THE EFFECT OF RATE OF STIMULATION OF THE AUDITORY NERVE ON PHONEME RECOGNITION

David B. Grayden and Graeme M. Clark
The Bionic Ear Institute
384-388 Albert St
East Melbourne VIC 3002
Australia
d.grayden@medoto.unimelb.edu.au

ABSTRACT: Five patients implanted with the Nucleus CI-24M cochlear implant were tested on consonant and vowel perception with three different average rates of stimulation: 250 pulses/s per channel, 807 pps/ch and 1615 pps/ch. There were no significant differences in phoneme recognition scores when learning effects were taken into account. Information transmission analyses of consonant confusion matrices revealed that, with higher rates of stimulation, manner of articulation features were better perceived but place of articulation features were more poorly perceived. The results and analyses suggest that high rates of stimulation provide improved information about temporal information and frication in speech, but mask the spectral detail required for the perception of place of articulation.

INTRODUCTION

A key question for cochlear implants is the impact of rate of electrical stimulation on speech perception performance. This paper examines the effect of rate of stimulation on the perception of consonant and vowel phonemes.

Increasing the rate of stimulation should improve the perception of waveform envelope and fine temporal structure, which is important for detection of voicing and manner of articulation (Van Tassel et al., 1987). Voicing creates periodicity in the speech waveform, while perception of the amplitude envelope of speech and the locations of transient events is important for recognising manner of articulation. For example, abrupt changes in energy at the onsets of closure and release are important temporal cues for stops, affricates and nasals. Finally, place of articulation distinctions between phonemes are perceived using spectral cues such as formant locations.

Previous studies have shown varying results when comparing rates of stimulation. Fu and Shannon (1999) demonstrated in the Nucleus 22 device that vowel and consonant recognition scores improved with rate of stimulation up to 150 pulses/s per electrode, but there was no further improvement as rate of stimulation was increased to 500 pulses/s per electrode. Stimulation rates of approximately 250, 800 and 1600 pulses/s per electrode with the CI-24M implant showed decreasing performance for monosyllabic word tests and open-set sentence recognition in noise with increasing rate of stimulation (Plant et al., 1999).

Five subjects using the continuous interleaved sampling (CIS) strategy as well as n-of-m strategies showed significant performance improvement when the rate of stimulation was increased from 250 pulses/s per electrode to 833 pulses/s per electrode (Wilson, 1997). However, increasing the rate of stimulation above 833 pulses/s per electrode did not increase performance further. However, Brill et al. (1997) demonstrated that with the COMBI 40+ cochlear implant using CIS, higher rates of stimulation resulted in improved speech understanding for rates between 1,515 and 9,090 pulses/s per channel.

In this study, the CI-24M cochlear implant was used at stimulation rates of approximately 250 pulses/s per electrode, 807 pulses/s per electrode and 1615 pulses/s per electrode. Consonant and vowel confusion matrices were obtained for 5 patients in order to evaluate overall phoneme perception performance and the different rates of stimulation and to study the impact of stimulation rate on phoneme confusion. The perception of manner of articulation and voicing, in particular, were expected to increase with increasing rate of stimulation.

METHOD

A signal processing and stimulation strategy similar to the Spectral Maxima Sound Processor (SMSP) strategy (McDermott et al., 1992) was used. The strategy commenced a cycle of stimulation by applying the fast Fourier transform (FFT) and then collapsing the frequency bands into up to 20 bins. The bins were spaced linearly up to 1125 Hz and then logarithmically up to 7885 Hz. The bins were assigned to electrodes in the implant in tonotopic order. Eight bins with the highest amplitude were selected for each cycle and the corresponding electrodes stimulated. This completed a stimulation cycle. Monopolar biphasic stimulation was used.

Patients were assessed in three conditions: 250 cycles/s, 807 cycles/s, and 1615 cycles/s. 10% random timing jitter was introduced to reduce possible rate-pitch perception effects. The inter-phase gap used in the 250 cycles/s and 807 cycles/s conditions was 25 ms. In the 1615 cycles/s condition, it was reduced to 8 ms in order to accommodate the desired stimulation rate. The SPRINT^{TM} processor was used for signal processing and implant stimulation. This device could only work at a maximum rate of about 750 cycles/s. Therefore, occasionally a stimulus frame was repeated for the 807 cycles/s rate, and each frame was repeated at least once for the 1615 cycles/s rate of stimulation.

Five patients implanted with the CI-24M implant manufactured by Cochlear Ltd participated in the experiment. The patients were chosen based on availability and willingness to participate. All patients had been using a 250 cycles/s stimulation strategy before this study commenced. The number of electrodes used by each patient varied between 15 and 20. The frequency range to electrode mapping was selected to provide similar frequency resolution between all five patients. This is summarised in Table 1.

Patient	Number of electrodes	Frequency range (Hz)	Evaluation order
1	16	160–5744	ABC ABC
2	20	116–7871	CAB CAB
3	15	244–4177	BCA BCA
4	20	116–7871	CBA CBA
5	18	142–7009	BAC BAC

Table 1: Patient map and evaluation summary

The speech material were 24 consonants in /aCa/ context and 19 vowels, including diphthongs, in /hVd/ context. The stimuli were recorded in an anechoic chamber by one male speaker and one female speaker who were both audiologists with extensive experience with live-voice testing and spoke Australian English. The utterances were normalised by placing 60 ms of silence before and after the syllables and a 20 ms ramp at each end to eliminate clicks. The tokens were then equalized to have the same RMS levels.

Two samples of each token were presented in each randomized test sequence. The level at each session was set to 75 dBA. Testing was performed using a repeated ABC protocol and the order of testing was balanced between the patients (see Table 1). The patients attended two testing sessions for each repetition of a strategy. The patients were given two weeks of take-home experience with each strategy followed by two weeks of testing before the next strategy was introduced. In the re-test phase, the patients used the strategy for one week before two further weeks of testing. The patients continued to use the appropriate strategy during the testing as well.

In a testing session the patients were given a printed list of the syllables in alphabetical order. The syllables were then played in the order displayed on the list to allow the patients to familiarize themselves with the speaker. Two randomised lists of syllables were then presented to the patients. Each phoneme was presented twice in each list. After completion of the two lists, the syllables of the same type were presented using the other speaker's voice. The full syllable set was first played in alphabetical order to familiarize the patients with the new voice and then the two test lists were presented.

RESULTS

Phoneme recognition results are displayed in Figures 1 and 2. Repeated measures ANOVA revealed a significant ($p < 0.05$) learning effect between the test and re-test repetitions that varied between strategies. This probably occurred because the patients had already been familiar with the 250 cycles/s rate and it took some time to become accustomed to the higher rates. Therefore, only the second trial was considered in further statistical analyses.

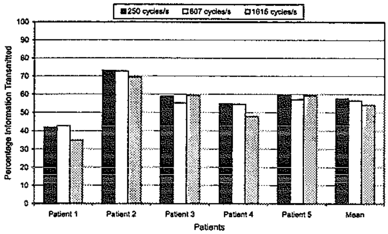

Figure 1: Consonant recognition results.

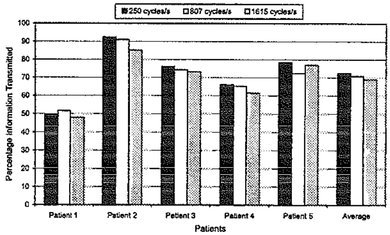

Figure 2: Vowel recognition results.

Consonants

There was no significant overall effect of rate of stimulation on consonant recognition performance. However, there were significant differences between the overall recognition abilities of the patients. There was a significant interaction between strategy and patient. When analysed separately, patient 1 had a significant strategy effect: performance was better for 807 cycles/s than for 1615 cycles/s. Patients 1, 2 and 4 followed the same trend as the overall average, although patients 1 and 4 had steeper declines with increasing rate of stimulation. Patients 3 and 5 had different trends, with the performance of strategy B tending to be less than strategies A and C.

There was a patient-speaker interaction, which indicates that the preferred speaker differed between patients. Overall, performance with speaker 1 was greater than with speaker 2, but the amount of difference varied between the patients.

Vowels

There was also no significant overall effect of rate of stimulation on vowel recognition performance. Again there were significant differences between the overall recognition abilities of the patients. There was a significant interaction between strategy and patient. When analysed separately, patient 3 showed an individual significant strategy effect: performance was better for 250 cycles/s than for 807 cycles/s.

There was also a patient-speaker interaction, which indicates that the preferred speaker differed between patients. Performance with speaker 2 was better for patients 1, 4 and 5, while speaker 1 was better for patient 3.

Log-Linear Modelling

The consonant confusion matrices were analysed using hierarchical log-linear modelling (Bell et al., 1986). Vowel confusion matrices were not analysed in this way because the high performances with these phonemes created sparse matrices. The models that were derived for each patient-speaker combination are shown in Table 2. The factors were rate of stimulation (rate, 3 levels), stimulus phoneme (stimulus, 24 levels), and response phoneme (response, 24 levels). Since the models are hierarchical, significant lower order terms are only displayed in Table 2 if they are not part of a significant interaction; in this case, there were no significant first-order terms satisfying this criterion.

In all cases, there were highly significant interactions between stimulus and response because all of the patients displayed some ability to distinguish between phonemes. Most patients exhibited interactions between rate of stimulation and response, showing that the patterns of confusions for the consonant phonemes varied significantly with rate of stimulation. Patient 4 with speaker 1 also required

Patient	Speaker 1	Speaker 2
1	S*R	rate * response + S*R
2	S*R	S*R
3	rate * response + S*R	rate * response + S*R
4	rate * stimulus + rate * response + S*R	rate * response + S*R
5	rate * response + S*R	rate * response + S*R

Table 2: Hierarchical log-linear models for consonant confusions. S*R is stimulus * response, a common interaction term in all models.

a rate * stimulus term showing that some phonemes were better perceived with a particular rate of stimulation.

Information Transmission Analysis

The relationship between rate of stimulation and phoneme confusions was investigated using information transmission analysis (Miller and Nively, 1955). Distinctive features were chosen from those described by Miller and Nively (1955), Singh (1968), Chomsky and Halle (1968) and Singh et al. (1972). Several features, such as *vocalic*, *rounded*, *low*, and *lateral* were not included since they distinguish few phonemes and *tense* was also not used since it is very similar to *voiced*.

The percentages of information transmitted for the distinctive features are plotted in Figures 3 and 4. Two manner of articulation features (see Figure 3), *nasal* and *continuant*, were both better transmitted with the 1615 cycles/s rate of stimulation . Recognising both of these types of phonemes required detection of the presence or absence of a discontinuity in the speech. The high rate of stimulation thus improved the perception of discontinuous sounds.

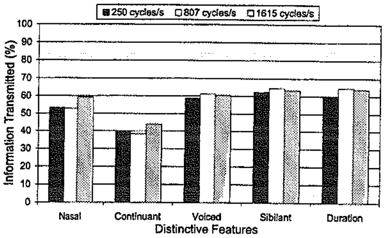

Figure 3: Information transmitted for manner of articulation features.

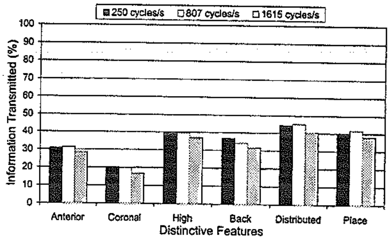

Figure 4: Information transmitted for place of articulation features.

Sibilant and *duration* are two different classifications of fricative phonemes. *Sibilant* phonemes contain considerable high-frequency noise and include the alveolar and palatal fricatives as well as the affricates. The *duration* feature includes the long duration fricatives only (the palatal fricatives and affricates), which contain the most frication. The higher-rate strategies tend to transmit more information for these phonemes that have considerable noise. Thus, as expected, a higher rate of stimulation is beneficial for perceiving the presence or absence of the highly fricative phonemes by increasing the stochastic nature of neural response.

There was little difference between the strategies in the perception of voicing. A small trend of increased information transmission with higher rates of stimulation was evident.

The place of articulation features show markedly different trends (see Figure 3). The final feature in Figure 3, *place*, is the four-place feature devised by Singh et al. (1972) with values: *front*, *alveolar*, *palatal* and *back*. The 1615 cycles/s rate was the worst for transmitting information for all features examined. There was little difference between the 250 cycles/s and 807 cycles/s rates. The rate of processing for the 1615 cycles/s rate was not different from 807 cycles/s, but the perception of spectral structure was much better for the latter strategy. The higher pulse rate, while able to convey good temporal information, was making the perception of spectral cues more difficult.

DISCUSSION

None of the different rates of stimulation was the best for conveying all information about phoneme identity. Envelope information and frication noise tended to be better perceived with higher rates of stimulation. However, place of articulation features were better recognised with lower rates of stimulation. Distinguishing these latter features requires perception of spectral detail such as formant locations and general spectral distributions. The highest rate of stimulation appears to have masked these fine spectral characteristics.

There are influences that may have a detrimental effect on performance at high stimulation rates. The impact of the refractory periods of the nerve fibres (Stypulkowski and Van Den Honert, 1984; Parkins, 1989; Bruce et al., 1999) on perception of electrical stimulation is not fully understood, especially for the complex electrical stimulation for speech sounds. High rates of stimulation may increase interaction between channels by summation of pulses from different electrodes (Wilson, 1997). The highest rates also have stimulation intervals well within neural refractory periods that, further complicated by electrical current spread, may lead to blurring of spectral detail.

The effect of spectral blurring was investigated by ter Keurs et al. (1991) who tested consonant and vowel perception with normal-hearing listeners when speech was smeared over bandwidths of up to 2 octaves in frequency regions from 100 to 8000 Hz. As blurring increased, consonant place of articulation became confused but manner of articulation was still well perceived. Spectral smearing is already a problem for cochlear implant users because of the small number of electrodes that are used. High rates of stimulation may exacerbate this problem.

The results of this study have suggested a new coding strategy, which is currently under development, where two stimulation rates are used simultaneously. A low rate of stimulation, between 250 and 800 cycles/s per electrode, is used in the 0–2700 Hz range, which encode the first two formants and, therefore, most place of articulation information. The spectral detail will thus be preserved in the region where it is most important. The remaining electrodes (2700–8000 Hz) are stimulated using a high rate of stimulation, above 800 cycles/s per electrode, to provide improved presentation of frication and information about precise temporal events. Representation of frication will benefit from the more stochastic nature of neural firing, which results from high-rate stimulation. Temporal events, such as onset of pitch periods and plosive bursts, tend to be present across the full spectrum and so there should be information transmission benefit when presenting high rate stimulation on the high-frequency electrodes.

CONCLUSION

Different rates of electrical stimulation provided different benefits for consonant recognition. High rates gave improved manner of articulation information because increased temporal resolution provided more information about variations in amplitude envelope. However, perception of place of articulation cues was reduced with high rates, probably because of spectral smearing due to refractory periods of auditory neurons. Therefore, there was no overall improvement in recognition between the three rates of stimulation that were investigated.

A new speech processing strategy is currently under development and testing that uses a low stimulation rate on apical (low frequency) electrodes of a cochlear implant, where place of articulation information is largely present, and a high stimulation rate on the basal (high frequency) electrodes, to better encode frication and envelope information.

ACKNOWLEDGEMENTS

This work was supported by the Sidney Myer Fund. We are grateful to the participants who were involved. We also thank Mr. Andrew Vandali, Ms. Kerry Plant and Ms. Lesley Whitford from the Cooperative Research Centre for Cochlear Implant, Speech and Hearing Research for their help with the research participants. We thank Dr. Chris James for recording and editing the speech stimuli and writing the testing program, and Dr. Richard Dowell and Ms. Karyn Galvin for uttering the speech samples.

REFERENCES

Bell, T. S., Dirks, D. D., Levitt, H. and Dubno, J. R. (1986), Log-linear modeling of consonant confusion data, *Journal of the Acoustical Society of America* **79**(2), 518–525.

Brill, S. M., Gstottner, W., Helms, J., V. Ilberg, C., Baumgartner, W., Muller, J. and Kiefer, J. (1997), Optimization of channer number and stimulation rate for the fast continuous interleaved sampling strategy in the COMBI 40+, *American Journal of Otology* **18**, S104–106.

Bruce, I. C., White, M. W., Irlicht, L. S., O'Leary, S. J. and Clark, G. M. (1999), The effects of stochastic neural activity in a model predicting intensity perception with cochlear implants: Low-rate stimulation, *IEEE Transactions on Biomedical Engineering* **46**(12), 1393–1404.

Chomsky, N. and Halle, M. (1968), *The Sound Pattern of English*, Harper & Row, New York.

Fu, Q.-J. and Shannon, R. V. (1999), Factors affecting speech performance in cochlear implant users, *Abstracts of the 1999 Conference on Implantable Auditory Prostheses*, Asilomar, CA.

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

Miller, G. A. and Nively, P. E. (1955), An analysis of perceptual confusions among some English consonants, *Journal of the Acoustical Society of America* **27**(2), 338–352.

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

Plant, K., Whitford, L. A., Psarros, C. E., Vandali, A. E. and Clark, G. M. (1999), Parameter selection and programming recommendations for the ACE and CIS speech processing strategies, *Abstracts of the 1999 Conference on Implantable Auditory Prosthesis*, Asilomar, CA.

Singh, S. (1968), A distinctive feature analysis of responses to a multiple choice intelligibility test, *International Review of Applied Linguistics* **6**, 37–53.

Singh, S., Woods, D. R. and Becker, G. M. (1972), Perceptual structure of 22 prevocaliv English consonants, *Journal of the Acoustical Society of America* **52**, 1698–1713.

Stypulkowski, P. H. and Van Den Honert, C. (1984), Physiological properties of the electrically stimulated auditory nerve. I. Compound action potential recordings, *Hearing Research* **14**, 205–223.

ter Keurs, M., Festen, J. M. and Plomp, R. (1991), Effect of spectral envelope smearing on speech reception. I, *Journal of the Acoustical Society of America* **91**(5), 2872–2880.

Van Tassel, D., Soli, S., Kirby, V. and Widin, G. (1987), Speech wave-form envelope cues for consonant recognition, *Journal of the Acoustical Society of America* **82**, 1152–1161.

Wilson, B. S. (1997), The future of cochlear implants, *British Journal of Audiology* **31**, 205–225.