Rate and onset cues can improve cochlear implant synthetic vowel recognition in noise

Myles Mc Laughlin[a)]
Hearing and Speech Research Laboratory, Department of Otolaryngology-Head and Neck Surgery, University of California, Irvine, California 92697-5320

Richard B. Reilly
Trinity Centre for Bioengineering, Trinity Biomedical Sciences Institute, Trinity College Dublin, Pearse Street, Dublin 2, Ireland

Fan-Gang Zeng
Hearing and Speech Research Laboratory, Departments of Anatomy and Neurobiology, Biomedical Engineering, Cognitive Sciences, and Otolaryngology—Head and Neck Surgery, University of California, Irvine, California 92697-5320

(Received 5 December 2011; revised 15 January 2013; accepted 16 January 2013)

Understanding speech-in-noise is difficult for most cochlear implant (CI) users. Speech-in-noise segregation cues are well understood for acoustic hearing but not for electric hearing. This study investigated the effects of stimulation rate and onset delay on synthetic vowel-in-noise recognition in CI subjects. In experiment I, synthetic vowels were presented at 50, 145, or 795 pulse/s and noise at the same three rates, yielding nine combinations. Recognition improved significantly if the noise had a lower rate than the vowel, suggesting that listeners can use temporal gaps in the noise to detect a synthetic vowel. This hypothesis is supported by accurate prediction of synthetic vowel recognition using a temporal integration window model. Using lower rates a similar trend was observed in normal hearing subjects. Experiment II found that for CI subjects, a vowel onset delay improved performance if the noise had a lower or higher rate than the synthetic vowel. These results show that differing rates or onset times can improve synthetic vowel-in-noise recognition, indicating a need to develop speech processing strategies that encode or emphasize these cues.
 [http://dx.doi.org/10.1121/1.4789940]

PACS number(s): 43.64.Me, 43.71.Es, 43.71.Ky [EB] Pages: 1546–1560

I. INTRODUCTION

The remarkable sound source segregation abilities of normal hearing (NH) people allow them to converse with relative ease in a noisy restaurant or at a cocktail party. This is not the case for most cochlear implant (CI) users who have reasonably good speech recognition in quiet (Balkany *et al*., 2007) but do extremely poorly in noise, even at moderate signal-to-noise ratios (SNR) (Friesen *et al*., 2001; Nie *et al*., 2005). In acoustic hearing, sound source segregation has been studied extensively (for a review, see Bregman, 1990). The acoustic cues for sound source segregation and the associated cochlear and neural processing are well studied. However, sound source segregation in electric hearing has been less well studied and, given the very different nature of electric and acoustic hearing, it may be expected that it will differ.

A CI bypasses any cochlear processing and, via a speech processing strategy, delivers acoustic information to the brain by direct electric stimulation of the auditory nerve. Current speech processing strategies do present some spectral and onset segregation cues to the nerve, but CI users performance in noise remains poor. The poor performance may be caused by an inability to fully extract these cues because of the way in which they are presented. An alternative explanation may be that CI users do not possess the same array of neural mechanisms to process these segregation cues. This explanation is perhaps most applicable to the prelingually deafened CI population (Hancock *et al*., 2010; Litovsky *et al*., 2010). Yet, a third explanation may be that sound source segregation could be improved in electric hearing by using different cues than those used in acoustic hearing.

A number of studies have investigated sound source segregation in CI users and in NH listeners using envelope-vocoder simulations. Nelson *et al*. (2003) and Stickney *et al*. (2004) both investigated the effect of fluctuating masking noise on sentence recognition and found that CI users were much worse at separating speech from noise than NH listeners. Qin and Oxenham (2005) showed the detrimental effects of envelope-vocoder processing and poor spectral resolution on a concurrent vowel recognition task and attributed the poor performance to the lack of the fundamental frequency (F0) segregation cue. Luo and Fu (2009) investigated concurrent Chinese syllable recognition in NH listeners and found significantly poorer results for CI processed speech than for unprocessed speech. In a separate study, Luo *et al*. (2009) showed that in Mandarin speaking CI users

[a)]Author to whom correspondence should be addressed. Also at: Trinity Centre for Bioengineering, Trinity Biomedical Sciences Institute, Trinity College Dublin, Pearse Street, Dublin 2, Ireland. Electronic mail: myles.mclaughlin@uci.edu

 0001-4966/2013/133(3)/1546/15/$30.00

concurrent syllable recognition scores were 40%–60% worse than single syllable recognition scores. They attributed poorer performance to the fact that CI speech processing severely degrades the F0 cues. Hong and Turner (2006, 2009) investigated stream segregation in CI users. They found that CI users could stream amplitude modulated noise at relatively low frequencies and that the ability to stream pure tones was significantly correlated with speech perception in noise. Cooper and Roberts (2009) also investigate stream segregation in CI users and found that it was poor, even with a comparatively simple task. In NH subjects using CI simulation, Gaudrain *et al.* (2008) showed that when F0 spectral cues were severely reduced, listeners may still make use of a weak F0 spectral cue to facilitate streaming. All of these studies investigated the segregation performance of CI users listening through a standard speech processing strategy or, in the case of NH listeners, a vocoder simulation designed to replicate the effects of a standard CI speech processing strategy. Therefore the results, to some degree, reflect the limitations of the speech processing strategy and not necessarily the sound source segregation limitations of electric hearing.

In the current study, a research interface is used to stimulate the implant directly and investigate the basic psychophysical cues underlying sound source segregation in electric hearing, independently from the speech processing strategy. This approach allows tight control of the cues present in the electrical stimulation, helping to delineate the confounding effects of the speech processing strategy and the neural processing abilities of the CI subject. It also gives us the flexibility to investigate the effect of stimulation rate on sound source segregation in electrical hearing.

Multichannel CIs make use of place pitch by partially mimicking basilar membrane cochlear processing: They present energy from different spectral regions of the acoustic signal to different electrodes in the implant, exciting a different region of the cochlea and eliciting a distinct place pitch percept. However, it is also possible to elicit a temporal pitch percept in CI users by changing the stimulation rate or number of pulses per second (pps) presented on one electrode (Zeng, 2002). Most modern speech processing strategies, based on the continuous interleaved sampling strategy (Wilson *et al.*, 1991), use a fixed rate pulse train and thus may not fully exploit stimulation rate as a potential segregation cue.

Data from NH listeners show that a sound's harmonic structure (both resolved and unresolved), defined by its F0, can be a good segregation cue in acoustic hearing (Assmann and Summerfield, 1990; Culling and Darwin, 1993). The brain can group together energy from resolved harmonics, which create a predictable pattern of excitation on the cochlea, to form one auditory object (a speakers voice for example) while ignoring energy from other frequency regions (e.g., background noise). The temporal pattern of excitation created by the unresolved harmonics may also contribute to the auditory object formation. In CI users, the limited spectral resolution provided by the 22 or less electrodes makes it impossible to accurately represent the resolved harmonic structure of a sound at the auditory nerve. This means that all the spectral regions of the output of the electrodes containing energy from a speaker's voice are no longer harmonically related according to the place pitch map; this may make it difficult for the brain to group them together. Currently, CI speech processing strategies use a fixed stimulation rate, but it may be possible to manipulate the stimulation rate cue to encode F0 information and thus improve sound source segregation.

In a preliminary report, Chatterjee *et al.* (2006) used a research interface to investigated auditory stream segregation in CI users. The results suggested that CI users could use electrode location and temporal envelope cues for stream segregation but did not specifically investigate the effect of stimulation rate. A study by Carlyon *et al.* (2007) investigated the effects of stimulation rate and onset delay on sound source segregation in CI listeners. The task used required CI subjects to detect the presence of a probe stimulus presented on one electrode together with masking stimuli presented on three other electrodes. The study found that when the onset of the probe stimulus was delayed by 200 ms relative to the masking stimulus detection thresholds were lower than when the probe stimulus was not delayed. However, when the stimulation rate of the probe stimulus was lower (77 pps) than that of the masking stimuli (100 pps), detection thresholds were not affected, leading the authors to conclude that CI users are unlikely to use a temporal pitch difference between adjacent electrodes to separate concurrent sounds.

In the current study, the effects of stimulation rate on sound source segregation were investigated, but the approach taken differs in a number of ways from that used by Carlyon *et al.* A five alternative forced choice synthetic vowel-in-noise recognition task was used. Synthetic vowels were represented by using a research interface to stimulate just two electrodes at a constant rate and constant amplitude. The two electrodes chosen to represent a vowel were those closest to the first and second formant frequencies of that vowel based on the subject's clinical map. The noise was represented by stimulating four other electrodes, meaning that six electrodes were used to represent one synthetic vowel-in-noise presentation. The vowels and noises were presented using three different stimulation rates (50, 145, or 795 pps), yielding in total nine different combinations of vowel and noise stimulation rate. This range of stimulation rates was chosen to test the entire range of rates available on the device. It was reasoned that the wider range of conditions tested in this experiment may reveal effects of stimulation rate not seen in Carlyon *et al.* Carlyon *et al.* tested a narrow range of stimulation rates (77 and 100 pps) where effects of rate may be weak or not present while here we tested a much larger range of rates. Carlyon *et al.* changed the rate of one out of four possible fixed electrodes while here rate variations across a larger number of electrodes were tested. In addition, this study examined the effect of combining stimulation rate with onset delay which was not tested by Carlyon *et al.*

II. METHODS

A. Subjects

Six adult (1 male and 5 female) postlingually deafened CI users participated in the study. Five used the Nucleus

CI24 implant (N24) and one used the Nucleus 22 implant (N22). They were between 38 and 74 yr old (mean: 64.3) and their details are provided in Table I. To participate in the study, CI users had to score above 80% correct on the vowel-in-quiet procedure described in Sec. II D. Nine subjects were initially tested, three of whom did not meet the 80% correct in quiet criteria and were not included in the study. Thus the CI subject population tested is not representative of all CI users, rather those who performed well in the vowel-in-quiet task and whom we expected would perform reasonably well in the vowel-in-noise task. Eight NH subjects (3 males and 5 females) between the ages of 18 and 31 yr old (mean: 22.6) participated in this study. The University of California Irvine's Institutional Review Board approved all experimental procedures for both the CI and the NH subjects. Informed consent was obtained from each subject. All subjects were native English speakers and were compensated for their participation.

B. Electric stimulation

All electrical stimuli were presented using a research interface (HEINRI, Wygonski and Robert, 2002) while the CI subject was seated in a quiet room. All the N24 CI subjects were stimulated in monopolar mode with the extra-cochlear electrodes MP1 and MP2 used as the return electrode. For N24 subjects, the pulse width was set at 50 μs per phase with a 10 μs interphase gap. The N22 system does not have extra-cochlear electrodes and was therefore used in pseudo-monopolar mode where all non-stimulation electrodes are used as return electrodes. For the N22 subject, the pulse width was set at 40 μs per phase with a 10 μs interphases gap. On the N22 device lowering, the pulse width allows the generation of higher pulse rates as described in the following text.

With both the Nucleus 22 and 24 devices, it is not possible to stimulate two or more electrodes simultaneously. Therefore when stimulating multiple electrodes, at the same stimulation rate, a small inter-electrode pulse interval was introduced. Figure 1 shows electrodes 10 and 12 being stimulated at 50 pps, and at this time scale, the pulses appear to be synchronous. However, there is a 10 μs gap between the end of the first pulse on electrode 10 and the beginning of the first pulse on electrode 12. This method of stimulation was used for all stimuli except the interleaved conditions tested at the end of experiment I.

TABLE I. Details of cochlear implant subjects. HLD, hearing loss duration measured in years and counted from the onset of profound hearing loss. CI, number of years of usage of the cochlear implant which was tested. Subject 5 has used a cochlear implant with a short insertion depth in the non-tested year for 10 yr. N24 means a Nucleus 24 device and N22 a Nucleus 22.

Subject	Gender	Age	Etiology	HLD (yr)	CI (yr)	CI Type
CI 1	F	68	Recessive Gene	36	8	N24
CI 2	F	69	Unknown	5	2	N24
CI 3	F	74	Unknown	20	7	N24
CI 4	M	38	Autoimmune	10	9	N24
CI 5	F	73	Hereditary	15	1	N24
CI 6	F	64	Meningitis	42	22	N22

1. Stimulation rate

To investigate the effect of stimulation rate on sound source segregation three different rates were selected for testing: 50, 145, and 795 pps per channel. However, because the N22 device uses a 2.5 MHz RF transmission coil (while the N24 uses a 5 MHz coil) and because it uses an expanded communication protocol (while the N24 uses the embedded), we could not achieve a stimulation rate of 795 pps per channel when using the HEINRI system for this particular stimulation paradigm (for a comparison of the N22 and N24 communication specifications, see Zeng *et al.*, 2008). Therefore when testing the N22 subject, a 365 pps, per channel, rate (the maximum we could reliably obtain with the HEINRI system for this multi-electrode paradigm) was substituted for the 795 pps rate. Hereafter, reference to the 795 pps rate, unless otherwise stated, means 795 pps in the N24 subjects and 365 pps in the one N22 subject.

2. Loudness balancing and amplitude roving

It has been shown that the stimulation rate has an effect on the perceived loudness in electrical stimulation (Fu, 2005; Zeng and Shannon, 1994), and it is expected that the perceived loudness of the vowel versus the noise will affect the difficulty of a vowel-in-noise recognition task. Therefore to limit the effects of loudness, we employed a loudness balancing and amplitude roving procedure as described in the following text.

Threshold and comfort levels were determined on all electrodes tested for each subject at each stimulation rate. The dynamic range for each electrode at each rate was defined as the difference between the comfort and threshold levels. A loudness balancing procedure was then applied where two electrodes (22 and 11 for subjects 2–5, 20 and 9 for subjects 1 and 6) were selected and stimulated at the same rate. By pressing a button on a graphical interface, the subject could listen to these two electrodes presented at 50, 145, and 795 pps. The subject was instructed to ignore the differences in pitch and try to match the loudness of the 50 and 795 pps stimuli to that of the 145 pps stimulus. The volume of the 50 and 795 pps stimuli were adjusted up or down by the experimenter until the subject was satisfied that all three stimuli had equal loudness. All subjects reported that the large differences in pitch made it difficult to loudness balance the stimuli and it necessitated considerable time and listening effort on the part of the subject. The procedure was repeated three times, and an average loudness correction factor was calculated for each rate based on the two electrodes tested (i.e., 22 and 11). This loudness correction factor was then applied to all electrodes when stimulated at that particular rate. This approach was chosen to save time with the caveat that it may not be as accurate as loudness balancing each electrode individually.

As a secondary precaution to limit the effects of loudness, an amplitude roving procedure was applied. Each time a multi-electrode stimulus was presented a roving factor was

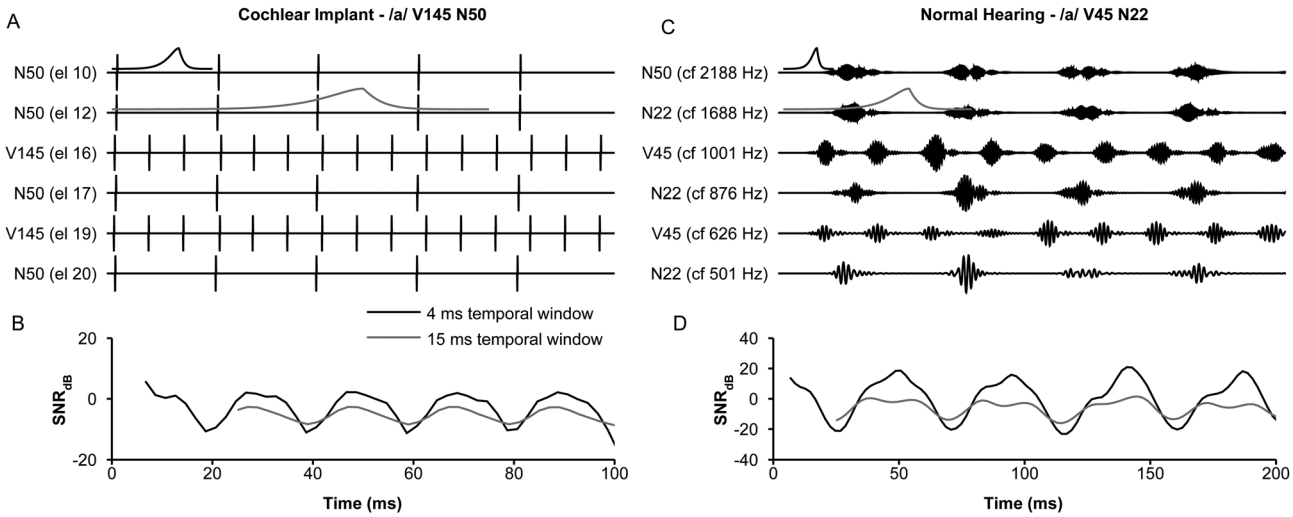

FIG. 1. Example of electrodogram and corresponding vocoder stimulus for the vowel /a/ in noise. (A) The first 100 ms of the electrodogram of the vowel-in-noise stimulus /a/ is shown. The vowel is presented on electrodes 16 and 19 at 145 pps and the noise on electrodes 10, 12, 17, and 20 at 50 pps. Electrode 10 shows a 4 ms rounded exponential (roex) temporal integration window (TIW) used in the model and in gray on electrode 12 is a 15 ms roex TIW. (B) The TIW model slides a roex function through each channel in 2 ms time steps, integrates the energy in each channel at each time step and thus calculates vowel or signal-to-noise ratio (SNR) specific to each vowel and noise rate combination. In the TIW model the final SNR_{TIW} is calculated as the mean of the 10 largest peaks in the SNR function. Smaller TIWs (4 ms in black) tend to give large peaks in the SNR function than longer TIWs (15 ms in gray). (C) The corresponding acoustic stimulus is shown. The vowel is represented with two 200 Hz wide noise-bands amplitude modulated at 45 Hz and the noise with four noise-bands at 22 Hz. The noise-bands are summed together to give the final vocoded stimulus. The same 4 and 15 ms roex TIWs are shown. (D) The same TIW model was used on the acoustic stimuli to calculate the SNR function.

calculated for each electrode. The roving factor randomly reduced the comfort level on each electrode by 0%, 6%, 12%, 18%, 24%, or 30% of the dynamic range with the constraint that the two vowel electrodes must always be reduced by 12%, 18%, 24%, or 30%. The amplitude roving procedure ensured that at least two, and sometimes all four, of the masking electrodes were presented at a larger percentage of their dynamic range than the vowel electrodes. Thus during all experiments, stimuli were presented at the comfort level determined for each electrode scaled by the loudness factor determined for the rate and again scaled by the amplitude roving factor calculated for each electrode on each new presentation. This combination of loudness balancing and amplitude roving ensured that while the masker electrodes may have a different stimulation rate than the vowel electrodes, the masker was always slightly louder than the vowel.

C. Acoustic simulation of electric stimulation

For the NH subjects, acoustic stimuli were presented to both ears at 70 dB SPL through Sennheiser headphones (HAD 200) using an external USB sound card (Creative Sound Blaster, 24 bit, 44.1 kHz) while seated in a sound booth (Industrial Acoustics). Electrical stimulation was simulated by using a procedure similar to a noise-band vocoder (Shannon *et al.*, 1995). Broadband white noise was amplitude modulated (AM) at one of three rates, 22, 45 or 100 Hz by multiplying it with a sine function raised to the power of 16. A fresh, randomly generated, noise sample was created for each stimulus presentation. The noise was modulated using a high exponent of the sine function as it concentrates more of the energy in the AM waveform near the peak. It was reasoned that concentrating more of the energy near the peak in the AM waveform would produce a simulation that is more similar to an electrical pulse than a simulation that distributes the energy around the peak. The lowest formant frequencies of a vowel used in this study was 251 Hz. A narrow band noise centered at 251 Hz with an AM rate greater than 125 Hz would mean that each peak present in the waveform used to perform the amplitude modulation (i.e., the sine function) would not be present in the amplitude modulated noise carrier (i.e. the simulated output of one electrode). The absence of peaks at some periods could introduce uncontrolled cues into the acoustic simulation. Therefore to faithfully represent the peak at each period of the modulation rate for a noise-band centered at 251 Hz the modulation rates needed to be below 125 Hz. To ensure that a peak was present at each period of the modulation rate a maximum modulation rate of 100 Hz was chosen for the NH experiments. Two lower rates, 45 and 22 Hz were arbitrarily selected to span a wide range of rates while ensuring each period of the modulation rate would be faithfully represented, even in the lowest frequency noise-band. Finally, the output of one electrode was simulated by passing AM noise through a 200 Hz bandpass filter (fourth order Butterworth) with a center frequency chosen to match that of the appropriate electrode as specified in Table II. To simulate multi-electrode stimulation at multiple stimulation rates, noise bands with different center frequencies, modulated at different rates, were summed together.

The level rove when testing CI subjects was designed to ensure that the vowel electrodes were not louder than the masker electrodes. This could arise because, in spite of the loudness balancing procedure, different stimulation rates may still give a loudness cue. With NH subjects, any potential loudness cues due to rate were controlled by equalizing the RMS amplitude of each noise-band in the simulation before summing them together, and so it was not necessary to introduce a level rove. Figure 1(C) shows an example of

TABLE II. First and second formant frequencies of the five vowels tested were mapped to two electrodes or two noise-band center frequencies in the vocoder simulations. The mapping was the done using a standarddefault Nucleus 24 frequency map. For CI subjects 1 and 6 all electrodes were shifted down by two places as electrodes 22 and 21 had very narrow dynamic ranges. CI, cochlear implant. NH, normal hearing.

	Formant (Hz)							
Vowel	First	Second	Filter on default N24 map pass-band (Hz)		CI subjects electrode No.		NH subjects center frequency (Hz)	
/i/	261	2032	188–313	1831–2063	22	11	251	1938
/a/	639	1016	563–688	938–1063	19	16	626	1001
/u/	264	794	188–313	688–813	22	18	251	751
/ɜ/	508	1240	438–563	1188–1313	20	14	501	1251
/ɛ/	585	1850	563–688	1813–2063	19	11	626	1938

the AM noise-bands at each center frequency before they are summed together. Due to the narrow bandwidth of each noise-band there is considerable fluctuation in the envelope.

D. CI subjects: Synthetic vowel-in-noise task

A five alternative forced choice recognition paradigm was used. The listener was seated in front of a computer screen displaying a graphical interface with one button representing each of the five vowels. They were presented with a stimulus and asked to use a mouse to press the button representing the vowel they had just heard. The five vowels were synthesized by stimulating just two electrodes, at a constant amplitude and rate, to represent the first and second formant frequencies of each vowel. The formant frequency to electrode mapping was determined based on a default N24 frequency map for all six subjects. Namely, the filter band within which the first formant frequency fell determined the electrode number for the first formant and the filter band within which the second formant frequency fell determined the electrode number of the second formant. Details of the formant frequencies, N24 filter bands and electrode numbers are given in Table II. In subjects 1 and 6, electrodes 22 and 21 had extremely narrow dynamic ranges, and so the formant to electrode mapping was shifted down by two electrodes. It is important to stress that these are synthetic vowel stimuli. Most of the CI subjects tested reported that the stimuli had a vowel like quality but did not sound like vowels heard through their everyday processor. They were, however, able to memorize the sounds and associate them with the vowel shown on the test interface. Using these impoverished vowel stimuli meant that the only cue available to distinguish between the five different vowels in quiet was a place pitch cue. To confirm that CI subjects could use this cue, they were first allowed to practice on a vowel-in-quiet test (i.e., just the two electrodes representing the vowel were presented) and only enrolled in the study if, after a number of practice sessions, they could consistently score above 80% in quiet.

The noise was represented by stimulating four other randomly chosen electrodes, restricted to electrode numbers between 11 and 22 (or 9 and 20 for subjects 1 and 6). Thus for a given vowel-in- noise presentation, six electrodes were stimulated, two representing the vowel and four representing the noise. An example of the vowel /a/ presented at 145 pps with noise at 50 pps is shown in Fig. 1(A). For each new vowel-in-noise presentation, the noise electrodes numbers were chosen at random with the result that with each new stimulus presentation the noise was slightly different but the vowel was constant.

Experiment I tested the effect of stimulation rate on synthetic vowel-in-noise recognition in the CI subjects using three different rates: 50, 145, and 795 pps. Presenting the vowel at one rate and the noise at another gave nine different experimental conditions, which were labeled as V50 N50, V50 N145, V50 N795, V145 N50,…, etc. Both vowel and noise were 500 ms in duration in all conditions. For a subset of CI subjects ($N = 4$), we tested three interleaved conditions where the stimulation rate of the vowel and masker were the same but the vowel pulses were shifted by half a stimulation period so that they occurred in the temporal gaps of the noise: V50I N50, V145I N50, and V795I N795. Experiment II tested the original nine non-interleaved rate conditions but now delayed the vowel by 200 ms and made the noise 700 ms long, so that the noise started first, followed by the vowel and they both stop at the same time: V50D N50, V50D N145, V50D N795, V145D N50,…, etc.

One test block consisted of 30 vowel-in-noise stimuli, six repetitions of each of the five vowels presented in random order, and each experimental condition consisted of four such blocks. The subject heard the stimuli and then responded by pressing one of five buttons on a MATLAB GUI (Mathworks, Natick, MA) indicating which vowel they heard. For each experimental condition, subjects were allowed to practice listening to the stimuli for as long as they wished before testing. During testing feedback was given after each response by highlighting the button representing the stimulus just played in green if they responded correctly and in red if they responded incorrectly. In half of the CI subjects tested, experiment II was run on the subject before experiment I.

E. NH subjects: Synthetic vowel-in-noise task

The stimulation rate cue used with the CI subjects is specific to electric hearing and would not normally occur in acoustic hearing. Therefore to test if the effects of stimulation rate on synthetic vowel-in-noise segregation were limited to electric hearing or were generalizable to acoustic hearing, NH subjects completed experiment I. It is should be noted that for the technical reasons described in the preceding text, much lower rates were used in the NH experiment than in the CI experiment. This will limit the usefulness of a

direct comparison of the results from both experiments. The motivation behind the NH experiment is simply to test if similar effects of electric stimulation rate on vowel-in-noise recognition can also be observed with acoustic stimulation. Before completing experiment I, NH listeners were tested on an acoustic simulation of the vowel-in-quiet test and all scored above 80% correct. Onset delay is a segregation cue, which is normally available in acoustic hearing but it is not normally combined with differences in rate. Therefore to examine the effect of combining onset delay cues with rate cues in acoustic hearing, experiment II was completed using the same group of NH listeners. Exactly the same five alternative forced choice vowel-in-noise paradigm, test GUI, and test procedures were used with the NH subjects, the only difference being that an acoustic simulation of electric hearing, as described in Sec. II C was used. In the NH subjects, the center frequency of a band-passed noise was set to the center frequency of the filter matching the electrode used in the CI subjects. Details are provided in Table II. All eight NH subjects participated in experiment I, which had nine different combinations of modulation rate: V22 N22, V22 N45, V22 N100, V45 N22,…, etc. As with the CI subjects, a subset of NH listeners ($n = 4$) participated in an interleaved condition were the vowel was shifted by half a period of the modulation rate: V22I N22, V45I N45, and V100I N100. Experiment II, as with the CI subjects, tested the effect of an onset delay of 200 ms on vowel-in-noise segregation using the same nine rate combinations yielding the following experimental conditions: V22D N22, V22D N45, V22D N100, V45D N22,…, etc. Six of the NH subjects participated in experiment II, with two subjects completing only the V45D conditions (e.g., V45D N22, V45D N45, and V45D N100).

F. Temporal integration window model

A temporal integration window (TIW) model was implemented to examine the effects temporal gaps present in the stimuli. The only segregation cue included in the model is a glimpsing or listening in the gaps cue. It was reasoned that if the TIW model could predict the general trends observed in the data from experiment I, it would support the hypothesis that the effects of rate on the vowel-in-noise task are mostly caused by a listening in the gaps cue, and not a temporal pitch cue. The model is based on the work of Moore *et al.* (1988), who showed that the TIW in NH humans is best described by the sum of two rounded exponentials (roex). Sliding a roex TIW through the stimulus allowed for the calculation of a vowel-to-noise ratio or SNR for each stimulus condition. The SNR of the stimulus could then be related to the mean subject score for that condition.

1. Electrodogram

An electrodogram for the CI stimuli from each of the nine experimental conditions used in experiment I (V50 N50, V50 N145,…, etc) and the interleaved conditions (V50I N50, V145I N145, and V795I N795) was collected. This was done by sampling the output of an implant-in-a-box that was stimulated using exactly the same stimuli used for the CI vowel-in-noise task. The sampled output was stored on a PC to be used in the TIW model. A different electrodogram was collected for each of the different vowel and noise rate conditions tested. It was not necessary to collect a different electrodogram for each of the five different vowels because electrode position was not taken into account in this model. The amplitude of the pulse train collected from each electrode [Fig. 1(A)] was normalized by giving each pulse train equal RMS amplitude. This normalization procedure was based on the assumption that they all have equal perceived loudness after the loudness balancing procedure.

2. Calculating the SNR function

To integrate the energy in a given time window, a roex function (W) described by Eq. (1) was employed,

$$W(t) = (1-w)\left(1+\frac{2t}{T_p}\right)e^{(-2t/T_p)} + w\left(1+\frac{2t}{T_s}\right)e^{(-2t/T_p)} \quad (1)$$

where t is time measured relative to the center of the window, w is a weighting parameter, T_p is a time constant determining the sharpness of the central part of the window, and T_s is a time constant determining the sharpness of the skirt of the window (Moore *et al.*, 1988; Patterson, 1976). Examples of roex functions are shown on Figs. 1(A) and 1(C). In TIW model in the current study, w had a fixed value of -17 dB and Tp and Ts had a fixed ratio of 5:11. The sum of T_p and T_s gives the equivalent rectangular duration of the window, which is reported here as TIW duration. It is the only free parameter in the model. The roex TIW was moved through each pulse train in 2 ms time steps to calculate the RMS amplitude within the TIW for each electrode at each time step. By summing the RMS amplitudes for both vowel electrodes (signal) and the 4 RMS amplitudes for each noise electrode (noise), the SNR for each time step was calculated. This SNR function (i.e., the SNR at each time step) was smoothed using a three point running average and is shown in Fig. 1(B). The final SNR (SNR_{TIW}) for each stimulus condition was calculated as the mean of the 10 largest peaks in smoothed SNR function. The smoothing applied through the three point running average was necessary to remove large spikes in SNR function that skewed the final SNR estimate to high SNR values.

It was found that smoothest SNR functions could be achieved by using the smallest possible time step (i.e., one sample point) and that with this time step it was not necessary to apply the three point running average. This small time step is probably a closer representation of a physiological continuous temporal integration. However, calculation times with this small time step were long, and it was found to be more computationally efficient and result in similar SNR functions to use a 2 ms time step followed by a three point running average.

3. Estimating the TIW duration

The only unknown in the model is the time duration of the rounded exponential window, i.e., the TIW. Figure 1

shows the effect of TIW duration on SNR: A broad TIW [Fig. 1(A), gray roex] gives an SNR with lower peaks and shallower valleys [Fig. 1(B), gray line], while a narrow TIW [Fig. 1(A), black roex] gives larger peaks and larger valleys SNR [Fig. 1(B), black line]. To estimate the TIW duration, a fitting procedure, as described in the following text, was employed.

It was assumed that the SNR_{TIW} was related to mean subjects scores by a psychometric function, i.e., below a certain SNR, the subject will always score at chance level while above a certain SNR, the subject will always score 100%, and SNRs between these two levels should show a reasonably linear relationship to subject score. TIW duration was systematically changed and the resulting SNR_{TIW} plotted against mean scores. Figure 5(B) shows such a plot for a TIW of 4 ms for the CI data. The SNR_{TIW} vs mean subject score data [Fig. 5(B), dots] was fitted with a psychometric function [Fig. 5(B), line] for each TIW that we tested. The psychometric function was described by a cumulative distribution function having 4 free parameters: The mean, the standard deviation, and the two asymptotic values. Finally, to test the goodness of the fit (Q), the fraction of the variance accounted for by the fitted psychometric function was calculated using the following equation,

$$Q = 1 - \sum_{k} \frac{(Yfit_k - Ydat_k a)^2}{\sigma^2_{data}}, \quad (2)$$

where *Yfit* is the psychometric function, *Ydata* are the mean subject scores, and σ^2_{data} is the variance of the in mean subject scores. After calculating the goodness of fit resulting from each TIW tested, the TIW that resulted in a fit that accounted for the largest fraction of the variance was selected.

4. Acoustic simulation

The same model was applied to the acoustic stimuli by sliding the roex function through each noise-band [Fig. 1(C)], calculating the RMS amplitude within the roex function, and then calculating a vowel-to-noise ratio or SNR [Fig. 1(D)]. The same procedure to estimate the TIW duration was applied by finding the best fit the NH data.

G. Statistics

For the reasons described in the electrical stimulation section, the 795 pps rate was substituted with a 365 pps rate in all experiments carried out by the sole N22 subject. In all figures and statistical analysis, this condition is grouped together with the 795 pps condition of the N24 subjects. Two points provided justification for grouping these differing high rate conditions together: (1) Both high rate conditions were above the generally accepted limit of temporal pitch discrimination of 300 pps for CI users (Zeng, 2002). (2) The pattern of results from the N22 subject followed the same general trends at the N24 subjects. Data were analyzed using a repeated measures ANOVA, and *post hoc* analysis was adjusted using Bonferroni corrections. Effects were considered statistically significant if $p < 0.05$. All statistical analysis was carried in SPSS (SPSS Inc., Chicago, IL). Scores are reported as mean percentage correct ± standard deviation. On all figures, bar graphs show the mean result for all subject scores in that condition, error bars show the standard deviation, and individual subject scores are shown as gray dots.

III. RESULTS

A. Experiment I: Stimulation rate

Experiment I investigated the effects of stimulation rate on synthetic vowel-in-noise recognition in CI subjects and in NH subjects using a CI simulation.

1. CI subjects

Figure 2 shows the results for all nine rate combinations of vowel and noise in the CI subjects. In Fig. 2(A), the vowel stimulation rate is always lower or equal to the noise rate, and the CI subjects score around chance level in each condition. CI subject 5 could not score above 80% correct when the vowel was presented at 50 pps in quiet and so did not complete the three conditions shown on Fig. 2(A) (V50 N50, V50 N145, and V50 N795). The main effects of stimulation rate are apparent in Fig. 2(B): If the vowel and noise have the same stimulation rate (V145 N145), CI subjects score only slightly above chance (20% indicated by the dashed line) in the task with a mean score of 27.4 ± 2.7, indicating that it is difficult to separate the vowel from the noise. If the vowel has a higher stimulation rate than the noise (V145 N50), CI vowel-in-noise recognition improves (55.3 ± 4.3). However, if the vowel has a lower stimulation rate than the noise (V145 N795), CI subjects in general have difficulty separating the vowel from the noise (33.2 ± 11.6). The results from the other experimental conditions show the same pattern of results. In Fig. 2(C) for conditions V795 N795, subjects score around chance level but as the stimulation rate of the noise lowers to 145 and then 50 pps, scores steadily improve. When the noise and vowel had equal stimulation rate (V50 N50, V145 N145, and V795 N795), subjects tended to score slightly above chance. In these conditions, the only cue available to the subjects may have been a weak place pitch cue as the vowel electrodes were constant for each presentation of that vowel but the maskers electrodes varied. This weak place pitch cue may have been enough to push performance over chance level. A repeated measures ANOVA where vowel rate and noise rate were treated as separate factors revealed a significant effect of vowel stimulation rate [$F(2,8) = 5.8$, $p = 0.028$] and a significant effect of noise stimulation rate [$F(2,8) = 23.5$, $p > 0.001$]. There was a significant interaction between vowel and noise stimulation rate [$F(4,16) = 6.44$, $p = 0.003$]. *Post hoc* analysis indicated that the V145 N50 condition was significantly different than the V145 N145 condition [marked with an asterisk on Fig. 2(B)] and that the V795 N50 condition was significantly different than the V795 N795 condition [marked with a double asterisk on Fig. 2(C)].

The results in Fig. 2 show that the noise must have a lower stimulation rate than the vowel for the subject to be

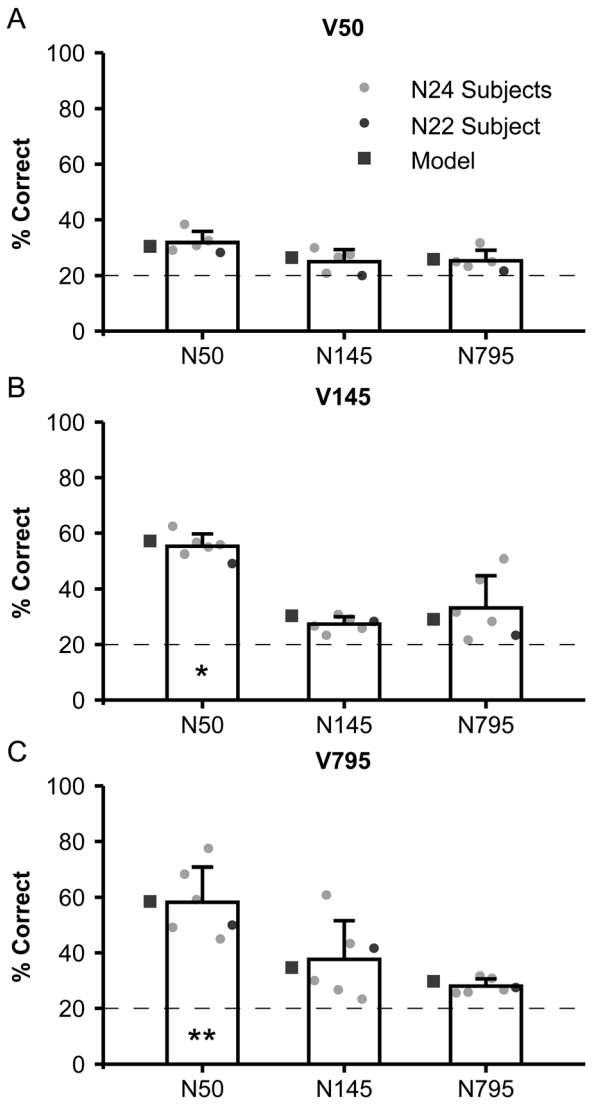

FIG. 2. Effect of stimulation rate on vowel-in-noise stimuli in the CI subjects. Unless otherwise stated in all figures: The bars indicate the mean score of all subjects and the error bars show the standard deviation; individual subject scores are shown as gray dots and predicted temporal integration window model scores as dark gray squares; dashed lines indicate chance performance. (A) The vowel is always presented at 50 pps and the noise at 50, 145, or 795 pps. (B) The vowel is always presented at 145 pps and the noise at 50, 145, or 795 pps. The asterisk indicates that the V145 N50 condition is significantly different than the V145 N145 condition. (C) The vowel is always presented at 795 pps and the noise at 50, 145, or 795 pps. The double asterisk indicates that the V795 N50 condition is significantly different than the V795 N145 condition and the V795 N795 condition.

able to separate the vowel from the noise. This effect is not reversible, i.e., the subject cannot separate a low rate vowel from a high rate noise. Thus it was postulated that the CI subjects were using the long temporal gaps in the low rate noise to listen for the vowel. To investigate this hypothesis experimentally, three interleaved conditions were tested in CI subjects 1 to 5; subject 5 did not participate in the V50I N50 condition. In the interleaved conditions, the vowel and noise had the same stimulation rate, but the vowel pulses were now shifted by half the stimulation period making the vowel pulses fall in gaps between the noise pulses. Figure 3(A) shows the percentage correct scores for the interleaved conditions, and Fig. 3(B) shows the improvement over the non-interleaved condition (i.e., V50I N50 - V50 N50, V145I N145 - V145 N145, and V795I N795 - V795 N795). The improvement seen in the V50I N50 condition over the V50 N50 condition indicates that the CI subjects may be using the temporal gaps in the masking noise to listen for the vowel. Such an improvement may also be caused by an aggregate pitch cue (McKay and McDermott, 1996) although pitch shifts caused by dual channel stimulation tend to be small (Macherey and Carlyon, 2010). There is a smaller improvement in the V145I N145 condition that may be due to the smaller the temporal gaps present in this interleaved condition. Finally, in the V795I N795 condition, there is a very weak effect of interleaving the vowel and noise. A repeated measures ANOVA was run where stimulation rate (of both the vowel and noise) and if the stimulus was interleaved or not were treated as factors. A significant effect of rate [$F(2,6) = 9.03$, $p = 0.015$] and a significant effect of interleaving the stimuli [$F(1,3) = 62.41$, $p = 0.004$] was found. The interaction between rate and interleaving the stimuli was not significant [$F(2,6) = 2.03$, $p = 0.21$]. *Post hoc* analysis comparing the interleaved and non-interleaved conditions did not reveal any significant difference between the individual means of each condition possibly due to the small sample size.

2. NH subjects

To examine whether the effects of stimulation rate were specific to electric hearing or could be generalized to acoustic hearing, the same experiment was carried out on a population of NH subjects using the CI simulation described in the methods. The results are summarized in Fig. 4. Figure 4(B) shows that for NH subjects if the vowel and noise have the same rate (V45 N45), subjects have difficulty separating the vowel from noise and score poorly on the task (30.5 ± 4.1). However, they are much better at separating the vowel from noise if the vowel is at a higher modulation rate than the noise (V45 N22, 70.0 ± 16.2). If the vowel is at lower stimulation rate than the noise (V45 N100), NH subjects find the task difficult and score poorly (28.7 ± 6.0). As with the CI subjects, when the noise and vowel had equal stimulation rate (V22 N22, V45 N45, and V100 N100), subjects tended to score slightly above chance. Again, the only cue available to the NH subjects may have been a weak place pitch cue, which may have been enough to push performance over chance level. It should be noted that because of the acoustic limitations of the CI simulation, the modulation rates used are much lower than the stimulation rates in the CI population. The higher scores measured in the NH population than in the CI population may be due to the lower rates used or they may also be due to the age difference in the two populations. Although the scores were higher in general in the NH population the pattern of results remain the same: If the masker had a lower rate than the vowel, the listener could separate the vowel from the noise, but if the masker had the same or a higher rate than the vowel, the listener found it difficult to separate the vowel from the noise. A repeated measures ANOVA where vowel rate and noise rate were treated as separate factors revealed a significant

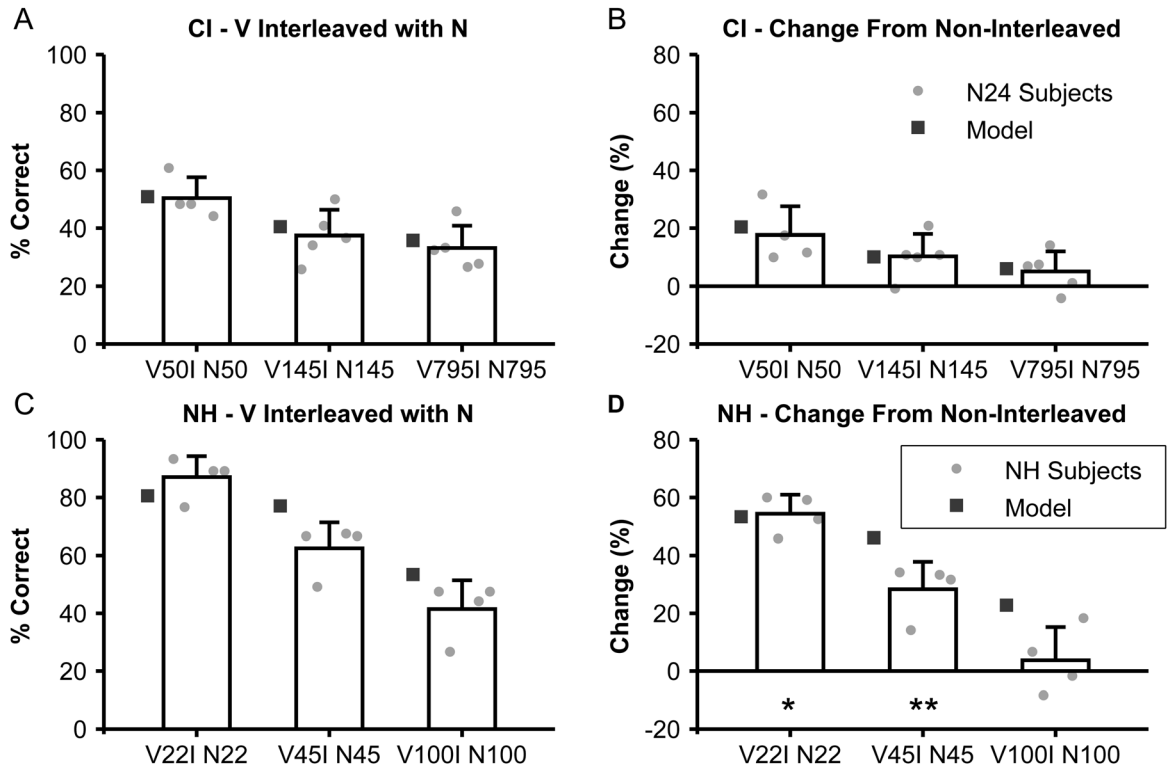

FIG. 3. Interleaved conditions: vowel and noise were presented at the same rate but vowel pulses were shifted by half a period to fall in the temporal gaps of the noise stimulus. (A) The percentage correct scores for the CI population. (B) Scores are shown as change from the non-interleaved condition for the CI population. A positive score represents a higher score in the interleaved condition. (C) and (D) show the same conditions and analysis for the NH population. The asterisk indicates that the V22 N22 condition is significantly different than the V22I N22 condition. The double asterisk indicates that the V45 N45 condition is significantly different than the V45I N45 condition.

effect of vowel stimulation rate [$F(2,14) = 5.8$, $p > 0.001$] and a significant effect of noise stimulation rate [$F(2,14) = 23.5$, $p > 0.001$]. There was a significant interaction between vowel and noise stimulation rate [$F(4,28) = 6.44$, $p > 0.001$]. *Post hoc* analysis indicated that the V45 N22 condition was significantly different than the V45 N45 and the V45 N100 condition [marked with an asterisk on Fig. 4(B)]. The V100 N22 condition was significantly different than the V100 N45 and the V100 N100 conditions [marked with a double asterisk on Fig. 4(C)]. The V100 N45 condition was also significantly different than the V100 N100 condition [marked with a triple asterisk on Fig. 4(C)]. Three interleaved conditions were also tested in four NH subjects, and the results are shown in Figs. 3(C) and 3(D). The same pattern of improved scores in the lower rate conditions that was found in the CI subjects is also evident in the NH subjects. A repeated measures ANOVA was run where stimulation rate (of both the vowel and noise) and if the stimulus was interleaved or not were treated as factors. A significant effect of rate [$F(2,6) = 66.70$, $p > 0.001$] and a significant effect of interleaving the stimuli [$F(1,3) = 52.96$, $p = 0.005$] was found. The interaction between rate and interleaving the stimuli was also significant [$F(2,6)=2.03$, $p>0.001$]. *Post hoc* analysis comparing the interleaved and non-interleaved conditions showed that the V22 N22 condition was significantly different than the V22I N22 condition [marked by an asterisk on Fig. 3(D)] and that the V45 N45 condition was significantly different than the V45I N45 condition [marked by a double asterisk on Fig. 3(D)].

B. Temporal integration window model

To examine the effect of the temporal gaps present in the stimuli, the TIW model described in Sec. II was implemented. This modeling approach does not attempt to take any temporal pitch or aggregate pitch segregation cues into account. In fact, the only cue examined in the model is that of the temporal gaps present in the stimuli.

The only free parameter in the model was TIW duration and the fitting procedure described in Sec. II was employed to find the TIW that gave the best fit to the mean subject scores from all CI data in experiment I. The same fitting procedure was applied to the NH data to estimate a TIW in the NH subjects. Figure 5(A) shows the Q factor (or goodness of fit), defined in Eq. (2), for a range of TIWs for both CI and NH subjects. The function peaks at 4 ms for the CI subjects and at 8 ms for the NH subjects. Figure 5(B) shows that for the CI subjects, the psychometric function (line) estimated for a TIW of 4 ms provides an excellent fit to the mean subject scores ($Q = 94\%$). Figure 5(C) shows that for the NH subjects the psychometric function (line) estimated for a TIW of 8 ms provides an reasonable fit to the mean subject scores ($Q = 88\%$). It should be noted that for the NH subjects, the model is less sensitive to changes in TIW duration than in the CI subjects. On Fig. 5(A) compare the relatively flat NH function between 1 and 12 ms with the large peak in the CI function. In spite of these differences, the best fit to the NH does occur for an 8 ms TIW, which is in agreement with previously reported estimates (Moore *et al.*, 1988).

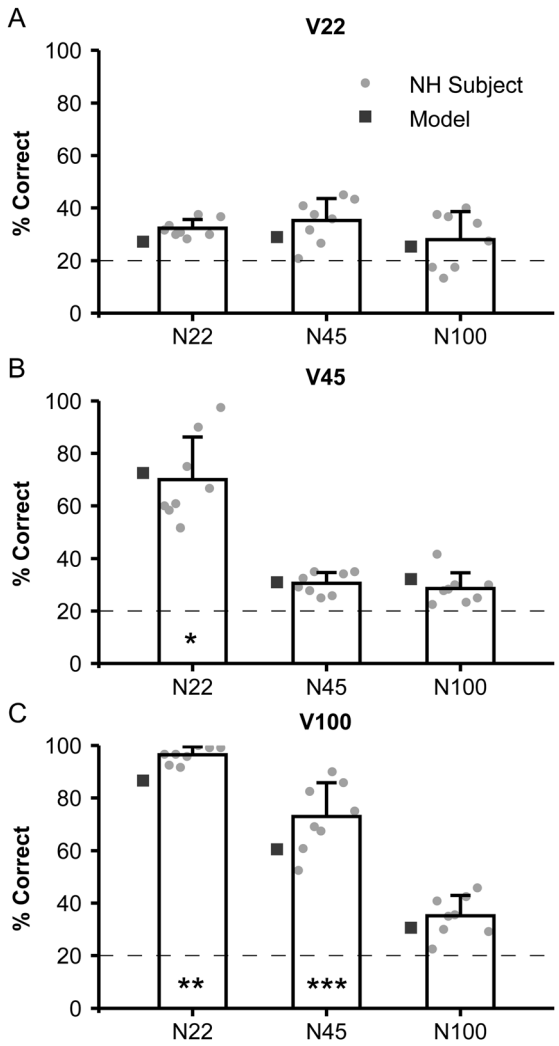

FIG. 4. Effect of modulation rate on vowel-in-noise in the NH subjects listening to stimuli presented through a noise-band vocoder. (A) The vowel is always presented at 22 Hz modulation rate and the noise modulated at 22, 45, or 100 Hz. (B) The vowel is always presented at 45 Hz modulation rate and the noise modulated at 22, 45, or 100 Hz. The asterisk indicates that the V45 N22 condition is significantly different than the V45 N45 condition and the V45 N100 condition. (C) The vowel is always presented at 100 Hz modulation rate and the noise is modulated at 22, 45, or 100 Hz. The double asterisk indicates that the V100 N22 condition is significantly different than the V100 N45 condition and the V100 N100 condition. The triple asterisk indicates that the V100 N45 condition is also significantly different than the V100 N100 condition.

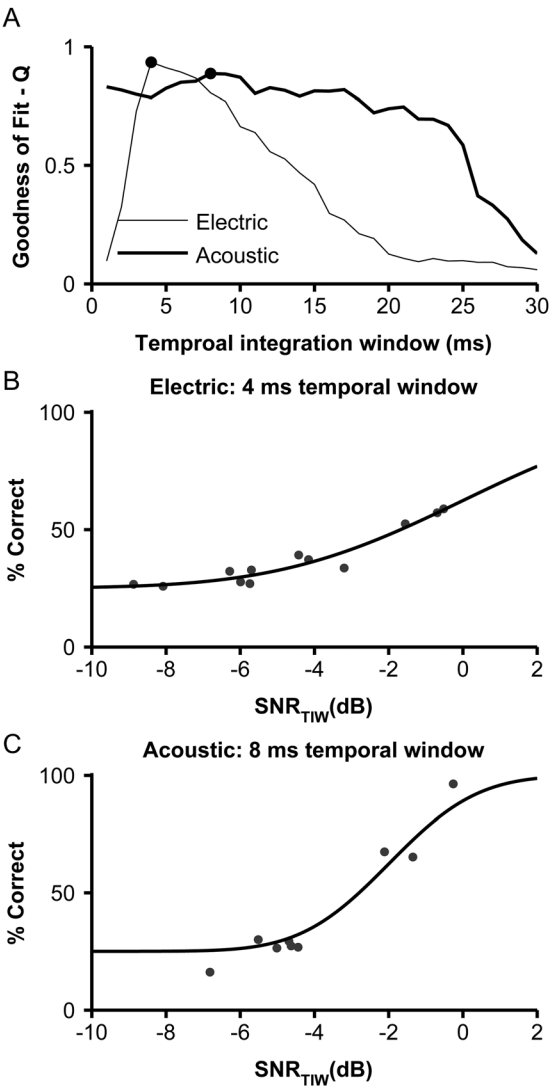

FIG. 5. A temporal integration window (TIW) model is used to predict the effect of stimulation rate on vowel-in-noise stimuli in both the CI and NH subjects. TIW duration (an unknown parameter) is estimated by assuming that the model estimated SNR is related to the mean subject scores via a psychometric function and then fitting the model to the data. (A) The goodness of fit of the psychometric function to the data (y axis) at different TIW durations (x axis). The thin line shows results from fitting the model to the CI data, and the thick line shows the same for the NH data. In both cases, the dot indicates the TIW duration at which the best fit to the data was obtained, 4 ms for the CI data and 8 ms for the NH data. (B) Fit of the psychometric function (line) to the CI data (dots) for a TIW duration of 4 ms. (C) The same plot for the NH data with a TIW of 8 ms.

However, it should be cautioned that this experiment was not specifically designed to measure differences in the TIWs of the CI and NH subjects. The TIW upon which the fitting procedure converges probably reflects the TIW that the auditory system used under those specific listening conditions and may not be representative of the minimum underlying TIW. Thus these are stimulus specific estimates of TIW and may be used to in a model to predict how listeners performed with these stimuli but should not be used in a more general stimulus independent model.

The model predicted scores for the CI subjects are shown as the gray squares on each panel of Figs. 2 and 3(A) and for the NH subjects as the gray squares on each panel of Figs. 3(B) and 4. The model gives a reasonably good prediction of the major trends observed in all conditions in experiment I in both the CI and the NH populations without taking into account any temporal or aggregate pitch mechanisms.

C. Experiment II: Onset delay

The second experiment investigated the effects of combining a stimulation rate cue with an onset delay cue.

1. CI subjects

Figures 6(A), 6(C), and 6(E) show the results for the delayed conditions for each of the nine rate combinations, and Figs. 6(B), 6(D), and 6(F) show the difference between

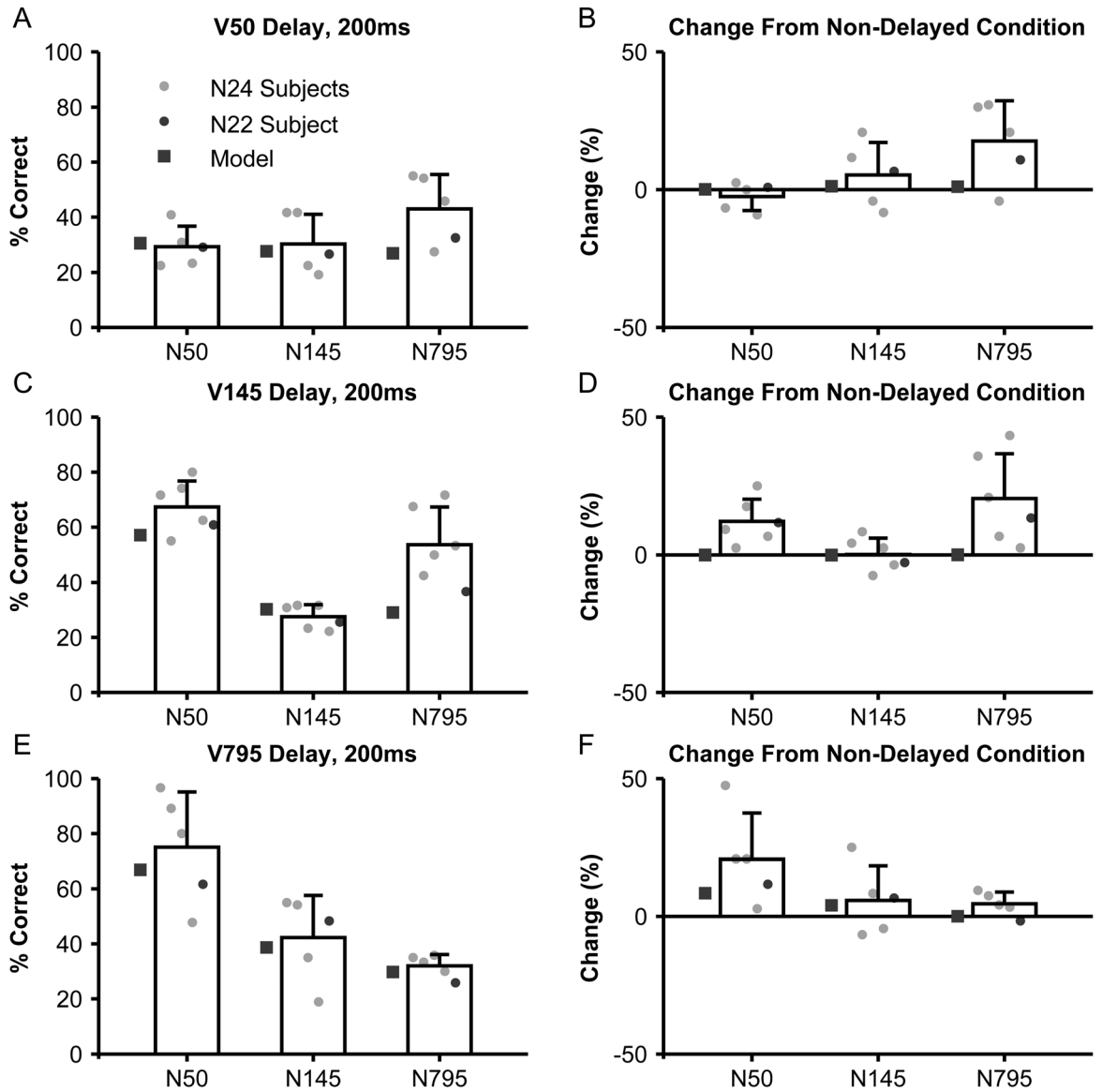

FIG. 6. Effect of onset delay in the CI subjects. The onset of the vowel is always delayed by 200 ms with respect to the noise. Scores for all nine rate combination are shown in the panels on the left-hand side. The panels on the right-hand side show the change in score which is calculated by subtracting the baseline non-delay (Fig. 2) form the delay condition for each subject for each rate combination. A positive score represents a higher score in the delay condition. (A) and (B) The vowel is always presented at 50 pps and the noise at 50, 145, or 795 pps. (C) and (D) The vowel is always presented at 145 pps and the noise at 50, 145, or 795 pps. (E) and (F) The vowel is always presented at 795 pps and the noise at 50, 145, or 795 pps.

the non-delay and delay condition for each of the nine rate combinations. A positive value on Figs. 6(B), 6(D), and 6(F) represents a higher score in the delay condition (e.g., V50D N145 minus V50 N145). The results reveal an interesting interaction between the onset delay cue and the rate cue. The pattern of results is apparent in Fig. 6(D): If the vowel was delayed by 200 ms but still had the same rate as the noise, no improvement was observed (V145D N145, 0.19 ± 5.8 improvement). However, an onset delay did improve synthetic vowel-in-noise recognition if the vowel and noise were presented at a different stimulation rates (V145D N50, 12.1 ± 8.1 improvement; V145D N795, 20.4 ± 16.3 improvement). The same pattern of results is observed in Figs. 6(D) and 6(F). In experiment I, where vowel and noise were presented with the same onset time, to observe an effect of rate, the vowel had to be at a higher stimulation rate than the noise. However, in experiment II, where the onset of the vowel was delayed with respect to the noise, an effect of rate was observed if the vowel was at a higher or a lower rate than the noise, but not if they had the same rate.

To compare the effects of delay and the different combinations of stimulation rate on vowel-in-noise recognition, a repeated measures ANOVA was used. The mean percentage correct scores from experiment I and experiment II were classified according to two grouping factors: (1) onset delay, having two levels: Delay or non-delay and (2) the vowel and noise stimulation rate, having nine levels: V50 N50, V50 N145, V50 N795, V145 N50, V145 N145,…, etc. The main effect of delay was significant [$F(1,4) = 7.85$, $p = 0.049$] as was the main effect of rate difference [$F(2,32) = 16.79$, $p < 0.001$]. The repeated measures ANOVA indicated that was a significant interaction between onset delay and stimulation rate [$F(8,32) = 3.84$, $p = 0.003$]. *Post hoc* analysis comparing the delayed and non-delayed conditions did not

reveal any significant differences between the individual means of each condition.

2. NH subjects

The same experiment was repeated in the NH subjects, but it was found that the onset delay affected synthetic vowel recognition differently than in the CI subjects. Figures 7(A), 7(C), and 7(E) show the results for the delayed conditions for each of the nine rate combinations, and Figs. 7(B), 7(D), and 7(F) show the difference in the delay and non-delay conditions for each of these experimental conditions. In general, delaying the onset of the vowel improved performance across all conditions. Figure 7 shows that in general the onset delay made the vowel-in-noise task easier for the subject even if the vowel and noise shared the same rate, which was not the case for the CI subjects. As with the CI subjects, a repeated measures ANOVA was used to compare the effects of delay and the stimulation rate on vowel-in-noise recognition. In the V100 N22 condition, most subjects already scored close to 100%. This ceiling effect made it difficult for the subjects to improve in the V100D N22 condition. To limit the impact of this ceiling effect on the statistical analysis, this condition was removed from the repeated measures ANOVA. The main effect of delay was significant [$F(1,3) = 89.33$, $p = 0.003$] as was the main effect of rate difference [$F(7,21) = 14.95$, $p < 0.001$]. The repeated measures ANOVA indicated that there was no interaction between onset delay and stimulation rate difference [$F(7,21) = 86.51$, $p = 0.103$]. *Post hoc* analysis comparing the delayed and non-delayed

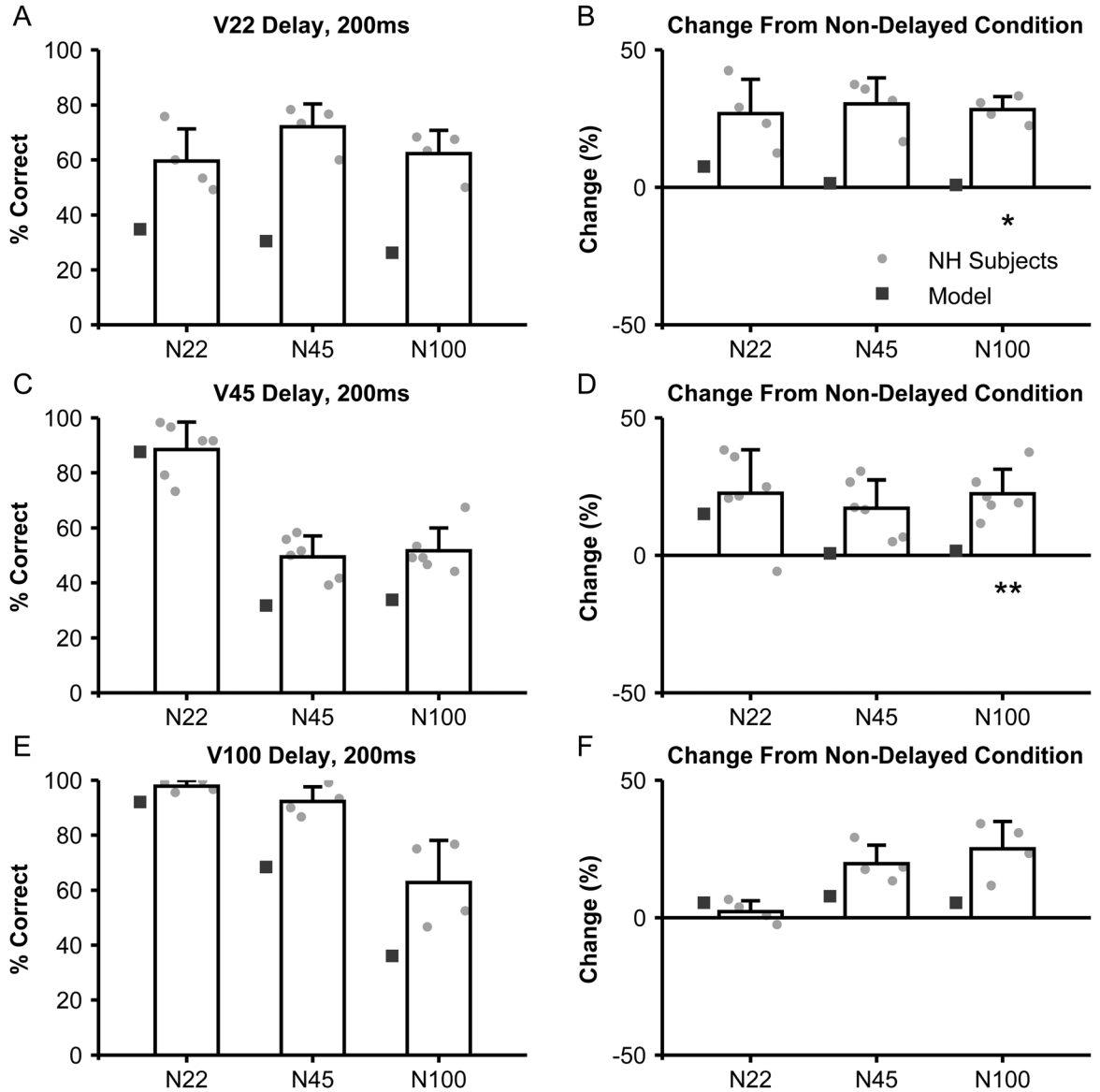

FIG. 7. Effect of onset delay in the NH subjects. The onset of the vowel is always delayed by 200 ms with respect to the noise. Scores for all nine rate combination are shown in the panels on the left-hand side. The panels on the right-hand side show the change in score which is calculated by subtracting the baseline non-delay (Fig. 4) form the delay condition for each subject for each rate combination. A positive score represents a higher score in the delay condition. (A) and (B) The vowel is always presented at 22 Hz modulation rate and the noise at 22, 45, or 100 Hz. The asterisk indicates that the V22 N100 condition is significantly different than the V22D N100 condition. (C) and (D) The vowel is always presented at 145 Hz modulation rate and the noise at 22, 45, or 100 Hz. The double asterisk indicates that the V45 N100 condition is significantly different than the V45D N100 condition. (E) and (F) The vowel is always presented at 795 Hz modulation and the noise at 22, 45, or 100 Hz.

conditions revealed that the V22 N100 condition was significantly different than the V22D N100 condition [marked by an asterisk on Fig. 7(B)] and that the V45 N100 condition was significantly different than the V45D N100 condition [marked by a double asterisk on Fig. 7(D)].

One possible explanation for the improved scores in both the CI and NH subjects may be that the onset delay changed the timing between the vowel and noise pulses and led to more vowel pulses occurring in the temporal gaps in the noise. To check this possibility, the TIW model was run on the CI and NH stimuli used in experiment II. The predicted values were largely similar to the model predictions for the non-delay conditions. The gray squares on Figs. 6 and 7 show the difference between the delay and non-delay model predictions and are normally close to zero. Thus the TIW model fails to explain the pattern of improved scores seen in either population, suggesting that the effects are driven by the onset delay itself and not by any changes in the temporal gaps between the stimuli introduced by the delay.

IV. DISCUSSION

The present study showed that in electric hearing, synthetic vowel-in-noise recognition is improved when the synthetic vowel and the noise have a different stimulation rate or when the onset of the synthetic vowel is delayed with respect to the noise. Specifically, if the vowel and noise had a simultaneous onset, the vowel needed to have higher stimulation rate than the noise to improve segregation. However, when an onset delay was present, segregation was improved if the vowel had a higher or lower stimulation rate than the noise. In the NH experiment, synthetic vowel recognition also improved when the vowel had a higher rate than the noise, although the rates used in NH experiment were lower than in than in the CI experiment. Section A compares these finding with the results from a previous study (Carlyon *et al*., 2007) that found no effect of stimulation rate. Sections B and C put forward an interpretation of these finding in terms of a glimpsing cue and a stimulation rate cue. Section D discusses how these results could impact on the design of speech processing strategies that perform better with speech in noise.

A. Comparison with previous study

Carlyon *et al*. (2007) investigated the effects of stimulation rate, onset delay, and asynchrony (similar to our interleaved condition) on sound source segregation. They found that onset delay improved sound source segregation in electric hearing, but their results differed from this study in that they did not find an effect of rate or asynchrony. There are a number of possible explanations for these differing findings. One difference, likely to have a significant effect on the results, is that this study used a wider range of stimulation rates (50, 145, 795 pps), whereas Carlyon *et al*. tested a narrower range (77 and 100 pps). Figure 2(C) clearly shows that the task gets easier as the stimulation rate of the noise becomes lower, i.e., the size of the effect is proportional to the rate difference. It is therefore likely that for the narrower range of rates used in Carlyon *et al*., this effect would not have been apparent. We also found that the stimulation rate segregation cue was strongest when it was combined with an onset delay cue; Carlyon *et al*. did not test this condition. The stimulation modes used also differed: Carlyon *et al*. stimulated in bipolar mode (BP+1) with 100 μs per phase, 43 μs interphase gap, whereas this study used monopolar mode (MP1+2) with 50 μs per phase and a 10 μs interphase gap. The tasks used were also very different. Carlyon *et al*. used a two-interval forced choice threshold detection task where one target pulse train was embedded in a masker consisting of three pulse trains. The current study used a five alternative forced choice task where two pulse trains represented a synthetic vowel and four other pulse trains as masking noise. The lexical meaning assigned to the stimulus may have invoked more "top-down" segregation cues, but these are likely to be weak given the short duration and highly synthesized nature of the stimuli. A final explanation for the different findings of the two studies may be subject selection. In Carlyon *et al*. (2007), all subjects use the N24 implant, but they do not report that the subjects were considered good users. In contrast, this study specifically selected subjects who should be good at the task (>80% vowel-in-quiet). The rationale behind the subject selection in the current study was to test the upper limits of sound source segregation in electric hearing.

B. Mechanism 1: Glimpsing

Results showed that a difference in stimulation rate between the synthetic vowel and noise can improve recognition. Two distinct mechanisms, a glimpsing cue (Howard-Jones and Rosen, 1993; Miller, 1950) and a temporal pitch cue, may provide an explanation for some of these results. The glimpsing cue is discussed in this section, the temporal pitch cue in the following.

When the stimulation rate of the noise is lower than the vowel or when the vowel and noise have an equal rate but the pulses are interleaved, the temporal gaps present may allow the listeners to make use of a glimpsing cue. For this cue to be effective, the temporal gaps in noise must be long enough for the CI subject to catch an auditory glimpse of the vowel. Results from the vocoder simulation study showed that NH listeners may make use of a similar glimpsing cue, although the rates used in the NH study were lower than in the CI study. Two recent studies by Li and Loizou (2007, 2008) investigated factors influencing the glimpsing of speech in noise, focusing on the importance of a glimpsing cue for CI users with residual low frequency acoustic hearing. These studies showed that useful glimpsing cues can be present in restricted time-frequency windows and that the most important spectral region for glimpsing cues was first and second formant region.

The TIW model worked by sliding a roex function through each channel of the electrodogram (or noise-band for the vocoder simulation) and then calculating the local RMS amplitude per channel. This allowed the calculation of a vowel-to-noise ratio or SNR that was different for each of the vowel and noise rate combinations. The SNR_{TIW} for each stimulus condition is clearly related to the mean subject

score in that condition by a psychometric function for both the CI and the NH subjects [Figs. 5(B) and 5(C), respectively]. The model is based on the idea that the subject listens to and integrates the information within short temporal windows (Moore *et al.*, 1988; Munson, 1947; Zwislocki, 1960) and so calculates a kind of running SNR [Fig. 1(B)]. When there is a temporal gap in the noise there is a localized peak in the SNR, which provides a glimpse of the vowel stimulus.

This approach did not model the possible effects of temporal pitch cues (discussed in the following text) or aggregate pitch cues caused by the summation of information on two neighboring electrode channels (Macherey and Carlyon, 2010; McKay and McDermott, 1996). The model only accounted for glimpsing cues caused by temporal gaps in the masker. In spite of the simplicity of the approach, the model showed excellent agreement with the mean subject scores. This provides evidence for the hypothesis that listeners make use of the temporal gaps in the low rate maskers to segregate the vowel from the noise. It also suggests that any temporal or aggregate pitch cues that may be present are probably weak and do not have a significant impact on performance in the listening conditions tested in experiment I.

C. Mechanism 2: Stimulation rate

In electric hearing, users can discriminate differences in temporal pitch between stimuli up to around 300 pps (Zeng, 2002). The 50 and 145 pps stimuli used in this study will elicit distinctly different temporal pitch percepts. The 795 pps stimuli will be discriminable from the 50 and 145 pps stimuli but may not elicit a distinct temporal pitch percept. Interestingly, a recent study suggests that this limit of temporal pitch may be extended with training beyond 300 pps (Goldsworthy and Shannon, 2011).

In the V145 N50 condition, the vowel and noise elicited different temporal pitch percepts, and the subjects were able to segregate the vowel from the noise. However, when this condition was reversed, V50 N145, subjects were able not able to perform the task. Thus in this condition, there is a difference in temporal pitch but the subject is unable to perform the task. This indicates that temporal pitch may not be a strong segregation cue. A glimpsing cue, as outlined in the section in the preceding text, provides a better explanation for the pattern of results observed.

The results from experiment II where a stimulation rate cue was combined with an onset delay cue were not well explained by the TIW model. Figures 6 and 7 show the mismatch between the TIW model estimated scores (gray squares) and the mean subject scores for both the CI and NH populations. Adding an onset delay cue does improve synthetic vowel-in-noise segregation in both the NH and CI populations. However, in the CI population, this effect was only present if the synthetic vowel had a different (higher or lower) stimulation rate than the noise. A weak stimulation rate cue that only becomes effective when combined with a stronger onset delay cue may offer some explanation for the pattern of results. This interaction between rate and onset delay was not seen in the NH population, where an onset delay improved synthetic vowel-in-noise recognition for all rates. This difference between the NH and CI populations may be caused by a more synchronized pattern of phase locking in the CI subjects. It is known that electrical stimulation of the auditory nerve with pulse trains leads to an enhanced phase-locking (Litvak *et al.*, 2001) compared to acoustic stimulation (Rose *et al.*, 1967). The enhanced phase locking may serve as a grouping cue, making it difficult to separate vowel electrodes from noise electrodes with the same stimulation rate, even when the vowel has an onset delayed cue. However, some the differences observed between the CI and NH populations may also be due to the different range of rates used in both experiments.

D. Improved speech processing strategies

One of the most common complaints from CI users is that they cannot follow a conversation if there is background noise present. A speech processing strategy that allowed them to separate speech from noise would be an enormous improvement. The results of the present study indicate that CI subjects can use differences in stimulation rate to separate synthetic vowels from noise. However, current speech processing strategies are not designed to encode these kinds of cues. A speech processing strategy in which amplitude and slow varying frequency modulations (FM) are separately extracted from a small number of frequency bands has been proposed by Nie *et al.* (2005) and Zeng *et al.* (2005). Using a vocoder simulation in NH listeners, they found that adding the FM cue gave a large improvement for speech-in-noise recognition. In a variable rate speech processing strategy, FM in the acoustic signal could be encoded as a channel specific RM in the electrical stimulation. Future speech processing strategies need to explore these differences in rate and modulation for optimal speech in noise recognition

ACKNOWLEDGMENTS

We would like to thank all of our subjects for their dedication. We thank Dr. Duo Zhang and Dr. Thomas Lu for their technical assistance. This work was partly supported by a Marie-Curie International Outgoing Fellowship (IOF 253047), the Hewitt Foundation, and the National Institutes of Health, National Institute on Deafness and Other Communication Disorders Grant Nos. R01-DC008858 and P30-DC008369.

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.

Balkany, T., Hodges, A., Menapace, C., Hazard, L., Driscoll, C., Gantz, B., Kelsall, D., Luxford, W., McMenomy, S., Neely, J. G., Peters, B., Pillsbury, H., Roberson, J., Schramm, D., Telian, S., Waltzman, S., Westerberg, B., and Payne, S. (**2007**). "Nucleus Freedom North American clinical trial," Arch. Otolaryngol. Head Neck Surg. **136**, 757–762.

Bregman, A. S. (**1990**). *Auditory Scene Analysis: The Perceptual Organization of Sound* (MIT Press, Cambridge, MA), pp. 1–773.

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. (**2009**). "Auditory stream segregation in cochlear implant listeners: Measures based on temporal discrimination and interleaved melody recognition," J. Acoust. Soc. Am. **126**, 1975.
Culling, J. F., and Darwin, C. J. (**1993**). "Perceptual separation of simultaneous vowels: Within and across-formant grouping by F0," J. Acoust. Soc. Am. **93**, 3454–3467.
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. (**2005**). "Loudness growth in cochlear implants: Effect of stimulation rate and electrode configuration," Hear. Res. **202**, 55–62.
Gaudrain, E., Grimault, N., Healy, E. W., and Béra, J.-C. (**2008**). "Streaming of vowel sequences based on fundamental frequency in a cochlear-implant simulation," J. Acoust. Soc. Am. **124**, 3076.
Goldsworthy, R., and Shannon, R. V. (**2011**). "Improvements in rate discrimination after training in adult cochlear implant recipients," *Conference on Implantable Auditory Prostheses*, p. 160.
Hancock, K. E., Noel, V., Ryugo, D. K., and Delgutte, B. (**2010**). "Neural coding of interaural time differences with bilateral cochlear implants: Effects of congenital deafness," J. Neurosci. **30**, 14068–14079.
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.
Hong, R. S., and Turner, C. W. (**2009**). "Sequential stream segregation using temporal periodicity cues in cochlear implant recipients," J. Soc. Acoust. Am. **126**, 291.
Howard-Jones, P. A., and Rosen, S. (**1993**). "Uncomodulated glimpsing in 'checkerboard' noise," J. Soc. Acoust. Am. **93**, 2915–2922.
Li, N., and Loizou, P. C. (**2007**). "Factors influencing glimpsing of speech in noise," J. Soc. Acoust. Am. **122**, 1165–1172.
Li, N., and Loizou, P. C. (**2008**). "A glimpsing account for the benefit of simulated combined acoustic and electric hearing," J. Soc. Acoust. Am. **123**, 2287–2294.
Litovsky, R. Y., Jones, G. L., Agrawal, S., and Van Hoesel, R. (**2010**). "Effect of age at onset of deafness on binaural sensitivity in electric hearing in humans," J. Soc. Acoust. Am. **127**, 400–414.
Litvak, L., Delgutte, B., and Eddington, D. (**2001**). "Auditory nerve fiber responses to electric stimulation: Modulated and unmodulated pulse trains," J. Soc. Acoust. Am. **110**, 368.
Luo, X., and Fu, Q.-J. (**2009**). "Concurrent-vowel and tone recognitions in acoustic and simulated electric hearing," J. Soc. Acoust. Am. **125**, 3223–3233.
Luo, X., Fu, Q.-J., Wu, H.-P., and Hsu, C.-J. (**2009**). "Concurrent-vowel and tone recognition by Mandarin-speaking cochlear implant users," Hear. Res. **256**, 75–84.
Macherey, O., and Carlyon, R. P. (**2010**). "Temporal pitch percepts elicited by dual-channel stimulation of a cochlear implant," J. Acoust. Soc. Am. **127**, 339–349.
McKay, C. M., and McDermott, H. J. (**1996**). "The perception of temporal patterns for electrical stimulation presented at one or two intracochlear sites," J. Acoust. Soc. Am. **100**, 1081–1092.
Miller, G. A. (**1950**). "The intelligibility of interrupted speech," J. Soc. Acoust. Am. **22**, 167.
Moore, B. C., Glasberg, B. R., Plack, C. J., and Biswas, A. K. (**1988**). "The shape of the ear's temporal window," J. Soc. Acoust. Am. **83**, 1102–1116.
Munson, W. A. (**1947**). "The growth of auditory sensation," J. Soc. Acoust. Am. **19**, 584.
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. Soc. Acoust. Am. **113**, 961–968.
Nie, K., Stickney, G., and Zeng, F.-G. (**2005**). "Encoding frequency modulation to improve cochlear implant performance in noise," IEEE Trans. Biomed. Eng. **52**, 64–73.
Patterson, R. D. (**1976**). "Auditory filter shapes derived with noise stimuli," J. Soc. Acoust. Am. **59**, 640–654.
Qin, M. K., and Oxenham, A. J. (**2005**). "Effects of envelope-vocoder processing on F0 discrimination and concurrent-vowel identification," Ear Hear. **26**, 451–460.
Rose, J. E., Brugge, J. F., Anderson, D. J., and Hind, J. E. (**1967**). "Phase-locked response to low-frequency tones in single auditory nerve fibers of the squirrel monkey," J. Neurophysiol. **30**, 769–793.
Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., and Ekelid, M. (**1995**). "Speech recognition with primarily temporal cues," Science, New York **270**, 303–304.
Stickney, G. S., Zeng, F.-G., Litovsky, R., and Assmann, P. (**2004**). "Cochlear implant speech recognition with speech maskers," J. Soc. Acoust. Am. **116**, 1081–1091.
Wilson, B. S., Lawson, D. T., Finley, C. C., and Wolford, R. D. (**1991**). "Coding strategies for multichannel cochlear prostheses," Am. J. Otol. **12** Suppl., 56–61.
Wygonski, J., and Robert, M. E. (**2002**). "HEI Nucleus research interface—HEINRI specification," House Ear Institute Internal Material (House Ear Institute, Los Angeles, CA).
Zeng, F. G. (**2002**). "Temporal pitch in electric hearing," Hear. Res. **174**, 101–106.
Zeng, F., and Shannon, R. (**1994**). "Loudness-coding mechanisms inferred from electric stimulation of the human auditory system," Science **264**, 564–566.
Zeng, F.-G., Nie, K., Stickney, G. S., Kong, Y.-Y., Vongphoe, M., Bhargave, A., Wei, C., and Cao, K. (**2005**). "Speech recognition with amplitude and frequency modulations," Proc. Natl. Acad. Sci. U.S.A. **102**, 2293–2298.
Zeng, F.-G., Rebscher, S., Harrison, W. V., Sun, X., and Feng, H. (**2008**). "Cochlear implants: System design, integration and evaluation," IEEE Rev. Biomed. Eng. **1**, 115–142.
Zwislocki, J. (**1960**). "Theory of temporal auditory summation," J. Soc. Acoust. Am. **32**, 1046.