ABSOLUTE IDENTIFICATION BY COCHLEAR IMPLANT PATIENTS OF SYNTHETIC VOWELS CONSTRUCTED FROM ACOUSTIC FORMANT INFORMATION

W.K. Lai, Y.C. Tong, and G.M. Clark

Dept. of Otolaryngology
University of Melbourne

ABSTRACT - Five speech processing schemes for presenting speech information to multiple-channel cochlear implant patients were investigated and compared. Tabulated data for formant frequencies of the natural vowels (i , ɪ , ɛ , æ, a, ɔ, ʊ , u , ʌ, ɜ , ɒ)) were coded into the parameters of the electric stimuli used in the cochlear implant, and these electric stimuli or synthetic vowels were presented to two patients in a single-interval absolute identification task. The results suggest that when first and second formant speech information is coded into the pulse rate as well as the electrode position, it is possible for the performance in the identification task to be significantly improved, compared to when the same information is coded into the electrode position only.

INTRODUCTION

In presenting useful acoustic speech information to a multiple-channel cochlear implant patient, the two most important electric signal parameters for coding this information are the electrode position and the electric pulse rate. Several methods of presenting this information through a combination of these two electrical signal parameters are investigated here.

In the speech coding scheme presently being used with the University of Melbourne/Nucleus cochlear implant, the speech processor converts the fundamental frequency, f0, of the acoustic speech signal into electric pulse rate, and the first and second formant frequencies, f1 and f2, into electrode positions. The two selected electrodes are stimulated in quick succession with a biphasic pulse each and this excitation of the two electrodes is repeated at the rate determined by f0. There is no temporal overlap between any of the pulses. This is called the f0-f1-f2 strategy. In this study, the performance of four speech coding schemes which utilise pulse rate as well as electrode position to code f1 and f2 information will be compared to the f0-f1-f2 strategy which uses only electrode position to code formant information. The pulse trains for these four schemes are specified such that there is also no temporal overlap between any of the pulses. The fundamental frequency, f0, is not used in the four coding schemes being investigated here.

METHOD

Electric stimuli constructed using data on the first and second formant frequencies of eleven male Australian vowels (i , ɪ , ɛ , æ, a , ɔ , ʊ , u , ʌ, ɜ , ɒ) as documented by Bernard (1970) were used in the absolute identification task. The formant frequencies of each of these vowels are used to derive the electrode position and pulse rate for producing the corresponding electrical stimulus. The duration of each stimulus is according to the corresponding value suggested by Bernard. The electric stimulus parameters are time invariant over the duration of the stimulus and, as such, are approximations of the steady-state portion of the corresponding natural vowel. These stimuli are henceforth referred to as 'synthetic vowels'.

The electrode array inserted into the scala tympani consists of twenty-two platinum electrodes equally spaced 0.75 mm apart, numbered 1 to 22 in an apical to basal direction. Electrodes are activated in bipolar pairs 1.5 mm apart with biphasic current pulses. A bipolar pair is numbered according to its basal member. As the cochlea is tonotopically organised, each electrode corresponds to a different location in the scala tympani with a different acoustic characteristic frequency. Using a formula described by Greenwood (1961), the corresponding acoustic characteristic frequency for each electrode is first calculated. A logarithmic-to-logarithmic transformation was then used to convert formant frequency to acoustic characteristic frequency, or in other words, electrode position. This procedure was used to code formant information in the f0-f1-f2 strategy, and is also used for selecting the electrode pairs in the other schemes under investigation here.

Four different logarithmic-to-logarithmic transformations, R1R2, R1+R2–, R1–R2+ and R1+R2+, were used for coding formant frequency into pulse rate. The + denotes an expanded range of pulse rates compared to R1R2, while the – denotes a reduced pulse rate range compared to R1R2. The four transformations are described as follows:

R1R2:	A straight-line function that transforms all formant information between 200 to 2500 Hz into pulse rates between 100 to 400 pulses per second (pps). The range of acoustic frequencies used here are representative of typical values for f1 (lower bound 200 Hz) through to f2 (upper bound 2500 Hz).
R1+R2–:	The transformation is described in two parts: f1 is mapped into a larger range of r1 values (80 to 250 pps) than in R1R2 above, while f2 has a correspondingly smaller range (250 to 400 pps). In graphical form, the transformation is described by a continuous function consisting of two straight lines of different slopes, with the change in gradient occuring at 800 Hz (typical upper limit of f1 and lower limit of f2).
R1–R2+:	Similar to R1+R2– above, except that the r1 range is now smaller (80 to 150 pps) compared to R1R2 while the r2 range is larger (150 to 600 pps). The f2 frequencies for the set of 11 synthetic vowels in the stimulus set ranges from 820 Hz to 2250 Hz. With the pulse rate range of 150 to 400 pps in R1+R2– above, a corresponding r2 range of only 151 to 282 pps is obtained. Increasing the upper limit of the pulse rates to 600 pps changes this range to 152 to 366 pps, allowing better representation of the range of pulse rates (80 to 400 pps) being investigated.
R1+R2+:	f1 is mapped onto the range 80 to 250 pps as in R1+R2– above, while f2 is mapped onto the range 150 to 600 pps) as in R1–R2+ above. The resultant transformation is now represented by two straight lines of different slopes.

The f0-f1-f2 strategy is abbreviated to R0R0, indicating that its two pulse rates are fixed according to f0, and are both set at 125 pps for the eleven synthetic vowels. All five coding schemes are illustrated in Figure 1 below.

Two post-lingually deaf patients, GW and MO, who have been implanted with the University of Melbourne/Nucleus cochlear implant were tested. GW was a 69-year-old male who suffered a progressive hearing loss after a bomb blast in 1944, and MO was a 46-year-old female who had a

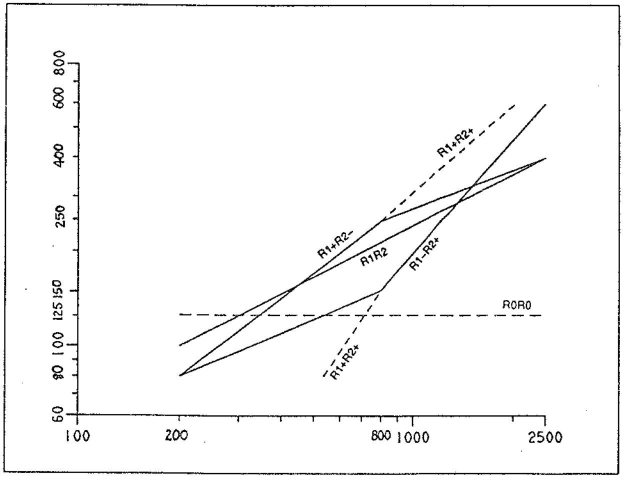

Figure 1. Functions for deriving pulse rates.

congenital hearing loss due to rubella and suffered further hearing impairment due to a head injury as well as recurrent otitis media. Pre-operatively, both patients showed that they suffered profound to total hearing loss and received no significant benefits from using conventional hearing aids. Both subjects have been using the cochlear implant system continuously for several years now and their current speech processor had been programmed with the f0-f1-f2 (R0R0) strategy at the time this study was being carried out. Both patients have also participated in other vowel studies using R0R0 such as that reported in Tong et al (1988). Thus, both subjects are familiar with this particular coding scheme.

The 11 synthetic vowels were presented to the subjects in a single-interval absolute identification task. The 11 stimuli were balanced for loudness as follows. The two electrode pairs of each signal were first balanced against each other for loudness. The current amplitude on one electrode pair was increased past threshold until a comfortably loud signal corresponding to a loudness estimate of 50 was heard. In the meantime, the current amplitude on the other electrode pair was kept to a minimum which is known to be inaudible. The current amplitude was then reduced until the loudness was estimated to be 30. This was repeated with the other electrode. The two electrode pairs were then presented together at these two respective current amplitudes to the subject. The current amplitudes on both electrode pairs were then increased or decreased together by the same amount until the loudness of resultant sound was estimated to be 50. This was done for all 11 synthetic vowels.

A written list of the 11 vowels to be identified was given to the subject to assist with the identification. The testing was done over several sessions, each session comprising of six or seven blocks of tests. A block consists of presenting each vowel a total of four times in random order with the other vowels. Thus, there are 44 presentations of the vowels in each block. For each presentation, the subject was also given feedback on whether the response was correct or wrong, and if it was wrong, what the right answer should have been. After every three or four consecutive blocks, the subject was given a five minute break.

The coding schemes were tested in the order R1R2, R1+R2–, R1–R2+, R0R0 and then R1+R2+ by both subjects. The first few sessions using a particular scheme were treated as training sessions until it was felt that the subjects had become sufficiently familiar with the set of synthetic vowels. Subsequent sessions were then taken to be proper tests. At least 10 blocks of tests were carried out in these proper test sessions, and testing for each scheme was terminated when the last three blocks of tests gave an overall percentage correct score within 5% of one another. The results from the final ten blocks were then used to construct the confusion matrices to be analysed for performance and comparison.

The results from both subjects for each of the four schemes, R1R2, R1+R2–, R1–R2+ and R1+R2+ were compared against the corresponding results for R0R0. The confusion matrices for each scheme were evaluated for the overall percentage-correct score, which is simply the ratio of the sum of the diagonal elements of a confusion matrix to the sum of all elements in the same matrix. The significance of differences in the score of each of the four schemes compared to R0R0 were assessed using a Student's t-test. An asterisk, *, indicates that the score is significantly better than the R0R0 score at a confidence level of $Pr > 0.95$. Two asterisks would correspond to a confidence level of > 0.975. Parentheses around the asterisks indicate that the score is worse than the R0R0 score.

The confusion matrices were also subjected to information transmission (IT) analysis according to the methods described by Miller & Nicely (1955), Wang & Bilger (1973) and Tong et al (1988). The confusion matrices were analysed firstly for the overall amount of information transmitted in binary bits. The amount of information transmitted on features of the stimuli such as the duration, f1 and f2 (in the form of electrode and pulse rate information) were evaluated next. Lastly, the amount of information transmitted on f1 (f2), with the effects of the duration information and f2 (f1) information respectively partialled out, was evaluated. The results for each of the four schemes were again compared to that from R0R0 and the differences assessed for statistical significance with t-tests. The same notation used above for the overall percentage-correct score is used here for the t-test results.

RESULTS

The overall percentage-correct scores for both subjects with all five schemes are shown in Table 2 below. The absence of an asterisk beside a score indicates that the score is not significantly different from the R0R0 score at confidence levels greater than 0.95. It should be noted that the results for MO, being very high (> 80%), are susceptible to 'ceiling effects' or 'saturation effects'.

	Subject	
Scheme	GW	MO
R1R2	79.93% *	81.26%
R1+R2–	71.21%	94.19% *
R1–R2+	62.35%	92.87%
R1+R2+	71.66%	90.53%
R0R0	69.25%	85.86%

Table 2. Overall percentage-correct scores

Table 3 below summarises the results from information transmission analysis. All figures shown are in units of binary bits of information.

	overall	dur	f1	f2	$f1_{dur,f2}$	$f2_{dur,f1}$
GW						
R1R2	2.8071	0.9754	0.5823	1.5256	0.1407**	0.8682
R1+R2–	2.4804	0.9793	0.3510	1.2357	0.0686	0.7423
R1–R2+	2.3206	0.9936	0.3480	0.8291(*)	0.0780	0.3887(**)
R1+R2+	2.5477	0.9757	0.4169	1.2267	0.0896	0.6863
R0R0	2.5348	0.9602	0.4354	1.2338	0.0663	0.6567
MO						
R1R2	2.8027(*)	0.9644	0.8422	1.5516(**)	0.2022	0.7475(**)
R1+R2–	3.2235**	0.9940	0.8626	1.7148	0.2478	0.8954
R1–R2+	3.1496	0.9735	0.8661	1.6747	0.2462	0.8843
R1+R2+	3.0945	0.9940	0.7428	1.6338	0.2389	0.8601
R0R0	2.9573	0.9940	0.8430	1.7182	0.2254	0.8854

Table 3. Summary of IT analysis

DISCUSSION

The overall percentage correct score serves only as a rough guide in comparing the four different schemes against R0R0. The observation that there is a significant improvement in performance observed with R1R2 for GW and with R1+R2– for MO indicates that the extra pulse rate information is being perceived by the subjects. However, as different subjects tend to be sensitive to different ranges of pulse rates (Tong & Clark, 1985), it is not a surprise that the four schemes, with R1R2 and R1+R2– in particular, produced different results for the two subjects. GW, who participated as subject PO6 in the test reported in Tong & Clark (1985), had difficulty discriminating between rates above 250 pps, suggesting a greater sensitivity towards lower pulse rates. That GW does not find the expanded r1 or r2 pulse rate ranges of R1+R2–, R1–R2+ and R1+R2+ above to be beneficial to the identification task at hand could be due to such a preference for lower pulse rates. MO did not participate in the above test and thus, no such comparison can be made.

Information transmission analysis results suggest that GW has made better use of f1 information coded into r1 with R1R2, as seen with the significant improvement in transmission of $f1_{dur,f2}$. This, however, is not reflected in a significant increase in the overall amount of information transmitted, although the percentage correct score indicates significant improvement. This implies that the overall amount of information transmitted is not directly affected by individual contributions from the identified features of the signals. Instead, the overall amount of information perceived probably depends on a more complex combination of the information transmitted on the different features. The results for R1–R2+, with its expanded r2 range as well as a reduced r1 range, indicate significantly poorer transmission of f2 information compared to R0R0. Although extra pulse rate information has been introduced, the amount of f2 information perceived by GW has decreased. A possible cause of this would be poor usage of the available pulse rate information due to insufficient familiarity of the subject with the r1 and r2 information being presented. For instance, confusion in differentiating r2 information from r1 information could then have produced the poorer perception of f2 information.

The results for MO are more difficult to analyse as the scores obtained with the various coding schemes are all very high. Such high scores mean that larger differences in the results have to be observed for the change in the amount of information transmitted to be considered different.

CONCLUSIONS

The results indicate that it is possible to present significantly more information to a cochlear implant patient, compared to the f0-f1-f2 strategy presently in use with the University of Melbourne/Nucleus cochlear implant system. In the presence of extra pulse rate information used to code the acoustic formant information, the amount of information transmitted on f1 can be significantly increased, as with GW using R1R2. A suitable coding scheme for further investigation should code f1 information into a pulse rate range simliar to that of R1+R2–.

ACKNOWLEDGEMENTS

The above research was carried out under a Deafness Foundation (Victoria) Postgraduate Scholarship, and was supported by NIH Grant NS-21027 and NIH Contract NOI-NS-5-2388.

REFERENCES

Bernard, J.R.L. (1970) *Toward the acoustic specification of Australian English,* Z. Phonetik 23, 113-128.

Greenwood, D.D. (1961) *Critical bandwidth and the frequency coordinates of the basilar membrane,* J. Acoust. Soc. Am. 27, 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. & Clark, G.M. (1985) *Absolute identification of electric pulse rates and electrode position by cochlear implant patients,* J. Acoust. Soc. Am. 77, 1881-1888.

Tong, Y.C., Lim, H.H. & Clark, G.M. (1988) *Synthetic vowel studies of cochlear implant patients,* J. Acoust. Soc. Am. (in press).

Wang, M.D. & Bilger, R.C. (1973) *Consonant confusions in noise: A study of perceptual features,* J. Acoust. Soc. Am. 54, 1248-1266.