



Evoked cortical activity and speech recognition as a function of the number of simulated cochlear implant channels

L.M. Friesen*, K.L. Tremblay, N. Rohila, R.A. Wright, R.V. Shannon, D. Başkent, J.T. Rubinstein

Department of Speech and Hearing Sciences, University of Washington, Seattle, WA 98105, USA

ARTICLE INFO

Article history:

Accepted 26 January 2009

Available online 27 February 2009

Keywords:

Cochlear implants
Acoustic change complex
Cortical auditory evoked potentials
Electrophysiology
Plasticity
Implant channel number
P100
N100
ERP
P1–N1–P2 complex
Cochlear implant simulations

ABSTRACT

Objectives: (1) To determine if consonant–vowel–consonant (CVC) syllables [Hillenbrand J, Getty L, Clark M, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am* 1995;97:3099–3111] could be used to evoke cortical far field response patterns in humans, (2) to characterize the effects of cochlear implant–simulated channel number on the perception and physiological detection of these same CVC stimuli, and (3) to define the relationship between perception and the morphology of the physiological responses evoked by these speech stimuli.

Methods: Ten normal hearing monolingual English speaking adults were tested. Unprocessed CVC naturally spoken syllables, containing medial vowels, as well as processed versions (2, 4, 8, 12, and 16 spectral channels) were used for behavioral and physiological testing.

Results: (1) CVC stimuli evoked a series of overlapping P1–N1–P2 cortical responses. (2) Amplitude of P1–N1–P2 responses increased as neural conduction time (latency) decreased with increases in the number of spectral channels. Perception of the CVC stimuli improved with increasing number of spectral channels. (3) Coinciding changes in P1–N1–P2 morphology did not significantly correlate with changes in perception.

Conclusions: P1–N1–P2 responses can be recorded using CVC syllables and there is an effect of channel number on the latency and amplitude of these responses, as well as on vowel identification. However, the physiological detection of the acoustic changes does not fully account for the perceptual performance of these same syllables.

Significance: These results provide evidence that it is possible to use vocoded CVC stimuli to learn more about the physiological detection of acoustic changes contained within speech syllables, as well as to explore brain–behavior relationships.

© 2009 International Federation of Clinical Neurophysiology. Published by Elsevier Ireland Ltd. All rights reserved.

1. Introduction

Multi-channel cochlear implants (CIs) divide acoustic input into frequency bands, extract temporal envelope information from each band, and electrically activate appropriate channels to stimulate multiple sites along the cochlea, using current pulses modulated by the temporal envelopes (Fu and Galvin, 1990; Loizou, 1999). Vocoded speech, a synthesized version of this signal, is commonly used to simulate the effects of signal processing provided by a cochlear implant (Shannon et al., 1995). When normal hearing listeners are tested using vocoded speech, researchers can evaluate how the auditory system processes degraded auditory signals, similar to those delivered by a cochlear implant. Simulations are often used

because they allow investigators to carefully control stimulus processing schemes when testing the perception and physiological representation of the test stimuli, while minimizing the many confounding variables that come with testing actual CI patients. With this information, investigators can then make inferences about perception in CI users and conduct similar studies in CI listeners for comparative purposes, all with the intention of developing or improving our basic understanding of speech perception in normal and disordered populations.

Using vocoded speech stimuli, we previously demonstrated (Friesen et al., 2001) that speech recognition performance improves as the number of active channels available to the listener is increased, perhaps because temporal and spectral resolution of the incoming signal improves (Fig. 1a). Take for example the consonant–vowel–consonant (CVC) syllables /hid/ (as in ‘heed’) and /hld/ (as in ‘hid’) used in the Friesen et al. (2001) study. Fig. 2 shows the acoustic waveform and spectrographs of these speech tokens in their natural (unprocessed), and processed/vocoded (2 and 8

* Corresponding author. Address: Sunnybrook Health Sciences Centre, Department of Otolaryngology, 2075 Bayview Avenue, Room M1-102 Toronto, Ont., Canada M4N 3M5. Tel.: +1 416 480 6100x89624; fax: +1 416 480 5761.

E-mail address: lendra.friesen@sunnybrook.ca (L.M. Friesen).

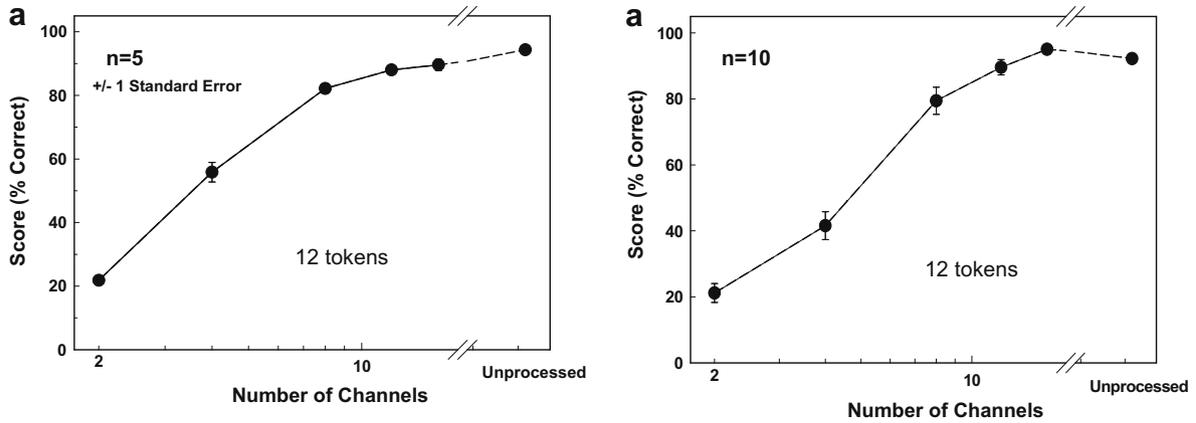


Fig. 1. Consonant–vowel–consonant (CVC) scores (12 tokens) vs. number of active channels averaged across normal hearing listeners (a), adapted from Friesen et al. (2001), and from the current study (b). Error bars represent ± 1 standard error.

channels) states. When comparing the spectrographs of each channel condition, multiple differences are seen: (1) Formants are clearly identifiable in the unprocessed vowel /i/ and /l/ but become less distinct in the 8 and 2 channel conditions. (2) In the unprocessed condition, the onset of /h/ is sharp but as the number of channels decreases, the onset for the glottal fricative /h/ is less defined. (3) Decreased channel number also blurs four CV (h–i, h–l), and VC (i–d, l–d) transitions.

Even though the effects of increased number of channels on perception has been documented, the physiological mechanisms underlying vocoded speech recognition are poorly understood. If we assume that perception is dependent on the physiological detection of spectral and temporal cues, it is logical to question if the number of CI channels also affects the physiological representation of spectral and temporal cues in the central auditory nervous system (CANS). For example, improved neural representation of acoustic transitions contributing to envelope encoding might underlie some of the perceptual gains seen with increasing channels of acoustic information.

Cortical auditory evoked potentials (CAEPs) are one method of measuring the neural representation of sound. CAEPs are non-invasive measures that are frequently used to examine the neural detection of sound in humans. Typically, clicks, tones, or other short duration sounds are used as stimuli to evoke CAEPs. While click stimuli are useful in that they simulate processor stimulation patterns, and brief synthetic speech sounds allow the investigator to control stimulus dimensions, these stimuli are not representative of everyday speech sounds heard by implant users. Moreover, because of the brevity of these stimuli, the evoked neural response patterns do not reflect some of the acoustic features that differentiate speech sounds (Friesen and Tremblay, 2006). Naturally produced speech sounds, however, are highly complex time-varying signals. They evoke complex neural response patterns (Polen, 1984; Ostroff et al., 1998), and might be more effective than clicks, tones, and short duration synthetic speech sounds for identifying neural processing problems in people with impaired speech understanding.

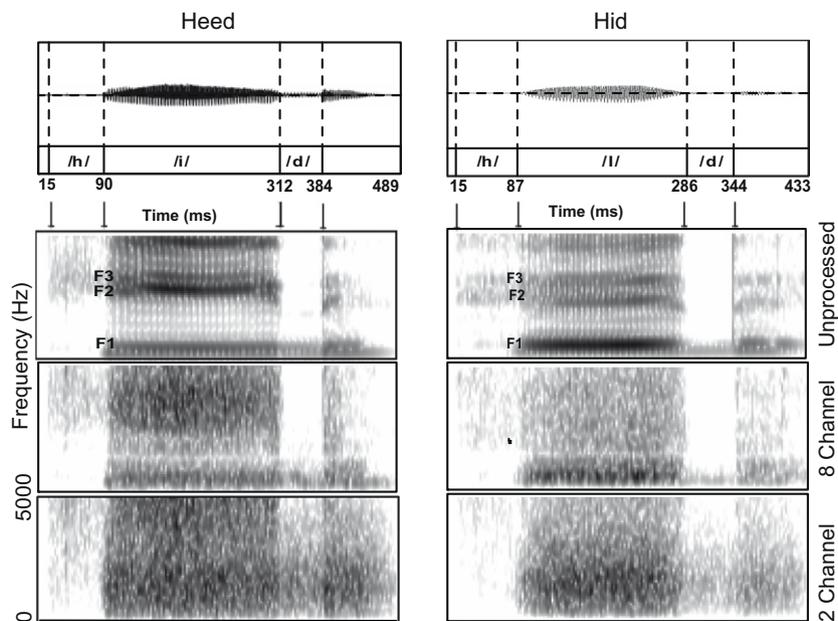


Fig. 2. Channel effects on the acoustics of ‘heed’ or /hid/ and ‘hid’ or /hld/. Top: Acoustic waveforms of syllables /hid/ and /hld/. The X-axis represents time in ms and the Y-axis represents amplitude. Acoustic transitions for all phonemes are labeled. Bottom: Three spectrographs of the syllables /hid/ and /hld/. The X-axis represents time in ms and the Y-axis represents frequency. The first row of spectrographs shows the unprocessed stimuli, the middle row is an 8 channel simulation, and the bottom row is a 2 channel simulation. Dark regions represent amount of acoustic energy present, with greater energy in darker areas. F1 = formant 1, F2 = formant 2, F3 = formant 3.

The P1–N1–P2 response is a CAEP consisting of a series of positive and negative peaks. The N1 component is an onset response that reflects synchronous neural activation of structures of the thalamic–cortical segment of the CANS in response to acoustic change (Naatanen and Picton, 1987). In other words, its presence reflects the neural response to acoustic change at the level of the cortex, and for this reason is sometimes called the acoustic change complex (ACC) (Martin and Boothroyd, 1999). Examples of acoustic change include silence to sound, or pitch changes within an ongoing signal. For example, the P1–N1–P2 is sensitive to acoustic frequency changes within a tone (McCandless and Rose, 1970; Naatanen and Picton, 1987), amplitude and frequency changes within a vowel (Martin and Boothroyd, 2000), changes in periodicity (Martin and Boothroyd, 1999), and within a CV syllable (Ostroff et al., 1998). When a CV syllable is used to stimulate the auditory system, multiple overlapping P1–N1–P2 waveforms are recorded (Kaukoranta et al., 1987; Ostroff et al., 1998; Tremblay et al., 2003; Friesen and Tremblay, 2006).

Because there is interest in examining brain–behavior relationships associated with speech perception, especially among cochlear implant users, it would be helpful to examine the neural representation of speech using the same stimuli that are often used clinically to measure speech perception in implant users (e.g., such as CVC syllables). However, acoustic changes contained in CVC syllables have not been reported in the literature and it is unknown if overlapping P1–N1–P2 responses, reflecting acoustic changes within each stimulus, can be recorded. Therefore, the first objective of this experiment was to determine if acoustic changes within a CVC syllable could be measured.

The second objective was to characterize evoked cortical neural response patterns and speech understanding as a function of the number of spectral channels. Our hypothesis was that the physiological detection of acoustic transitions would be altered by the number of simulated implant channels. More specifically, because increased amplitudes (strength of the synchronized response) and decreased latencies (neural conduction time) are associated with clearer transmission of acoustic information, we hypothesized that N1 amplitudes would increase and N1 latencies would decrease as the number of CI channels increased. Finally, the third objective was to determine if there was a relationship between speech understanding and these physiological responses. We hypothesized that shorter latencies and larger amplitudes would coincide with improved speech understanding.

2. Methods

2.1. Subjects

Listeners were 10 normal hearing monolingual American English speakers (22–40 years; five women and six men). Auditory thresholds fell within normal limits bilaterally (<25 dB HL) and were symmetrical across frequencies of 250–8000 Hz. Tympanometric findings also fell within normal limits [admittance (≥ 0.2 ml); tympanometric width (<200 daPa)]. No history of Menieres or neurological disorders was present.

2.2. Stimuli

Similar to our previous study (Friesen et al., 2001), 12 naturally produced CVC tokens in/h/v/d/ context (heed, hid, head, had, who'd, hood, hod, hud, hawed, heard, hoed, and hayed), initially recorded and analyzed by Hillenbrand et al. (1995), were used. Each token was produced by a female talker. The original Hillenbrand stimuli were digitally normalized to the same RMS level. In addition to the naturally produced (unprocessed) stimuli, listeners

were tested using a noise band simulation of implant processing (see Shannon et al., 1995). Acoustic processors were designed with 16, 12, 8, 4, and 2 bands. For processors with up to 8 frequency bands, the total frequency range was partitioned into bands based on the Clarion 1.0 device (Clarion Reference Manual, 1998). Similar to Friesen et al. (2001), for processors with greater than 8 frequency bands, the entire frequency range from 250 Hz to 6.8 kHz was divided into equal parts in terms of cochlear distance in mm using the cochlear tonotopic formula of Greenwood (1990). The envelope was extracted from each band by half-wave rectification and low-pass filtering at 160 Hz. This envelope signal was then used to modulate a wide-band noise and then band-pass filtered with the same filter set used on the original speech signal. The modulated noise bands were then summed and presented through a calibrated loudspeaker in a sound treated room.

2.3. Procedure

Participants were seated in a sound-attenuated booth, one meter in front of a JBL Professional LSR25P Bi-Amplified speaker at 0° azimuth. In all conditions, the stimulus presentation level was 65 dBA. The behavioral and electrophysiologic test sessions were conducted on different days and the order of testing was counterbalanced.

2.4. Behavioral testing

Using a PC computer, each subject identified the 12 medial vowels as they were presented in random order using custom software (Robert, 1997). Each token was presented in text, in alphabetical order, on a computer screen as the response set. To ensure subjects understood the task, they were initially given practice sessions where perception was measured using the 12 unprocessed tokens. One session included each stimulus being presented 5 times. No feedback was provided. Once the participant was able to attain two consecutive scores within 5% of each other, they moved on to the test stage where all six conditions were presented (16, 12, 8, 4, and 2 bands, as well as the unprocessed stimuli). During testing, each of the 12 tokens was presented 20 times, for a total of 240 trials per condition. Percent correct scores were calculated for each individual. Also, for the purpose of comparing perceptual and physiological responses, percent correct scores for the tokens 'hid' and 'heed' were extracted from the overall performance score.

2.5. Electrophysiological testing

Because it is not feasible to record responses for all 12 tokens, as well as their processed conditions, evoked responses were recorded using two randomly selected CVC syllables used during behavioral testing, 'heed' and 'hid'. P1–N1–P2 responses were evoked by each of these two speech tokens in six stimulus conditions (2, 4, 8, 12 and 16 channels, as well as the unprocessed stimuli). The stimulus 'heed' was 490 ms and 'hid' was 433 ms in duration. Participants were instructed to ignore the stimuli and watch a closed-captioned video of their choice. In each condition, the stimulus was presented 300 times in a homogenous sequence. For example, 'hid' simulated through the 2-channel condition was presented 300 times in order to generate a single averaged response for this 2-channel condition. This procedure was repeated for each channel condition. The presentation order of stimulus conditions was randomized across subjects. Each participant was given a 5 min listening break in-between the presentation of different stimulus conditions.

CAEPs were recorded using a 32-channel Neuroscan™ Quik-Cap system. The ground electrode was located on the forehead and the

reference electrode was placed on the nose. Eye blink activity was monitored using an additional channel with electrodes located superiorly and inferiorly to one eye and at the outer canthi of both eyes. Ocular artifacts exceeding $\pm 80 \mu\text{V}$ were rejected from averaging. The recording window consisted of a 100 ms pre-stimulus period and a 1400 ms post-stimulus time. Evoked responses were analog band-pass filtered on-line from 0.15 to 100 Hz (12 dB/octave roll off). Using a Neuroscan™ recording system, all EEG channels were amplified with a gain $\times 500$, and converted using an analog-to-digital rate of 1000 Hz. Following eye-blink rejection, the remaining sweeps were averaged and filtered off-line from 1.0 Hz (high-pass filter, 24 dB/octave) to 20 Hz (low-pass filter, 12 dB/octave). Averaged data files were linear detrended.

Quantifying electrophysiological changes as a function of channel number began with partitioning the stimuli into their constituent phonemes, similar to Ostroff et al. (1998), to identify latency regions that corresponded to transition areas shown in Fig. 2 and expected to generate N1 peak responses. These time windows were then confirmed using butterfly plots, which display evoked activity across the entire scalp, as well as mean global field power measurements (MGFP; Fig. 3). The butterfly plot displays CAEP recordings from all electrode sites superimposed upon one another while MGFP is a measure defined as the standard deviation across

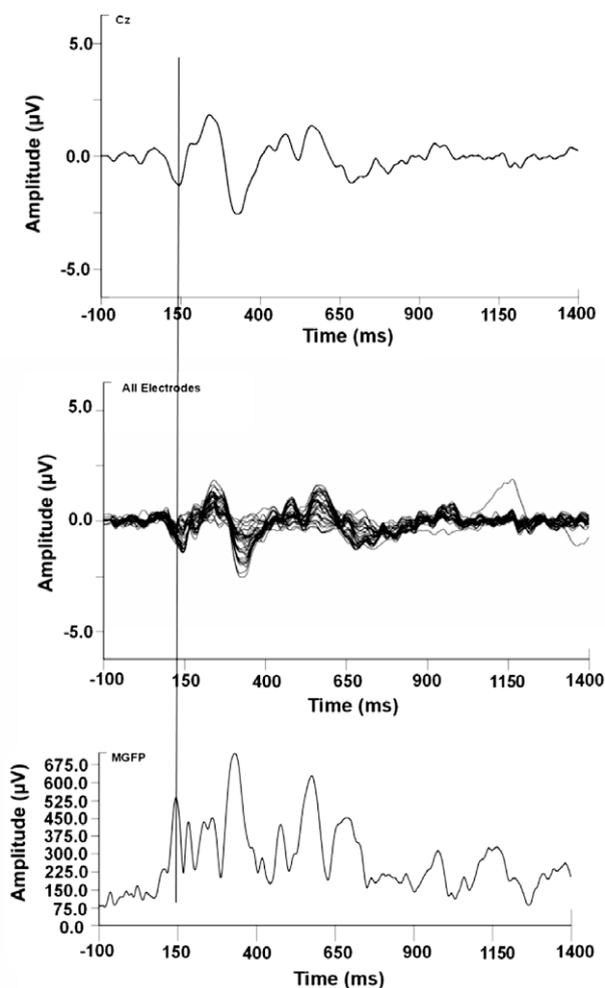


Fig. 3. Group averaged CAEPs to stimulus 'heed'. CAEPs recorded from a group averaged response to the stimulus 'heed'. Recordings from electrode site Cz are shown on top. A butterfly plot of these same data appears in the middle. The mean global field power is shown below. The line connecting all three figures occurs at approximately 150 ms and coincides with the onset of the syllable 'h'.

multiple channels as a function of time within a sample interval (Lehmann and Skrandies, 1980). Because this particular CAEP is optimally recorded from electrode site Cz (Näätänen and Picton, 1987), and this electrode site is typically used to assess implant users in clinical settings, peak responses were analyzed from this electrode.

For each stimulus type ('heed' vs. 'hid'), a separate factorial repeated measures ANOVA was completed (6 channel conditions and 4 peak conditions) where number of channels was the independent variable and peak amplitude or latency was the dependent variable. Greenhouse–Geisser corrections (Greenhouse and Geisser, 1959) were used where an assumption of sphericity was not appropriate. When corrections were used, epsilon (ϵ) values are included in the text.

3. Results

3.1. Behavioral results

Percent correct scores averaged across subjects for the 12 tokens are shown in Fig. 1b. Repeated measures ANOVA results using identification scores for the six different channel number conditions were similar to those of Friesen et al. (2001). Namely, as the number of channels increased, identification of the 12 tokens improved significantly from 21% (2 channels) to 92% correct (unprocessed) [$F(5,45) = 134.064$, $p < 0.001$, $\epsilon = 0.46$]. However, as shown in Fig. 4, the performance functions seen in grouped data do not apply to all speech tokens. Listeners perceived the word 'heed' remarkably well with only 2 channels (66%) of information and performance continued to increase to 100% with additional acoustic information. This result illustrates that the averaged percent correct score, for the 12 tokens, does not represent each individual token. Even though performance is good with only 2 channels, performance still improves as a function of increasing channel number. A repeated measures ANOVA, comparing identification scores for the different channel number conditions, revealed a significant main effect for channel number for each token ['heed' $F(5,45) = 11.857$, $p < 0.001$, $\epsilon = 0.393$] and 'hid' [$F(5,45) = 46.44$, $p < 0.001$, $\epsilon = 0.383$].

3.2. Electrophysiological results

Overlapping P1–N1–P2 responses, representing acoustic transitions, can be recorded in response to CVC syllables. The same two tokens ('heed' and 'hid') described above were used to elicit P1–N1–P2 responses. As can be seen in the group-averaged neural responses to the stimuli 'heed' and 'hid' for the 2-channel and unprocessed condition in Fig. 5, evoked cortical activity was significantly altered by channel number. As expected, there was a significant main effect for channel number with peak amplitude for both stimuli ['heed' $F(5,180) = 2.685$, $p = 0.023$ and 'hid' $F(5,180) = 2.411$, $p = 0.038$], indicating that as the number of channels of acoustic information increases, there is also a significant increase in peak amplitude for the evoked neural responses. The number of CI simulation channels also affects the latency of these peaks. There was a significant effect for peak latency ['heed' $F(5,180) = 3.216$, $p = 0.008$, and 'hid' $F(5,180) = 7.127$, $p < 0.001$, $\epsilon = 0.600$] suggesting that as the number of channels of acoustic information increased, CAEP latencies decreased.

Because the effects of channel number might affect some acoustic transitions more than others, we also examined the interaction between the 4 N1 peaks reflecting the 4 acoustic transitions vs. number of channels (peak \times channel interaction). In short, the only significant peak \times channel interaction identified was for amplitude with the 'heed' stimulus [$F(15, 180) = 2.346$, $p = 0.004$]. There was a

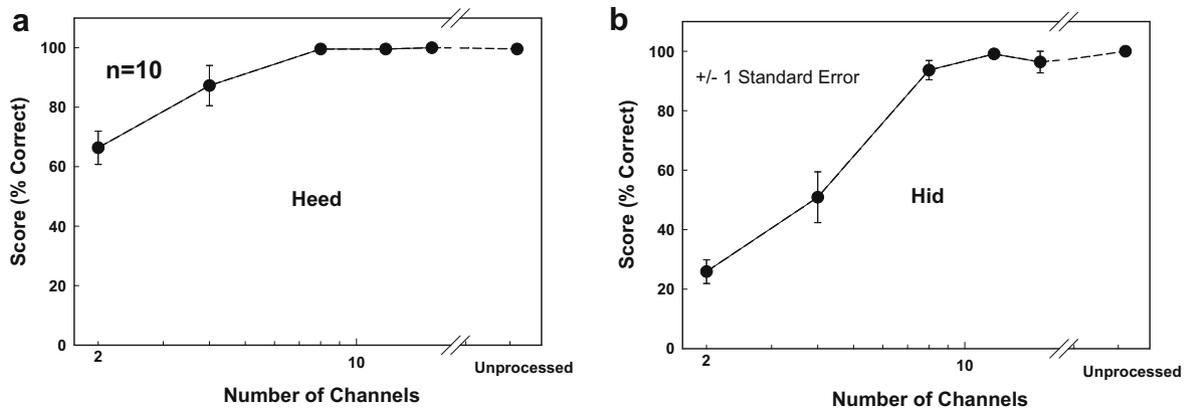


Fig. 4. CVC percent correct scores vs. number of active channels averaged across normal hearing listeners for 'heed' (a) and 'hid' (b). Error bars represent ± 1 standard error.

significant effect for channel on peaks 1 [$F(5,45) = 3.013, p = 0.020$] and 4 [$F(5,45) = 5.862, p < 0.001$], meaning the magnitude (amplitude) of peaks representing stimulus onset and offset were most affected by channel number. No other significant interactions were obtained.

3.3. Comparison of behavioral and electrophysiologic results

To determine whether there was a relationship between behavioral and electrophysiologic results, partial correlations, collapsed across channel and peak, were computed between speech understanding scores vs. latency and amplitude results obtained for 'heed' and 'hid' (Table 1). Percent correct scores were only associated with the amplitude of the CAEP response recorded to 'heed'.

3.4. Summary of results

Acoustic transitions contained within CVC syllables are reflected as multiple, overlapping P1–N1–P2 responses recorded from the surface of the human scalp. With increasing numbers of channels, perception improved and physiological processing of acoustic changes becomes shorter in time (decreased latency) and larger in magnitude (increased amplitude). The growth functions of these two measures (behavior and electrophysiology) were somewhat different, however, and were only associated with the amplitude of the CAEP response recorded to 'heed', suggesting that acoustic transitions alone cannot fully explain the perceptual performance of our listeners.

4. Discussion

Despite the fact that many CI users perform well with their implants, performance variability exists across individuals. For this reason, there is interest in learning more about potential sources of variability and how we can improve speech understanding for those people who are unsatisfied with their device. One approach is to examine the neural detection of speech sounds in individual listeners and study potential brain–behavior relationships that might exist. If it is possible to determine if acoustic cues contained in speech are (or are not) being detected by the brain, it might also be possible to modify implant settings to improve physiological detection patterns and in turn improve perception.

We therefore set out to determine if CVC syllables that are often used to measure speech identification in CI users in CI research could be used to characterize neural detection patterns associated with these same speech sounds. P1–N1–P2 responses have been traditionally recorded using short duration stimuli such as clicks, tones, pulse trains, and these stimuli evoke primarily onset responses. But the P1–N1–P2 response is also sensitive to acoustic changes contained within complex signals such as speech.

Results from this feasibility study provide evidence that the P1–N1–P2 response can be used to record acoustic transitions in CVC syllables. When examining the effects of simulated channel number on perception and physiology, the expected findings were obtained. Similar to the Friesen et al. (2001) study, performance recognition for all 12/h/V/d/ tokens improved with increasing numbers of channels providing more acoustic information. How-

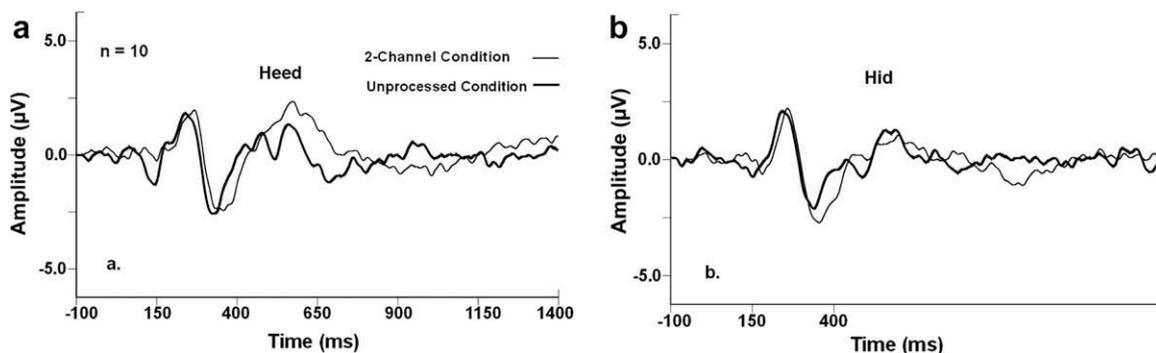


Fig. 5. Representative responses shown at electrode Cz for unprocessed and 2 channel stimuli 'heed' (a) and 'hid' (b). Thin black lines represent 2-channel conditions while thick black lines represent the unprocessed conditions.

Table 1
Correlations between percent score and amplitude and latency.

Heed amplitude vs. perception		Heed latency vs. perception		Hid amplitude vs. perception		Hid latency vs. perception	
<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>	<i>r</i>	<i>p</i>
–0.11	0.04	0.04	0.25	–0.04	0.26	–0.02	0.40

Significant correlations ($p < 0.05$).

ever, the performance results for the individual stimuli ‘heed’ and ‘hid’ appear different from each other. While the behavioral function for ‘hid’ was similar to that of the 12-token function (poor perception with only 2 channels), listeners did remarkably well identifying ‘heed’ with only 2 channels (66% correct). Because of this finding, data from original Friesen et al. (2001) study was re-analyzed by the authors and a similar performance function was found. Therefore, it is important to point out that averaged percent correct scores based on all 12 tokens does not necessarily represent the performance function of individual words.

What might account for the different performance scores (‘heed’ vs. ‘hid’) when only 2 channels of information are available? Vowel identification is dependent on more than just the neural response to acoustic transitions, and we speculate that individuals were likely able to make use of spectral cues present in ‘heed’ (but not ‘hid’) with only two channels of information. One possibility is vowel duration; the duration of the vowel in ‘heed’ (222 ms) was longer than the vowel duration in ‘hid’ (159 ms). Other cues might include the distance between the processes indexed by the first (F1) and second formant (F2). In the stimulus ‘heed’, the process represented by F2 is the highest in frequency of all 12 tokens, while that representing F1 is the lowest. Therefore, it may be that listeners are able to use some spectral information with only 2 channels of information, inherent to one or both of the processes represented by the F1 and F2 spectral peaks, together with vowel duration, to help recognize ‘heed’ from all other tokens. The F1 and F2 distance explanation would be in agreement with results from studies examining vowel recognition in CI listeners (Tyler et al., 1989; Skinner et al., 1996, 1997; van Wieringen and Wouters, 1999). The explanation of using a combination of F1, F2, and vowel duration is in agreement with study results examining information transmitted during vowel recognition in CI listeners (Tyler et al., 1989; Skinner et al., 1996) and normal hearing subjects presented with CI simulations (Xu et al., 2005).

The neural response to acoustic changes, contained in CVC speech tokens, was also affected by channel number. As was hypothesized, the magnitude of evoked neural activity increased (amplitude) while neural conduction time (latency) decreased. Increased amplitudes and decreased latencies likely reflect increased neural synchrony, signaling the response to acoustic changes contained in the speech signal (Fig. 2), with increasing channel number. In general, it can be said that the effects of channel number affect the acoustic representation of the signal. However, there were significant peak \times channel interactions for the amplitude of the CAEP response to ‘heed’, suggesting that the effects of channel number appeared to affect response peaks differently. For example, for the stimulus ‘heed’, N1 responses corresponding to stimulus onset and offset appear to be most affected by the number of channels of acoustic information, possibly conveying stimulus duration information.

Even though changes in CAEP morphology coincided with improved performance (with increasing channel number), latency and amplitude changes do not reflect the perceptual functions shown in Fig. 4a and b. Correlations between amplitude and percent correct were only significant for ‘heed’. Together, these results suggest that the amount of acoustic information available to the

listener affects the physiological representation of sound and perception, but the physiological effects as measured by CAEPs do not fully explain the perceptual performance.

A disconnect between brain and behavior measures is not entirely surprising given that speech contains many different types of acoustic information, and the CAEPs used in this study are primarily sensitive to temporal cues that convey envelope-like information about the signal. A stronger relationship between speech understanding and neural responses might be found if a measure that is sensitive to frequency information is also included. Also, higher level processing is necessary for perception, and separate processing streams have been indicated in auditory association areas in the temporal lobes using positron emission tomography (PET; Scott et al., 2006), but the CAEPs used in this study reflect relatively early, sensory stages of processing, namely the physiological detection of sound. Moreover, the analysis of CAEP patterns was limited to electrode site Cz because there is precedence in the literature to record this response from midline electrodes (Näätänen and Picton, 1987) to evaluate CI users (Martin et al., 2008). However, analyzing patterns of brain activity recorded from other electrode locations might provide additional information and a future goal will be to examine the effects of one evoked response on the others that follow. As shown in Fig. 5 the resultant patterns of activity are complex, containing multiple overlapping positive and negative peaks. It is likely that preceding processes alter subsequent amplitude/latency values by overlapping with positive (or negative) waves as well as by changing the strength of between-sweep synchronization. It will be important to understand these interactions as they might be confounding or informative. For example, some individuals might exhibit stronger, synchronized, responses to the initial onset of sound than others, and this could be advantageous for perception or problematic if it introduces refractory or masking-like effects. It is therefore important to consider alternative ways of defining and quantifying patterns of N1 responses as well as their interactions.

It is also important to keep in mind that top down processes represented by efferent connections also contribute to speech identification (Kumar and Vanaja, 2004; Harkrider and Smith, 2005; Harkrider and Tampas, 2006) and although their functions are not completely understood, it is believed that efferent fibers play an important role in audition including speech understanding, especially with degraded speech. The noise-vocoded speech stimuli used in this experiment were degraded signals that were new to our subjects, and likely stimulated novel detection patterns that required cognitive processes to decipher. In this respect, perception scores reflect top-down cognitive processes such as memory and word closure that are very different from the pre-attentive recording of CAEPs. Taking this point one step further, the context of each stimulus differed during brain and behavior testing. In order to measure a person’s ability to correctly identify a CVC token of interest, a task that includes more than one CVC option is necessary. In the present experiment each person compared the sound they heard to other CVC options. However, it is not feasible to record neural detection patterns from all versions of the stimuli used during perceptual testing so representative samples ‘heed’ and ‘hid’ were used instead. This means the context of ‘heed’ and ‘hid’ were different between CAEP and behavioral testing. In this respect, the context of brain and behavior measures was different from one another which in turn might contribute to some of the differences between brain and behavior results.

A final note is that we intentionally completed the channel study with normal hearing individuals listening to CI acoustic simulations because it enabled us to carefully control and understand stimulus processing schemes on the perception and physiological representation of the test stimuli while minimizing the many confounding variables that come with testing actual CI patients. This process allows us to determine the effects of CI processing on per-

ception in individuals with normal hearing and provides an indication of potential outcomes in individuals with a CI. However, the generalization of performance with CI simulations to performance of actual CI listeners is limited because of the effects of auditory deprivation and electrical stimulation on the implant listener's auditory system. With that said, the present study does suggest that the physiological detection of acoustic changes, at the CVC syllable level, can be recorded in individual subjects and groups. Moreover, the physiological detection of these CVC syllables is modified with the number of channels in a way that coincides with improved perception. Namely, decreased peak latencies and increased amplitudes suggest improved synchronous representation with increasing channel number. Collectively, these results suggest that the ability to make these types of recordings in individual people permits the ability to study the brain–behavior relationships in individuals with impaired speech understanding.

5. Conclusions

In this study, we demonstrate that it is possible to record the physiological detection of acoustic changes contained in CVC syllables in individuals. Moreover, perception and physiological recordings are both affected by the number of channels. This means that the number of active channels should be taken into consideration when describing evoked CAEP activity, and that future studies can be designed around this fact by exploring further effects of speech processor settings on evoked brain activity. Although we used vocoded speech signals to examine neural response patterns in normal hearing listeners, our results suggest that this type of technique might be useful for examining the physiological detection of acoustic transitions in people who have CIs. Although not yet suitable for clinical settings, a goal will be to develop similar techniques that could be used in clinical settings to measure the neural representation of acoustic cues in persons unable to participate in traditional behavioral tests, and also changes in neural activity following implantation.

Acknowledgments

The authors thank the subjects who participated in this study for their time and effort. We also thank Kate McClanahan and Katie Faulkner for their help with testing subjects. This work was supported by a personnel training Grant (T32 DC00033) from the National Institutes of Health, and the University of Washington, Department of Speech and Hearing Sciences, an American Academy of Audiology student research grant, and the ASH Foundation that provided support for L. Friesen. We also acknowledge funding from the National Organization for Hearing Research (NOHR), National Institutes of Health Grant No. NIDCD 0007705, and National Institutes of Health Grant No. P30 DC04661. Portions of these data were presented at the Association for Research in Otolaryngology and Conference on Implantable Auditory Prostheses meetings in 2005 and at the International Evoked Response Audiometry Study Group in 2007.

References

- Clarion by advanced bionics: SCLIN 98 for windows device fitting manual. Sylmar, California; 1998.
- Friesen LM, Tremblay KT. Acoustic change complexes (ACC) recorded in adult cochlear implant listeners. *Ear Hear* 2006;27(6):678–85.
- Friesen LM, Shannon RV, Baskent D, Xiaosong W. Speech recognition in noise as a function of the number of spectral channels: comparison of acoustic hearing and cochlear implants. *J Acoust Soc Am* 2001;110(2):1150.
- Fu Q, Galvin JJ. The effects of short-term training for spectrally mismatched noise-band speech. *J Acoust Soc Am* 1990;113(2):1065–72.
- Greenhouse WW, Geisser S. On methods in the analysis of profile data. *Psychometrika* 1959;24:95–112.
- Greenwood DD. A cochlear frequency-position function for several species – 29 years later. *J Acoust Soc Am* 1990;87:2592–605.
- Harkrider AW, Smith SB. Acceptable noise level, phoneme recognition in noise, and measures of auditory efferent activity. *J Am Acad Audiol* 2005;16(8):530–45.
- Harkrider AW, Tampas JW. Differences in responses from the cochleae and central nervous systems of females with low versus high acceptable noise levels. *J Am Acad Audiol* 2006;17(9):667–76.
- Hillenbrand J, Getty L, Clark M, Wheeler K. Acoustic characteristics of American English vowels. *J Acoust Soc Am* 1995;97:3099–111.
- Kaukoranta E, Hari R, Lounasamaa O. Responses of the human auditory cortex to vowel onset after fricative consonants. *Exp Brain Res* 1987;69(1):19–23.
- Kumar UA, Vanaja CS. Functioning of olivocochlear bundle and speech perception in noise. *Ear Hear* 2004;25(2):142–6.
- Lehmann D, Skrandies W. Reference-free identification of components of checkerboard-evoked multichannel potential fields. *Electroencephalogr Clin Neurophysiol* 1980;48:609–21.
- Loizou P. Signal-processing techniques for cochlear implants. *IEEE EMB* 1999;18:34–46.
- Martin BA, Boothroyd A. Cortical, auditory, event-related potentials in response to periodic and aperiodic stimuli with the same spectral envelope. *Ear Hear* 1999;20(1):33–44.
- Martin BA, Boothroyd A. Cortical, auditory, evoked potentials in response to changes of spectrum and amplitude. *J Acoust Soc Am* 2000;107(4):2155–61.
- Martin BA, Tremblay KL, Korczak P. Speech evoked potentials: from the laboratory to the clinic. *Ear Hear* 2008;29:285–313.
- McCandless GA, Rose DE. Evoked cortical responses to stimulus change. *J Speech Hearing* 1970;13:624–34.
- Naatanen R, Picton T. The N1 wave of the human electric and magnetic response to sound: a review and analysis of the component structure. *Psychophysiology* 1987;24:375–425.
- Ostroff J, Martin B, Boothroyd A. Cortical evoked responses to spectral change within a syllable. *Ear Hear* 1998;19(4):290–7.
- Polen SB. Auditory event related potentials. *Sem Hear* 1984;5(2):127–41.
- Robert ME. AIPSS-ID – Phoneme Identification Software. Los Angeles: House EarInstitute; 1997.
- Scott SK, Rosen S, Lang H, Wise RJ. Neural correlates of intelligibility in speech investigated with noise vocoded speech – a positron emission tomography study. *J Acoust Soc Am* 2006;120(2):1075–83.
- Shannon RV, Zeng FG, Kamath V, Wygonski J, Ekelid M. Speech recognition with primarily temporal cues. *Science* 1995;270(5234):303–4.
- Skinner MW, Fourakis MS, Holden TA, Holden LK, Demorest ME. Identification of speech by cochlear implant recipients with the multipeak (MPEAK) and spectral peak (SPEAK) speech coding strategies. I. Vowels. *Ear Hear* 1996;17:182–97.
- Skinner MW, Holden LK, Holden TA, Demorest ME, Fourakis MS. Speech recognition at simulated soft, conversational, and raised-to-loud vocal efforts by adults with cochlear implants. *J Acoust Soc Am* 1997;101:3766–82.
- Tremblay KL, Friesen L, Martin BA, Wright R. Test–retest reliability of cortical evoked potentials using naturally produced speech sounds. *Ear Hear* 2003;24(3):225–32.
- Tyler RS, Tye-Murray N, Otto SR. The recognition of vowels differing by a single formant by cochlear-implant subjects. *J Acoust Soc Am* 1989;86:2107–12.
- van Wieringen A, Wouters J. Natural vowel and consonant recognition by Laura cochlear implantees. *Ear Hear* 1999;20:89–103.
- Xu L, Thompson CS, Pfingst BE. Relative contributions of spectral and temporal cues for phoneme recognition. *J Acoust Soc Am* 2005;117:3255–67.