavioral speech identification was administered in a three alternative forced choice (3AFC) procedure having alternatives of /
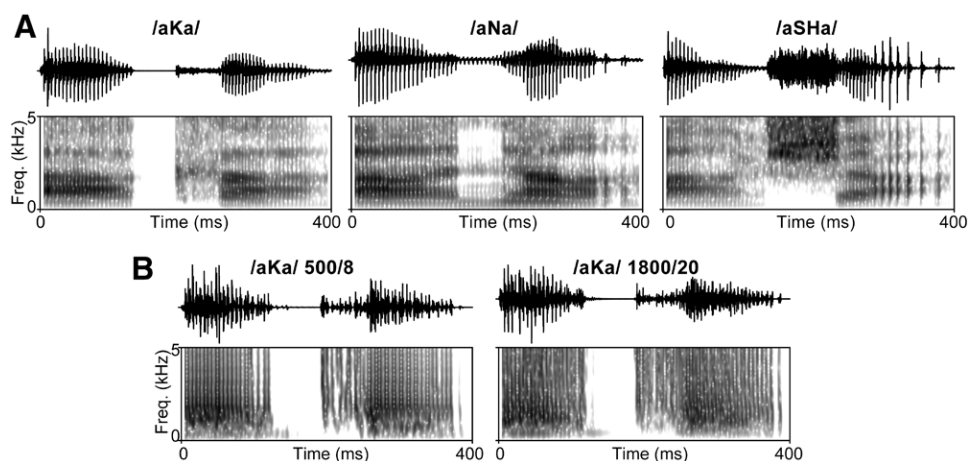

Fig. 1. Stimulus time waveforms and spectrograms. A, Unprocessed speech tokens /aKa/, /aNa/, and /aSHa/. B, Example of CI vocoding of the token /aKa/ with different rate/maxima combinations.

TABLE 1. Cochlear implant simulated stimulus set consisting of three vocoded speech sounds varying in high/low stimulation rates and high/low number of spectral maxima

	/aKa/	/aNa/	/aSHa/
Stimulation rate (pps)/# maxima	1800/20	1800/20	1800/20
	1800/8	1800/8	1800/8
	500/20	500/20	500/20
	500/8	500/8	500/8

aKa/, /aNa/, and /aSHa/. Participants were seated in the middle of the sound booth with access to a monitor and a keyboard, and asked to press the key that matched the stimulus they perceived. They were encouraged to respond as accurately and quickly as possible. For familiarization, a short practice session preceded the main task using unprocessed (non-vocoded) stimuli. No feedback was provided during actual experimental trials.

The 12 CI-simulated speech stimuli were presented in a random sequence controlled by MATLAB 2013 (The Math-Works, Inc., Natick, MA) routed through a TDT RP2 digital interface (Tucker Davis Technologies, Alachua, FL). Each individual token was presented 150 times, resulting in a total of 1800 (=12 tokens × 150 repetitions) speech stimuli for each subject. The task was divided into 2 to 3 blocks to provide short periodic breaks and avoid fatigue. Stimuli were delivered binaurally via insert earphones (ER-2, Etymotic Research) using fixed, rarefaction polarity at an intensity of 81 dB SPL. Following participants' behavioral response, the proceeding stimulus commenced with a jittered interstimulus interval between 800 and 1000 msec (uniform distribution). The entire experiment lasted ~1 hr per listener.

Evoked Potential Recording

During the speech identification task, neuroelectric responses were recorded differentially between electrodes placed on the high fore-head (~Fpz) referenced to linked mastoids (A1/A2) (Bidelman et al. 2013; Bidelman 2015; Krishnan et al., 2010). We used this single channel montage given its applicability to clinical use over multichannel configurations (Picton et al. 1977) and the frontocentral distribution of the auditory cortical ERPs (Picton et al. 1999). Another electrode placed on the mid-forehead served as the common ground. Impedances of all the electrode contacts were maintained below 5 kΩ. In accordance with the number of repetitions in the identification task, individual subjects' ERPs were composed of averages of 150 epochs for each parameter set. A total of 1800 (=12 subjects × 150 repetitions) neural recordings were obtained for each of the 12 parameter conditions. Continuous EEGs were digitized at 5 kHz (Curry 7; SynAmps RT amplifiers; Compumedics Neuroscan) using an online passband of 0.05 to 2000 Hz. EEGs were then epoched 100 msec before to 500 msec after the stimulus onset (i.e., −100 to 500 msec window), baseline corrected to the prestimulus interval, and averaged in the time domain to obtain ERPs for each stimulus condition and participant. Epochs in which neural activity exceeded ±50 to 75 μV (e.g., blinks) were rejected as artifacts before averaging (artifact thresholds were determined for individual subjects). Grand average ERPs were then further band pass filtered off-line between 3 to 30 Hz before peak measurement to isolate cortical activity from the EEG and quantify evoked responses (Bidelman et al. 2013; Bidelman 2015; cf. Fig. 3).

Behavioral Data Analysis (Accuracy and RTs)

Listeners' behavioral identification responses were recorded in MATLAB. Accuracy (%-identification) and reaction times (RTs) were computed for each stimulus and VCV class. Accuracy was determined as the average percent of speech items correctly identified as the appropriate consonant class (/aKa/, /aNa/, and /aSHa/). RTs were computed as the median time lapse between the onset of stimulus presentation and the listener's behavioral response (i.e., button press).

Evoked Potential Analysis

P1–N1–P2 Peak Quantification • We quantified prominent deflections (i.e., "waves") of each listener's cortical ERPs using well-established procedures from our laboratory (Bidelman et al. 2013; Bidelman & Alain 2015). The auditory ERPs typically consist of a series of obligatory wave (P1–N1–P2) occurring over the first ~150 to 200 msec after the time-locking stimulus. However, initial visual inspection of the data revealed that the P1 was weak and highly variable across listeners, consistent with previous reports examining the neural correlates of other speech degradations (Parbery-Clark et al. 2011; Bidelman & Howell 2016). To this end, we measured amplitude and latencies for the N1 and P2 waves only. The N1–P2 complex was of particular interest in the present study as it is highly sensitive to speech perception tasks (Tremblay et al. 2001; Alain et al. 2007; Bidelman et al. 2013, 2014a) and prone to the neuroplastic effects of speech sound training (Tremblay et al. 2001; Alain et al. 2007). N1 and P2 peaks were automatically determined by Curry 7 (Compumedics Neuroscan) as the minimum/maximum peaks within a search window of 75 to 220 msec. Automatic peak selections were then manually confirmed by the first author and corrected in instances of spurious selection (e.g., bifid peak) based on a comparison with the grand average trace (Fig. 3).

Discriminant Analysis and Neural Confusion Matrices • In addition to conventional peak analysis, we aimed to determine which combination of CI parameters (SR and number of maxima) produced the most distinct neural responses. We reasoned that parameters which produce maximally different cortical ERPs represent optimal coding strategies and would allow robust perceptual identification of the three VCV stimuli (cf. Chang et al. 2010; Bidelman et al. 2013). To this end, a quadratic DA (QDA) was conducted to determine the best weighted combination of the four neural predictor variables (N1 and P2 amplitudes and latencies) that best discriminate between the three VCV classes (i.e., /aKa/, /aNa/, and /aSHa/). The QDA was developed at the group level with a set of quadratic equations that was designed to classify listeners' neural responses into one of several mutually exclusive groups on the basis of the four predictor variables extracted from their ERPs (peak and latencies of the N1 and P2 waves). In the QDA, each listener's dependent measures (i.e., N1 and P2 amplitudes/latencies) were treated as an independent set of observations. QDA was implemented in MATLAB using the "fitdiscr" function with prior probabilities set to $p = 0.33$ (i.e., chance level for three stimulus classes).

Classification performance of the discriminant functions were determined by comparing the predicted VCV from each ERP against the ground truth stimulus used to actually evoke the response. This resulted in both a classification accuracy and error rate that were then used to construct a "neural confusion matrix." The confusion matrices quantitatively indicate the degree (%-correct) to which the combined neural measures could correctly classify the input vowel

stimulus (/aKa/, /aNa/, /aSHa/). We further assessed generalizability of the QDA predictions using *k*-fold crossvalidation. *K*-fold crossvalidation is a commonly used machine-learning technique for evaluating the performance of a classifier and generalization to an independent dataset. In this technique, the dataset is partitioned into *k* subsamples (folds). In this study, we set $k = N$, where *N* is number of observations (subjects), which is equivalent to a leave-one-out resampling. For each *k* sample, was the QDA function was recomputed from the remaining participants. This was repeated until every one of the 12 participants was used for validation. Results were then averaged across the *k*-folds to arrive at a cross-validated estimate of accuracy. Although cross-validated performance is typically lower than that obtained on the training set, it helps prevent overfitting. Neural confusion matrices were constructed for each of the four CI parameter combinations (Table 1). The rate/maxima settings producing the largest (cross-validated) classification accuracy were taken as the optimal combination of CI settings that maximize perceptual identification.

Information Transfer Analysis • We used information transfer (IT) (Miller & Nicely 1955) to more directly assess the correspondence between stimulus input and behavioral/neural output. IT is defined as the ratio of transmitted information between *x* and *y* [i.e., $T(x;y)$] to the input entropy (H_x), expressed as a percent. $T(x;y)$ represents the transmission of information (in an information theoretic sense) from *x* to *y*, measured in bits per stimulus, and was computed from confusion matrices via Equation 1:

$$T(x;y) = -\sum_{ij} p_{ij} \log_2 \frac{p_i p_j}{p_{ij}}, \quad (1)$$

where p_i and p_j are the probabilities of the observed input and output variables, respectively, and p_{ij} is the joint probability of occurrence for observing input *i* with output *j*. These probabilities were computed from the confusion matrices as $p_i = n_i/N$, $p_j = n_j/N$, and $p_{ij} = n_{ij}/N$, where n_i is the frequency of stimulus *i*, n_j is the frequency of response *j*, and n_{ij} is the frequency of the joint occurrence of stimulus *i* and response *j* (i.e., diagonal elements of the confusion matrix) in the sample of *N* total observations (Miller & Nicely 1955). The input entropy H_x is given by Equation 2:

$$H_x = -\sum_x p_x \log_2 p_x \quad (2)$$

In the present study, all stimuli occurred with equal probability (i.e., $p_x = 0.33$). IT was then computed as (Equation 3)

$$IT = \frac{T(x;y)}{H_x} \quad (3)$$

This metric varies from 0 to 1. Intuitively, if the transmission is poor and a listener's response does not closely correlate with the stimulus, then IT will approach zero; alternatively, if the response can be accurately predicted from the stimulus then IT will approach 1 (i.e., 100% IT). IT was computed from both the behavioral and neural confusion matrices. Comparisons of IT across stimulus conditions allowed us to assess which combination of CI settings optimally transmitted information in listeners' neural and perceptual responses.

Statistical Analysis

Unless otherwise noted, repeated measures (rm) analysis of variance (ANOVAs) were conducted separately on each dependent measure. Three within-subject factors (VCV, rate, and maxima) were used to determine significant behavioral (RT and percent correct) and neural (latency and amplitude of N1/P2) response variations as a function of combinations of SRs and number of maxima using IBM SPSS (v.23). Following omnibus analyses, post hoc multiple comparisons were employed using Bonferroni corrections to control Type I error inflation. An a priori alpha level was set at $\alpha = 0.05$ for all statistical tests.

Brain–Behavior Relations • We also aimed to assess the relation between brain and behavioral measures. To this end, a generalized linear mixed effects model (GLME) (logistics regression) was used to evaluate the relation between ERP response measures and behavioral speech identification accuracy. In this analysis, we considered the combination of all ERP amplitude and latency measures, as well as their interactions as predictors of speech identification scores; in addition to this fixed effect, subjects were modeled as a random factor by allowing for a random intercept per subject [i.e., correct ~ n1a mp*p2amp*n1lat*p2lat+(1|subject)]. Logistics regression was achieved using the "fitglme" function in MATLAB. The behavioral responses we aimed to measure were a binary response (i.e., correct/incorrect). Consequently, we used a binomial distribution and canonical logit link function to model the relation between neural and perceptual responses (Jaeger 2008)*.

RESULTS

Behavioral Results

Despite the degraded clarity of speech tokens, listeners showed high perceptual performance on the speech identification task. Group mean average speech identification accuracy (percent correct scores) for all 12 conditions were greater than 97% (Fig. 2A; see also Fig. S2A of Supplemental Digital Content, which shows boxplots and individual subject data, http://links.lww.com/EANDH/A324). Overall speech identification did not differ with changes in either SR [$F(1, 11) = 0.688$, $p = 0.425$] or maxima [$F(1, 11) = 0.104$, $p = 0.753$]. However, we found a significant main effect of VCV on identification scores [$F(2, 22) = 3.73$, $p = 0.04$]. Bonferroni pairwise comparisons revealed that /aNa/ was significantly higher than /aKa/ (but not /aSHa/) in terms of speech identification scores. We did not find interaction effects of VCV × Rate [$F(2, 22) = 2.345$, $p = 0.119$] or VCV × Maxima [$F(2, 22) = 1.911$, $p = 0.172$]. However, we found a significant interaction effect of Rate × Maxima [$F(1, 11) = 5.427$, $p = 0.04$]. Follow-up contrasts revealed that when

---

*The analysis of near-ceiling effects would be problematic using traditional Gaussian-based correlations. Even the RAU-transformed data (Studebaker 1985) does not mitigate proportions close to floor or ceiling. Consequently, we chose to analyze brain–behavior relationships using a logistics GLME with a logit link function that modeled the categorical nature of the response (0 = incorrect vs. 1 = incorrect) via a binomial distribution. This greatly improves validity and is the preferred method for data approaching 0 or 100% (Jaeger 2008).
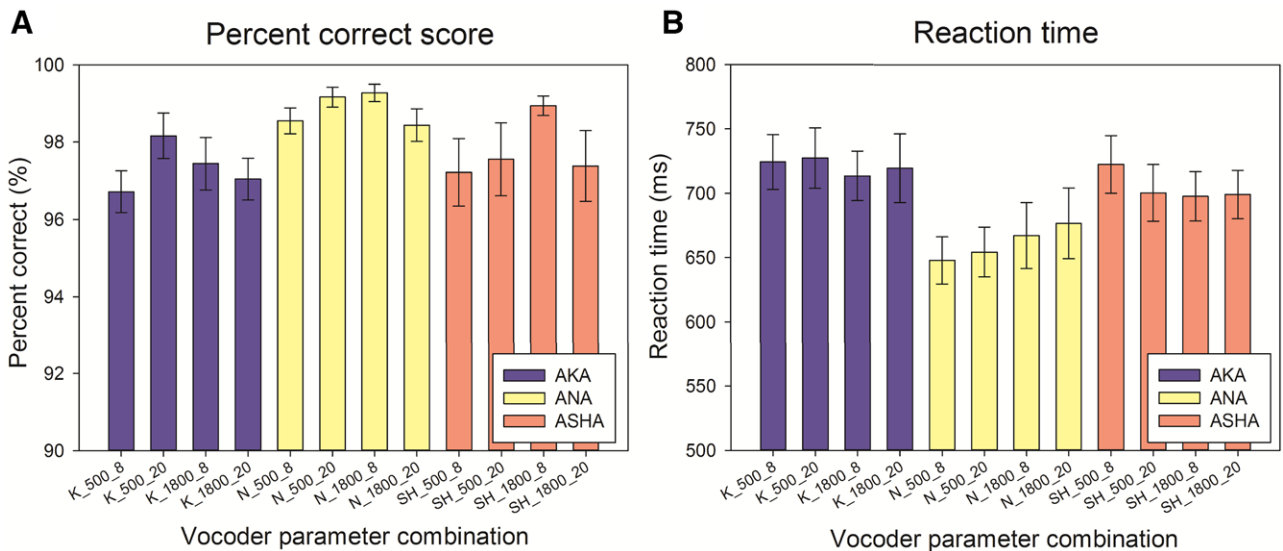
Fig. 2. Group mean behavioral speech identification accuracy scores (A), and reaction times (B) for the 12 CI-simulated speech conditions. The 12 types of stimuli are shown with a format of rate_maxima along the abscissa. Error bars denote ± 1 SEM.

maxima is fixed at a small number (8), higher rates (1800 pps) resulted in better speech identification than the lower rates (500 pps). In contrast, for a larger number of maxima (20), better performance is achieved with a lower (500 pps) compared with a higher rate (1800 pps).

RTs for speech identification are shown in Figure 2B (see also Fig. S2B, Supplemental Digital Content, http://links.lww.com/EANDH/A324). No main effects of either rate or maxima were found in terms of RTs across the stimulus conditions [rate: $F(1, 11) = 0.036$, $p = 0.852$; maxima: $F(1, 11) = 0.088$, $p = 0.773$]. However, there was again a main effect of VCV [$F(2, 22) = 9.96$, $p = 0.001$]. Bonferroni pairwise comparison indicated that /aNa/ was identified significantly faster than either /aKa/ or /aSHa/. The analysis indicated no interaction effects of either VCV × Maxima [$F(2, 22) = 1.768$, $p = 0.194$] or Rate × Maxima [$F(1, 11) = 1.627$, $p = 0.228$]. However, we found an interaction effect between VCV and rate [$F(2, 22) = 3.62$, $p = 0.044$]. For /aKa/ and /aSHa/, higher rates (1800 pps) lead to faster RTs than lower rates (500 pps), whereas for /aNa/, the lower rate (500 pps) caused faster RTs than the higher rate (1800 pps). Collectively these findings suggest that both the accuracy and speed of speech identification is dependent on a specific combination (interaction) of CI SR and number of maxima.

Electrophysiological Results

N1–P2 Responses • Grand average ERPs for each stimulus VCV token and across the four CI vocoder conditions are shown in Figure 3A, B, respectively. Peak amplitude and latency results are shown in Figures 4 and 5, respectively (see also Figs. S4 and S5, Supplemental Digital Content, http://links.lww.com/EANDH/A324). The latencies and amplitudes of the individual N1 and P2 waves did not show
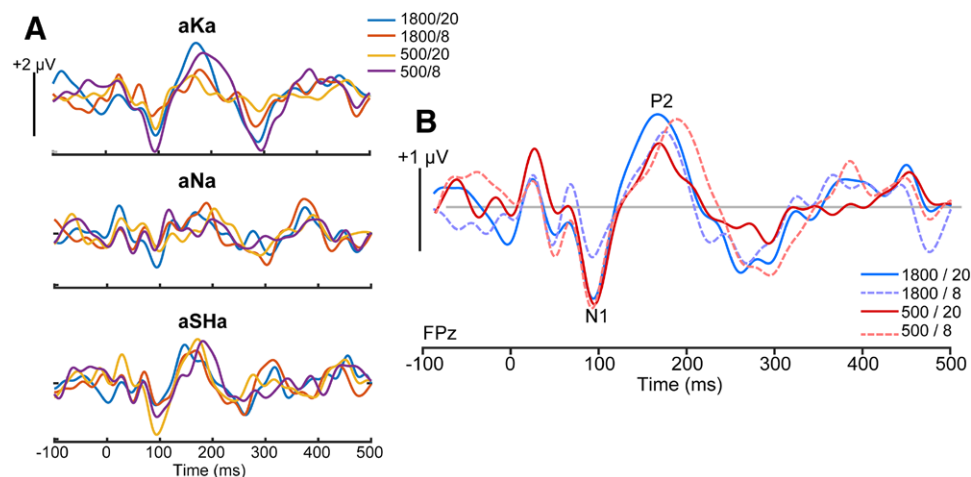

Fig. 3. Grand average cortical ERPs to CI-vocoded speech varying in stimulation rate (1800 vs. 500 pps) and spectral maxima (20 vs. 8 peaks). ERP waveforms (A) by stimulus token and (B) by CI rate/maxima settings (pooled across VCV tokens). Prominent wave deflections of the neural responses (N1, P2) are marked within the traces. $t = 0$ marks the onset of the time-locking speech stimulus. Positive is plotted up.
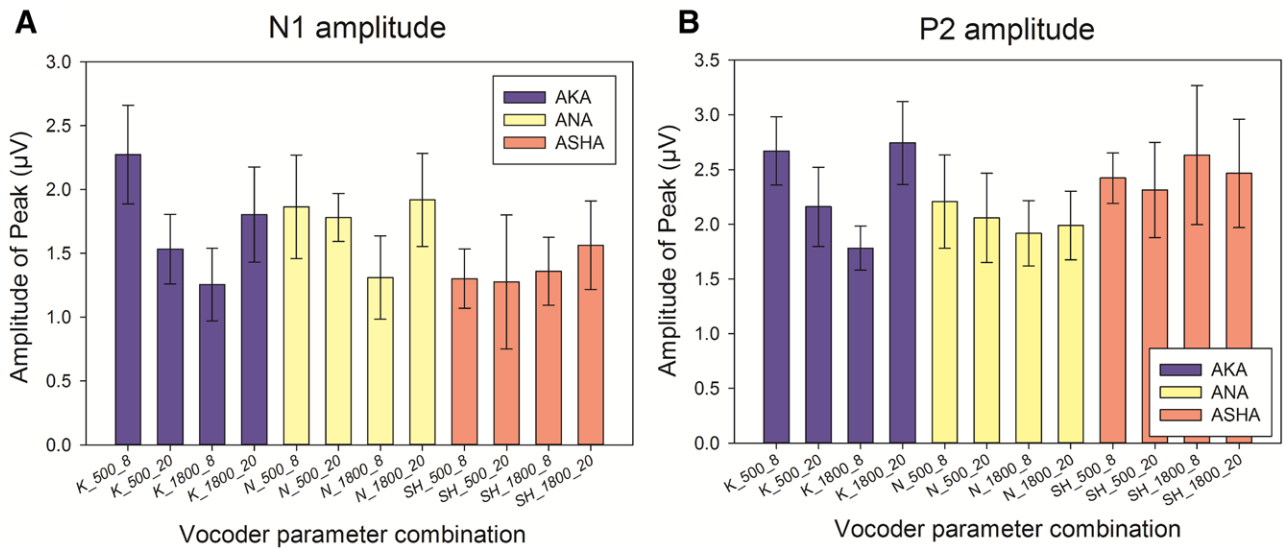
Fig. 4. Group mean amplitudes of the individual N1 (A), and P2 response (B) evoked by CI-simulated speech. Note for the N1, the modulus of the amplitude is shown. Error bars denote ±1 SEM.

a systematic pattern across stimulus conditions. This was confirmed by a rmANOVA, which revealed no significant difference in amplitudes and latencies of each peaks across the conditions; no significant main effects of three within-subject factors (VCV, rate, and maxima) were observed for amplitude of N1[VCV: $F(2, 22) = 2.14$, $p = 0.142$; rate: $F(1, 11) = 0.94$, $p = 0.353$; maxima: $F(1, 11) = 0.493$, $p = 0.50$], amplitude of P2 [VCV: $F(2, 22) = 1.557$, $p = 0.23$; rate: $F(1, 11) = 0.11$, $p = 0.742$; maxima: $F(1, 11) = 0.012$, $p = 0.91$], latency of N1[VCV: $F(2, 22) = 0.83$, $p = 0.45$; rate: $F(1, 11) = 3.69$, $p = 0.081$; maxima: $F(1, 11) = 0.014$, $p = 0.91$], and latency of P2 [VCV: $F(2, 22) = 2.60$, $p = 0.097$; rate: $F(1, 11) = 0.29$, $p = 0.60$; maxima: $F(1, 11) = 0.30$, $p = 0.60$]. That is, we found no systematic modulation in individual

ERP components (i.e., isolated wave measures) with our CI parameter manipulations. Consequently, we next examined whether a multidimensional combination of neural measures could (1) accurately segregate the speech-evoked ERPs and (2) correctly predict listeners' behavioral identification and perceptual confusions.

Brain–Behavior Relations • We used a GLME logistics regression model to assess the relation between ERP measures and behavioral speech identification. Results revealed a significant four-way interaction (i.e., N1amp* P2amp* N1lat* P2lat) between the amplitude and latency measures in predicting behavioral speech identification performance [$t(28) = 10.70$, $p = 0.0014$]†. This indicates that behavioral speech identification was indeed predicted by the combination of
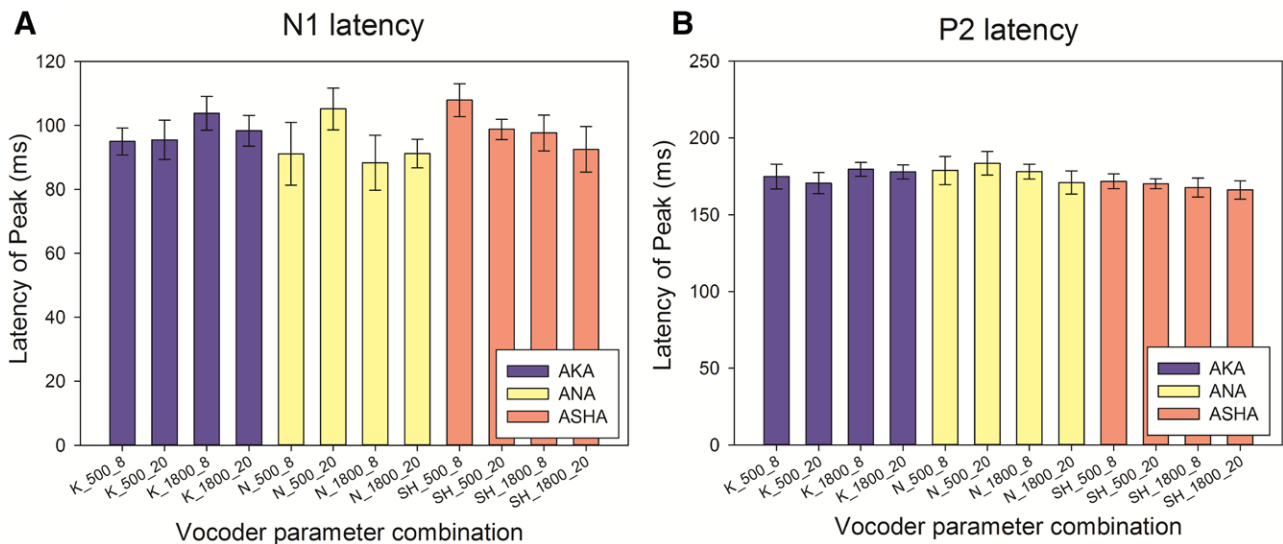

Fig. 5. Group mean latencies of the individual N1 (A), and P2 response (B) evoked by CI-simulated speech. Error bars denote ±1 SEM.

neural measures rather than any singular wave in isolation. Thus, subsequent analyses focused on additional multivariate techniques that considered the combination of measures extracted from ERP responses.

Discriminant Function Analysis • Figure 6 shows neural confusion matrices derived from QDA along with those from the behavioral identification for each of the four pairs of CI rate/maxima parameters. As noted earlier, approaching ceiling effects were observed for behavioral identification scores over all conditions (>97%; cf. Fig. 2). Of particular interest is the performance of the neural data to classify the stimulus VCV class. Neural confusion matrices considered the four variables extracted from listeners' ERP responses (amplitude and latency of N1/P2) at the group level, and how accurately they could segregate the three VCV stimulus inputs. Neural classification matrices output from QDA are shown in the bottom row of Figure 6. Overall, neural responses predicted the evoking stimulus class (/aKa/, /aNa/, /aSHa/) with ~60 to 80% accuracy. Because we can expect to get only 33% of the classifications correct by chance, these accuracy rates represent a considerable improvement above random guess. On average, 77.8% of VCVs were accurately classified in the 1800/8 condition, followed by 500/8 (75%), 500/20 (61.1%), and 1800/20 (61.1%). Thus, neural predictions of perceptual confusions were found to be most accurate for the 1800/8 setting pair.

As expected, considerably fewer correct classifications were made in the cross-validated analysis: 1800/8 (50.0%), 500/8 (47.2%), 500/20 (50.0%), and 1800/20 (27.8%). Nevertheless, classification performance was still well above chance for three of the four parameter settings and highest for the 1800/8 pairing. These findings indicate that speech sounds could be reliably predicted in an objective manner from multivariate measures of listeners' ERPs but not necessarily via conventional univariate (peak analysis) approaches.

IT Analysis • On the basis of the behavioral and neural confusion matrices, we computed the amount of information accessible to listeners via IT analysis. IT is obtained from the confusion matrix by calculating the number of bits accurately predictable by the listener and dividing this result by the number of bits available in the input stimulus (Miller & Nicely 1955; Sagi & Svirsky 2008).

IT results are shown in Figure 7 for behavioral and neural responses. IT scores approaching 100% indicate perfect prediction of the response given the input; values approaching 0 indicate total independence of the stimulus and response. Results show that IT is near ceiling for listeners' behavioral confusions. In contrast, IT computed on neural confusions indicates lower (but nonetheless reliable) rates of IT from ERP responses. IT was largest for the CI parameter combination 1800/8 followed by 500/8, 500/20, and 1800/20, consistent with classification results (Fig. 6).

DISCUSSION

In the present study, we examined the effects of parametric variations in CI processor settings (number of spectral maxima and SR) on the auditory cortical ERPs with the aim of using evoked responses to identify optimal combinations of settings that maximize speech encoding and perception. Univariate

analysis of latencies and amplitudes of the N1 and P2 ERPs failed to determine the best set of parameters that robustly differentiate speech. In contrast, by considering these four evoked responses simultaneously, multivariate analysis (discriminant function and IT analyses) identified a particular set of parameters that best distinguished speech-evoked neural activity. Our findings suggest that the auditory cortical ERPs might be used as an objective technique to identify relevant CI settings that yield maximally contrastive speech coding and are predictive of listeners' behavioral speech identification performance.

Behavioral Identification of CI Speech

Behavioral identification and RTs did not show strong modulations with CI rate or maxima. However, behavioral measures did vary in response to certain VCV tokens (i.e., /aNa/ versus /aKa/ and /aSHa/). This finding is undoubtedly due to acoustic/phonetic features of these consonants. The consonant, /n/, in /aNa/ which is more distinct in terms of phonetic features, such as nasality and voicing, may result in better performance. Indeed, voicing information has been found to be a robust cue in consonant identification studies (Van Tasell et al. 1987; Dannhauer et al. 1990; Dorman et al. 1990; Munson et al. 2003), and is reflected not only in recognition tests but also RT paradigms (Laguitton et al. 2000). The better behavioral performance for /aNa/ likely results from the fact that our other two VCV stimuli (/aKa/ and /aSHa/) share a voiceless feature which could be a source of confusion between these two tokens. This notion is supported by our behavioral confusion matrices (Fig. 6), which showed slightly higher rates of confusions between /aKa/ and /aSHa/.

Identification of Optimal Spectral Maxima and SR CI Settings Via ERPs

Several previous studies have examined the interaction between spectral and temporal cues with regard to CI users' speech perception. Nie et al. (2006) found a trade-off between spectral and temporal information in five MED-EL CI recipients. Their results showed that consonant and sentence recognition scores remained unchanged with increasing number of electrodes coupled with decreasing SR. Such a trade-off between spectral and temporal cues has frequently been observed using variations of different parameters (e.g., cut-off frequency of the envelope filter) in other previous studies (Shannon et al. 2001; Xu et al. 2002, 2005). This implies that the two crucial factors for speech perception, spectral and temporal cues, interact with one another to compensate for the other diminished cue. In the present study, we examined different combinations of spectral (maxima) and temporal (SR) CI settings to examine these possible interactions at a neural level and identify combinations which maximized the encoding and perceptual identification of speech.

Our findings further extend recent studies examining the effects of CI signal processing settings on the auditory ERPs. Using acoustically vocoded consonant–vowel–consonants, Friesen et al. (2009) found a decrease in peak latencies and an increase in amplitudes with increasing number of channels from 2 to 16, indicative of improved speech encoding. However, changes in P1–N1–P2 responses did not correlate with behavioral identification scores (Friesen et al. 2009). Although our study design differed in the CI processing under investigation (e.g., we did not vary channel numbers), our data indicate that it is the interaction of these components (rather than isolated

---

†The GLME also revealed main effects of N1amp ($p < 0.0001$), N1lat ($p = 0.021$), P2lat ($p = 0.021$), and P2amp ($p = 0.016$). However, in light of the significant four-way interaction between measures, these main effects become meaningless.
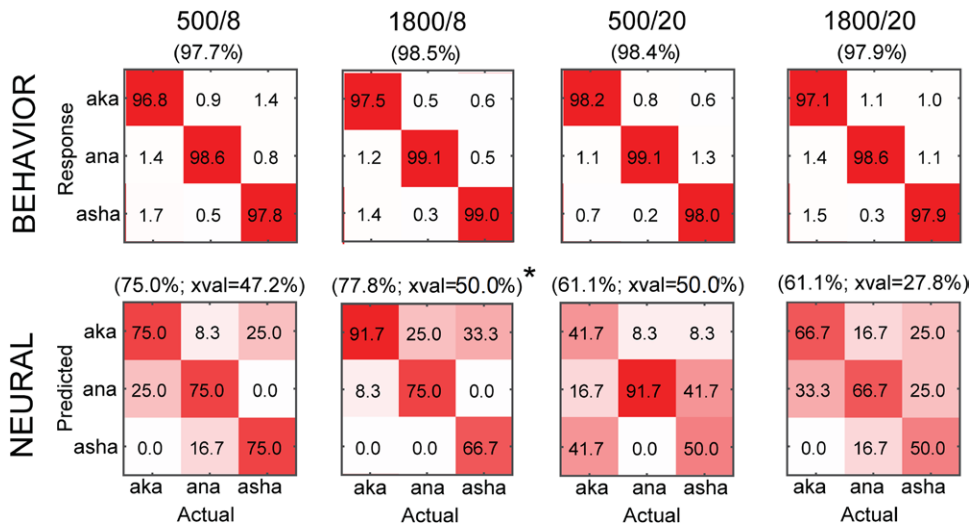
Fig. 6. Behavioral (top row) and neural (bottom row) confusion matrices for speech identification under four different CI rate/maxima parameter combinations. For behavioral confusion matrices, columns represent the true stimulus and rows are participants' perceptual response. For neural confusion matrices, rows represent the predicted VCV class and columns the actual input stimulus. Diagonals show correct responses. Both raw and cross-validated (xval) classification accuracy are shown above each panel. Chance level = 33%. *Rate/maxima combination producing the best cross-validated classification accuracy.

waves) that best predict behavioral responses. Moreover, multivariate analyses of the ERPs lead us to infer that objective determination of CI settings might be best performed by considering a larger set of neural features than what is supplied by individual response components alone (cf. Friesen et al. 2009).

Our findings identified that the SR/maxima combination of 1800/8 produced the most distinctive neural representations
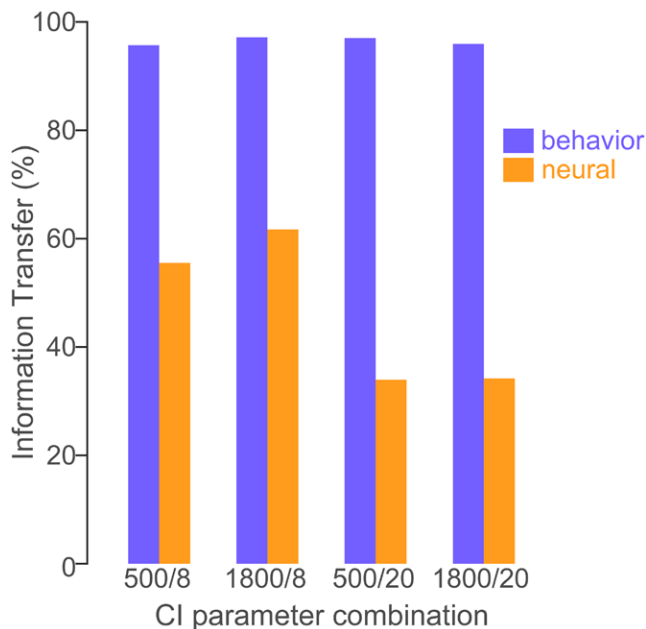

Fig. 7. IT for perceptual and neural confusion matrices. IT represents the degree to which listeners' responses can be accurately predicted given the known input stimulus (Miller & Nicely 1955). Values approaching 100% indicate perfect prediction of the response given the input; values approaching 0 indicate total independent of the stimulus and response. IT is near ceiling for behavioral confusions. IT computed on neural confusions indicate that the CI parameter combination 1800/8 yields maximal IT, consistent with classification results (Fig. 6).

as measured via the ERPs (at least within the parameter space under study, Table 1). Even though the 1800/8 is neither a set having higher spectral cues nor a default setting of manufacturers, we found it was the optimal set via discriminant analysis of the N1/P2 responses. This agrees with anecdotal reports of some CI manufacturers. For example, in the Nucleus 24 system, 1800 SR with 8 channels/maxima is a consistent parameter combination selected by participants in informal listening tasks (Cochlear-corporation 2010). Based on formal empirical CI studies, it is now well accepted that a larger number of CI channels produces better speech identification (Shannon et al. 1995), but that perceptual performance does not change (i.e., asymptotes) once a patient has saturated in their performance (Loizou et al. 1999; Dorman et al. 2002). Our findings, however, reveal that a parameter set with a relatively small number of maxima (8) might be superior to those with a higher number of maxima (20), at higher SRs (and in normal-hearing listeners). This again highlights an important interaction between spectral and temporal processing of CI devices (Nie et al. 2006).

In general, we did not find that the fastest SR and highest number of spectral maxima yielded the highest accuracy in neural classifications. Instead, the highest SR coupled with the lowest maxima under study (i.e., 1800/8) produced optimal results. A plausible implication of this finding is that both maxima and SR have reciprocal influence on the quality of CI-processed sounds resulting in different speech perception performance. This is consistent with findings from other studies in which optimal SR or channels are not simply at the extreme end of a given parameter space (Friesen et al. 2005). Instead, our data are most consistent with the notion that there are ranges of useful settings and interactions between CI parameters which lead to maximum speech perception (Holden et al. 2002; Cochlear-corporation 2010).

Univariate Versus Multivariate Techniques for CI Mapping Via ERPs

Our findings highlight the fact that (1) there is considerable behaviorally-relevant information contained in scalp-recorded

potentials and (2) these correlates are only revealed when considering the complete series of brain responses unfolding over time. In this regard, our study corroborates several recent attempts to apply multivariate approaches (e.g., principal component analysis, DA) to the analysis of ERPs and their prediction of speech behaviors (Do & Kirk 1999; Chang et al. 2010; Davies et al. 2010; Bidelman et al. 2013). In particular, we have previously shown that using multivariate techniques (clustering, multidimensional scaling), listeners' speech-ERPs can be correctly classified into perceptually-meaningful groups which mimic their categorical perception (Bidelman et al. 2013; Bidelman & Alain 2015; Bidelman & Lee 2015). That is, (multivariate) brain activity clusters according to phonetic rather than acoustic rules and mirrors listeners' behavioral speech classification. These neurophysiological correlates of speech identification are not always revealed when considering only individual ERP components (Bidelman et al. 2014b). While we did not find reliable modulations of the individual ERP waves, we also note that our use of highly degraded (vocoded) tokens may have precluded strong effects at the single component level. We would however predict much stronger stimulus-related changes in the N1 and/or P2 for clean (undegraded) speech (Agung et al. 2006; Bidelman, et al. 2014b).

In this regard, we infer that prominent peaks of the ERPs should be considered a series of related (rather than independent) neural events that mark the temporal evolution of auditory speech processing. While traditional univariate approaches (e.g., ANOVA) are concerned only with independently assessing variation in singular peak characteristics, multivariate approaches (adopted here) consider multiple properties of the neural response simultaneously and better characterize the overall pattern of neural response to speech. This notion is supported by current data: optimal CI settings were identified by considering a combination of neural measures whereas individual ERP waves in isolation failed to show systematic modulations across CI speech conditions.

Although the best-parameter setting identified via behavioral and neural responses were identical (1800/8) in the present study, some caution is warranted when drawing conclusions between the neural and behavioral data. First, the ranking from best to worst in accuracy was somewhat discrepant between behavioral and neural data (Fig. 6); 1800/8 > 500/20 > 1800/20 > 500/8 for behavioral versus 1800/8 > 500/8 > 500/20 = 1800/20 for neural responses. This latter neural pattern was also identified in our ERP IT results. We assume that this discrepancy is presumably caused by the approaching ceiling effect in behavioral identification which muted differences in performance across the four parameter settings. However, we would argue that identifying the most optimal parameters that maximize neural/behavioral speech identification accuracy is most important for potential application to CI fitting rather than the exact ordering of conditions according to their diminishing performance. Second, neural confusion patterns also differed from those observed in behavioral identification test. Behaviorally, /aKa/ or /aSHa/ were less confused with /aNa/ (and elicited slower RTs) due to the latter's more distinctive phonetic features (e.g., voicing) that /aKa/ and /aSHa/ do not contain (Fig. 6). This pattern, however, was not identified in the confusion matrix from our neurophysiological data. The N1–P2 are generally considered to reflect preperceptual processing and stimulus coding whereas RTs/identifications reflect post perceptual processing after the decision event. It is possible that auditory coding and the decision reflect distinct channels

of information processing (Binder et al. 2004). Alternatively, a causal chain of events might predict that stimulus encoding (N1–P2) would be related to the later decision (RTs were also ~500 msec after the P2). Indeed, we have observed correlations between N1–P2 and RTs/identifications in our previous studies (Bidelman et al. 2014b; Bidelman & Alain 2015) but this was for clean speech perception tasks. It is possible that signal degradation (noise, vocoding) produce further delays in the decision process and poorer identification of certain VCVs (/aKa/ and /aSHa/) that decouple sensory coding (N1–P2) from behavior. In addition, the ERPs are known to largely index transient events of the acoustic signal (Picton et al. 1974, 1978). Thus, some discrepancies between behavioral and neural confusion patterns and accuracy might be expected if ERPs miss idiosyncrasies of the speech signal that are exploited at the behavioral level and only reflect the initial onset of speech tokens. However, we find this explanation unlikely in light of our recent ERP work in categorical perception, which demonstrates that speech sounds (with identical stimulus onsets) can be used to decode speech identification from the auditory ERPs well above chance levels (Bidelman et al. 2013).

Clinical Implications, Limitations, and Directions for Future Research

CI mapping is a two-way communication between an audiologist and patient that seeks the best fit by adjusting system parameters under time constraints. Our study highlights a potential application of ERP responses in the objective fitting and optimization of CI parameters. Presumably, this technique could be extended to map implant settings in difficult-to-test patients or young infants where traditionally behavioral responses are impractical.

Nevertheless, several limitations of the present study are worth noting. First, our behavioral identification scores showed considerable ceiling effects due to the small number of VCV alternatives and use of normal-hearing listeners. While we did observe a similar match between neural and behavioral speech confusions (Fig. 6), little variation in the behavioral data (other than a VCV effect) preclude firm conclusions regarding a strong brain–behavior correspondence.

Second, the present study assessed acoustically vocoded CI speech in normal-hearing listeners using an active listening paradigm. Although vocoded speech is thought to represent an accurate acoustic proxy of real CI stimulation (Shannon et al. 1995), it remains to be tested whether our approach will generalize to actual implant patients. Although ERPs are often prolonged and weaker in implant patients, the N1 and P2 waves are still identifiable (Friesen & Picton 2010; Atcherson et al. 2011; Mc Laughlin et al. 2013), suggesting that the current approach should be applicable to actual CI users. In general, the N1 and P2 also vary between active and passive recording conditions; amplitudes often increase in active conditions compared with passive conditions (Billings et al. 2011a). For potential objective mapping for uncooperative patients, future studies using passive paradigms could be conducted to determine if speech classification from the ERPs is still robust without having listeners actively engaging with sound. Several other practical restrictions, including the removal of electromagnetic artifact of the implant (Sporns et al. 2004; Hoeft et al. 2007; Mc Laughlin et al. 2013), would need to be overcome for our approach to be directly translated to CI patients.

Third, the CI settings used in our study were selected from a closed set of two parameters (number of spectral maxima and SR) and three speech stimuli. Maxima is also a parameter only available in n-of-m strategy in Nucleus devices (Cochlear Corporation, Macquarie University, NSW, Australia) and the role of SR may vary depending on signal processing strategies. Future studies are needed to confirm that the proposed approach generalizes and can reliably identify optimal settings from among an open set and other relevant implant parameters (e.g., threshold and upper stimulation levels).

Finally, optimal parameter settings typically vary across individuals (Plant et al. 2002; Arora et al. 2009). Our approach was based on DA which determines accuracy at the group level. While a DA allows for classification (prediction) of new observations, vetting the classifier developed in the present study would require application to a different (independent) dataset. Nevertheless, to reach a fuller understanding of the neural consequences of CI processor settings, it would be worthwhile to analyze patients' responses on an individual or single trial basis.

REFERENCES

Agung, K., Purdy, S. C., McMahon, C. M., et al. (2006). The use of cortical auditory evoked potentials to evaluate neural encoding of speech sounds in adults. *J Am Acad Audiol*, 17, 559–572.

Alain, C., Snyder, J. S., He, Y., et al. (2007). Changes in auditory cortex parallel rapid perceptual learning. *Cereb Cortex*, 17, 1074–1084.

Alain, C., Roye, A., Salloum, C. (2014). Effects of age-related hearing loss and background noise on neuromagnetic activity from auditory cortex. *Front Syst Neurosci*, 8, 8.

Arora, K., Dawson, P., Dowell, R., et al. (2009). Electrical stimulation rate effects on speech perception in cochlear implants. *Int J Audiol*, 48, 561–567.

Atcherson, S. R., Damji, Z., Upson, S. (2011). Applying a subtraction technique to minimize cochlear implant artifact with soundfield and direct audio input stimulations. *Cochlear Implants Int*, 12, 234–237.

Bidelman, G. M. (2015). Towards an optimal paradigm for simultaneously recording cortical and brainstem auditory evoked potentials. *J Neurosci Methods*, 241, 94–100.

Bidelman, G. M., & Alain, C. (2015). Musical training orchestrates coordinated neuroplasticity in auditory brainstem and cortex to counteract age-related declines in categorical vowel perception. *J Neurosci*, 35, 1240–1249.

Bidelman, G. M., & Howell, M. (2016). Functional changes in inter- and intra-hemispheric cortical processing underlying degraded speech perception. *Neuroimage*, 124(Pt A), 581–590.

Bidelman, G. M., & Lee, C. C. (2015). Effects of language experience and stimulus context on the neural organization and categorical perception of speech. *Neuroimage*, 120, 191–200.

Bidelman, G. M., Moreno, S., Alain, C. (2013). Tracing the emergence of categorical speech perception in the human auditory system. *Neuroimage*, 79, 201–212.

Bidelman, G. M., Villafuerte, J. W., Moreno, S., et al. (2014a). Age-related changes in the subcortical-cortical encoding and categorical perception of speech. *Neurobiol Aging*, 35, 2526–2540.

Bidelman, G. M., Weiss, M. W., Moreno, S., et al. (2014b). Coordinated plasticity in brainstem and auditory cortex contributes to enhanced categorical speech perception in musicians. *Eur J Neurosci*, 40, 2662–2673.

Billings, C. J., Bennett, K. O., Molis, M. R., et al. (2011a). Cortical encoding of signals in noise: Effects of stimulus type and recording paradigm. *Ear Hear*, 32, 53–60.

Billings, C. J., Tremblay, K. L., Miller, C. W. (2011b). Aided cortical auditory evoked potentials in response to changes in hearing aid gain. *Int J Audiol*, 50, 459–467.

Binder, J. R., Liebenthal, E., Possing, E. T., et al. (2004). Neural correlates of sensory and decision processes in auditory object identification. *Nat Neurosci*, 7, 295–301.

Boersma, P., Weenink, D. (2013). Praat: Doing phonetics by computer (Version 5.3.51) [Computer program]. Retrieved September 1, 2013, from http://www.praat.org.

Buechner, A., Frohne-Buechner, C., Boyle, P., et al. (2009). A high rate n-of-m speech processing strategy for the first generation Clarion cochlear implant. *Int J Audiol*, 48, 868–875.

Casserly, E. D. (2015). Effects of real-time cochlear implant simulation on speech production. *J Acoust Soc Am*, 137, 2791–2800.

Chang, E. F., Rieger, J. W., Johnson, K., et al. (2010). Categorical speech representation in human superior temporal gyrus. *Nat Neurosci*, 13, 1428–1432.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychol Assess*, 6, 284–290.

Cochlear-corporation. (2010). Nucleus® cochlear implants - Physician's Package Insert.

Cullington, H. E., & Zeng, F. G. (2011). Comparison of bimodal and bilateral cochlear implant users on speech recognition with competing talker, music perception, affective prosody discrimination, and talker identification. *Ear Hear*, 32, 16–30.

Danhauer, J. L., Ghadialy, F. B., Beck, D. L., et al. (1990). Audio-visual consonant recognition with the 3M/House cochlear implant. *J Rehabil Res Dev*, 27, 247–254.

Davies, P. L., Chang, W. P., Gavin, W. J. (2010). Middle and late latency ERP components discriminate between adults, typical children, and children with sensory processing disorders. *Front Integr Neurosci*, 4, 16.

de la Torre Vega, Á., Martí, M., de la Torre Vega, R., et al. (2004). Cochlear Implant Simulation version 2.0: Description and usage of the program. *University of Granada, Spain*.

Do, K. A., & Kirk, K. (1999). Discriminant analysis of event-related potential curves using smoothed principal components. *Biometrics*, 55, 174–181.

Dorman, M. F., Loizou, P. C., Rainey, D. (1997). Simulating the effect of cochlear-implant electrode insertion depth on speech understanding. *J Acoust Soc Am*, 102(5 Pt 1), 2993–2996.

Dorman, M. F., Loizou, P. C., Spahr, A. J., et al. (2002). A comparison of the speech understanding provided by acoustic models of fixed-channel and channel-picking signal processors for cochlear implants. *J Speech Lang Hear Res*, 45, 783–788.

Dorman, M. F., Soli, S., Dankowski, K., et al. (1990). Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant. *J Acoust Soc Am*, 88, 2074–2079.

Firszt, J. B., Chambers, R. D., Kraus And, N., et al. (2002). Neurophysiology of cochlear implant users I: Effects of stimulus current level and electrode site on the electrical ABR, MLR, and N1-P2 response. *Ear Hear*, 23, 502–515.

Fishman, K. E., Shannon, R. V., Slattery, W. H. (1997). Speech recognition as a function of the number of electrodes used in the SPEAK cochlear implant speech processor. *J Speech Lang Hear Res*, 40, 1201–1215.

Friesen, L. M., & Picton, T. W. (2010). A method for removing cochlear implant artifact. *Hear Res*, 259, 95–106.

Friesen, L. M., Shannon, R. V., Cruz, R. J. (2005). Effects of stimulation rate on speech recognition with cochlear implants. *Audiol Neurootol*, 10, 169–184.

Friesen, L. M., Tremblay, K. L., Rohila, N., et al. (2009). Evoked cortical activity and speech recognition as a function of the number of simulated cochlear implant channels. *Clin Neurophysiol*, 120, 776–782.

Ghrissi, M., & Cherif, A. (2012). Comparison of IIR filterbanks and FFT filterbanks in cochlear implant speech processing strategies. *J. Electrical Systems*, 76–84.

Gilley, P. M., Sharma, A., Dorman, M., et al. (2006). Minimization of cochlear implant stimulus artifact in cortical auditory evoked potentials. *Clin Neurophysiol*, *117*, 1772–1782.

Gordon, K. A., Tanaka, S., Papsin, B. C. (2005). Atypical cortical responses underlie poor speech perception in children using cochlear implants. *Neuroreport*, *16*, 2041–2045.

Groenen, P. A., Beynon, A. J., Snik, A. F., et al. (2001). Speech-evoked cortical potentials and speech recognition in cochlear implant users. *Scand Audiol*, *30*, 31–40.

Hoeft, F., Meyler, A., Hernandez, A., et al. (2007). Functional and morphometric brain dissociation between dyslexia and reading ability. *Proc Natl Sci U S A*, *104*, 4234–4239.

Holden, L. K., Skinner, M. W., Holden, T. A., et al. (2002). Effects of stimulation rate with the Nucleus 24 ACE speech coding strategy. *Ear Hear*, *23*, 463–476.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *J Mem Lang*, *59*, 434–446.

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.

Kiefer, J., Hohl, S., Stürzebecher, E., et al. (2001). Comparison of speech recognition with different speech coding strategies (SPEAK, CIS, and ACE) and their relationship to telemetric measures of compound action potentials in the nucleus CI 24M cochlear implant system. *Audiology*, *40*, 32–42.

Kim, J. R., Brown, C. J., Abbas, P. J., et al. (2009). The effect of changes in stimulus level on electrically evoked cortical auditory potentials. *Ear Hear*, *30*, 320–329.

Krishnan, A., Gandour, J. T., & Bidelman, G. M. (2010). Brainstem pitch representation in native speakers of Mandarin is less susceptible to degradation of stimulus temporal regularity. *Brain Res.*, *1313*, 124–133.

Laguitton, V., De Graaf, J. B., Chauvel, P., et al. (2000). Identification reaction times of voiced/voiceless continua: A right-ear advantage for VOT values near the phonetic boundary. *Brain Lang*, *75*, 153–162.

Loizou, P. C., Dorman, M. F., Powell, V. (1998). The recognition of vowels produced by men, women, boys, and girls by cochlear implant patients using a six-channel CIS processor. *J Acoust Soc Am*, *103*, 1141–1149.

Loizou, P. C., Dorman, M., Tu, Z. (1999). On the number of channels needed to understand speech. *J Acoust Soc Am*, *106*(4 Pt 1), 2097–2103.

Loizou, P. C., Poroy, O., Dorman, M. (2000). The effect of parametric variations of cochlear implant processors on speech understanding. *J Acoust Soc Am*, *108*, 790–802.

Luck, S. J. (2005). *An Introduction to the Event-Related Potential Technique*. MIT Press, Cambridge MA.

Martin, B. A. (2007). Can the acoustic change complex be recorded in an individual with a cochlear implant? Separating neural responses from cochlear implant artifact. *J Am Acad Audiol*, *18*, 126–140.

Mc Laughlin, M., Lopez Valdes, A., Reilly, R. B., et al. (2013). Cochlear implant artifact attenuation in late auditory evoked potentials: A single channel approach. *Hear Res*, *302*, 84–95.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *J Acoust Soc Am*, *27*, 338–352.

Munson, B., Donaldson, G. S., Allen, S. L., et al. (2003). Patterns of phoneme perception errors by listeners with cochlear implants as a function of overall speech perception ability. *J Acoust Soc Am*, *113*, 925–935.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. *Psychophysiology*, *24*, 375–425.

Nie, K., Barco, A., Zeng, F. G. (2006). Spectral and temporal cues in cochlear implant speech perception. *Ear Hear*, *27*, 208–217.

Novak, G. P., Kurtzberg, D., Kreuzer, J. A., et al. (1989). Cortical responses to speech sounds and their formants in normal infants: Maturational sequence and spatiotemporal analysis. *Electroencephalogr Clin Neurophysiol*, *73*, 295–305.

Parbery-Clark, A., Marmel, F., Bair, J., et al. (2011). What subcortical-cortical relationships tell us about processing speech in noise. *Eur J Neurosci*, *33*, 549–557.

Picton, T. W., Hillyard, S. A., Krausz, H. I., et al. (1974). Human auditory evoked potentials. I. Evaluation of components. *Electroencephalogr Clin Neurophysiol*, *36*, 179–190.

Picton, T. W., Woods, D. L., Baribeau-Braun, J., et al. (1976). Evoked potential audiometry. *J Otolaryngol*, *6*, 90–119.

Picton, T. W., Woods, D. L., Proulx, G. B. (1978). Human auditory sustained potentials. I. The nature of the response. *Electroencephalogr Clin Neurophysiol*, *45*, 186–197.

Picton, T. W., Alain, C., Woods, D. L., et al. (1999). Intracerebral sources of human auditory-evoked potentials. *Audiol Neurootol*, *4*, 64–79.

Plant, K. L., Whitford, L. A., Psarros, C. E., et al. (2002). Parameter selection and programming recommendations for the ACE and CIS speech-processing strategies in the Nucleus 24 cochlear implant system. *Cochlear Implants Int*, *3*, 104–125.

Sagi, E., & Svirsky, M. A. (2008). Information transfer analysis: A first look at estimation bias. *J Acoust Soc Am*, *123*, 2848–2857.

Samira, A., Bharath, C., Han-Gyol, Y., et al. (2010). Cortical-evoked potentials reflect speech-in-noise perception in children. *Eur J Neurosci*, *32*, 1407–1413.

Sandman, C. A., & Patterson, J. V. (2000). The auditory event-related potential is a stable and reliable measure in elderly subjects over a 3 year period. *Clin Neurophysiol*, *111*, 1427–1437.

Shannon, R. V., Zeng, F. G., Kamath, V., et al. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Shannon, R., Fu, Q., Wang, X., et al. (2001). Critical cues for auditory pattern recognition in speech: Implications for cochlear implant speech processor design. In *Physiological and psychological bases of auditory function: Proceedings of the 12th International Symposium on Hearing* (pp. 500–508). Maastricht: Shaker Publishing BV.

Shannon, R. V., Cruz, R. J., Galvin, J. J. 3rd. (2011). Effect of stimulation rate on cochlear implant users' phoneme, word and sentence recognition in quiet and in noise. *Audiol Neurootol*, *16*, 113–123.

Sharma, A., & Campbell, J. (2011). A sensitive period for cochlear implantation in deaf children. *J Matern Fetal Neonatal Med*, *24*, 151–153.

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

Sharma, A., Dorman, M. F., Kral, A. (2005). The influence of a sensitive period on central auditory development in children with unilateral and bilateral cochlear implants. *Hear Res*, *203*, 134–143.

Skinner, M. W., Holden, L. K., Whitford, L. A., et al. (2002). Speech recognition with the nucleus 24 SPEAK, ACE, and CIS speech coding strategies in newly implanted adults. *Ear Hear*, *23*, 207–223.

Sporns, O., Chialvo, D. R., Kaiser, M., et al. (2004). Organization, development and function of complex brain networks. *Trends Cogn Sci*, *8*, 418–425.

Studebaker, G. A. (1985). A "rationalized" arcsine transform. *J Speech Hear Res*, *28*, 455–462.

Tremblay, K., Kraus, N., McGee, T., et al. (2001). Central auditory plasticity: Changes in the N1-P2 complex after speech-sound training. *Ear Hear*, *22*, 79–90.

Tremblay, K. L., Piskosz, M., Souza, P. (2003). Effects of age and age-related hearing loss on the neural representation of speech cues. *Clin Neurophysiol*, *114*, 1332–1343.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., et al. (1987). Speech waveform envelope cues for consonant recognition. *J Acoust Soc Am*, *82*, 1152–1161.

Vandali, A. E., Whitford, L. A., Plant, K. L., et al. (2000). Speech perception as a function of electrical stimulation rate: Using the Nucleus 24 cochlear implant system. *Ear Hear*, *21*, 608–624.

Xu, L., Tsai, Y., Pfingst, B. E. (2002). Features of stimulation affecting tonal-speech perception: Implications for cochlear prostheses. *J Acoust Soc Am*, *112*, 247–258.

Xu, L., Thompson, C. S., Pfingst, B. E. (2005). Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am*, *117*, 3255–3267.

Zhang, F., Samy, R. N., Anderson, J. M., et al. (2009). Recovery function of the late auditory evoked potential in cochlear implant users and normal-hearing listeners. *J Am Acad Audiol*, *20*, 397–408.

Zhang, F., Anderson, J., Samy, R., et al. (2010). The adaptive pattern of the late auditory evoked potential elicited by repeated stimuli in cochlear implant users. *Int J Audiol*, *49*, 277–285.