COMPARISON OF TWO COCHLEAR IMPLANT SPEECH-PROCESSING STRATEGIES

GRAEME M. CLARK, PHD, FRACS
YIT CHOW TONG, PHD
RICHARD C. DOWELL, DIP AUD

EAST MELBOURNE, AUSTRALIA

Speech processors extracting either the fundamental frequency (F0) alone, or the fundamental frequency combined with second formant information (F0-F2), have been evaluated on a totally deaf patient using a multiple-channel cochlear implant. A closed set test using 16 spondees and a modified rhyme test showed that for electrical stimulation alone the F0-F2 speech processor was significantly better than the F0 processor. The open set tests using phonetically balanced words and Central Institute for the Deaf everyday sentences showed that for electrical stimulation alone and electrical stimulation combined with lipreading, the results with the F0-F2 speech processor were all significantly better than with the F0 processor. Information transmission for consonant speech features was also better when using the F0-F2 processor.

INTRODUCTION

Electrical stimulation of residual auditory nerve fibers can help postlingually deaf patients communicate, and it also offers hope of assisting the prelingually deaf. For further progress in the field it is important to evaluate the various speech-processing and stimulus systems available, and to design and develop others.

For this reason it was considered of value to compare the effects of two speech-processing strategies using a wearable speech processor in the home environment. During this time the processor extracted and presented him with both F0 and F2 frequencies. In March 1982 a study was undertaken to compare the speech information perceived when extracting either the F0 alone, or both the F0 and F2. To be sure the patient was receiving maximal information in the F0 mode, equal exposure to an F0 and F2 frequen-

had been constructed by ranking the electrodes from dullest to sharpest, and assigning frequency subbands to these electrodes in an order from lowest to highest. Similarly, the current level was determined on the basis of an A2-to-current-level map. The speech parameters were determined every 10 ms and only one electrode was activated within a 10-ms time frame.

Since May 1980 the patient has been continuously using a wearable speech processor in his home environment. During this time the processor extracted and presented him with both F0 and F2 frequencies. In March 1982 a study was undertaken to compare the speech information perceived when extracting either the F0 alone, or both the F0 and F2. To be sure the patient was receiving maximal information in the F0 mode, equal exposure to an F0 and F2 processor would have been desirable. It was difficult to get the patient to accept this training routine as he had experienced better speech perception with a F0-F2 processor right from the outset when informal comparisons were made with an F0 processor. It was therefore necessary to assume that in the combined F0-F2 mode, the amount of information obtained through the F0 channel was optimal. An audiological evaluation of the two processors using closed sets of 16-spondee words, open sets of phonetically balanced words, and open sets of Central Institute for the Deaf (CID) sentences showed better results for the F0-F2 than the F0 processor.

The present study has been undertaken to extend the audiological assessment of these two speech-processing strategies, and to make a statistical comparison of the results. Since nonsense syllables have

From the Department of Otolaryngology, University of Melbourne, and The Royal Victorian Eye and Ear Hospital, East Melbourne, Victoria, Australia. This study was funded by the National Health and Medical Research Council of Australia, and the Department of Science and Technology of the Australian government.

REPRINTS — Graeme M. Clark, PhD, FRACS, Department of Otolaryngology, University of Melbourne, The Royal Victorian Eye and Ear Hospital, East Melbourne, Vic 3002, Australia.
been of value in providing information about the speech features perceived with electrical stimulation, they have also been used in the present study to compare the two speech-processing strategies.

METHODS

Over a period of three days in March 1982, the patient was tested with both the FO and the FO-F2 speech processors for the test conditions electrical stimulation plus lipreading (EL) and electrical stimulation alone (EA). During this period lipreading alone (LA) scores were obtained for the FO processor test series, but, due to a shortage in list material, the LA scores for the FO-F2 processor were obtained from a test series performed some months previously. The comparison was made using closed sets of 16-spondee (16-SP) words, closed sets of modified rhyme test (MRT) words, open sets of Arthur Boothroyd (AB) words, the word intelligibility by picture identification (WIPI) lists as open sets of words, and open sets of CID everyday sentences. It has been shown that for the AB words and the CID sentences certain lists should not be used to ensure list equivalence. These requirements have been met in this study so that any differences in the results are due to variations in test difficulty. Furthermore, it has been shown that the four lists of WIPI words are highly equivalent. The word and sentence tests were all adapted for Australian conditions and presented by a male audiologist with an Australian accent and experience in monitoring the intension of words. The patient was placed in a sound-attenuated room with the tester outside. Lipreading evaluations were performed using a video camera to display the tester's face on a monitor in the room. The testing of the FO speech processor was performed on March 15 and the FO-F2 processor on March 19 and 22. During this period there was no change in the thresholds at the different electrodes. The subject had previous experience with the tests in May 1980, May 1981, and October 1981.

The order of presentation of the test conditions was EL, EA, and LA. This means that any practice effect tended to reduce the expected difference between the scores for LA and EL. It should also be noted that the better results obtained for the FO-F2 processor in the EL condition could lead to a greater learning effect and a carryover to a better performance for the EA condition. This could, in turn, result in improved scores for EA with the FO-F2 compared to the FO speech processor.

With the 16-SP word test each word was presented randomly five times until a total of 80 were given. For the MRT, 50 ensembles of six rhyming words were used for each evaluation, and the patient was asked to identify the test word selected from each ensemble. He was not informed about the correctness of his response. A computer program randomized the order of the 50 ensembles of rhyming words; the order of the six words within each ensemble was printed on the test form as was the order of the test word from each ensemble. In this way the material could be used repeatedly, without learning or a response bias affecting the results.

There were 15 AB word lists. Each list contained ten words and 30 phonemes, and the phoneme composition of the lists was standardized. Two practice words were presented before each list. For each condition two lists of ten words (a total of 20 words and 60 phonemes) were presented, except for lipreading alone, when one list was given for each speech processor. The WIPI words were randomly assigned to four test lists, and in each list there were 25 monosyllabic words and 80 phonemes. With children these are usually used in a closed set, but for the speech-processing evaluation the words were presented as an open set and the patient was not given any clues about their identity. For each condition 25 words and 80 phonemes were presented. With the CID sentences there were ten lists; each list contained ten sentences and 50 score words. For each condition ten sentences containing 50 score words were presented.

The statistical significance of the differences in the results for the FO and FO-F2 speech processors was evaluated by modeling the scores as a binomial variable. An are sine transformation was first performed on the scores. The variance of a difference between transformed scores was then calculated, and the probability of an observed difference occurring by chance estimated using a Z-table. In this way critical differences were estimated for each of the pairs of scores obtained.

Finally, the identification of nonsense syllables was investigated using the 12 consonants /b/, /d/, /g/, /j/, /l/, /m/, /n/, /p/, /t/, /v/, /z/, /d/ and /k/ in a vowel-consonant-vowel context with the vowel /a/. This was undertaken for the FO speech processor so that the results could be compared with those obtained previously for the FO-F2 processor. The test was performed in a similar manner and the closed sets of randomized syllables were presented for the conditions EL, EA, and LA. There were ten presentations of each consonant for each condition, or a total of 120 per condition. For the FO-F2 processor evaluation there were 240 presentations per condition. The information transmission for the speech features voicing, nasality, affrication, duration, and place, was also calculated.

Percentage information transmission was defined by Miller and Nicely as the ratio of the number of bits of information detected by the patient to the information available in the stimulus, expressed as a percentage, or \( \frac{\text{Percent}}{\text{Transmitted}} = \frac{T(x;y)}{H(x)} \times 100 \), where \( T(x;y) \) is the information transmission from \( x \) (stimulus) to \( y \) (responses) in bits per stimulus, and \( H(x) \) is the information available in the stimuli in bits per stimulus.

For example, if, in the present study, there were an equal number of voiced and unvoiced consonants, one bit of information would be present in the stimulus presented. One bit of information being transmitted is a choice between two alternatives. If all the voiced consonants were correctly distinguished from the unvoiced consonants, one bit of information would have been transmitted, and the information transmission would be 100%. If the patient was to guess the result on the basis of prior knowledge, he/she would at best be right 50% of the time, and in this case 0% information would have been transmitted. Consequently, information transmission is a more meaningful way of assessing a patient's response than recording a percentage-correct score. In the present study information transmission was obtained from the confusion matrices using each cell to calculate the bits of information transmitted from the stimulus to the response, and the bits of information available in the stimulus. The results were obtained from both 120 observations/matrix and 240 observations/matrix, which are smaller than previous studies. This would tend to be an underestimate of the information transmission statistics. As twice as many observations were made for the FO-F2 compared to the FO processor, the FO results could also be an overestimate. The application of the information-transmission analysis to our studies is described in more detail by Dowell et al.

RESULTS

The results of the 16-SP word, MRT word, AB word, WIPI word, and CID sentence tests for the FO and FO-F2 speech processors are shown in Table 1. It can be seen that, for the closed set 16-SP words and EA condition, the FO processor gave a 28% score. This increased to a score of 100% for the FO-F2 processor, and was statistically significant at the 0.05% level. With closed sets of MRT words the scores for EA were 24% for the FO processor and 50% for the FO-F2 processor; this difference was significant. For the EL condition the improvement with the FO-F2 processor was not sig-
significant. With the open sets of AB words, scored on the basis of words identified, the patient obtained a significant improvement when using the FO-F2 processor compared to the FO processor. For EA the score increased from 0% to 25%, and for EL from 20% to 50%. When the test was scored phonemically, significant improvements also occurred. The EA result went from 10% to 42%, and the EL result from 53% to 76%. A highly significant improvement for FO-F2 versus FO processing was also seen with the open set of WIPI words scored both on the words and phonemes identified. For the EA condition the word score increased from 0% to 28%, and the phoneme score from 8% to 50%. Finally, the results for the CID everyday sentences showed significantly improved performances for the FO-F2 versus the FO processor for both the EA and EL conditions. The score increased from 0% to 22% for EA, and from 60% to 98% for EL.

Table 2 shows the information transmission obtained with both the FO and FO-F2 processors for the speech features: voicing, nasality, affrication, duration, and place, and the consonants /b/, /p/, /m/, /v/, /f/, /d/, /t/, /n/, /z/, /s/, /g/ and /k/. With the EA condition the percentage information transmission was essentially the same for both processors for the voicing distinction (26% FO; 25% FO-F2). On the other hand, for all other speech features, better scores were obtained for the FO-F2 processor compared to the FO processor. The overall transmission also increased from 35% for the FO processor to 42% for the FO-F2 processor. The speech feature scores were all higher for the FO-F2 processor compared to the FO processor for the EL condition. Overall transmission also increased from 63% for the FO processor to 75% for the FO-F2 processor.

**DISCUSSION**

The results of the evaluation of the two speech-processing strategies showed that for EA the FO-F2 processor was always significantly better than the FO processor. This finding is consistent with our previous psychophysical results, where spectral information could be perceived by our multiple-channel cochlear implant patients on the basis of the site of electrode stimulation, and where site and rate of stimulation are perceived as two separate percepts. Consequently, the FO-F2 speech-processing strategy, where voicing is represented as rate and the second formant as site of stimulation, conveys more information than the FO processor, where only voicing is transmitted.

This was confirmed by our consonant-confusion study where the results showed that voicing was transmitted equally well by both the FO and FO-F2 speech processors, but nasality, affrication, duration and place were all better transmitted by the FO-F2 processor. Fundamental frequency alone is not the only cue for voicing. Voice onset time and first formant (F1) transition are also important. Further improvements in speech processing should be possible if these cues are also maximized. The better performance for nasality with the FO-F2 processor was probably due to the fact that a rising or falling F2 is important for the perception of these phonemes and is conveyed by the FO-F2 processor.
cessor. With affrication, concentrations of frequency energy in the F2 region are the important cues, and this would explain the better performance of the F0-F2 processor. Furthermore, for fricatives, the duration of the burst of noise, the closure interval, and the duration of the associated vowels also help to distinguish them from plosives and from each other. The improved performance with the F0-F2 processor for the duration feature is interesting. Both processors should convey timing information equally well. The important cue, however, is mostly timing in relation to the F2 and this would, therefore, explain the better performance with the F0-F2 processor. The most significant cue for place is the F2 transition and, since this was transmitted by the F0-F2 processor, this would explain the better results. Finally, as F1 information can also be useful in the perception of consonants and vowels, it is to be expected that improvements will result if speech processors transmit F1 as well as F0 and F2 information. Psychophysical studies on our first patient have, in fact, shown that this is possible with multichannel stimulation.

When electrical stimulation was combined with lipreading, the scores obtained with the F0-F2 processor were significantly better for all tests except the MRT word lists, where the improvement from 76% to 84% was not significant. In this case, the relatively high score obtained with F0 suggests that the test is not sensitive. The significantly better results of the F0-F2 processor in the combined EL condition for all the AB word, WIPI word, and CID sentence tests are a measure of differences between the processors for normal discourse.

The better performance of the F0-F2 compared to the F0 speech processor when used in combination with lipreading was also seen when consonant speech-feature information transmission was measured. As shown in Table 2, information transmission was better for all speech features using the F0-F2 processor. Improvements obtained for nasality, affrication, and duration may not be as great as those recorded, as the F0-F2 processor also showed better scores for these features for LA. The scores for LA in the F0-F2 evaluations were obtained some months before those with the F0 processor, and minor differences in the lighting conditions and the small numbers of contrasting speech features in the test could account for the discrepancies. There were, however, a greater number of contrasting features in the test for voicing and place of articulation, and the results are therefore more reliable. They indicate that for place of articulation, the F2 information supplemented the lipreading cues and improved speech comprehension. The results obtained with voicing are more difficult to explain. They are, however, consistent with a previous study using the F0-F2 processor where the voicing results for EL were also better than EA. This is probably due to the design of the test. The test also depends on the recognition of other cues. For example, in detecting voicing in the above set of consonants, nasality cues can also be used.

REFERENCES


Minerva Access is the Institutional Repository of The University of Melbourne

Author/s:
Clark, Graeme M.; Tong, Yit Chow; Dowell, Richard C.

Title:
Comparison of two cochlear implant speech-processing strategies

Date:
1984

Citation:

Persistent Link:
http://hdl.handle.net/11343/27200