Speech Perception as a Function of Electrical Stimulation Rate: Using the Nucleus 24 Cochlear Implant System

Andrew E. Vandali, Lesley A. Whitford, Kerrie L. Plant, and Graeme M. Clark

Objective: To investigate the effect of varying electrical stimulation rate on speech comprehension by cochlear implant users, while keeping the number of stimulated channels constant.

Design: Three average rates of electrical stimulation, 250, 807, and 1615 pulses per second per channel (pps/ch), were compared using a speech processing strategy that employed an electrode selection technique similar to that used in the Spectral Maxima Sound Processor strategy (McDermott, McKay, & Vandali, 1992; McDermott & Vandali, Reference Note 1; McKay, McDermott, Vandali, & Clark, 1991) and the Spectral Peak strategy (Skinner et al., 1994; Whitford et al., 1995). Speech perception tests with five users of the Nucleus 24 cochlear implant system were conducted over a 21-wk period. Subjects were given take-home experience with each rate condition. A repeated ABC evaluation protocol with alternating order was employed so as to account for learning effects and to minimize order effects. Perception of open-set monosyllabic words in quiet and open-set sentences at signal to noise ratios ranging from +20 to 0 dB, depending on the subject's ability, were tested. A comparative performance questionnaire was also administered.

Results: No statistical differences in group performance between the 250 and 807 pps/ch rates were observed in any of the speech perception tests. However, significantly poorer group performance was observed for the 1615 pps/ch rate for some tests due predominantly to the results of one subject. Analysis of individual scores showed considerable variation across subjects. For some subjects, one or more of the three rate conditions evaluated provided benefits on some speech perception tasks. The results of the comparative performance questionnaire indicated a preference for the 250 and 807 pps/ch rates over the 1615 pps/ch rate for most listening situations.

Conclusions: For the speech processing strategy, implant system, and subjects evaluated in this study, the group results indicated that the use of electrical stimulation rates higher than 250 pps/ch (up to 1615 pps/ch) generally provided no significant improvement to speech comprehension. However, individual results indicated that perceptual benefits could be obtained by adjusting rate of stimulation optimally to suit each subject. Results from one subject, together with tinnitus problems arising from high-rate stimulation for another subject, indicated that high rates of stimulation may in fact be undesirable for some subjects.

(Ear & Hearing 2000;21;608–624)

Advances in cochlear implant and speech processing technology have provided the means for electrical stimulation of the cochlea at higher rates and with increased numbers of channels and electrodes. The challenge we face now is to determine how to best make use of this improved transmission bandwidth. Use of higher electrical stimulation rates, above that of 250 pulses per second per channel (pps/ch) as employed in the Spectral Maxima Sound Processor (SMSP) strategy (McDermott et al., 1992; McDermott & Vandali, Reference Note 1; McKay et al., 1991) and the Spectral Peak (SPEAK) strategy (Skinner et al., 1994; Whitford et al., 1995) may provide improvements to speech perception. Possible reasons for this expectation are that higher rates may provide increased temporal detail in the stimulus signal and/or may provide neural firing patterns that closely approximate patterns from acoustic stimulation.

The response of auditory neurons to acoustic stimulation is quite different to that of electrical stimulation, partially due to differences in the transduction mechanisms. Cochlear mechanics and hair cell transduction involved with acoustic stimulation are bypassed by direct electrical activation of auditory neurons. The probabilistic nature of hair cell transduction (owing to the nature of neurotransmitter availability and release) gives rise to a stochastic relationship between acoustic stimulation and neural response. This allows population responses to show phase locking for frequencies of up to approximately 2 to 4 kHz, which is well beyond the firing rate of individual neurons (Rose, Brugge, Anderson, & Hind, 1968). In contrast, neural responses elicited electrically are governed more by interactions between the stimulus intensity and the refractory properties of the neurons. A more deterministic

response behavior with a high degree of phase locking and entrainment for low rates of stimulation (up to approximately 800 pps) is observed, as shown in data from animal studies (Hartmann & Klinke, 1990; Javel, Tong, Shepherd, & Clark, 1987; Kiang, Eddington, & Delgutte, 1979; Van den Honert & Stypulkowski, 1987). At higher rates of stimulation, where relative refractory effects become more dominant (800 to 2000 pps), entrainment decreases and population responses, which are based on averages of single-unit responses, exhibit poorer phase locking than that seen at lower rates to each burst in the stimulus pulse train (Dynes, & Delgutte, 1992; Parkins, 1989). This is most noticeable during approximately the first 50 msec of the pulse train, with the exception of the onset response, which is highly synchronized to the initial stimulus pulse. At higher rates again (2000 to 4000 pps), measurements of intracochlear evoked potentials in cat and human have shown behavior more similar to that of the acoustic case (Matsuoka, Abbas, Rubinstein, & Miller, Reference Note 2; Wilson, Finley, Lawson, & Zerbi, 1997). Other studies have suggested that similar response patterns might be obtained by the use of high-rate conditioning stimuli or by the addition of noise to a lower rate stimulus signal (Matsuoka et al., Reference Note 2; Morse & Evans, 1996; Moss, Chioutan, & Klinke, 1996; Rubinstein, Wilson, Finley, & Abbas, 1999). It is therefore possible that higher electrical stimulation rates can provide response patterns that are more stochastic in nature, but the importance of this to speech perception has yet to be determined. However, it may be that optimal rates required to achieve stochastic response patterns may be higher than those used in the study that is described in this paper.

Although it would seem from the temporal response patterns of auditory neurons to acoustic stimulation that relatively high-frequency information such as fine temporal structure in the speech envelope can be coded, the relevance of such information to speech perception is not well understood. Numerous speech perception studies with normally hearing listeners have demonstrated that low-frequency temporal envelope information in speech provides adequate temporal cues for speech perception and that finer temporal structure such as the fundamental frequency (F0) adds little advantage to comprehension (Drullman, 1995; Shannon, Zeng, Kamath, Wygonski, & Ekelid, 1995; Van Tasell, Greenfield, Logemann, & Nelson, 1992; Van Tasell, Soli, Kirby, & Widin, 1987). However, the relevance of these findings to speech perception in noise may not be straightforward. Perception of fine temporal structure such as F0 can assist in tasks such as speaker identification and intonation, as has been shown for normal-hearing listeners (Carrell, Reference Note 3; Liberman, 1967) and for cochlear implant users (McKay & McDermott, 1993; Tong, Vandali, Busby, Harrison, Van Hoesel, & Clark, Reference Note 4). It may subsequently be the case that better separation of speech from competing background noise and/or voices may be assisted by the listener's ability to attend to the voice of the target speaker. This idea is supported by data from Brokx and Nooteboom (1982), who showed that for Dutch nonsense sentences presented in background speech, intelligibility for normally hearing listeners varied as a function of F0 differences between the target and background speech. Best intelligibility was observed for F0 differences of approximately three semitones. These results were attributed to a decrease in the probability of perceptual fusion of pitch and attention switching to the wrong voice. Further evidence that F0 differences aid the perceptual segregation of concurrent sounds has been obtained in studies by Assmann and Summerfield (1990), and Scheffers (Reference Note 5). In their studies, perception of two simultaneously presented vowels was explored as a function of F0 difference between the vowels. A marked increase in intelligibility was observed as the F0 difference increased from 0 to 0.5 semitones. It may be the case, however, that for a multitude of background speakers covering a range of F0s, the effect may be somewhat reduced because competing F0 trajectories can overlap with those of the target speaker. Furthermore, for positive signal to noise ratios (SNRs) typical of many listening situations, other cues in the signal independent of F0, such as temporal envelope and spectral amplitude cues in each channel, are likely to assist in speaker separation.

Psychoacoustic studies using amplitude-modulated noise with normally hearing listeners have demonstrated that pitch information coded in temporal patterns of acoustic stimulation can be perceived well for frequencies up to approximately 300 Hz (Burns & Viemeister, 1976, 1981; Houtsma, Wicke, & Ordubadi, 1980). Rate pitch identification studies with cochlear implant users also have demonstrated a similar limit known as the "pitch saturation limit," which typically occurs at rates of 200 to 500 pps but can be as high as 1000 pps for some subjects (Eddington, Dobelle, Brackmann, Mladejovsky, & Parkin, 1978; Shannon, 1983; Simmons, 1966; Simmons, Mathews, Walker, & White, 1979; Tong, Blamey, Dowell, & Clark, 1983; Townshend, Cotter, Van Compernolle, & White, 1987). These results indicate that fine temporal detail in the F0 range can be perceived using electrical stimulation. The Multipeak strategy as employed in the Nucleus MSP speech processor (Skinner et al. 1991) coded F0

TABLE 1. Subject details: age in years at commencement of study, duration of implantation (in months) at commencement of study, number of years of severe-to-profound sensorineural deafness at time of implantation, etiology of deafness, and 3-mo postoperative CNC word score using the SPEAK strategy.

Subject	Age (yr)	Months Implanted	Duration of Severe-to-Profound Deafness (yr)	Etiology	CNC Word Test (3-mo Postoperative)
1	63	8	0.75	Unknown	14%
2	68	10	0.75	Unknown	38%
3	44	6	<10	Otosclerosis	22%
4	70	5	5	Chronic otitis media	20%
5	46	18	0.5	Unknown progressive	24%
6	57	4	10	Unknown	34%

via rate of stimulation. For strategies such as the SMSP, SPEAK, and Continuous Interleaved Sampling (CIS) strategy (Wilson, Finley, Lawson, Wolford, Eddington, & Rabinowitz, 1991), which employ a constant rate of stimulation, fine temporal detail (F0) in the speech signal is coded by way of amplitude modulation of the stimulus/carrier signal, as has been demonstrated using the SMSP strategy by McKay and McDermott (1993) and Tong et al. (Reference Note 4). More recent psychophysical studies by Busby, Tong, and Clark (1993) and McKay, McDermott, and Clark (1994), examining the detection and pitch perception of amplitude-modulated current pulse trains, have shown that detection of the modulation frequency (for frequencies in the F0 range) can in fact be carried out reliably only when the carrier frequency is at least four times that of the modulation frequency. Typically, F0 can vary from approximately 80 to 350 Hz; thus, a stimulation rate of about 320 to 1400 pps/ch or higher may be desirable.

A number of speech perception studies exploring rate of stimulation with the CIS strategy have been described in a series of NIH quarterly progress reports. Early work by Wilson, Lawson, Finley, and Zerbi (Reference Note 6) with one subject suggested that rates of stimulation above the "pitch saturation limit" provided improved speech perception results and eliminated pitch percepts associated with fixed low-rate stimulation. Further results (Wilson, Lawson, Zerbi, & Finely, Reference Note 7) showed that best fitting of CIS processor parameters (based on a closed-set consonant identification task) varied widely among a group of 11 subjects. The number of channels, pulse duration, rate of stimulation and integration filter cutoff frequency were all optimized on an individual basis. Optimal rates of stimulation ranged from 500 to 2941 pps/ch. In another report by Wilson, Lawson, and Zerbi (Reference Note 8), manipulation of pulse duration and stimulation rate was explored more systematically, but with only one subject. Results from closed-set consonant identification tests indicated that best performance was

obtained with short-duration pulses, presented at stimulation rates of 833 pps/ch or lower. In a later report by Lawson, Wilson, Zerbi, and Finley (Reference Note 9), manipulation of a greater number of coding parameters was explored with a group of five subjects implanted with a percutaneous version of the Nucleus 22 electrode array. Parameters of CIS that were varied included rate of stimulation (250, 833, 2525 pps/ch) and upper frequency range (5500, 9500 Hz). An "n-of-m" channel selection strategy (which employs an electrode selection scheme similar to the SMSP and SPEAK strategies) was also investigated. Results for the CIS strategy for closed-set consonant identification in quiet generally showed best performance for the extended frequency range (i.e., 9500 Hz) and for stimulation rates of 833 pps/ch. Results for the "n-of-m" strategy also generally provided best performance using a stimulation rate of 833 pps/ch compared with 250 pps/ch. Unfortunately, no take-home experience with each strategy was provided and order of evaluation of strategies were not taken into consideration in the analysis of these results. In addition, some variability in perceptual performance among subjects for the range of parameters tested was observed.

The aim of this investigation was to study the effects of electrical stimulation rate on speech perception with an experimental design that provided take-home experience with each rate condition and that allowed statistical analysis of learning effects and within-subject data.

PROCEDURE

Subjects

Six subjects implanted with the Nucleus 24 (CI24M) receiver/stimulator (see Appendix A) participated in this study. However, subject 6 was unable to complete the study due to increased levels of tinnitus associated with the use of high stimulation rates. Subject details are summarized in Table 1. At the commencement of the study, subjects had been implanted with the CI24M for between 4 and

TABLE 2. Rate conditions evaluated: (A) 250, (B) 807, and (C) 1615 average pulses per second per channel (pps/ch), frequency at which filter-bank channels were analyzed/updated in Hz, duration of biphasic pulses in μsec, duration between biphasic pulses/phases in μsec, and average duration between stimuli (shorting time) in μsec.

Rate Condition	Average Channel Rate (pps/ch)	Analysis Frequency (Hz)	Pulse Width (μsec)	Inter-Phase Gap (μsec)	Average Inter-Stimulus Gap (μsec)
A	250	250	25	25	380
B	807	760	25	25	65
C	1615	760	25	8	12

18 mo and had used the SPEAK speech processing strategy implemented in the SPrint speech processor since implantation. The duration of severe-to-profound deafness for subjects varied between 0.5 and 10 yr. The subjects speech recognition scores overlapped those of the larger population of Australian adult implant users and were representative of average to good performers. Three-month postoperative CNC word scores using the SPEAK strategy have been included in Table 1 for each subject.

Speech Processing Strategy and Stimulus Conditions

The speech processing strategy evaluated in this study employed an electrode selection technique similar to that of the SMSP or n-of-m strategies (which is similar to the technique employed by the SPEAK strategy) and it was implemented in the SPrint speech processor (see Appendix B). Three stimulation rate conditions were explored. The average rates tested were 250 pps/ch (condition A), 807 pps/ch (condition B) and 1615 pps/ch (condition C), as summarized in Table 2. Jitter in the absolute timing of each stimulus was employed so as to minimize possible fixed rate-pitch percepts, which may interact with place-pitch and F0 perception, especially for rates below the "pitch saturation limit" (Wilson et al., Reference Note 6). The stimulation rates quoted are thus average rates with $\pm10\%$ time jitter. For all three rate conditions the temporal information from the analysis filter-bank was limited to less than approximately 250 Hz (-3.3 dB@190 Hz) by a low-pass envelope smoothing filter.* For the 250 pps/ch rate condition, the frequency at which the analysis filter-bank channels were sampled (i.e., analysis frequency) was 250 Hz. Thus for a speaker F0 of greater than 125 Hz, some aliasing will have resulted. For both the 807 and 1615 pps/ch rate conditions, the analysis frequency was 760 Hz, which will have adequately encoded F0 up to 190 Hz (recall that an analysis/stimulation rate of four times the modulation frequency is required as per the findings of Busby et al. 1993; McKay et al., 1994). The temporal envelopes in the stimulus signals provided by the 807

and 1615 pps/ch rate conditions were thus expected to be very similar. This was confirmed by observing the output of the speech processor for typical input speech signals using "electrodograms" (see Appendix C). These are somewhat similar to spectrograms for acoustic signals and show how the excitation on each of the electrodes varies with time. Note that for the 807 pps/ch condition the stimulation rate slightly exceeds the analysis rate. To make up the required stimulation rate, after every 17 or 18 analysis intervals the last stimulus group was repeated once. Similarly, to make up the 1615 pps/ch rate each stimulus group was repeated once per analysis interval.

The number of maxima selected was fixed at eight for all conditions and was chosen to give a spectral representation similar to the SPEAK strategy in which the number of maxima selected ranges from 6 to 10. Monopolar stimulation was used so as to minimize the pulse-width required to elicit auditory perception, hence allowing for higher rates of stimulation. For all rate conditions, biphasic current pulses of 25 μsec per phase were used. For the 250 and 807 pps/ch rate conditions, the inter-phase gap (the interval between the first and second phases of the stimulus) was 25 μsec, whereas for the 1615 pps/ch rate condition it was reduced from 25 to 8 μsec to accommodate the desired stimulation rate. For each of the three rates, the inter-stimulus gap was adjusted so as to evenly spread eight stimulus pulses across the required rate interval. The number of active electrodes assigned to each of the subjects varied from 15 to 20. The frequency-to-electrode mapping was based on default settings available for the SPEAK strategy given the number of active electrodes assigned (see Table 3). For each subject and rate condition, the stimulation levels for threshold and maximum comfortable loudness on each active electrode were measured and mapped into their processors.

Study Design

The study design involved a repeated ABC protocol carried out over a 21-wk period with the evaluation order balanced across subjects. It should be noted, however, that due to the withdrawal of sub-

---

*This is also typical of clinical applications of SPEAK, ACE, and CIS.

TABLE 3. MAP and evaluation details: number of active electrodes, frequency range, order of evaluation of the three rate conditions (A, B, C), and signal to noise ratios used in CUNY sentence tests for each subject.

Subject	Number of Electrodes in MAP	Frequency Range (Hz)	Rate Condition Evaluation Order	SNR used in CUNY Sentence Tests (dB)
1	16	160–5744	ABC–ABC	+15, +10, +5
2	20	116–7871	BCA–BCA	+10, +5, 0
3	15	244–4177	CAB–CAB	+20, +15, +10
4	20	116–7871	ACB–ACB	+15, +10, +5
5	18	142–7009	CBA–CBA	+20, +15, +10
6	20	116–7871	BAC–BAC	+15, +10, +5

ject 6, the evaluation order was in fact not balanced across subjects. A repeated design protocol was employed so that learning effects could be included in the analysis of individual results and to minimize order effects across subjects. The order in which the rate conditions were evaluated for each subject is shown in Table 3. At the commencement of the study, subjects were programmed with the first rate condition (e.g., condition A for subject 1) and used this stimulation rate exclusively for a period of 3 wk, during which time stimulation current threshold and maximum comfortable loudness levels were adjusted if necessary. Speech perception performance was evaluated at the end of week 3 and again at the end of week 4. This procedure was then repeated for the next two rate conditions (e.g., condition B and then C for subject 1) giving a total of 4 wk of experience with each rate condition. At the end of this first evaluation stage (ABC_1) the protocol was repeated (ABC_2). However, in this case only 2 wk of experience was provided before the two evaluation sessions. In total, for each rate condition and subject, four evaluation sessions were conducted and 7 wk of experience was provided.

Speech Perception Evaluation

In each evaluation session, speech perception was assessed using open-set monosyllabic CNC words (Peterson, & Lehiste, 1962) presented in quiet and open-set CUNY sentences (Boothroyd, Hanin, & Hnath, Reference Note 10) presented with multitalker noise (Auditec, St. Louis, catalogue No. C146-MT MULTITALKER) at three SNRs. The speech perception tests were conducted in a sound-attenuated booth using prerecorded material played back via a loudspeaker at an azimuth of zero degrees. The word and sentence material was presented at a level of 70 dB SPL, as measured at the approximate location of each subject's behind-the-ear microphone. The presentation level of the speech material (including the multi-talker noise) was measured using a Rion Co. Ltd. (NA-20) sound level meter with dBA weighting, using peak level read-

ings and fast response measurements. The test material was recorded using two female speakers. The fundamental frequency of the speaker used in the CNC word tests was approximately 180 Hz and in the CUNY sentence tests approximately 190 Hz. The fundamental frequencies were therefore low enough to be coded in the temporal envelope of the system, given an analysis rate of 760 Hz.

Two lists of CNC words containing 50 words per list were presented in each evaluation session (a total of eight lists for each rate condition) and these were scored as whole words correct and phonemically as percentage of vowels and consonants correct. Two lists of CUNY sentences containing 102 words in 12 sentences were presented at each SNR in each evaluation session (a total of eight lists per SNR for each rate condition) and were scored as percentage words correct. To reduce ceiling and floor effects, the SNRs administered to each subject were varied according to their ability (see Table 3). This was measured before the study using the SPEAK strategy. The order of the lists was randomized across subjects and conditions so as to minimize list effects. No feedback was given during or after testing.

Data Analysis

Multi-way analysis of variance was applied both to group and individual results to determine whether performance for any of the rate conditions was significantly different on each of the speech perception tests, and whether there was a significant effect of learning between results obtained in the first (ABC_1) and second (ABC_2) evaluation stage. Post hoc comparisons were conducted using the Scheffé test (Scheffé, 1959) when significant effects of dependent variables (rate condition, evaluation stage) were observed. Separate analyses were carried out for the vowel, consonant, and word scores within the CNC word tests as well as for each SNR used in the CUNY sentence tests.

Tinnitus Evaluation and Comparative Performance Questionnaire

At the commencement of the study, subjects were administered a tinnitus questionnaire to assess baseline tinnitus levels. The questionnaire was re-administered at the end of the first evaluation stage for each rate condition. Subjects were asked to report any increases in tinnitus or nonauditory percepts throughout the duration of the study.

On completion of the repeated ABC protocol, all three rate conditions were programmed into separate program locations in each subject's speech processor. Subjects were asked to use each program for equal lengths of time over a 3- to 4-wk period and to compare each program in a variety of listening situations in everyday life. At the end of this period, they each completed a comparative performance questionnaire to assess which rate condition, if any, subjectively provided the best hearing and understanding in a variety of listening situations. These included conversation in quiet, conversation in noise, conversation using the telephone, listening to the television, listening to the radio, listening to music and listening to environmental sounds. An overall preference judgment was also obtained. If subjects preferred more than one rate condition in a given situation, they were encouraged to indicate this and the preference score was equally divided between the preferred rate conditions.

RESULTS

Threshold and Maximum Comfortable Loudness Levels

The electrical stimulation levels measured at the three stimulation rates are shown in Figure 1. Threshold levels (T-Level) and maximum comfortable loudness levels (C-Level) are shown with dotted lines and electrical dynamic ranges (D-Range) are shown with solid lines for each rate condition and electrode. Current levels are shown in clinical units and range from 1 to 255 where a level of 1 is nominally 10 uA with logarithmic increments of 2% between steps. For all subjects very few changes to threshold and comfortable levels were made over the duration of the study. The mean threshold levels (averaged across all subjects and electrodes for the 250, 807, and 1615 pps/ch rates, respectively) were 162.9, 145.5, and 132.7 and comfortable levels were 190.7, 183.2, and 178.6. Most subjects showed a decrease in threshold levels for increasing rate of stimulation. Comfortable levels also decreased slightly with rate, but to a smaller extent than threshold levels, so that the electrical dynamic range increased with increasing rate of stimulation. The mean dynamic range (averaged across all sub-

jects and electrodes) for the three rate conditions were 27.7, 36.9, and 44.8, respectively. Two-way analysis of variance (using subject and rate as factors) on these data showed a significant effect of rate ($F(2, 309) = 183$; $p < 0.001$). For subject 5, however, comfortable levels decreased by a greater amount than threshold levels with increasing rate of stimulation, so that electrical dynamic range did not increase with increasing rate of stimulation. For this subject, no significant differences (using a post hoc Tukey test) in mean dynamic range between the three rates of stimulation were observed.

Speech Perception Results

Mean scores averaged across the five subjects and over the two evaluation stages (ABC_1 and ABC_2) are shown in Figure 2. The results are presented for each rate condition and for each of the test categories. For the CNC word test, the scores are presented as percentage of vowels, consonants and whole words correct. For the CUNY sentences in noise, the scores are presented as percentage words correct for each of the three SNRs examined. The SNRs have been labeled highest, middle, and lowest SNR because absolute SNR varied for each subject. The group mean results of the speech perception evaluations revealed little difference between the 250 and 807 pps/ch rate conditions for both the CNC word tests in quiet and the CUNY sentence tests in noise. However, for the 1615 pps/ch rate condition, decreased performance was observed for vowels in the CNC word tests and for CUNY sentences at the middle and lowest SNR. The individual mean scores for the five subjects evaluated are shown in Figure 3. Again, the scores have been averaged over the two evaluation stages and are presented for each rate condition and for each of the test categories. Significant differences observed among the three rate conditions are shown in the tables presented below each bar graph in Figures 2 and 3. A detailed analysis of the results follows.

CNC Word Test—Results

Analysis of the group results for the CNC word tests revealed no significant differences across the three rate conditions for the word and consonant scores. However, the group mean vowel scores were significantly different ($F(2, 90) = 6.9$; $p < 0.05$). The mean score at 250 pps/ch was 3.5 percentage points higher than the score at 1615 pps/ch ($p^s < 0.05$, note p^s denotes the p-level derived by the post hoc Scheffé test). Note that percentage points relate to the difference in mean percent correct scores, which in this case were 65.5% and 62.0% for the 250 and 1615 pps/ch rates, respectively. In addition, the mean
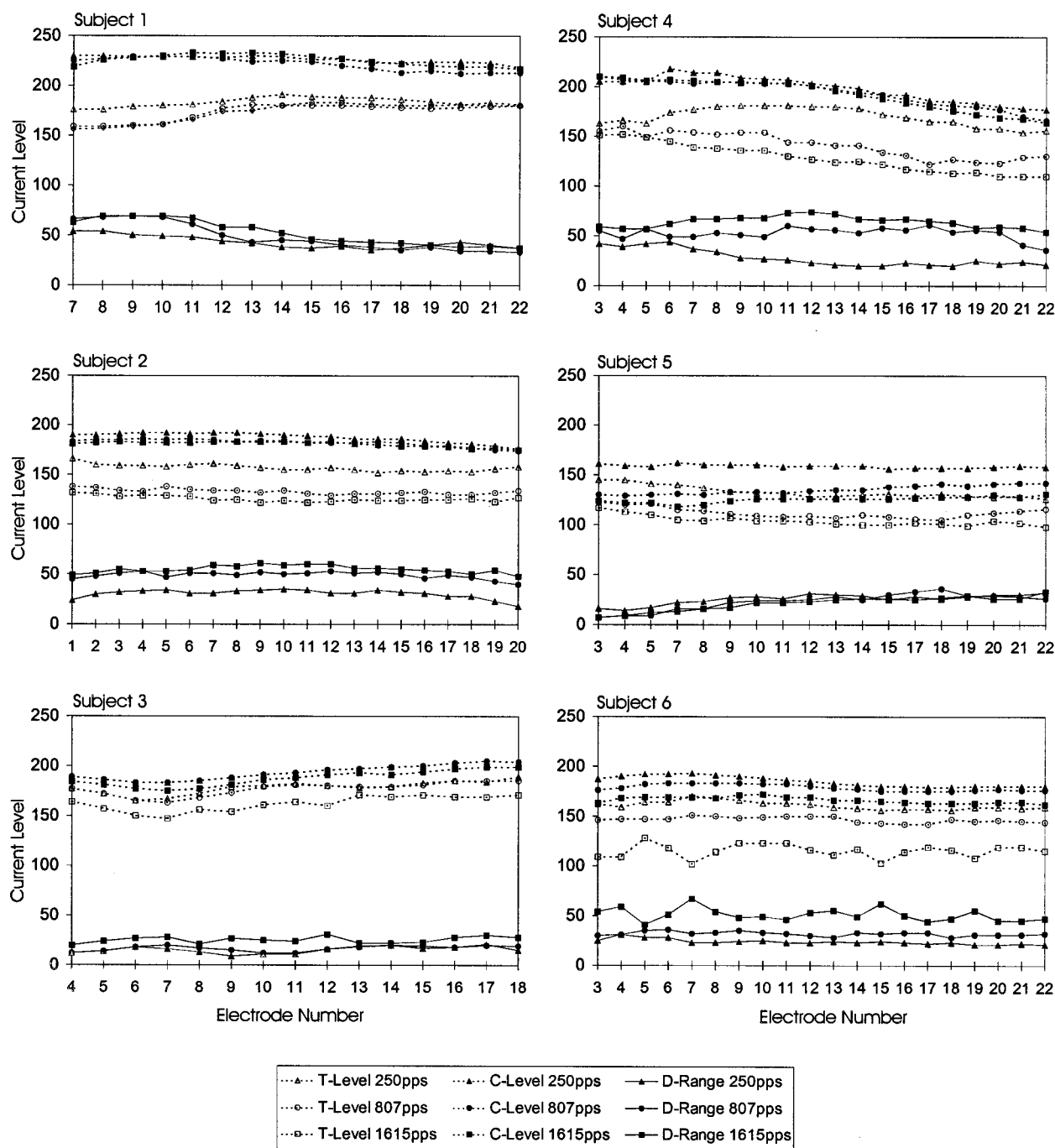
Figure 1. Stimulus levels (in clinical units) measured at each rate condition and mapped into each subject's speech processor are shown in panels 1 to 6. Stimulus current threshold levels (T-Level) and maximum comfortable loudness levels (C-Level) are shown with dotted lines, and electrical dynamic ranges (D-Range) are shown with solid lines for each rate condition and electrode. Triangle symbols are used for levels measured at 250 pps, circles for 807 pps, and squares for 1615 pps.

vowel score at 807 pps/ch was 4.2 percentage points higher than the score at 1615 pps/ch ($p^s < 0.01$). A significant interaction between subject and rate was also observed in the vowel results ($F(8, 90) = 5.1$; $p < 0.001$), indicating that the relationship between rate and vowel scores varied across subjects. Referring to Figure 3, for subject 1 the vowel score at 807 pps/ch was 11.7 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 8.0$; $p^s < 0.01$). In contrast for subject 2 the vowel score at 250 pps/ch was 7.5 percentage points higher than the scores at 807 pps/ch ($F(2,18) = 28.5$; $p^s < 0.001$), and also 10.2 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 28.5$; $p^s < 0.001$). Scores for subject
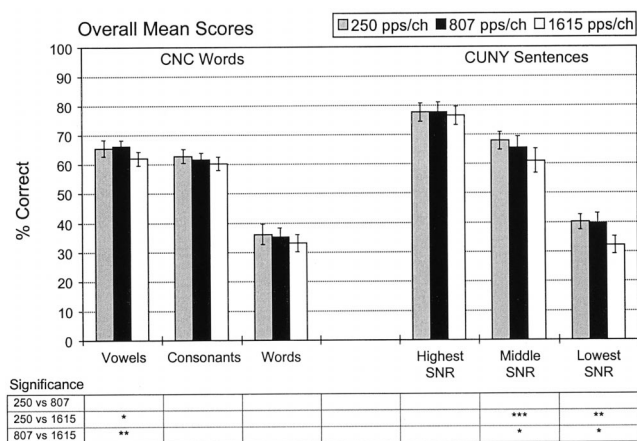
Figure 2. Overall group results averaged across five subjects for the open-set CNC word tests conducted in quiet and open-set CUNY sentence tests conducted in competing multi-talker noise. The word test results are scored as percentage of vowels, consonants, and whole words correct. The sentence test results are scored as percentage words correct for the three SNRs examined from highest to lowest signal to noise ratio (SNR) used for each subject. The overall mean scores are an average of 40 lists for each rate condition. Statistically significant differences (post hoc Scheffé test) are shown in the tables presented below each bar graph. Significance at the 5% level is indicated by *, at the 1% level by **, and at the 0.1% level by *. The error bars indicate the standard error (that is, standard deviation divided by the square root of sample size).**

3 differed again. The vowel score for this subject at 1615 pps/ch was 8.7 percentage points higher than the score at 250 pps/ch ($F_{(2,18)}$ = 5.4; p^s < 0.05). No significant differences were observed for subjects 4 and 5 in all categories of the CNC word evaluations. Although no significant interaction between word score and rate condition was observed in the group mean results, some significant differences in scores across rate were observed for two subjects. The word score for subject 2 at 250 pps/ch was 9.1 percentage points higher than the score at 1615 pps/ch ($F_{(2,18)}$ = 4.2; p^s < 0.05). In contrast for subject 3 the word score at 1615 pps/ch was 6.7 percentage points higher than the score at 250 pps/ch ($F_{(2,18)}$ = 6.8; p^s < 0.01).

CNC Word Test—Learning Effects • Small but statistically significant learning effects in the group mean scores were observed in all categories of the CNC word tests. For the word results, the mean score for all three rate conditions increased by 2.8 percentage points from the first evaluation stage (ABC$_1$) to the second (ABC$_2$) ($F_{(1,90)}$ = 5.9; p < 0.05). For the consonant results, means scores increased by 3.0 percentage points over the two evaluation stages ($F_{(1,90)}$ = 12.3; p < 0.001) and a significant interaction between rate and evaluation

stage was also observed ($F_{(2, 90)}$ = 5.9; p < 0.01). For the vowel results, mean scores increased by 3.3 percentage points over the two evaluation stages ($F_{(1,90)}$ = 11.5; p < 0.01). Post hoc analysis of the effect of evaluation stage for each rate condition alone revealed that learning effects in the CNC word tests were most significant for the 1615 pps/ch rate condition. Mean scores for the 1615 pps/ch rate condition increased over the two evaluation stages for the word, vowel and consonant results by 5.7 (p^s < 0.01), 4.5 (p^s < 0.05) and 5.5 (p^s < 0.01) percentage points, respectively. For the 807 pps/ch rate condition, only consonant scores showed significant learning effects of 4.5 percentage points (p^s < 0.01).

Some effects of learning were also observed for individual subjects. For subject 1, a significant increase of 7.8 percentage points over the two evaluation stages in the mean vowel scores (averaged across all rate conditions) was observed ($F_{(1,18)}$ = 10.5; p^s < 0.01). For subject 3, a significant increase of 3.6 percentage points in the mean consonant scores was observed ($F_{(1,18)}$ = 4.6; p^s < 0.05). For subject 5 an increase of 5.4 percentage points in the mean consonant scores was observed ($F_{(1,18)}$ = 5.8; p^s < 0.05). For each subject and speech perception test condition, mean score differences over the two evaluation stages suggested that the majority of group learning effects occurred with the 1615 pps/ch rate and were mainly evident in the results of subjects 5 and 1.

CUNY Sentence Test—Results

Analysis of the group results for the CUNY sentence tests in noise revealed no significant differences across the three rate conditions for tests conducted at the highest SNR. However, for tests conducted at the middle and lowest SNRs, the mean scores for the 1615 pps/ch rate were significantly lower than those at the other rates. For the middle SNR the mean score at 250 pps/ch was 6.8 percentage points higher than the score at 1615 pps/ch ($F_{(2, 90)}$ = 7.8; p^s < 0.001). In addition, the mean score at 807 pps/ch was 4.4 percentage points higher than the score at 1615 pps/ch ($F_{(2, 90)}$ = 7.8; p^s < 0.05). Similarly, for the lowest SNR the mean score at 250 pps/ch was 7.9 percentage points higher than the score at 1615 pps/ch ($F_{(2, 90)}$ = 6.4; p^s < 0.01), and the score at 807 pps/ch was 7.4 percentage points higher than the score at 1615 pps/ch ($F_{(2, 90)}$ = 6.4; p^s < 0.05).

Significant interactions between subject and rate condition were observed for tests at both the middle SNR ($F_{(8, 90)}$ = 11.4, p < 0.001) and at the lowest SNR ($F_{(8, 90)}$ = 6.6; p < 0.001) indicating that the scores varied from subject to subject with rate of
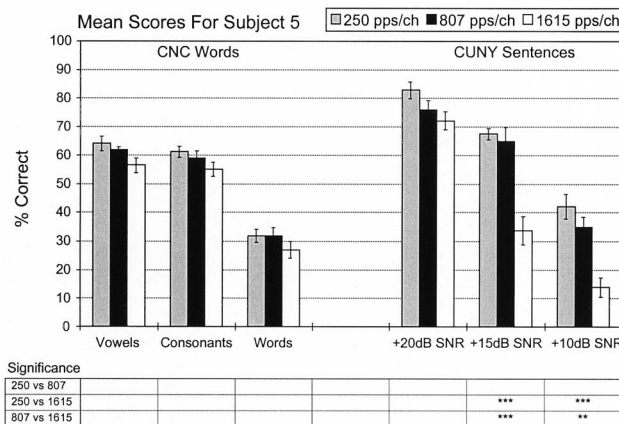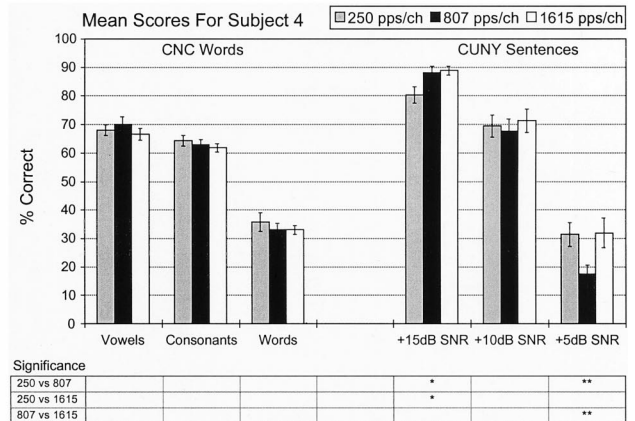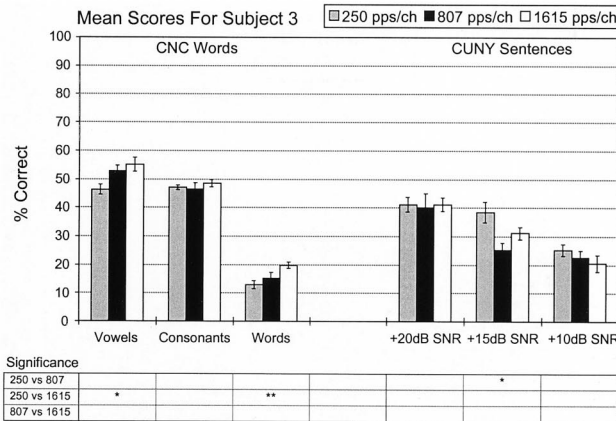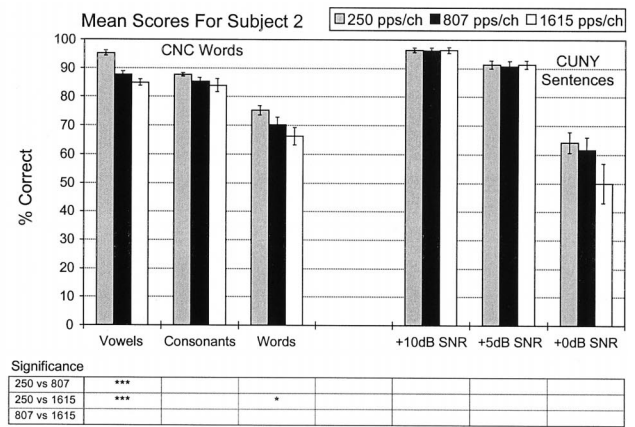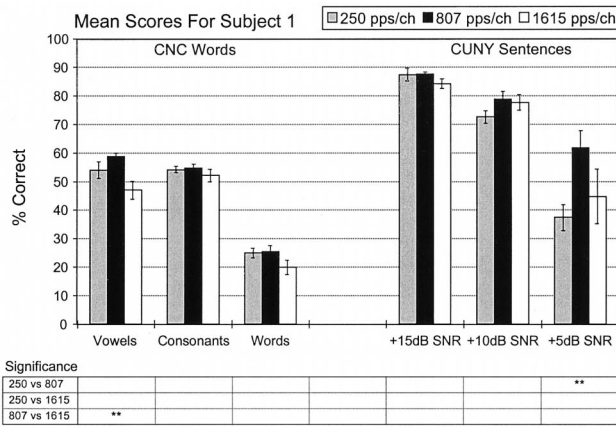
Figure 3. Results for individual subjects of open-set CNC word tests conducted in quiet and open-set CUNY sentence tests conducted in competing multi-talker noise. The word test results are scored as percentage of vowels, consonants, and whole words correct. The sentence test results are scored as percentage words correct for the three signal to noise ratios (SNRs) examined. The individual mean scores are an average of eight lists for each rate condition. Statistically significant differences (post hoc Scheffé test) are shown in the tables presented below each bar graph. Significance at the 5% level is indicated by *, at the 1% level by **, and at the 0.1% level by ***. The error bars indicate the standard error.

stimulation. For subject 1, at the lowest SNR the score at 807 pps/ch was 24.4 percentage points higher than the score at 250 pps/ch ($F(2,18) = 6.5; p^s < 0.01$). For subject 2 no significant differences in results were observed for the tests at all three SNRs. For subject 3, at the middle SNR the score at 250 pps/ch was 13.2 percentage points higher than the score at 807 pps/ch ($F(2,18) = 5.7; p^s < 0.05$). For

subject 4, at the highest SNR the score at 807 pps/ch was 7.9 percentage points higher than the score at 250 pps/ch ($F(2,18) = 5.3$; $p^s < 0.05$), and the score at 1615 pps/ch was 8.6 percentage points higher than the score at 250 pps/ch ($F(2,18) = 5.3$; $p^s < 0.05$). In addition, at the lowest SNR for this subject the score at 250 pps/ch was 14.0 percentage points higher than the score at 807 pps/ch ($F(2,18) = 11.0$; $p^s < 0.01$), and the score at 1615 pps/ch was 14.5 percentage points higher than the score at 807 pps/ch ($F(2,18) = 11.0$; $p^s < 0.01$). Results for subject 5 were different again with poor performance being obtained with the 1615 pps/ch rate condition. For this subject, at the middle SNR the score at 250 pps/ch was 33.7 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 57.9$; $p^s < 0.001$), and the score at 807 pps/ch was 31.1 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 57.9$; $p^s < 0.001$). Similarly for the lowest SNR the score at 250 pps/ch was 28.2 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 14.7$; $p^s < 0.001$), and the score at 807 pps/ch was 21.0 percentage points higher than the score at 1615 pps/ch ($F(2,18) = 14.7$; $p^s < 0.01$). In fact the significantly lower group mean scores obtained with the 1615 pps/ch rate compared with those of the two lower rates can be accounted for predominantly by the results of this subject.

CUNY Sentence Test—Learning Effects • No significant learning effects in the group results for the CUNY sentence tests conducted at all SNRs were observed. However, significant interactions were observed between subject and evaluation stage ($F(4, 90) = 4.6$; $p < 0.01$) for tests conducted at the lowest SNR. In addition, a significant group interaction between subject, rate condition and evaluation stage ($F(8, 90) = 6.3$; $p < 0.001$) was observed for tests conducted at the middle SNR. This indicates that learning effects varied between subjects and across the three rate conditions examined. However, the analysis of individual data revealed significant learning effects and/or interactions between evaluation stage and rate in only two subjects. For subject 4, this effect was observed at the lowest SNR ($F(2, 18) = 8.6$; $p < 0.01$). For this subject a significant decrease in scores of 14.7 percentage points from the first evaluation stage to the second was observed ($p^s < 0.001$), which was predominately due to a decrease in scores for all three rate conditions. For subject 5, a significant interaction between rate and evaluation stage ($F(2, 18) = 21.7$; $p < 0.001$) was observed for the middle SNR. For this subject scores with the 1615 pps/ch rate increased from the first evaluation stage to the second, by approximately 10 and 20 percentage points for tests at the lowest and middle SNRs, respectively.
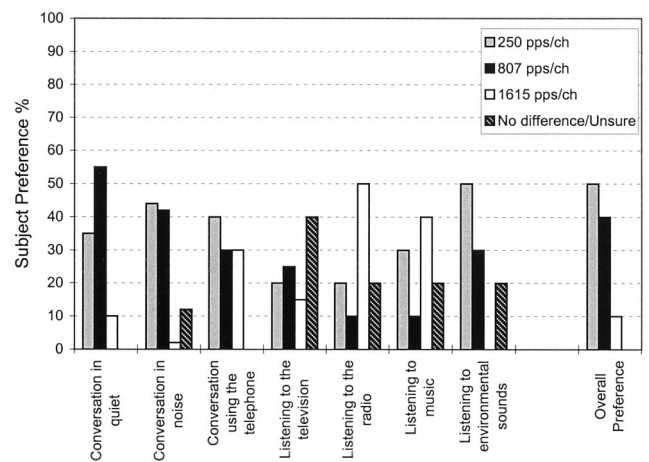

Figure 4. Summary of responses from the comparative performance questionnaire. Percentage of subjects who reported that in the following listening situations, the 250, 807, or 1615 pulses per second per channel (pps/ch) rate condition provided the best hearing and understanding, or there was no difference/they were unsure.

Tinnitus Evaluation

Subjects 1 to 5 reported no changes in tinnitus levels over the duration of this study. However, subject 6 withdrew from the study due to increased levels of tinnitus, which developed as a result of high-rate stimulation. For this subject, increased tinnitus was initially not reported during use of the first rate condition evaluated (807 pps/ch). However, on changing to the 250 pps/ch rate condition, a decrease in tinnitus levels was reported after approximately 1 wk of use. Subsequently, within a week of changing to the 1615 pps/ch rate condition, high levels of tinnitus were reported with the subject commenting that "head noises were very noticeable" and that "she felt very confused" and at this point the subject was withdrawn from the study.

Comparative Performance Questionnaire

A summary of the responses to the comparative performance questionnaire administered to the five subjects who completed this study are shown in Figure 4. For the overall preference category, it appears that most subjects preferred the 250 and 807 pps/ch rate conditions. For conversation in quiet and in noise the 250 and 807 pps/ch rate conditions were preferred. Subjects were divided in their preference for conversation using the telephone with all three rate conditions polling fairly equally. When listening to music or the radio it appears that subjects preferred the 1615 pps/ch rate condition. Most subjects reported little difference between the three rate conditions or were unsure of their preference when listening to the television. For environ-

mental sounds most subjects preferred the 250 pps/ch rate condition.

Individually, subject 1 indicated an overall preference for the 807 pps/ch rate condition, subjects 2 and 3 preferred the 250 pps/ch rate condition, subject 4 preferred both the 807 and 1615 pps/ch rate conditions, and subject 5 preferred both the 250 and 807 pps/ch rate conditions. At the completion of the study (i.e., after the comparative performance questionnaire was administered), subject 1 continued to use both the 250 and 807 pps/ch stimulation rates in everyday life and subjects 2, 3, and 5 elected to use the 250 pps/ch rate exclusively. Subject 4 chose to use the 807 and 1615 pps/ch stimulation rate. However, his decision may have been influenced by the fact that during the comparative performance phase of the investigation, he developed voltage compliance problems with the CI24M when using the 250 pps/ch stimulation rate. This subject's CI24M was subsequently explanted due to medical complications unrelated to implant function and therefore it has not been possible to further investigate this issue.

DISCUSSION

The group results showed that for the speech processing strategy evaluated, the use of stimulation rates higher than 250 pps/ch (up to 1615 pps/ch) generally did not improve speech perception scores in quiet or in noise. However, individual data showed that in some speech tests, some subjects obtained significantly better results with one (or more) of the three rate conditions. This suggests that stimulation rate should be optimally selected to suit each subject.

Based on previous speech perception studies and psychophysical experiments discussed in the introduction, an improvement in mean performance when changing from the 250 pps/ch rate condition to the 807 and 1615 pps/ch rate conditions was anticipated. However, limitations of current cochlear implant technology, which may restrict the amount of acoustic information delivered to implantees compared with that of normally hearing listeners, could account for our findings. For example, in present cochlear implant systems there is less redundancy as well as considerable quantization in amplitude, time and frequency of the signal when compared with normal hearing. Consequently, corruption of the electrical stimulus signal by the addition of noise (as used in the sentence evaluations) may have a much larger effect, and thus reduce listeners' ability to make use of any additional fine structure and F0 cues encoded by the high rate conditions. Furthermore, electrical stimulus interactions such as for-

ward, backward and simultaneous masking, temporal interactions between successive stimuli and charge summation across adjacent channels may have limited perception of any additional fine temporal and spectral detail provided by the high rate conditions, especially when many closely spaced sites of stimulation are provided. Differences between neural firing responses to acoustic and electrical stimulation, and the particular electrical coding techniques currently available, may also in part account for the lack of improvement in overall mean results. It may be the case that for higher electrical stimulation rates than those used in this study (e.g., 5000 pps/ch), neural responses can be made to more closely approximate the acoustic case (Matsuoka et al., Reference Note 2; Rubinstein et al., 1999; Wilson et al., 1997). However, perceptual benefits of these very high rates have not been reported to date. Similarly, stimulation techniques such as those proposed by Morse and Evans (1996), Moss et al. (1996), and Parkins (1989), which also aim to provide electrical neural responses more like the acoustic case, may be of some benefit.

If coding of F0 does indeed improve separation of speakers with differing F0s (Assmann & Summerfield, 1990; Brokx & Nooteboom, 1982; Scheffers, Reference Note 5), it is possible that the speech material used in this study were not effective in revealing any advantages. For our tests in noise, perhaps the F0 frequencies present in the multi-talker noise were too similar to the F0 of the speaker. Thus, any advantages of using high rates to separate the signal from the background noise may have been compromised. Alternatively, it may be that even if fundamental frequencies are aliased (as was the case for the 250 pps/ch rate), the subject can make use of the information, as long as the coded F0 differs from that of the noise. This is supported by results of McKay and McDermott (1993) and Tong et al. (Reference Note 4) who showed that speaker and intonation pattern identification were possible with a strategy coding F0 modulation using a 250 pps/ch stimulation rate (i.e., where F0 was aliased).

On the other hand, it may in fact be the case that coding of fine temporal structure offers no advantage to speech intelligibility for implant users and that the temporal information in the stimulus (together with spectral information) for the 250 pps/ch rate condition already provided sufficient information despite the psychophysical findings (as discussed in the introduction) suggesting otherwise. Given that a positive SNR was used with the speech-in-noise tests, temporal envelope and spectral amplitude cues present in each channel of stimulation, which were similar for all three rate conditions, may therefore have been the more dominant cue (in

preference to F0 cues) for signal detection in the presence of noise. This is supported for speech in quiet and in the presence of speech-weighted noise by the results of Drullman (1995), Shannon et al. (1995), and Van Tasell et al. (1987, 1992). Each of these studies demonstrated that for normal-hearing listeners, fine temporal structure in the F0 region provides little advantage to speech comprehension.

The result showing that speech perception for sentences in noise decreased for the 1615 pps/ch rate condition was somewhat unexpected. This result, however, can be attributed mainly to the performance of subject 5. If this subject's data are omitted from the analysis, no significant difference in group results between the three rate conditions is observed for the CUNY sentence tests in noise or the word and consonant scores in the CNC word tests. The vowel scores in the CNC word tests, however, are still significantly different ($F(2, 72) = 4.4$; $p < 0.05$), but only show a small decrease in mean scores of 2.5 and 3.9 percentage points for the 1615 pps/ch rate condition compared with the 250 and 807 pps/ch rate conditions, respectively. During the course of the study, subject 5 commented on the harshness of high-pitched sounds produced by high rates of stimulation and stimulation at basal sites. She found it difficult to use the 1615 pps/ch rate condition especially when confronted with high levels of background noise. She was therefore provided with an option to revert to the 250 pps/ch rate condition in difficult circumstances. This option was estimated to have been used approximately 1 to 2 hr every day while working in a noisy environment. Although this decreased her experience with the 1615 pps/ch rate condition slightly, it seems more likely that her dislike of the harsher sounds produced by the higher rate conditions may have contributed to the decreased speech perception performance observed. This is also reflected by the subject's shifts in threshold and maximum comfortable levels with increased rate of stimulation. Referring to Figure 1, it was observed that for most subjects, the electrical dynamic range increased with increasing rate of stimulation. For subject 5, however, the electrical dynamic range decreased with increasing rate of stimulation, which may have both reduced intelligibility and increased the harshness of the signal.

In addition to subject 5's problems with the high rate conditions, subject 6 withdrew from the study as a result of elevated tinnitus associated with high-rate stimulation. Even if this is a rare occurrence, it would seem that some subjects can suffer adverse effects with high rates of electrical stimulation. The underlying mechanisms responsible for this are beyond the scope of this paper, but may be related to issues such as over-stimulation of audi-

tory neurons, undesirable interactions between closely spaced stimuli (in time and place), and revised loudness growth function/mapping required for high electrical stimulation rates. Some consideration should also be given to long-term safety issues associated with the use of high-rate stimulation. Research using the Nucleus 24 (CI24M) stimulator (Huang, Shepherd, Seligman, & Clark, 1998; Huang, Carter, Shepherd, Seligman, Tabor, & Clark, Reference Note 12; Tykocinski, Shepherd, & Clark, 1997; Xu, Shepherd, Millard, & Clark, 1997) has indicated that for high-rate monopolar stimulation (up to 3625 pps/ch on four channels, total 14500 pps), excitability and survival of auditory neurons are not adversely affected, and DC levels remain well within known clinical safety levels. Nevertheless it may be appropriate to proceed with some caution in this area (Tykocinski, Shepherd, & Clark, 1995).

Learning effects in the data may in part account for the poorer performance obtained with the highest rate conditions. Before commencement of the study, all subjects had gained at least 4 mo of experience with the SPEAK strategy, which was most similar to the 250 pps/ch rate condition tested. In the CNC word test, a statistically significant learning effect in group results across the two evaluation stages was observed for the vowel, consonant and word scores. Post hoc analysis showed that learning effects were greatest for the 1615 pps/ch rate condition, where word, consonant and vowel scores all increased by approximately 5 percentage points. Given that the mean vowel score at 1615 pps/ch, averaged across the two evaluation stages, was approximately 4 percentage points lower than scores with the lower rates, it is possible that with further experience results with the highest rate may become comparable with those of the other two rates.

For the CUNY sentence tests in noise no significant group learning effects were observed. However, some interactions in the data between evaluation stage, subject and rate condition were observed. The individual data indicated the greatest degree of learning, which occurred with the 1615 pps/ch rate and was mainly evident in the results for subject 5. However, in this case the size of the learning effect was substantially smaller than the difference in mean scores between the 1615 pps/ch rate and those at the two lower rates. Given the other issues discussed earlier for this subject, it seems very likely that other factors in addition to learning may have contributed to the poorer results with the 1615 pps/ch rate.

It is interesting to compare these results with those of Lawson et al. (Reference Note 9), which

generally showed improved performance for stimulation rates of 833 pps/ch as compared with 250 pps/ch using both the CIS strategy and an "n-of-m" strategy. However, as was the case in our study, some variation among subject performance for the parameter set examined was observed. More recently a within-subject comparison of three coding strategies (SPEAK, CIS, and Advanced Combination Encoder [ACE]) using the Nucleus 24 cochlear implant system (Arndt, Staller, Arcaroli, Hines, & Ebinger, Reference Note 11) was conducted. The group mean results for 62 subjects showed that best performance was obtained using the ACE strategy (which used a higher stimulation rate than SPEAK but not CIS). However, individual results indicated that the optimal coding strategy and optimal parameter set (e.g., number of channels, rate of stimulation, etc) varied widely from subject to subject. Aspects of their results do concur with those of our study in that individual subject variability was evident that suggested that an optimally selected stimulation rate may benefit some subjects.

The overall results of the comparative performance questionnaire were consistent with the speech perception evaluations. Most subjects indicated a preference for the 250 and 807 pps/ch rate conditions for everyday conversation and listening in a variety of situations. Interestingly, the highest rate condition (1615 pps/ch) was preferred by the majority of the subjects when listening to music or the radio. It may be that the higher rate condition provided a quality improvement to the signal that was not directly related to speech intelligibility. This suggests that alternative means of measuring sound quality for cochlear implant subjects are required.

CONCLUSION

For the speech processing strategy employed (prototype ACE strategy) in this study, group results for the five subjects who completed the study indicated that stimulation rates higher than 250 pps/ch (up to 1615 pps/ch) provided no significant improvement to speech perception in quiet or in noise. However, individual subject variability was evident, and some subjects showed increased scores with particular rates of stimulation on some speech perception tasks, suggesting that some subjects may benefit from an optimally selected stimulation rate. In addition to these findings, results for the high stimulation rate condition (1615 pps/ch) for one subject, together with tinnitus problems arising from high-rate stimulation for another subject, indicated that high rates of stimulation may be undesirable for some subjects. The results of the comparative performance questionnaire indicated that most subjects

preferred the 250 and 807 pps/ch stimulation rates for conversation and listening in a variety of situations, which is in agreement with the group results of the speech perception tests. However, the 1615 pps/ch rate condition was generally preferred for music perception.

APPENDIX

A: The Nucleus 24 Cochlear Implant System

The Nucleus 24 cochlear implant system consists of an implantable receiver/stimulator (CI24M), a wearable speech processor (SPrint) and a Windows programming and diagnostic system (WinDPS). The CI24M implant is capable of delivering up to 14400 pps. The receiver/stimulator is connected to an array of 22 banded intra-cochlear electrodes and two extra-cochlear electrodes, one being a plate electrode located on the implant package and the other a ball electrode located on a lead, which is positioned under the temporalis muscle during surgery. The electrodes can be stimulated in a bipolar, monopolar or common ground configuration. The extra-cochlear electrodes are activated during monopolar stimulation and can be used individually or connected together. Alternatively, for bipolar and common ground stimulation modes they are isolated from the rest of the circuit. Biphasic stimulus pulses are generated with electrode shorting during the inter-stimulus gap to remove any residual charge. The extra-cochlear electrodes also incorporate series capacitance to block any DC current during monopolar stimulation.

B: The Advanced Combination Encoder (ACE) Speech Processing Strategy

The ACE strategy is a flexible speech processing strategy capable of providing high stimulation rates and up to 22 channels of stimulation. It can implement the electrode selection technique employed in the SPEAK strategy or alternatively be pro-

grammed to stimulate at fixed electrode sites (i.e., CIS). The strategy estimates the incoming signal spectrum using a Fast Fourier Transform (FFT) filter-bank, which simulates a bank of band-pass filters followed by envelope smoothing filters. The FFT bins, which are linearly spaced in frequency, are used to produce N filter bands, which are typically linearly spaced from 188 to 1312 Hz and then logarithmically spaced up to 7938 Hz. Summation of power is used to combine adjacent FFT bins to form larger band-pass filter bands. A subset of these N bands are then selected in each stimulation time frame on the basis of highest signal level (spectral maxima). Electrodes corresponding to the selected bands are then stimulated in a tonotopic basal to apical order. The stimulus intensities of the activated electrodes are derived from a mapping between the users measured threshold and comfortable levels and the signal intensities within the selected bands. The overall loudness of the stimulation can also be controlled by the user via the volume control.

The filter-bank can be chosen from a selection of specifications that provide from 1 to 22 filter bands covering a maximum frequency range of 188 to 7938 Hz. The number of selected maxima can range from 1 to 20 (or from 1 to N, if N < 20) and the stimulation rate can be chosen from a range of 250 to 2400 pps/ch (limited by a maximum rate of 14400 pps across all channels). The stimulation rate can either be constant or jittered in time by a percentage of the average rate. In the case of constant rate, the inter-stimulus gap (the interval between consecutive stimuli in which the electrodes are not activated) is set to evenly space the selected stimuli (i.e., those with the highest amplitude) within one stimulation interval (the stimulation interval is the inverse of the stimulation rate). In the case of jittered rate, the inter-stimulus gap (which is equal for all stimuli within one stimulation interval) is adjusted every stimulation interval by a random amount (i.e., ± percentage jitter as programmed by the clinician). The stimulation rate thus produced varies between consecutive stimulation intervals but has a fixed average rate. For stimulation rates of less than approximately 760 pps/ch, the filter-bank analysis rate is set to equal the stimulation rate. However, for higher stimulation rates, the analysis frequency is limited by the system to approximately 760 Hz and higher stimulation rates are obtained by repeating stimulus frames (stimuli in one stimulation interval) when necessary. For the 807 pps/ch rate, approximately one in every 17 or 18 stimulation frames is repeated. For the 1615 pps/ch rate, approximately every stimulus frame is repeated.

The filter-bank is implemented using a 128 point Hanning Window and FFT. A sampling rate of 16 kHz is used, thus providing an FFT channel spacing of 125 Hz and a low-pass filter cut-off frequency (-3 dB) of 180 Hz. However, the attenuation of the low-pass filter at twice the FFT channel spacing (i.e., 250 Hz) is only approximately 6 dB. The roll-off of the low-pass filter then becomes much steeper and by three times the FFT channel spacing (375 Hz) attenuation is greater than 30 dB. The temporal envelope derived from each FFT bin (or summed FFT bins) is thus expected to contain frequency components up to approximately 250 Hz and beyond this they will be greatly attenuated. The average F0 of the speakers used in this study ranged from approximately 180 to 190 Hz (at 190 Hz the smoothing filter attenuation is approximately 3.3 dB). Thus we would expect substantial amounts of F0 modulation in the temporal envelopes of the filter-bank channels. Furthermore, given an analysis rate and stimulation rate of at least four times this frequency (that is, $4 \times 190 = 760$ Hz), we would expect these frequency modulations to be adequately coded for implant users.

The strategy evaluated in this study was a prototype of the ACE strategy, which differed from the strategy commercialized by Cochlear Limited in two ways. Firstly, the maximum frequency range of the prototype strategy ranged from 80 to 7885 Hz and typically employed linear spaced filter bands in the region 80 to 1125 Hz followed by logarithmic spacing up to 7885 Hz. Secondly, the commercial version uses a default inter-phase gap of 8 μsec for all rates of stimulation, whereas in the prototype strategy a 25 μsec gap was used for the 250 and 807 pps/ch rates.

C: Stimulus Output Patterns (Electrodograms)

Stimulus output patterns (electrodograms) for the speech processing strategy evaluated in this study are shown in Figure C1. Time is shown along the abscissa and electrode number along the vertex. For each stimulus pulse recorded from the output of the processor, a vertical bar is shown in the electrodogram at the time and electrode position of the stimulus. The height of the vertical bar represents the stimulus level (log current in clinical units) where minimum amplitude corresponds to threshold, and maximum amplitude corresponds to comfortable level. The speech token presented in these recordings was taken from the sentence material used in this study. The token "boss" was presented in quiet via a loudspeaker to the ear level microphone of the speech processor. The output signal of the processor was recorded and plotted in the elec-
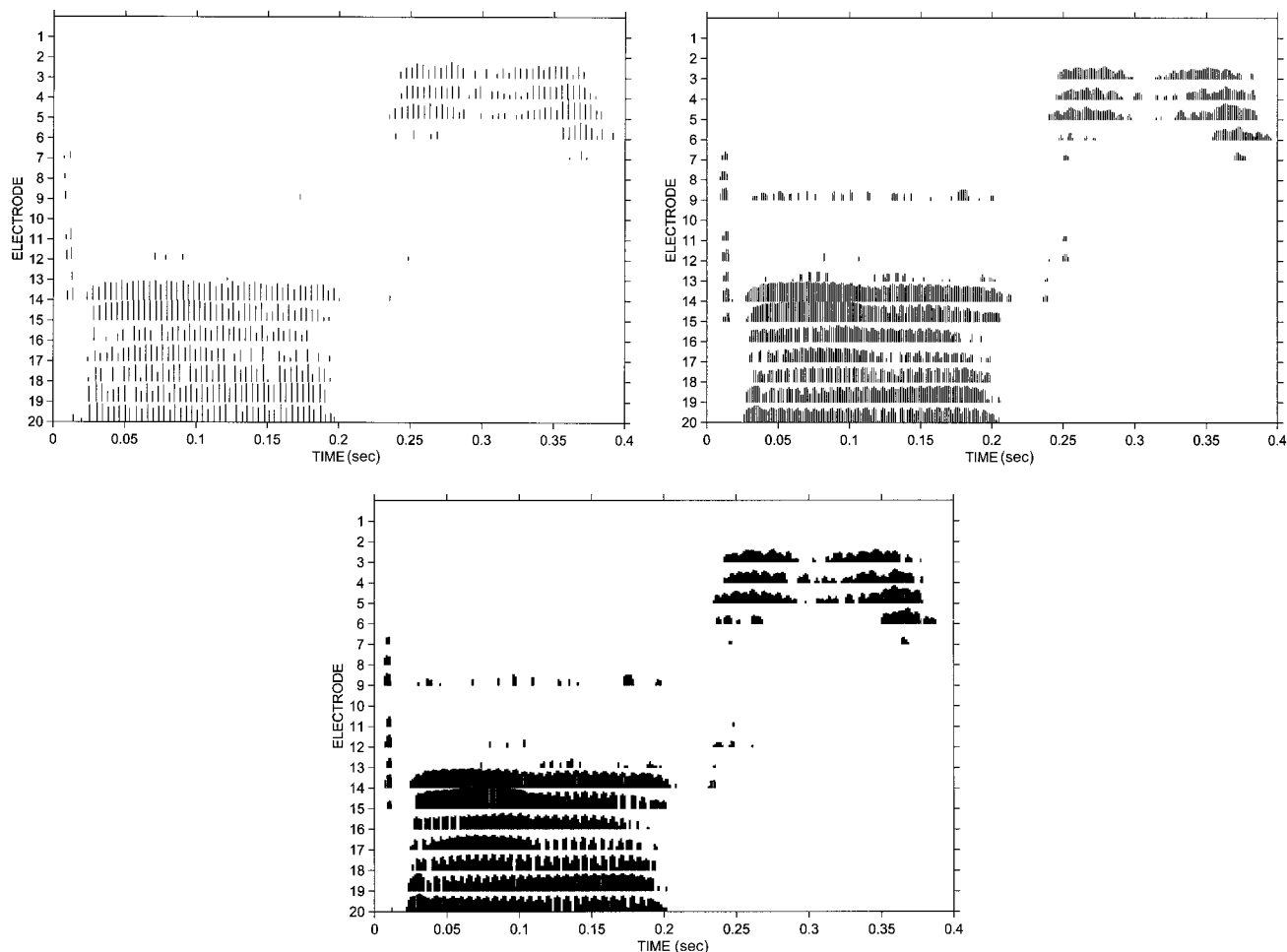
Figure C1. Electrodogram (stimulus output pattern) recordings from the speech processing strategy evaluated in this study. Time is shown along the abscissa and electrode number along the vertex. For each stimulus pulse, a vertical bar is shown at the time and electrode position of the stimulus. The amplitude of the bar represents the stimulus log current. Panels 1, 2, and 3 show the electrodograms for the token "boss" presented in quiet for the rate conditions 250, 807, and 1615 pps/ch, respectively.

trodograms shown in Figure C1 (panels 1, 2, and 3) for the 250, 807 and 1615 pps/ch rate conditions, respectively. The effect of stimulation rate can clearly be seen by the density of the stimulus pulses on the time axis. As expected, channel/electrode selection and the temporal envelope of the stimulus (excluding fine temporal structure, e.g., F0) for the three rate conditions are very similar.

Comparing the electrodograms, one major difference between the three rate conditions apart from the stimulation rate, is that of the fine temporal detail in the stimulus signal. Although all three rate conditions employ the same low-pass smoothing filter, only the 807 and 1615 pps/ch rate conditions provide a high enough stimulation rate to avoid aliasing. For these rate conditions, F0 modulation in the stimulus amplitude can clearly be seen during voiced portions in the speech signal (F0 is approximately 180 Hz in this token). Also, temporal details in other parts of the signal, such as unvoiced noise

bursts, are more accurately represented. On the other hand, less temporal detail for the 250 pps/ch rate condition compared with the higher rate conditions is provided. In this case, F0 modulation is aliased to a lower frequency due to the lower carrier/stimulation rate. Another more subtle difference visible in the electrodogram is that short duration transient features in the speech signal for the 250 pps/ch rate condition are often only represented by a small number of stimuli, which could make their perception by cochlear implantees difficult. The higher rate conditions provide better reconstruction of this fine temporal structure, although this does not necessarily imply that it will be better perceived.

REFERENCES

Assmann, P. F., & Summerfield, Q. (1990). Modeling the perception of concurrent vowels: vowels with different fundamental frequencies. *Journal of the Acoustical Society of America, 88,* 680–697.

Brokx, J. P. L., & Nooteboom, S. G. (1982). Intonation and the perceptual separation of simultaneous voices. *Journal of Phonetics*, *10*, 23–36.

Burns, E. M., & Viemeister, N. F. (1976). Nonspectral pitch. *Journal of the Acoustical Society of America*, *60*, 863–869.

Burns, E. M., & Viemeister, N. F. (1981). Played-again SAM: Further observations on the pitch of amplitude-modulated noise. *Journal of the Acoustical Society of America*, *70*, 1655–1660.

Busby, P. A., Tong, Y., C., & Clark, G. M. (1993). The perception of temporal modulations by cochlear implant patients. *Journal of the Acoustical Society of America*, *94*, 124–131.

Drullman, R. (1995). Temporal envelope and fine structure cues for speech intelligibility. *Journal of the Acoustical Society of America*, *97*, 585–592.

Dynes, S. B. C., & Delgutte, B. (1992). Phase-locking of auditory-nerve discharges to sinusoidal electric stimulation of the cochlea. *Hearing Research*, *58*, 79–90.

Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G., & Parkin, J. L. (1978). Auditory prosthesis research with multiple channel intracochlear stimulation in man. *Annals of Otology Rhinology & Laryngology*, *87*, (Suppl. 53), 5–39.

Hartmann, R., & Klinke, R. (1990). Response characteristics of nerve fibers to patterned electrical stimulation. In J. M. Miller & F. A. Spelman (Eds.), *Cochlear Implants: Models of the Electrically Stimulated Ear* (pp. 135–159). New York: Springer-Verlag.

Houtsma, A. J. M., Wicke, R. W., & Ordubadi, A. (1980). Pitch of amplitude-modulated low-pass noise and predictions by temporal and spectral theories. *Journal of the Acoustical Society of America*, *67*, 1312–1322.

Huang, C. Q., Shepherd, R. K., Seligman, P. M., & Clark, G. M. (1998). Reduction in excitability of the auditory nerve following acute electrical stimulation at high stimulus rates: III. Capacitive versus non-capacitive coupling of the stimulating electrodes. *Hearing Research*, *116*, 55–64.

Javel, E. Tong, Y. C., Shepherd, R. K., & Clark, G. M. (1987). Responses of cat auditory-nerve fibers to biphasic electrical current pulses. *Annals of Otology Rhinology & Laryngology*, *96*, (Suppl. 128), 26–30.

Kiang, N. Y. S., Eddington, D. K., & Delgutte, B. (1979). Fundamental considerations in designing auditory implants. *Acta Oto-laryngologica (Stockholm)*, *87*, 204–218.

Liberman, P. (1967). *Intonation, Perception and Language*. Cambridge, MA: Massachusetts Institute of Technology.

McDermott, H. J., McKay, C. M., & Vandali, A. E. (1992). A new portable sound processor for the University of Melbourne/Nucleus Limited multielectrode cochlear implant. *Journal of the Acoustical Society of America*, *91*, 3367–3371.

McKay, C. M., & McDermott, H. J. (1993). Perceptual performance of subjects with cochlear implants using the Spectral Maxima Sound Processor (SMSP) and the Mini Speech Processor (MSP). *Ear and Hearing*, *14*, 349–367.

McKay, C. M., McDermott, H. J., & Clark, G. M. (1994). Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implants. *Journal of the Acoustical Society of America*, *96*, 2664–2673.

McKay, C. M., McDermott, H. J., Vandali, A. E., & Clark, G. M. (1991). Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant. *Journal of the Oto-Laryngological Society of Australia*, *6*, 354–359.

Morse, R. P., & Evans, E. F. (1996). Enhancement of vowel coding for cochlear implants by addition of noise. *Nature Medicine*, *2*, 928–932.

Moss, F., Chioutan, F., & Klinke, R. (1996). Will there be noise in their ears? *Nature Medicine*, *2*, 860–862.

Parkins, C. W. (1989). Temporal response patterns of auditory-nerve fibers to electrical stimulation in deafened squirrel monkeys. *Hearing Research*, *41*, 137–168.

Peterson, G., & Lehiste, I. (1962). Revised CNC lists for auditory test. *Journal of Speech and Hearing Disorders*, *27*, 62–70.

Rose, J. E., Brugge, J. F., Anderson, D. J., & Hind, J. E. (1968). Patterns of activity in single auditory fibers of the squirrel monkey. In A. V. S. de Reuck & J. Knight (Eds.), *Hearing Mechanisms in Vertebrates* (pp. 185–194). London: Churchill.

Rubinstein, J. T., Wilson, B. S., Finley, C. C., & Abbas, P. J. (in press). Pseudospontaneous activity: Stochastic independence of auditory nerve fibers with electrical stimulation. *Hearing Research*, *127*, 108–118.

Scheffé, H. (1959). *The Analysis of Variance*. New York: Wiley.

Shannon, R. V. (1983). Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. *Hearing Research*, *11*, 157–189.

Shannon, R. V., Fan-Gang Zeng, Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science*, *270*, 303–304.

Simmons, F. B. (1966). Electrical stimulation of the auditory nerve in man. *Archives of Otolaryngology*, *84*, 2–54.

Simmons, F. B., Mathews, R. G., Walker, M. G., & White, R. L. (1979). A functional multichannel auditory nerve stimulator. *Acta Oto-laryngologica (Stockholm)*, *87*, 170–175.

Skinner, M. W., Clark, G. M., Whitford, L. A., Seligman, P. A., Staller, S. J., Shipp, D. B., Shallop, J. K., Everingham, C., Menapace, C. M., Arndt, P. L., Antogenelli, T., Brimacombe, J. A., & Beiter, A. L. (1994). Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system. *The American Journal of Otology*, *15*, (Suppl. 2), 15–27.

Skinner, M. W., Holden, L. K., Holden, T. A., Dowell, R. C., Seligman, P. M., Brimacombe, J. A., & Beiter, A. L. (1991). Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and the Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant. *Ear and Hearing*, *12*, 3–22.

Tong, Y. C., Blamey, P. J., Dowell, R. C., & Clark, G. M. (1983). Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant. *Journal of the Acoustical Society of America*, *74*, 73–80.

Townshend, B., Cotter, N., Van Compernolle, D., & White, R. L. (1987). Pitch perception by cochlear implant subjects. *Journal of the Acoustical Society of America*, *82*, 106–115.

Tykocinski, M., Shepherd, R. K., & Clark, G. M. (1995). Reduction in excitability of the auditory nerve following electrical stimulation at high stimulus rates. *Hearing Research*, *88*, 124–142.

Tykocinski, M., Shepherd, R. K., & Clark, G. M. (1997). Reduction in excitability of the auditory nerve following electrical stimulation at high stimulus rates. II. Comparisons of fixed amplitude with amplitude modulated stimuli. *Hearing Research*, *112*, 147–157.

Van den Honert, C., & Stypulkowski, P. H. (1987). Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. *Hearing Research*, *14*, 225–243.

Van Tasell, D. J., Greenfield, D. G., Logemann, J. J., & Nelson, D. A. (1992). Temporal cues for consonant recognition: training, talker generalization, and use in evaluation of cochlear implants. *Journal of the Acoustical Society of America*, *92*, 1247–1257.

Van Tasell, D. J., Soli, S. D., Kirby, V. M., & Widin, G. P. (1987). Speech waveform envelope cues for consonant recognition. *Journal of the Acoustical Society of America*, *82*, 1152–1161.

Whitford, L. A., Seligman, P. M., Everingham, C. E., Antognelli, T., Skok, M. C., Hollow, R. D., Plant, K. L., Gerin, E. S., Staller, S. J., McDermott, H. J., Gibson, W. R., & Clark, G. M. (1995).

Author/s:
Vandali, Andrew E.;Whitford, Lesley A.;Plant, Kerrie L.;Clark, Graeme M.

Title:
Speech perception as a function of electrical stimulation rate: using the nucleus 24 cochlear implant system

Date:
2000

Citation:
Vandali, A. E., Whitford, L. A., Plant, K. L., & Clark, G. M. (2000). Speech perception as a function of electrical stimulation rate: using the nucleus 24 cochlear implant system. Ear and Hearing, 21(6), 608-624.

Persistent Link:
http://hdl.handle.net/11343/27580