Effects of electrode separation between speech and noise signals on consonant identification in cochlear implants^{a)}

Bom Jun Kwon^{b)}

Department of Communication Sciences and Disorders, University of Utah, 390 S 1530 E, Salt Lake City, Utah 84112

(Received 18 October 2007; revised 31 August 2009; accepted 25 September 2009)

The aim of the present study was to examine cochlear implant (CI) users' perceptual segregation of speech from background noise with differing degrees of electrode separation between speech and noise. Eleven users of the nucleus CI system were tested on consonant identification using an experimental processing scheme called "multi-stream processing" in which speech and noise stimuli were processed separately and interleaved. Speech was presented to either ten (every other electrode) or six electrodes (every fourth electrode). Noise was routed to either the same (the "overlapped" condition) or a different set of electrodes (the "interlaced" condition), where speech and noise electrodes were separated by one- and two-electrode spacings for ten- and six-electrode presentations, respectively. Results indicated a small but significant improvement in consonant recognition (5%–10%) in the interlaced condition with a two-electrode spacing (approximately 1.1 mm) in two subjects. It appears that the results were influenced by peripheral channel interactions, partially accounting for individual variability. Although the overall effect was small and observed from a small number of subjects, the present study demonstrated that CI users' performance on segregating the target from the background might be improved if these sounds were presented with sufficient peripheral separation. © 2009 Acoustical Society of America. [DOI: 10.1121/1.3257200]

PACS number(s): 43.71.Ky, 43.66.Ts, 43.71.Es, 43.71.Gv [KWG]

Pages: 3258–3267

I. INTRODUCTION

Everyday sounds typically consist of several spectral components in a wide frequency range. Grouping these components is one of the first steps necessary to capture the meaning of an auditory stimulus. Two primary factors in auditory grouping are the harmonic relations and the resolvability of components by the auditory filters (e.g., Moore *et al.*, 1986) and the temporal arrangement of components (such as synchronous onset; Bregman, 1990). Grouping is particularly important when multiple sound stimuli exist; in the context of speech recognition, decoding the fundamental frequency (F0) of each sound and/or tracking the contour of F0 facilitates segregation of one from the other and improves speech recognition in background sounds (Assmann and Summerfield, 1990; Brox and Nooteboom, 1982). In addition, amplitude fluctuations of the background also aid speech recognition through masking release (i.e., "listening-in-the-valley;" e.g., Bacon *et al.*, 1998). These mechanisms of grouping/segregation in the context of *auditory scene analysis* (Bregman, 1990) partially account for the robust speech understanding in noisy environments enjoyed by normal hearing (NH) listeners compared to the less impressive results obtained from automatic speech recognition by ma-

chine based on pattern-matching algorithms (Allen, 1994).

Speech understanding by individuals implanted with cochlear implants (CIs) is often comparable to that of NH listeners in quiet listening environments. However, CI users' performance sharply drops compared to that of NH listeners if there is a noise or a competing background masker (Skinner *et al.*, 1994; Battmer *et al.*, 1997; Dorman *et al.*, 1998a; Fu *et al.*, 1998; Fetterman and Domico, 2002). For over a decade there have been a myriad of studies of speech recognition in noise by CI users or by NH listeners through CI simulations under a variety of conditions (Nelson *et al.*, 2003; Stickney *et al.*, 2004, 2007; Fu and Nogaki, 2005; Cullington and Zeng, 2008). The findings in the literature can be summarized as follows. First, compared to NH listeners, CI users show a reduced ability to segregate the target voice from a background voice based on F0 differences such as voice gender difference (Stickney *et al.*, 2004, 2007). Second, CI users may not experience or utilize masking release; in contrast, the opposite effect, modulation interference, can sometimes occur (i.e., poorer speech recognition in a modulated background; Nelson *et al.*, 2003; Stickney *et al.*, 2004). In short, the ability of CI users to segregate the target voice from the background based on the cues known in NH listening has not been demonstrated. However, several observations suggest the possibility that CI users may still use auditory grouping or segregation to process sounds. Fu and Galvin (2001) showed similar recognition of spectrally desynchronized sentences by CI users and NH listeners, implying that CI and NH listeners might utilize common temporal grouping mechanisms. In addition, recent studies have suggested that CI users may use stream segregation to process

^{a)} A portion of the pilot data for this study was presented in poster sessions at the 149th Meeting of the Acoustical Society of America, Vancouver, Canada, May, 2005 and at the Conference on Implantable Auditory Prostheses, Pacific Grove, CA, July, 2005.

^{b)} Author to whom correspondence should be addressed. Present address: Department of Otolaryngology-Head and Neck Surgery, Eye and Ear Institute, Ohio State University, 915 Olentangy River Rd., Columbus, OH 43221. Electronic mail: bjkwon@gmail.com

non-speech stimuli (Chatterjee *et al.*, 2006; Hong and Turner, 2006; Carlyon *et al.*, 2007). Interestingly, in these studies the separation between stimulating electrodes influenced the results in a way that suggests that stream segregation in CI users, like in NH listeners, appears to be affected by peripheral channeling (Hartmann and Johnson, 1991; Beauvois and Meddis, 1996), although the mechanisms underlying these observations may still be a matter of debate (see Cooper and Roberts, 2007).

If CI users can indeed perform auditory grouping or segregation, the underlying cause behind CI users' poor understanding of speech in noise deserves further attention. To this end, it is worth noting that, in commonplace CI settings multiple signals where are acoustically mixed and processed by a set of conventional clinical parameters, the issues of auditory grouping/segregation are not directly addressed; in other words, potentially important attributes of CI users' abilities are not revealed due to the limitations in conventional signal processing. For example, while it is reasonably well accepted that CI users are unable to segregate a male target voice from a female voice background (Stickney *et al.*, 2004, 2007), this effect might be due to the lack of resolvability in the filterbank of the CI device rather than an inability to segregate *per se*. The filter slopes in the filterbank used for CI speech processing are generally shallow so that when the input is a pure tone, actual stimulation takes place on multiple channels. For example, as can be easily seen in clinical programming equipment and/or software in the nucleus system, three or four electrodes are activated even for a pure tone input. Thus, CI users are not provided with the signal with a sufficient resolvability. In addition to the issue of filter resolution, the degradation of signal quality due to only limited spectral channels imposes a further challenge to CI users. Speech is represented in CIs by modulated pulse trains delivered to a small number of electrodes; fine structure cues are not available in each channel. In acoustic hearing, when noise is added to speech, the auditory periphery is presented with the mixture of the two signals; it is the task of the auditory system to parse the mixture and segregate one from the other with the available cues mentioned above. In the electric representation of speech in CIs, however, the consequence of noise is not just an addition of a distracting or masking signal, but rather a severe distortion or alteration in the pattern of speech pulses. This distortion involves a high compression in loudness mapping between the acoustic input and its electric output, which is necessary to compensate for a substantially reduced electric dynamic range (Nelson *et al.*, 1996). Such distortion is usually nonlinear and irrecoverable, yet another reason noise negatively impacts speech intelligibility more through a CI system than in acoustic hearing. In summary, due to the technical limitations in clinical CI systems, examination of speech understanding by CI users in competing backgrounds through conventional processing mechanisms does not provide sufficient insight with regard to auditory grouping or segregation in electric hearing.

Given this background, the present study aimed to examine CI users' ability to understand speech in the noise background with a setting where both signals are presented in their entirety without the limitations imposed in conven-

FIG. 1. A block diagram of multi-stream processing with an example of ten-channel CIS processing. Two CIS blocks, one for speech and the other for noise, are processed in parallel and interleaved. For the “interlaced” condition, the speech and noise sequences are delivered on different channel sets (odd-even). For the “overlapped” condition, speech and noise are delivered on the same channel sets (odd-odd). Multi-stream processing allows each of the speech and noise signals to be delivered entirely at the auditory periphery without the limitations imposed by signal processing in conventional CI systems.

tional clinical processing such as resolvability in the filterbank or non-linear compression. In order to achieve this, as shown in Fig. 1, the speech and noise signals were processed separately. Then two separate pulse trains, the output of processing, could be simply interleaved in time, i.e., the speech and noise signals would be presented as two interleaved signals at the periphery (the electrode-nerve interface) rather than one distorted signal as in conventional processing. In the present study, this experimental method is referred to as “multi-stream processing” (Kwon and van den Honert, 2006). Admittedly this method might have only limited practical implications with the current device architecture, as the presumption that each of the multiple input signals is available separately for parallel processing is often violated in real applications. However, at least for a laboratory testing, the signal presentation through multi-stream processing provides an advantage when studying CI users' ability to organize multiple inputs (e.g., Deeks and Carlyon, 2004). The effects of peripheral separation between speech and noise electrodes were of particular interest in the present study and examined by comparing the following two-electrode assignment layouts: “overlapped” and “interlaced,” akin to the “same” and “different” conditions in Deeks and Carlyon (2004). In the overlapped layout the speech and noise pulses were presented on the same set of electrodes, whereas in the interlaced layout they were presented on different sets of electrodes. In both layouts, speech and noise pulses were interleaved in time, and the only difference was how the electrodes were assigned. Therefore, if speech understanding improves in the interlaced layout compared to the overlapped layout, it would indicate that CI users would be potentially capable of performing segregation of multiple signals (such as segregating voices with F0 cues) if those signals were separately delivered to the auditory periphery. If not, it would mean that the peripheral manipulation of the separate presentation of speech and noise is not capitalized upon by CI users, implying that any applications built upon an expectation of improved segregation via separately controlled use of electrodes would be only minimal. The present study is

TABLE I. Subjects' demographic information, history, devices used, CNC word recognition scores in quiet, and clinical MAP parameters. Subjects S1–S6 were tested with the ten-channel CIS, whereas S7–S11 were tested with the six-channel CIS.

ID	Gender	Age at testing (years)	CI experience at testing (years)	Device implanted (ST: straight electrode array; C: contour array)	Etiology	CNC word recognition score (%)	Channel stimulation rate in clinical MAP (Hz)
S1	M	72	6	CI24M-ST	Noise exposure	46	900
S2	M	55	3	CI24R-ST	Otosclerosis	63	900
S3	M	74	4	CI24M-ST	Familial	54	250
S4	M	60	1	CI24R-C	Hereditary	59	1200
S5	M	69	1	CI24RE-C	Noise exposure	34	900
S6	F	60	1	CI24RE-C	Unknown	42	720
S7	M	76	4	CI24R-C	Unknown/progressive	35	1200
S8	F	49	3	CI24R-C	Unknown	38	250
S9	F	81	3	CI24R-C	Ototoxic drug	75	900
S10	F	52	1	CI24RE-C	Ototoxic drug	52	720
S11	F	56	1	CI24RE-C	Hereditary	45	1200

not a feasibility test of a strategy based on multi-stream processing, as it does not address any issue of source separation algorithm (e.g., Kokkinakis and Loizou, 2008) that should be in place in real implementation. Yet, the findings of the present study might provide us the insight of whether such algorithm would be worthwhile to pursue.

II. METHODS AND PROCEDURE

A. Background in CI signal processing and terminology

The continuous interleaved sampling (CIS) strategy (Wilson *et al.*, 1991) consists of the following steps: (i) a spectral analysis of the input signal is done by a series of band-pass filters and the signal envelope (or the momentary fluctuation of spectral energy) in each band is obtained, (ii) a pulse train at a constant repetition rate is modulated by the envelope in each band (with a nonlinear loudness-mapping function), (iii) the pulse trains of all bands are temporally interleaved to avoid simultaneous current field summation, and (iv) the resulting pulse train is delivered to pre-allocated electrodes with the same number of processing bands. While the CIS strategy is an option that is available from all three manufacturers of widely used CI systems, the strategy is implemented in a slightly different way in each brand, emphasizing different aspects of processing due to unique hardware/electrode features.

Before proceeding to the description of the details of signal processing for the present study, some definition of terminologies would be beneficial to avoid confusion. First, the term “channel” in the present paper is mostly used to refer to the processing band of acoustic input, i.e., the ordinal division of discrete frequency regions, or *acoustic* channel. This is in contrast to *electric* channel, another usage of the same word channel in the literature, which refers to the pathway of the current flow (i.e., electric channel) created by the chosen stimulation configuration in the CI system (e.g., monopolar and bipolar), as established by the ordinal division of electrodes. Electric channel is referred to as “electrode” in the present paper, which is proper because in monopolar stimulation each electric channel is implemented by one electrode. This clarification is particularly important in the

present paper because the information processed in each (acoustic) channel is to be delivered to a specifically defined set of electrodes. Whereas the concept of electric channel is directly bound with the use of electrodes, the binding is not as strong for an acoustic channel: for example, the lowest frequency band is not necessarily mapped with the apical-most electrode for various reasons in clinical applications, and relatively unique manipulation for the frequency-to-electrode binding is attempted in the present study. Second, a MAP is a clinical term (note that it is not an acronym) that describes a complete set of parameters for signal processing including the information on not only the basic pulse parameters such as threshold, stimulation rate, and pulsewidth but also the usage of channels such as the total number of acoustic processing bands, the (de)selection of electrodes, and the loudness-mapping function. Third, the total stimulation rate of a MAP is defined by the stimulation rate per channel multiplied by the number of channels delivered to represent the signal in each analysis time window.

B. Subjects

A total of 11 adult CI users participated in the present study. They had been implanted with the Nucleus[®] CI24M, CI24R, or CI24RE (known as “freedom” implant) device at least 1 year prior to entry into this study. The study was conducted at the laboratories of Cochlear Americas (Englewood, CO) and at the University of Utah (Salt Lake City, UT). The use of human subjects in this study was reviewed and approved by the Western Institutional Review Board and the University of Utah IRB, respectively. Table I provides subject-specific information, including patient history, the implanted device, the channel stimulation rate used in each subject’s clinical MAP, and the consonant-nucleus-consonant (CNC) monosyllabic word recognition scores in quiet (Peterson and Lehiste, 1962) measured with the clinical MAP. Six subjects were tested in the ten-channel CIS condition and subsequently five subjects were tested in the six-channel CIS condition (see Sec. II C).

FIG. 2. Electrode layout for the “overlapped” and “interlaced” conditions for each CIS processing condition.

C. Signal processing and experimental conditions

Throughout the present study, many processing parameters were chosen from the default settings in clinical applications. For example, the monopolar stimulation configuration was used, the phasewidth was 25 μ s, and the interphase gap was 8 μ s. Experimental conditions were designed to create varying degrees of speech-noise separation on electrodes. Given 22 intra-cochlear electrodes in the nucleus system, the following two conditions of speech presentation were chosen: ten-channel CIS (using every other electrode: E21, E19, E17, E15, E13, E11, E9, E7, E5, and E3) and six-channel CIS (using every fourth electrode: E21, E17, E13, E9, E5, and E1). Noise was presented either on the same electrodes as speech in the overlapped condition or different electrodes in the interlaced condition. Therefore, in the interlaced condition, speech and noise electrodes were separated by one-electrode spacing in the ten-channel CIS or by two-electrode spacing in the six-channel CIS, as shown in Fig. 2. It was not possible to use six electrodes for noise in the interlaced condition with a two-electrode spacing margin (note that the number of electrodes is 22, not 24). Therefore, for a fair comparison between the overlapped and interlaced conditions, only five electrodes were used for noise; these electrodes were either E21, E17, E13, E9, and E5 (for the overlapped condition) or E19, E15, E11, E7, and E3 (for the interlaced condition). Note that this configuration might slightly favor the overlapped condition because if the channel spread were wider than two-electrode spacing (but not as wide as four-electrode spacing), the speech information from E1 would be masked in the interlaced, but not in the overlapped condition. Therefore, if speech understanding were substantially better in the interlaced condition, this comparison could serve as evidence, suggesting CI users’ capability of perceptual segregation.

In the present study the total stimulation rate was limited to 14 400 Hz due to the hardware/software restrictions in the nucleus system. For example, if the channel stimulation rate was 900 Hz, then up to 16 channels could be used for a MAP ($900 \times n \leq 14\,400$ for $n \leq 16$). Ideally, to maintain the usual speech intelligibility and quality, the same stimulation rate as in the clinical MAP should have been used in all conditions. Unfortunately, this was not possible in some subjects for whom the total stimulation rate in their clinical MAP exceeded 7200 Hz (S4, S7, and S10), one-half of the limit, because the total bandwidth, 14 400 Hz, was shared by the interleaved speech and noise streams. For these subjects, all of whom were tested with ten-channel CIS, a channel stimulation rate of 720 Hz was chosen to make the total stimulation rate of 7200 Hz. Although those subjects noticed some speech quality difference with this rate, subjective speech intelligibility was not particularly degraded. Table II displays the frequency boundaries for processing bands. Frequency boundaries of noise processing in the six-channel CIS condition were different (not shown) from those of speech processing because noise was actually processed with the five-channel CIS, as described above. It is unlikely that the use of different acoustic frequency boundaries for noise in this condition undermined the rationale of the present study because the interaction between the stimuli in the acoustic domain was intentionally avoided and the aim was to compare two conditions manipulating the stimulating electrodes of speech and noise pulses (overlapped or interlaced).

D. Apparatus

Custom-made software written in C++ and MATLAB was used to process the speech material, generate the stimuli, and administer the testing. The software was developed with the NICstream library (Cochlear Ltd., Sydney, Australia), which

TABLE II. Nominal frequency boundaries for the filterbank used for the ten-channel and six-channel CISs.

Frequency boundaries (Hz)											
Ten-channel	188	438	688	1063	1438	1938	2563	3438	4563	6063	7938
Six-channel	188	563	1063	1813	2938	4813	7938				

included signal processing routines used for the nucleus implant system in the form of MATLAB function files (“m” files) and a series of dynamic link library (DLL) files for communication between a personal computer (PC) and the implant device. The default m files were modified and restructured to implement the multi-stream processing scheme for the present study and were converted to a DLL through MATLAB COMPILER (version 4.7); this DLL was used as the main software that administered the experiment. The pulse trains, generated from speech and noise wave files by the PC software, were transmitted to the subject’s internal device transcutaneously through a radio-frequency (rf) signal via one of the following two interfaces: (1) the nucleus clinical programming system with the SPrint speech processor used for a desktop PC or (2) a USB-based POD interface with the L34 speech processor used for a laptop. In both cases SPrint and L34 were used as relays that generated a rf signal from the PC rather than actual processors; neither the microphone connected to the implant nor the processing capability of the implanted processor was used. The speech input level was adjusted at the front-end stage so that the digitized waveform samples would be emulating an acoustic input with 70 dB sound pressure level rms with the sensitivity setting at 8 to ensure sufficient loudness in all conditions. The subject was given a separate PC for making responses and receiving feedback. This PC was connected via the TCP/IP protocol to the operator’s PC, which performed signal processing and test administration.

E. Stimuli and procedure

Recognition of the following consonants was measured: b, d, g, p, t, k, z, s, m, n, f, v, θ , δ , β , and $\int$. The material was recorded in the /aCa/ context by four talkers—two males and two females—and a total of 64 (i.e., 16×4) sound samples (Turner *et al.*, 1995) was used. Each testing session consisted of 128 presentations (two presentations of each sound sample) in a pseudo-random order. The subject was given 16 choices of consonants on a computer screen after each presentation and was asked to respond by a mouse or a finger touch on a touch screen. The average percent-correct score was obtained for each subject after repeating the session at least four times. Feedback was given only prior to data collection during the practice session. Speech-shaped noise was generated by the following method. The long-term spectrum of the aggregate sound file (concatenation of the 64 sound files) was analyzed in third-octave bands in the range 100–8064 Hz. A narrowband noise was generated by filtering a white noise with a bandpass filter for each of the third-octave bands in this range, and the level of each narrowband noise was adjusted to match the relative magnitude of the long-term spectrum of speech materials. The narrowband noise in each band was then summed to generate the speech-shaped noise. Finally, the spectrum of this noise was compared again with the long-term spectrum of speech sounds to ensure that a peculiar summation of noise at transition bands did not occur.

For the ten-channel CIS, noise was combined with speech so that the onset of the noise signal preceded that of

FIG. 3. The percent consonant identification scores of each subject for the “overlapped” (dark bars) and “interlaced” (light bars) conditions in ten-channel CIS. The interlaced condition represented the separation between the speech and noise electrodes by one-electrode spacing. The tested SNR is indicated for each subject. The scores marked with “Q” indicate the scores in quiet. Error bars indicate one standard deviation.

the speech signal by a lead time randomly determined in each presentation by a uniform distribution between 0 and 700 ms. For the six-channel CIS, the lead time of the noise signal was fixed at 1 s. The noise level, or the signal-to-noise ratio (SNR), was set such that the identification score reasonably dropped relative to the quiet condition.¹ As the effect of noise on speech understanding is usually variable across subjects, the SNR was selected between 5 and 20 dB for each subject, instead of testing at a fixed SNR. Six subjects were tested in the ten-channel CIS condition and five different subjects were tested in the six-channel CIS condition. In both conditions, the order of overlapped versus interlaced testing sessions was randomized. There was an additional test with two subjects using three different levels of noise after the initial data collection (the rationale and procedure are described in Sec. III B).

III. RESULTS

A. Effect of electrode separation

The results of consonant identification in the ten- and six-channel CIS conditions are plotted in Figs. 3 and 4, respectively. The percent-correct identification scores of the overlapped (dark bars) and interlaced (light bars) conditions

FIG. 4. The percent consonant identification scores of each subject for the “overlapped” (dark bars) and “interlaced” (light bars) conditions in six-channel CIS. The interlaced condition represented the separation between the speech and noise electrodes by two-electrode spacing. The tested SNR is indicated for each subject. The scores marked with “Q” indicate the scores in quiet. Error bars indicate one standard deviation.

FIG. 5. The mean difference in scores between the “overlapped” and “interlaced” conditions for each electrode separation. Error bars indicate one standard deviation.

are plotted for each subject. Percent-correct scores in a quiet background with the ten-channel CIS ranged from 29% to 68% (mean: 49.2%), and with the six-channel CIS were from 25% to 60% (mean: 45.2%). The observed decrease in the quiet score is generally consistent with previous studies (e.g., Friesen *et al.*, 2001). The effect of noise was variable across subjects: while some subjects showed a relatively higher resilience to noise at 5 dB SNR (e.g., S4), other subjects suffered greatly from an even smaller amount of noise (e.g., S7 at 20 dB SNR). For the ten-channel CIS, the difference in consonant identification scores was negligible between the overlapped and interlaced conditions (0.1% on average; cf. Fig. 5). In this case, the speech and noise signals were presented with a spacing of one electrode in the interlaced condition. In contrast, the difference was greater between the two conditions in the six-channel CIS, in which speech and noise were separated by a two-electrode spacing in the interlaced condition. Using a mixed-ANOVA model with one within-subject factor, the electrode layout (overlapped or interlaced), and two between-subject factors (the two CIS conditions and SNR), a significant interaction was found between the CIS conditions and electrode layout ($p=0.003$), suggesting that the two CIS conditions be examined separately. The effect of SNR was not significant ($p=0.055$). As displayed in Fig. 5, the average difference between the overlapped and interlaced conditions for six-channel CIS, after collapsing across SNR conditions, was 4.7%: this was statistically significant in a paired-t test ($p=0.012$). Thus, the subjects tested with the six-channel CIS speech (S7–S11) on average performed better in the interlaced condition than in the overlapped condition, indicating improved perceptual segregation with the separated presentation. However, due to the large inter-subject variability, the small number of subjects for this analysis ($N=5$), and the small overall effect (4.7%), caution should be exercised when interpreting this statistical analysis. While the statistical trend of an effect of segregation is revealed in the observed group, it is difficult to judge the validity of the effect on an individual basis (except for S7 showing almost no effect). The statistical analysis still suggests that the effects of electrode separation and perceptual segregation be considered in further analyses.

FIG. 6. Additional consonant recognition test for two subjects (S8 and S9) at three SNR levels. Error bars indicate one standard deviation.

B. Additional experiment

Given that a large inter-subject variability with regard to the observed effect exists, as seen in Fig. 4, which is not unusual in CI studies, perhaps it would be sensible to examine the effect in each individual rather than trying to establish a universal trend. The effect of electrode separation appeared to be almost non-existent in S7, for example, who showed the lowest score without noise and suffered from noise in general most severely (note that the SNR tested was 20 dB). While it is unclear whether the lack of effect was due to the compromise in the peripheral nerve system (such as uneven or simply poor neural survival) or in a more central mechanism in this subject, it is apparent that any potential effect was likely overshadowed by the factor(s) causing poor general performance. Therefore, it would be appropriate to reason that the effect of peripheral separation in the context of speech understanding in noise might not be observed in all CI users. For this reason, instead of trying to demonstrate the effect in a larger number of subjects, it was worthwhile to strengthen the statistical power of the within-subject effect by further testing those who had shown a strong effect of separation. Particularly it was necessary to test whether the effect shown in the main experiment was due to the particular choice of SNR for each subject. Therefore, two subjects (S8 and S9) were called back for post-testing sessions 7 or 8 months after the main experiment. The same speech material in the main experiment was used but only two relatively clear voices were chosen (one male and one female). There was a total of six listening conditions (three SNRs and two layouts: overlapped or interlaced). In order to prevent an order effect from influencing the results, all six conditions were mixed into one session consisting of 192 stimuli (16 stimuli for each of the two voices times 6 conditions). This session was repeated three times to obtain the final results.

Results are displayed in Fig. 6 (triangles and circles indicate the interlaced and overlapped conditions, respectively). In both subjects, the effect of electrode separation, as indicated by better performance in the interlaced condition than in the overlapped condition, was observed at all three SNRs; the difference in score between the overlapped and interlaced conditions was 5.2% or more. A two-way ANOVA tested the effect of separation and SNR as fixed factors with subject as a random factor. Both effects were significant ($p < 0.01$), and the effect of subject was not ($p=0.177$). None

TABLE III. Relative information transmitted for each consonant feature (Miller and Nicely, 1955) calculated from aggregate presentation-response matrices across several SNR conditions for each subject.

Condition	S8			S9		
	Voicing	Manner	Place	Voicing	Manner	Place
Quiet	0.59	0.56	0.07	0.72	0.88	0.18
Overlapped	0.1	0.15	0.04	0.31	0.35	0.04
Interlaced	0.45	0.35	0.05	0.68	0.7	0.1

of the interactions were significant. Therefore, at least in these two subjects, the effect of electrode separation was significant regardless of the SNR. The amount of relative information transmitted for voicing, manner, and place features was calculated using feature transmission analysis (Miller and Nicely, 1955). Table III shows these calculations made from aggregate presentation-response matrices in all three SNR conditions for each subject. This analysis reveals that subject performance shows substantial differences between the overlapped and interlaced conditions in voicing and manner features, but not in the place feature. Even in the quiet condition, the transmission of place information was poor, as expected from previous CIS processing studies (Friesen *et al.*, 2001; Donaldson and Kreft, 2006). Therefore, it was not surprising to observe a weak effect of speech-noise electrode separation on the transmission of place information. On the other hand, voicing and manner features are relatively better utilized by CI users in the quiet condition (Donaldson and Kreft, 2006). It was notable that while a substantial reduction in information about these features was seen in the overlapped condition, similar to what was observed in previous studies based on conventional processing (Friesen *et al.*, 2001), the effect of interlaced noise on these features was overall not as severe as had been previously shown. This contrast provides additional evidence of perceptual segregation from electrode separation in these subjects. While the differences in overall percent-correct consonant identification scores are modest (from 5.7% to 9.9%), the effect of electrode separation is further confirmed by the fact that the identification errors made during incorrect trials were still within the same voicing and manner categories more often in the interlaced condition.

IV. DISCUSSION

A. Electrode spacing

The electrode spacing between two center points of adjacent electrodes in the banded (straight) nucleus electrode array is 0.75 mm and is constant across electrodes. On the other hand, the electrode spacing in the nucleus contour array, which is pre-curved toward the modiolus, or the peripheral process of the auditory nerve is not uniform across the array. It is 0.81 mm at the basal end and becomes smaller toward the apical end, where it is 0.4 mm for the three apical-most electrode pairs; the average of 21 electrode-to-electrode distances is 0.56 mm (van den Honert, *personal communication*). The five subjects tested with six-electrode CIS who demonstrated the effect of perceptual segregation with two-electrode spacing as a group were implanted with

the contour array (cf. Table I). Thus, an average electrode distance of 1.12 mm might be considered to be the separation required for perceptual segregation between speech and noise. Although not tested in the present study, a condition testing two-electrode spacing in the straight electrode array (1.5 mm) might have made it easier to demonstrate the beneficial effect of electrode separation. Furthermore, experiments with other CI brands such as the HiFocus electrode array from Clarion II, for which the electrode spacing is 1.1 mm, might provide additional insights regarding electrode spacing and perceptual segregation. There is a caveat, however, in comparing or projecting the results from different electrode arrays because each electrode array has a unique curvature and the electrodes are placed at a different distance from the modiolus. For example, in the straight array, which does not curl inward toward the modiolus, apical electrodes are farther away from the modiolus than basal electrodes. In contrast, apical electrodes in the contour array are closer to the modiolus than basal electrodes. Therefore, the constant electrode spacing in the straight array does not imply that there is a spatially constant tonotopic mapping to the neuronal population excited by each electrode; similarly, the non-uniform spacing in the contour array does not necessarily lead to narrower spacing of nerve populations in the apical region excited by adjacent electrodes. Likewise, although the spacing of the HiFocus array is known to be 1.1 mm and it is clear that this spacing is in general greater than that in nucleus arrays, the implications of this difference are not definitive unless an accurate geometric configuration of the electrode array is available and compared. Thus, in addition to the inter-subject variability, which is common in CI experiments, the actual values of electrode spacing indicated in this paper should be interpreted only within the context of the specific electrode array and should be taken only as a guideline.

B. Number of effective channels and channel interactions in CIs

Although speech processing with only six to eight bands of spectral information provides reasonably good speech intelligibility in quiet for both CI users and NH listeners using CI simulation (Dorman *et al.*, 1998b; Friesen *et al.*, 2001; Fu *et al.*, 1998), it has also been suggested that more bands might be beneficial under adverse listening conditions, such as listening to difficult materials or speech in a competing background (Shannon *et al.*, 2004). While the benefit of an increased number of channels has been observed up to 24 channels in CI simulation with NH listeners (Qin and Oxen-

ham, 2003), there is little evidence demonstrating a similar trend in real CI users. Yet there seems to be a consensus to equip CI systems with more than eight bands, as all CI systems currently manufactured and used for regular clinical applications have at least 12 electrodes. The discrepancy between CI simulations and real implant performance can be often well accounted for by channel interactions limiting CI users' functional spectral resolution. (Note that the word "channel" in the term "channel interactions" refers to electric channel.) Although the term "channel interaction" has originally been used to refer to vector summation among current fields occurring simultaneously in multiple channels (Shannon, 1983; White *et al.*, 1984), interactions may still occur in interleaved stimulation when the interleaved delays are within the range of the neural refractory period of nerve fibers (Eddington *et al.*, 1978; Tong and Clark, 1986; Zeng *et al.*, 2005). Therefore, channel interactions have been identified as one of the primary limiting factors affecting spectral resolution in CIs (Henry *et al.*, 2005; Hughes and Abbas, 2006; Won *et al.*, 2007). The spatial boundaries of the neural populations excited by each electrode are not clearly defined, and there would likely be a substantial overlap in the populations excited by adjacent electrodes, compromising spectral resolution. This might be the basis of an upper limit to the "effective" number of channels available for each CI user. Also, the limited availability of healthy neural populations proximal to CI electrodes could be a further complicating factor. Beyond this number, little additional benefit would be obtained because any detailed spectral information available at this stage of signal processing would not pass through the electrode-neural interface. The number of effective channels in most CI users, presuming that no serious irregularities in neural survival patterns exist, is likely between 8 and 12 (with an average of 10) inferred from the fact that at least eight channels of speech information are required for successful performance in a number of clinical studies, and in certain cases the increase in channels beyond 8 (such as up to 12) leads to additional benefits (Nie *et al.*, 2006). Present-day technology can produce an electrode array with electrodes at a higher spatial density; yet currently CI systems do not utilize more than 22 electrodes (in fact, other devices that were developed later than the nucleus system adopt fewer electrodes, e.g., 16 electrodes for Clarion and 12 for Medel), and there is no evidence of disadvantage associated with fewer electrodes in these systems, as it is difficult to demonstrate an improvement in speech understanding beyond more than 12 channels.

The present results, suggesting that electrodes should be separated by a two-electrode spacing to enhance the effect of perceptual segregation in some subjects, are an interesting coincidence given that the limiting number of effective channels, as discussed above, is about 10. Subjects in the present study demonstrated as a group that they could segregate six speech channels from five different noise channels, while they were unable to segregate ten speech channels from ten different noise channels. Therefore, in the present experiment, subjects were able to effectively use up to 11 channels. This result could be an alternative demonstration of the limiting number of effective channels. Hypothetically, had there

been an effect of segregation in the ten-channel CIS, it would have implied that the upper limit would be about 20 channels. This result also suggests that the number of effective channels in S7 (who showed the least effect of separation among those five subjects) might be substantially lower than 11. Furthermore, the present and previous results in speech recognition performance, demonstrating little benefit beyond 12 channels, further suggest that an attempt to present information to substantially more than 12 channels might not necessarily provide a benefit, for example, 43-channel ACE processing using 22 monopolar (MP) modes and 21 dual-MP modes in the nucleus freedom system (Busby and Plant, 2005), unless it used a novel approach that reduced or minimized channel interactions between electrodes. Along the same line of argument, an adjustment of the frequency boundaries between two channels by altering the frequency-to-electrode table or filterbank settings, as described in some studies (e.g., Geurts and Wouters, 2004), might be effective only if used between electrodes with sufficient spacing.

C. Summary, caveats, and further implications

The present study demonstrated a modest improvement in consonant recognition in noise by CI users when speech and noise electrodes were sufficiently separated, such as a two-electrode spacing in the nucleus contour array (equivalent to 1.1 mm on average). This serves as evidence that some CI users could segregate speech from background noise if these signals are presented with sufficient peripheral separation. While it might be problematic to prescribe two-electrode spacing as a necessary and sufficient condition for perceptual segregation, as the clear effect was demonstrated only in two subjects; this measure could serve as a guideline for future investigations. It appears that the performance of perceptual segregation is influenced by channel interactions occurring at the periphery, which might be one source of the individual variability observed in the present study.

1. Two CIS conditions (ten-channel versus six-channel)

The two separation conditions in the present study (one-electrode and two-electrode) were tested with different CIS processing conditions (one in ten-channel and the other in six-channel) due to the layout of nucleus electrode arrays: ten-channel CIS processing could not be tested with two-electrode spacing (a total of 40 electrodes would be required in order to test it); also testing six-channel CIS with one-electrode spacing would not be realistic because then the tonotopic representation of speech information would be highly compressed into a narrow region of the cochlea and the intelligibility of such processing would be severely limited (Baskent and Shannon, 2004). Therefore, the interpretation of the present results is based on a presumption that the effect of electrode separation observed was due to the amount of electrode separation rather than the number of CIS channels per se. On a related note, as the two conditions were tested with different subject groups, the data might have been affected by the grouping of subjects; i.e., had S8 and S9 been tested in the ten-channel CIS condition, a stron-

ger effect might have been observed with one-electrode spacing. Overall, this is in line with the suggestion earlier that the findings in terms of electrode spacing be taken as a guideline and the results be subject to the individual variability.

2. Temporal versus spatial separation

In both the overlapped and interlaced conditions, although pulse trains of speech and noise signals were temporally interleaved, the temporal separation alone does not imply perceptual segregation of those pulse trains because the temporal gap between successive pulses in the present experiment layouts was relatively small (on the order of 0.1 ms). (The temporal gap between successive pulses is the reciprocal of the total stimulation rate: e.g., it is 69 μ s with a total stimulation rate of 14 400 Hz). It is unlikely that CI users can segregate streams of input on the same or close electrodes with such a short temporal gap. It was of primary interest in the present study to assess the effect of spatial separation (i.e., separated electrodes) between those two conditions.

3. Implications for application

As mentioned earlier, the conception of multi-stream processing was rationalized for a proper laboratory testing of CI users' ability to manage multiple signals and was not intended to emulate signal enhancement algorithms. The crux of this method is to provide multiple sources of input with as much integrity as possible, thereby allowing CI users to utilize these input streams to their advantage. The present data imply that a multi-stream processing approach might be beneficial to some CI users under some limited circumstances. For example, a dual-microphone system can be utilized, as Deeks and Carlyon (2004) proposed, in which the signals from two microphones are routed to different electrode sets. This application might be useful in applications where there is an uncertainty regarding source locations and selective listening in space is required, as in a cocktail party listening environment. However, a future study of selective listening tasks would be necessary to properly address the issue.

ACKNOWLEDGMENTS

The author is grateful to Wendy Parkinson, Lisa Dahlstrom, and Frank Warren for access to the clinical population. The author is heavily indebted to Chris van den Honert for his guidance and feedback. The following individuals were deeply appreciated for invaluable comments, suggestions, and constructive feedback at early stages of this study and preparation of the manuscript: Chris Turner, Robert Shannon, Monita Chatterjee, Robert Carlyon, John Galvin III, two anonymous reviewers, and Ken Grant. This work was supported by Cochlear Americas, the University of Utah, a seed money grant from the National Organization for Hearing Research Foundation, and the National Institutes of Health (NIDCD Grant No. R03DC009061).

¹The level of noise was set during the pilot study to decrease the score about 10%–15% from the quiet condition through conventional processing

of speech and noise mixture. This reduction might be different from what is shown in Figs. 2 and 3, where the scores were obtained with multi-stream processing.

- Allen, J. B. (1994). "How do humans process and recognize speech?," *IEEE Trans. Speech Audio Process.* **2**, 567–577.
- Assmann, P. F., and Summerfield, Q. (1990). "Modeling the perception of concurrent vowels: Vowels with different fundamental frequencies," *J. Acoust. Soc. Am.* **88**, 680–697.
- Bacon, S. P., Opie, J. M., and Montoya, D. Y. (1998). "The effects of hearing loss and noise masking on the masking release for speech in temporally complex backgrounds," *J. Speech Lang. Hear. Res.* **41**, 549–563.
- Baskent, D., and Shannon, R. V. (2004). "Frequency-place compression and expansion in cochlear implant listeners," *J. Acoust. Soc. Am.* **116**, 3130–3140.
- Battmer, R. D., Reid, J. M., and Lenarz, T. (1997). "Performance in quiet and in noise with the Nucleus Spectra 22 and the Clarion CIS/CA cochlear implant devices," *Scand. Audiol.* **26**(4), 240–246.
- Beauvois, M. W., and Meddis, R. (1996). "Computer simulation of auditory stream segregation in alternating-tone sequences," *J. Acoust. Soc. Am.* **99**, 2270–2280.
- Bregman, A. S., *Auditory Scene Analysis* (The MIT Press, Cambridge, MA, 1990).
- Brokx, J. P. L., and Nöteboom, S. G. (1990). "Intonation and the perception of simultaneous voices," *J. Phonetics* **19**, 23–26.
- Busby, P. A., and Plant, K. L. (2005). "Dual electrode stimulation using the Nucleus CI24RE cochlear implants: Electrode impedance and pitch ranking studies," *Ear Hear.* **26**, 504–511.
- Carlyon, R. P., Long, C. J., Deeks, J. M., and McKay, C. M. (2007). "Concurrent sound segregation in electric and acoustic hearing," *J. Assoc. Res. Otolaryngol.* **8**, 119–133.
- Chatterjee, M., Sarampalis, A., and Oba, S. I. (2006). "Auditory stream segregation with cochlear implants: A preliminary report," *Hear. Res.* **222**, 100–107.
- Cooper, H. R., and Roberts, B. (2007). "Auditory stream segregation of tone sequences in cochlear implant listeners," *Hear. Res.* **225**, 11–24.
- Cullington, H. E., and Zeng, F. G. (2008). "Speech recognition with varying numbers and types of competing talkers by normal-hearing, cochlear-implant, and implant simulation subjects," *J. Acoust. Soc. Am.* **123**, 450–461.
- Deeks, J. M., and Carlyon, R. P. (2004). "Simulations of cochlear implant hearing using filtered harmonic complexes: Implications for concurrent sound segregation," *J. Acoust. Soc. Am.* **115**, 1736–1746.
- Donaldson, G. S., and Kreft, H. A. (2006). "Effects of vowel context on the recognition of initial and medial consonants by cochlear implant users," *Ear Hear.* **27**, 658–677.
- Dorman, M. F., Loizou, P. C., and Fitzke, J. (1998a). "The identification of speech in noise by cochlear implant patients and normal-hearing listeners using 6-channel signal processors," *Ear Hear.* **19**, 481–484.
- Dorman, M. F., Loizou, P. C., Fitzke, J., and Tu, Z. (1998b). "The recognition of sentences in noise by normal-hearing listeners using simulations of cochlear-implant signal processors with 6–20 channels," *J. Acoust. Soc. Am.* **104**, 3583–3585.
- Eddington, D. K., Dobene, W. H., Brackmann, D. E., Mladejovsky, M. G., and Parkin, J. L. (1978). "Auditory prostheses research with multiple channel intracochlear stimulation in man," *Ann. Otol. Rhinol. Laryngol.* **87**, 5–38.
- Fetterman, B. L., and Domico, E. H. (2002). "Speech recognition in background noise of cochlear implant patients," *Otolaryngol.-Head Neck Surg.* **126**, 257–263.
- Friesen, L. M., Shannon, R. V., Baskent, D., and Wang, X. (2001). "Speech recognition in noise as a function of the number of spectral channels: Comparison of acoustic hearing and cochlear implants," *J. Acoust. Soc. Am.* **110**, 1150–1163.
- Fu, Q. J., and Galvin, J. J., III (2001). "Recognition of spectrally asynchronous speech by normal-hearing listeners and Nucleus-22 cochlear implant users," *J. Acoust. Soc. Am.* **109**, 1166–1172.
- Fu, Q. J., and Nogaki, G. (2005). "Noise susceptibility of cochlear implant users: The role of spectral resolution and smearing," *J. Assoc. Res. Otolaryngol.* **6**, 19–27.
- Fu, Q. J., Shannon, R. V., and Wang, X. (1998). "Effects of noise and spectral resolution on vowel and consonant recognition: acoustic and electric hearing," *J. Acoust. Soc. Am.* **104**, 3586–3596.

- Geurts, L., and Wouters, J. (2004). "Better place-coding of the fundamental frequency in cochlear implants," *J. Acoust. Soc. Am.* **115**, 844–852.
- Hartmann, W. M., and Johnson, D. (1991). "Stream segregation and peripheral channeling," *Music Percept.* **9**, 155–184.
- Henry, B. A., Turner, C. W., and Behrens, A. (2005). "Spectral peak resolution and speech recognition in quiet: Normal hearing, hearing impaired, and cochlear implant listeners," *J. Acoust. Soc. Am.* **118**, 1111–1121.
- Hong, R. S., and Turner, C. W. (2006). "Pure-tone auditory stream segregation and speech perception in noise in cochlear implant recipients," *J. Acoust. Soc. Am.* **120**, 360–374.
- Hughes, M. L., and Abbas, P. J. (2006). "Electrophysiologic channel interaction, electrode pitch ranking, and behavioral threshold in straight versus perimodiolar cochlear implant electrode arrays," *J. Acoust. Soc. Am.* **119**, 1538–1547.
- Kokkinakis, K., and Loizou, P. C. (2008). "Using blind source separation techniques to improve speech recognition in bilateral cochlear implant patients," *J. Acoust. Soc. Am.* **123**, 2379–2390.
- Kwon, B. J., and van den Honert, C. (2006). "Independent and concurrent processing of multiple audio input signals in a prosthetic hearing implant," U.S. Patent Application 20060265061, filed Jan. 20, 2006; published Nov. 23, 2006.
- Miller, G. A., and Nicely, P. E. (1955). "An analysis of perceptual confusions among some English consonants," *J. Acoust. Soc. Am.* **27**, 338–352.
- Moore, B. C., Glasberg, B. R., and Peters, R. W. (1986). "Thresholds for hearing mistuned partials as separate tones in harmonic complexes," *J. Acoust. Soc. Am.* **80**, 479–483.
- Nelson, D. A., Schmitz, J. L., Donaldson, G. S., Viemeister, N. F., and Javel, E. (1996). "Intensity discrimination as a function of stimulus level with electric stimulation," *J. Acoust. Soc. Am.* **100**, 2393–2414.
- Nelson, P. B., Jin, S. H., Carney, A. E., and Nelson, D. A. (2003). "Understanding speech in modulated interference: Cochlear implant users and normal-hearing listeners," *J. Acoust. Soc. Am.* **113**, 961–968.
- Nie, K., Barco, A., and Zeng, F. G. (2006). "Spectral and temporal cues in cochlear implant speech perception," *Ear Hear.* **27**, 208–217.
- Peterson, G. E., and Lehiste, I. (1962). "Revised CNC lists for auditory tests," *J. Speech Hear. Disord.* **27**, 62–70.
- Qin, M. K., and Oxenham, A. J. (2003). "Effects of simulated cochlear-implant processing on speech reception in fluctuating maskers," *J. Acoust. Soc. Am.* **114**, 446–454.
- Shannon, R. V. (1983). "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," *Hear. Res.* **12**, 1–16.
- Shannon, R. V., Fu, Q. J., and Galvin, J. (2004). "The number of spectral channels required for speech recognition depends on the difficulty of the listening situation," *Acta Oto-Laryngol., Suppl.* **552**, 50–54.
- Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. M., Staller, S. J., Shipp, D. B., Shalloo, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antongenelli, T., Brimacombe, J. A., Pijl, S., Daniels, P., George, C. R., McDermott, H. J., and Beirer, A. L. (1994). "Evaluation of a new spectral peak coding strategy for the Nucleus 22 channel cochlear implant system," *Am. J. Otol.* **15**, 15–27.
- Stickney, G. S., Assmann, P. F., Chang, J., and Zeng, F. G. (2007). "Effects of cochlear implant processing and fundamental frequency on the intelligibility of competing sentences," *J. Acoust. Soc. Am.* **122**, 1069–1078.
- Stickney, G. S., Zeng, F. G., Litovsky, R., and Assmann, P. (2004). "Cochlear implant speech recognition with speech maskers," *J. Acoust. Soc. Am.* **116**, 1081–1091.
- Tong, Y. C., and Clark, G. M. (1986). "Loudness summation, masking, and temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea," *J. Acoust. Soc. Am.* **79**, 1958–1966.
- Turner, C. W., Souza, P. E., and Forget, L. N. (1995). "Use of temporal envelope cues in speech recognition by normal and hearing-impaired listeners," *J. Acoust. Soc. Am.* **97**, 2568–2576.
- van den Honert, C., and Kelsall, D. C. (2007). "Focused intracochlear electric stimulation with phased array channels," *J. Acoust. Soc. Am.* **121**, 3703–3716.
- White, M. W., Merzenich, M. M., and Gardi, J. N. (1984). "Multichannel cochlear implants: Channel interactions and processor design," *Arch. Otolaryngol.* **110**, 493–501.
- Wilson, B. S., Finley, C. C., Lawson, D. T., Wolford, R. D., Eddington, D. K., and Rabinowitz, W. M. (1991). "Better speech recognition with cochlear implants," *Nature (London)* **352**, 236–238.
- Won, J. H., Drennan, W. R., and Rubinstein, J. T. (2007). "Spectral-ripple resolution correlates with speech reception in noise in cochlear implant users," *J. Assoc. Res. Otolaryngol.* **8**, 384–392.
- Zeng, F. G., Chen, H., and Han, S. (2005). "Temporal masking in electric hearing," *J. Assoc. Res. Otolaryngol.* **6**, 390–400.