AN EXTENSION OF THE MULTIPEAK SPEECH PROCESSING STRATEGY FOR THE MSP/MINI 22 COCHLEAR IMPLANT SYSTEM.

P.A. Jones, H.J. McDermott, P.M. Seligman*, J.B. Millar+.
Department of Otolaryngology, University of Melbourne.
*** Cochlear Pty. Ltd. Lane Cove, Sydney.**
+ Research School of Physical Sciences, Australian National University.

ABSTRACT - The speech perception of three post-linguistically deaf adults using the Nucleus MSP/Mini System 22 cochlear implant system programmed with a new speech processing strategy, MPEAK+A0, was evaluated. The MPEAK+A0 strategy retains all the information of the standard Multipeak speech processing strategy and additionally presents acoustic components below 400Hz to the most-apical electrode. This extra spectral information may help implantees understand speech, particularly in noise. Since the estimated fundamental frequency is presented as the rate of stimulation at a fixed intracochlear site and is thereby potentially perceived more easily, and the amplitude of the stimulation on the apical electrode, associated with the voice fundamental, is directly determined from the estimated energy in the relevant spectral region, these coding factors may provide a better representation of the prosodic information in speech and a more complete auditory feedback signal. The comparison between Multipeak and MPEAK+A0 included tests of vowel, consonant and CNC word recognition. Speech materials were presented with both a male and female speaker. Sentence material, presented with background masking noise (four-speaker babble), was also used. The results showed that the new strategy significantly improved the ability of these MSP users to recognise words in open-set sentences in noisy conditions.

INTRODUCTION

A cochlear implant is a prosthetic device designed to restore some degree of auditory-alone speech recognition to the sensorineurally deaf(Clark 1990). The system generally consists of a wearable speech processor, an implanted receiver-stimulator and electrode array which together can provide an electrical representation of the speech signal to the residual nerve fibres of the peripheral auditory system. The current commercial Melbourne/Nucleus MINI 22 cochlear implant system is composed of an array of 22 electrodes which is inserted into the scala tympani of the cochlea and the MSP speech processor. The MSP is usually programmed with the Multipeak speech processing strategy. Multipeak extracts the the fundamental frequency of the voice, F0; determines whether the signal represents a voiced segment of speech; estimates the first and second formant frequencies (F1,F2), the amplitudes of these formants (A1,A2) and the amplitudes in three higher frequency regions of the signal (F3: 2000-2800Hz, F4: 2800-4000Hz, F5:4000-6000Hz). These features are presented as electrical stimulation by: mapping the formant frequencies to electrode pairs along the array according to the tonotopic organisation of the cochlea; assigning the outputs of the high frequency filters to fixed basal electrodes; presenting acoustic amplitudes as the current level/pulse width of biphasic current pulses. Voicing is signalled by the presence of periodic stimulation at the F0 rate and unvoiced segments by a random stimulation rate of approximately 250Hz. The Multipeak scheme presents four pulses in an F0 period. If the signal is voiced the stimulation sequence is F4, F3, F2, and F1. For an unvoiced segment the sequence is F5, F4/F3, F2, F1 where the F4/F3 notation indicates that F4 or F3 is selected based on the estimated frequency of the second formant. If F2 is high, F4 is presented; F3 otherwise. Previous speech processing strategies have included F0/F2 and F0/F1/F2 whose structures can be deduced from the description of Multipeak above. Together with Multipeak, the trend in the results achieved with these speech processing schemes shows that the addition of more spectral information provides implantees with better speech perception(Blamey 1987, Skinner 1991). In this context, an extension of the Multipeak strategy, MPEAK+A0, was proposed which involves the addition of an extra current pulse per F0 period which presents an estimate of the energy in the frequency region below 400Hz to the most-apical usable electrode in the implantee's array.

METHOD

Strategy Implementation

The new feature added to Multipeak to create MPEAK+A0, A0, is an already existing feature within the speech processor which is used in the determination of the voicing status of the signal. It is a measure of the energy in the signal below 400Hz whose temporal character is limited by the smoothing applied by a 35 Hz low pass filter. Due to noise internal to the speech processor in this frequency region, the first 8 non-zero values of the amplitude estimate of A0 are ignored. The temporal sequence for the MPEAK+A0 strategy is - voiced: F4 A0 F3 F2 F1 at the F0 rate; unvoiced: F5 A0 F4/F3 F2 F1 at a random rate. In implementation, the most-apical electrode is selected by using the electrode associated with the lowest F1 frequency. Therefore, the A0 pulse has been placed so that it is temporally distant from the F1 pulse to minimise any effect due to double stimulation in the apical region when F1 is low.

Subjects

Three post-linguistically deaf adult subjects were used in this study. When fitted with the MPEAK+A0 strategy using their existing threshold (T) and maximum comfortable (C) levels for electrical stimulation, speech was louder than it was with Multipeak. The C levels were modified globally by an amount such that the MPEAK+A0 strategy was of equal loudness to the standard strategy. The three subjects required similar amounts of C level modification ranging between 10-15% of the dynamic range. Subjects reported that the MPEAK+A0 strategy sounded deeper than Multipeak and found it comfortable to use.

Protocol

The auditory-alone speech perception of the subjects when using MPEAK+A0 and Multipeak was compared over 6 evaluation sessions for each subject in an alternating design. At each session the subject had worn the evaluation strategy for at least one week immediately beforehand. Within each session, two lists of 4 repetitions of 11 steady state vowels in a /h V d/ context, two lists of 4 repetitions of 12 consonants in an /a C a/ context and two lists of 50 CNC words were presented. One list of each pair was presented with a female speaker, the other with a male speaker. Both speakers were familiar to the subjects. During the study, the balanced design for the vowel and consonant tests was slightly modified when a small amount of extra data was collected for the two better performing subjects. Two CUNY sentence lists of 12 sentences each, spoken by the female speaker were administered, one at each of two signal-to-noise ratios (SNR). The higher SNR was always tested first. Because of the different speech perception ability of the subjects, the SNRs for two of the subjects were chosen to be 15 and 10 dB. The other subject performed at SNRs of 10 and 5 dB. All tests were conducted in a sound treated room, with the speech material presented from audio tape at a signal level of 65 dBA. The results were subjected to analysis of variance for repeated measures within subjects to determine whether the subjects' performance was significantly different for the two strategies or the two speakers in the case of the vowel, consonant and CNC word tests, and the strategy and signal-to-noise ratio for the CUNY sentence test in noise. Where significant factors were observed, analysis of variance was applied to individual subject scores and t-tests were used with appropriately pooled data.

RESULTS

Vowels

A three-way analysis of variance showed that there was no significant effect for the strategy or speaker factors. The subject factor was highly significant for this and all other tests, indicating that the subjects had different speech perception abilities with their implants. A significant interaction term between the strategy and subject factors was evident [F(2,25) = 8.13, p=0.002]. Individual t-tests with the speaker data pooled showed that this interaction term was due to contrary trends in the performance of two subjects. For subject 1, a significant advantage was demonstrated for MPEAK+A0 (t=2.29; df=9; p=0.045) while a significant disadvantage was recorded for subject 2 (t=-2.99; df=5; p=.0031). The third subject showed no effect for strategy but performed significantly better with the male speaker (t=3.99; df=9; p=0.0032). A summary of the information transmission analysis(Miller 1955) for the strategy condition is shown in figure 1.

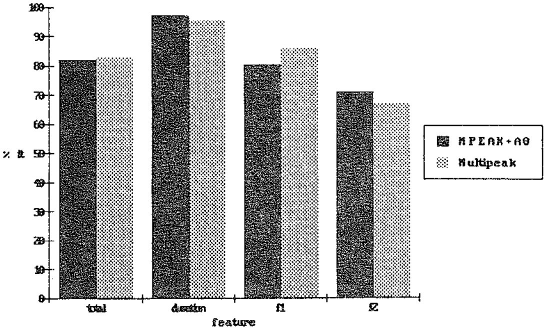

Figure 1. Information transmission for vowels with pooled speaker data.

Consonants

For the consonant test, neither the strategy or the speaker factors were statistically significant although higher mean scores were observed for the MPEAK+A0 strategy and the male speaker. This was reflected in the analysis of variance of the results from the information transmission analysis, and illustrated for the strategy conditions in figure 2. The total information [F(1,7)=21.66, p<0.0005] and place [F(1,7)=19.77, p=0.003] feature scores were significantly higher with the MPEAK+A0 strategy compared to Multipeak. Similarly, total information [F(1,7)=37.13, p<0.001], place [F(1,7)=7.99, p=0.026], manner [F(1,7)=10.38, p=0.015] and the amplitude envelope [F(1,7)=26.64, p=0.001] were significantly higher with the male speaker.

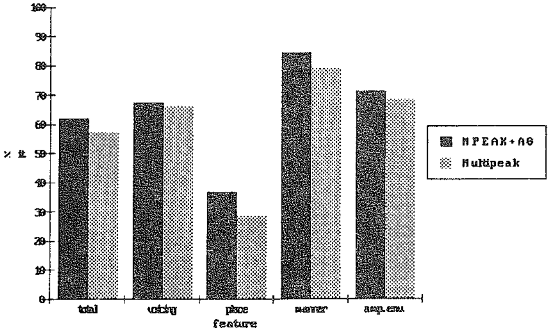

Figure 2. Information transmission for consonants with pooled speaker data.

CNC words

The analysis of variance for the CNC word test scores showed that there was no significant effect for strategy. This was also the case for all the information transmission features. For the speaker factor, the performance on the identification of the vowel portion of the CNC words was significantly better [F(1,24)=9.17, p=0.006] with the male speaker. This was reflected in the information transmission analysis where the following features were significant: total vowel information [F(1,7)=6.02, p=0.044], f1 [F1(1,7)=27.23, p=0.0005], f1r/f [F(1,7)=8.26, p=0.024] and f2r/f[F(1,7)=17.57, p=0.004]. The f1r/f and f2r/f features describe the ability of the subject to identify whether the formants are rising(r), falling(f) or steady. It was also found that the initial consonant voicing feature score was significantly higher for the female speaker [F(1,7)=5.56, p=0.049].

CUNY sentences

The results of the CUNY sentences test is shown in figure 3. For the purpose of analysis, the higher SNR condition for each subject was grouped and noted as snr1. The lower SNR for each subject was grouped into the snr2 condition. The analysis of variance showed significant differences for the strategy [F(1,24)=24.21, p<0.0005], SNR [F(1,24)=90.21, p<0.0005] and subject [F(2,24)=8.85, p= 0.001] factors. The analysis of variance for individual subjects consistently showed a significant effect for strategy: subject1 [F(1,8)=9.93, p=0.014], subject2 [F(1,8)=6.41, p=0.035], subject3 [F(1,8)=8.79, p=0.018] in favor of MPEAK+A0 over Multipeak. The SNR factor was also highly significant in each case. Examining the two SNRs separately, the strategy factor was significant in both conditions: snr1 [F(1,12)=10.27, p=0.008], snr2 [F(1,12)=16.29, p=0.002].

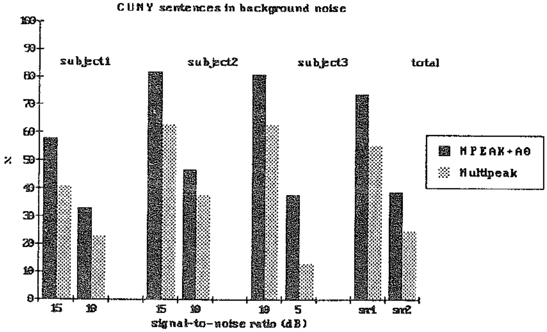

Figure 3. CUNY sentence results for 3 subjects at 2 SNRs.

DISCUSSION

The result for the CUNY sentence test was encouraging in at least two respects: this open-set sentence recognition task in background noise best measures the utility of a strategy to the implantees in their real-world communications; and the statistical significance was achieved with a small number of sentence lists despite the alternating strategy design. This design can have a smoothing effect on strategy evaluations because of the subjects' relative lack of experience with the new strategy and disruption caused by the strategy being changed frequently. The explanation for this improved performance with sentences in background noise is not obvious from the vowel, consonant and CNC word results. As mentioned in the introduction, previous improvements in speech perception with Melbourne/Nucleus speech processing schemes have been achieved with the addition of more spectral information. This developmental pathway is however, limited by electrical masking effects(Tong 1986). For example, a potential disadvantage with this strategy was that it may have affected vowel recognition performance, especially for vowels with low first formant frequencies. This masking effect was observed in the vowel test(see figure 1) and the vowel portion of the CNC words but not to a significant extent. Some improvement was seen in the consonant features in the isolated consonant test, notably in the place feature. This consonant advantage was not seen in the CNC word test however. This indicates that the effect of A0 is marginal at the phonetic level. It is possible that the benefit of A0 was counteracted by the reduction in the dynamic range caused by the global reduction in C levels applied in the fitting of MPEAK+A0. To an extent, this reduction in dynamic range and the masking of the F1 component by A0, can be addressed by excluding the F1 component from the apical electrode and applying the C level modification to the apical electrode only. Another option to reduce stimulation in the apical region is to present the A0 component only when the signal is voiced. These factors are being investigated. Along with the presentation of the additional low frequency information in MPEAK+A0, its coding onto a fixed apical electrode has the effect of presenting the F0 rate with its associated energy at a fixed, tonotopically appropriate stimulation site.

Since it can be assumed that the A0 feature will be strongly correlated with the speech envelope itself, this coding has the potential to provide a better presentation of the prosodic information in the speech signal through both the time-intensity amplitude and F0 contour cues. This may contribute to the observation that the benefit for MPEAK+A0 is more evident in sentence materials than in isolated vowel and consonant or word contexts. It was shown that the vowel portion of the CNC words was better perceived with the male voice. This is in accord with the subjects' statements about the better intelligibility of the male speaker. In practical use, two of the three subjects reported that the MPEAK+A0 strategy was better than Multipeak for female voices because it lessened the 'shrillness' of some high pitched female voices. One the other hand, one subject found that very deep voices became 'mumbly' because MPEAK+A0 was too deep in this situation. Finally, anecdotal reports suggest that with MPEAK+A0 more environmental sounds are audible, music perception may be better and it may provide a more appropriate auditory feedback signal for voice monitoring.

CONCLUSION

The evaluation of the MPEAK+A0 strategy showed that the strategy produced improved speech recognition for open-set sentences in background noise, a condition which most closely mirrors the acoustic environment of the implantee in everyday life. The analysis of the results at a phonetic level showed only a small advantage for the MPEAK+A0 strategy in isolated consonants and this effect was not evident in words. Although the relation between performance at the phonetic and sentence levels is complex, these results suggest that the improvement in sentence level performance is due in part to the addition of extra spectral information and also due to the better transmission of prosodic information inherent in the new coding of the voice source information onto the most-apical electrode.

ACKNOWLEDGEMENTS

This work has been supported by a DEET APRA (Industry) scholarship and Cochlear Pty. Ltd, the Human Communication Research Centre and the Australian Bionic Ear and Hearing Research Institute. The authors wish to acknowledge the assistance provided by Lesely Whitford and Professor Graeme Clark.

REFERENCES

Clark, G.M. Tong, Y.C. Patrick, J.F.(Eds) (1990), *Cochlear Prostheses*, Churchill Livingstone.

Blamey, P.J. Dowell, R.C. Brown, A.M. Clark, G.M. and Seligman, P.M. (1987), *Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors.* J. Acoust. Soc. Am. 82(1): 48-57.

Lim, H.H. Tong, Y.C. and Clark, G.M. (1989), *Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea.* J. Acoust. Soc. Am. 86(3): 971-980.

Miller, G.A. Nicely, P.E. (1955), *An analysis of perceptual confusions among some English consonants.* J. Acoust. Soc. Am. 27: 338-352.

Skinner, M.W. Holden, L.K. Holden, T.A. Dowell, R.C. Seligman, P.M. Brimacombe, J.A. Beiter A.L. (1991), *Performance of Postlinguistically Deaf Adults with the Wearable Speech Processor (WSPIII) and the Mini Speech Processor (MSP) of the Nucleus Multi-Electrode Cochlear Implant.* Ear and Hearing, Vol. 12, No. 1, 3-22.

Tong, Y.C. Clark, G.M. (1986), *Loudness summation, masking, and temporal interaction for sensations produced by electric stimulation of two sites in the human cochlea.* J. Acoust. Soc. Am. 79 (6): 1958-1966.

FORMANT-BASED PROCESSING FOR HEARING AIDS

P.J. Blamey[*], G.J. Dooley[*], P.M. Seligman[+], J.I. Alcantara[*], and E.S. Gerin[*]

[*] Human Communication Research Centre
University of Melbourne

[+] Cochlear Proprietary Limited

ABSTRACT - A body-worn hearing aid has been developed with the ability to estimate formant frequencies and amplitudes in real time. These parameters can be used to enhance the output signal by "sharpening" the formant peaks, by "mapping" the amplitudes of the formants onto the available dynamic range of hearing at each frequency, or by resynthesizing a speech signal that is suited to the listener's hearing characteristics. Initial evaluations have indicated small improvements in speech perception for three groups of subjects: users of a combined cochlear implant and speech processing hearing aid, normally hearing listeners in background noise, and a hearing aid user with a severe hearing loss.

SOME PROBLEMS WITH HEARING AIDS

Most conventional hearing aids amplify sounds with a fixed linear gain function that varies across frequencies to compensate for the hearing loss. Commonly, the maximum power output is also limited to avoid uncomfortably loud sounds. These aids make sounds louder, but do not overcome all of the problems that are associated with hearing-impairment. For example, recruitment is a narrowing of the dynamic range of hearing caused by threshold levels being raised more than discomfort levels for sound. This effect is a consequence of the loss of outer hair cells that produce the very sensitive and highly tuned cochlear response to pure tones at low levels (Patuzzi, 1990). When the vulnerable outer hair cells are lost through exposure to loud sound, disease, or other trauma, the sensitivity is lost, but responses to louder sounds are relatively unaffected. Fine frequency resolution and selectivity can also be reduced by loss of the outer hair cells (Evans, 1975). These distortions of the hearing sensations operate to reduce the intelligibility of speech even when it is amplified to a comfortable level in quiet conditions. Background noise is also a major problem for hearing aid users who are usually affected much more than normally hearing listeners. In the last two decades, it has been proposed that hearing aids should be designed to compensate for the effects of recruitment, poor frequency resolution, and background noise as well as providing amplification (e.g. Villchur, 1973).

Multiband compression has been a common method used to compensate for recruitment. The amplitude levels for fixed frequency bands are compressed nonlinearly to match the dynamic range of hearing (e.g. Lippmann, Braida & Durlach, 1981). At most, this method has led to modest improvements in speech intelligibility compared with linear gain. The limiting factors appear to be distortions in the processed signals and reduced loudness contrasts between different temporal and spectral components of the speech. Spectral enhancement is a procedure that increases the amplitude differences between peaks and valleys of a spectrum with the aim of compensating for reduced frequency selectivity (Simpson, Moore & Glasberg, 1990). A narrowband enhancement algorithm can increase the relative amplitude of the harmonic components of the voice. A wideband procedure increases the relative amplitude of the formants. These procedures have produced small improvements in speech intelligibility for hearing impaired listeners in laboratory studies. Directional microphones have been used in hearing aids for some time to reduce the effects of background noise. Processing to implement adaptive beamforming microphones also shows promise (e.g. Peterson, Durlach, Rabinowitz & Zurek, 1987). Adaptive filtering based on long-term noise spectra is less effective in improving intelligibility, but can decrease the annoying effects of some types of noise (e.g. Levitt, Neuman, Mills & Schwander, 1986).

In addition to the fundamental problems outlined above, there are practical difficulties in determining and fitting ideal linear gain functions to the characteristics of an individual user's hearing loss.