Psychophysical studies investigating the use of pulse rate to encode acoustic speech information for a multiple-electrode cochlear implant

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

University of Melbourne, Department of Otolaryngology, The Royal Victorian Eye and Ear Hospital, East Melbourne, Vic. 3002.

ABSTRACT

Two psychophysical studies were conducted on cochlear implant recipients using a place/rate speech coding strategy which encodes f1 and f2 formant information into two different electrode positions as well as the pulse rates presented respectively on the two selected electrode pairs. The results indicated that a significant increase was achieved in the transmission of information encoded into pulse rates between 80 to 250 pps, but not for higher pulse rates. Also, the pulse rates presented on each of the two electrode pairs were found to be perceptually partially independent, which means that the pulse rates presented on two electrode pairs may be used for encoding more than one acoustic speech feature that could be useful for speech perception. Furthermore, the apical pulse rate was found to be perceptually more dominant than the basal pulse rate.

INTRODUCTION

The studies presented in this paper investigate what happens when acoustic formant information such as F1 and F2 is also encoded into the pulse rates presented on two electrode pairs selected in the above manner. One speech coding strategy to do this may be described as follows: F1 is used to select a bipolar electrode pair as well as the pulse rate to be presented on this electrode pair. Similarly, F2 is encoded into the location of a second electrode pair as well as the pulse rate presented on this electrode pair. The two pulse rates delivered respectively to the two electrode pairs are therefore different from one another. This coding strategy is referred to as the "place/rate" strategy.

Two psychophysical studies investigating the feasibility of encoding formant information using the place/rate strategy are presented in this paper. In particular, the place/rate strategy is compared with the place/F0 strategy.

The first study was carried out as a pilot study involving absolute identification of synthetic vowels to compare the relative merits of a number of speech coding schemes implementing the place/rate strategy. The F1 and F2 frequencies of eleven male Australian acoustic vowels were encoded into the electric signal parameters of eleven corresponding electric stimuli using a place/rate strategy. Each stimulus involved stimulating two different electrode pairs, each at its own respective pulse rate. Such electric stimuli, simulating synthetic vowels, were used to examine the effects of the place/rate strategy on the perception of speechlike stimuli. The identification results were then subjected to information trnamsission analysis to determine how the encoded F1 and F2 information was transmitted by the different coding schemes.

Having completed the first study, a second study was carried out to investigate in greater detail the perceptual interactions that occur between two pulse rates presented respectively on two different electrode pairs. The second study investigated the perceptual independence of pulse rates presented respectively on two fixed electrode pairs. The pulse rates presented on either electrode pair were varied in an orthogonal fashion between 100 and 200 pulses per second. The results were analysed using two different methods to determine how each of the two different pulse rates was perceived and how the perception of one influenced the perception of the other. The analysis results were then used to help determine whether useful acoustic speech information should be encoded into the pulse rate of one electrode pair or the other, or into the pulse rates of both electrode pairs.

I. HARDWARE FOR ELECTRICAL STIMULATION

The electrode array consists of twenty-two platinum electrodes, equally spaced 0.75 mm apart, inserted into the basal turn of the scala tympani. The electrodes are numbered 1 to 22 in an apical to basal direction and are stimulated in bipolar pairs 1.5 mm apart with biphasic current pulses at specified rates using the Cochlear Pty. Ltd. CI1 receiver-stimulator. A bipolar pair is identified by an ordered pair (b,a) where b is the more basal member and a is the more apical member. The basal member acts as the source electrode during the positive phase of the current pulse and as the sink electrode during the negative phase.

The stimuli used in these studies all involved stimulating two bipolar electrode pairs over the duration of the stimulus. For each of these stimuli, the pulses on the two electrode pairs were presented in a temporally non-overlapping manner.

II. SUBJECTS

A total of three cochlear implantees, GW, MO and VZ participated in the two studies reported here. Prior to implantation, all were found to have profound-to-total hearing loss and received no significant benefits from conventional hearing aids.

The first study was carried out with two implantees, GW and MO, while the other study was conducted with all three implantees.

III. METHOD

A. Synthetic vowel study

This study was carried out to compare the performance and relative merits of several speech coding schemes implementing the place/rate strategy. For each coding scheme tested, a set of stimuli was constructed using tabulated data on the formant frequencies of 11 Australian acoustic vowels for male speakers.

A logarithmic-to-logarithmic transformation was used to map acoustic formant frequencies to twenty consecutive electrode positions within the scala tympani. This transformation was in turn based on an average insertion of the electrode array into the scala tympani. Using such a transformation, the F1 and F2 frequencies of each of the eleven acoustic vowels were used to select two corresponding electrode pairs, e1 and e2 respectively. Each stimulus thus involved a different pair of e1 and e2.

The pulse rate to be presented on e1, called r1, was also determined using F1. Similarly, the pulse rate on e2, called r2, was determined using F2. To convert a given range of acoustic formant frequencies into a corresponding range of pulse rates, a logarithmic-to-logarithmic (log-log) transform was used. Separate transforms were used for converting F1 and F2 frequencies to the corresponding rates, r1 and r2. Thus, each coding scheme involved one F1 transform and one F2 transform, and in general, r1 was different from r2 for each stimulus. Furthermore, as each acoustic vowel consists of a different pair of F1 and F2, each stimulus also comprised a different set of e1 and r1 as well as e2 and r2. The electric stimulus parameters were invariant over the duration of the stimulus, and therefore approximated the steady state portion of the corresponding acoustic vowel. These electric stimuli are referred to as "synthetic vowels".

Five different speech coding schemes were tested. Four of them, labelled R1R2, R1+R2–, R1–R2+ and R1+R2+, implemented the place/rate strategy and each of them thus had its own F1 and F2 transforms. The labels indicate that the pulse rates for each coding scheme were derived from both F1 and F2. The "+" after R1 in R1+R2– indicates that the range of r1 values used in that coding scheme was greater than the range of r1 values used in R1R2. Similarly, the "–" after R2 in R1+R2– indicates that the range of r2 values used in that coding scheme was smaller than the range of r2 values used in R1R2. The same notation is applied to R1–R2+ and R1+R2+. The fifth coding scheme, labelled R0R0, was representative of the place/F0 strategy and was tested as a control. The label indicates that the pulse rates for this coding scheme were derived from F0.In the R0R0 scheme, the two electrode pairs of each synthetic vowel were both stimulated at the same pulse rate of 125 pps.

The five coding schemes were tested by both subjects in the order R1R2, R1+R2–, R1–R2+, R0R0 and R1+R2+. Each of these coding schemes were tested in turn according to the following procedure. The eleven synthetic vowels from each coding scheme were first balanced in loudness using the method described above. Testing was then performed over several consecutive sessions. Each session comprised six or seven blocks of single interval trials. A block consisted of four presentations of each synthetic vowel in random order, giving a total of 44 presentations within the block. A written list of the eleven vowels to be identified was given to the implantee as a reminder of the labelling used for the stimuli. For each presentation of a synthetic vowel or stimulus, feedback was given on whether the response was correct or wrong, and if it was wrong, what the correct answer should have been. After every three of four blocks, the implantee was given a five minute break.

The first few sessions were treated as training sessions. At least ten blocks of tests were then carried out and the results stored in the form of confusion matrices. Testing was terminated when the last three blocks showed overall percentage correct scores within 5% of one another. The results from the last ten blocks were cumulated for subsequent analysis.

B. Perceptual independence study

Two fixed electrode pairs, A and B, were selected for each implantee tested. Electrode pairs with similar dynamic ranges were used to facilitate the balancing of loudness between them. Each of the two electrode pairs were to be stimulated at one of four different pulse rates logarithmically spaced between 100 to 200 pps. A total of sixteen fixed duration (200 ms) stimuli, comprising all possible combinations of the four pulse rates were presented respectively on the two electrode pairs. The sixteen stimuli were first balanced in loudness and then presented to the implantee in single intervals in an absolute identification task.

The sixteen stimuli were labelled 1 to 16. The implantee was first familiarized with the stimuli over a number of sessions. Testing was then carried out over a number of subsequent sessions, each session comprising three or four blocks of testing. A block consisted of presenting each of the sixteen stimuli a total of three times in random order, giving a total of 48 presentations within a block. For each presentation, the implantee was required to identify the stimulus using one of the labels 1 to 16. Feedback was given as to whether the response was correct or wrong, and if it was wrong, what the correct response should have been. At least ten blocks of tests were carried out. Testing was terminated when the final three blocks had percentage correct scores within 5% of one another. The results were stored in the form of confusion matrices and the results from the last ten blocks cumulated and used in the data analysis.

IV. RESULTS

A. Synthetic vowel study

The results indicated that F1 information encoded into r1 was better perceived than F2 information encoded into r2. For GW, F1 information was significantly better transmitted when encoded using R1R2. There was no corresponding increase observed in the transmission of F2 information. For MO, however, F2 information was significantly more poorly transmitted when encoded using R1R2 while no significant changes in F1 transmission were seen with any of the four coding schemes.

It should be noted at this point that MO performed very well in the identification task and as a result, the analysis figures are all very high (>80% in all cases). Such high levels of performance are susceptible to ceiling effects which make it difficult to observe significant differences in performance. GW's results, on the other hand, are much more varied and enables differences in performances to be more easily seen.

B. Perceptual independence study

The results from this study were analysed for perceptual independence and also subjected to multidimensional scaling.

Analysis for perceptual independence indicated that for the pulse rates tested (which were below 200 pps), the two pulse rates presented respectively to two different electrode pairs are perceptually partially independent. The results also suggest that the pulse rates presented on the apical electrode pair were better perceived than their basal counterparts.

Multidimensional scaling (MDS) of the data showed that the apical pulse rate was transmitted more independently of the basal pulse rate than vice-versa. In either the one- or two-dimensional MDS solutions, the apical pulse rate appeared to be well correlated with one of the MDS dimensions while the basal pulse rate was well correlated with the MDS dimensions only in the presence of a higher apical pulse rate.

The results from perceptual independence analysis and multidimensional scaling agree well with one another. In both cases, the apical component of the stimuli is seen to be better transmitted than the basal component. The overall conclusion from these two forms of analysis is that the two pulse rates are perceptually partially independent.

V. DISCUSSION

The synthetic vowel study was carried out primarily to investigate how information encoded using the place/rate strategy was transmitted, particularly when compared with the place/F0 strategy. The improvements in the percentage correct scores have indicated that the place/rate strategy can transmit the encoded vowel information better than the place/F0 strategy. The differences in scores between the two implantees for a given coding scheme suggest that the information encoded into the pulse rates are perceived differently by different implantees. However, the fact that a significant improvement in the percentage correct score was obtained suggests that the place/rate strategy may be used to produce better identification of the synthetic vowels compared to the place/F0 strategy when optimally adjusted for a particular implantee. Further analysis revealed that F1 information encoded into r1 was better transmitted than F2 information encoded into r2. Thus, it may be possible to improve the perception of F1 information by encoding it into r1 as well as e1, while there may be little advantage, or even a disadvantage, in doing the same with F2 information. Note also that in the four place/rate coding schemes investigated in this study, the acoustic formant information was well transmitted when encoded into pulse rates within the range 80 to 250 pps but not in a higher range of pulse rates.

With the perceptual independence study, the results from both forms of analysis indicated that the apical and basal pulse rates were perceptually partially independent. Perception of the apical pulse rate was influenced by the basal pulse rate to a lesser degree than the reverse case. The results also suggested that the apical pulse rate was better perceived than its basal counterpart. In both the one-dimensional and two-dimensional MDS solutions, there was poor correspondence between the basal pulse rate and the MDS dimensions at the lower apical pulse rates of 100 and 126 pps, but at higher apical pulse rates (i.e. 158 and 200 pps), the basal pulse rate corresponded well with the MDS dimensions.

The results from both studies suggest that the information encoded into the apical and basal pulse rates of the stimuli could both be perceived, with the apical pulse rate being better perceived than its basal counterpart. The MDS results from the perceptual independence study indicate that at lower apical pulse rates (i.e. 100 and 126 pps), there was much confusion in trying to identify the stimuli with different basal pulse rates. This confusion in identifying the basal pulse rate could be caused by the lower apical pulse rate being perceptually more dominant, to the extent that it actually affected the identification of the basal pulse rate. At higher apical pulse rates (i.e. 158 and 200 pps), the MDS results indicate that both pulse rates were well identified. Thus, it appears that the apical pulse rate could be perceptually dominant over the basal pulse rate at lower pulse rates, but not so much at higher pulse rates. These results suggest that both the apical and basal pulse rates can be used to convey encoded speech information when the apical pulse rate is greater than say 150 pps. If the apical pulse rate is low (<150 pps) it will be perceptaully dominant.