IDENTIFICATION OF SYNTHETIC VOWEL NUCLEI BY COCHLEAR IMPLANT PATIENTS

H.H. Lim, Y.C. Tong and G.M. Clark

Department of Otolaryngology,
University of Melbourne.

ABSTRACT - Six speech processing schemes, differing in the formant frequency-to-electrode position map and the number of formant frequencies encoded were investigated . The six schemes consisted of two single-formant (F2) schemes, three two-formant (F1 and F2 or F2 and F3) schemes and one three-formant (F1, F2 and F3) scheme. Eleven steady state Australian vowel nuclei ([i], [a], [ɔ], [u], [3], [I], [ε], [æ], [Λ], [ɒ] and [ʊ]) synthesised as electrical signals were used to evaluate the relative merits of the six schemes on three cochlear implant patients. The first five vowels are long vowels and the remaining six are short vowels. The steady state formant frequencies of these vowel nuclei (Bernard, 1970) were transformed to steady state electrode positions using different formant frequency-to-electrode position maps. The confusion matrices were subjected to conditional information transmission analysis. The results showed that : (1) training, experience and adaptability to a new speech processing scheme were the main factors influencing the identification of the synthetic vowels ; and (2) adding an extra formant vowel feature to a speech processing scheme tended to decrease the amount of information transmission about the existing formant feature(s). From these synthesis results, the three-formant (F0/F1/F2/F3/B) speech processing scheme appeared to be the logical choice for future implementation in speech processors for cochlear implant patients.

INTRODUCTION

The University of Melbourne/Nucleus multichannel cochlear implant employs a speech processing strategy that uses electric repetition rate to encode the fundamental frequency, electrode position to encode formant frequency and electric current to encode intensity of the acoustic speech signals (Tong *et al.*, 1983 and Blamey *et al.*, 1987). There are, however, many ways to convert formant frequencies to electrode positions. This paper evaluates the performance of three postlingually deaf cochlear implant patients for six speech processing schemes by varying the formant frequency-to-electrode position maps and the number of formant frequencies encoded. Eleven steady state Australian vowel nuclei ([i], [a], [ɔ], [u], [3], [I], [ε], [æ], [Λ], [ɒ] and [ʊ]) from Bernard (1970) were used for the evaluation of the relative merits of the six speech processing schemes. The first five vowels were long vowels while the remaining six were short vowels. These vowel nuclei were synthesised as electrical signals using an IBM microcomputer. The electrode positions, current levels and repetition rate were time invariant for the duration of each synthetic vowel.

The use of computer synthesised vowels had several advantages over the use of live voice or recorded vowels : (1) it would eliminate the problems of speaker dependent results ; and (2) the electrical parameters (electrode positions, duration, repetition rate and electric currents) of each synthetic vowel could be precisely controlled. This would by-pass any imperfections in the formant frequency extraction process or filter design in a wearable speech processor.

METHOD

Three postlingually deaf cochlear implant patients (LL, MO and GW) participated. An array of 22 banded electrodes spaced 0.75 mm was inserted through the round window for a distance of up to 25 mm into the scala tympani. The electrodes were numbered from 1 to 22 in an apical to basal direction. The electrodes were activated in bipolar mode, in which the stimulation current was passed from a source electrode to a different sink electrode. Biphasic, charge balanced current pulses were used to activate the residual nerve fibers.

On the basis of the place-frequency formula published by Greenwood (1961), the electrode array resided in the region of the scala tympani with characteristic frequencies ranging from about 1 kHz to 11 kHz. Unnatural sensation may therefore be produced when speech signal ranging from 200 Hz to 4 kHz were encoded on these electrode positions. A period of training was expected before the patients could make optimal use of the speech information provided by a speech processor. Three formant frequency-to-electrode position maps (A, B and C) were used (figure 1). Map A allocated 13 electrodes for F2 and 4 electrodes for F3 ; map B allocated 6 electrodes for F1, 10 electrodes for F2 and 3 electrodes for F3 ; and map C allocated 8 electrodes for F1 and 8 electrodes for F2. There were two speech processing schemes implemented using map A (F0/F2/A and F0/F2/F3/A) ; three speech processing schemes using map B (F0/F2/B, F0/F1/F2/B and F0/F1/F2/F3/B) ; and one speech processing scheme using map C (F0/F1/F2/C). F0/F2/A was a subset of F0/F2/F3/A in that F3 information was not encoded in F0/F2/A. Similarly, F0/F2/B and F0/F1/F2/B were subsets of F0/F1/F2/F3/B in that F1 and F3 information was not encoded in F0/F2/B and F3 information was not encoded in F0/F1/F2/B. The F1-F2 electrode plot of the eleven vowels for map B are shown in figure 2. F0/F2/B and F0/F1/F2/B schemes are similar to the F0F2 and F0F1F2 schemes used in the University of Melbourne/Nucleus speech processing schemes. LL had had 3.5 years of experience with F0/F2/A. MO had had 6 weeks of experience with F0/F2/A and 1 year of experience with F0/F1/F2/B ; while GW had had 4 years of experience with F0/F2/A and 15 months of experience with F0/F1/F2/B. The six schemes were tested in a randomised order on MO and GW. For LL, F0/F1/F2/B and F0/F1/F2/F3/B were tested on her first while she was using a F0/F2/A wearable speech processor. F0/F1/F2/B was retested on LL 5 months after her speech processor was converted to F0/F1/F2/B speech processor.

For all six schemes, electric repetition rate was fixed at 125 rep/s. A fixed pulse width of 100 µs/phase was used. All eleven vowels were activated at comfortable loudness level for all schemes. For the two single-formant (F0/F2/A and F0/F2/B) speech processing schemes, each vowel was synthesised by activating a single bipolar electrode pair activated at comfortable loudness during a stimulus period (inverse of repetition rate) ; for the two-formant (F0/F2/F3/A, F0/F1/F2/B and F0/F1/F2/C) speech processing schemes, each vowel was synthesised by activating two bipolar electrode pairs at half comfortable loudness in quick succession during a stimulus period ; and for the three-formant (F0/F1/F2/F3/B) scheme, each vowel was synthesised by activating three bipolar electrode pairs at about 33 % comfortable loudness in quick succession during a stimulus period. For multi-formant speech processing schemes, the onset times of the two or three biphasic electric current pulses within a stimulus period were separated by 800 µs. Loudness balance was conducted for all synthetic vowels within each scheme by adjusting the current levels of the component electrode pairs as described in Tong and Clark (1985 and 1986).

At least six consecutive sessions of testing was conducted on each scheme. Each session consisted of three or four blocks of single interval trials. Each block consisted of four presentations of the eleven vowels in random order. Before each block of test, the patient was allowed to listen to the eleven vowels in a fixed order. The test for each scheme was terminated if the number of consecutive sessions was at least six and the difference in percentage correct scores between the last session and the average of the two sessions prior to the last was less than twice the standard error of the percentage correct score of the last session. These termination criteria corresponded to a 95 % probability that the score of the last session was worse than the average of the two previous sessions. The results were subjected to information transmission analysis.

PERCENTAGE CORRECT SCORES

The overall percentage correct scores for all three patients in the six processing schemes are shown in figure 3. For F0/F1/F2/B, symbol LL represents the performance of LL before her F0/F1/F2/B speech processor conversion and symbol (LL) represents the performance of LL after 5 months of experience with her F0/F1/F2/B speech processor. All patients performed worst in F0/F1/F2/C ; this poor performance could be due to the unfamiliar ranges of electrode positions used for encoding F1 and F2 in F0/F1/F2/C. LL performed best with F0/F2/A and F0/F1/F2/B (after experience with her F0/F1/F2/B wearable speech processor). The substantial increase in F0/F1/F2/B score of LL after experience with her F0/F1/F2/B wearable speech processor compared to the score before her F0/F1/F2/B speech processor conversion indicated that training and experience were influencing the identification performance of LL. MO performed best with F0/F1/F2/F3/B ; she was easily adapted to this new scheme. It could be possible that LL may performed better with F0/F1/F2/F3/B than F0/F1/F2/B after gaining experience as indicated by her better score of F0/F1/F2/F3/B over F0/F1/F2/B before gaining experience. There was very little variation in the scores of GW for all six schemes. Since GW was reluctant to changes in speech processing scheme, one might speculate that if F0/F1/F2/F3/B had first been introduced to him instead of F0/F2/A, his performance would have been best for F0/F1/F2/F3/B.

INFORMATION TRANSMISSION ANALYSIS

The transmitted information about vowel features (F1, F2, F3 and DUR) were calculated using the procedure described in Miller and Nicely (1955). To analyse the redundancy among the features, the conditional transmitted information about a feature of interest was calculated by partialling out the effects of one or more of the other features using the procedure described in Wang and Bilger (1973). The transmitted information was expressed in bits per stimulus instead of percentage as different schemes used different ranges and number of electrode positions to encode each formant frequency. $(F1)_{DUR}$ denotes the conditional transmitted information about F1 with the effects of duration being partialled out. Similarly $(F1)_{DUR,F2}$ denotes the conditional transmitted information about F1 with the effects of duration and F2 being partialled out. The percentage correct scores and information transmission results are shown in Table I. Pairwise performance comparisons between schemes were conducted using Student t-tests. The comparison results were evaluated at 0.05 level of significance. The results indicated that : (1) duration feature was efficiently transmitted (at least 75 % for all schemes except MO in F0/F1/F2/C). There was few confusions of long vowels with short vowels. This was of little surprise as all patients were postlingually deaf ; and (2) adding an extra formant feature to a speech processing scheme tended to decrease the amount of information transmission about the existing formant feature(s) ; this could be illustrated by the significantly smaller F2 transmission for GW and MO in F0/F2/F3/A compared to F0/F2/A ; the significantly smaller F2 transmission for all patients in F0/F1/F2/B compared to F0/F2/B ; the significantly smaller F2 transmission for GW and the significantly smaller F1 transmission for all patients in F0/F1/F2/F3/B compared to F0/F1/F2/B ; (3) the performance for F0/F1/F2/B in which the patients had had previous experience was better than that for F0/F1/F2/C without previous experience ; (4) for F0/F1/F2/B, LL showed a substantial increase in all information transmission after 5 months of experience with her F0/F1/F2/B wearable speech processor compared to the performance before her speech processor was converted to F0/F1/F2/B. This indicated that training and experience could influence the identification performance ; (5) Comparing F0/F1/F2/B and F0/F1/F2/F3/B, LL (before her speech processor conversion to F0/F1/F2) did not appear to have utilised the information provided by the F3 signal component in F0/F1/F2/F3/B while GW and MO were able to make use of the information provided by the F3 signal component.

CONCLUSIONS

The results confirmed that experience, training and adaptability were major factors influencing the identification of the steady-state synthetic vowels. Although multi-formant schemes may out-perform single-formant schemes, adding extra formant features tended to decrease the amount of information transmission for existing formant feature(s). Taking into considerations the effects of training, experience, and adaptability and the pairwise comparison results of scheme performance, the three-formant (F0/F1/F2/F3/B) speech processing scheme appear to be the logical choice for future implementation in speech processors for cochlear implant patients.

REFERENCES

BERNARD, J.R. (1970) "Toward the Acoustic Specification of Australian English", Zeit. Fur Phonetick 23, 113-128.

BLAMEY, P.J., SELIGMAN, P.M., DOWELL, R.C. & CLARK, G.M. (1987) "Acoustic parameters measured by a formant-based speech processor for a multiple channel cochlear implant", J. Acoust. Soc. Am. 76, 104-110.

GREENWOOD, D.D. (1961) "Critical bandwidth and the frequency coordinates of the basilar membrane", J. Acoust. Soc. Am. 33, 1344-1356.

MILLER, G. A. & NICELY, P.E. (1955) "An analysis of perceptual confusions among some English consonants", J. Acoust. Soc. Am. 27, 338-352.

TONG, Y.C., DOWELL, R.C., BLAMEY, P.J. & CLARK, G.M. (1983) "Two component hearing sensations produced by two-electrode stimulation in the cochlea of a totally deaf patient:, Science 219, 993-993.

TONG, Y.C. & CLARK, G.M. (1985) "Absolute identification of electric pulse rates and electrode positions by cochlea implant patients", J. Acoust. Soc. Am. 77, 1881-1888.

TONG, Y.C. & CLARK, G.M. (1986) "Loudness summation, masking and temporal interaction for sensations produced by electric stimulation of temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea", J. Acoust. Soc. Am. 79, 1958-1966.

WANG, M.D. & BILGER, R.C. (1973) "Consonant confusions in noise : A study of perceptual feature", J. Acoust. Soc. Am. 54, 1248-1266.

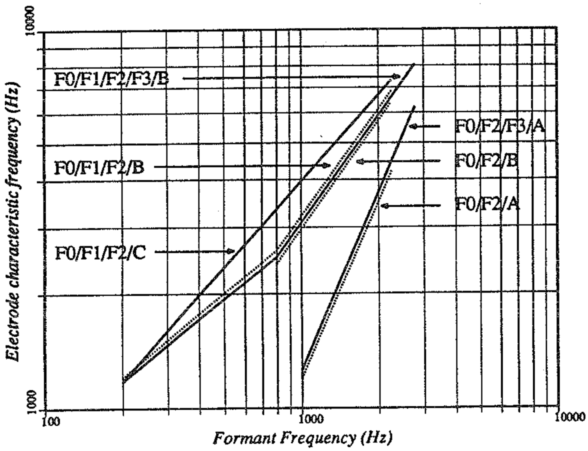

Figure 1. The formant frequency-to-electrode position functions

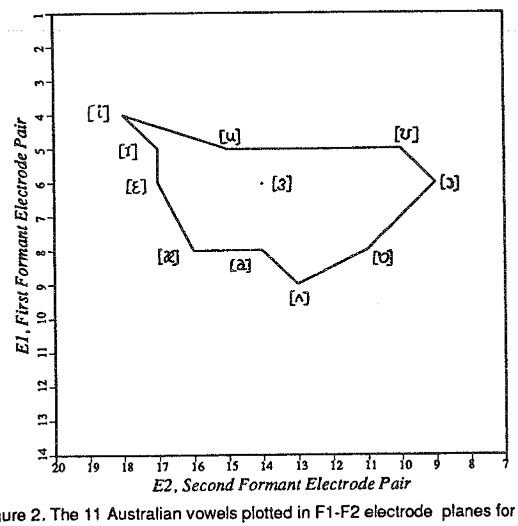

Figure 2. The 11 Australian vowels plotted in F1-F2 electrode planes for map B

Table I. Information transmission analysis results of all patients in six speech processing strategies.

Patient	Scheme	% correct	Information transmitted (bits): TOTAL	DUR	$(F2)_{DUR}$	$(F1)_{DUR,F2}$	$(F3)_{DUR,F2}$	$(F3)_{DUR,F2,F1}$	% a/c for
GW	F0/F2/A	70	2.49	0.87	1.42				92
MO		77	2.73	0.99	1.52				92
LL		73	2.64	0.96	1.42				90
GW	F0/F2/B	61	2.37	0.92	1.20				90
MO		62	2.42	0.94	1.19				88
LL		58	2.38	0.99	1.19				92
GW	F0/F2/F3/A	60	2.31	0.91	0.82		0.06		75
MO		66	2.46	0.94	1.21		0.11		92
LL		69	2.56	0.96	1.33		0.04		90
GW	F0/F1/F2/B	58	2.29	0.95	0.83	0.19			86
MO		76	2.64	0.94	0.92	0.36			84
LL		49	1.93	0.95	0.39	0.15			77
(LL)		73	2.55	0.98	0.85	0.45			89
GW	F0/F1/F2/C	53	2.08	0.92	0.60	0.14			80
MO		41	1.70	0.66	0.30	0.08			61
LL		47	1.83	0.93	0.44	0.06			77
GW	F0/F1/F2/F3/B	57	2.13	0.99	0.63	0.10		0.07	84
MO		85	2.95	0.99	0.87	0.53		0.16	86
LL		57	2.13	0.98	0.50	0.33		0.01	85

Figure 3. The overall vowel percentage correct scores.