A FORMANT-ESTIMATING SPEECH PROCESSOR FOR COCHLEAR IMPLANT PATIENTS

P.J. Blamey, R.C. Dowell, A.M. Brown, P.M. Seligman and G.M. Clark

Department of Otolaryngology
University of Melbourne

ABSTRACT - A simple formant-estimating speech processor has been developed to make use of the "hearing" produced by electrical stimulation of the auditory nerve with a multiple-channel cochlear implant. Thirteen implant patients were trained and evaluated with a processor that presented the second formant frequency, fundamental frequency, and amplitude envelope of the speech. Nine patients were trained and evaluated with a processor that presented the first formant frequency and amplitude as well. The second group performed significantly better in discrimination tasks and word and sentence recognition through hearing alone. The second group also showed a significantly greater improvement when hearing and lipreading was compared with lipreading alone in a speech tracking task.

INTRODUCTION

A multiple channel cochlear implant is designed to stimulate different parts of the cochlea with electrical currents to produce different hearing sensations for the patients. The electrical stimulation patterns are controlled by an externally worn speech processor to produce hearing sensations that represent the speech sounds picked up by the processor's microphone. This "speech coding" has been implemented in different ways by different cochlear implant research groups (Millar, Tong and Clark, 1984).

The 22 channel cochlear implant developed by the University of Melbourne and Cochlear Pty Ltd uses a speech coding scheme that requires the estimation of formant frequencies and amplitudes (Clark et al, 1984). This information is then coded in terms of the place of stimulation in the cochlea and the current level of the electrical pulses. The rate of electrical pulses is equal to the fundamental frequency of the speaker's voice. The initial speech processor (F0F2) estimated the second formant frequency and determined the the place of stimulation in the cochlea by choosing the electrode to be stimulated. The 22 electrodes were used to represent the range of estimated frequencies (EF2) from 800 to 4000 Hz. In a later speech processor (F0F1F2) estimates of the first and second frequencies (EF1 and EF2) were used to select two electrodes that were stimulated in quick succession. The 22 electrodes were then used to represent the range from 300 to 4000 Hz.

This study compares the results of two groups of cochlear implant patients that used the two processors in a ten session training program immediately after implantation. Results for simple discrimination tasks, word and sentence recognition, and speech tracking are considered. These results confirm the preliminary studies in which the F0F1F2 speech processor was shown to be more effective than F0F2 using an acoustic model of the implant (Blamey, Martin and Clark, 1985) and seven patients who switched from F0F2 to F0F1F2 (Dowell et al, 1986).

Proceedings of the 1st Australian Conference of Speech Science and Technology, 1986 272-277

METHOD

The implanted prosthesis used by all patients in this study consisted of an intracochlear electrode array and a receiver/stimulator unit (Clark et al, 1984). The 22 electrodes were platinum bands 0.3 mm wide and spaced at 0.7 mm intervals along a Silastic carrier. The electrodes were inserted into the scala tympani, extending along a region that would normally correspond to a frequency range from about 600 to 8000 Hz. The receiver/stimulator was controlled and powered by an externally applied radio-frequency transmission through the skin. The device was capable of providing pulsatile electrical stimulation between any pair of electrodes at current levels from 20 uA to 1.5 mA, with pulse durations from 1 to 400 us and pulse rates up to 1400 pps.

The parameter extraction circuit of the wearable speech processor is shown in block form in Figure 1. The speech processor and its function have been described in detail by Blamey et al (1986a). The F0F1F2 speech processor was developed from the F0F2 processor by adding the F1 extraction circuits. For some speech sounds, the formant frequency estimates (EF1 and EF2) do not represent true formant frequencies. They are more correctly thought of as the frequencies of the largest spectral peaks in the frequency ranges 300 to 1000 Hz and 800 to 4000 Hz respectively. For example, the EF2 value can rise to its maximum value of 4000 Hz during the frication of an /s/ or during the burst of a /t/.

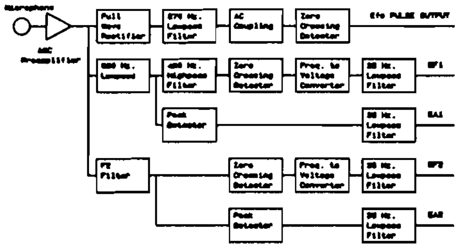

Figure 1. Parameter extraction circuits of the F0F1F2 wearable speech processor.

Twenty two postlingually deaf adult patients implanted over a three year period provided the data for this study. The main selection criteria for implantation were: 1. Profound bilateral deafness acquired after the development of normal language skills. 2. No practical benefit from conventional hearing aids. 3. No radiological contraindication to placement of the prosthesis. 4. No medical contraindication to cochlear implant surgery. 5. Positive response to electrical stimulation of the promontory. Thirteen of the patients used the F0F2 speech processor and nine used the F0F1F2 processor for their initial training program. Table 1 shows the mean age and duration of deafness for each group. The groups did not differ significantly with respect to these parameters. Etiologies in both groups included meningitis, head trauma, cochlear otosclerosis and progressive loss of unknown origin.

Following surgery and a two-week recuperation period, patients were tested to determine the threshold and dynamic range of hearing for each electrode. Then, patients underwent a ten session training program which included counselling, speech discrimination practice, evaluation and continuing adjustment of the current levels used for each electrode. During training, the patients were assessed using the speech tracking procedure (de Filippo and Scott, 1978) to monitor speech communication rates in the lipreading alone (L) and hearing plus lipreading (HL) conditions. At the completion of the training program, patients were assessed with a series of recorded speech discrimination and recognition tests in the hearing alone (H) condition and a videotaped sentence recognition test in the three conditions H, L, and HL. Most of the tests were from the Minimal Auditory Capabilities (MAC) battery (Owens et al, 1980). The tests were recorded by an adult Australian male speaker. All tests were administered at an average peak sound pressure level of 70 dBA in a sound treated room. The patients were positioned one metre from the loudspeaker sound source and used the standard ear level microphone (Knowles EB 1863) as the input to the wearable speech processor.

Table 1. Details of patient groups.

	FOF2 GROUP	FOF1F2 GROUP
No of patients	13	9
Mean age	51.3 years	50.1 years
Mean duration of profound deafness	16.7 years	12.3 years

Mean scores and standard errors were calculated for each test for the two groups of patients. The significance of differences in the mean scores were assessed using t-tests. In the case of tests involving lipreading, the mean improvement (HL score - L score) was compared.

RESULTS

Table 2. Results of prosodic tests presented 3 months post surgery in the hearing alone condition at 70 dBA.

TEST	CHANCE SCORE	FOF2 GROUP	FOF1F2 GROUP	SIGNIFICANT DIFFERENCE
one/two syllables	50%	97.0%	99.3%	$p < 0.01$
noise/voice	50%	97.2%	95.4%	NO
male/female speaker	50%	88.9%	95.1%	NO
spondee same/different	50%	90.0%	95.0%	NO
accented word	25%	72.7%	87.8%	$p < 0.01$
question/statement	50%	65.8%	87.2%	$p < 0.01$

Table 2 shows the mean results for the two patient groups on the prosodic discrimination tests of the MAC battery in the H condition. Three of the tests showed a significant difference between the two groups at the 99% confidence level. All scores for both groups were well above chance.

Table 3 shows the mean results for the closed set word recognition tests in the H condition. Three of the tests show a significant difference between the two groups, with the largest improvements occurring in the consonant recognition tests.

Table 3. Results of closed set word recognition tests presented 3 months post surgery in the hearing alone condition at 70 dBA.

TEST	CHANCE SCORE	FOF2 GROUP	FOF1F2 GROUP	SIGNIFICANT DIFFERENCE
4 choice spondee	25%	86.9%	92.2%	NO
medial vowel	25%	51.2%	57.9%	$p < 0.05$
initial consonant	25%	53.5%	66.3%	$p < 0.01$
final consonant	25%	53.5%	66.7%	$p < 0.01$

Table 4 shows the results for the open set tests of the MAC battery in the H condition. These tests required the patients to recognize words or sentences without any contextual information or lists from which to choose. These tests provide an assessment of whether patients are able to use the prosthesis as a communication device in the hearing alone (H) condition. Each of these tests showed a significantly higher score for the FOF1F2 group.

Table 4. Results of open set word and sentence recognition tests presented 3 months post surgery in the hearing alone condition at 70 dBA.

TEST	CHANCE SCORE	FOF2 GROUP	FOF1F2 GROUP	SIGNIFICANT DIFFERENCE
spondee recognition	0%	13.6%	26.0%	$p < 0.01$
CID sentences	0%	15.9%	35.4%	$p < 0.01$
monosyllabic words	0%	4.9%	12.4%	$p < 0.01$
phonemes in monosyllables	approx 5%	23.1%	33.4%	$p < 0.01$

Table 5 shows the results for lipreading assessments. The CID sentence test was a recorded test carried out at the end of the training program. The speech tracking results were obtained during the training program using a live speaker to present the material. The results represent the rate at which words read from a book ("Danny, Champion of the World", by Roald Dahl, A.A. Knopf, 1975), were correctly recognized by the patient. Every sentence had to be recognized perfectly before going on to the next one.

The rates were measured in ten minute sessions, alternating the order of L and HL conditions. The results in Table 5 are averages over the last 4 sessions for each condition. Both lipreading assessments show significantly better results for the F0F1F2 group when the differences between the HL and L scores are compared. The differences between HL and L scores were used to remove any effect of differences in lipreading skills between the groups.

Table 5. Lipreading assessment results.

CONDITION	F0F2 GROUP	F0F1F2 GROUP	SIGNIFICANT DIFFERENCE
A. CID SENTENCES: 3 months post surgery			
Lipreading alone	58.5%	46.5%	
Hearing plus lipreading	88.0%	91.4%	
Improvement (HL-L)	29.5%	44.9%	$p < 0.01$
B. SPEECH TRACKING: during training			
Lipreading alone	16.4 wpm	20.2 wpm	
Hearing plus lipreading	44.6 wpm	58.2 wpm	
Improvement (HL-L)	28.2 wpm	38.0 wpm	$p < 0.05$

DISCUSSION

This study demonstrated the superiority of the F0F1F2 speech processor over the F0F2 processor for patients who had equal exposure to the implant and equal training and usage opportunities. There were improvements in both prosodic and phonetic tests and these combined to give highly significant improvements in the open set speech recognition tests. Lipreading was also enhanced to a greater extent by the F0F1F2 processor. These results should not be interpreted as ultimate limits on the performance of patients using either processor as many of them have continued to improve over a year or more after surgery.

A detailed study of the vowel and consonant confusions of patients using the two processors (Blamey et al, 1986b) showed that the F0F1F2 processor improved vowel recognition primarily by increasing the F1 frequency information presented. Consonant recognition was improved primarily by increasing the voicing information presented. Both processors present some voicing information through the pulse rate and amplitude envelope variations during the consonants. The F0F1F2 processor provides an extra direct cue to voice onset time in stop consonants because the amplitudes of the first and second formants may rise at different times. In the F0F2 processor, only a single amplitude parameter was presented and so this cue was not available. The F0F1F2 processor also gives the patient the opportunity of comparing the energy in high and low frequency regions during fricatives.

Further improvements can still be made. At present, about one third of the patients are able to understand enough speech in the hearing alone condition to have a simple telephone conversation (Brown et al, 1985). Hearing alone speech tracking rates between 20 and 40 words per minute have also been achieved by some patients (Dowell et al 1986). These signs suggest the feasibility of developing the speech processor further as a device to be used without lipreading. Introduction of third formant information may increase the "naturalness" of the sounds produced and contribute to better vowel and consonant recognition. It is hoped that variations of the basic speech coding schemes will also make it easier for patients to perceive the EF1 and EF2 information that is extracted.

Perhaps the most important problem at present is the degradation of performance that occurs in moderate levels of background noise. Improvements in this area will probably require changes to both the parameter extraction hardware and the speech coding algorithm.

REFERENCES

BLAMEY, P.J., MARTIN, L.F.A. & CLARK, G.M. (1985) "A Comparison of Three Speech Coding Strategies Using an Acoustic Model of a Cochlear Implant", J. Acoust. Soc. Am. 77, 209-217.

BLAMEY, P.J., SELIGMAN, P.M., DOWELL, R.C. & CLARK, G.M. (1986a) "Acoustic Parameters Measured by a Formant-based Speech Processor for a Multiple-channel Cochlear Implant", J. Acoust. Soc. Am. (under review).

BLAMEY, P.J., DOWELL, R.C., BROWN, A.M., SELIGMAN, P.M. & CLARK, G.M. (1986b) "Vowel and Consonant Recognition of Cochlear Implant Patients Using Formant-estimating Speech Processors", J. Acoust. Soc. Am. (under review).

BROWN, A.M., CLARK, G.M., DOWELL, R.C., MARTIN, L.F.A. & SELIGMAN, P.M. (1985) "Telephone Use by a Multi-channel Cochlear Implant Patient. An Evaluation Using Open-set CID Sentences", J. Laryngol. Otol. 99, 231-238.

CLARK, G.M., TONG, Y.C., PATRICK, J.F., SELIGMAN, P.M., CROSBY, P.A., KUZMA, J.A. & MONEY, D.K. (1984) "A Multi-channel Hearing Prosthesis for Profound-to-total Hearing Loss", J. Med. Eng. Technol. 8, 3-8.

DE FILIPPO, C.L. & SCOTT, B.L. (1978) "A Method for Training and Evaluating the Reception of Ongoing Speech", J. Acoust. Soc. Am. 63, 1186-1192.

DOWELL, R.C., SELIGMAN, P.M., BLAMEY, P.J. & CLARK, G.M. (1986) "Evaluation of a Two-formant Speech Processing Strategy for a Multi-channel Cochlear Prosthesis", Ann. Otol. Rhinol. Laryngol. (in press).

MILLAR, J.B., TONG, Y.C. & CLARK, G.M. (1984) "Speech Processing for Cochlear Implant Prostheses", J. Speech Hear. Res. 27, 280-296.

OWENS, E., KESSLER, D.K., TELLEEN, C.C. & SCHUBERT, E.D. (1980) "The Minimal Auditory Capabilities Battery", Dept. of Otolaryngology, Univ. of Calif., San Francisco, CA.