Twelfth Quarterly Progress Report

June 27 through September 26, 1988

NIH Contract N01-NS-5-2396

Speech Processors for Auditory Prostheses

Prepared by

Blake S. Wilson, Charles C. Finley and Dewey T. Lawson

Neuroscience Program Office
Research Triangle Institute
Research Triangle Park, NC 27709

This QPR is being sent to you before it has been reviewed by the staff of the Neural Prosthesis Program

CONTENTS

I. Introduction

The purpose of this project is to design and evaluate speech processors for auditory prostheses. Ideally, the processors will extract (or preserve) from speech those parameters that are essential for intelligibility and then appropriately encode these parameters for electrical stimulation of the auditory nerve. Work in the present quarter included the following:

1. Continued upgrade of the Cochlear Implant Laboratory at Duke, along the lines indicated in our last progress report for this project;

2. Development of interactive software for studies with the laser videodisc implementations of the Iowa Cochlear Implant Battery and of the Speech Pattern Contrast (SPAC) test;

3. Further development and application of integrated field/neuron models of the electrically stimulated cochlea;

4. Installation and application of sophisticated software tools for analysis of results from tests of consonant and vowel identification, with the help of Sig Soli who instructed us on the use of these tools in a visit to RTI in August, 1988;

5. Preparation for two invited lectures for presentation at the 25th Anniversary Symposium of the Kresge Hearing Research Institute, to be held in Ann Arbor, October 3-5, 1988;

6. Presentation of project results at the World Congress on Medical Physics and Biomedical Engineering, in San Antonio, August 6-12, 1988;

7. Continued collaboration with the UCSF team on the development of the speech processor and transcutaneous transmission system for a next-generation cochlear prosthesis; and

8. Continued preparation of manuscripts, including two for the book Models of the Electrically Stimulated Cochlea (to be edited by J.M. Miller and F.A. Spelman).

The main part of the present report, on "Representations of Speech Features with Cochlear Implants," is a slightly modified version of one of the papers indicated in point 8 above. The other paper, on "Models of Neural Responsiveness to Electrical Stimulation," will be presented in the next progress report for this project. Details on points 1, 2, 4 and 7 also will be presented in future reports.

II. Representations of Speech Features with Cochlear Implants

A. Introduction

The development and application of cochlear prostheses have improved the quality of life for many deaf individuals. Much work remains to be done, however, in order to achieve high levels of speech recognition in a majority of patients. In particular, we do not fully understand how different speech processor and implanted electrode designs affect the perception of speech and other sounds. Also, we have only fragmentary and primitive knowledge of how various differences among patients affect their outcomes.

The principal characteristics of five different cochlear implant systems are presented in Table 1. These characteristics include the type of electrode used, the number of stimulation channels, and the method for processing speech inputs. As is evident from the Table, a tremendous diversity exists in current approaches to prosthesis design. Further details on the systems listed in Table 1 can be found in recent reviews [10], [25], [31], [34] and in primary sources [2], [3], [8], [14], [20].

A remarkable finding from evaluations of the disparate approaches to prosthesis design is that several of these approaches can produce high levels of speech perception in some patients. Unfortunately, the converse is also true in that poor levels of performance are found for other patients using the same approaches.

These facts are illustrated in Table 2, which compares the average values and standard deviations of speech recognition scores for the three multichannel systems listed in Table 1. All the data presented in Table 2 were obtained from studies in which single presentations of recorded material were used. The data reported for the Nucleus device are those for the new "F0/F1/F2" processing strategy. Note that the standard deviations are large for all measures and devices. This variability reflects the

Table 1. Cochlear implant systems.

Device	Electrode(s)	Stimulus Channels/ Stimulus Sites	Processing Strategy
3M/House	monopolar	1 / 1	amplitude modulation of 16 kHz carrier
3M/Vienna	longitudinal bipolar or extracochlear	1 / 1	compressed analog with frequency equalization
Nucleus	longitudinal bipolar	2 / 21	feature extraction (F0, F1 and F2), pulsatile stimulation
Symbion	monopolar	4 / 4	compressed analog with filtering
Storz	radial bipolar	4 / 4	compressed analog with filtering

observation that, for any of the listed devices, excellent performance is found for some patients while poor performance is found for others. Finally, it is worth noting that experience with an implant can have a profound effect on the scores for these (and other) tests. With the Storz device, for example, both the NU6 word and CID sentence scores show significant increases ($p < .02$ for the NU6 test and $p < .01$ for the CID test) from the 3 month to the 12 month studies.

Much of this heterogeneity of outcomes may be attributed to variation among patients. The importance of learning effects associated with experience has just been mentioned. Other potentially important variables include differences among patients in the survival of neural elements in the

Table 2. Average percent correct scores for recognition of NU6 words and CID sentences with multichannel cochlear prostheses. The standard deviations of these scores are given in parentheses.

Device	NU6	CID	N	Months Postimplant	Citation
Nucleus	6.5 (6.4)	20.7 (19.5)	53/49*	3	Brimacombe et al., 1988
Symbion	8.5 (10.9)		11	>7	Gantz, 1987
Storz	7.6 (8.2)	21.0 (16.5)	14	3	Wilson et al., 1988d
	20.2 (15.0)	45.9 (26.8)	12	12	"

*N = 53 for the NU6 test and N = 49 for the CID test.

implanted ear, integrity of the central auditory pathways, and cognitive and language skills. The likely influence of patient variables on outcome obviously complicates comparisons aimed at identifying the best types of electrode and speech processor designs for cochlear prostheses. Indeed, such comparisons are not meaningful when small populations of subjects are used for the evaluation of each system, or when test procedures used at the various laboratories involved in the evaluation of individual systems are different.

Recognition of the problem just described helped to initiate two major studies in which relatively large populations of patients implanted with different devices will be tested in a uniform and consistent manner. One of these studies is a cooperative effort among VA medical centers and the other study is being conducted at the University of Iowa [11]. Results from these studies should be of great value in establishing expected levels of performance for contemporary cochlear prostheses.

A complementary method for the comparison of prosthesis systems is to evaluate different processing strategies and electrode coupling configurations in tests with the same implant patient. A key advantage of such tests is that controls are provided for patient variables. Thus, for a given patient with a fixed pattern of nerve survival and fixed levels of cognitive skill, etc., the performance levels of different prosthesis systems can be compared directly.

In recent studies conducted in collaboration with investigators at the University of California at San Francisco (UCSF) and at Duke University Medical Center, our team compared a variety of speech processing strategies in tests with patients implanted with the UCSF/Storz electrode array [51]-[54]. Some of the largest differences in performance among processing strategies were found in comparisons between the compressed analog (CA) processor of the present UCSF/Storz prosthesis and a type of "interleaved pulses" (IP) processor which delivers pulses in sequence to the different channels in the implanted electrode array. To show how the design of the processor can affect the outcome for individual patients, we will restrict ourselves here to descriptions of tests to evaluate and compare the performance of these two processors. The emphasis will be on the results obtained in tests of consonant and vowel identification. In addition, results for the open-set tests of the Minimal Auditory Capabilities (MAC) battery [29] will be mentioned. Results for the full MAC battery and other tests of speech perception are presented elsewhere [52]-[54].

B. Methods

Processing Strategies

In the clinical UCSF/Storz device alternate pairs of the 16 available electrodes are stimulated simultaneously with the CA outputs of a four-channel speech processor. The basic functions of this processor are to compress the wide dynamic range of input speech signals into the narrow dynamic range available for electrical stimulation of the cochlea, and then to filter the compressed signal into individual frequency bands for presentation to each pair of stimulated electrodes. Typical waveforms of the CA processor are shown in Fig. 1. The top trace in each panel is the input signal, which in this case is the word "bought." The other waveforms in each panel are the filtered output signals for 4 channels of intracochlear stimulation. The bottom left panel shows an expanded display of waveforms during the initial part of the vowel in "bought," and the bottom right panel shows an expanded display of waveforms during the final /t/. The lower panels in Fig. 1 thus exemplify differences in waveforms for voiced and unvoiced intervals of speech.

In the voiced interval the relatively large outputs of channels 1 and 2 reflect the low-frequency formant content of the vowel, and in the unvoiced interval the relatively large outputs of channels 3 and 4 reflect the high-frequency noise content of the /t/. In addition, the clear periodicity in the waveforms of channels 1 and 2 reflects the fundamental and first formant frequencies of the vowel during the voiced interval, and the lack of periodicity in the output of any channel reflects the noise-like quality of the /t/ during the unvoiced interval. As has been described elsewhere [36]-[38], [54], this representation of speech features can support high levels of open-set recognition for many (but not all) of the patients implanted with the UCSF/Storz prosthesis.

A concern associated with the use of multichannel CA processors is that of channel interactions [49]. Simultaneous stimulation of two or more channels with continuous waveforms results in summation of the electric

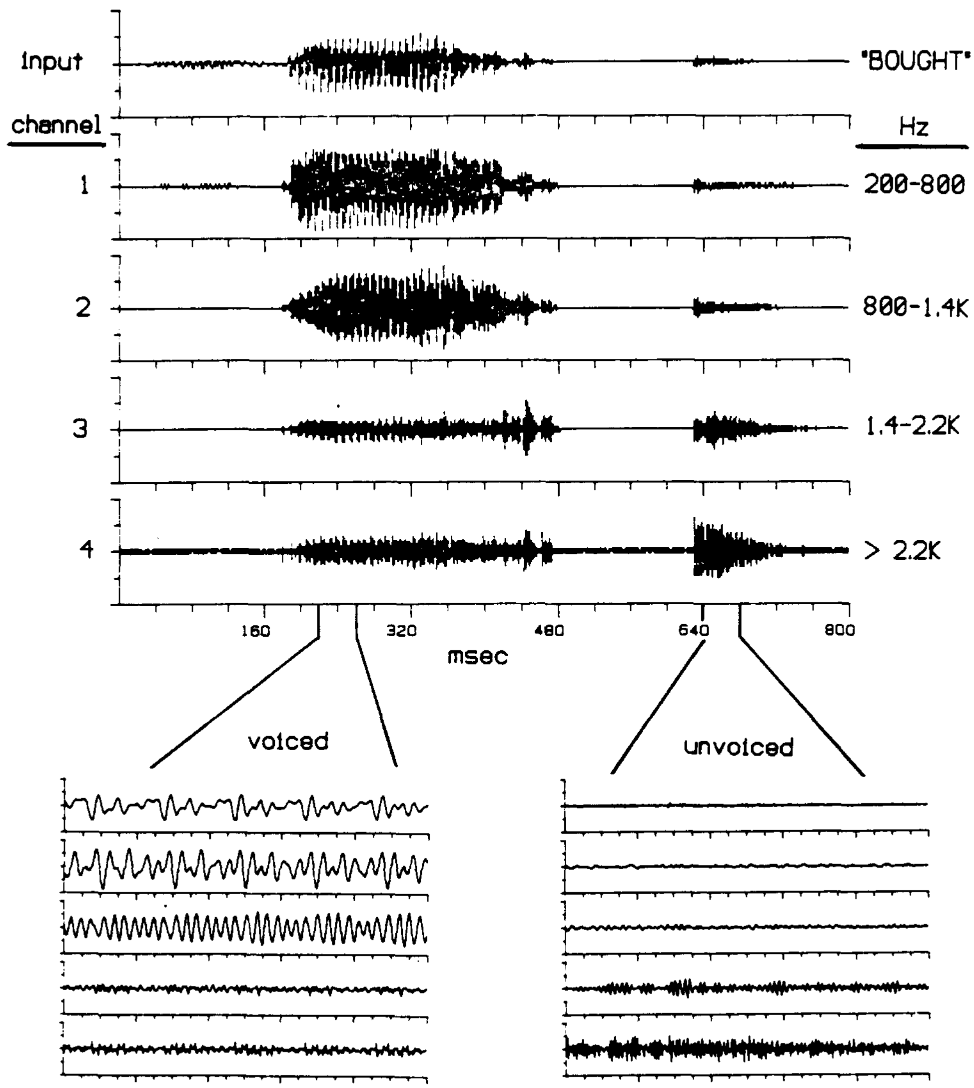

Fig. 1. Waveforms of a compressed analog (CA) processor.

fields from the individual bipolar pairs of electrodes. This summation can exacerbate interactions among channels, especially for patients who require high stimulation levels. Summation of stimuli from multiple channels also depends on the phase relationships among the waveforms. Because these relationships are not controlled in a multichannel CA processor, representation of the speech spectrum may be further distorted by continuously changing patterns of channel interaction. A reduction of channel interactions might increase the salience of channel-related cues for implant patients.

The problem of channel interactions is addressed in the IP processor of Fig. 2 through the use of nonsimultaneous stimuli. There is no temporal overlap between stimulus pulses so that direct summation of electric fields produced by different electrode channels is avoided. The energy in each frequency band of the input signal is coded as the amplitude of the pulses delivered to the corresponding stimulus channel. Distinctions between voiced and unvoiced segments of speech are represented by the timing of cycles of stimulation across the electrode array. In this particular processor stimulation cycles are timed to occur in synchrony with the detected fundamental frequency for voiced speech sounds and at the maximum rate (with one stimulation cycle immediately following its predecessor) for unvoiced speech sounds. The timing of stimulation cycles for voiced and unvoiced intervals can be seen in the lower panels of Fig. 2.

Comparison of Figs. 1 and 2 shows large differences in the stimuli presented for the CA and IP processing strategies. One might expect that the CA processing strategy would provide the greatest benefits to patients who could appreciate details in the stimulus waveforms (see above) and who have low thresholds of stimulation. In contrast, the IP processing strategy might be expected to provide superior performance for less-fortunate patients who cannot make use of such details in CA stimulus waveforms and who have high thresholds of stimulation. We note that high thresholds of stimulation and high levels of measured channel interactions with simultaneous stimuli are both regarded as signs of poor nerve survival in the implanted ear [12], [21], [35], [49]. Thus, application of an IP processor may confer special benefits for patients with poor nerve survival.

Subjects

Six patients implanted with the UCSF/Storz cochlear prosthesis [19], [20], [38], [54] participated as subjects in this study. Tests with the CA

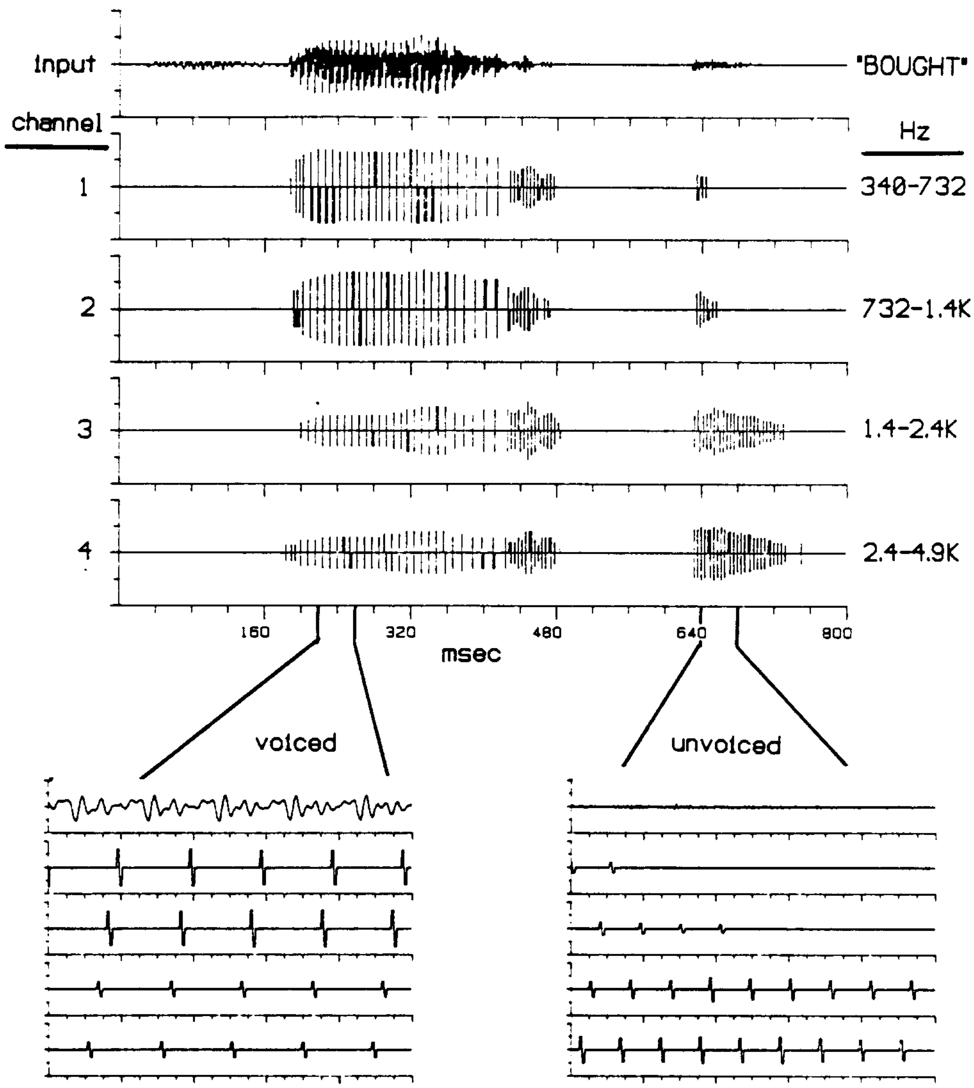

Fig. 2. Waveforms of an interleaved pulses (IP) processor.

processor were conducted with each patient's clinical device, and tests with the IP processor were conducted either with computer simulations [50] or a real-time, microprocessor-based instrument [9].

It is important to note that certain attributes of these subjects favored the CA processor in comparisons of the CA and IP strategies. First, all six subjects entered the study with substantial experience using the CA processor. The average experience with this processor approximated one year of daily use. In contrast, experience with the IP processor was limited to that obtained in a six day period of testing a variety of processors with each subject. As mentioned in the Introduction (and discussed in detail elsewhere, see references [7], [42] and [54]), such a disparity in experience might strongly favor the CA processor.

An additional factor weighing against the IP processor was the use of the 4-channel, UCSF/Storz transcutaneous transmission system (TTS). The principal limitations of that system for IP processors were (a) inadequate levels of voltage compliance for stimulation with short-duration pulses, (b) the small number of channels, and (c) lack of current control in the stimulus waveforms. Half of the subjects were further limited to fewer than four channels due to a mode of device failure [38]. Because optimized fittings of IP processors require at least six channels of stimulation and short-duration pulses [51], [52], compromises had to be made in the fitting of IP processors for the subjects of this study.

The parameters selected for the IP processors used by each of the six subjects are presented in Table 3. The best fulfillments of the fitting criteria for IP processors [52] were obtained for subjects HE and MC2. Each had the use of all four stimulation channels and the average pulse width across channels was 0.5 ms/phase or less for these two subjects.

In contrast, relatively poor sets of parameters had to be used for the remaining subjects. Subjects MC1 and JM had only three usable channels and subject RC only two. In addition, long pulse durations (1.0 ms) had to be used for subjects JM and ET.

Tests

Most of the results reported in this QPR are from tests of consonant and vowel identification. The consonants were presented in an /aCa/ context and the vowels in a /bVt/ context. Two consonant tests were used. The

Table 3. Parameters of IP processors.*

Subject	Channels	Pulse Widths/Phase (ms)	Pulse Sep. (ms)	Cycle Time (ms)
MC1	3	0.5	0.5	4.5
HE	4	0.5	0.5	6.0
JM	3	1.0	0.1	6.3
RC	2	0.5	0.1	2.2
ET	4	1.0	0.1	6.4
		1.0		
		0.5		
		0.5		
MC2	4	0.3	0.5	5.2
		0.7		
		0.3		
		0.3		

*All six processors used symmetric biphasic pulses with positive phase leading and with the channels stimulated in base-to-apex order. Stimulation cycles were presented at the fundamental frequency for voiced intervals and at maximum rate (period equal to cycle time) during unvoiced intervals.

first was the one developed at the University of Iowa for measurement of audiovisual consonant perception [43]. A video tape of an adult male speaker provided the visual component of each presentation. The audio track of the tape provided an input to the UCSF/Storz processor or the real-time

IP processor via direct connection. The consonants were /p, b, m, f, v, ʃ, dʒ, s, z, t, d, n, g, k/. Each consonant was presented five times in a randomized list of stimulus presentations. After each presentation, the subject responded by pointing to one choice in a table of the 14 response options. No feedback on correct or incorrect responses was provided. Finally, the order of testing for the different conditions was designed to confer any benefits of learning on the CA processor. The order was first to test the IP processor plus vision, then vision alone, and then the CA processor plus vision.

A matrix of stimuli and responses was compiled for each subject and condition. The matrices were then summed across subjects for each of the conditions. These summed matrices provided the inputs to the analyses described in the Results section of this report. The raw summed matrices may be found in Appendix 1.

The second consonant test was one suggested by Earl Schubert [39] to assess the ability to distinguish the nonlabial consonants with the greatest frequencies of occurrence in spoken English. These consonants include / ð, s, z, t, d, n, k, l/ and are difficult or impossible to distinguish with speechreading alone (particularly /s, z, t, d, n/). Schubert reasoned that a pragmatic approach to processor design and evaluation would be to concentrate on these eight important (but largely invisible) consonants. The primary purpose of the vowel test was to measure the ability to discriminate relatively-large differences among the selected vowels in the frequencies of the first and second formants. The vowels included /i, I, ɔ, o, u/.

Single exemplars of the tokens in the last two tests (hereafter called the "RTI Tests") were recorded and digitized from representative utterances of an adult male speaker. The digitized tokens were used as inputs to the UCSF/Storz processor (after appropriate digital-to-analog conversion) or the computer simulation of the IP processor. A single block of trials included three presentations of each of the consonants or five presentations each of the vowels in random order. Multiple repetitions of a token were available at regular intervals during each presentation. At the beginning of each presentation a display of response options was shown on a computer terminal used by the subject. The subject responded by touching a key on the terminal. Usually a response was entered after the first or second repetition. At the end of a block, the subject was given the overall percent correct score and an indication of the principal confusions made during the test. With few exceptions, no feedback was given during a block. In the exceptional cases (12 out of 137 blocks), feedback was provided across conditions so that no processor would receive an advantage over another.

The conditions for both RTI tests included vision only, CA processor plus vision, IP processor plus vision, CA processor only, and IP processor only. For the conditions with a visual component, speechreading information was provided by miming the tokens in synchrony with the stimulus repetitions. The same person (DTL) mimed the tokens for all subjects.

Blocks of trials were repeated as time permitted during the six days of testing with each subject. Because many other tests were being conducted during this same period [52]-[54], the total number of trials for the RTI tests was not uniform across subjects and conditions. The actual totals are presented in Table 4. For the great majority of subjects and conditions, the number of trials with each token for the consonant test was 6 or more and the number for the vowel test was 10 or more.

As with the Iowa test, matrices of stimuli and responses were compiled for all subjects and conditions. Each RTI matrix was normalized to show the fraction of responses in each cell, and the normalized matrices were then summed across subjects for each of the conditions. The estimates of matrix responses calculated in this way reflect balanced contributions from all subjects for each condition while still using all of the available data. The final matrices for the RTI tests may be found in Appendix 1.

In addition to the tests of consonant and vowel identification, the CA and IP processors were further evaluated with an extensive series of speech

Table 4. Number of presentations of each token in the RTI tests for the indicated subjects.

Test	Subject	Condition*				
		V	CA+V	IP+V	CA	IP
Consonant	MC1	3	9	15	9	18
	HE	6	6	9	3	9
	JM	9	9	6	9	6
	RC	6	6	9	6	12
	ET	3	6	6	9	6
	MC2	6	3	6	3	6
Vowel	MC1	10	15	10	15	10
	HE	10	10	30	15	15
	JM	15	10	5	10	5
	RC	10	10	10	10	10
	ET	5	10	5	10	5
	MC2	10	10	10	10	10

*Abbreviations are V for Vision, CA+V for compressed analog plus vision, IP+V for interleaved pulses plus vision, CA for compressed analog only, and IP for interleaved pulses only.

perception tests. These additional tests included all subtests of the Minimal Auditory Capabilities (MAC) battery [29]; the Diagnostic Discrimination Test (DDT) of consonant confusions [13]; and connected discourse tracking with and without the prosthesis [4], [30]. The results from the subtests of the MAC battery designed to measure open-set recognition will be discussed in this report. Results from the full MAC battery, along with the results from the other tests, are presented elsewhere [52]-[54].

C. Results

Percent Correct Scores

The means and standard deviations of overall percent correct scores for the tests of consonant and vowel identification are presented in Table 5. To compare results among the conditions for each test, a randomized-blocks analysis of variance (ANOVA) was conducted with the subjects as blocks and the conditions as the second factor. For the Iowa consonant test a significant effect of conditions was found ($F(2,10) = 20.56$; $p < .001$). Post hoc comparison of the means using the Tukey multiple comparisons procedure showed that the means for both processor plus vision conditions are significantly higher than the mean for the vision only condition ($p <$ processor plus vision conditions is not significant ($p > .05$).

A significant effect of conditions was also found for the RTI consonant test ($F(4,20) = 22.23$; $p < .001$). Post hoc comparison of the means, with the Tukey procedure, can be summarized as follows:

1. The means for the processor plus vision conditions are significantly higher than the mean for the vision only condition ($p < .01$, both processors);

2. The difference between means for the processor plus vision conditions is not significant;

3. The mean for the IP only condition is significantly higher than the mean for the CA only condition ($p < .05$);

4. The means for both processor plus vision conditions are significantly higher than the means for the processor only conditions ($p < .01$ for the CA processor and $p < .05$ for the IP processor);

Table 5. Means and standard deviations (SD) of consonant and vowel identification scores in percent correct.

	Iowa consonants		RTI consonants		RTI vowels	
condition*	Mean	SD	Mean	SD	Mean	SD
Vision only	33.3	5.3	42.0	8.3	83.4	8.7
CA + Vision	54.5	16.6	69.6	13.6	94.0	2.5
CA only			51.8	12.6	86.0	5.1
IP + Vision	64.0	6.7	79.7	13.0	87.3	9.9
IP only			65.8	14.6	83.3	10.8

*Abbreviations are CA for compressed analog processor and IP for interleaved pulses processor.

5. The mean for the IP only condition is significantly higher than the mean for the vision only condition ($p < .01$); and

6. None of the differences among means for the remaining combinations of conditions is significant.

Finally, no significant differences among conditions were found for the RTI vowel test ($F(4,20) = 2.81$; $p > .05$). We note that scores are quite high for all conditions of this test. Possibly, true differences among conditions may have been masked by ceiling effects and, if so, a more difficult test might demonstrate such differences.

Information Transmission (IT) Analysis

Although overall percent correct scores can serve as a rough indication of processor performance, they provide little or no insight into the strengths and weaknesses of specific strategies. That is, the pattern of confusions (and correct responses) in a consonant or vowel identification test can provide much more detailed information on processor performance than the overall percent correct score.

To evaluate the patterns of confusions for the conditions of this study, the combined matrix for the responses of all subjects for each condition (see Appendix 1) was used as an input to the information transmission (IT) analysis described by Miller and Nicely [23]. In this analysis the "relative transinformation" is calculated for selected articulatory or acoustic features of the phonemes in the identification tests. The relative transinformation score for each feature, expressed here as percent information transfer, indicates how well that feature was transmitted to the subjects. The consonant features selected for the present study were voicing (voice), nasality (nasal), place of articulation (place), duration (durat), frication (fric), envelope cues (envel) and visual cues (viseme). The vowel features were first formant frequency (F1), second formant frequency (F2), duration (durat) and visual cues (viseme). Assignments of these features for the phonemes in the Iowa and RTI tests may be found in Appendix 2.

The results from IT analysis of the Iowa consonant matrices are presented in Fig. 3. The open bars show IT scores for the vision only condition, the bars with diagonal lines show the scores for the CA processor plus vision condition, and the solid bars show the scores for the IP processor plus vision condition. Note that the viseme and place features are transmitted equally well for all three conditions. The high score for place in the vision only condition is indicative of the high redundancy between assignments for the place and viseme features. That is, a front (bilabial and labiodental) place of articulation usually can be distinguished from other places of articulation through speechreading alone

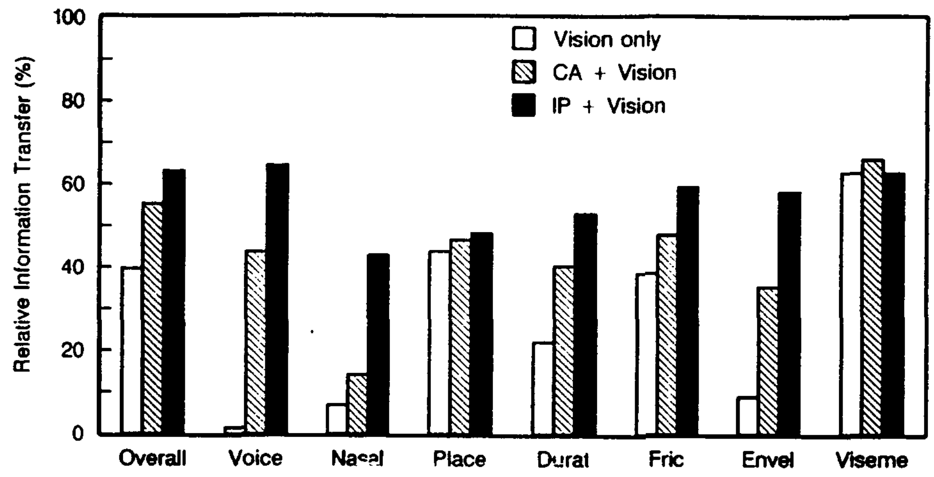

Fig. 3. Relative information transfer of speech features for the Iowa consonant test.

[28], and this ability is reflected in the choices for the viseme groupings. Thus, if subjects can distinguish the groups /p,b,m,f,v/, /ʃ,dʒ/ and /d,s,z,t,d,n,g,k/ through speechreading, then the scores for both viseme and place will be high.

Other features that exhibit some redundancy with the viseme groupings are duration and frication. The relatively high scores for these features with vision alone reflect this overlap. On the other hand, the scores for voicing, nasality and envelope are all low for the vision only condition. These features are invisible on the lips and have little or no redundancy with the viseme groupings.

The scores for both processor plus vision conditions demonstrate increases over the scores for the vision only condition. Especially large increases are found for the features of voicing, duration and envelope. In addition, the scores for overall information transfer are higher for the processor plus vision conditions.

Comparison of the scores obtained with the two processors indicates superiority of the IP processor for all features except place and viseme, where the scores are about the same. Scores for the IP processor are much higher for the features of voicing, nasality and envelope. Inasmuch as these are the very features that have little or no redundancy with the viseme groupings, the IP processor might be expected to provide a highly effective supplement to speechreading.

The general finding of superior performance with the IP processor is also evident in the results from IT analysis of the RTI consonant matrices. Results for the vision only and processor plus vision conditions are presented in Fig. 4, and results for the processor only conditions are presented in Fig. 5. In Fig. 4 the open, diagonally lined and solid bars again show IT scores for vision only, CA processor plus vision and IP processor plus vision, respectively. In Fig. 5 the stippled bars show IT scores for the CA processor only, and the vertically lined bars show the scores for the IP processor only.

For the conditions with a visual component (Fig. 4), high scores again are obtained for the viseme feature. Because the consonants in the RTI test all have a nonlabial place of articulation, however, high scores for the viseme feature merely show that the groups /s,z,t,d,n/, /k,l/ and /ð/ can be distinguished. /ð/ and /l/ usually are visible through tongue protrusion and tongue flap, respectively, even though they have nonlabial places of articulation. Perception of these cues for /ð/ and /l/ can produce relatively high viseme scores for the consonants in the RTI test.

Another effect of the choice of consonants for the RTI test is to hold place of articulation essentially constant. All consonants except /k/ have a mid place of articulation [41]. Thus, the only distinction that has to be made to produce high place scores is the one between /k/ (back place of articulation) and the remaining consonants. The low place score for the vision only condition in Fig. 4 reflects the fact that the place and viseme features are not redundant for the particular consonants of the RTI test.

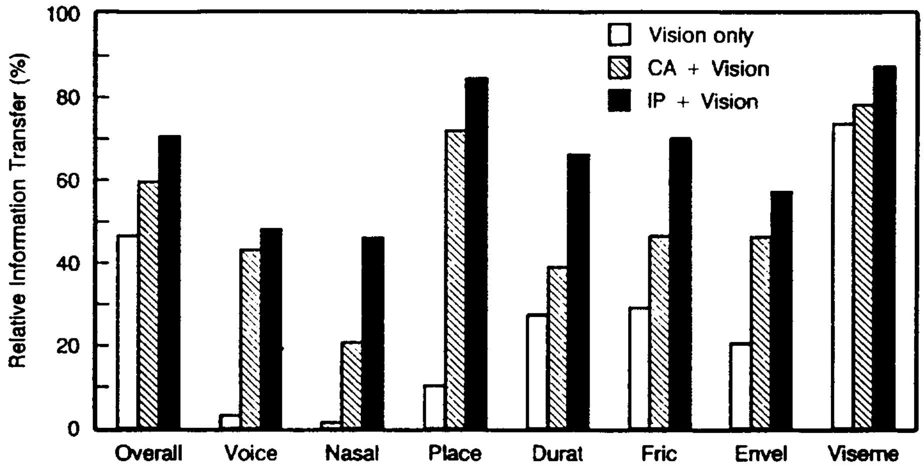

Fig. 4. Relative information transfer of speech features for the vision only and vision plus processor conditions of the RTI consonant test.

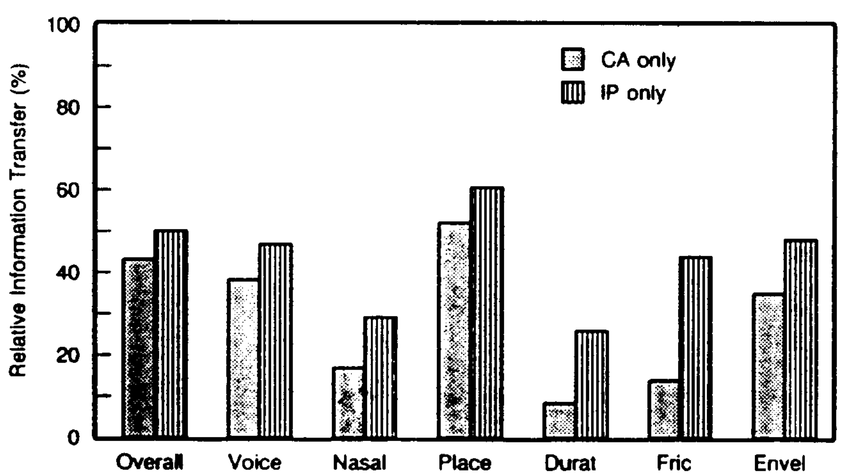

Fig. 5. Relative information transfer of speech features for the processor only conditions of the RTI consonant test.

The scores for all other features (voicing, nasality, duration, frication and envelope) are generally consistent with the scores for the vision only condition of the Iowa test.

Comparison of results across conditions again shows increases over the the vision only scores when either processor is used with speechreading. The largest increases are found for the features of voicing, nasality, place and envelope. The increases for voicing and envelope are quite similar to those found for the Iowa test. The increases for nasality and place, however, are not seen (place) or not as large (nasality) in the Iowa results. The difference in the increases for place can be attributed to the particular choice of consonants in the RTI test, as outlined above. The difference in the increases for nasality is one of degree in that increases are found for both tests, but the relative increase for the CA processor plus vision over vision only is not as large for the Iowa test compared to the increase for the RTI test. This difference between tests again might be a consequence of the different choices of consonants: the only nasal in the RTI test is /n/, while the Iowa test contains /n/ and /m/. The single distinction of /n/ from the remaining consonants may allow higher scores for the nasal feature in the RTI test.

As with the Iowa test, large increases are found in feature transmission scores when the IP processor is used instead of the CA processor for the vision plus processor conditions. The IP processor produces at least some increase in the score for every studied feature, and substantial increases are demonstrated for the features of nasality, duration and frication. The same pattern of increases is evident in the scores for the Iowa test; however, the relative increases for the voicing and envelope features are greater with the Iowa test while the relative increases for the duration and frication features are greater with the RTI test. These differences probably can be attributed to the differences in the consonant sets and to test variability. In all, the patterns of results from the Iowa and RTI tests are remarkably consistent. Both patterns demonstrate substantial gains over vision alone when either processor is used in conjunction with speechreading, and both patterns show superiority of the IP processor. In addition, the particular differences in feature scores found between conditions for one of the tests usually are found for the other test as well.

The results from the RTI test for the processor only conditions (Fig. 5) mirror those reviewed above for the processor plus vision conditions (Fig. 4). Specifically, the IP processor again produces an increase in the score for every studied feature, and substantial increases are found for the features of nasality, duration and frication. Moreover, for all features the ratios of the scores for the CA processor plus vision and IP processor plus vision conditions (Fig. 4) closely approximate the ratios for the CA processor only and IP processor only conditions (Fig. 5). These findings suggest that the IP processor provides additional cues which are utilized by the subjects in both the hearing only and hearing plus vision conditions.

In contrast to the results from the Iowa and RTI consonant tests, the IT scores from the RTI vowel test indicate superiority of the CA processor. These scores for the vowel test are presented in Fig. 6, where the coding of the bars for the various conditions is identical to the coding used in Figs. 3 - 5. Comparison of the IT scores between processors shows that the CA processor produces higher or equivalent scores for every feature. For the processor plus vision conditions higher scores are obtained for overall transmission, F1 and duration, and for the processor only conditions higher scores are obtained for these features and F2. In the cases where equivalent scores are found (F2 and viseme features, processor plus vision conditions), ceiling effects may have masked true differences between the processors. A more difficult test (with, for example, more vowels and less redundancy between assignments for the F2 and viseme features) would provide a more sensitive detector of any difference between processors. In any event, the present results show that the CA processor is superior at least for the transmission of F1 and duration information.

The most general observations from the IT data reviewed above are that (a) the IP processor produces higher or essentially equivalent scores for every studied feature of the phonemes in the Iowa and RTI consonant tests

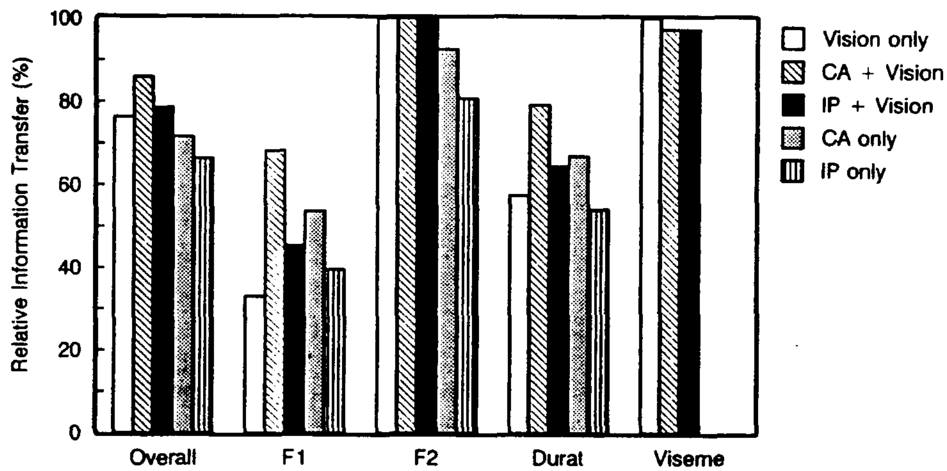

Fig. 6. Relative information transfer of speech features for the RTI vowel test.

and (b) the opposite is found for every studied feature of the phonemes in the RTI vowel test. To evaluate the significance of these observations, a blocked ANOVA was conducted for each test with the features as blocks and the processor plus vision and processor only conditions as the second factor. Because the objective was to compare the contributions of the two processors, the viseme feature was not included in the analysis. The ANOVA results supported the general observations for each of the three tests. In particular, significant effects of conditions were found for all three tests (for the Iowa consonant test $F(1,6) = 18.19$ and $p < .01$; for the RTI consonant test $F(2,18) = 19.92$ and $p < .001$; and for the RTI vowel test $F(3,9) = 18.33$ and $p < .001$). The significant effect for the Iowa test demonstrated superior performance of the IP processor ($p < .01$). Post hoc comparison of the means for the four conditions of the RTI consonant test (using the Tukey procedure) also demonstrated superiority of the IP processor (for the processor plus vision conditions $p < .01$, and for the processor only conditions $p < .05$). Finally, post hoc comparison of the means for the RTI vowel test demonstrated superiority of the CA processor (for both sets of processor conditions, with and without vision, $p < .05$).

Sequential Information Analysis (SINFA)

As noted above, interpretation of the results from IT analysis is complicated by the presence of redundancies in feature assignments for the phonemes in the identification test. Because some features are highly redundant with others (e.g., viseme and place for the Iowa consonant test), IT analysis cannot determine the extent to which each separate feature is used by the subjects in making their judgments. To address this problem of IT analysis, Wang and Bilger [46] developed a procedure designed to remove the effects of redundancies among features. In this procedure, called Sequential Information Analysis (SINFA), the unconditional IT scores are first calculated for all features, as in the standard procedure for IT analysis. Then the feature with the highest score for percent information transfer is identified and held constant for subsequent iterations of IT analysis. In the subsequent iterations any redundancies between the previously-identified feature and the remaining features are removed by holding the identified feature constant. At the conclusion of each iteration the remaining feature with the highest information transmission score is identified and added to those being held constant for all subsequent iterations. Iterations continue until all features have been examined or until the remaining features can account for less than 1% of the total received information. The final output of SINFA thus consists of a sequence of the features that are most salient at each successive iteration. In addition, the SINFA results indicate the relative contribution of each feature (in the same sequence) to the pattern of judgments made by the subjects.

SINFA results for the Iowa consonant test are presented in Fig. 7. For each of the three conditions the order of identified features is indicated from left to right (where the leftmost bar in the panel for each condition shows the first-identified or most salient feature for that condition) and the relative contribution of each feature is indicated by the length of its bar. The full width of each panel corresponds to 100%. Thus, the rightmost

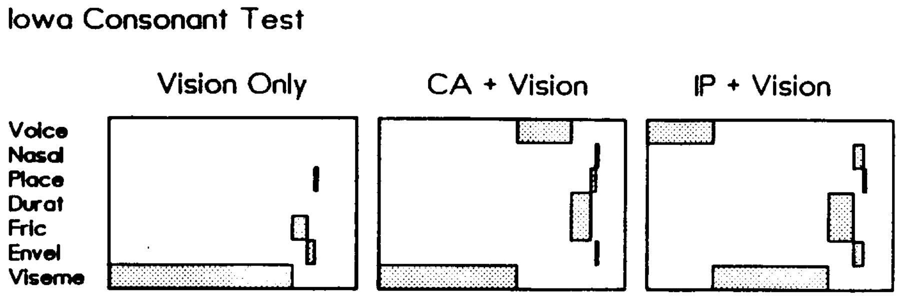

Fig. 7. Sequential information analyses of results from the three conditions of the Iowa consonant test.

extent of the bar for the last-identified feature indicates the proportion of the received information that is explained by all identified features. This proportion, in turn, indicates how completely the identified features can account for the subjects' observed judgments. The assignments across phonemes for two or more features (see Appendix 2) can become identical in subsequent iterations of SINFA as more and more features are held constant. In cases where multiple features with identical assignment patterns are identified as most salient in a SINFA iteration, bars are shown for all such features for that iteration. For example, the third bar (from left to right) in the panel for the CA processor plus vision condition indicates that (a) the features of duration and frication had identical assignment patterns at the third SINFA iteration and (b) the combined "duration/frication" feature was most salient for that iteration.

Returning now to the results of Fig. 7, SINFA demonstrates large contributions of visual inputs to the judgments made for all three conditions. For the vision only condition the viseme feature accounts for 74% of the received information, and for the CA processor plus vision and IP processor plus vision conditions this feature accounts for 56 and 47% of the received information, respectively. As might be expected, the viseme feature is most salient for the vision only condition. In addition, the remaining features identified by SINFA for this condition account for only 10% of the received information.

The viseme feature is also most salient for the CA processor plus vision condition. However, other features now make substantial contributions to the judgments. Most salient among these other features are voicing (22%) and frication/duration (8%).

In contrast to the dominance of the viseme feature for the judgments of the vision only and CA processor plus vision conditions, the first-identified feature for the IP processor plus vision condition is voicing (27%). This finding suggests that an auditory cue is most important for making the judgments of this latter condition, even when speechreading information is presented. Other auditory cues contributing to the judgments are frication/duration (10%) and nasal/envelope (4%).

As can be appreciated from the rightmost extents of the final bars in each panel of Fig. 7, the selected features account well for the patterns of judgments. The proportion of received information explained by these features for the vision only condition is 84%, and the proportions for the two processor plus vision conditions are both 89%.

SINFA results for the five conditions of the RTI consonant test are presented in Fig. 8. Large contributions of visual inputs are again demonstrated for the three conditions with a visual component. The viseme feature is the most salient feature for each of these conditions. The proportions of received information accounted for by this feature are 68, 57 and 54% for the vision only, CA processor plus vision and IP processor plus vision conditions, respectively. As might be expected, the greatest proportion is found for the vision only condition.

In addition to the contributions made by visual inputs, results for the two processor plus vision conditions show that substantial contributions are made by auditory inputs. In the CA processor plus vision condition these latter contributions include place (13%), frication/duration (10%) and

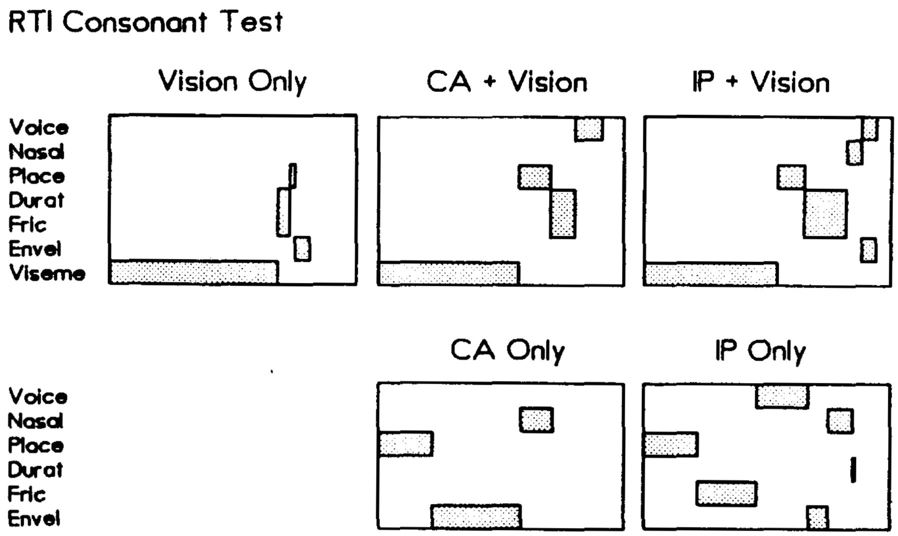

Fig. 8. Sequential information analyses of results from the five conditions of the RTI consonant test.

voicing (11%), and in the IP processor plus vision condition these contributions include place (11%), frication/duration (17%), nasality (6%) and voicing/envelope (6%). Note that place and frication/duration are salient features for both processor plus vision conditions. Also note that features other than the viseme feature make a greater contribution to the judgments made for the IP processor plus vision condition (40%) compared with the contribution for the CA processor plus vision condition (34%). A large part of the difference between the two conditions is the demonstrated access to the nasality feature with the IP processor, which is not found with the CA processor. The IP processor also appears to provide greater access to frication/duration information and to envelope information.

Greater access to nonvisual features of consonants is also demonstrated for the IP processor in the processor only conditions. Three features are identified by SINFA for the CA processor (place, 22%; envelope, 36%; nasality, 13%), while all six nonvisual features are identified for the IP processor (place, 22%; frication, 24%; voicing, 21%; envelope, 8%; nasality, 10%; duration, 1%). In addition, the identified features explain a greater proportion of the received information with the IP processor (86%) than with the CA processor (71%).

To summarize the SINFA results for the consonant tests, we note the following:

1. The viseme feature accounts for a large percentage of the received information for all conditions with a visual component;

2. The viseme feature is the most salient feature for all of these conditions except the IP processor plus vision condition, for which the feature of voicing is most salient;

3. The addition of processor inputs increases access to features other than the viseme feature;

4. Among the most salient of these other features are voicing (both processors, Iowa test), duration/frication (both processors, Iowa and RTI tests) and place (both processors, RTI test);

5. The IP processor further provides access to the features of nasality and envelope for the processor plus vision conditions (Iowa and RTI tests);

6. Both processors provide access to place, nasality and envelope features for the processor only conditions (RTI test); and

7. The IP processor further provides access to the features of frication, voicing and duration.

Finally, SINFA results for the RTI vowel test are presented in Fig. 9. As with the consonant tests, the viseme feature is most salient, and

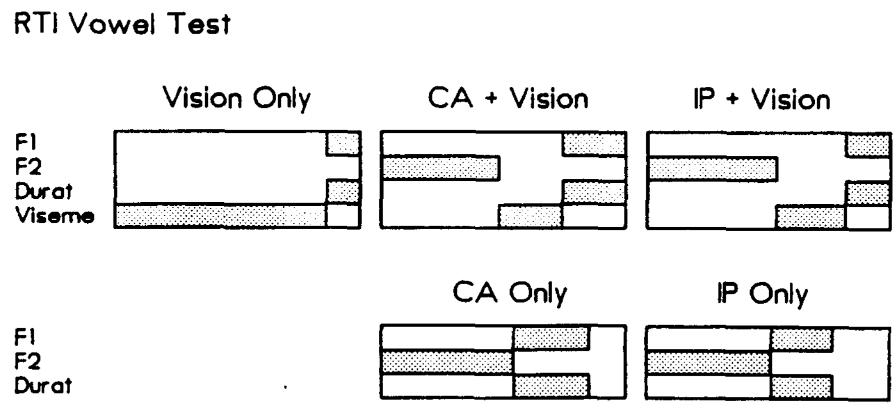

Fig. 9. Sequential information analyses of results from the five conditions of the RTI vowel test.

accounts for the great majority of received information, for the vision only condition. Unlike the results from the consonant tests, however, an acoustic feature (F2) is more salient than the viseme feature for the processor plus vision conditions. Also, an identical pattern of feature rankings is found for these latter conditions, i.e., F2, viseme and F1/duration. The only difference between the processors is in the proportions of received information explained by the different features. The proportions of the F2 and viseme features are higher for the IP processor (53 vs. 48% for F2, and 28 vs. 26% for viseme), and the proportion of the combined F1/duration feature is substantially higher for the CA processor (26 vs. 18%). These differences are consistent with the clearly superior IT scores of the CA processor for F1 and duration (Fig. 6).

Greater access to F1 and duration cues is also demonstrated for the CA processor in the SINFA results for the processor only conditions. As shown in the bottom panels of Fig. 9, the amount of received information accounted for by the combined F1/duration feature is substantially higher for the CA processor (31%) than for the IP processor (25%). Also, the amount of received information explained by the F2 feature is somewhat higher for the

CA processor (54 vs. 51%).

In summary, the CA processor provides greater access to the combined F1/duration feature in both the processor plus vision and processor only conditions. This processor may also provide greater access to the F2 feature in the processor only conditions.

Open-Set Recognition Scores

The most difficult tests normally administered to assess the performance of patients with cochlear implants are tests of open-set recognition. Good performance on these tests probably requires a host of linguistic and cognitive skills that are not tapped in tests of consonant and vowel identification. That is, the open-set tests help to evaluate the integration of segmental identification (consonants and vowels), prosodic cues and contextual information. The open-set tests thus provide complex measures of the representation of speech sounds at the auditory periphery and the interpretation of this representation in the central nervous system. Also, in terms of the "bottom line" for implant patients, the open-set tests mimic many aspects of everyday listening situations.

Results from the open-set tests of the MAC battery for the subjects and processors of this study are presented in Fig. 10. The tests include those of spondee recognition (Sp), recognition of monosyllabic words from Northwestern University list six (NU6), recognition of everyday sentences from lists prepared at the Central Institute for the Deaf (CID), and recognition of single words in the context of sentences (WIC). The results for the CA processor are indicated by the stippled bars, and the results for the IP processor are indicated by the vertically lined bars.

Comparison of the results across subjects for each of the open-set tests demonstrates that there are no significant differences between processors (paired $t < 1.71$ and $p > .10$ for all tests). However, substantial differences are found among subjects both in terms of overall performance and in terms of the scores for the two processors. Subjects

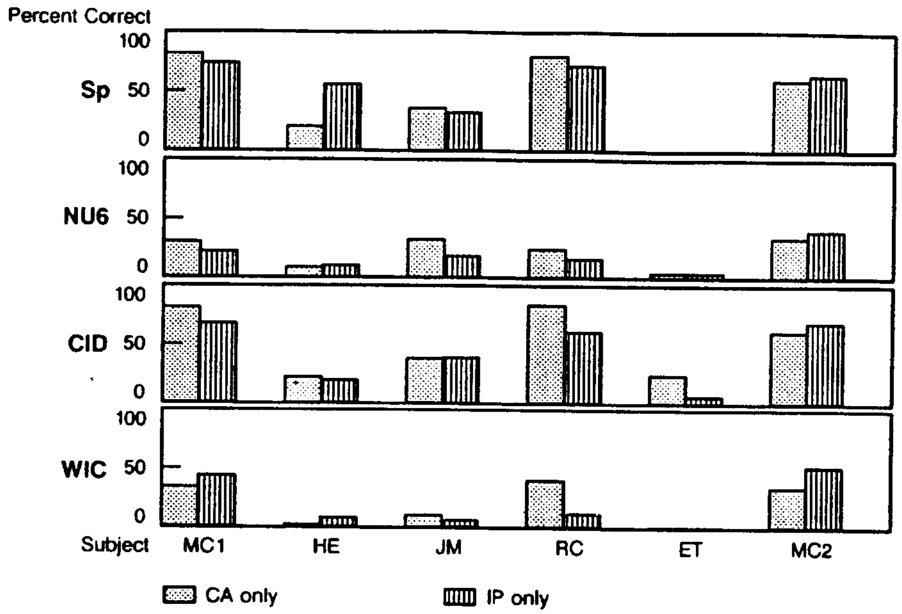

Fig. 10. Results from subtests of the Minimal Auditory Capabilities (MAC) battery designed to measure open-set recognition of speech. Abbreviations for the subtests are Sp for Spondee recognition, NU6 for recognition of monosyllabic words from Northwestern University list six, CID for recognition of everyday sentences from lists prepared at the Central Institute for the Deaf, and WIC for recognition of Words in Context.

MC1, RC and MC2 have excellent performance with both processors, while the remaining subjects have either moderate (HE and JM) or poor (ET) performance with both processors. Between processors, subject RC has higher scores with the CA processor for all four tests and subject MC2 has higher scores with the IP processor for all four tests. Paired t comparisons between processors for these tests show that the CA processor is significantly better for subject RC (paired $t = 3.25$; $p < .05$) and that the IP processor is marginally better for subject MC2 (paired $t = 2.90$; $p < .10$). No significant differences are found between processors for the remaining subjects (paired $t < 1.67$; $p > .10$).

D. Discussion

In this study the compressed analog (CA) and interleaved pulses (IP) processors were compared in tests with six subjects implanted with the UCSF/Storz cochlear prosthesis. The tests included those of consonant and vowel identification and of open-set recognition. Each subject had the use of 2 - 4 channels of intracochlear stimulation via the UCSF/Storz transcutaneous transmission system (TTS). Also, each subject had had considerable experience with the CA processor when the processors were compared.

Large differences between processors were demonstrated in the results from the consonant and vowel tests. In general, the IP processor produced superior results for consonant identification and the CA processor produced superior results for vowel identification. The principal gains for consonants were in the transmission of information on voicing, nasality, envelope, frication and duration, and the principal gains for vowels were in the transmission of information on F1 and duration. Results from the tests of open-set recognition indicated an overall equivalence of the two processors across subjects. For particular subjects, however, one processor provided better performance than the other. Subject RC obtained higher scores with the CA processor for all four open-set tests and subject MC2 obtained higher scores with the IP processor for all four tests.

Representations of Speech Features

The differences in results for the consonant and vowel tests are consistent with differences in the ways in which each processor presents speech stimuli. The CA processor simultaneously presents continuous analog waveforms to all of the stimulation channels in a multielectrode array, while the IP processor presents nonsimultaneous pulses to the same channels. In addition, the type of IP processor used in the present study provides explicit coding of fundamental frequency (F0) and voiced/unvoiced intervals. F0 is coded by initiating stimulation cycles in synchrony with the detected F0 during voiced intervals, and unvoiced intervals are signaled by initiating stimulation cycles at randomly varied times or at rates above the normal range of F0.

A possible advantage of the CA processor is in the preservation of details in the stimulus waveforms. Details that may be perceived by some implant patients include: (a) frequency changes up through the range of F1 [8], [15], [48], (b) rapid temporal variations in the envelopes of speech and speechlike stimuli [16], and (c) subtle waveshape changes produced by the addition of frequency components beyond F1 [14].

The results from the vowel test are consistent with the observation that at least some patients can discriminate frequency changes through the range of F1 with CA stimuli. This ability would explain the high information transmission (IT) scores obtained with the CA processor for the F1 feature. Because waveform details are discarded in the IP processor, one might also expect lower scores for F1 with that processor. Differences in scores for transmission of F1 information thus may be attributed to the presence (CA processor) or absence (IP processor) of details in analog waveforms.

Improved perception of F1 also could explain the high IT scores found for the duration feature with the CA processor. The roles of the F1 and duration features are highly correlated for the vowels in the RTI test. This leads to assignment redundancies (see Appendix 2), that could produce high IT scores for duration even if duration _per se_ were not well transmitted. The present results do not allow discrimination among three possibilities: (a) excellent transmission of F1 information only, (b) excellent transmission of duration information only, or (c) excellent transmission of both. Results from previous studies, on the perception of F1 information with CA waveforms [8], [15], [48], support possibilities (a) and (c).

The chief advantage of the IP processor may be the reduction of channel interactions through the use of nonsimultaneous stimuli. This reduction may increase the salience of channel-related cues and therefore improve the representation of frequency components across the speech spectrum. An increased salience of channel cues is likely to become progressively more important as the number of stimulation channels is increased. Thus, this advantage of the IP processor may be more evident in subjects with many channels (e.g., 6 - 10) than in subjects with relatively few channels (e.g., 2 - 4). Indeed, results from previous studies conducted by our group, using a subject fitted with a percutaneous cable, demonstrated large increases in both consonant and vowel identifications when the number of stimulation channels in an IP processor was increased from 4 to 6 [51].

For the subjects of the present study, with only 2 - 4 channels of stimulation, the representation of frequency components with channel cues is necessarily coarse. However, even a coarse representation may provide a significant advantage for the perception of overall spectral shape. This advantage might be reflected in the improved IT scores for many of the consonant features, especially nasality and voicing. The nasals are characterized by an intense formant at approximately 250 Hz and a relatively weak set of higher formants above 800 Hz. The spectral region between the first and second formants is almost devoid of energy. Recognition of this spectral shape could allow identification of the nasals as a class in the consonant identification tests.

A superior representation of spectral shape also would provide greater access to voicing information. An important cue to voicing is the ratio of energy at low frequencies (e.g., below 1500 Hz) to energy at high frequencies (e.g., above 1500 Hz). A high ratio is typically found for voiced intervals of speech and a low ratio is typically found for unvoiced intervals. Faithful transmission of the relative energies in these bands would help to produce correct voiced/unvoiced decisions.

The voiced/voiceless distinction might also be aided by the explicit coding of voiced and unvoiced intervals used in the IP processor. Indeed, percepts produced during unvoiced intervals -- with stimulus pulses presented at randomly varied times or at rates above the normal range of F0 -- should be easily discriminated from the percepts produced during voiced intervals -- with stimulus pulses presented at the F0 rate (see, e.g., refs [6], [33], [40]). Such coding of voiced and unvoiced intervals complements the representation of spectral balance between low and high bands. This multiplicity of cues to voicing could increase the subjects' access to voicing information. Improved transmission of voicing information would be expected to produce increases in the IT scores for voicing, envelope and frication.

A final difference that may favor the IP processor involves the stimuli used for the high frequency (basal) channels. The presence of speech energy at high frequencies is signaled in the CA processor by the delivery of high-frequency analog stimuli to these channels. In contrast, low-frequency pulse trains are used in the IP processor for all channels, avoiding a potential problem with adaptation. Results from both psychophysical [40] and single-unit [17], [26], [31], [44] studies have demonstrated strong adaptation to stimuli with frequencies much above 300 Hz. Thus, the use of such stimuli in the CA processor could produce substantial temporal distortions in the representation of sustained high-frequency components in speech, a problem that may be avoided through the use of an IP processor. A faithful representation of sustained high-frequency components would allow discrimination of the long-duration consonants /ʃ, s, z/ from the remaining consonants in the Iowa and RTI tests. This discrimination would improve the IT score for the duration feature. The IT scores for the envelope and frication features may also be enhanced by a representation that maintains perception of sustained high-frequency sounds.

To summarize the observations on the consonant and vowel results, we note that (a) waveform cues may be used to perceive the F1 feature of vowels with the CA processor and (b) reduced channel interactions and explicit coding of voice/unvoiced information may produce increased scores for the voicing, nasality, frication and envelope features of consonants with the IP processor. In addition, the use of low-frequency pulse trains for the stimulation of basal channels in the IP processor may reduce or eliminate possible deleterious effects of adaptation. This last advantage of IP processors would be expected to produce increases in the transmission of duration, frication and envelope information for consonants.

Comparison of Tests

Although the consonant and vowel tests show clear differences between the CA and IP processors, results from the open-set tests do not demonstrate an overall superiority of one processor over the other. This latter finding is a little surprising in that consonant identification is more important than vowel identification for overall speech intelligibility (see, e.g., refs [5], [23], [24]). Thus, in the absence of other factors, one might expect that the IP processor would produce superior scores on the open-set tests. Many other factors, however, may have affected the results of the present open-set tests. These factors include (a) the disparity in the subjects' experience with the two processors and (b) the fact that good performance on the open-set tests probably involves a host of linguistic and cognitive skills that are not tapped in tests of consonant and vowel identification. In addition, the superior performance obtained with the CA processor for vowel identification may have offset to some degree the superior performance obtained with the IP processor for consonant identification. Finally, the results from the consonant and vowel tests may not be fully representative inasmuch as only a limited number of consonant and vowel tokens were used.

Among these possibilities, we regard the last as least likely because (a) the results from all consonant tests were highly consistent with each other and together included most of the consonants with high frequencies of occurrence in English [5] and (b) the formant space sampled by the vowels in the vowel test spanned a considerable portion of the range of formant frequencies for all vowels in English [32]. The remaining factors are all likely contributors to the open-set results.

The one of these factors that may have favored one processor over the other is the disparity in experience. For the subjects of this study typical experience with the CA processor approximated one year of daily use, while experience with the IP processor was limited to the tests conducted with that processor (among several) during a one-week period. Results from many previous studies have demonstrated large learning effects associated with the experience gained from using a particular prosthesis system (see, e.g., Table 2 and refs [36], [42], [54]). To the extent that such learning is not transferred to a new system (in this case, a different speech processor) [7], [42], one might expect the disparity in experience to influence test scores in favor of the CA processor. In any event, equivalent or superior results on the open-set tests are found with the IP processor for five of the six subjects in the present study. This finding suggests (a) that the IP processor could be applied to these five subjects without any initial deficit, and (b) that, with equivalent experience, the IP processor might emerge as the superior processor for most subjects.

An additional observation from the open-set results is that the performance of the IP processor appears to be highly sensitive to limitations on the optimization of processor parameters. In particular, the subject with the consistently better performance using the IP processor (MC2) had the best fulfillments of the fitting criteria established for that processor [51], [52], [54]. Reference to Table 3 shows that she had the use of all four stimulation channels, a relatively long time between sequential pulses (0.5 ms), and a relatively short time for a complete stimulation cycle (5.2 ms).

In contrast, the subject with the consistently better performance using the CA processor (RC) had only two functional channels of intracochlear stimulation. This number is certainly too few for even a gross representation of the speech spectrum with an IP processor. The relatively poor performance of the IP processor therefore could be attributed at least in part to a poor match to its fitting criteria.

Another possible explanation for RC's superior performance with the CA processor is that he made especially good use of the information present in the CA stimulus waveforms. Indeed, the impressive results obtained with RC (two channels), MC1 (three channels), certain patients in the Vienna series (one channel; see [15], [16]), and certain patients in the Symbion series (four monopolar channels with relatively poor isolation; see [8]) support the hypothesis that the major bearer of information in CA processors is the waveform itself. Although results from studies conducted at UCSF demonstrate that additional information can be provided with four channels of CA stimulation using the UCSF electrode array [27], [37], [47], this additional information is not required for excellent performance in some patients. Most likely, the best results are obtained for patients who have the greatest access to information in the CA waveform(s).

If this second interpretation is pertinent to RC's case, then patients with such special abilities might be best served with a CA processor. Optimal implementations of this processor would provide any additional information the patient might be able to utilize in multiple channels of stimulation. The maximum number of useful channels is likely to be limited, however, by the severe interactions that can occur among channels when simultaneous stimuli are used.

In summary, the results from the open-set tests provide additional insights into the processors and patients of this study. They demonstrate an overall equivalence of the two processors despite considerable experience with the CA processor, and they indicate that performance with the IP processor is highly sensitive to the choice of processor parameters. Finally, the open-set results show that either processor can be superior for particular subjects. This last finding further suggests that substantial gains in speech recognition can be made across a population of implant patients by (a) selecting the best type of processor for each patient and (b) using implanted and external hardware capable of supporting a wide range of processor options.

Hybrid Processor

The results of the consonant and vowel tests show that both the CA and IP processors have strengths. The main strength of the CA processor may be in its presentation of waveform details, and the main strength of the IP processor may lie in its reduction of channel interactions with the use of nonsimultaneous stimuli. It may be possible to combine these strengths in a single "hybrid" processor using both CA and IP stimuli. Recall that the principal advantages of the CA strategy may be realized in a single channel of stimulation. If so, one channel of CA stimulation could be used in a hybrid processor to good effect, while leaving the remaining channels available for IP stimuli. The channel for CA stimulation could be selected on the basis of the tonotopic position corresponding to the range of F1 (i.e., the apical-most channel for present prostheses) or on the basis of some desirable combination of low threshold, wide dynamic range, and measured access to waveform details (such as discrimination of frequency changes in the F1 range). If the channel used for the CA stimulus has a relatively low threshold, the presence of the continuous CA signal might not interfere too much with the IP stimuli. That is, the advantage of having a CA stimulus may outweigh the increased level of channel interactions that it introduces. The IP stimuli could be used to code either the entire spectrum of speech or the spectrum above F1. In the first case frequency components in the F1 range would be redundantly coded by CA and IP stimuli, and the region above F1 would be coded with a relatively coarse resolution using the IP stimuli. In the second case the redundancy of coding in the F1 range would be traded for a higher resolution of coding (with IP stimuli) for frequency components above the F1 range. One or more of these possible implementations of hybrid processors may produce results that are superior those obtained with either parent processor. The hybrid processor might be most useful for patients who (a) can discriminate frequency changes throughout a substantial portion of the F1 range and (b) have low thresholds of stimulation for at least one channel.

Although the hybrid processor may provide significant benefits for certain patients, we believe the number of such patients is likely to be small. First, the lack of waveform detail in the IP processor will be of little or no consequence to the many patients who cannot discriminate frequency changes above 300 Hz [33], [52]. Also, patients with high thresholds for all channels, or with high levels of measured interactions among channels, probably will obtain superior results with the IP processor [51], [52]. Finally, when four or more channels are available the IP processor may be superior for all patients because (a) the resolution of frequency representation in the F1 range will approximate or exceed the resolution obtained under optimal conditions with CA stimuli, (b) the resolution of frequency representation above the F1 range is likely to be much better than that obtained with single or multichannel CA processors, and (c) the additional advantages of IP processors, such as explicit coding of voicing information and the use of low-frequency stimuli for the basal channels, will be retained as the number of channels is increased.

E. Conclusions

Major conclusions from the studies reviewed in this report are the following:

1. For the present subjects, each of whom had 2 - 4 channels of intracochlear stimulation with the UCSF/Storz cochlear prosthesis, consonant identification is superior with an interleaved pulses (IP) processor and vowel identification is superior with a compressed analog (CA) processor.

2. Strengths of the IP processor for consonant identification include better representations of voicing, nasality, envelope, frication and duration, and strengths of the CA processor for vowel identification include better representations of F1 and duration.

3. These strengths might be combined in a hybrid CA/IP processor that could be superior to either parent processor for some subjects (e.g., subjects with 2 - 4 channels of stimulation and with a low threshold for at least one of these channels).

4. Even in the face of considerable experience with the CA processor, five of the six subjects immediately had similar (4 subjects) or superior (1 subject) performance on tests of open-set recognition using the IP processor. The subject with the superior performance using the IP processor had the best match to the fitting criteria established for that processor, and the subject with the superior performance using the CA processor had the worst match. The ability to match IP processor parameters to the fitting criteria was limited in the present study by the number of functional channels and by the properties of the UCSF/Storz transcutaneous transmission system.

5. These findings from the open-set tests indicate that different implant patients generally require different processors for optimal performance. This conclusion in turn suggests that improvements across a population of patients can be made by (a) selecting the best type of processor for each patient and (b) using implanted and external hardware capable of supporting a wide range of processor options.

F. Acknowledgments

We thank the patients who participated in the described studies for their dedicated effort and pioneering spirit. We are pleased to acknowledge the important scientific contributions of FT Hambrecht, DK Kessler, GE Loeb, MM Merzenich, SD Soli and RD Wolford. This work was supported by NIH contract N01-NS-5-2396, through the Neural Prosthesis Program.

G. Appendix 1

Matrices for the consonant and vowel tests of this study are presented in Tables A1.1 - A1.3. The stimuli for each matrix are indicated along the left margin and the responses for each matrix are indicated along the top margin. Matrix values for the RTI tests (Tables A1.2 and A1.3) are given as the fraction of response for each cell, as described in the Methods section.

H. Appendix 2

Assignments of articulatory and acoustic features for the consonants and vowels of the Iowa and RTI tests are shown in Table A2.1. For the two consonant tests, the assignments for voicing, nasality, duration and frication are the same as those used by Miller and Nicely [23] and by Singh and Black [41]. The assignments for place of articulation are identical to those of Singh and Black [41]. Envelope cues are assigned according to the major groupings described by Van Tasell and coworkers [45]. Finally, the visual cues are assigned according to previous findings on viseme groupings for these particular consonants [28] and the patterns demonstrated in the vision only matrices of this study.

The assignments for the vowel features of first and second formant frequencies and of duration reflect the results from direct measurement of these acoustic parameters for the tokens in the RTI vowel test. The viseme groupings are those of Jeffers and Barley [18].

Table A1.1. Confusion matrices for Iowa consonant test.

a) Vision only

	p	b	m	f	v	ʃ	dʒ	s	z	t	d	n	g	k
p	10	15	5											
b	8	13	8								1			
m	4	13	12								1			
f				18	3		1	2	1	3	1			1
v				18	6					1	3		1	
ʃ			1			7	10	7	3				2	
dʒ				1		9	14	4	1				1	
s		1			1			12	10	2	3			1
z						3	2	7	12	1	4	1		
t					1	2	3	6	2	8	6	1		1
d	1							6	3	4	11	2		3
n				1	1			1	2	5	10	7	2	
g	1						1	3	2	4	10	3	4	2
k								1	1	5	8	7	3	5

Table A1.1, continued.

b) CA + Vision

	p	b	m	f	v	ʃ	dʒ	s	z	t	d	n	g	k
p	28	2												
b	13	16	1											
m	8	11	11											
f				18		1				9		1	1	
v				5	18			3	2	1	1			
ʃ						15		9	1	4				1
dʒ						4	19	1	3	1			2	
s						4		21	2	2				1
z						1	4	3	16	2	2		2	
t								4	2	21	1			2
d							2		3		18	2	4	1
n						1					13	8	8	
g			1			1	1		1	1	19	3	2	1
k								3		6	1		2	18

Table A1.1, continued

c) IP + Vision

	p	b	m	f	v	ʃ	dʒ	s	z	t	d	n	g	k
p	29	1												
b	1	17	11									1		
m			30											
f		1		8	4	4	1	1		8	3			
v		1		2	19				4	1	3			
ʃ						20	3	5	2					
dʒ						1	22		5		2			
s						7		17	1	1		1	1	2
z						1	2		25		1	1		
t										20				10
d		1							2		18	7	2	
n				1	1				2		7	19		
g									2		20	6	2	
k										4	1		2	23

Table A1.2. Confusion matrices for the RTI consonant test.

a) Vision only

	ð	s	z	t	d	n	k	l
ð	1.000							
s		.398	.417	.130	.056			
z		.519	.333	.019	.083	.019	.028	
t		.269	.241	.278	.194	.019		
d	.019	.213	.444	.222	.102			
n	.056	.056		.074	.583	.111		.120
k					.019	.185	.194	.602
l						.056		.944

b) CA + Vision

	ð	s	z	t	d	n	k	l
ð	.981		.019					
s	.037	.463	.222	.194	.028		.056	
z	.019	.269	.528	.083	.037	.019	.028	
t	.028		.046	.731	.056		.139	
d		.102	.194	.056	.435	.185		
n	.056		.046	.056	.324	.454		.065
k						.028	.944	.028
l								1.000

Table A1.2, continued

c) CA only

	ð	s	z	t	d	n	k	l
ð	.426	.083	.028	.120	.204		.139	
s	.130	.278	.037	.463			.093	
z	.157	.204	.259	.019	.111	.130	.083	.037
t	.074		.056	.500		.019	.352	
d		.111	.213		.250	.278	.148	
n	.074	.056	.111		.222	.463	.056	
k				.019			.981	
l					.019			.981

d) IP + Vision

	ð	s	z	t	d	n	k	l
ð	.981	.019						
s		.544	.281	.137	.019		.019	
z		.248	.722	.019	.011			
t		.019		.817	.106	.022	.037	
d		.074	.037	.074	.631	.183		
n	.028		.019		.163	.735		.056
k				.028			.954	.019
l						.011		.989

Table A1.2, continued

e) IP only

	ð	s	z	t	d	n	k	l
ð	.417	.218	.102	.069	.120	.037	.009	.028
s	.093	.620	.111	.148	.009		.019	
z	.250	.074	.403	.060	.065	.120		.028
t	.060	.037		.690	.083	.056	.074	
d	.037	.019	.019	.028	.699	.181	.019	
n	.056	.042			.204	.625		.074
k				.171			.829	
l						.019		.981

Table A1.3. Confusion matrices for the RTI vowel test.

a) Vision only

	i	I	ɔ	o	u
i	.661	.339			
I	.261	.739			
ɔ			1.000		
o				.967	.033
u				.194	.806

b) CA + Vision

	i	I	ɔ	o	u
i	.872	.128			
I	.050	.950			
ɔ			1.000		
o			.028	.933	.039
u				.056	.944

c) CA only

	i	I	ɔ	o	u
i	.900	.083			.017
I	.128	.883			.011
ɔ			.933	.067	
o			.144	.706	.150
u		.011		.089	.878

d) IP + Vision

	i	I	ɔ	o	u
i	.828	.172			
I	.217	.783			
ɔ			1.000		
o			.033	.950	.017
u				.194	.806

e) IP only

	i	I	ɔ	o	u
i	.822	.094	.028	.011	.044
I	.189	.744		.033	.033
ɔ			.972	.028	
o			.039	.861	.100
u				.233	.767

Table A2.1. Classification of phonemes used in the Iowa and RTI tests.

a) Iowa test

	p	b	m	f	v	ʃ	dʒ	s	z	t	d	n	g	k
Voicing	0	1	1	0	1	0	1	0	1	0	1	1	1	0
Nasality	0	0	1	0	0	0	0	0	0	0	0	1	0	0
Place	0	0	0	0	0	1	1	2	2	2	2	2	3	3
Duration	0	0	0	0	0	1	0	1	1	0	0	0	0	0
Friction	0	0	0	1	1	1	1	1	1	0	0	0	0	0
Envelope	0	1	2	3	1	3	1	3	1	0	1	2	1	0
Viseme	0	0	0	1	1	2	2	3	3	3	3	3	3	3

b) RTI tests

	ð	s	z	t	d	n	k	l
Voicing	1	0	1	0	1	1	0	1
Nasality	0	0	0	0	0	1	0	0
Place	0	0	0	0	0	0	1	0
Duration	0	1	1	0	0	0	0	0
Friction	1	1	1	0	0	0	0	0
Envelope	0	1	0	2	0	3	2	3
Viseme	0	1	1	1	1	1	2	2

	i	I	ɔ	o	u
F1	0	1	1	1	0
F2	1	1	0	0	0
Duration	1	0	2	2	1
Viseme	0	0	1	2	2

I. References

[1] Brimacombe JA, Webb RL, Dowell RC, Mecklenburg DJ, Beiter AL, Barker MJ, Clark GM (1988) Speech recognition abilities in profoundly deafened adults using the Nucleus 22 channel cochlear implant system. In: Banfai P (ed) Cochlear implant: Current situation. Erkelenz, West Germany: Rudolf Bermann GmbH, pp. 487-490

[2] Clark GM, Blamey PJ, Brown AM, Gusby PA, Dowell RC, Franz BK-H, Pyman BC, Shepherd RK, Tong YC, Webb RL, Hirshorn MS, Kuzma J, Mecklenburg DJ, Money DK, Patrick JF, Seligman PM (1987) The University of Melbourne-Nucleus multi-electrode cochlear implant. Basel: Karger.

[3] Danley MJ, Fretz RJ (1982) Design and functioning of the single-electrode cochlear implant. Ann. Otol. Rhinol. Laryngol. 91, Suppl. 91: 21-26

[4] De Filippo CL, Scott BL (1978) A method for training and evaluating the reception of ongoing speech. J. Acoust. Soc. Am. 63: 1186-1192

[5] Denes PB (1963) On the statistics of spoken English. J. Acoust. Soc. Am. 35: 892-904

[6] Dobie RA, Dillier N (1985) Some aspects of temporal coding for single-channel electrical stimulation of the cochlea. Hearing Res. 18: 41-55

[7] Dowell RC, Seligman PM, Blamey PJ, Clark GM (1987) Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. Ann. Otol. Rhinol. Laryngol. 96, Suppl. 128: 132-134

[8] Eddington DK (1983) Speech recognition in deaf subjects with multichannel intracochlear electrodes. Ann. N.Y. Acad. Sci. 405: 241-258

[9] Finley CC, Wilson BS, Lawson DT (1987) Speech processors for auditory prostheses. Seventh Quarterly Progress Report, NIH project N01-NS-5-2396. Bethesda, MD: Neural Prosthesis Program, National Institutes of Health

[10] Gantz BJ (1987) Cochlear implants: An overview. Adv. Otolaryngol. Head & Neck Surg. 1: 171-200

[11] Gantz BJ, Tyler RS, Knutson JF, Woodworth G, Abbas P, McCabe BF, Hinrichs J, Tye-Murray N, Lansing C, Kuk F, Brown C (1988) Evaluation of five different cochlear implant designs: Audiologic assessment and predictors of performance. Laryngoscope 98: 1100-1106

[12] Gardi JN (1985) Human brain stem and middle latency responses to electrical stimulation: Preliminary observations. In: Schindler RA, Merzenich MM (eds) Cochlear implants. New York: Raven Press, pp. 351-363

[13] Grether CB (1970) Psychoacoustic assessment of speech communication systems: The diagnostic discrimination test. Project Themis. Air Force Office of Scientific Research, contract F44620-69-C-0033

[14] Hochmair ES, Hochmair-Desoyer IJ (1985) Aspects of sound signal processing using the Vienna intra- and extracochlear implants. In: Schindler RA, Merzenich MM (eds) Cochlear implants. New York: Raven Press, pp. 101-110

[15] Hochmair-Desoyer IJ, Burian K (1985) Reimplantation of a molded scala tympani electrode: Impact on psychophysical and speech discrimination abilities. Ann. Otol. Rhinol. Laryngol. 94: 65-70

[16] Hochmair-Desoyer IJ, Hochmair ES, Stiglbrunner HK (1985) Psychoacoustic temporal processing and speech understanding in cochlear implant patients. In: Schindler RA, Merzenich MM (eds) Cochlear implants. New York: Raven Press, pp. 291-304

[17] Javel E (1989) Acoustic and electrical encoding of temporal information. In: Miller JM, Spelman FA (eds) Models of the electrically stimulated cochlea, New York: Springer-Verlag, to be published

[18] Jeffers J, Barley M (1971) Speechreading. Springfield, Ill.: Charles C. Thomas

[19] Loeb GE, Byers CL, Rebscher SJ, Casey DE, Fong MM, Schindler RA, Gray RF, Merzenich MM (1983) Design and fabrication of an experimental cochlear prosthesis. Med. & Biol. Eng. & Comput. 21: 241-254

[20] Merzenich MM (1985) UCSF cochlear implant device. In: Schindler RA, Merzenich MM (eds) Cochlear implants. New York: Raven Press, pp. 121-129

[21] Merzenich MM, Leake-Jones P, Vivion M, White M, Silverman M (1978) Development of multichannel electrodes for an auditory prosthesis. Fourth Quarterly Progress Report, NIH project N01-NS-7-2367. Bethesda, MD: Neural Prosthesis Program, National Institutes of Health

[22] Millar JB, Tong YC, Clark GM (1984) Speech processing for cochlear implant prostheses. J. Speech & Hear. Res. 27: 280-296

[23] Miller GA, Nicely PE (1955) An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27: 338-352

[24] Minifie FD (1973) Speech acoustics. In: Minifie FD, Hixon TJ, Williams F (eds) Normal aspects of speech, hearing and language. Englewood Cliffs, NJ: Prentice-Hall, pp. 235-284

[25] Moore BCJ (1985) Speech coding for cochlear implants. In: Grey RF (ed) Cochlear implants. San Diego, CA: College-Hill Press, pp. 163-179

[26] Moxon EC (1971) Neural and mechanical responses to electric stimulation of the cat's inner ear. Doctoral dissertation, MIT, Cambridge, MA

[27] Ochs MT, White MW, Merzenich MM, Schubert ED (1985) Speech recognition in single- and multichannel cochlear implants. J. Acoust. Soc. Am. 77 (Suppl. 1): S81

[28] Owens E, Blazek B (1985) Visemes observed by hearing-impaired and normal-hearing adult viewers. J. Speech & Hear. Res. 28: 381-393

[29] Owens E, Kessler DK, Raggio M, Schubert ED (1985) Analysis and revision of the Minimal Auditory Capabilities (MAC) battery. Ear Hear. 6: 280-287

[30] Owens E, Raggio M (1987) The UCSF tracking procedure for evaluation and training of speech reception by hearing-impaired adults. J. Speech & Hear. Disorders 52: 120-128

[31] Parkins CW (1986) Cochlear prostheses. In: Altschuler RA, Hoffman DW, Bobbin RP (eds) Neurobiology of hearing: The cochlea. New York: Raven Press, pp. 455-473

[32] Peterson GE, Barney HL (1952) Control methods used in a study of vowels. J. Acoust. Soc. Am. 24: 175-184

[33] Pfingst BE (1985) Psychophysical data from cochlear implants: Relevance to strategies for rehabilitation. Seminars in Hearing 6: 7-21

[34] Pfingst BE (1986) Stimulation and encoding strategies for cochlear prostheses. Otolaryngol. Clin. N. Amer. 19: 219-235

[35] Pfingst BE, Sutton D (1983) Relation of cochlear implant function to histopathology in monkeys. Ann. N.Y. Acad. Sci. 405: 224-239

[36] Schindler RA, Kessler DK (1987) The UCSF/Storz cochlear implant: Patient performance. Am. J. Otol. 8: 247-255

[37] Schindler RA, Kessler DK, Rebscher SJ, Jackler RK, Merzenich MM (1987) Surgical considerations and hearing results with the UCSF/Storz cochlear implant. Laryngoscope 97: 50-56

[38] Schindler RA, Kessler DK, Rebscher SJ, Yanda JL, Jackler RK (1986) The UCSF/Storz multichannel cochlear implant: Patient results. Laryngoscope 96: 597-603

[39] Schubert ED (1985) Some limitations on speech coding for implants. In: Schindler RA, Merzenich MM (eds) Cochlear implants. New York: Raven Press, pp. 269-276

[40] Shannon RV (1983) Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics. Hearing Res. 11: 157-189

[41] Singh S, Black JW (1966) Study of twenty-six intervocalic consonants as spoken by four language groups. J. Acoust. Soc. Am. 39: 372-387

[42] Tyler RS, Preece JP, Lansing CR, Otto SR, Gantz BJ (1986) Previous experience as a confounding factor in comparing cochlear-implant processing schemes. J. Speech & Hear. Res. 29: 282-287

[43] Tyler RS, Preece JP, Lowder MW (1983) The Iowa cochlear-implant test battery. Laboratory Report, University of Iowa at Iowa City, Department of Otolaryngology - Head and Neck Surgery

[44] van den Honert C, Stypulkowski PH (1987) Temporal response patterns of single auditory nerve fibers elicited by periodic electrical stimuli. Hearing Res. 29: 207-222

[45] Van Tasell DJ, Soli SD, Kirby VM, Widin GP (1987) Speech waveform envelope cues for consonant recognition. J. Acoust. Soc. Am. 82: 1152-1161

[46] Wang MD, Bilger RC (1973) Consonant confusions in noise: A study of perceptual features. J. Acoust. Soc. Am. 54: 1248-1266

[47] White M, Ochs M, Raggio M, Morledge D, Merzenich M (1985) Formant discrimination in a multichannel cochlear prosthesis. J. Acoust. Soc. Am. 77 (Suppl. 1): S81

[48] White MW (1983) Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes. Ann. N.Y. Acad. Sci. 405: 348-359

[49] White MW, Merzenich MM, Gardi JN (1984) Multichannel cochlear implants: Channel interactions and processor design. Arch. Otolaryngol. 110: 493-501

[50] Wilson BS, Finley CC (1985) A computer-based simulator of speech processors for auditory prostheses. ARO Abstracts, 8th Midwinter Res. Conf., St. Petersburg, FL, p. 209

[51] Wilson BS, Finley CC, Farmer JC Jr, Lawson DT, Weber BA, Wolford RD, Kenan PD, White MW, Merzenich MM, Schindler RA (1988a) Comparative studies of speech processing strategies for cochlear implants. Laryngoscope 98: 1069-1077

[52] Wilson BS, Finley CC, Lawson DT, Wolford RD (1988b) Speech processors for cochlear prostheses. Proc. IEEE 76: 1143-1154

[53] Wilson BS, Finley CC, Lawson DT, Wolford RD (1988c) Direct comparisons of analog and pulsatile coding strategies with six cochlear implant patients. In prep.

[54] Wilson BS, Schindler RA, Finley CC, Kessler DK, Lawson DT, Wolford RD (1988d) Present status and future enhancements of the UCSF cochlear prosthesis. In: Banfai P (ed) Cochlear implant: Current situation. Erkelenz, West Germany: Rudolf Bermann GmbH, pp. 395-427

III. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Present project results at the 25th Anniversary Symposium of the Kresge Hearing Research Institute (Oct. 3-5) and at the Nineteenth Neural Prosthesis Workshop (Oct. 26-28);

2. Conduct additional studies in follow-up visits with two patients (MC1 and MC2) implanted with the UCSF/Storz cochlear prosthesis (these studies are scheduled for the the first two weeks of November);

3. Continue work to upgrade the Cochlear Implant Laboratory at Duke, with emphasis on software conversion for the new 80386 machine and on development of real-time signal processing code for the TMS320C25 system;

4. Continue preparation for studies with patients implanted with the Nucleus cochlear prosthesis, with emphasis on evaluation of alternative systems for external control of the subcutaneous receiver for that prosthesis (the alternative systems include one developed by Bob Shannon, one developed by Norbert Dillier, and a TMS320-based system developed by Charlie Finley of our RTI group); and

5. Continue preparation of manuscripts for publication.

Appendix 1

Summary of Reporting Activity for the Period of
June 27 through September 26, 1988
NIH Contract N01-NS-5-2396

Four papers and one abstract describing work on this project were published in the present reporting period. The citations are:

Wilson, B.S., Finley, C.C., Lawson, D.T., and Wolford, R.D.: Speech processors for cochlear prostheses. Proc. IEEE, vol. 76, 1143-1154, 1988.

Wilson, B.S., Finley, C.C., Farmer, J.C., Jr., Lawson, D.T., Weber, B.A., Wolford, R.D., Kenan, P.D., White, M.W., Merzenich, M.M., and Schindler, R.A.: Comparative studies of speech processing strategies for cochlear implants. Laryngoscope, vol. 98, 1069-1077, 1988.

Wilson, B.S., Schindler, R.A., Finley, C.C., Kessler, D.K., Lawson, D.T., and Wolford, R.D.: Present status and future enhancements of the UCSF cochlear prosthesis. In: P. Banfai (Ed.), Cochlear Implant: Current Situation, Rudolf Bermann GmbH, Erkelenz, West Germany, 1988, pp. 395-427.

Wilson, B.S. (moderator), Dent, L.J., Dillier, N., Eddington, D.K., Hochmair-Desoyer, I.J., Patrick, J., Pfingst, B.E., Sürth, W., and Walliker, J. (panelists): Round table discussion on speech coding. In: P. Banfai (Ed.), Cochlear Implant: Current Situation, Rudolf Bermann GmbH, Erkelenz, West Germany, 1988, pp. 693-704.

Soli, S.D., and Wilson, B.S.: Within-subject comparisons of analog and pulsatile speech processors for cochlear implants. J. Acoust. Soc. Am., vol. 84 (Suppl. 1), S41, 1988.

In addition to the publications, the following presentations were made in the present reporting period:

Finley, C.C.: Design considerations for auditory prostheses. Invited paper presented at the World Congress on Medical Physics and Biomedical Engineering, San Antonio, TX, Aug. 6-12, 1988.

Finley, C.C.: Co-Chairman, Session on Auditory System Research. World Congress on Medical Physics and Biomedical Engineering, San Antonio, TX, Aug. 6-12, 1988.