



Published in final edited form as:

J Speech Lang Hear Res. 2009 April ; 52(2): 385–395. doi:10.1044/1092-4388(2008/07-0219).

The effect of temporal gap identification on speech perception by users of cochlear implants

Elad Sagi¹, Adam R. Kaiser², Ted A. Meyer³, and Mario A. Svirsky¹

¹Department of Otolaryngology, New York University School of Medicine, New York, New York, 10016

210021 Dupont Circle Ct., Fort Wayne, Indiana, 46845

³Department of Otolaryngology – HNS, Medical University of South Carolina, Charleston, South Carolina, 29425

Abstract

Purpose—This study examined the ability of listeners using cochlear implants (CIs) and listeners with normal-hearing (NH) to identify silent gaps of different duration, and the relation of this ability to speech understanding in CI users.

Method—Sixteen NH adults and eleven postlingually deafened adults with CIs identified synthetic vowel-like stimuli that were either continuous or contained an intervening silent gap ranging from 15 to 90 ms. Cumulative d' , an index of discriminability, was calculated for each participant. Consonant and CNC word identification tasks were administered to the CI group.

Results—Overall, the ability to identify stimuli with gaps of different duration was better for the NH group than for the CI group. Seven CI users had cumulative d' scores that were no higher than those of any NH listener, and their CNC word scores ranged from 0 to 30%. The other four CI users had cumulative d' scores within the range of the NH group, and their CNC word scores ranged from 46% to 68%. For the CI group, cumulative d' scores were significantly correlated with their speech testing scores.

Conclusions—The ability to identify silent gap duration may help explain individual differences in speech perception by CI users.

I. INTRODUCTION

Cochlear implants (CIs) are electronic devices that have enabled individuals with severe to profound hearing loss to regain some hearing. Nearly all of those who receive cochlear implants regain the sensation of sound. Some implanted individuals receive enough perceptual benefit that they can even communicate using a standard telephone (Dorman, Dove, Parkin, Zacharchuk, and Dankowski, 1991). This task is difficult because there are no visual cues and the auditory signal itself is not optimal. However, many users of cochlear implants do not receive such benefit, and demonstrate substantial variability in speech perception performance (Staller et al., 1997). Unfortunately, there is much we do not know about the exact mechanisms that are responsible for this variability or about the perceptual cues that are used by listeners with cochlear implants to understand speech. A better understanding of these cues may eventually guide the development of better speech processing algorithms, and provide insights into the peripheral and central processes involved in speech perception by users of cochlear implants.

Although there are many outstanding questions concerning how CI users understand speech, there are also some hypotheses that have received considerable experimental support. For example, it is thought that one of the reasons why CI users have more difficulty understanding speech sounds than listeners with normal hearing is that the former are limited in their ability to discriminate frequency. This limitation may be a problem because different speech sounds are produced by different articulatory gestures, resulting in different spectral envelope peaks. Steady-state values of these spectral peaks, or formant frequencies, provide important information for vowel and consonant identification, as do some types of formant transitions. Poor frequency discrimination can make it difficult to identify formant frequency values, thus making it more difficult to identify speech sounds.

Kewley-Port and Watson (1994) found that listeners with normal-hearing (NH) could detect differences in formant frequency of about 14 Hz in the range of frequencies around F1 and about 1.5% in the range of frequencies around F2. In contrast, Donaldson and Nelson (2000) measured the ability of fourteen users of the Nucleus-22 cochlear implant (standard electrode) to pitch-rank stimuli delivered to different cochlear locations and found that, on average, a physical separation of about 1.5 mm was necessary to discriminate between two stimulated locations. The Nucleus electrode array has an interelectrode distance of about 0.75 mm, so the 1.5 mm mentioned above is the distance between two electrodes that are not adjacent, but instead separated by an intermediate electrode. Given the maps that are typically used with the Nucleus device to relate analysis filter bands to intracochlear electrodes, a distance of 1.5 mm is equivalent to 300–320 Hz in the F1 frequency range and about 30% in the F2 frequency range. More recently, Fitzgerald et al. (2007) assessed formant frequency discrimination in twenty users of the Nucleus-24 device (9 implanted with the standard straight electrode and 11 implanted with the Contour™ modiolar hugging electrode). On average, these listeners could discriminate on the order of 50–100 Hz in the F1 frequency range and about 10% in the F2 frequency range. In terms of physical separation along the electrode array, these formant discrimination values amount to about 0.5 mm. Hence, notwithstanding improvements in performance with advances in device technology, these data are consistent with the generally held belief that CI users' ability to discriminate formant frequency is much worse than that of normal-hearing listeners. Donaldson and Nelson also found that variability in their CI users' ability to utilize available spectral cues for consonant place-of-articulation was correlated with these users' ability to discriminate between electrodes based on pitch. Hence some of the variability in CI users' speech perception outcome can be explained by limitation in these users' ability to discriminate frequency.

In contrast to their relatively poor frequency discrimination ability, listeners with CIs are thought to perform nearly as well as normal-hearing listeners on some temporal processing tasks. Shannon (1993) reviewed studies of temporal processing by CI users. This review included the following measures: forward masking, temporal integration, modulation detection, rate discrimination, and gap detection. Forward masking uses pairs of stimuli to assess the refractory properties of the auditory system. Two stimuli are used, a high level masker and a lower level test stimulus. The signal level of the test stimulus required to achieve perceptual threshold is measured as a function of the length of time that has elapsed since the presentation of the masker stimulus. In effect, the first stimulus “masks” the second and the level of the test stimulus required to achieve threshold as a function of time since masker offset is a measure of the listener's recovery from masking. After normalizing for differences of scale in signal level between acoustic and electrical stimulation, Shannon (1990) demonstrated that the forward masking functions for listeners using cochlear implants and for listeners with normal-hearing were approximately the same. Another psychophysical process, temporal integration, can be estimated by measuring signal detection thresholds for a stimulus with fixed amplitude but varying in duration. Up to a point, longer signals can be detected at lower amplitudes. The signal duration at which further temporal increases fail to reduce the detection

threshold is similar for listeners with CIs and listeners with normal-hearing, i.e. between 100–200 ms (Shannon, 1993). Modulation detection, the ability to detect amplitude modulation in either a carrier of pulse-trains or a continuous signal can also be used to compare listeners with cochlear implants to those with normal-hearing. Both groups can detect modulation at roughly the same modulation depth, and both groups show a decreased ability to detect modulation at higher modulation frequencies (Shannon, 1992), though modulation detection by CI users is much more sensitive to stimulation level than for listeners with normal-hearing (Fu, 2002). Rate discrimination measures the ability of listeners to perceive pitch differences in stimuli delivered at different stimulation rates. When presented with pulse trains, listeners with CIs can perceive pitch differences for stimulation rates up to 300 Hz. This ability compares well with that of listeners with normal-hearing who can perceive pitch differences in amplitude modulated broadband noise for modulation rates up to 300 to 500 Hz (Shannon, 1983).

Many studies suggest that detection of temporal gaps is comparable in adults with cochlear implants who were deafened postlingually and listeners with normal-hearing. For both groups, gap detection threshold decreases as stimulus level is increased. From low to high level stimuli, gap detection thresholds can improve from 20–50 ms to 2–5 ms in CI users (Shannon, 1989; Preece and Tyler, 1989; Moore and Glasberg, 1988), and from about 25 ms to 2 ms in listeners with normal-hearing (Fitzgibbons and Gordon-Salant, 1987; Fitzgibbons and Wightman, 1982; Florentine and Buus, 1984) when stimuli are presented under quiet listening conditions, though these values can also vary depending on one's choice of stimulus marker (Tyler, Moore, and Kuk, 1989; Wei, Cao, Jin, Chen, and Zeng, 2007). It should be noted, however, that the similarity of gap detection thresholds for adults using CIs and adults with normal-hearing may be specific to CI users who were deafened *postlingually* with at least a minimal amount of open-set speech recognition. For example, a study of CI users who were prelingually deafened has shown substantial intersubject variability in gap detection threshold values (Busby and Clark, 1999). Similarly, gap detection thresholds for CI users without any, or with very little, open-set speech recognition can be greater than 50 ms (Moore and Glasberg 1988; Muchnik, Taitelbaum, Tene, and Hildesheimer, 1994; Tyler, Moore, and Kuk, 1989; Wei et al., 2007).

Temporal processing abilities are potentially important in speech perception. For example, some intervocalic consonants, such as those in /aga/ and /ata/, contain silent gaps (short time intervals with little or no acoustic energy), while others such as /aza/ and /ama/ do not. Thus, detection of an acoustic gap may be useful to distinguish between speech sounds. Additionally, the average duration of the gap may be a useful cue for phoneme identification. For example, by our measurements of electrical stimulation patterns (presented below), the gap in /aga/ is, on average, shorter than the gap in /ata/ by about 20ms and may therefore be a potential cue to the identification of these phonemes. Yet another example of how silent gaps are important for differentiating between phonemes is demonstrated by Munson and Nelson (2005). Using a categorical perception task, they demonstrated in both listeners with normal hearing and those with CIs that silent gap durations of 20–50 ms were critical in producing a perceptual shift between synthetic productions of 'say' and 'stay'. Temporal cues such as silent gap duration may, in fact, be particularly important for CI users because their perception of spectral cues is significantly worse than that of listeners with normal hearing, and they may be forced to rely more heavily on those acoustic cues that are relatively well perceived with a CI. Indeed, a study by Teoh, Neuburger, and Svirsky (2003) suggests that gap duration and other temporal cues (such as noise duration and total duration of the phoneme) may be employed by CI users for consonant identification.

The purpose of the present study is to measure the ability of CI users to identify silent gaps of different duration, to compare this ability with that of listeners with normal hearing, and to assess the relation, if any, between gap identification ability and speech perception in CI users. Although some evidence suggests that the ability to detect the presence or absence of a gap

(i.e. gap detection) is similar for listeners with CIs and listeners with normal-hearing, the ability to distinguish between several gaps of different duration (i.e. gap identification) has not been studied and may be different between these groups of listeners. In turn, identification of silent gap duration may be pertinent to categorizing different consonants, and hence words, especially for CI users who are limited in their ability to utilize the spectral cues that help identify these speech sounds. Furthermore, if a relationship exists between temporal gap identification ability and speech perception in CI users, then this relationship could also help explain some of the large individual differences in speech perception outcomes that exist for this group. A secondary motivation for this study was to obtain estimates of just noticeable differences (JND's) for temporal gap identification. These JND measurements are an important input to our multidimensional phoneme identification model, a quantitative framework that has been proposed to explain the mechanisms employed by CI users to understand speech sounds (Svirsky, 2000, 2002).

II. METHOD

A. Participants

Twenty seven adult listeners were tested in this study. Sixteen of the listeners were adults with normal-hearing (21 to 61 years of age with audiometric thresholds < 25 dB HL at 1, 2, and 4 kHz), and the other eleven were adult users of cochlear implants who were postlingually deafened (35 to 73 years of age). All of the CI users, recruited from the clinical population at Indiana University, had profound bilateral sensorineural hearing losses and at least one year of experience with their device (Table 1). Each participant provided informed consent and was reimbursed for travel to and from testing sessions and for the time of participation. Five of the participants used the Nucleus 22 device, one used the Nucleus 24 device, one used the Clarion S device, three used the Clarion 1.0 device, and one used the Clarion 1.2 device. All users of the Nucleus device used the SPEAK processing strategy, with the exception of C3 who used MPEAK at the time of testing. All users of the Clarion device used the CIS processing strategy.

B. Stimuli and equipment

Seven stimuli were created using the Klatt 88 speech synthesizer software (Klatt & Klatt, 1990). The first stimulus in the continuum was a synthetic three-formant steady state vowel with a duration of 1 second. Formant frequencies were 500, 1500, and 2500 Hz with a fundamental frequency of 100Hz. Onset and offset of the vowel envelope occurred over a 10 ms period, and were linear in dB. The other six stimuli were similar to the first, with the exception that they contained an intervening silent gap placed at the center of the stimulus. For these stimuli, the silent gap duration was 15, 30, 45, 60, 75, or 90 ms. Stimuli were presented at a level of at least 70 dB C-weighted SPL. For the silent gap, the stimulus offset (transition from maximum sound level to silence) and stimulus onset (from silence back up to maximum sound level) were made over a period of 10 ms, and were linear in dB. The gap duration was specified as the interval between stimulus offset and stimulus onset, starting from the midpoint of the upward and downward slopes of the stimulus offset envelope represented in dB, and ending at the same midpoint of the stimulus onset.

The total amount of energy in each of the seven stimuli was identical so that the duration of the stimulus markers that preceded and followed each silent interval remained constant while the silent gap was varied. Hence, the stimuli with longer gaps also had a correspondingly longer total duration (see Figure 1). Although differences in total duration could potentially serve as a cue for identifying stimuli, the change in total duration is small (at most 90 ms) relative to the duration of the markers (i.e. 500 ms each) and presumably less prominent perceptually than the salience of a silent gap interval placed at the center of the stimulus. In principle, one could fix the total duration of the stimuli and insert silent gaps of varying duration at the center of

each stimulus, but this would result in stimuli that vary in marker duration, i.e. total energy, which could also serve as a potential cue for stimulus identification. Between the two alternatives, fixed total duration or fixed total energy, stimuli with a fixed amount of total energy were chosen for the present study.

The stimuli were digitally stored using a sampling rate of 11025 Hz at 16 bits of resolution. They were presented from an Intel® based PC equipped with a SoundBlaster compatible sound card to an Acoustic Research loudspeaker. Custom software was used to present stimuli and record responses.

Speech perception stimuli included CNC word lists and 16 consonants from the Iowa Consonant Identification Task (female speaker) (Tyler, Preece, & Lowder, 1987). Each CNC word list consists of a 50 item monosyllabic open-set word identification task (Peterson & Lehiste, 1962). The consonant identification task is a closed-set 16 alternative task that uses 16 consonants in an /a/consonant/a/ format. For example, the consonant *m* would be presented as /ama/. Both of these tests used natural speech and were presented in an auditory only, quiet condition at 70 dB C-weighted SPL to listeners seated approximately 1 meter from the loudspeaker.

C. Procedure

All participants were tested using a seven-alternative absolute identification task. This task was chosen because in order to identify a consonant, listeners presumably have to estimate the value of different acoustic cues rather than discriminate two different values. Thus, the gap duration identification and consonant identification tasks employed in the present study are essentially the same, with the main difference between them being only the type and number of stimuli. The task of identifying silent gaps of different durations (just like identifying different phonemes) is more difficult than a categorical perception task, as it involves heavier reliance on auditory memory and other cognitive resources in addition to basic psychoacoustic abilities.

Each of the seven stimuli was randomly presented ten times during each block of testing for a total of 70 presentations per block. Prior to testing, participants were allowed to listen to the stimuli at will and were then walked through a practice session of 14 presentations, 2 per stimulus, to become familiar with the procedure and the testing software. On the computer display, seven interactive buttons were labeled 'Stimulus 1' to 'Stimulus 7'. 'Stimulus 1' produced the stimulus with no gap, 'Stimulus 2' produced the stimulus with the 15 ms gap, and so forth, up to 'Stimulus 7' which produced the 90 ms gap. During testing, a listener would select a button on screen to play a stimulus, and then select one of the Stimulus buttons s/he felt corresponded best with the stimulus that was heard. If desired, the listener had the option to repeat the stimulus before making an identification. After each response, feedback was provided on the computer monitor before moving on to the presentation of the next stimulus. Listeners were tested until asymptotic performance was achieved, as determined by failure to improve their cumulative d' -scores (next section). The number of testing blocks ranged from 6 to 10, i.e. the total number of stimuli ranged from 420 to 700. The average of the best two cumulative d' -scores during the first eight blocks was typically sufficient for most listeners to provide an estimate of asymptotic performance, and this average score is the parameter that is reported in the present study. Listeners were allowed time for breaks between testing blocks as they deemed necessary. Including breaks, listeners with normal-hearing completed this task within 1.5 to 2 hours, and CI users required up to 3 hours.

CI users were asked to return at a later date to complete a speech perception battery. Nine CI users performed at least 5 repetitions of the consonant identification task for a total of fifteen presentations per consonant. Two CI users, C3 and C7, performed 4 and 2 repetitions

respectively. All CI users were administered at least three CNC word lists. Word lists were selected so that the difficulty of words across lists varied less than 3% on average when the lists selected for a given subject were combined (estimated from the CNC list equivalency data of Skinner et al., 2006). For these tasks, all runs were averaged to arrive at a percent correct score for each CI user.

D. Analysis

For each block of testing, a sensitivity index d' (Braid and Durlach, 1972; Tong and Clark, 1985) was calculated for each pair of adjacent stimuli (1 vs. 2, 2 vs. 3, etc.) using Equation 1.

$$d_n = \frac{\bar{x}_{n+1} - \bar{x}_n}{\frac{\sigma_{n+1} + \sigma_n}{2}} \quad (1)$$

d' is a parameter that indicates the discriminability between two distributions. It was used here to describe the discriminability between two adjacent stimuli, say the n^{th} and the $n+1^{\text{th}}$ stimuli for $n = 1$ to 6. If one assumes that responses for a given pair of adjacent stimuli follow normal distributions and have equal variance, then an optimal observer trying to identify one of two stimuli with a d' of 0 would perform at chance level, responding correctly 50% of the time. Similarly, a d' of 1 would result in correct responses 69% of the time in the same two-stimulus discrimination task, and a d' of 3 would result in correct responses 93% of the time. To avoid the possibility of d' approaching infinity due to lack of overlap between adjacent distributions, especially between the stimulus with no gap and the stimulus with the 15 ms gap (i.e. 1 vs. 2), any d' greater than 3 was assigned a value of 3 (Tong and Clark, 1985), as is customary. The discrimination index was calculated for each adjacent set of stimuli, 1 vs. 2, 2 vs. 3, 3 vs. 4 etc. The results from these individual comparisons were then added to arrive at a measure of total sensitivity, i.e. 'cumulative d' ' (Braid and Durlach, 1972) which is a global measure of the participant's discrimination abilities across all the gap durations tested. That is, cumulative d' provides an estimate of the degree of confusion among the seven stimuli, where a larger cumulative d' implies relatively less confusion (i.e. good discrimination) and a smaller cumulative d' implies relatively more confusion among the range of silent gap durations tested (i.e. from 0 ms to 90 ms in steps of 15 ms).¹

A 2-factor Analysis of Variance was conducted to test for differences in cumulative d' scores between listeners with normal hearing and those with CIs, and between younger and older participants. For the latter, participants under the age of 45 comprised the younger group and participants 45 years and older comprised the older group. For participants with normal hearing, the younger group was comprised of 11 listeners and the older group was comprised of 5 listeners. For participants with CIs, the younger group was comprised of 4 listeners and the older group was comprised of 7 listeners. The Holm-Sidak method was employed for post-hoc comparisons within and between groups.

The gap duration in each consonant token used in the Iowa 16-consonant test was measured after tokens were processed through two CI speech processors; once for the Nucleus 22 device and once for the Clarion 1.2 device. All three repetitions of the 16 consonants in the test were played at 70 dB SPL (measured at the level of the speech processor's microphone) and the corresponding electrical stimulation patterns coming out of the speech processor were recorded to disk. Speech processor parameters for each device were set to default settings. For the Nucleus device, the stimulation pattern was obtained using sCILab (Bögli, Dillier, Lai, Rohner,

¹Note that the cumulative d' scores in the present study were derived from an identification task and therefore the results are not necessarily identical to those that would be obtained in a discrimination task.

& Zillus, 1995; Lai, Bögli, & Dillier, 2003), a program which captures and decodes the radio-frequency signal that is transmitted by the external speech processor. For the Clarion device, tokens were delivered to the speech processor connected to an 'implant in a box', and the stimulation pattern was digitized directly from the electrode array contacts (which were connected using 5 kOhm resistors) into a computer. For the purpose of calculating gap duration we disregarded voicing energy, which showed up as stimulation pulses delivered to the lowest frequency channel. Using these gap measurements, each CI user's consonant matrix was reduced to a 2×2 gap-feature matrix indicating the number of times consonants without a silent gap were mistaken for consonants with a silent gap, and vice-versa. The percent correct score of this gap-feature matrix was calculated for each CI user. The percentage of time each gap consonant was confused as a no-gap consonant was also calculated across listeners, to determine if any gap consonants were potentially masked relative to others.

Three Pearson product-moment coefficients were calculated to assess the correlation, if any, between the CI users' gap identification cumulative d' scores and speech perception percent correct scores (i.e. consonants, gap-feature, and CNC words). A Fisher's z-test was then used to test for any significant differences between correlation coefficients.

III. RESULTS

Table 2 shows cumulative d' scores for all listeners as well as speech perception scores for the listeners using CIs. Cumulative d' scores for listeners with normal-hearing fell within the range 5.4 to 11.9, and scores for CI users fell within the range 3.2 to 9.8. Figure 2 illustrates, for both groups of listeners, the cumulative d' score (total bar height) as well as the fraction due to the comparison between the first two stimuli. That is, the total cumulative d' is broken down into two parts: black bars represent a d' comparing responses to stimuli 1 and 2 (i.e., the non-gapped stimulus and the stimulus with a 15 ms gap), and the stacked white bars represent cumulative d' for stimuli 2–7 (i.e., stimuli with gaps ranging from 15 ms to 90 ms). It is very clear from the figure that all listeners, CI users and listeners with normal hearing alike, had perfect or near perfect ability to discriminate the non-gapped stimulus from the stimulus with a 15 ms gap, because the corresponding d' values between these two stimuli are all equal, or nearly equal, to 3. In contrast, the ability to identify gaps of different durations (cumulative d' for stimuli 2–7) ranged from 2.4 for NH1 to 8.9 for NH16, and from near zero for C1 to 6.8 for C11. In fact, seven out of the eleven CI users had cumulative d' scores that were no higher than the lowest cumulative d' score among the listeners with normal-hearing. The other four CI users (C8 to C11) performed within the range of the group with normal-hearing. This difference in performance between the NH and CI groups was confirmed by the 2-factor analysis of variance, which revealed that cumulative d' scores for the CI group were significantly lower than for the NH group ($p = 0.0105$).

The ANOVA also revealed that age was a significant factor affecting cumulative d' scores ($p = 0.0303$), such that younger participants (i.e., < 45 years) scored significantly higher on average than older participants (≥ 45 years). This small but significant effect is consistent with previous studies that demonstrate age-related differences in gap-detection thresholds between younger and older listeners with normal hearing (Schneider and Hamstra, 1999; Snell and Frisina, 2000). No significant interaction was obtained between hearing status and age.

Speech perception tests show the typically wide performance range found in CI users. On the 16-consonant task, scores ranged from 6% to 75% correct with chance performance at about 6%. When consonant confusion matrices were partitioned into 2×2 gap feature matrices, i.e. 10 consonants without a silent gap and 6 gapped consonants (see Figure 3), the scores ranged from 51% to 95%. With this analysis, purely chance performance would yield a score of about 53% assuming that each consonant is selected with equal probability. On the CNC word

recognition task, the scores ranged from 0% to 68%, where chance performance would yield a score near 0% since this is an open-set task. This range of performance on the CNC task is consistent with previous studies that tested listeners with earlier generation CI devices similar to those tested in the present study (Waltzman, Cohen, and Roland, 1999; Zwolan et al., 2005), but lower on average than CNC scores obtained with more recent devices (Balkany et al., 2007; Fitzgerald et al., 2007; Koch, Osberger, Segel, and Kessler, 2004; Skinner et al., 2006). Note that the group of seven CI users with poor cumulative d' scores had CNC word identification scores of 0% to 30%, whereas the four CI users whose cumulative d' scores were within the normal-hearing range had CNC word scores of 46% to 68%. In contrast to the listeners with CIs, virtually all listeners with normal-hearing would be expected to score close to 100% in these speech perception tests under the same conditions.

Figure 3 shows the average silent gap durations measured for each one of the 16 consonants used in this test. Each bar represents the average of three tokens. There are six consonants containing a silent gap, /g/, /k/, /b/, /p/, /d/ and /t/, with average gap durations of 59, 60, 62, 65, 73, and 76 ms, respectively. Based on these measurements, one would expect that all of the participants with CIs would be able to perceive the difference between a consonant with a gap and a consonant without a gap. This is because all of the gapped consonants contain gaps that are more than 50 ms in duration, and all of these participants showed that they were able to easily distinguish continuous sounds from those containing a 15 ms gap. However, percent correct scores for the gap/no-gap feature (Table 2 and Figure 4) were less than 100% for all CI users. Furthermore, in Table 3, the percentage of the number of times (across CI users) where each gapped consonant was presented and identified as a consonant without a gap is not uniform across gapped consonants. Rather, the consonants /g/ and /b/ were confused 50–60% of time as consonants without gaps, whereas the other gapped consonants were mistaken for consonants without gaps only 15–25% of the time. Hence, the gaps in the consonants /g/ and /b/ were potentially masked to a larger extent than the other gapped consonants.

Figures 4A and 4B show scatter-plots of percent correct scores for the 16-consonant task and the gap/no-gap feature for this task, respectively, against cumulative d' scores for the 11 CI users tested in this study. Figure 4C is a scatter-plot of percent correct scores for CNC words against cumulative d' for these CI users. All speech perception scores were significantly correlated with cumulative d' : $r = 0.68$ and $p = 0.021$ for the 16-consonant task; $r = 0.69$ and $p = 0.019$ for the gap/no-gap feature; and $r = 0.88$ and $p < 0.001$ for CNC words. The relatively larger correlation between cumulative d' and CNC words is not significantly different from the other two correlation coefficients (Fisher's z -test, $p = 0.27$).

IV. DISCUSSION

Speech perception by users of CIs is highly variable from one user to the next and is also, on average, worse than that of listeners with normal-hearing. There is a great deal of interest in psychophysical and other capabilities that might explain these two facts. A generally held view in the cochlear implant literature is that CI users suffer from poor frequency discrimination, but that their temporal processing skills are comparable to those of listeners with normal-hearing. The latter view that CI users have temporal processing capabilities similar to those of listeners with normal-hearing is supported by several types of experiments (summarized by Shannon, 1993) including some experiments measuring gap detection thresholds (but see Busby & Clark, 1999; Moore & Glasberg, 1988; Muchnik et al., 1994) and is consistent with our finding that all of our participants with CIs were clearly able to label and differentiate a continuous stimulus from one with a 15 ms gap. For the presentation level and type of marker stimuli used in this study, it is reasonable to expect that gap thresholds would be below 15 ms for these CI users (Shannon, 1989; Moore & Glasberg, 1988; Muchnik et al., 1994). However, the ability of CI users to identify silent gaps of different duration was extremely variable,

ranging from almost nil to normal. In addition, for the CI group, this ability (as measured by cumulative d' scores for stimuli 2–7) was significantly correlated with speech perception scores as shown in Figure 4. These results suggest that CI users may use silent gap duration as a cue to consonant identity, and that their differing abilities to identify such gaps may help explain some of their individual differences in speech perception.

The relation between gap identification ability and speech perception reported here contributes to a small, but growing body of evidence suggesting that certain aspects of temporal processing contribute to speech perception in CI users. For example, Fu (2002) demonstrated a strong correlation between phoneme identification and CI users' mean modulation detection thresholds calculated across each subject's dynamic range. The contribution of temporal processing ability to phoneme identification receives additional, indirect support from the modeling studies of Svirsky (2002), whose model of consonant identification can fit consonant confusion data much better when it includes a temporal dimension. The temporal dimension used by Svirsky is the duration of a silent gap. In any case, it seems clear that the gap duration identification abilities of less successful CI users lag substantially behind those of listeners with normal-hearing or those of more successful CI users. What may be the physiological reason behind this variability in gap duration identification? In our opinion, the auditory periphery is an unlikely locus for individual differences in temporal processing (Abbas, 1993) because the electrically stimulated auditory nerve is quite capable of encoding the gross temporal differences that were present in our stimuli, such as the difference between a 15 ms silent gap and a 90 ms silent gap, which some CI users were unable to perceive. Instead, it may be that the differences in cumulative d' that were observed in this study have their origin in more central differences in auditory processing skills.

Another observation is that the CI users' identification of gapped and non-gapped consonants was much lower than would be expected based on their psychophysical performance. We say this because their discrimination of continuous sounds from sounds with a 15 ms or greater gap was virtually perfect in the psychophysical task (where they had to perform absolute identification of synthetic stimuli), but their ability to identify consonants with and without gaps was far from perfect (as evidenced by their percent correct scores for the gap-no gap feature). One possible account for this result is that listeners may have been paying more attention to acoustic cues other than gap duration, or at least weighting those other cues more heavily. In other words, they may have known that a given stimulus contained a silent gap, but other acoustic cues led them to identify the stimulus as a non-gapped consonant nonetheless. Two prominent examples of this are the consonants /g/ and /b/. Although these consonants contain a silent gap, they were confused by the CI listeners as consonants without a silent gap more often than the other gapped consonants. Both of these consonants have a relatively short silent gap duration (Figure 3), and they are also voiced. Remember that for the purpose of calculating gap duration we disregarded voicing energy, which showed up as stimulation of the most apical (low frequency) channel. It is possible, however, that the presence of voicing energy decreased the listener's certainty that a silent gap occurred, particularly if the silent gap was relatively short. In contrast to /b/ and /g/, the gapped consonant /d/ was confused far less frequently as a consonant without a gap, even though it is voiced, perhaps because it has a relatively longer silent gap duration. Alternatively, it may be that the cumulative d' estimates obtained in this study may only be achieved by most participants under relatively ideal conditions, that is, with carefully synthesized acoustic stimuli that are identical to each other except for the duration of the silent gap. When participants have to process gap information in conjunction with other acoustic cues (spectral and amplitude cues, for example) their processing of temporal cues may suffer to some extent.

It is important to note that the CI listeners in the present study were users of somewhat older generation CIs and processing strategies. The possibility does exist that gap identification

ability could improve with newer generation devices. Newer CI devices operate at higher stimulation rates and are thus capable of providing more temporal resolution when encoding a sound stimulus. In principle, this improved resolution could provide listeners with better discrimination for silent gap duration. However, we suspect that the magnitude of this improvement would be small.

Clearly, perception of speech sounds by human listeners in general and by users of CIs in particular is a very complex phenomenon. In the case of CI users, we hope to develop a comprehensive quantitative model of speech perception based in part on the individual listener's discrimination abilities. The present study, exploring temporal gap identification, is one step in that direction. The manner in which other abilities known to be important for speech perception, such as formant frequency discrimination, combine with temporal processing ability, such as temporal gap identification, and contribute to speech perception in CI users is currently under investigation. The existence of a theoretical framework for speech perception by CI users could help guide the search for improved signal processing and aural rehabilitation strategies for this population.

ACKNOWLEDGEMENTS

This research was supported by Grants R01-DC003937 (PI: Mario A. Svirsky) and T32- DC000012 (PI: David B. Pisoni) from the National Institute on Deafness and Other Communication Disorders (NIH, USA). We express thanks to Advanced Bionics Corporation for providing us with a Clarion 1.2 'implant in a box', and to Norbert Dillier from the ENT Department at University Hospital, Zürich, Switzerland, for providing a copy of sCILab software.

References

- Abbas, PJ. Electrophysiology. In: Tyler, RS., editor. Cochlear implants: Audiological foundations. San Diego: Singular Publishing Group, INC.; 1993. p. 317-356.
- Bögli, H.; Dillier, N.; Lai, WK.; Rohner, M.; Zillus, BA. Swiss Cochlear Implant Laboratory (Version 1.4) [Computer software]. Zürich, Switzerland: Authors; 1995.
- Balkany T, Hodges A, Menapace C, Hazard L, Driscoll C, Gantz B, Kelsall D, Luxford W, McMenomy S, Neely JG, Peters B, Pillsbury H, Roberson J, Schramm D, Telian S, Waltzman S, Westerberg B, Payne S. Nucleus freedom north american clinical trial. *Otolaryngology – Head and Neck Surgery* 2007;136:757–762. [PubMed: 17478211]
- Braida LD, Durlach NI. Intensity perception. II. Resolution in one-interval paradigms. *Journal of the Acoustical Society of America* 1972;51:483–502.
- Busby PA, Clark GM. Gap detection by early-deafened cochlear-implant participants. *Journal of the Acoustical Society of America* 1999;105:1841–1852. [PubMed: 10089607]
- Donaldson GS, Nelson DA. Place-pitch sensitivity and its relation to consonant recognition by cochlear implant listeners using the MPEAK and SPEAK speech processing strategies. *Journal of the Acoustical Society of America* 2000;107:1645–1658. [PubMed: 10738818]
- Dorman M, Dove HJ, Parkin J, Zacharchuk S, Dankowski K. Telephone use by patients fitted with the Ineraid cochlear implant. *Ear and Hearing* 1991;12:368–369. [PubMed: 1783245]
- Fitzgerald MB, Shapiro WH, McDonald PD, Neuburger HS, Ashburn-Reed S, Immerman S, Jethanamest D, Roland JT, Svirsky MA. The effect of perimodiolar placement on speech perception and frequency discrimination by cochlear implant users. *Acta Oto-Laryngologica* 2007;127:378–383. [PubMed: 17453457]
- Fitzgibbons PJ, Gordon-Salant S. Temporal gap resolution in listeners with high-frequency sensorineural hearing loss. *Journal of the Acoustical Society of America* 1987;81:133–137. [PubMed: 3819171]
- Fitzgibbons PJ, Wightman FL. Gap detection in normal and hearing-impaired listeners. *Journal of the Acoustical Society of America* 1982;72:761–765. [PubMed: 7130534]
- Florentine M, Buus S. Temporal gap detection in sensorineural and simulated hearing impairments. *Journal of Speech and Hearing Research* 1984;27:449–455. [PubMed: 6482414]

- Fu QJ. Temporal processing and speech recognition in cochlear implant users. *NeuroReport* 2002;13:1635–1639. [PubMed: 12352617]
- Kewley-Port D, Watson CS. Formant-frequency discrimination for isolated English vowels. *Journal of the Acoustical Society of America* 1994;95:485–496. [PubMed: 8120259]
- Klatt DH, Klatt LC. Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America* 1990;87:820–857. [PubMed: 2137837]
- Koch DB, Osberger MJ, Segel P, Kessler D. HiResolution™ and conventional sound processing in the HiResolution™ Bionic Ear: using appropriate outcome measures to assess speech recognition ability. *Audiology and Neuro-Otology* 2004;9:214–223. [PubMed: 15205549]
- Lai WK, Bögli H, Dillier N. A software tool for analyzing multichannel cochlear implant signals. *Ear and Hearing* 2003;24:380–391. [PubMed: 14534409]
- Moore CBJ, Glasberg BR. Gap detection with sinusoids and noise in normal, impaired, and electrically stimulated ears. *Journal of the Acoustical Society of America* 1988;83:1093–1101. [PubMed: 3356814]
- Muchnik C, Taitelbaum R, Tene S, Hildesheimer M. Auditory temporal resolution and open speech recognition in cochlear implant recipients. *Scandinavian Audiology* 1994;23:105–109.
- Munson B, Nelson PG. Phonetic identification in quiet and in noise by listeners with cochlear implants. *Journal of the Acoustical Society of America* 2005;118:2607–2617. [PubMed: 16266181]
- Peterson G, Lehiste I. Revised CNC lists for auditory tests. *Journal of Speech and Hearing Disorders* 1962;27:62–70. [PubMed: 14485785]
- Preece JP, Tyler RS. Temporal-gap detection by cochlear prosthesis users. *Journal of Speech and Hearing Research* 1989;32:849–856. [PubMed: 2601315]
- Schneider BA, Hamstra SJ. Gap detection thresholds as a function of tonal duration for younger and older listeners. *Journal of the Acoustical Society of America* 1999;106:371–380. [PubMed: 10420628]
- Shannon RV. Multichannel electrical stimulation of the auditory nerve in man: I. Basic psychophysics. *Hearing Research* 1983;11:157–189. [PubMed: 6619003]
- Shannon RV. Detection of gaps in sinusoids and pulse trains by patients with cochlear implants. *Journal of the Acoustical Society of America* 1989;85:2587–2592. [PubMed: 2745882]
- Shannon RV. Forward masking in patients with cochlear implants. *Journal of the Acoustical Society of America* 1990;88:741–744. [PubMed: 2212298]
- Shannon RV. Temporal modulation transfer functions in patients with cochlear implants. *Journal of the Acoustical Society of America* 1992;91:2156–2164. [PubMed: 1597606]
- Shannon, RV. Psychophysics. In: Tyler, RS., editor. *Cochlear implants: Audiological foundations*. San Diego: Singular Publishing Group, INC.; 1993. p. 357-388.
- Skinner MW, Holden LK, Fourakis MS, Hawks JW, Holden T, Arcaroli J, Hyde M. Evaluation of equivalency in two recordings of monosyllabic words. *Journal of the American Academy of Audiology* 2006;17:350–366. [PubMed: 16796301]
- Snell KB, Frisina DR. Relationships among age-related differences in gap detection and word recognition. *Journal of the Acoustical Society of America* 2000;107:1615–1626. [PubMed: 10738815]
- Staller S, Menapace C, Domico E, Mills D, Dowell RC, Geers A, Pijl S, Hasenstab S, Justus M, Bruelli T, Borton AA, Lemay M. Speech perception abilities of adult and pediatric Nucleus implant recipients using spectral peak (SPEAK) coding strategy. *Otolaryngology-Head and Neck Surgery* 1997;117:236–242. [PubMed: 9334771]
- Svirsky MA. Mathematical modeling of vowel perception by cochlear implantees who use the "compressed-analog" strategy: Temporal and channel-amplitude cues. *Journal of the Acoustical Society of America* 2000;107:1521–1529. [PubMed: 10738806]
- Svirsky, MA. The multidimensional phoneme identification (MPI) Model: A new quantitative framework to explain the perception of speech sounds by cochlear implant users. In: Serniclaes, W., editor. *Etudes et Travaux*. Vol. Vol. 5. Brussels, Belgium: Institut de Phonetique et des Langues Vivantes of the ULB (Free University of Brussels; 2002. p. 143-186.
- Teoh SW, Neuburger HS, Svirsky MA. Acoustical and Electrical Pattern Analysis of Consonant Perceptual Cues Used by Cochlear Implant Users. *Audiology and Neuro-Otology* 2003;8:269–285. [PubMed: 12904682]

- Tong YC, Clark GM. Absolute identification of electric pulse rates and electrode positions by cochlear implant patients. *Journal of the Acoustical Society of America* 1985;77:1881–1888. [PubMed: 3839004]
- Tyler RS, Preece JP, Lowder MW. The Iowa audiovisual speech perception laser video disc. *Laser Videodisc and Laboratory Report, University of Iowa at Iowa City, Department of Otolaryngology – Head and Neck Surgery*. 1989
- Tyler RS, Moore BCJ, Kuk FK. Performance of some of the better cochlear-implant patients. *Journal of Speech and Hearing Research* 1989;32:887–911. [PubMed: 2601319]
- Waltzman SB, Cohen NL, Roland JT. A comparison of the growth of open-set speech perception between the Nucleus 22 and Nucleus 24 cochlear implant systems. *The American Journal of Otology* 1999;20:435–441. [PubMed: 10431883]
- Wei C, Cao K, Jin X, Chen X, Zeng FG. Psychophysical performance and mandarin tone recognition in noise by cochlear implant users. *Ear and Hearing* 2007;28:62S–65S. [PubMed: 17496650]
- Zwolan TA, Kileny PR, Smith S, Waltzman S, Chute P, Domico E, Firszt J, Hodges A, Mills D, Whearty M, Osberger MJ, Fisher L. Comparison of continuous interleaved sampling and simultaneous analog stimulation speech processing strategies in newly implanted adults with a Clarion 1.2 cochlear implant. *Otology and Neurotology* 2005;26:455–465. [PubMed: 15891649]

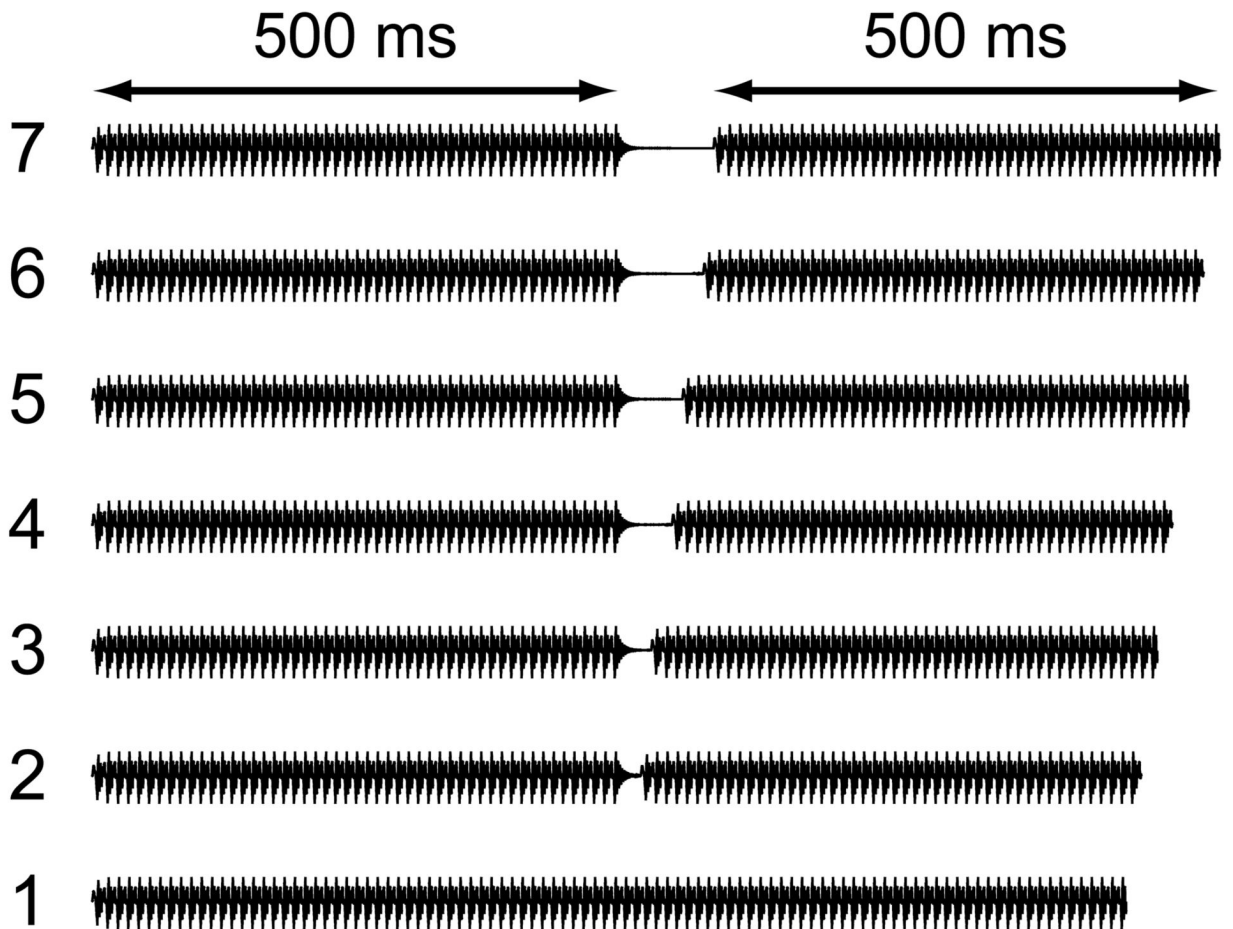


FIG. 1.
Graphical representation of stimulus waveforms used for the gap-identification task arranged by stimulus number.

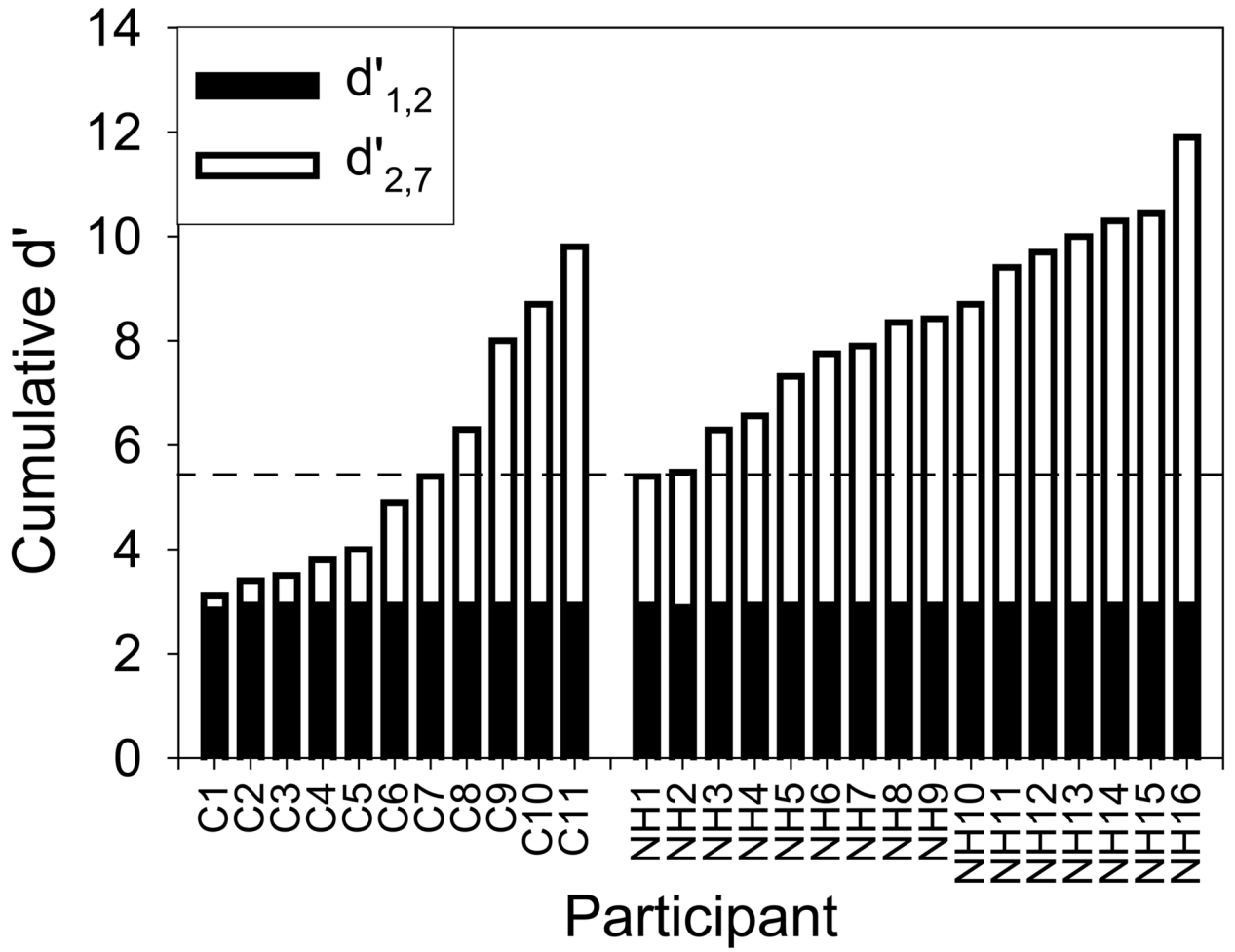


FIG. 2. Cumulative d' measurements for each participant. The black part of the bar represents the portion of the cumulative d' that is due to the participant's ability to discriminate between stimulus 1 and stimulus 2. The horizontal dashed line indicates how scores for 7 listeners with CIs fell below the NH group. C = participants with cochlear implants and NH = participants with normal hearing

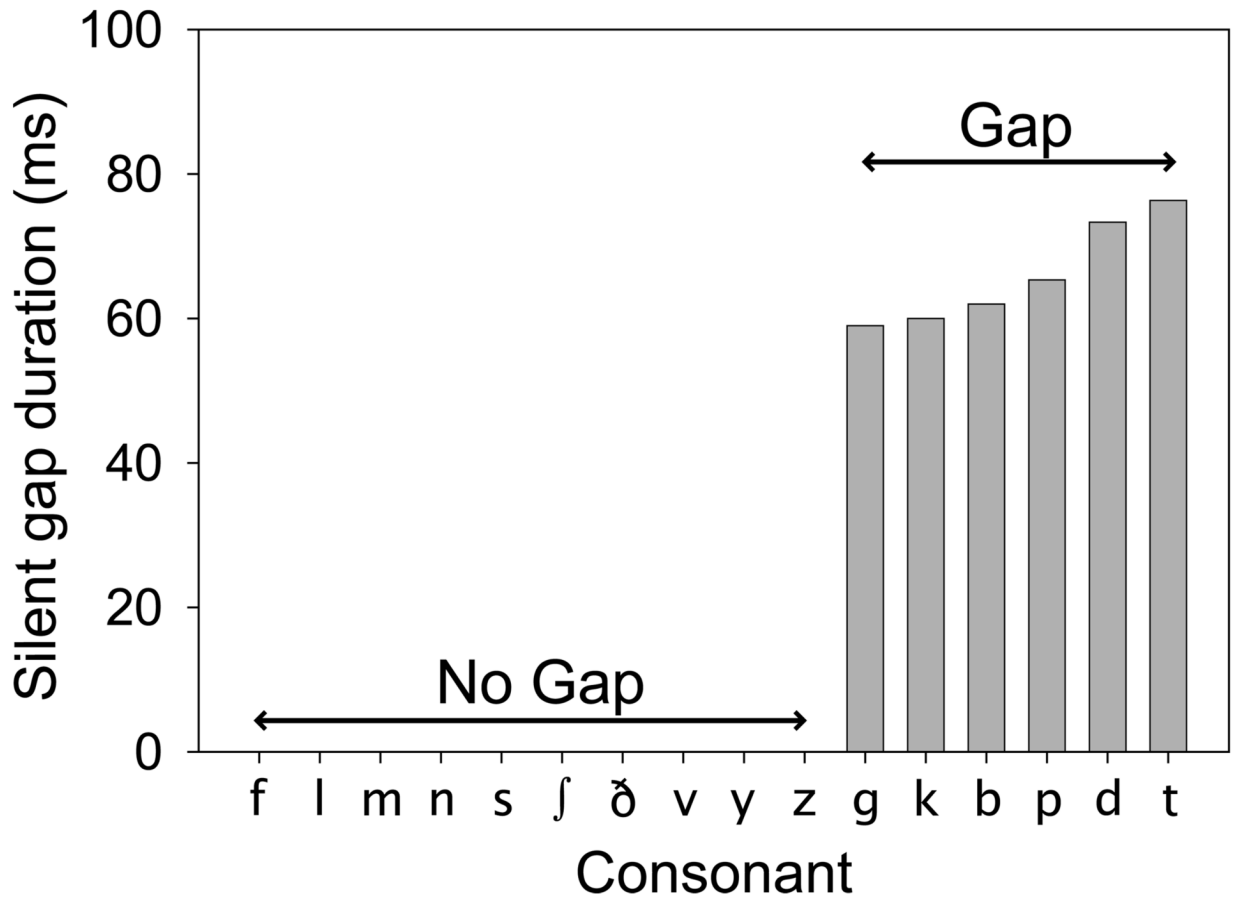


FIG. 3.

Gap durations for stimuli in the 16-consonant task as measured from the consonant stimuli processed through a Spectra 22 cochlear implant speech processor and visually analyzed using sCILab software.

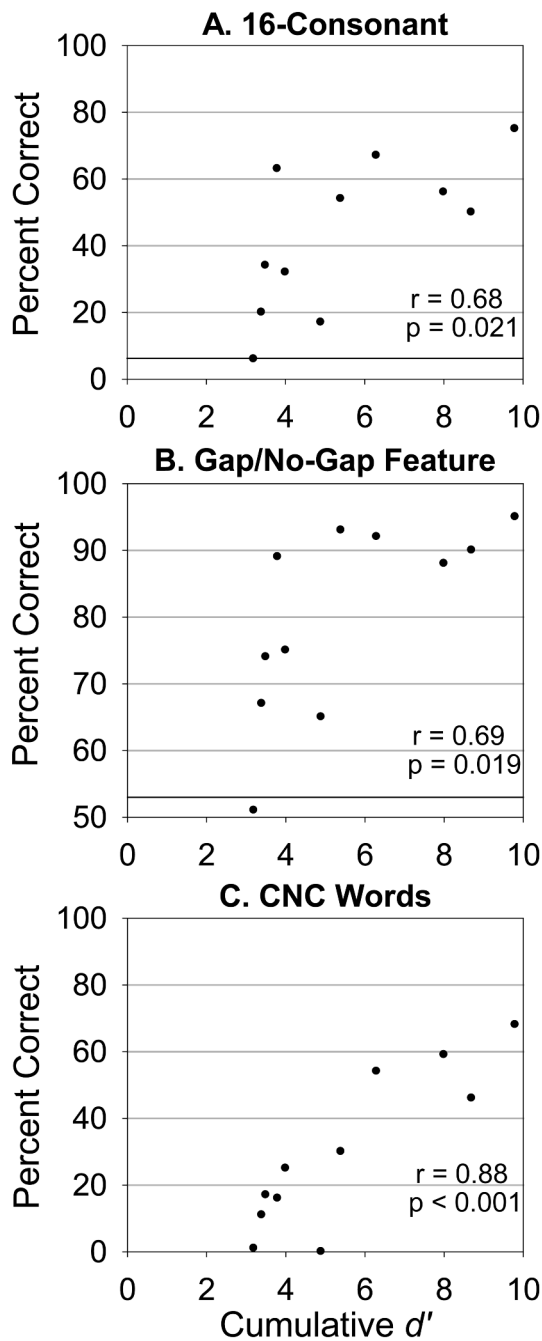


FIG. 4. Correlation between percent correct scores from the (A) 16-Consonant identification task, (B) gap/no-gap feature on the consonant task, (C) CNC wordlists, and cumulative d' scores (abscissa) from the gap-identification task for the 11 CI users tested in this study. All percent correct scores were significantly correlated with cumulative d' scores. In panels A and B, black horizontal lines indicate chance performance. In panel C the task is open-set, so chance performance is near 0%.

Table 1
Demographics of participants with cochlear implants (C) and normal-hearing (NH).

Normal Hearing		CI Users					
Listener	Age (Years)	Listener	Age (Years)	Age at Onset of Deafness (Years)	Age at Implant (Years)	Implant Use (Years)	Implant Type
NH1	56	C1	67	* ^a	65	2	Nucleus 22
NH2	57	C2	62	56	57	5	Nucleus 22
NH3	21	C3	41	* ^a	38	2	Clarion 1.0
NH4	61	C4	69	55	66	2	Nucleus 22
NH5	45	C5	73	27	71	2	Nucleus 22
NH6	24	C6	58	* ^a	52	5	Clarion 1.0
NH7	35	C7	41	32	34	7	Nucleus 22
NH8	48	C8	66	43	61	5	Clarion 1.0
NH9	34	C9	45	42	43	1.7	Clarion S
NH10	29	C10	37	34	36	1.1	Nucleus 24
NH11	30	C11	35	29	31	3	Clarion 1.2
NH12	25						
NH13	22						
NH14	22						
NH15	33						
NH16	32						

^a C1, Progressive; C2, Progressive since childhood; C6, Progressive since childhood

Table 2

Performance in the gap duration identification task (all participants) and in the speech perception tasks (CI users). Listeners with normal-hearing (NH) are listed first, and CI users follow (C). Within each group, participants are listed in order of increasing ability to identify silent gap duration.

Normal Hearing		CI Users				
Listener	Cum-d'	Listener	Cum-d'	16-Cons (%)	Gap/No-gap (%)	CNC (%)
NH1	5.4	C1	3.2	6	51	1
NH2	5.5	C2	3.4	20	67	11
NH3	6.3	C3	3.5	34	74	17
NH4	6.6	C4	3.8	63	89	16
NH5	7.3	C5	4.0	32	75	25
NH6	7.8	C6	4.9	17	65	0
NH7	7.9	C7	5.4	54	93	30
NH8	8.4	C8	6.3	67	92	54
NH9	8.4	C9	8.0	56	88	59
NH10	8.7	C10	8.7	50	90	46
NH11	9.4	C11	9.8	75	95	68
NH12	9.7					
NH13	10.0					
NH14	10.3					
NH15	10.4					
NH16	11.9					
Mean	8.4	Mean	5.5	43	80	30
SD	1.9	SD	2.3	2.3	14	24

Table 3

Percentage (%) of the number of times where each gapped consonant was presented and identified as a consonant without a gap, as measured across CI users.

Cons.	%
g	59
k	20
b	52
p	15
d	23
t	25