original
sent to printing
2/4/92

Tenth Quarterly Progress Report

August 1 through October 31, 1991

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

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 encode these parameters for electrical stimulation of the auditory nerve or central auditory structures. Work in the present quarter included the following:

1. Initiation of studies with Ineraid subject SR9, to evaluate *continuous interleaved sampling* (CIS) processors with a patient who has relatively poor results using her clinical *compressed analog* (CA) processor. (This continues a new series of studies with patients who have poor clinical performances. Results for another subject in the series were presented in QPR 5 for this project. A total of approximately six subjects is planned for the series.)
2. Continued studies with Ineraid subject SR2, to complete work begun in the last quarter. The studies included (a) evaluation of a new variation of CIS processors in which the order of pulse presentations is randomized for each succeeding sequence of stimulation across the electrode array, (b) evaluation of CIS processors that present pulses at relatively low rates on each channel, (c) measurement of consonant identification for CIS processors with different numbers of channels, (d) evaluation of a new processor structure, the hybrid *peak picker*/CIS strategy, and (e) measurement of consonant identification for the CA and CIS processors in the presence of multitalker speech babble.
3. Presentation of project results in invited lectures at the *International Symposium on Natural and Artificial Control of Hearing and Balance*, held in Rheinfelden, Switzerland, September 4-8 (Lawson), the *Annual Meeting of the American Neurotology Society*, held in Kansas City, MO, September 21 (Wilson), and the *Neural Prosthesis Workshop*, held in Bethesda, MD, October 22-24 (Wilson).
4. Continued preparation of manuscripts for publication.

In this report we describe work related to point 2 above. Results from studies indicated in point 1 will be presented in a future report, when data from additional subjects can be included.

In addition to the material related to point 2, we present in a final section a discussion of the use and possible development of new tests for evaluating the speech reception abilities of implant patients. Many of our better subjects now are approaching or hitting the 100% ceilings of our standard tests when using the CIS and other new processors. Thus, even more difficult tests will be needed in future studies to demonstrate further improvements in performance. Several such tests are listed in the last section as candidates under current consideration.

II. Randomized Update Orders

Many multichannel cochlear prosthesis patients achieve higher levels of speech perception performance when their electrodes are stimulated in nonsimultaneous sequences of pulses rather than multiple channels being stimulated at once. Some of those patients achieve even higher performances when the order in which their electrodes are stimulated in each sequence is controlled so that adjacent electrodes are never stimulated sequentially. We have referred to such stimulation schemes as presenting pulses in a *staggered* order, a typical example being 6-3-5-2-4-1 for a six-channel CIS processor, where channel 1 corresponds to the most apical electrode. All such fixed orders stimulate each channel at a constant pulse rate.

Depending on the pulse widths and separations used and on individual psychophysical characteristics, some patients may perceive such a fixed stimulation rate as a pitch. In some cases adaptation may reduce the salience of such a percept within the first minute or so of processor operation. In other cases the perceived pitch may compromise the processor's effectiveness.

Design of Stimulation Sequences

Relief from such a pitch percept may be obtained in some cases without changing the long term stimulation rate per channel. This may be accomplished by randomizing the order of channel stimulation from sequence to sequence, subject to such constraints as never stimulating adjacent electrodes sequentially and requiring a minimum number of intervening pulses on other channels between successive stimulations of any one channel.

A number of issues arise in the implementation of such randomizations of stimulation order. The design of filters for the signals being conveyed by each channel must take into account a range of effective stimulation rates rather than a single uniform one, for instance. The minimum effective rate will correspond to a channel's being the first stimulated in one sequence cycle and the last stimulated in the next, while the maximum effective rate will be dictated by the required minimum number of intervening pulses on other channels. Specifying a minimum number of intervening pulses may also dramatically bias the statistical distribution among the remaining possible intervals between successive stimulations of any one channel.

To demonstrate some of these effects, we shall examine a six-channel randomized order processor.

One of the first considerations in designing such a processor is determining, for the given number of channels, how many independent orders of stimulation exist that avoid stimulating adjacent channels sequentially. (The real-time processing overhead involved can be greatly reduced by making random selections among available valid sequences, rather than applying a full set of channel selection rules to each individual pulse.) The number of such independent valid sequences is a strong function of number of channels. There are only two such sequences for a five-channel processor: 1-3-5-2-4 and 1-4-2-5-3. (Any beginning point, of course, may be chosen for either of these sequences: 3-5-2-4-1, 5-2-4-1-3, 2-4-1-3-5, and 4-1-3-5-2 are all variants of the first sequence.) The number of such independent valid sequences is 66 for seven-channel processors, 490 for eight-channel versions, and 10 for our six-channel example, to wit:

1-3-5-2-4-6
1-3-5-2-6-4
1-3-6-4-2-5
1-4-2-5-3-6
1-4-2-6-3-5
1-4-6-2-5-3
1-5-2-4-6-3
1-5-3-6-2-4
1-6-3-5-2-4
1-6-4-2-5-3

If these ten available six-channel sequences are arranged in a 6 x 10 matrix, a candidate for the next sequence to be presented can be obtained by generating two random integers -- a row 1..10 and a starting column 1..6. We must then compare the channel(s) stimulated at the end of the previous sequence with the channel(s) to be stimulated at the beginning of the candidate sequence, to be certain that the minimum intervening pulse requirement will be met for each channel. (If only one intervening pulse on a different channel is required, this test consists merely of verifying that the channel last stimulated by the previous sequence is not the same one to be stimulated first by the proposed next sequence. A two-intervening-pulse requirement will mean consideration of the last two channels of the previous sequence and the first two of the candidate, *etc.*) The random selection of candidate starting points (row,column) continues until one is found that meets the intervening pulse criteria. This procedure can be carried out in real time by the speech processor, or can be used in advance to prepare a long list for the processor's use.

In our example, the interval between successive pulses on the same channel can vary between a maximum corresponding to a rate that is 6/11 = 0.54 of the long-term average rate, and a minimum corresponding to a rate 3, 2, or 1.5 times the long-term average for criteria of 1, 2, or 3 minimum intervening pulses respectively. This range of effective rates on each channel must be taken into account when designing processor filters -- to avoid aliasing problems, for instance.

Figure 1 shows the effect of requiring a minimum of 1, 2, or 3 intervening pulses on interpulse interval histograms for each channel in our example processor. The statistics are for lists of 1365 consecutive sequences in each case (a number chosen to fill the available buffer of a particular real-time processor). As the criteria become more restrictive, the histogram rapidly becomes more sharply peaked about the long-term average rate (five intervening pulses).

Preliminary Evaluation of Processors Using Randomized Update Orders

We have evaluated several processor implementations using randomized update orders in recent tests with subject SR2. One of these presented pulses at typical rates for CIS processors studied to

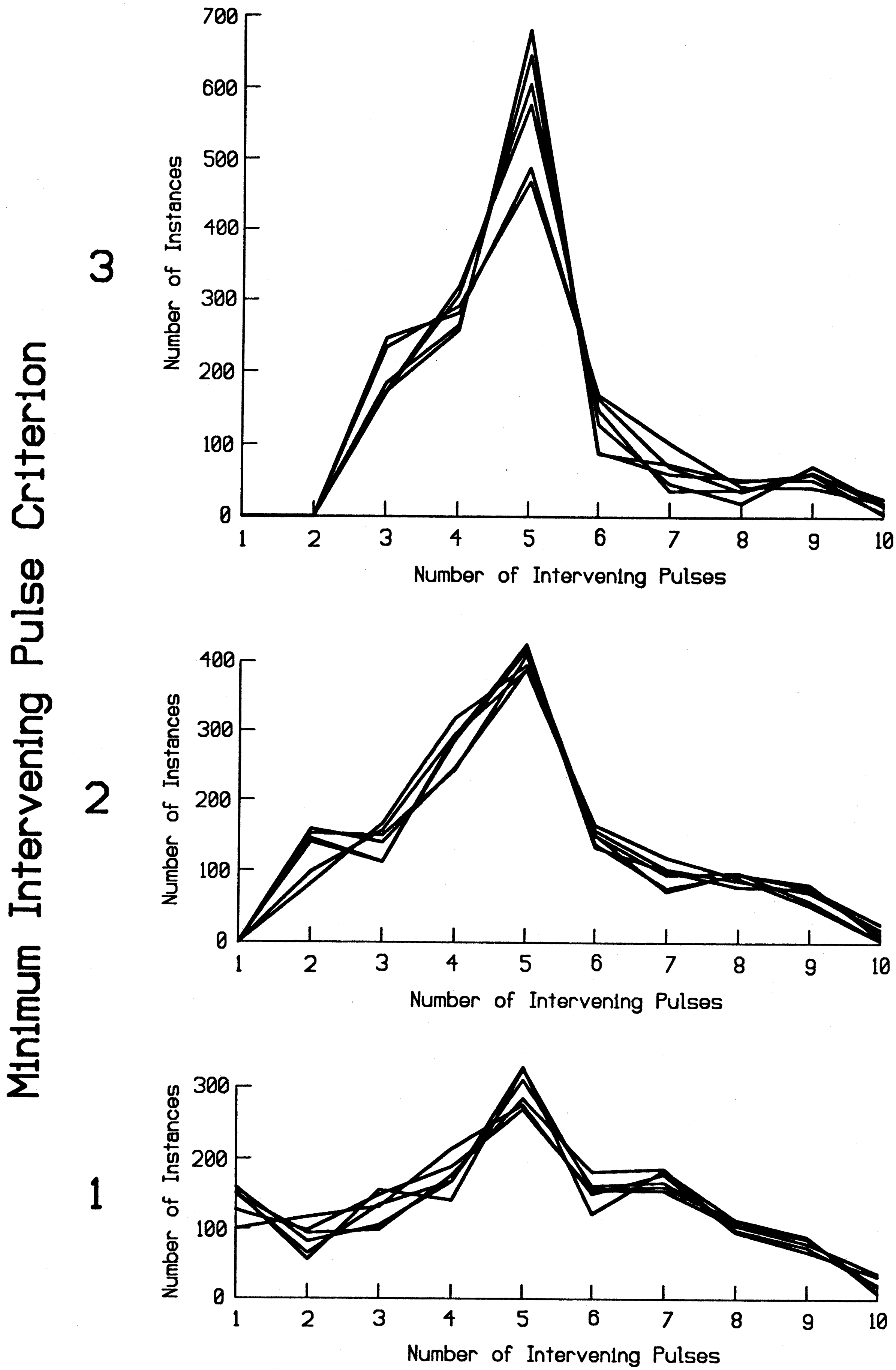

FIG. 1. Each curve represents one of the six channels in each case.
1365 sequences tallied in each case.

date (800 to 2525 pps on each channel), and another presented pulses at a much lower rate, as will be described below under the heading of "Slow Rate CIS Implementations."

A possible advantage of randomized update orders for CIS processors using typical rates of stimulation is that effects of forward and backward masking may be balanced across channels by averaging. In a fixed order of stimulation one particular electrode always precedes another particular electrode in the stimulation sequence. Forward masking produced by stimulation with the first electrode, and backward masking produced by stimulation with the second and subsequent electrodes, may distort the representation of channel intensities. For example, forward masking produced by stimulation with the first electrode could reduce the number of nonrefractory neurons available for stimulation by the second electrode. As suggested above, a randomization of the stimulus update order, with consequent randomization of masking patterns, may allow a more balanced representation of channel intensities.

Tests

The comparison tests included identification of 24 consonants (/b, d, f, g, dʒ, h, j, k, l, m, n, ŋ, p, r, s, ʃ, t, tʃ, ð, θ, v, w, z, ʒ/) in an /a/-consonant-/a/ context and the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery [Owens et al., 1985].

In the consonant tests multiple exemplars of the tokens were played from laser videodisc recordings of a male or female speaker [Tyler et al., 1987; QPR 14, NIH project N01-NS-5-2396]. A single block of trials consisted of five randomized presentations of each consonant for one of the speakers.

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 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, without feedback as to correct or incorrect responses. Results for the consonant identification test 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. A table of feature assignments for the 24 consonants is presented in Appendix 2.

Results

Results from a comparison of a standard CIS processor with a randomized list CIS processor are presented in Fig. 2. Both processors used 33 μs/phase pulses, presented either at a fixed (standard CIS) or average (random list CIS) rate of 2525 pps. Other parameters shared by the two processors included the use of 6 channels, 6th order bandpass filters, fullwave rectification, and 400 Hz smoothing filters (4th order). The standard CIS processor used a staggered update order (6-3-5-2-4-1), and the randomized list processor used a minimum intervening pulse criterion of one (see Fig. 1).

The results demonstrate high levels of performance with both processors. Because all scores except those for the NU-6 test either approach or encounter the 100% ceiling, no difference between the processors is evident in these particular results. Further, the NU-6 scores are quite close, with each exceeding 80% correct.

We note that novel lists, with different recorded speakers, were used for the CID, SPIN and NU-6 tests. In addition, 24 consonants were used instead of our standard 16 (see above and QPRs 4 and 9 for this project) to increase the difficulty of the consonant identification test. Despite these precautions and the increase in the number of consonants, though, SR2 achieved quite high scores on all of these tests.

While even these tests are no longer sensitive enough to distinguish among some of the better processors for SR2 and other subjects, SR2 did note that the percepts produced with the random list processor sounded even more intelligible and natural than the percepts produced with the standard CIS processor. We plan to investigate further the use of randomized update orders in future studies with subjects whose clinical performances span the range from poor to excellent. In addition, we plan to add tests with greater sensitivities for demonstrating differences among processors in the studies with subjects who have high levels of clinical performance (see section VII on identification and possible development of more sensitive tests).

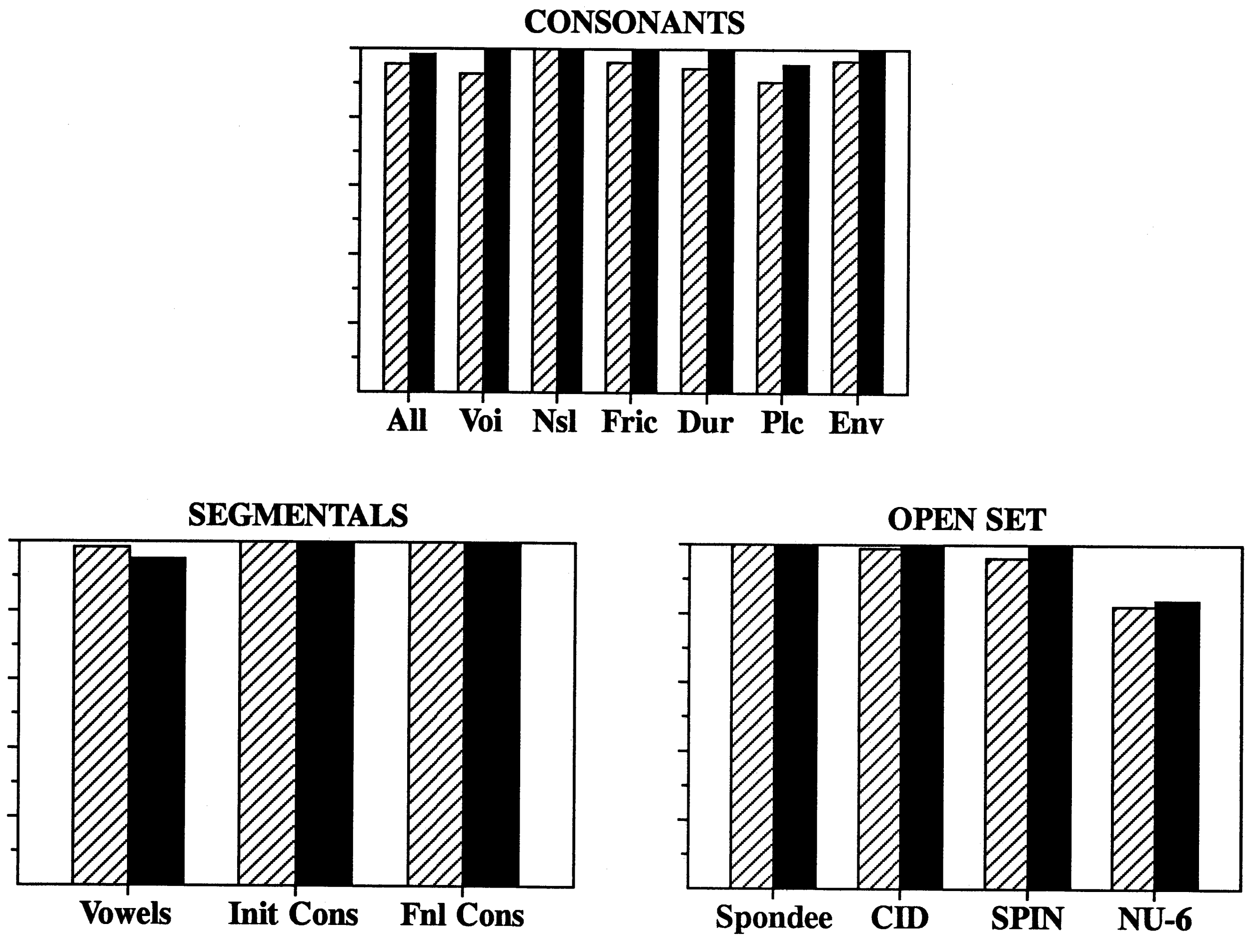

FIG. 2. Comparison of speech test scores for CIS processors that use fixed (striped bars) versus randomized (solid bars) orders of channel updates. The top panel shows relative information transfer for articulatory and acoustic features of consonants. The features include overall transmission (All), voicing (Voi), nasality (Nsl), frication (Fric), duration (Dur), place of articulation (Plc), and envelope cues (Env). Full scale corresponds to 100% information transfer. Ten presentations of each of 24 consonants were used in the consonant identification test for the fixed order processor (striped bars), and five presentations of each consonant were used for the randomized order processor (solid bars). The male speaker was used for both tests. The bottom panels show scores from the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery. See text for abbreviations. Full scale corresponds to 100% correct. Data are from tests with Ineraid subject SR2.

III. Slow Rate CIS Implementations

As noted above, our principal interest in randomized list processors relates to a continuous whistling percept produced with relatively low rates of stimulation in CIS processors. We suspect that, for these lower pulse pates, subjects can hear the pulse train carrier separately from the modulation of the carrier. That is, pulse rates below 800 pps probably are within the "pitch saturation limit" of some of the better subjects, and stimuli within this limit may be heard as having a distinct "rate pitch."

Subject SR2, for instance, when being switched between two similar CIS processors -- from one with a pulse rate exceeding his pitch saturation limit to one with a lower rate -- reported that the speech sounded essentially the same, but that a high-pitched whistle or buzz had been added, a new component that came and went in synchrony with the speech cadence.

When we increase rates of stimulation above 800 pps (or above 400-500 pps for some subjects), the whistling disappears and only natural-sounding speech is perceived. These changes are consistent with the idea that only the modulation is heard when the pitch saturation limit is exceeded.

Our first studies with processors using some of the attributes of our recent CIS designs date back to the mid 80s, when we applied a constant rate pulsatile processor in tests with one of the UCSF percutaneous subjects [Wilson et al., 1988]. This subject had extremely narrow dynamic ranges (on the order of 1 dB for pulsatile stimulation at 300 pps) and extremely poor performance with his standard CA processor (using the UCSF design). Indeed, he refused to describe any percepts elicited with the analog strategy as sounding like speech. The application of the pulsatile processor immediately produced speech percepts for this subject, and immediately produced scores significantly above chance on tests of consonant and vowel identification (scores for the CA processor were at chance levels). Unfortunately, medical complications intervened and this subject's implant device had to be removed before he could use the new strategy in his daily life, and before we could conduct any more tests.

That processor implementation included the use of relatively low rates of stimulation by our present standards. The envelope signals from six bandpass channels were scanned for each frame of stimulation. Pulses then were delivered to electrodes corresponding to the two channels with the highest envelope signals among the six. 300 μs/phase pulses were used, with 1.0 ms of "dead time" between sequential pulses. The maximum rate on any one channel was 313 pps, which is well below the 800 pps or higher that we now typically use for CIS processors.

We note that this early processor, while CIS-like in some respects, also was similar in design to the new Spectral Maxima Sound Processor (SMSP) described by the Melbourne team [McKay et al., 1991; McKay and McDermott, 1991; McDermott et al., 1991].

This particular subject did not report any whistling percepts with use of the 313 pps processor. Possibly, his limit of "pitch saturation" was at or below the typical limit of 300-400 pps found for many implant patients (see footnote 1 at the beginning of the references section and Eddington et al., 1978; Shannon, 1983; Simmons, 1966; Tong et al., 1983).

In subsequent studies, beginning in the summer of 1989 with subject SR2, we evaluated a CIS-like processor that presented its pulses at 278 pps (in this processor all six channels were updated in each stimulus frame; see QPR 1). SR2 enjoyed immediate improvements in speech recognition with the use of that processor, compared to the levels of recognition obtained either with his clinical CA processor or with an *interleaved pulses* (IP) processor. However, SR2 heard a distracting whistle along with his speech percepts using the "slow rate" CIS processor.

In later studies with SR2 we used rates in excess of 800 pps. The whistling disappeared and speech scores improved. Results from studies with many subsequent patients have confirmed these early findings.

Despite these advantages of relatively high rates of stimulation, we remain interested in the development of processors that use lower rates. The principal reason for this interest is that many patients are implanted with the Nucleus device, which has a transcutaneous transmission system (TTS) that will not allow the rapid sequencing of one pulse after its predecessor. In fact, the maximum rate of stimulation that can be sustained with the TTS for CIS-like stimulation is around 500 pps on each of six channels. If a way could be found to eliminate the whistling percept associated with such low rates of stimulation (at least for some subjects), then the benefits of the CIS strategy might be made available to the large number of people implanted with the present Nucleus device.

In recent studies we have evaluated the possibility that use of a randomized list processor might reduce or eliminate the whistling percepts of slow rate CIS processors while maintaining high levels of speech recognition. The studies again were conducted with Ineraid subject SR2. As a control, a CIS processor using a fixed rate of 500 pps on each channel was first compared with the standard CIS processor described above, using a fixed rate of 2525 pps. Other than the different rates of stimulation, these two processors had identical sets of parameters.

In a second experiment we compared the CIS processor using a fixed rate of 500 pps with a CIS processor using randomized update orders, with an average rate of 500 pps on each channel. The randomized list processor again used a minimum intervening pulse criterion of one.

Results for the first experiment are presented in Fig. 3. The consonant tests for this comparison included 24 consonants, for both the male and female speakers. As shown in the figure, both processors produced high scores for all tests. Indeed, ceiling effects do not allow distinctions between these processors on the basis of the administered tests.

In contrast to results obtained with other subjects, very high levels of speech recognition are maintained with the slow rate CIS processor. With other subjects, recognition scores have dropped with reductions in rates below 800 pps (and concurrent reductions in the cutoff frequency of the lowpass smoothing filters to prevent aliasing). While more sensitive tests may have demonstrated a difference for these two processors for SR2, the fact remains that the speech recognition scores for both processors are quite high.

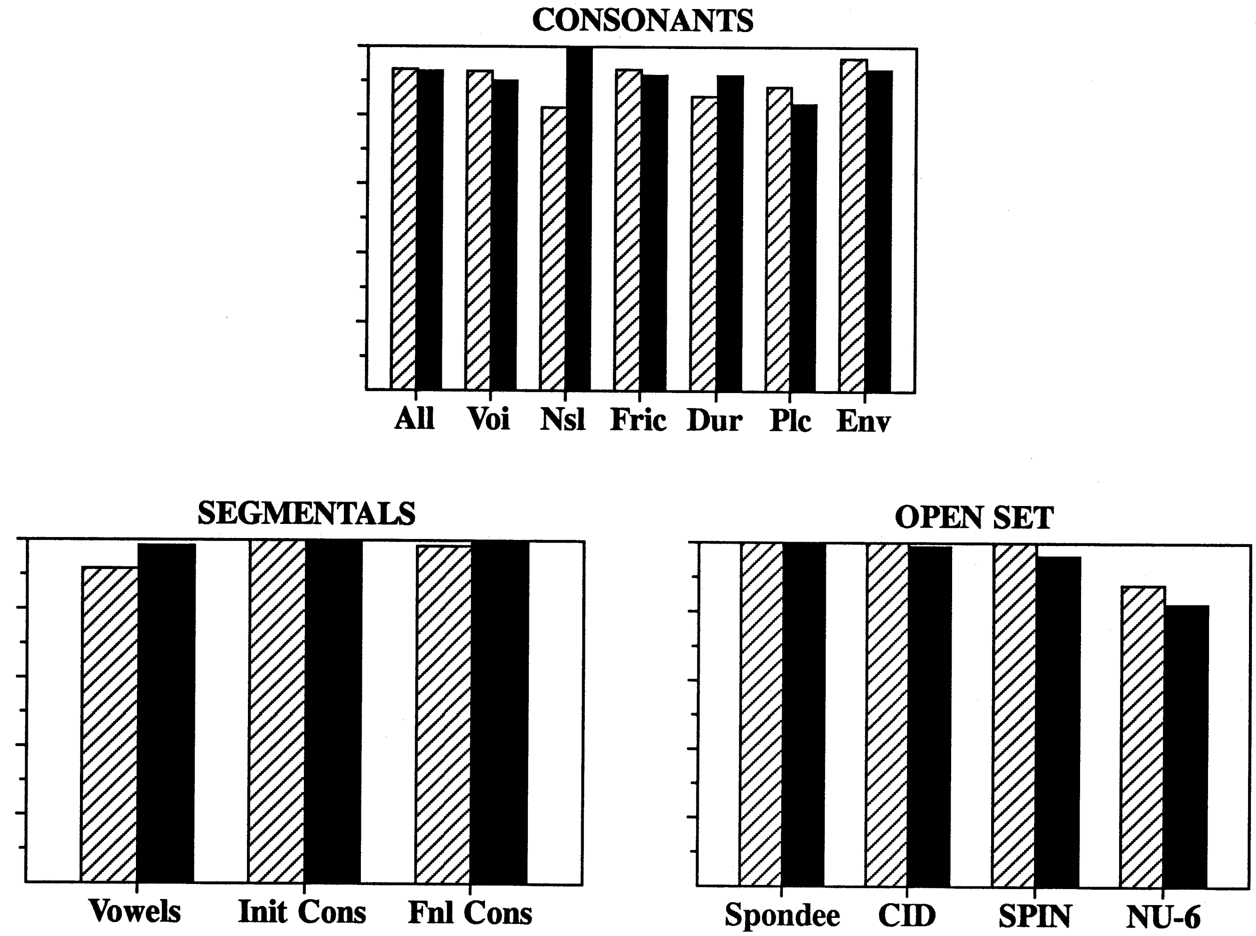

FIG. 3. Comparison of speech test scores for CIS processors that use slow (500 pps, striped bars) versus typical (2525 pps, solid bars) rates of stimulation. Ten presentations of each consonant were used in the consonant identification test for the slow rate processor (striped bars), and twenty presentations were used for the standard rate processor (solid bars). The presentations for each processor were equally divided between the male and female speakers.

As with previous slow rate CIS processors, SR2 heard a whistle along with speech when using the fixed rate 500 pps processor.

Results for the second experiment are presented in Fig. 4. The consonant scores are presented for the male speaker only, as tests with the female speaker were not conducted with the randomized list processor. Also, time limitations did not permit completion of the spondee, CID or segmental tests for the randomized list processor.

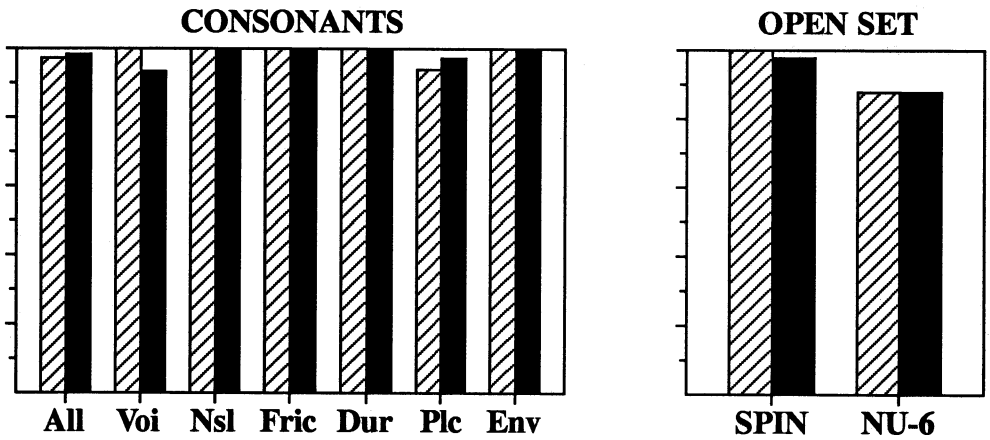

FIG. 4. Comparison of speech test scores for slow rate (500 pps) CIS processors that use fixed (striped bars) versus randomized (solid bars) orders of channel updates. Five presentations of each consonant by the male speaker were used in the consonant identification tests for both processors.

High scores again are found for both processors. Indeed, the processors cannot be distinguished on the basis of these scores, as all scores hit or approach the 100% ceilings, including the scores for the NU-6 test (88% correct for both processors).

In contrast to the fixed rate processor, however, no whistle was heard with the randomized list processor. Apparently, the randomization of intervals between pulses on any one channel was sufficient to eliminate completely the perception of the carrier.

As indicated in recent results reported by Norbert Dillier [1991], CIS-like strategies might confer immediate and substantial gains in speech recognition to users of the standard Nucleus device. The present results further indicate that the whistling percepts that accompany some implementations of slow rate CIS processors might be eliminated by randomizing the update order from one stimulation sequence to the next. We plan additional parametric studies with randomized list processors, and we expect to include subjects implanted with the standard Nucleus device.

IV. Channel Number Manipulations

Another parametric study of special interest is illustrated in Figs. 5 and 6. Here we show the results of consonant identification tests for CIS processors with 6, 5, 4, 3, and 2 channels, and for a CS processor with 1 channel (the CS processor is a single channel variation of "CIS" processors, but is named CS because interleaving is not applicable to a single channel). Each n-channel processor used the n apical-most electrodes and filtered the same total frequency range into n logarithmically-spaced bands. For example the three channel processor used apical electrodes 1, 2 and 3. All processors used 33 μs/phase pulses, presented at the rate of 2525 pps on each channel (delays were interposed between sequential pulses for processors other than the six-channel processor to maintain a constant rate of stimulation across numbers of channels). In addition, each processor used 6th order bandpass filters, fullwave rectifiers, and 400 Hz smoothers (4th order). For consistency, a fixed base-to-apex update order was used for all processors. For example, the three channel processor stimulated its electrodes in the sequence 3-2-1.

We note that none of the processors in this series was optimized for the individual subject. The six channel version, for instance, was inferior to staggered-order processors for this subject. Also, processors using less than six channels probably would have benefited from use of specific electrodes other than the most apical n (e.g., use of more widely spaced electrodes may have produced a better result). The purpose of the present manipulation was to evaluate effects of changes in the number of channels, while maintaining a consistency in all other CIS parameters.

The results show a strong effect of channel number on consonant identification and feature transmission. Overall percent correct scores decline monotonically, for both the male and female speakers, with reductions in the number of channels (Fig. 5). Also, transmission of place information declines precipitously for the male speaker as the number of channels is reduced from 6 to 3, and drops precipitously for the female speaker as the number of channels is reduced from 5 to 4 (Fig. 6). In all cases the transmission of place information declines monotonically as the number of channels is reduced. In contrast, transmission of envelope information is relatively well maintained when the number of channels is reduced, as is the transmission of voicing, frication and nasality information for the male speaker (indeed, the transmission of voicing information remains high even for a single channel). Results for the female speaker are somewhat different in that the transmission of voicing and nasality information drops sharply when the number of channels is reduced from 2 to 1, and the transmission of frication information drops precipitously over the range of channel reductions from 5 to 1.

A consistent finding in the data is the dependence of place transmission on the number of stimulation channels. In addition, results from the female speaker suggest that transmission of frication information may depend on number of channels, at least up to 5 channels, and at least for certain speakers. Further increases in channel number may improve the transmission of place information and other important cues for the correct identification of consonants. As indicated in Table 5 of QPR 4 and elsewhere (e.g., Dorman et al., 1990; Tye-Murray and Tyler, 1989; Tyler, 1990), such identification is highly correlated with open-set recognition of words, sentences and running speech.

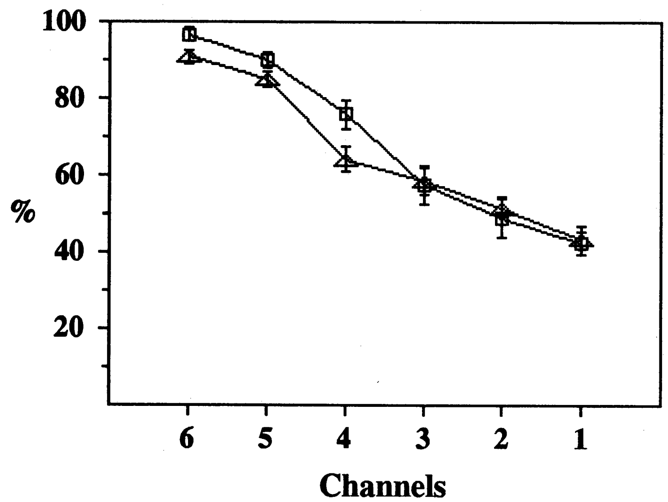

FIG. 5. Percent correct scores from processors using different numbers of channels. Five presentations of each of 24 consonants by the male speaker, and five presentations of each consonant by the female speaker, were used in the tests with each processor. The presentations were arranged in block randomized order, providing a percent correct score after each set of randomized presentations of all 24 consonants. The square symbols show the averages of these scores (from 5 randomized sets) for the male speaker, and the triangles show the averages for the female speaker. Standard errors of the mean are indicated with the vertical bars.

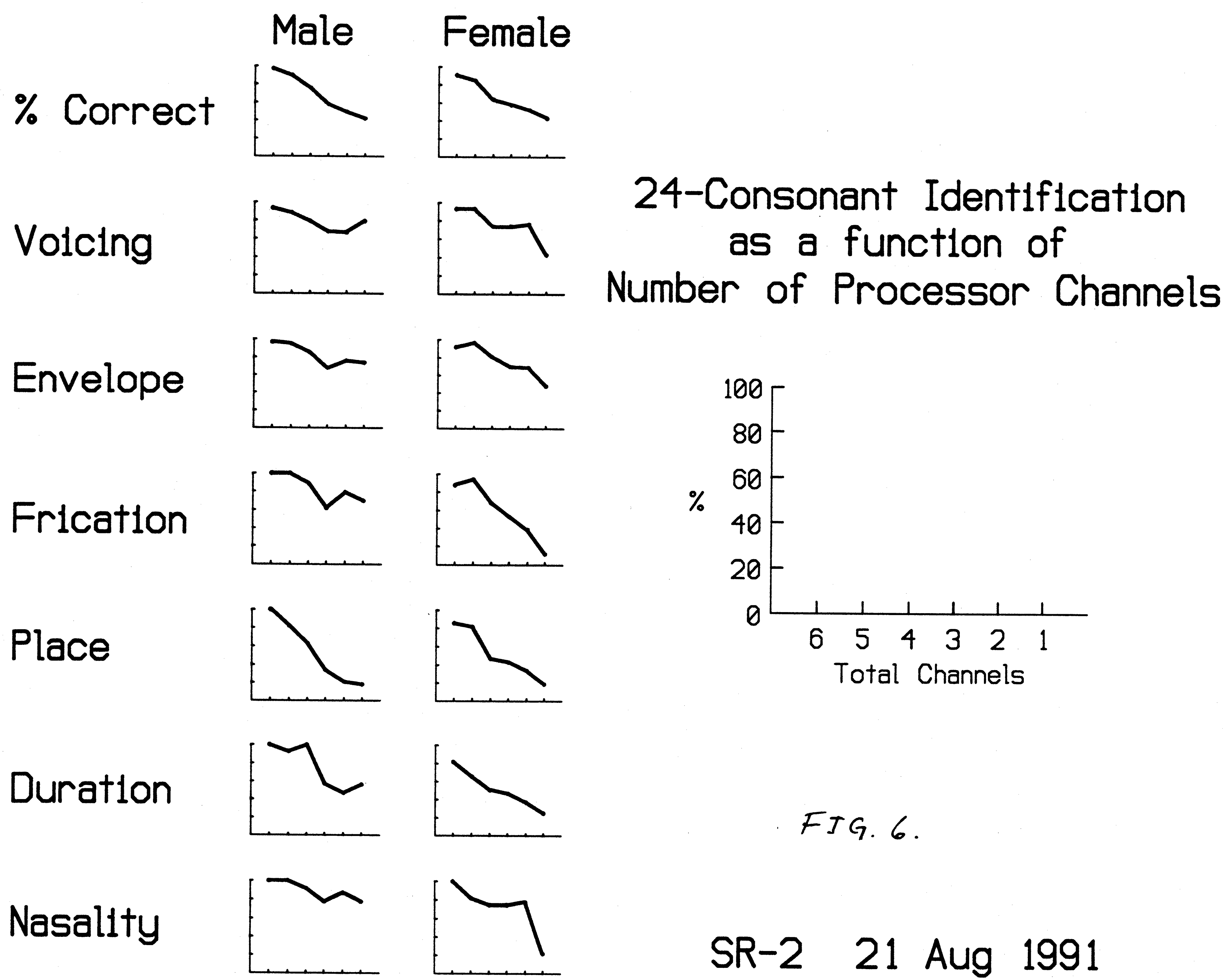
Male
Female
% Correct
Voicing
Envelope
Frication
Place
Duration
Nasality
24-Consonant Identification
as a function of
Number of Processor Channels
100
80
60
40
20
0
%
6
5
4
3
2
1
Total Channels
SR-2 21 Aug 1991

FIG. 6.

V. Evaluation of Other Promising Strategies

While very high levels of speech recognition have been obtained with the CIS strategy, other strategies may well be better, at least for certain classes of patients. One possibility is the *peak picker* (PP) strategy first described in QPR 3 for this project. In studies with one of our Ineraid patients, this strategy produced transmission scores for several consonant features that were higher than the scores obtained with the CIS strategy. Overall transmission of consonant information was approximately the same for the PP and CIS strategies. Transmission of vowel features to this patient by the PP strategy was perfect for our eight vowel test (compared with high, but not perfect, scores for the CIS strategy). The PP strategy obviously has promise and should be investigated further with additional tests and subjects.

A possible advantage of the PP strategy is that it uses generally lower rates of stimulation than the CIS strategy. This may allow useful implementations of the PP strategy for patients implanted with the Nucleus device, whose transcutaneous transmission system (TTS) does not permit the rapid sequencing of pulses required by typical implementations of the CIS strategy.

In addition to the PP strategy, we have conducted preliminary studies to evaluate a hybrid PP/CIS strategy. In this strategy PP stimuli are delivered to the apical-most electrodes (usually the two most apical electrodes in an array of six), and CIS stimuli are delivered to the remaining electrodes. This hybrid strategy attempts to combine attributes of the PP and CIS approaches.

In this section we summarize results reported in QPR 3 for the PP strategy, and we describe preliminary studies with the new PP/CIS strategy.

Peak Picker (PP) Processor

The design of the PP processor is illustrated in Fig. 7. In this processor the position of a peak in either the bandpass or envelope detector output is signaled by the presentation of a pulse. Also, as in the IP and CIS processors, the exact timing of the pulses is adjusted so that there is no temporal overlap of stimuli across channels.

In Fig. 7 the middle panel shows bandpass outputs for each of four channels along with the stimulus pulses derived from those outputs. In addition, the positions of the peaks in the bandpass outputs are marked by short vertical lines above each trace. The lower panel shows the stimulus pulses only.

In this particular implementation of the PP strategy, a timer is advanced for each channel in a sequence of stimulation across the electrode array. At each time step a pulse is delivered if a peak occurred in the bandpass output between the previous and present time steps for that channel. The amplitude of the pulse is determined with the same logarithmic transformation used in the IP and some CIS processors (i.e., the actual pulse amplitudes would be computed using a logarithmic transformation of the amplitudes shown in the figure). A fixed time is reserved for each channel in the stimulation sequence whether or not a pulse is delivered. As indicated in Fig. 7, this variation of the processor

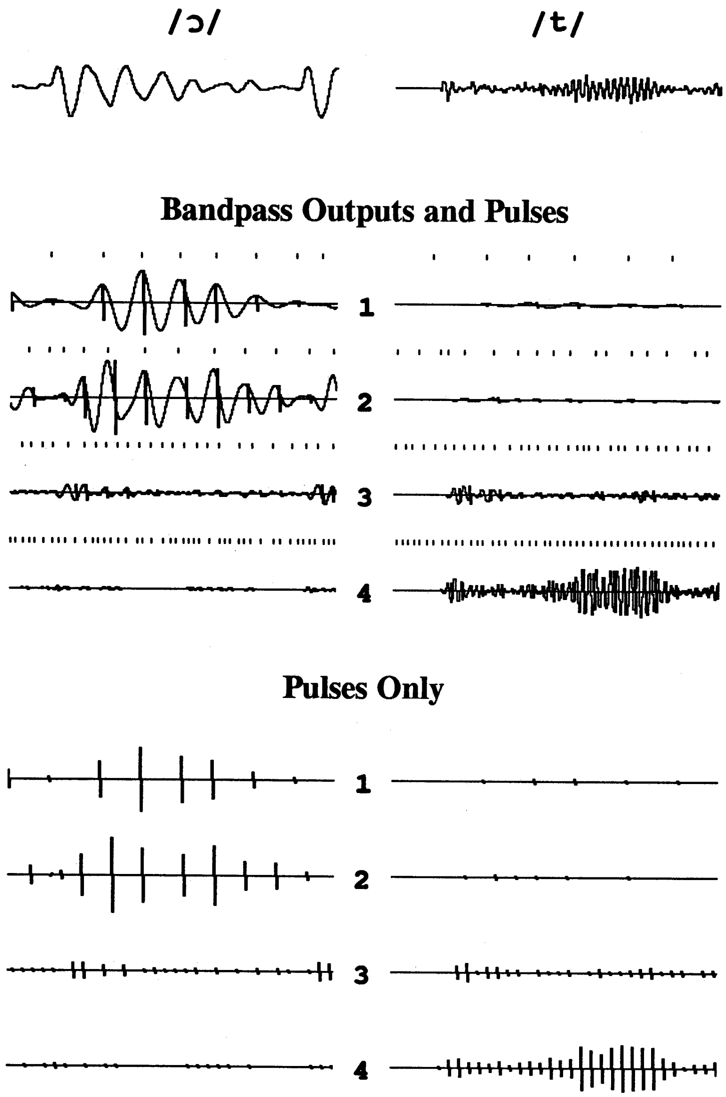

FIG. 7. Waveforms of the "peak picker" (PP) processing strategy. Equalized speech inputs are shown at the top and processor waveforms are shown in the middle and bottom panels. The middle panel shows the bandpass outputs and stimulus pulses for each of four channels. The location of peaks in the bandpass outputs are marked with short vertical lines above each trace. The bottom panel shows the stimulus pulses only. The duration of each trace is 12.25 ms.

produces clusters of pulses at the F0 rate and individual pulses at the F1 rate for voiced speech sounds (left panels). Because the pulses must be presented nonsimultaneously, though, higher frequencies in the bandpass outputs are not followed with pulses at those frequencies. Notice, for instance, that many peaks are missed in channels 3 and 4, and that large offsets between the positions of peaks and subsequent pulses are seen in the waveforms of channel 2.

Alternative implementations of the PP processor are illustrated in Fig. 8. The uppermost panel beneath the input signals shows the waveforms of the implementation just described ("Bandpass Outputs, Time for Each Channel"); the next panel down shows an implementation in which the time allocated for each channel is *not* used if a pulse is not delivered ("Bandpass Outputs, Channels Skipped"); and the bottom panel shows an implementation in which the outputs of the envelope detectors are used instead of the bandpass outputs ("Envelope Detector Outputs, Channels Skipped").

A summary of waveforms for various types of pulsatile processor is presented in Fig. 9. All three types of processor use nonsimultaneous stimuli. Among these, the CIS processor delivers the greatest number of pulses per unit time, and the IP processor the least. The PP processor provides an intermediate level of temporal detail, with a representation of F1 in the apical channel(s). In addition, the PP processor presents different rates of stimulation on each channel, which might increase the salience of channel-related cues for some patients (i.e., channel cues might be represented both by place of stimulation and by rate of stimulation).

One implementation of a PP processor was evaluated in preliminary tests with Ineraid subject SR2. The design illustrated in Fig. 7 was used, with time taken for each channel whether or not a pulse is delivered. The PP processor used six channels, with the staggered update order found to be best in (contemporaneous) evaluations of CIS processors (6-3-5-2-4-1).

The tests included identification of 16 consonants and 8 vowels, for male and female speakers (see QPR 3 for a complete description of the tests and related procedures).

Results for a contemporaneous implementation of the CIS strategy (processor RTSS8) and the PP strategy (processor RTPP1) are presented in Fig. 10. As noted above, both processors used a staggered update order. In addition, both used pulse durations of 55 μs/phase, no time delay between sequential pulses, and a 600 Hz corner frequency for the input equalization filter (present versions of CIS processors generally use shorter pulses and a 1200 Hz corner frequency for the equalization filter). Finally, the envelope detectors in the CIS processor used halfwave rectifiers and 800 Hz lowpass filters (present CIS processors generally use lower cutoff frequencies for the lowpass filter).

Comparison of the scores for the two processors shows a similarity in feature transmission for consonants. Overall transmission approximates 90% for both strategies. The PP strategy produces somewhat higher scores for the temporal features of voicing, duration, and envelope cues, and the CIS strategy produces somewhat higher scores for the features of nasality and place of articulation. Scores for frication are quite similar for the two processors.

In contrast to the overall picture for consonants, transmission of vowel information appears to be better with the PP processor. In fact, the 8 vowels of our vowel identification test were perfectly identified for both the male and female speakers when the PP processor was used.

The large gain in the transmission of vowel feature information found for F1 is consistent with the explicit representation of F1 in the apical channels with the PP processor.

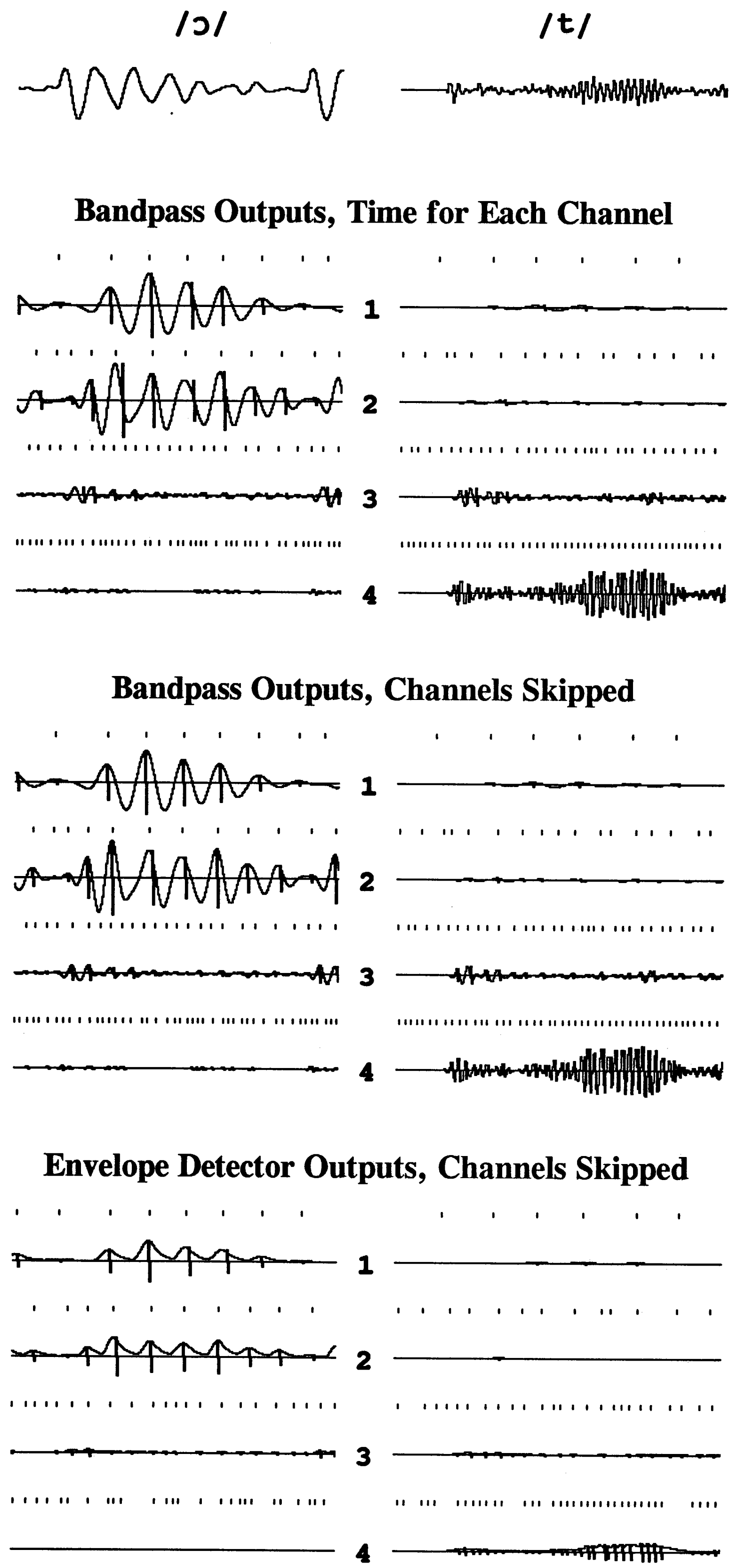

FIG. 8. Various implementations of PP processors. See text for details.

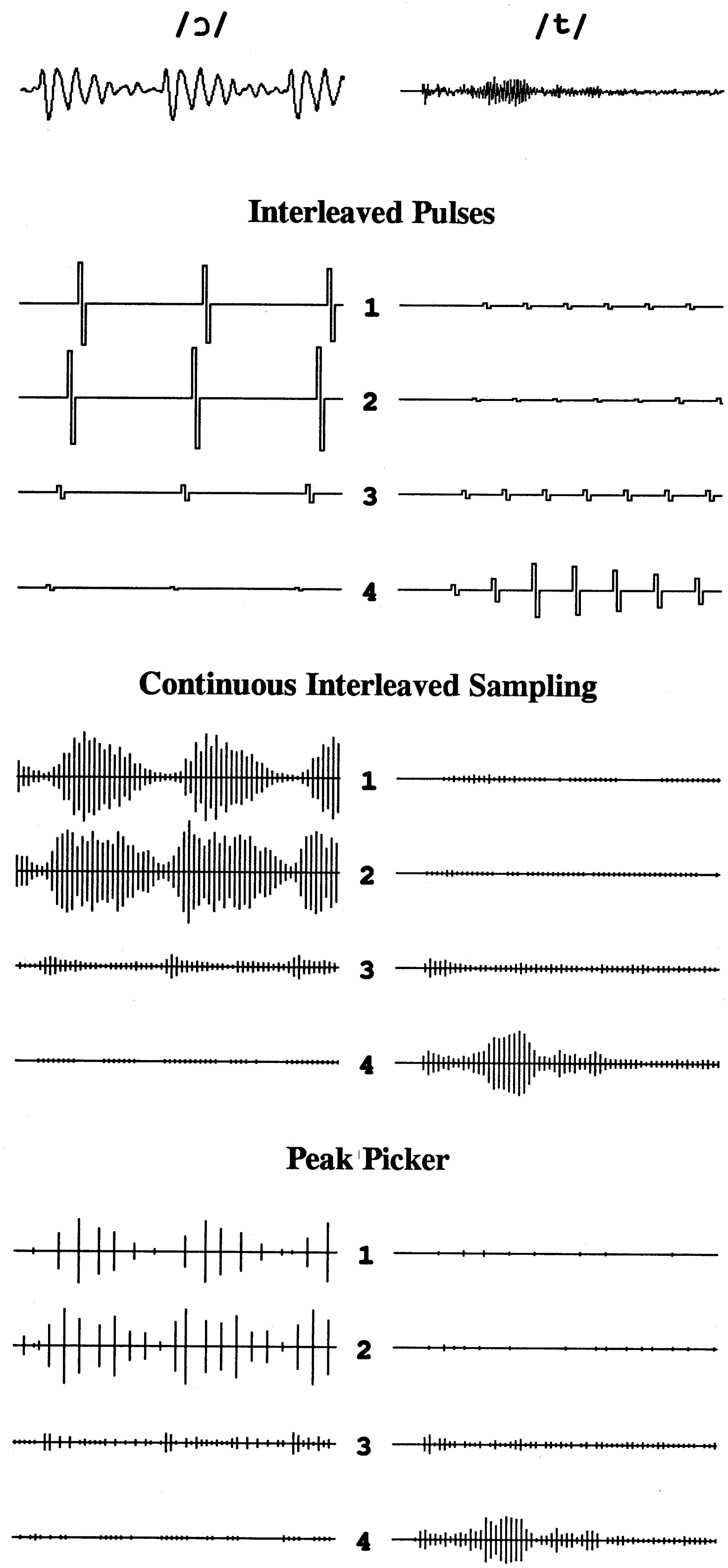

FIG. 9. Waveforms of three types of pulsatile processors. The duration of each trace is 25.4 ms. Note that the prolonged stimulation in the IP processor for the /t/ burst is a consequence of the long time constant of the lowpass filters in the envelope detectors (25 Hz cutoff versus 400 Hz cutoff for the other processors).

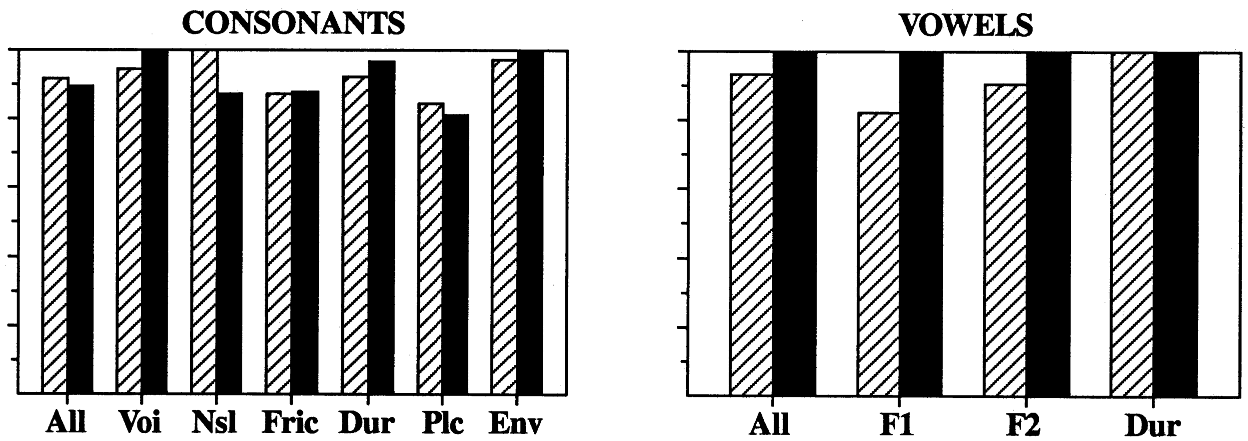

FIG. 10. Comparison of speech test scores for an early implementation of a CIS processor (striped bars) and a peak picker (PP) processor (solid bars) evaluated at the same time. Twenty presentations of each of 16 consonants were used in the consonant identification test for both processors. The presentations were equally divided between the male and female speakers. For the vowel identification tests 18 presentations were used for evaluation of the CIS processor (striped bars) and 12 were used for the PP processor (solid bars). The presentations were equally divided between the male and female speakers.

With the exception of this one feature, and possibly nasality for the consonant test, no obvious differences are found in the results for the two processors. We note, however, that many of the scores approach or encounter the 100% ceiling for these particular tests. More difficult tests will be needed to detect additional differences between the processors, if indeed such differences exist.

SR2 remarked that percepts produced with the PP processor were more "pitch appropriate" than percepts produced with the contemporaneous versions of the CIS processor, particularly for low frequency sounds such as the fundamental frequency of voiced speech. While the PP processor sounded a bit more natural to SR2, both processors were judged by him to be highly intelligible, with no clear difference between processors for recognizing connected speech. (SR2 has scored substantially higher with subsequently developed CIS versions.)

Hybrid PP/CIS Processor

In much more recent studies we have evaluated a hybrid of the PP and CIS strategies. In this PP/CIS processor PP stimuli were delivered to the apical two electrodes in the Ineraid array and CIS stimuli were delivered to the remaining four electrodes. The speech processor was programmed to examine the signals from the envelope detectors in the apical-most two channels just before the scheduled delivery of a CIS pulse on one of the more basal channels. If the processor detected a peak in one or both of the apical channels, then the CIS pulse would be delayed to allow the delivery of a (nonsimultaneous) PP pulse for each channel with a detected peak. This process was repeated for each CIS pulse.

Results from an initial evaluation of this hybrid PP/CIS processor are presented in Fig. 11. In addition, results from a CIS processor with parameters similar or identical to those of the CIS channels in the PP/CIS processor are shown. The consonant test used for these evaluations included 24 consonants. Tests were conducted with the male speaker only. The CID, SPIN and NU-6 tests were conducted using novel lists of recorded sentences and words (with a different male speaker).

Clearly, both processors support high levels of speech recognition. All scores except those for the NU-6 test approach or hit the 100% ceiling. The NU-6 scores are 84 and 80% correct for the CIS and PP/CIS strategies, respectively.

Anecdotally, the PP/CIS processor sounded quite natural, especially for music. Indeed, SR2, who was a musician before he lost his hearing, indicated that percepts produced with the PP/CIS processor had greater "pitch appropriateness" and musical clarity than percepts produced with his clinical CA processor or with the CIS processor (this observation was made while listening to small ensemble jazz recordings, which were familiar to the subject only through use of his clinical CA processor). Because the sound was so enjoyable to SR2 ("music is wonderful through this [PP/CIS] processor, like nothing I've even dreamed of ever hearing again with my implant"), we took some time off from testing so that he could listen to tapes of music he remembered from years ago (rock band material, which was familiar to the subject before loss of his normal hearing). Again, SR2 heard nuances in the material that could not be perceived with the other processors.

These high levels of performance for the PP/CIS strategy are encouraging. We plan further studies, with additional subjects and tests, to evaluate further the PP and PP/CIS strategies. The preliminary results, along with anecdotal comments, suggest that use of PP stimuli, at least for the apical channels, may provide an improved representation of frequencies in the F0 and F1 ranges, at least for some of the better subjects.

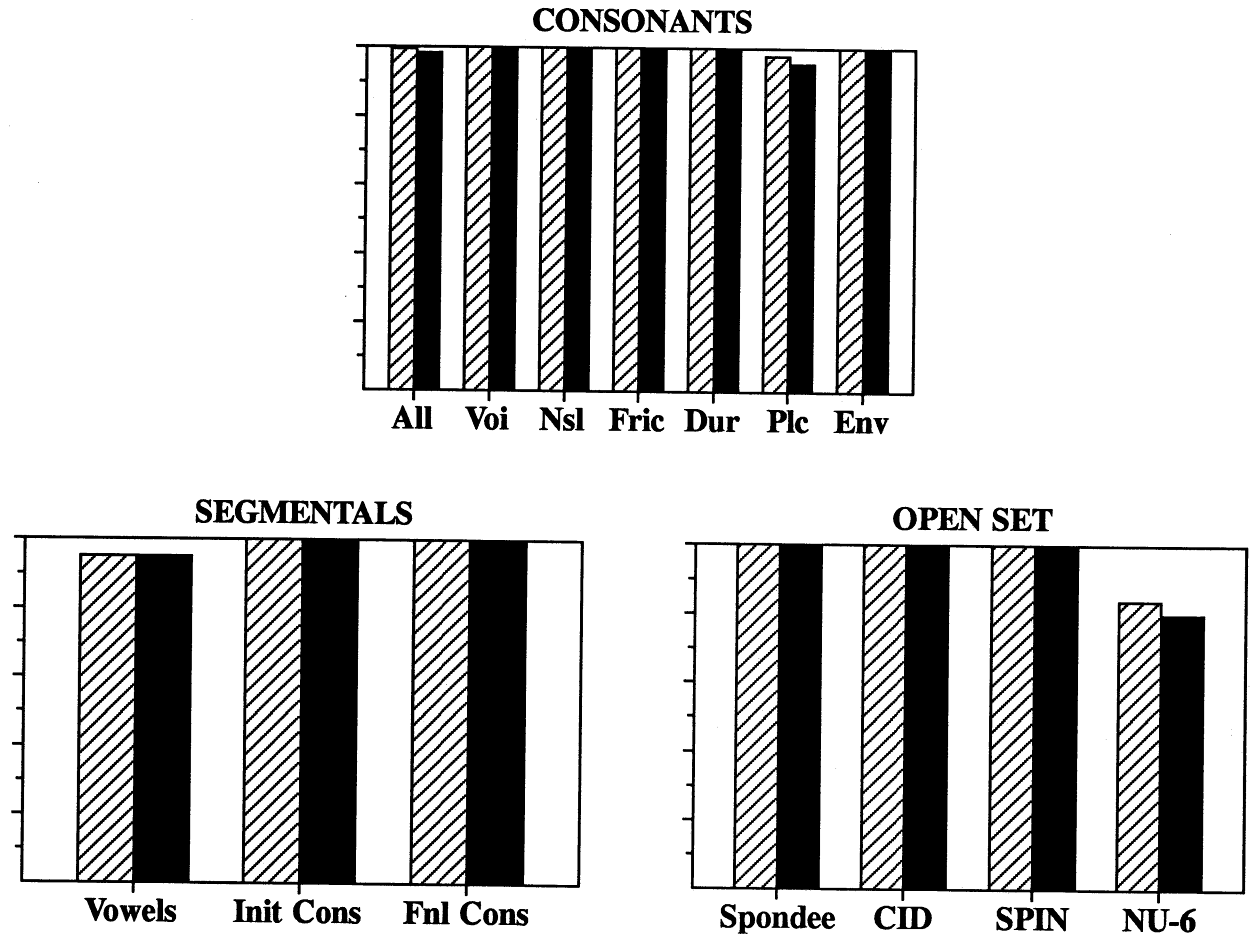

FIG. 11. Comparison of speech test scores for a standard CIS processor (striped bars) and a hybrid PP/CIS processor (solid bars). Five presentations of each of 24 consonants by the male speaker were used in the consonant identification tests for both processors.

VI. Performance of CIS and CA Processors in Noise

The primary goal of our studies to date has been to define a processing strategy that would support high levels of speech recognition for implant users. That is, we have focused on identifying peripheral representations adequate for speech understanding. This goal has been realized in large measure, at least for some patients, with the CIS strategy. Now that high levels of speech recognition have been achieved in quiet environments, issues of noise and noise reduction deserve a higher priority. Indeed, perhaps the most common complaint of implant patients relates to the difficulty of using their devices in noisy surroundings.

Noise reduction is a difficult problem. Many approaches have been tried without much success for both hearing aids [CHABA, 1991; Tyler and Kuk, 1989; Van Tasell et al., 1988] and cochlear implants [Skinner et al., 1991].

Among the various sources of noise, the most destructive to intelligibility is that of multiple talkers. This type of noise is extremely difficult to treat because it is nonstationary and its spectrum overlaps that of the speech signal of interest. Unfortunately, users of cochlear implants often find themselves in the presence of such noise, for example at parties or meetings.

We plan to evaluate several strategies for reducing the deleterious effects of noise interference. As a preliminary to those studies, we have measured the performances of the CIS and CA strategies in noise without any special provisions for noise reduction. The tests were conducted with Ineraid subject SR2, and included identification of 24 consonants in an /a/-C-/a/ context uttered by the male speaker on the Iowa videodisc materials. Consonant identification first was measured under quiet conditions, and then progressively greater amounts of multitalker speech babble were added to the primary speech signal. The noise was obtained from a babble tape recorded for use in SPIN sentence tests [Bilger et al., 1984; Kalikow et al., 1977]. Signal-to-noise ratios (SNRs) included 15, 10, 5 and 0 dB, with 0 dB corresponding to the babble signal amplitude exceeding the maximum consonant waveform amplitude briefly about once per second on average. No effort had been made to optimize the CIS processor for performance in noise.

The results are presented in Figs. 12 and 13. While the presence of noise clearly degrades the performance of both processors, relatively high percent correct scores are maintained down to a SNR of 5 dB. The scores for the CIS processor are higher than those for the CA processor at all SNRs (Fig. 12). This is especially encouraging inasmuch as the Iowa team has identified the CA processor in the Ineraid device as the most resistant to the deleterious effects of noise among several tested processing strategies and devices (including the multichannel "F0F2" strategy of Nucleus device; see Gantz et al., 1987).

One possible factor underlying the high levels of CIS performance in the presence of interfering speech babble is a good representation of envelope cues. In particular, covariation in envelope information across channels may help maintain high levels of speech recognition in noise (Smoorenburg, personal communication to Wilson, June, 1991; Moore, 1991). Such across-channel information may allow a listener to follow the correlated cues of the primary speech signal, while rejecting the uncorrelated variations produced by the noise.

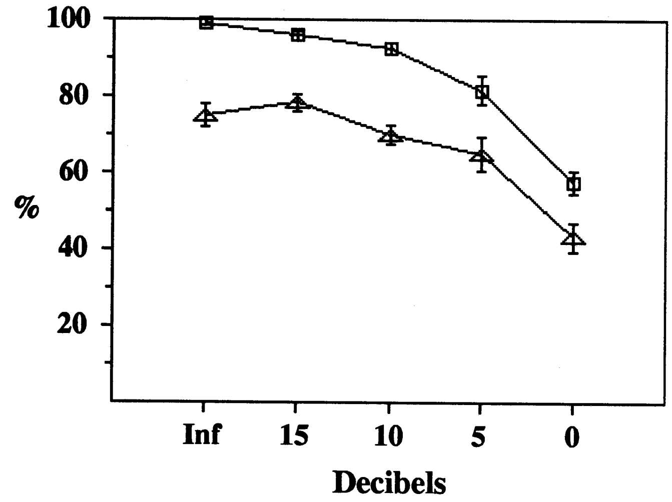

FIG. 12. Average percent correct scores for the CIS (squares) and CA (triangles) processors as a function of signal-to-noise ratio (SNR). Five presentations of each of 24 consonants by the male speaker were used in the consonant identification tests for both processors, at each SNR. Standard errors of the average percent correct scores are indicated with the vertical bars.

Another factor that may underlie the relatively high levels of performance found for both the CA and CIS strategies is the fact that neither relies on feature extraction. The accuracy of such extraction can be severely degraded by even modest amounts of noise.

A key lesson in the present results is that the choice of a basic processing strategy *per se* can have large effects on performance in noise.

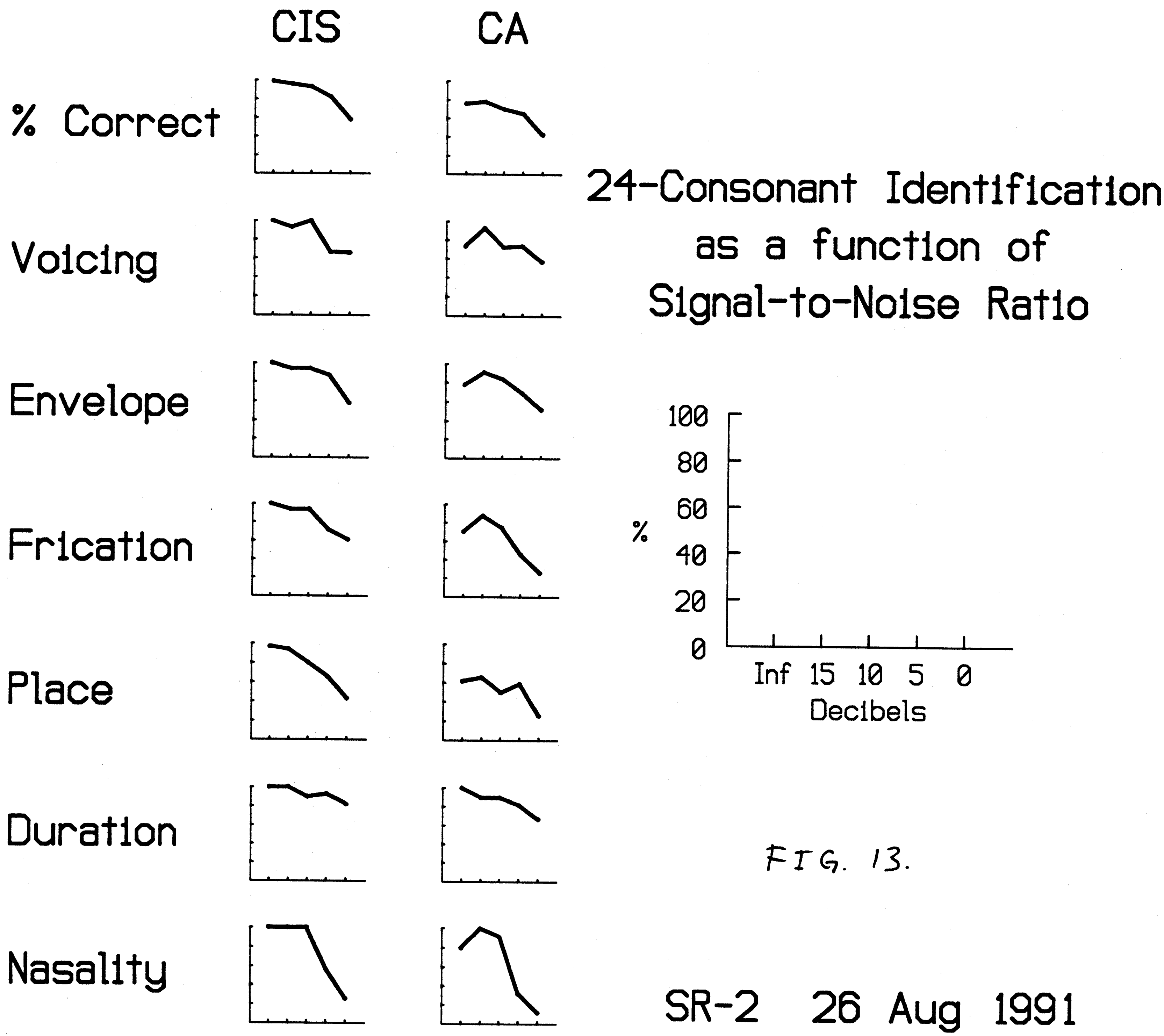
CIS
CA
% Correct
Voicing
Envelope
Frication
Place
Duration
Nasality
24-Consonant Identification
as a function of
Signal-to-Noise Ratio
100
80
60
40
20
0
%
Inf 15 10 5 0
Decibels
SR-2 26 Aug 1991

FIG. 13.

VII. Use and Possible Development of New Test Materials

As indicated elsewhere in this report, many of our better subjects are approaching or hitting the 100% ceilings of our standard tests when using the CIS and other new processors. New test materials must be identified to fill two distinctly different needs: (a) the rapid, sensitive, and reliable assessment of relevant differences in processor performance during fitting and optimization, and (b) the measurement of levels of performance in terms that will allow comparisons with tests of other patients, with other processors, and at other times and places.

Until recently, our tests of medial consonant identification have fulfilled need (a) and the open set subtests of the MAC battery have played role (b). The limitations of such tests in the face of rising patient performance have become a serious issue in our laboratory and soon will be impacting other investigators as well.

In our studies comparing the CA and CIS strategies, for instance, a majority of seven subjects (chosen for high levels of performance with the CA processor) approached the ceiling for most of our open-set tests [Wilson et al., 1991]. Among the seven subjects, five scored 96% or higher for the spondee test using the CIS processor; all seven scored 95% or higher for the CID test; and five scored 92% or higher for the SPIN test. Finally, two of the subjects scored 80% on the NU-6 test, which is in the range of scores obtained by people with mild-to-moderate hearing losses when taking the same test [Bess and Townsend, 1977; Dubno and Dirks, 1982].

In more recent studies one of our better subjects has achieved NU-6 scores in the high 80s or low 90s with a variety of CIS processor implementations. This same subject scored 100% correct on all other open-set tests (SPIN, spondee, CID), using novel lists and a variety of recorded speakers.

Obviously, new and even more difficult tests will be needed in future studies to demonstrate improvements beyond those recorded with the CIS strategy. This is a nice problem to have, in that its mere existence signifies a major advance in the performance of cochlear prostheses and shows that high levels of open-set recognition are possible with these devices.

In our first attempt to improve the sensitivity of the consonant identification test, we increased the number of consonants from 16 to 24. Even with multiple exemplars, randomized orders, and multiple speakers, though, the above subject scores in the mid to high 90s with several CIS processors, using hearing alone. Indeed, with the male speaker on the Iowa laserdisc recordings of those 24 consonants, this subject has achieved block scores of 119 out of 120 on more than a few occasions. His scores are only somewhat lower with the female speaker.

Similarly, our better subjects score at or near 100% correct on our medial vowel test (which includes 8 vowels, see QPR 4).

We need ~~to address this problem of test sensitivity~~ before many more subjects are studied. Among the options under consideration are the following:

- Use of more difficult open-set tests, such as the low context SPIN sentences [Bilger et al., 1984; Kalikow et al., 1977], the low context IEEE/Harvard sentences [IEEE, 1969], or the zero context nonsense sentences recently developed by Bill Rabinowitz and coworkers at MIT (Rabinowitz, personal communications to Wilson and Lawson, spring, 1991).
- Use of materials from the battery of relatively difficult tests recently developed by Sig Soli and coworkers at the HEI [Nilsson, Sullivan and Soli, 1990] for hearing aid assessment. This battery is designed to detect differences among aids for listeners with mild-to-moderate hearing losses.
- Use of CV syllables for the consonant identification test, rather than VCV syllables, to deprive the listener of any transitional cues from the terminal part of a leading vowel into a following consonant.
- Use of more than one vowel context for either the CV or VCV syllables (our 24 consonant test, for instance, now supports /i/-C-/i/ as well as /a/-C-/a/ testing with sound alone).
- Use of synthetic steady-state vowels, such as the test of 15 synthetic vowels recently developed by Michael Dorman (Dorman, personal communication to Wilson, September, 1991).
- Use of more than 8 vowels in tests of vowel identification, with either synthetic or natural tokens.
- Use of the MAC segmental test recordings with alternative sets of foils chosen to probe more subtle distinctions.

We plan to evaluate at least several of these options in pilot studies with one or more of our better subjects. The results should provide guidance on the selection of more sensitive tests for future studies, or indicate the need to develop new tests for those studies. Suggestions from readers of this report on other options would be welcomed.

VIII. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Initiation of studies with Ineraid subject SR10, to evaluate CIS processors with another patient who has relatively poor performance with his clinical CA processor (week starting November 18).
2. Preparation for further studies of speech reception in noise and of noise reduction, including completion of software for controlled presentations of consonant tokens in reproducible noise contexts and software for real-time implementations of noise reduction algorithms.
3. Visit by consultant Bob Shannon on December 9 and 10, to review software and procedures we have developed for various psychophysical studies, including channel ranking and frequency discrimination and scaling.
4. Further development and refinement of that software and those procedures, according to Dr. Shannon's recommendations.
5. Continued analysis of data from all subjects in our Ineraid series, to evaluate effects of single parameter changes on the performance of CIS processors.
6. Continued preparation of manuscripts for publication.

IX. Footnotes and References

[1] Typical patients have limits between 200 and 500 Hz [Eddington et al., 1978; Müller, 1983; Shannon, 1983; Simmons, 1966; Tong et al., 1983] and exceptional patients have limits in the region of 1 kHz [Hochmair-Desoyer et al., 1983; Townshend et al., 1987].

Bess, F.H., and Townsend, T.H. (**1977**). "Word discrimination for listeners with flat sensorineural hearing losses," J. Speech Hear. Disorders **42**, 232-237.

Bilger, R.C., Nuetzel, J.M., Rabinowitz, W.M., and Rzeczkowski, C. (**1984**). "Standardization of a test of speech perception in noise," J. Speech Hear. Res. **27**, 32-48.

CHABA (**1991**). "Speech-perception aids for hearing-impaired people: Current status and needed research," Committee on Hearing and Bioacoustics -- Working Group on Communication Aids for the Hearing Impaired, J. Acoust. Soc. Am. **90**, 637-684.

Dillier, N. (**1991**). "Speech encoding strategies for multielectrode cochlear implants: A digital signal processor approach," International Symposium on Natural and Artificial Control of Hearing and Balance, Rheinfelden, Switzerland, September 4-7.

Dorman, M.F., Soli, S., Dankowski, K., Smith, L.M., McCandless, G., and Parkin, J. (**1990**). "Acoustic cues for consonant identification by patients who use the Ineraid cochlear implant," J. Acoust. Soc. Am. **88**, 2074-2079.

Dubno, J.R., and Dirks, D.D. (**1982**). "Evaluation of hearing-impaired listeners using a nonsense syllable test. I. Test reliability," J. Speech Hear. Res. **25**, 135-141.

Eddington, D.K., Dobelle, W.H., Brackmann, D.E., Mladejovsky, M.G., and Parkin, J.L. (**1978**). "Auditory prosthesis research with multiple channel intracochlear stimulation in man," Ann. Otol. Rhinol. Laryngol. **87**, Suppl. 53, 1-39.

Gantz, B.J., McCabe, B.F., Tyler, R.S., and Preece, J.P. (**1987**). "Evaluation of four cochlear implant designs," Ann. Otol. Rhinol. Laryngol. **96**, Suppl. 128, 145-147.

Hochmair-Desoyer, I.J., Hochmair, E.S., Burian, K., and Stiglbrunner, H.K. (**1983**). "Percepts from the Vienna cochlear prosthesis," Ann. N.Y. Acad. Sci. **405**, 295-306.

IEEE (**1969**). "IEEE recommended practice for speech quality measurements," Institute of Electrical and Electronic Engineers, Inc.

Kalikow, D.N., Stevens, K.N., and Elliott, L.L. (**1977**). "Development of a test of speech intelligibility in noise using sentence materials with controlled word predictability," J. Acoust. Soc. Am. **61**, 1337-1351.

McDermott, H., McKay, C., Vandali, A., and Blamey, P. (**1991**). "An improved sound processor for the University of Melbourne/Nucleus multielectrode cochlear implant," Poster Presentation at the International Symposium on Natural and Artificial Control of Hearing and Balance, Rheinfelden, Switzerland, September 4-7.

McKay, C.M., and McDermott, H.J. (**1991**). "Speech perception ability of adults with multiple-channel cochlear implants, using the spectral maxima sound processor," Poster Presentation at the 121st Meeting of the Acoustical Society of America, Baltimore, MD, April 29 to May 3 (abstract published in J. Acoust. Soc. Am. **89**, Suppl. 1, 1959).

McKay, C., McDermott, H., Vandali, A., and Clark, G.M. (**1991**). "Preliminary results with a six spectral maxima sound processor for the University of Melbourne/Nucleus multiple-electrode cochlear implant," J. Otolaryngol. Soc. Austral. **6**, 354-359.

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

Moore, B.C.J. (**1991**). "Comodulation masking release and across-channel masking," Workshop on the Psychophysics of Speech Perception -- II, Utrecht, Netherlands, July 1-5.

Müller, C.G. (**1983**). "Comparison of percepts found with cochlear implant devices," Ann. N.Y. Acad. Sci. **405**, 412-420.

Nilsson, M., Sullivan, J., and Soli, S.D. (**1990**). "Development of a speech intelligibility test for hearing aid research," J. Acoust. Soc. Am. **88**, S175.

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.

Shannon, R.V. (**1983**). "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hear. Res. **11**, 157-189.

Simmons, F.B. (**1966**). "Electrical stimulation of the auditory nerve in man," Arch. Otolaryngol. **84**, 2-54.

Skinner, M.W., Holden, L.K., Holden, T.A., Dowell, R.C., Seligman, P.M., Brimacombe, J.A., and Beiter, A.L. (**1991**). "Performance of postlinguistically deaf adults with the Wearable Speech Processor (WSP III) and Mini Speech Processor (MSP) of the Nucleus multi-electrode cochlear implant," Ear Hear. **12**, 3-22.

Tong, Y.C., Blamey, P.J., Dowell, R.C., and Clark, G.M. (**1983**). "Psychophysical studies evaluating the feasibility of a speech processing strategy for a multiple-channel cochlear implant," J. Acoust. Soc. Am. **74**, 73-80.

Townshend, B., Cotter, N., Van Campernolle, D., and White, R.L. (**1987**). "Pitch perception by cochlear implant patients," J. Acoust. Soc. Am. **87**, 106-115.

Tye-Murray, N., and Tyler, R.S. (**1989**). "Auditory consonant and word recognition skills of cochlear implant users," Ear Hear. **10**, 292-298.

Tyler, R.S. (**1990**). "What should be implemented in future cochlear implants?," Acta Otolaryngol. Suppl. **469**, 268-275.

Tyler, R.S., and Kuk, F.K. (**1989**). "The effects of 'noise suppression' hearing aids on consonant recognition in speech-babble and low-frequency noise," Ear Hear. **10**, 243-249.

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.

Van Tasell, D.J., Larsen, S.Y., and Fabry, D.A. (**1988**). "Effects of an adaptive filter hearing aid on speech recognition in noise by hearing-impaired subjects," Ear Hear. **9**, 15-21.

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. (**1988**). "Comparative studies of speech processing strategies for cochlear implants," Laryngoscope **98**, 1069-1077.

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.

Appendix 1

Summary of Reporting Activity for the Period of

August 1 through October 31, 1991

NIH Project N01-DC-9-2401

Reporting activity for the last quarter included three invited lectures, listed below. The abstract for the first lecture is reproduced on the next page.

Wilson, BS, DT Lawson and CC Finley (1991). A new processing strategy for multichannel cochlear prostheses. *International Symposium on Natural and Artificial Nervous Control of Hearing and Balance*, Rheinfelden, Switzerland, September 4-8. [Presented by DT Lawson]

Wilson, BS (1991). New coding strategy for cochlear implants. *American Neurotology Society*, Kansas City, MO, September 21.

Wilson, BS (1991). Speech processors for auditory prostheses. *Neural Prosthesis Workshop*, National Institutes of Health, Bethesda, MD, October 22-24.

41 SPEECH ENCODING STRATEGIES FOR MULTIELECTRODE COCHLEAR IMPLANTS: A DIGITAL SIGNAL PROCESSOR APPROACH

N. DILLIER, Department of Otorhinolaryngology, University Hospital, Zürich, Switzerland

Hearing sensations can be restored for profoundly deaf patients via artificial electrical stimulation of the auditory nerve. Present electrode technology and electrophysiological constraints however allow, at best, a very crude and limited approximation of the normal neural excitation pattern. Signal processing for cochlear implants therefore, is confronted with the problem of a severely restricted channel capacity and the necessity to select and encode a subset of the information contained in the sound signal reaching the listener's ear.

The search for new signal processing schemes has to consider the specific perceptual attributes of various electrical stimulation waveforms and patterns. In order to convey the contents of a particular transmitted message, its primary information elements (e.g. phonetic or acoustic speech features) should be transformed into those physical stimulation parameters which can generate distinctive perceptions for the listener.

With single-chip digital signal processors (DSP's) incorporated in personal computers, different speech coding strategies can be evaluated in relatively short laboratory experiments. In contrast to conventional electronic signal processing with filters, amplifiers and logic circuits; a DSP approach allows the implementation of much more complex algorithms such as nonlinear multiband loudness correction, speech feature contrast enhancement, and adaptive noise reduction. Although many aspects of speech encoding can be efficiently studied using a laboratory digital signal processor it would be desirable to allow subjects more time for adjustment to a new coding strategy. Several days or weeks of habituation are sometimes required until a new mapping can be fully exploited. Thus, for scientific as well as practical purposes the miniaturization of wearable DSP's will be of great importance.

Two different processing strategies have been implemented on an experimental DSP-system: The first approach (PES, Pitch Excited Sampler) is based on the classical channel vocoder concept, the second approach (CIS, Continous Interleaved Sampler) uses a stimulation pulse rate which is independent of the input signal. The algorithm scans continuously all frequency bands and samples their energy levels. To date, evaluation experiments have been conducted with 5 patients. Performance in consonant identification tests was significantly better with the new processing strategies than with the patient's own wearable speech processors, whereas improvements in vowel identification tasks were less obvious. Several modifications of the basic PES- and CIS-strategies were tested, resulting in large variations of identification scores. Information transmission analysis of confusion matrices revealed a rather complex pattern across conditions and speech features. Optimization and fine-tuning of processing parameters for these coding strategies requires more data both from speech identification and discrimination, as well as psychophysical experiments.

42 A NEW PROCESSING STRATEGY FOR MULTICHANNEL COCHLEAR PROSTHESES

B.S. WILSON[1,2], D.T. LAWSON[1] and C.C. FINLEY,[1,2], [1]Neuroscience Program, Research Triangle Institute, Research Triangle Park, NC 27709, USA; [2]Division of Otolaryngology, Duke University Medical Center, Durham, NC 27710, USA

Various strategies for representing speech information with multichannel cochlear prostheses were compared in tests with implant patients. The strategies included the *compressed analog* (CA) approach of a standard clinical device, and the alternative *interleaved pulses* (IP) and *continous interleaved sampling* (CIS) strategies. The CA and IP strategies had been compared in previous studies with a wide range of subjects. The present studies compared all three strategies in tests with one subject and the CA and CIS strategies in tests with six additional subjects. The subjects for the present studies were selected for their excellent performance with the clinical CA processor, and the tests included closed-set identification of consonants and open-set recognition of words, sentences, and paragraph mâterial. For every test, every subject obtained the highest score, or repeated a score of 100 % correct, using the CIS strategy. In the comparisions of all three strategies, the IP processor produced scores between those of the CA and CIS processors. These differences in results are discussed in terms of their implications for processor design.

Appendix 2

Tables of Feature Assignments for 24 Consonants and 8 Vowels

TABLE A.2.1. 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
r	2	1	1	1	4	4	1
w	2	1	1	1	1	4	1
j	2	1	1	1	4	4	1
ŋ	2	2	1	1	5	4	1
h	1	1	2	1	5	3	1
ʒ	2	1	2	2	4	2	1
θ	1	1	2	1	2	3	1
tʃ	1	1	1	1	4	2	2

TABLE A.2.2. 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