original X
sent to printing
5/11/92

Eleventh Quarterly Progress Report

November 1, 1991 through January 31, 1992

NIH Contract N01-DC-9-2401

Speech Processors for Auditory Prostheses

Prepared by

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

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

CONTENTS

I. Introduction 3

II. Efficacy of CIS Processors for Patients with Poor Clinical Outcomes 4

III. Plans for the Next Quarter 14

Appendix 1: Summary of Reporting Activity for this Quarter 16

Appendix 2: Tables of Feature Assignments for 16 Consonants and 8 Vowels 18

I. Introduction

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

1. Continued studies with Ineraid patients who have poor outcomes with their clinical devices. The studies have included comparisons of the clinical *compressed analog* (CA) processors with various implementations of *continuous interleaved sampling* (CIS) processors.
2. A detailed review of software and procedures for future psychophysical studies, including studies of channel ranking, frequency discrimination, and frequency scaling. Consultant Robert Shannon visited RTI on December 9 and 10 to participate in the review and to offer his expert advice on testing issues.
3. Continued analysis of data from all subjects in our Ineraid series, to evaluate effects of single parameter changes on the performance of CIS processors.
4. Initial studies with a patient implanted with the new MiniMed device.
5. Preparation and submission of a manuscript on "Importance of Patient and Processor Variables in Determining Outcomes with Cochlear Implants."
6. Preparation and submission of an invited manuscript on "New Processing Strategies for Multichannel Cochlear Prostheses."
7. Continued preparation of other manuscripts for publication.

In this report we present interim results from our present Ineraid series, with subjects selected for their relatively poor performances with CA processors. Work related to points 2, 3 and 4 will be described in future reports.

II. Efficacy of CIS Processors for Patients with Poor Clinical Outcomes

Studies to compare the CA and CIS strategies have been conducted with eleven subjects implanted with the Ineraid electrode array. Seven of these subjects (SR2-8) enjoyed high levels of speech recognition with their clinical CA processors, having been selected as representative of patients with the best results using this or any other clinical implant system [Wilson et al., 1991]. The remaining subjects were chosen for their relatively poor performances with the clinical processors (SR1 and SR9-11).

Subjects

Near the end of studies with the latest subject in the "poor performance" group we discovered that the alignment pin of the plug mating the cable from our stimulation hardware to the patient's percutaneous connector had been pushed into its case. The tip of the alignment pin was flush with the mating surface of the plug, allowing connection to be made when the plug was rotated 180° with respect to its correct orientation. This incorrect orientation had been used consistently by this subject (SR11) in all testing before we identified the problem with the plug. All results obtained under that condition are invalid in that the electrode assignments and coupling configurations were quite different from the intended assignments and monopolar coupling.

When the correct orientation was established, by turning the plug 180° prior to insertion, measured electrode impedances were consistent with independent measures made at the referring clinic in Salt Lake City (data kindly supplied by Michael Dorman at a later date) and also were consistent with the typical pattern of relatively low impedances for the four electrodes used with the clinical CA processor and relatively high impedances for the two electrodes not used with the clinical processor. In addition, a clear tonotopic ranking of electrodes was observed with the correct orientation, in contrast to the complex and anomalous ranking observed with the incorrect orientation.

In retrospect these signatures of incorrect orientation are obvious; however, at the time there was no reason to suspect a problem with the plug. We accepted the unusual results as perhaps characteristic of a subject in the poor-performance group, perhaps with an unusual pattern of nerve survival in the implanted cochlea.

Following our experience with SR11 we reviewed the cases of all previous subjects studied in our Ineraid series. All but one of the subjects had (a) a normal pattern of impedances across electrodes, (b) normal ranking of electrodes in a tonotopic order, and (c) large increases in speech test scores with use of a CIS processor. The exceptional subject, SR9, had an impedance pattern just like the one seen for SR11 with her plug in the incorrect orientation. In addition, SR9's ranking of percepts by electrode showed anomalous departures from a tonotopic order. As with SR11, impedance measures made at the referring clinic in Salt Lake City demonstrated a normal pattern for SR9 (data again kindly supplied by Michael Dorman).

Like SR11 with the incorrect plug orientation, SR9's improvements on speech tests with a CIS processor were markedly less than those for all other patients in the series thus far.

These findings indicate with high probability that (a) the alignment pin was pushed in prior to, or at the beginning of, SR9's visit, (b) subject SR9 consistently used the incorrect orientation, and (c) even though the pin was pushed in, subject SR10 happened to use the correct orientation consistently. The correct orientation was used by SR11 during the final afternoon and morning sessions of her week-long visit. In those final hours we were able to test only one implementation of a CIS processor, bypassing our standard procedures for optimizing CIS parameters for each subject (see below). Use of this first processor produced large and immediate improvements in open-set recognition, consistent with the improvements observed for other subjects. We would expect further gains in speech test scores with optimization of the CIS processor. We plan to complete a normal week of studies with SR9 and SR11 using a replacement for the damaged plug (SR11 already has accepted our invitation to return the first week of June, 1992).

Tests

As outlined above, valid measures from a full week of testing already have been obtained in studies with two of the four subjects in our low-performance series (SR1 and SR10). Results for SR1 have been presented in QPR 5 for this project, but are repeated here for completeness. Results for SR10 have not been presented in prior reports.

In each set of comparisons the clinical processor was used for evaluation of the CA strategy while a laboratory processor was used for evaluation of the CIS strategy. Both subjects were studied for a one-week period in which (a) basic psychophysical measures were obtained on thresholds and dynamic ranges for pulsatile stimuli, (b) a variety of CIS processors (with different choices of processor parameters) were evaluated with preliminary tests of consonant identification, and (c) performance with the best of the CIS processors and the clinical CA processor was documented with a broad spectrum of speech tests. Experience with the clinical processor exceeded one year of daily use for both subjects. In contrast, experience with the CIS processors was limited to no more than several hours before formal testing.

The comparison tests included identification of 16 consonants (/b, d, f, g, dʒ, k, l, m, n, p, s, ʃ, t, ð, v, z/) in an /a/-consonant-/a/ context; identification of 8 vowels (/i, ɔ, ɛ, u, I, U, ʌ, æ/) in a /h/-vowel-/d/ context; and the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery [Owens et al., 1985].

In both the consonant and vowel tests multiple exemplars of the tokens were played from laser videodisc recordings of male and female speakers [Tyler et al., 1987; Lawson et al., 1989]. A single block of trials consisted of five randomized presentations of each consonant or three randomized presentations of each vowel for one of the speakers. At least two blocks were administered for each speaker, processor and subject in the consonant tests, and at least three blocks were administered for each speaker, processor and subject in the vowel tests.

The segmental tests included identification of the word containing the correct vowel, initial consonant (Init Cons), or final consonant (Fnl Cons) among four options for each test item. The vowel test contained 60 items, the initial consonant test 64 items, and the final consonant test 52 items.

The open-set tests included recognition of 50 one-syllable words from Northwestern University Auditory Test 6 (NU-6), 25 two-syllable words (spondees), 100 key words in the Central Institute for the Deaf (CID) sentences of everyday speech, and the final word in each of 50 sentences from the Speech Perception in Noise (SPIN) test (here presented without noise). In both the segmental and open-set tests single presentations of the words or sentences were played from cassette tape recordings of a male speaker.

All tests were conducted with hearing alone, and without feedback as to correct or incorrect responses. Results for the tests of consonant and vowel identification were expressed as percent information transfer for various articulatory and acoustic features [Miller and Nicely, 1955], and results for the remaining tests were expressed as the percentage of correct responses. Tables of feature assignments for the 16 consonants and 8 vowels are presented for reference in Appendix 2.

Results

In this section we review results from the subjects in the "high performance" group (subjects SR2-8) and present individual scores for SR1 and SR10. Results from the high-performance subjects provide a context for interpretation of results from the two low-performance subjects.

Subjects SR2-8

Results for the subjects in the high performance group are presented in Figure 1. The top two panels show the information transfer (IT) scores for consonant and vowel features, and the bottom panels show average scores for the segmental and open-set tests of the MAC battery.

Large increases in IT scores for the consonant features of overall transmission (All), nasality (Nsl), and frication (Fric) are produced with use of the CIS processors. In addition, substantial increases are found for consonant duration (Dur), place of articulation (Plc), and envelope (Env) features. The scores for nasality, frication and envelope each exceed 82% when the CIS processors are used, and the scores for all remaining features except place (62.7%) exceed 72%.

Scores from IT analyses of the aggregate matrices for vowels are quite high for both processing strategies and for all features. IT scores are nearly identical for the vowel features of first formant (F1) and duration (Dur), and somewhat higher with the CA processors for overall transmission (All) and second formant (F2).

The averaged scores from the segmental tests are also all quite high and mirror, to some extent, the results from the tests of vowel and consonant identification. In particular, the scores for the vowel test are indistinguishable, while the scores for the final consonant test indicate superiority of the CIS processors ($p < 0.02$). The scores for the initial consonant test do not favor either processing strategy. However, ceiling effects may have masked a true difference between strategies for that test (absolute scores are greater than 92% for both strategies).

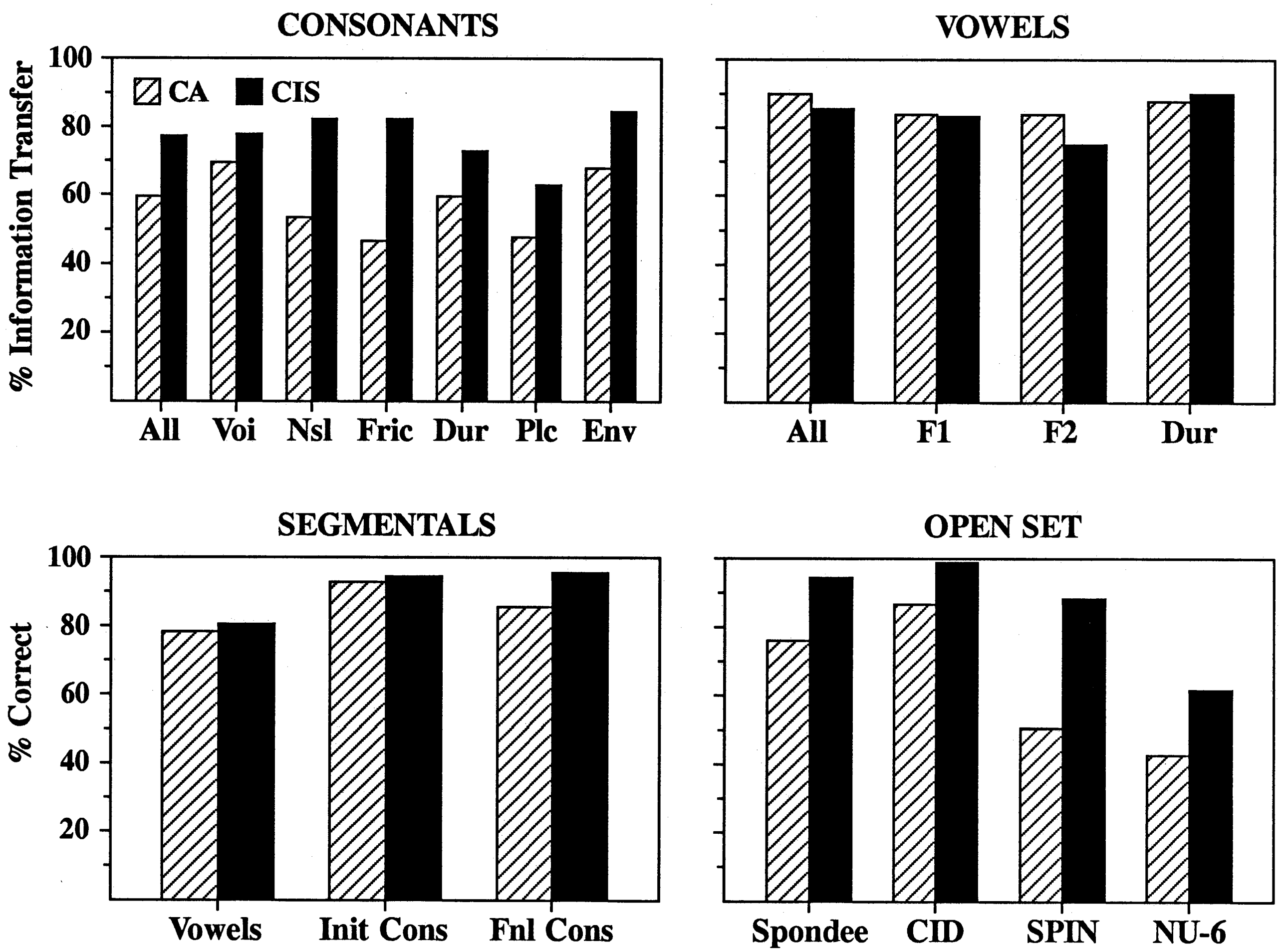

FIG. 1. Speech test results for the seven subjects in the "high performance" group. Scores for the CA processors are indicated by the striped bars, and those for the CIS processors by the solid bars. The top panels show relative information transfer for articulatory and acoustic features of consonants and vowels. The features for consonants include overall transmission (All), voicing (Voi), nasality (Nsl), frication (Fric), duration (Dur), place of articulation (Plc), and envelope cues (Env). The features for vowels include overall transmission (All), first formant frequency (F1), second formant frequency (F2), and duration (Dur). 205 presentations of each of 16 consonants were used in the consonant identification test with the CA processors, and 145 presentations were used for the CIS processors. For the vowel identification tests 132 presentations of each of 8 vowels were used for the CA processors and 126 presentations were used for the CIS processors. Presentations for both the consonant and vowel tests were balanced between male and female speakers for both processors. The bottom panels show scores from the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery. See text for abbreviations. Parameters of the CIS processors used by each of the subjects may be found in Table 1 of Quarterly Progress Report 4 for this project.

Finally, the open-set results indicate clear superiority of the CIS processors. Remarkable gains are found for all tests not limited by ceiling effects. Paired-t comparisons show that the increases across subjects are significant for the spondee, SPIN, and NU-6 tests.

Subject SR1

Results for the first subject in the "low performance" group are presented in Figure 2. The pattern of these results is remarkably similar to the pattern of results for the subjects with high levels of initial performance with their clinical CA processors (Figure 1). In particular, large gains are found in the transmission of consonant features and in the recognition of open-set material when the CIS processor is used instead of the CA processor. The increases in IT scores for subject SR1 mirror those for subjects SR2-8. Especially large increases in overall transmission and in the transmission of nasality, frication and envelope information are seen in both sets of data. However, the magnitudes of the increases are larger for subject SR1.

As with the prior subjects in the high-performance group, SR1's scores for the transmission of vowel features are similar for the two processors. The previous pattern, of slightly higher scores for overall transmission, F1 and F2 with the CA processors, and of a slightly higher score for the transmission of duration information with the CIS processors, is repeated in the results obtained with subject SR1. Why?

In addition, results for SR1 demonstrate substantial increases in the scores for the segmental tests of the MAC battery. As with the prior subjects, a large increase is found for the final consonant test. Unlike the prior subjects, though, large increases also are found for the vowel and initial consonant tests.

Finally, the open-set results for SR1 again demonstrate clear improvements with the CIS processor. A large gain in the recognition of spondee words is realized with the CIS processor, and scores for the CID, SPIN and NU-6 tests are more than doubled with the use of that processor.

The results for SR1 show that a patient starting with low levels of performance with one processing strategy may receive large benefits from substitution of another strategy. Indeed, with the CIS strategy SR1 has scores that fall within the ranges of those obtained by subjects SR2-8 with their CA processors. As noted before, such scores were among the best previously recorded for cochlear implant patients.

Subject SR10

Subject SR10 also enjoyed large gains in speech recognition with use of a CIS processor, as shown in Figure 3. IT scores for every feature of consonants are at least doubled with the CIS processor, as are the IT scores for the vowel features of F1 and F2. Further, scores for the MAC segmental tests of vowel identification (Vowels) and final consonant identification (Fnl Cons) also are doubled. The score for the initial consonant test (Init Cons), while not doubled, is greatly improved with the use of the CIS processor.

Perhaps the most striking results are those from the open-set tests. Performances on these tests with the clinical CA processor were quite poor. In contrast, use of the CIS processor immediately produced

Why don't vowels do as well CIS - How can this be improv

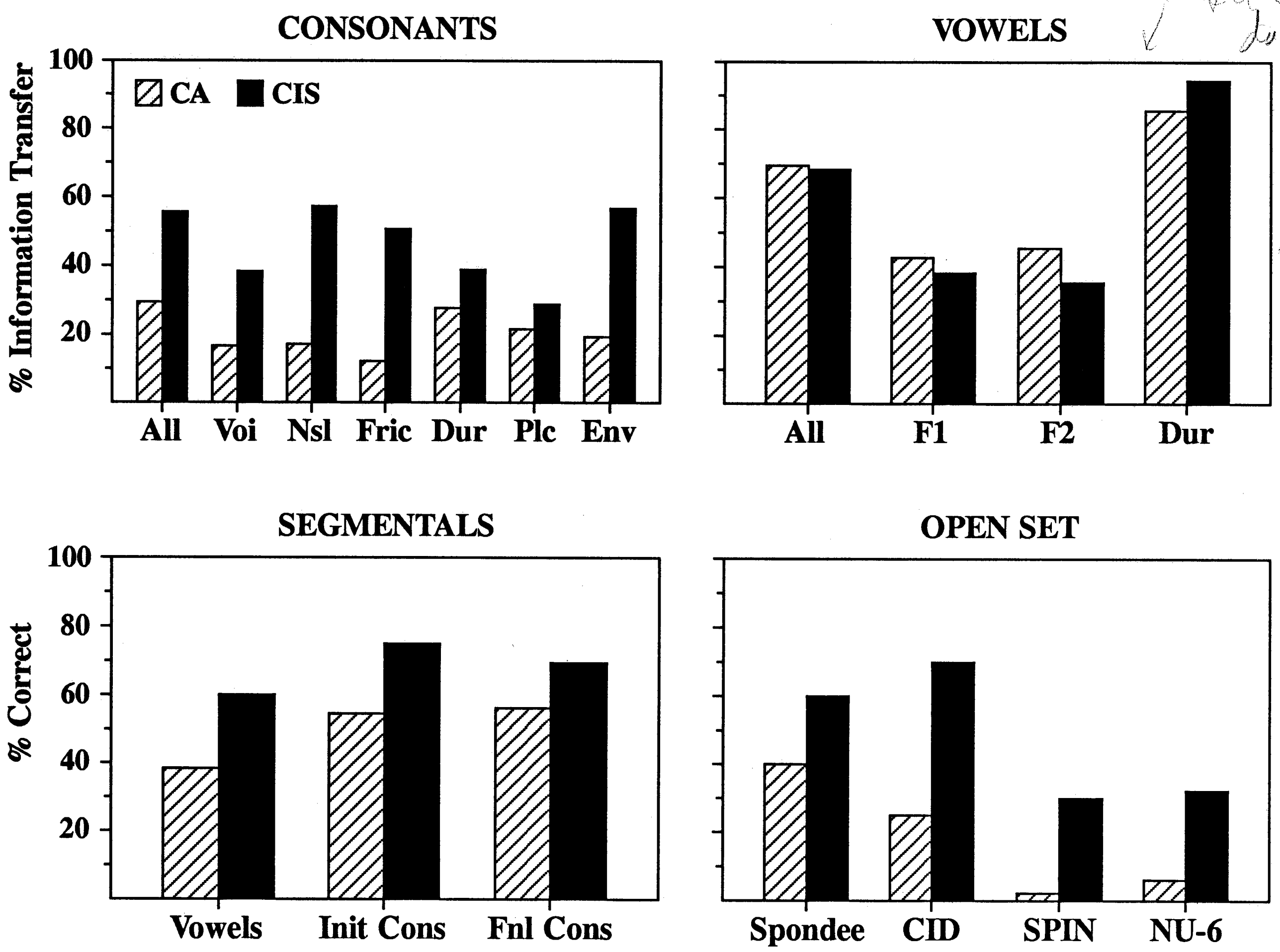

FIG. 2. Speech test results for subject SR1. Scores for the CA processor are indicated by the striped bars, and those for the CIS processor by the solid bars. Forty presentations of each of 16 consonants were used in the consonant identification test with the CA processor, and twenty presentations were used for the CIS processor. For the vowel identification tests 18 presentations of each of 8 vowels were used for both processors. Presentations for both the consonant and vowel tests were balanced between male and female speakers for both processors. The CIS processor used by SR1 (RB26m3) had the following parameters: pulse duration of 34 μs/phase, 5 channels (one electrode was not used because of its high thresholds), pulse rate of 833 pps on each channel, staggered order of channel updates (6-3-1-5-2), halfwave rectifiers in the envelope detectors, cutoff frequency of 400 Hz for the lowpass filters (1st order) in the envelope detectors, and a logarithmic mapping of envelope signals onto pulse amplitudes (our "mapping law 3").

relatively high levels of open-set recognition. The score on the spondee test increased from 0 to 56% correct, on the CID test from 1 to 55% correct, on the SPIN test from 0 to 26% correct, and on the

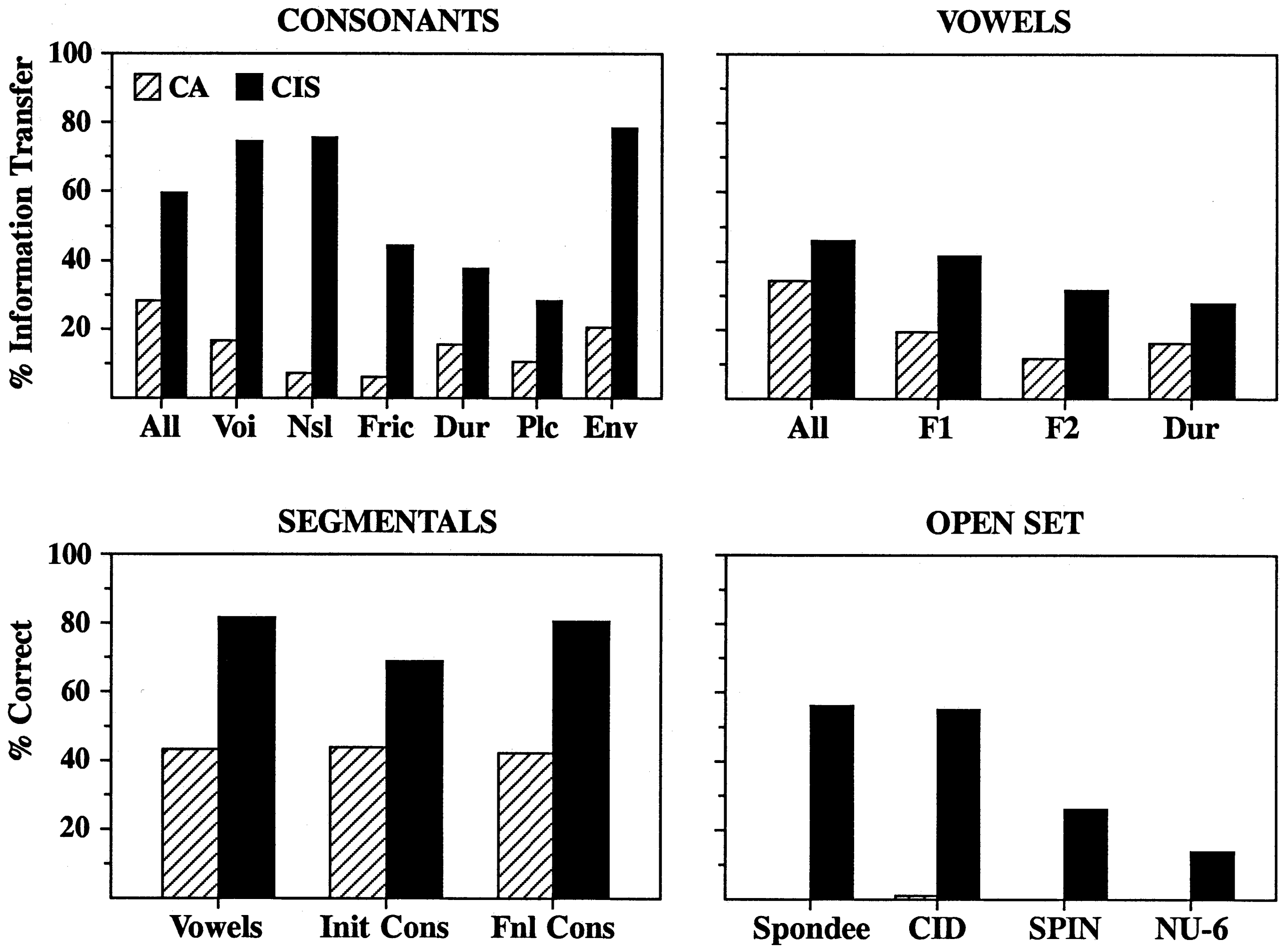

FIG. 3. Speech test results for subject SR10. Scores for the CA processor are indicated by the striped bars, and those for the CIS processor by the solid bars. Twenty-five presentations of each of 16 consonants were used in the consonant identification test with the CA processor, and twenty presentations were used for the CIS processor. For the vowel identification tests 18 presentations of each of 8 vowels were used for both processors. Presentations for both the consonant and vowel tests were balanced between male and female speakers for both processors. The CIS processor used by SR10 (MK14m4) had the following parameters: pulse duration of 167 μs/phase, 6 channels, pulse rate of 500 pps on each channel, staggered order of channel updates (6-3-5-2-4-1), fullwave rectifiers in the envelope detectors, cutoff frequency of 200 Hz for the lowpass filters (4th order) in the envelope detectors, and a power-law mapping of envelope signals onto pulse amplitudes (exponent of 0.2, our "mapping law 4").

NU-6 test from 0 to 14% correct. These increases were obtained with no more than several hours of aggregated experience with CIS processors, compared to more than a year of daily experience with the clinical CA processor.

Will this subject get a CIS process to take home

While these gains are large, they are not atypical of results from other subjects in our Ineraid series. Figure 4 shows the open-set scores for all nine subjects. The darker lines show results for the two subjects who started with relatively poor outcomes with their clinical CA processors. Results for SR10 are highlighted with circles at each endpoint. His improvements follow the pattern of other subjects, i.e., generally large gains in the scores of tests that are not limited by ceiling effects. The unique aspect of SR10's results is that he enjoys such gains even though he started at or near zero on all four tests. Thus, on a relative (ratio) scale, improvements for SR10 are larger than those for any other subject in the series thus far.

Discussion

An interesting aspect of the studies with SR1 and SR10 is that their best CIS processors used parameters distinct from those of the best processors for subjects in the high-performance group. The best processor for SR1 used short-duration pulses (33 μs/phase) presented at a relatively-low rate (833 pps), and the best processor for SR10 used long-duration pulses (167 μs/phase) presented at an even lower rate (500 pps). In contrast, many of the subjects in the high-performance group obtained their best scores with processors using short pulses and high rates (e.g., 33 μs/phase pulses presented at 2525 pps).

Use of shorter pulses and higher rates allows representation of higher frequencies in the modulation waveform for each channel, i.e., the cutoff frequency of the lowpass filter in the envelope detectors for each channel may be raised to 1/2 the pulse rate without introducing aliasing effects. In addition, the dynamic range (DR) of electrical stimulation -- from threshold to most comfortable loudness (MCL) -- is a strong function of pulse rate and a weaker function of pulse duration (see section VI of QPR 9 for this project). Large increases in DR generally are found with increases in pulse rates from about 400 pps to 2500 pps. Smaller increases often (but not always) are observed with increases in pulse duration (at a fixed rate of stimulation) from roughly 50 μs/phase to higher values (e.g., out to 200 μs/phase for practical CIS designs).

For some patients, however, these advantages may be outweighed by other factors. For several subjects in our Ineraid series, for instance, we have observed that the salience of channel ranking can decline with decreases in pulse widths below 100 μs/phase. A favorable tradeoff for such subjects might involve the use of long-duration pulses (e.g., 100 μs/phase or greater) to preserve channel cues, while sacrificing some of the potential DR, that could be obtained with shorter pulses and higher rates of stimulation.

Another possible advantage of relatively low rates of stimulation is reduction of channel interactions. Interposition of time between pulses on sequential channels can reduce the "temporal integration" component of channel interactions. Thus, use of time delays between short-duration pulses in the stimulation sequence across electrodes may reduce interactions. Alternatively, use of long-duration pulses with no time delay also might reduce temporal interactions in that a relatively long period still is interposed between the excitatory phases of successive pulses. The possible equivalence of such

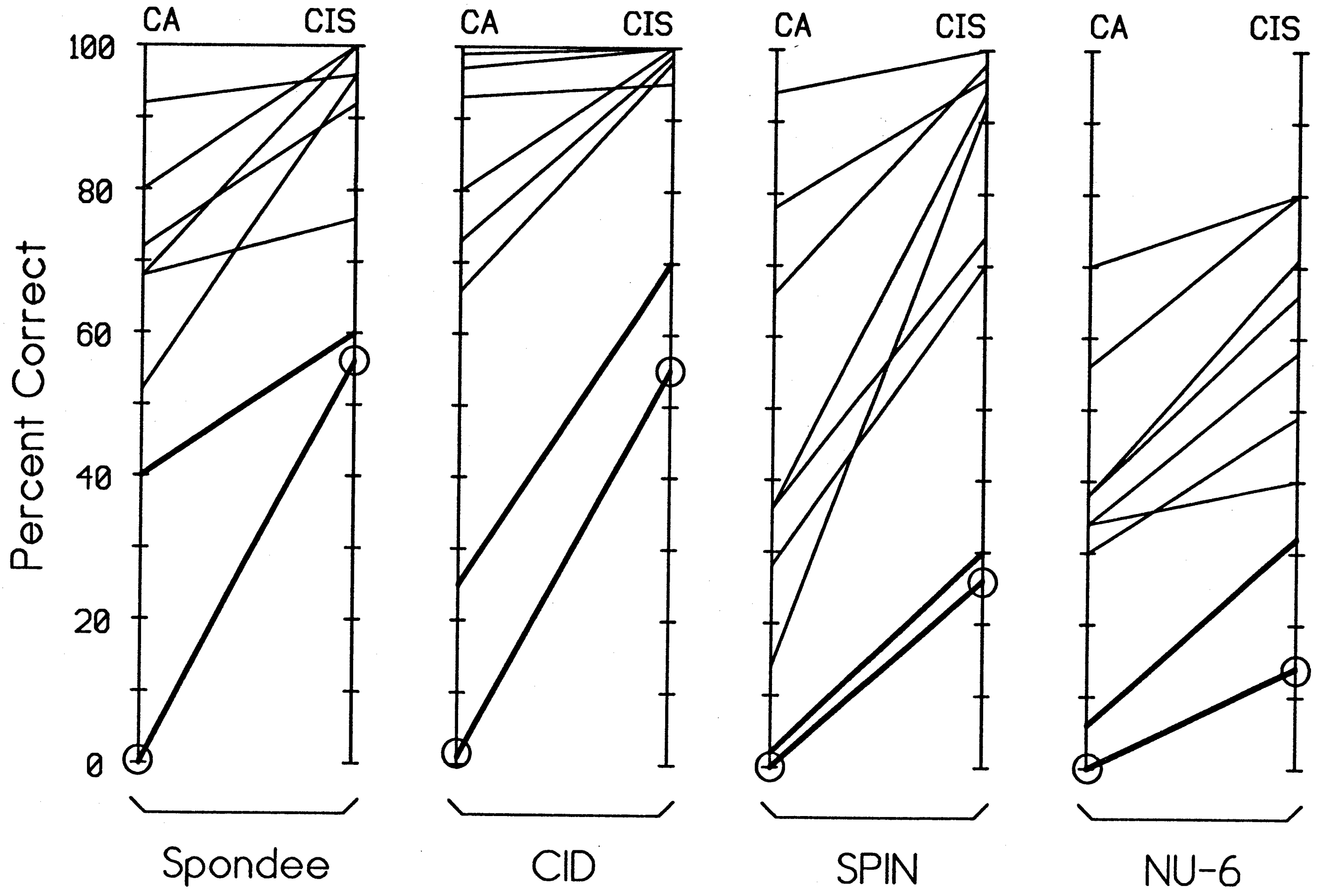

FIG. 4. Speech recognition scores for CA and CIS processors. The light lines show scores from tests with seven subjects who were selected for their excellent performances with the clinical CA processor, and the dark lines show scores from tests with two subjects who were selected for their relatively poor performances with that processor. Scores for subject SR10 are highlighted with circles.

alternatives will be explored in the planned further studies with SR9 and SR11.

Collectively the results presented in this report indicate that (a) the performance of at least some patients with poor clinical outcomes can be improved substantially with use of a CIS processor, (b) use of long-duration pulses produced a marked gain in speech test scores for one such subject, (c) use of short-duration pulses presented at a low rate produced similar improvements in another such subject, and (d) the optimal tradeoffs among pulse duration, pulse rate, interval between sequential pulses, and cutoff frequency of the lowpass filters appear to vary from patient to patient.

Acknowledgement

The use of pulse durations in excess of 100 μs/phase was suggested by Bob Shannon in a discussion with Blake Wilson several weeks before our studies with subject SR10. Bob reminded us of a prediction from one of our early models of ensemble neural responses to intracochlear electrical stimulation (see, e.g., Wilson et al., 1985; Wilson and Finley, 1986). The prediction is that the longitudinal spread of the neural excitation field is inversely related to pulse duration. That is, short-duration pulses produce a wider spread of excitation than long-duration pulses of the same charge. Thus, use of long-duration pulses may sharpen neural excitation fields and thereby increase channel independence and the salience of channel-related cues.

References

Lawson, D.T., Wilson, B.S., and Finley, C.C. (1989). Speech processors for auditory prostheses. Fourteenth Quarterly Progress Report, NIH project N01-NS-5-2396.

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

Owens, E., Kessler, D.K., and Raggio, M., and Schubert, E.D. (1985). Analysis and revision of the Minimal Auditory Capabilities (MAC) battery. Ear Hear. **6**, 280-287.

Tyler, R.S., Preece, J.P., and Lowder, M.W. (1987). The Iowa audiovisual speech perception laser videodisc. Laser Videodisc and Laboratory Report, Department of Otolaryngology -- Head and Neck Surgery, University of Iowa Hospital and Clinics, Iowa City, IA.

Wilson, B.S., and Finley, C.C. (1986). Latency fields in electrically evoked hearing. ARO Abstracts, 9th Midwinter Research Conference, pp. 170-171.

Wilson, B.S., Finley, C.C., and Lawson, D.T. (1985). Speech processors for auditory prostheses. Eighth Quarterly Progress Report, NIH project N01-NS-5-2356.

Wilson, B.S., Finley, C.C., Lawson, D.T., Wolford, R.D., Eddington, D.K., and Rabinowitz, W.M. (1991). Better speech recognition with cochlear implants. Nature **352**, 236-238.

III. Plans for the Next Quarter

We plan systematic parametric studies to evaluate tradeoffs among pulse duration, pulse rate, and interval between sequential pulses, with subjects representing a wide range of clinical and CIS performances. The studies with each subject will involve some or all of the conditions indicated in Table 1.

Among the studied subjects will be SR11, who will return to the laboratory in early June, and SR2, who will return sometime in late spring or early summer. We are in the process of scheduling additional subjects for these and other tests.

We also plan follow-up studies with one of the new MiniMed patients and with Ineraid subject SR9.

Our plans also include:

1. Studies with SR2 to evaluate techniques of noise reduction, used in conjunction with CIS processors.
2. Presentation of project results in an invited lecture at the *Fourth Symposium on Cochlear Implants in Children*, to be held in Kansas City, MO, February 14 and 15.
3. Continued preparation of manuscripts for publication, including an invited chapter for the book *Cochlear Implants: Audiological Foundations* (Edited by R. Tyler).

TABLE 1. Study to evaluate effects and possible interactions of changes in pulse duration, pulse rate, and interval between sequential pulses. Durations and rates to be used are presented in the Table. All other parameters of the processors will be held constant and will include fullwave rectifiers and 200 Hz lowpass filters in the envelope detectors. The lowpass filters will have a 4th order rolloff. These parameters will prevent aliasing in the "low rate" processors. Where necessary, delays will be interposed between pulses on sequentially stimulated channels to produce the indicated rates. Because coverage of the entire matrix will require a considerable amount of testing time, some rows or columns may be eliminated for some subjects. At least several subjects will be tested with all 21 conditions. The recent results with SR10 emphasize the need for this study.

	Pulse Duration (μs/phase)					
Pulse Rate (pps)	33	67	100	133	167	200
417	*	*	*	*	*	*
500	*	*	*	*	*	
627	*	*	*	*		
833	*	*	*			
1244	*	*				
2525	*					

Appendix 1

Summary of Reporting Activity for the Period of

November 1, 1991 through January 31, 1992

NIH Project N01-DC-9-2401

Reporting activity for the last quarter included preparation and submission of the two papers cited below. The first paper now has been accepted for publication, pending the incorporation of minor revisions.

Wilson, B.S., Lawson, D.T., Finley, C.C., and Wolford, R.D. (1992). Importance of patient and processor variables in determining outcomes with cochlear implants. *Journal of Speech and Hearing Research*, accepted for publication with revisions.

Lawson, D.T., Wilson, B.S., and Finley, C.C. (1992). New processing strategies for multichannel cochlear prostheses. *Progress in Brain Research*, submitted.

Appendix 2

Tables of Feature Assignments for 16 Consonants and 8 Vowels

TABLE A2.1. Assignment of consonant features.

Consonant	Voicing	Nasality	Frication	Duration	Place	Envelope	Affric
m	2	2	1	1	1	4	1
n	2	2	1	1	3	4	1
f	1	1	2	1	1	3	1
v	2	1	2	1	1	2	1
s	1	1	2	2	3	3	1
z	2	1	2	2	3	2	1
ʃ	1	1	2	2	4	3	1
ð	2	1	2	1	2	2	1
p	1	1	1	1	1	1	1
b	2	1	1	1	1	2	1
t	1	1	1	1	3	1	1
d	2	1	1	1	3	2	1
g	2	1	1	1	5	2	1
k	1	1	1	1	5	1	1
dʒ	2	1	1	1	4	2	2
l	2	1	1	1	3	4	1

TABLE A2.2. Assignment of vowel features.

Vowel	F1	F2	Duration
i	1	1	1
ɔ	2	2	1
ɛ	2	1	2
u	1	2	1
I	1	1	2
U	1	2	2
ʌ	2	3	2
æ	2	1	1