The Journal of Laryngology and Otology
November 1980. Vol. **94.** pp. 1241–1256.

Psychophysical and speech perception studies on two multiple channel cochlear implant patients

By Y. C. TONG, J. B., MILLAR, G. M. CLARK, L. F. MARTIN, P. A. BUSBY and J. F. PATRICK
(Melbourne)

Introduction

PRELIMINARY studies on our first patient with a multiple channel cochlear implant (Tong *et al.*, 1979) showed that he was able to associate the hearing sensations produced by electrical stimulation of residual auditory nerve fibres with different vowel (spectral) colours. The frequencies characterizing the vowel colours produced by stimulation at single scala tympani electrodes were, in general, closely related to the tonotopical organization of the cochlea. Furthermore, results from these studies indicated that the pitch produced by single-electrode stimulation increased with pulse rate, and that the loudness increased with current level.

Based on these results, a speech coding scheme for the multiple-channel cochlear implant was proposed (Tong *et al.*, 1979). It was suggested that the spectral emphases of the speech signal could be transformed to a pattern of activated electrodes, the fundamental frequency to pulse rate, and the amplitude envelope to current level.

This paper describes the results of a series of psychophysical and speech perception experiments conducted with both our first and second cochlear implant patients. The first objective of these studies was to make a quantitative assessment of the speech coding scheme by evaluating the performance of the patients listening through a laboratory speech processor which was a partial implementation of the proposed scheme. The second objective was to determine the generality of some of the observations described in the previous report by comparing the results obtained from the two patients in the present studies.

The clinical history and management of the two patients are presented in Section I of this report. Section II describes the procedure and results of a sharpness ranking experiment developed to obtain an indication of the variation of spectral colours amongst electrodes. These results form a major part of the specifications for the laboratory speech processor which is described in Section III. A quantitative assessment of the speech coding scheme is presented in Section IV, and vowel and consonant confusion test procedures and results are given and discussed. Section V

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

presents results observed in informal studies using sentences. Conclusions and future directions are presented in Section VI. A list of phonetic symbols of English phonemes used in this paper is given in the Appendix.

I. Clinical history and management

The first patient (MC1) selected for a multiple-channel cochlear implantation was a 47-year-old man who lost hearing completely, following a head injury 18 months prior to the cochlear implant operation. Audiological tests under head-phones established that there was no hearing up to the limits of the audiometer for either ear. There were also no thresholds for pure tones when tested free-field up to 116 dB, and for environmental noises up to 110 dB. Electrical stimulation of the promontory indicated that some auditory nerve fibres were intact as the patient could perceive a tonal sensation with stimulation at rates of 50, 100, 200, 400 and 800 pulses/second (pps).

At operation on the 1st August 1978, the electrode array was inserted for a distance of 25 mms. Subsequently, due to inadequate fixation the electrode array was found to have slipped out 10 mms, as a post-operative x-ray showed the most basal electrode lying in the region of the round window niche.

The second patient (MC2) was a 63-year-old man who lost his hearing in 1940, following a bomb blast during the war, but was still able to use a hearing aid. Over the years there was a progressive deterioration in the hearing in both ears, and in 1966 he was found to have a profound hearing loss in both ears, and could no longer get any help from a hearing aid. A pure-tone audiogram under headphones in 1978 showed no hearing in the left ear up to the limit of the audiometer, and in the right ear a threshold of 85 dB at 250 Hz, 105 dB at 500 Hz and a threshold beyond the limit of the audiometer at all other frequencies. There was a speech discrmination score of 0 per cent in both ears under headphones at an intensity of 120 dB, and also of 0 per cent with free-field testing 110 dB.

Environmental sounds could not be distinguished up to an intensity of 110 dB. Electrical stimulation of the promontory in the left ear indicated that the patient could distinguish pitch sensation at rates of 50, 100, 200, 400, 800 and 2,000 pps. He also reported that 800 pps produced a higher pitch than 50 pps, and that 2,000 pps produced a higher pitch than 800 p.p.s. A left cochlear implant operation was performed on the 17th July 1979, the electrode array inserted 19·5 mms, and well fixed in position. The electrode array was also designed so that a stimulating electrode lay in the region of the round window membrane.

For both patients, the electrodes are numbered from 0 to 9 according to the physical location of the electrodes. Electrode 0 corresponds to the most apical electrode, and electrode 9 the most basal one and the closest to the round window region. More detailed descriptions of the electrode

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

array, implanted receiver-stimulator and related external hardware can be found in previous reports (Clark *et al.*, 1977; Tong *et al.*, 1979).

II. Sharpness ranking studies

The central concept of the proposed speech coding scheme is that different spectral colours can be cued by the activation of individual electrodes. This implies that the relationship between the spectral colours and the electrodes forms a major part of the specifications of the speech processor. In the previous study, this relationship was determined by the first patient's subjective judgement on percepts produced by steady-state pulse trains. This approach resulted in the association of vowel categories with groups of electrodes. However, the detailed variation of spectral colour within a group of electrodes was not indicated.

In the present studies, the following indirect procedure was developed to obtain an indication of the variation of spectral colour amongst electrodes. The ranking of sharpness amongst single electrodes was investigated. The stimuli delivered to single-electrodes were 300 ms pulse trains with rise-decay times of 50 ms. The pulse rate was fixed at 100 pps and the current levels were obtained by balancing the loudness across electrodes. During each trial, the patients were presented with two 300 ms pulse trains separated in time by a 1s interval, and were asked to indicate whether the sensation produced by the second stimulus was duller or sharper than the first. The trials were presented in blocks. Within each block, a reference stimulus—a 100 pps pulse train at a reference electrode—was always presented as the first stimulus in each trial; this reference stimulus was paired at random with each of the pulse trains at the remaining electrodes. MC1 received nine blocks of trials where a different electrode served as the reference in each block, while MC2 received ten blocks of trials. Only nine electrodes were used in the study for MC1 because of the high current threshold of his most apical electrode (Tong *et al.*, 1979).

TABLE I

SHARPNESS RANKING AMONGST ELECTRODES

	Patient		
	MC1	MC2	
	9	0	Dull
	1	1	
	2	2, 9	
	3	3	
Electrode number	4	4	
	5	5	
	6	6	
	7	7	
	8	8	Sharp

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

The results of the sharpness ranking experiments are shown in Table I. For both patients, sharpness increased monotonically with electrode placement in the basal direction. The only exception is the most basal electrode (9). For patient MC1 the sensation at electrode 9 was the dullest, while for patient MC2 the sharpness associated with electrode 9 was approximately equivalent to that associated with electrode 2.

The relatively dull sensation associated with electrode 9 for both patients indicated that this response to stimulation in the region of the round window is reproducible and consistent across patients. As discussed in the previous report, this dull sensation could be a result of extra-cochlear current spread and excitation of auditory nerve fibres with lower characteristic frequencies.

It has been shown that sharpness increases monotonically with the frequency characterizing the spectral colour of an acoustic signal (Plomp, 1976, page 109). The ranking of sharpness amongst electrodes obtained in this study can therefore be considered as an indication of the ranking of frequencies characterizing the spectral colours associated with the single electrodes. Since the sharpness associated with a single electrode increased in the basal direction, one may further assume that the frequency characterizing the spectral colours for a single electrode increased in the same direction. This variation of frequency is consistent with the tonotopical organization of the cochlea.

For the laboratory speech processor to be described in the next section, the assignment of single-electrodes to subbands of second formant estimates was based on frequency ranking data derived from these sharpness ranking results. The subbands of low to high second formant estimates were respectively assigned to electrodes characterized by dull to sharp sensations.

III. Description of the laboratory speech processor

The laboratory speech processor accepts the speech waveform as an input and produces a pattern of electrical stimulus data as output. A transmitter unit, external to the speech processor, receives the stimulus data, formats these data and subsequently generates radio frequency transmitter signals which are inductively coupled to the implanted receiver-stimulator. The receiver-stimulator reconfigures the transmitter signals, and presents the electrical stimuli to the appropriate electrodes.

The laboratory speech processor consists of three basic sections: speech signal parameter extraction; encoding of speech parameters to electrical stimulus parameters; and digital configuration of electrical stimulus data.

Four speech signal parameters were estimated every 20 ms in the parameter extraction section. The fundamental (voicing) frequency (FO) and a low frequency energy measure (AO) were estimated by measuring the period and the average amplitude of the output waveform of a low-

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

pass filter. The second formant frequency (F2) and its amplitude (A2) were estimated by measuring the zero crossings and average amplitude of the output waveform of a bandpass filter. The measurements at the outputs of the analogue filters were performed by a digital computer. The encoding and digital configuration sections of the speech processor were also implemented on the same digital computer.

The speech parameter estimates (FO, AO, F2, A2) were transformed to electrical stimulus parameters every 20 ms in the encoding section. In the present speech processor, only one electrode was activated in any 20 ms time frame. For a given F2 estimate an electrode was selected according to a predetermined F2-to-electrode transformation map: the F2 frequency range was divided into nine subbands; each subband was assigned to a particular electrode. The subband of lowest frequency was assigned to the electrode with the dullest sensation, while the subband of highest frequency was assigned to the electrode with the sharpest sensation. Only nine electrodes were used in the speech processor because of the high current threshold of electrode 0 in MC1 and the almost equivalent sharpness between electrodes 2 and 9 in MC2. The current level for the single-electrode pulse train was determined from A2. A 20 ms speech segment was classified as voiced if AO exceeded a pre-selected threshold, and unvoiced otherwise. For unvoiced speech segments, a constant low pulse rate for electrical stimulation was used. This low rate was used as it produced a sensation described as 'rough' which is the closest response to that of noise received by normal hearing subjects. For voiced speech segments, the pulse rate was proportional to FO, and was higher than the pulse rate used for the unvoiced segments.

Given the electrode selection, current level and pulse rate for a 20 ms time frame, the digital configurations section of the speech processor formats the electrical stimulus data, and transfers these data to the external transmitter unit.

IV. Vowel and consonant confusion studies

The confusion studies reported in this Section were conducted with the following objectives in mind:

(i) to determine the kinds of vowel and consonant confusions experienced by the patients when receiving electrical stimulation through the laboratory speech processor; and

(ii) to suggest directions for future improvement of the speech processor design based on the results of these studies.

A. *Method and test materials*

The test materials and procedures employed in these studies were chosen to allow the patients to perform the identification tasks with a minimum amount of specific training and a minimum of discouragement.

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

It was argued that in this early stage of speech processor development the emphasis should be on the demonstration of the potential of the speech coding scheme, rather than striving for an exhaustive evaluation of the perceptual confusions among all phonemes.

Two members of the research staff, one female with a British accent and one male with an Australian accent, served as speakers. Both patients have had some exposure to the female speaker's speech processed through the laboratory speech processor, while the male speaker's speech represented an unfamiliar sample. The confusion tests were conducted in a quiet room with no sound treatments. A directional microphone with windshield was used throughout the four consecutive days of testing.

The test materials were presented live without lipreading being involved.

In order to familiarize the patient with the range of hearing sensations and responses which were to be included in a test, a typed list of the test words (six for vowel studies and ten for consonant studies) was given to the patient, and the test words were presented in the order as shown in the list. This procedure was repeated two or three times before each block of trials.

Six vowels were used in the vowel confusion studies. These included /i/ /I/ /a/ /ʌ/ /ɔ/ and /ɞ/. The six test words were of "hVd" structure—each word beginning with /h/ and ending with /d/ and differing only in the intermediate vowel. The patients heard ten presentations of each of the six "hVd" words in a block of trials. Two blocks of trials, one for each speaker, were conducted with each patient. The ordering of the sixty words in a block was randomized. The patients were instructed to identify each presentation as one of the six "hVd" words, or as a "Don't know" category.

The two back vowels, /ɔ/ and /a/ and the front vowel /i/ were chosen because these vowels can be closely approximated by single-formant synthetic acoustic signals, which were considered to be approximate acoustical equivalents for single electrode stimulations (Tong *et al.*, 1979). Delattre *et al.* (1952) showed that single formant approximations to back vowels were easier to identify than approximations to middle and front vowels, with the exception of the front vowel /i/. Furthermore, the frequencies of the single—formant approximations to these three vowels are very close to their respective second formant frequencies (F2), and F2 is the only parameter reflecting the spectral composition of the speech signal in the present speech processor. As a result of this limitation of the speech processor the other back vowels, with the frequencies of their single formant approximations significantly lower than their second formant frequencies, were omitted.

The other three vowels /ɞ/, /ʌ/ and /I/ were included in order to establish the extent to which the duration of the vowel nucleus was used as a cue for vowel identification by the patients listening through the

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

speech processor. It has been shown that the second formant frequencies of /ɓ/. /ʌ/, and /I/ differ only slightly from those of /ɔ/, /a/ and /i/ (Ainsworth, 1976; Petersen and Barney, 1952). With the present speech processor and electrode array, these small differences in second formant frequencies could not be reliably detected and encoded and the only consistent cue available for the distinction between these two groups of vowels was the duration of the vowel nucleus.

The consonants in the test materials were selected from all major classes of English consonants characterized by acoustic features which are considered to be sufficiently detected and encoded by the present speech processor. As a result, the following classes of consonants were omitted:

(i) voiced plosives with short voice onset time;

(ii) voiced fricatives involving the detection and encoding of a mixed mode of source excitation, and

(iii) the semi-vowel /l/ which may be easily confused with /r/ in the absence of the third formant (Ainsworth, 1976).

The ten consonants /k/, /p/, /s/, /ʃ/, /tʃ/, /f/, /w/, /j/, /r/, and /n/, were used. The ten test words were of CV structure with V set equal to /a/ as in f*a*ther. The patient heard ten presentations of each of the ten CV words in a block of trials. Four blocks of randomized trials, two for each speaker, were conducted with each patient. The patients were instructed to identify each presentation as one of the ten CV words, or as a "Don't Know" category.

TABLE II

PERCENTAGE CORRECT SCORES: VOWEL STUDIES

	Female speaker	Male speaker
Patient MC1	83·3%	81·66%
Patient MC2	83·3%	60%

Mean score = (83·3+83·3+81·66+60)/4
= 77%

TABLE III

PERCENTAGE CORRECT SCORES: CONSONANT STUDIES

	Female speaker	Male speaker
Patient MC1	44%	30%
Patient MC2	33%	33%

Mean score = (44+33+30+33)/4
= 35%

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

B. *Results—overall performance*

The percentage correct scores for vowel and consonant confusion studies for the two patients with the two speakers are tabulated in Tables II and III. The mean percentage correct scores across patients and speakers were 77 per cent for vowels and 35 per cent for consonants.

The consistently high scores for MC1 in the vowel studies with both the female (83·3 per cent) and the male (81·66 per cent) speakers suggested that MC1 could readily identify the 'features' which characterize vowels produced by a familiar (female) speaker or by an unfamiliar (male) speaker. In addition he could readily grasp the 'rules' governing the ways these features were combined in different vowels for both speaker conditions. In terms of the speech coding scheme, the major vowel features were the spectral colour and pitch associated with the single-electrode pulse train.

Both features were speaker-specific, but were not independent for a particular speaker. An example to illustrate the rules by which these features were combined is the following: a high pitch sensation combined with a high frequency spectral colour characterized a female /i/, while a lower pitch sensation combined with a lower frequency spectral colour represented a male /i/.

Patient MC2, however, did not exhibit the same sophistication in feature identification and rule learning for vowels. Although MC2's vowel score with the female speaker (83·3 per cent) was high, his score with the unfamiliar male speaker (60 per cent) was much lower than the score achieved by MC1. These results suggested that familiarity with the processed speech of a particular speaker does not necessarily lead to the learning of features and rules for another speaker. A more important factor is likely to be the amount of exposure to electrical stimuli each patient had received. In this regard, MC1 was more experienced in the discrimination and description of the hearing sensations produced by steady state pulse trains as a result of his involvement in psychophysical experiments for twelve months. Furthermore, MC1 had a longer period of exposure to the female speaker's (processed) speech in comparison to MC2 who received his implant only six weeks before these experiments.

For the consonant studies, MC1 scored better with the female speaker (44 per cent) than with the unfamiliar male speaker (30 per cent). It should be noted that the consonant scores across speakers for MC1 follow the same trend as the vowel scores for MC2. Again, these results emphasized the importance of the amount of exposure in determining the performance with an unfamiliar speaker, as discussed above. However, with further exposure to electrical stimuli with transitional characteristics (either through the speech processor or in psychophysical testing) an improvement of MC1's performance with consonants for both familiar and unfamiliar speakers might be possible.

The identical consonant scores for MC2 with both speakers suggested

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

that the test vowels were much easier to get familiar with than the test consonants. In this case, MC2 had not appeared to have familiarized himself with the consonants of the female speaker. As a result, the consonants of both speakers appeared to be equally difficult to identify.

Vowel confusions

Vowel confusion details for the two patients are given in Tables IV and V. The data for the two speakers were pooled. The choice of the six vowel stimuli and the present speech processor design suggested that duration and second formant frequency were the major cues available to the patients. The discussions that follow are directed towards the performance of the patients in relation to these two available cues.

It is noticeable that both patients were able to make use of the duration of the vowel nucleus as a cue for vowel identification. Confusions between the long vowels (/ɔ/, /a/, /i/,) and short vowels (/ɒ/, /ʌ/, /I/) were rare—

TABLE IV

VOWEL CONFUSIONS FOR SUBJECT MC1

		STIMULUS					
		/i/	/a/	/ɔ/	/I/	/ʌ/	/ɒ/
Response	/i/	18					
	/a/		19	2			
	/ɔ/		1	18			1
	/I/	2			20		
	/ʌ/					20	15
	/ɒ/						4

TABLE V

VOWEL CONFUSIONS FOR SUBJECT MC2

		STIMULUS					
		/i/	/a/	/ɔ/	/I/	/ʌ/	/ɒ/
Response	/i/	19					
	/a/		11	4			1
	/ɔ/		9	16			3
	/I/				19	4	
	/ʌ/				1	14	9
	/ɒ/	1				2	7

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

3 errors out of 200 presentations for MC1 and 5 errors for MC2. It has been found that the mean difference in duration between short and long vowels approaches 100 ms (House, 1961) which is encodable by the present speech processor.

The major confusions for both patients were between vowels with a relatively small difference in second formant frequency. Results in Tables IV and V showed that there were a noticeable number of confusions between /ɔ/ and /a/, and between /ʌ/ and /ɓ/, while /i/ and /I/ were rarely confused with the other vowels and were correctly identified for more than 80 per cent of the time for both patients. It has been shown that the difference in second formant frequency between /i/ and the two back vowels /a/ and /ɔ/ /ɔ/ is much greater than the difference between /a/ and /ɔ/. These differences in second formant frequency are illustrated in the following set of mean second formant frequencies given in Peterson and Barney (1952):

/i/	(2290 Hz, for male; 2790 Hz, for female)
/a/	(1090 Hz, for male; 1220 Hz, for female)
/ɔ/	(840 Hz, for male; 920 Hz, for female)

It has been mentioned in Section III that the electrode which was activated in a 20 ms time frame was selected on the basis of the second formant frequency estimate (F2), and that each electrode was assigned to a specific F2 subband. A small difference in F2 estimate resulted in a small separation between the activated electrodes and therefore an increase in the number of confusions. Similar reasoning applies to the confusions amongst the short vowels.

Another factor that exerts an influence on the vowel results is the variability of the F2 estimate. As a result of this variability, more than one single electrode would have been assigned to a particular vowel, and these single-electrodes would overlap with those assigned to other vowels.

Consonant confusions

The general patterns of consonant confusions generated by the two patients were similar. The only consonant which was identified correctly a large percentage of the time was /w/—MC1 (female speaker: 95 per cent, male speaker: 100 per cent); MC2 (female: 90 per cent, male speaker: 80 per cent). With the exception of MC2 with the female speaker, the other consonants which were generally perceived correctly were (ʃ) and /tʃ/ with percentage correct scores ranging from 40 per cent to 90 per cent. For MC2 with the female speaker, /ʃ/ and /tʃ/ were mostly confused with /s/.

In terms of the strategy adopted in the speech processor design, consonant confusions can be analysed on the basis of the following two patterns: voicing confusions and transition confusions. The former refers to the voiced/unvoiced distinction, the latter to consonant identi-

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

fication based on the transitional characteristics of the second formant frequency.

In order to study the pattern of voicing confusions, the ten consonant stimuli were, according to the articulation process for speech production, classified into three groups: voiced (/n/, /r/, /w/, /j/), long unvoiced (/f/, /s/, /ʃ/, /tʃ/) and short unvoiced (/p/, /k/). In articulatory terms, the voiced consonants are produced by vocal cord excitation of the vocal tract, while the unvoiced consonants are produced from the excitation of the tract by noise which is generated by air flow at some point of constriction. Acoustically, voiced consonants correspond to periodic signals, while unvoiced consonants are noisy in character.

Within the unvoiced consonant group, the duration of the consonant segment varies. For the consonant stimuli employed in the present studies the two plosives /p/ and /k/ are shorter in duration than the fricatives, /f/, /s/, /ʃ/ and /tʃ/.

Voicing confusions among the three groups (VOICED, LONG UNVOICED, SHORT UNVOICED) are given in Table VI. Data for the two patients and two speakers were pooled so that the results from 800 presentations are summarized in the confusion matrix. The overall percentage correct for the above voicing-duration classification was 70 per cent. The individual percentage correct was 32 per cent for the short unvoiced group, 79·5 per cent for the long unvoiced group, and 80 per cent for the voiced group.

The high percentage correct scores for the long unvoiced group and the voiced group indicated that the patients were able to distinguish between voiced and unvoiced speech segments by paying attention to the roughness of the hearing sensations produced by electrical stimulation. As described earlier in Section III, the present speech processor codes unvoiced segments as electrical stimuli at a low pulse rate, and codes voiced segments as stimuli with higher pulse rates. The present results suggested that this pulse rate differential is indeed an effective strategy for voiced/unvoiced encoding.

The low percentage correct score for the plosives (/p/, /k/) in the short unvoiced group was the result of incorrect voiced/unvoiced decisions made by the present speech processor. Subsequent analysis of the electrical stimuli showed that the initial unvoiced portions of these consonants had been incorrectly classified as voiced. These decision errors made by the processor were reflected in the results given in the confusion matrix. The results in Table VI showed that more than 50 per cent of the short unvoiced stimuli were perceived as belonging to the voiced category. The correct answers given by the patients for the short unvoiced consonants were likely to be influenced by information provided by the second formant transitions.

For transiton confusions, the following groups are classified as: level transition (/n/, /f/), falling transition (/k/, /s/, /ʃ/, /tʃ/, /j/) and rising

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

TABLE VI

CONSONANT CONFUSIONS: VOICING STIMULUS

		STIMULUS		
		Short-unvoiced	Long-unvoiced	Voiced
Response	Short-unvoiced	51	20	22
	Long-unvoiced	24	255	41
	Voiced	85	45	257
		$\Sigma = 160$	$\Sigma = 320$	$\Sigma = 320$

Short-unvoiced = /p/, /k/.
Long-unvoiced = /f/, /s/, /ʃ/, /tʃ/.
Voiced = /n/, /j/, /w/, /r/.

Percentage correct:

Short-unvoiced: $\frac{51}{160} = 32\%$

Long-unvoiced: $\frac{255}{320} = 79{\cdot}5\%$

Voiced: $\frac{257}{} = 80\%$

Overall percentage correct = $(51+255+257)/800$
$= 70\%$

transition (/p/, /w/, /r/). The direction of second formant frequency transition from consonant to vowel has been found to be an important cue for the distinctions amongst consonants. In the present classification, level transition refers to a minimal second formant frequency shift from the consonant to the vowel in a CV combination, falling transition refers to a high-to-low frequency shift, and rising transition corresponds to a low-to-high frequency trajectory. The general basis on which the above groupings were classified is well documented in the literature (see for example, Ainsworth, 1976). However, the classification of the semi-vowel /r/ requires further comments in the context of the present speech processor design. /r/ was included in the rising formant group because the F2 estimate of this phoneme was contaminated by the third formant. The third formant for /r/ starts quite low, as low as 1500 Hz, which then rises through to the third formant of the vowel at about 2500 Hz, well within the range of 750 to 3000 Hz of the analyzing filter employed for F2 estimation.

Transition confusions amongst the three groups (level transition, falling transition, rising transition) are given in Table VII. Data for the patients and speakers were pooled. The overall percentage correct for transition classification was 67·5 per cent. The individual percentage correct was 30 per cent for the level transition group, 82 per cent for the falling transition group, and 68 per cent for the rising transition group.

The results given in Table VII showed that the patients were able to

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

TABLE VII

CONSONANT CONFUSIONS: TRANSITIONS
STIMULUS

		No. variation	Falling	Rising
Response	No. variation	48	51	42
	Falling	79	329	35
	Rising	33	20	163
		$\Sigma = 160$	$\Sigma = 400$	$\Sigma = 240$

No variation = /n/, /f/.
Falling = /k/, /s/, /ʃ/, /tʃ/, /j/.
Rising = /p/, /w/, /r/.

Percentage correct:

$$\text{No variation} = \frac{48}{160} = 30\%$$

$$\text{Falling} = \frac{329}{400} = 82\%$$

$$\text{Rising} = \frac{163}{240} = 68\%$$

$$\text{Overall percentage correct} = (48+329+163)/800 = 67{\cdot}5\%$$

distinguish different classes of consonants on the basis of the direction of electrode shift. The present speech processor assigns different electrodes to individual ranges of the second formant estimate; a second formant transition is therefore transformed to a shift in the electrode being activated.

The low percentage score for the level transition group could be the result of the variability of the second formant estimate. Although the consonants in the level transition group were defined as having a minimal shift in second formant frequency from the consonant to the vowel in a CV syllable, the trajectory of the F2 estimate was influenced by a certain amount of variability, and could be actually level, rising or falling. Furthermore, the perceptual variability of the patients should also be considered. In this regard, it is of interest to note that MC1 had preference for falling transitions in response to the level transition stimuli, while MC2 had no preference for either rising or falling transitions.

IV. Performance with running speech and sentence

The ability of the patients to identify certain articulatory features as demonstrated in the last section suggested that they should be able to comprehend some running speech by audition alone. Our informal observations with MC1 have shown that he could comprehend without

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

lip-reading some sentences and phrases that are used everyday in the testing situation—for example, "How loud is it", "What is the pitch", "Let's have a cup of coffee", "Who drove in this morning", "Did you watch T.V. last night", "What are the names of your children", "Do you have any pets".

Furthermore, informal studies have shown that MC1 could identify fifteen sentences from a closed set with an accuracy of 100 per cent by audition alone. The following sentences were used for MC1—1. Eat the apple; 2. We work hard; 3. We'll go now; 4. Petrol is expensive; 5. Let's go home; 6. You tell funny jokes; 7. Do you read a lot; 8. Anything on T.V. tonight; 9. I can't play football; 10. This is hard work; 11. I went to the football; 12. It was cold this morning; 13. Let's go for a drink; 14. Will it snow this year; 15. What will you do today.

MC1 was first given training in identifying these sentences. This consisted of reading out the sentences in the order they were written on the page, then taking groups of 3 or 4 successive sentences, presenting these randomly for him to identify and giving him feedback on his response. Initial testing (presenting the 15 sentences randomly) showed that MC1 could identify with a 80 per cent accuracy and by the 4th or 5th testing sessions with 100 per cent accuracy. The performance of MC1 with these sentences was not surprising considering the contextual constraints within the material.

Similar results with sentences have been obtained with MC2. In addition he was tested by reading from a daily newspaper. For example, on the first occasion he was tested with the speech processor four weeks after the operation, he was read the following passage once through the speech processor with assistance from lip-reading:

> "A meeting of Union officials today is likely to call a public transport strike of at least twenty-four hours this week in Victoria, South Australia and Tasmania. Officials from nine Unions will meet in Adelaide to plan action in support of increases of about $50 a week".

The passage was read to him a second time in three to five word segments without help from lip reading, and he was instructed to repeat what he heard as it was spoken. He was able to repeat most of the passage back accurately. However, some phrases and words (example: "today", "is likely to call", "nine Unions", "Adelaide", "in support of") caused difficulty.

V. Conclusions and future directions

The results obtained from the sharpness ranking experiments and vowel and consonant studies demonstrate the generality of some of the observations described in the previous report (Tong *et al.*, 1979). The present results indicated that the hearing sensations generated by electrical stimulation were similar in character for the two patients. The monotonic

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

variation in sharpness from apical to basal electrodes, the spectral colours perceived in the vowel studies, and the consistent pattern of consonant confusions observed in both patients were evidence for the similarities across patients. The duller sensation associated with the electrode in the region of the round window reported by MC2 also confirmed the previous observation by MC1.

The results from the vowel and consonant confusion studies indicated that the principles of the speech-coding scheme are basically sound. The use of a pulse rate differential to signal the voiced/unvoiced distinction, electrode position to convey steady-state second formant frequency information, and electrode trajectory to indicate the direction of the second formant transition were shown to be effective.

However, the low percentage correct scores for the consonants and the restrictiveness of the vowel and consonant test materials indicated room for improvement, both in terms of the present speech processor design and the basic speech coding scheme. As far as the speech processor design is concerned, the crude estimation techniques employed in the speech parameter (FO, AO, F2, A2) extraction section are known to be inaccurate. New techniques are being implemented to improve their accuracy. For the basic speech coding scheme, further strategies for the transformation of the speech signal parameter to electrical parameter are being investigated to encode speech features which require the detection of the first or third formant component or a mixed mode (voiced and unvoiced) of sourse excitation for their identification.

As it stands, however, the speech processor appears to be helpful for communication in real life situations. The observations that the patients' percentage correct scores in the vowel and consonant studies are likely to be related to the amount of exposure to electrical stimuli, and the patients' ability to learn sentences encourage us to believe that the abilities of the patients to communicate by audition alone will be improved with further training.

Appendix

English Phonemes with Word Examples:

/i/	beat
/I/	bit
/a/	bart
/ʌ/	but
/ɔ/	bought
/ɓ/	bot
/p/	rip
/k/	back
/f/	luff
/s/	hiss

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press

/ʃ/	bash
/tʃ/	chew
/n/	no
/w/	we
/j/	you
/r/	read

Acknowledgements

We would like to acknowledge financial assistance from the National Health and Medical Research Council of Australia, the Department of Productivity of the Federal Government, the Channel O Deafness Appeal, and Lions International.

We would also like to thank our colleagues: P. Blamey, P. Seligman, R. Walkerden, G. Cook and R. Black for their support, and Miss J. Maher for the typing.

REFERENCES

AINSWORTH, W. A. (1976) Mechanisms of Speech Recognition. Pergamon Press, Oxford.

CLARK, G. M., BLACK, R., DEWHURST, D. J., FORSTER, I. C., PATRICK, J. G., and TONG, Y. C. (1977) *Medical Progress through Technology*, **5**, 127.

DELATTRE, P. C., LIBERMAN, A. M., COOPER, F. S., and GERSTMAN, L. J. (1952) *Word*, **8**, 195.

HOUSE, A. S. (1961) *Journal of the Acoustical Society of America*, **33**, 1174.

PETERSON, G. E., and BARNEY, H. L. (1952) *Journal of the Acoustical Society of America*, **24**, 175.

PLOMP, R. (1976) Aspects of Tone Sensation. Academic Press, London, p. 109.

TONG, Y. C., BLACK, R. C., CLARK, G. M., FORSTER, I. C., MILLAR, J. B., O'LOUGHLIN, B. J., and PATRICK, J. F. (1979) *Journal of Laryngology and Otology*, **93**, 679.

Department of Otolaryngology,
University of Melbourne,
32 Gisborne Street,
East Melbourne, 3002,
Victoria, Australia.

https://doi.org/10.1017/S002221510009006X Published online by Cambridge University Press