Third Quarterly Progress Report

February 1 through April 30, 1993

NIH Contract N01-DC-2-2401

Speech Processors for Auditory Prostheses

Prepared by

Blake S. Wilson, Mariangeli Zerbi and Dewey T. Lawson

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

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

CONTENTS

I. Introduction

The purpose of this project is to design and evaluate speech processors for 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 *virtual channel interleaved sampling* (VCIS) processors, including (a) evaluation of different types of VCIS processors with reduced numbers of electrodes, to reduce scores on the 24 consonant test into a range of sensitivity for Ineraid subject SR2; (b) evaluation of "sharpened field" processors, in which small reversed-phase pulses are presented simultaneously on adjacent electrodes for a principal pulse on the central electrode for each channel; and (c) preliminary evaluation of the VCIS approach with a subject who does not enjoy high levels of clinical performance with his Ineraid *compressed analog* (CA) processor (subject SR10).
2. Psychophysical studies relating to the identification of virtual channel conditions on the basis of pitch. These studies were conducted with Ineraid subject SR2.
3. Development of new software for support of the psychophysical studies.
4. Continued collaborative studies with Michael Dorman to evaluate representations of frequency cues through CA, *continuous interleaved sampling* (CIS) and VCIS processors. The studies also now include measures of vowel identification with each type of processor, using 12 synthesized vowels that are identical in all respects except for differences in formant frequencies.
5. Preliminary studies of the representations of complex tones by CIS and VCIS processors. These studies were conducted with Ineraid subject SR2.
6. Continued analysis of data from prior and current studies, to evaluate effects of single parameter changes on the performance of CIS processors.
7. Invited participation in a "mini symposium" at the House Ear Institute on envelope representations with cochlear implants (Los Angeles, CA, February 25-28).
8. Presentation of project results in invited lectures at the *1993 Cherry Blossom Conference: Current and New Applications in Hearing and Equilibrium* (Washington, D.C., April 2-4) and at the *Third International Cochlear Implant Conference* (Innsbruck, Austria, April 4-7).
9. Continued transfer of CIS processor technology to the Innsbruck team, in a visit to North Carolina by three engineers from the Innsbruck team, and in a subsequent visit by Wilson to the Innsbruck laboratory, following the *Third CICI.*
10. Continued preparation of manuscripts for publication.

In this report we present results from the psychophysical studies relating to the identification of virtual channel conditions on the basis of pitch (point 2 above). In addition, our paper for the Innsbruck conference, "Recent Developments with the CIS Strategies," is presented in Appendix 2. Work indicated in points 1, 4, 5 and 6 above will be described in future reports. Also, we note that several of these areas will be addressed in invited lectures at the *1993 Conference on Implantable Auditory Prostheses*, to be held in Smithfield, RI, July 11-15. Results from work indicated in points 1, 4 and 5 will be included in lectures by Blake Wilson, Michael Dorman and Dewey Lawson, respectively.

II. Identification of Virtual Channel Conditions on the Basis of Pitch

The concept of virtual channels was introduced in the First Quarterly Progress Report for this project. Briefly, virtual channels involve simultaneous stimulation of two or more electrodes and may be used in *virtual channel interleaved sampling* (VCIS) processors to provide pitch percepts that are different from those elicited by stimulation of single electrodes. VCIS processors thereby offer the possibility of increasing the number of effective channels beyond the number of available electrodes.

The construction of various types of virtual channels is illustrated in Fig. 1. The top curve in each panel is a hypothetical sketch of the number of neural responses, as a function of position along the cochlea, for a given condition of stimulation. The condition of stimulation is indicated by the pulse waveform(s) below each dot, which represent the positions of three adjacent electrodes. Conditions involving stimulation of one electrode only are shown in panels *a* and *b*, and conditions involving simultaneous stimulation of more than one electrode are shown in panels *c* through *g*.

As indicated in our prior report, Ineraid subject SR2 has reported that many of these conditions can be ranked according to their distinct pitches. In informal listening tests stimulation of apicalmost electrode 1 alone (*a*) produced a low pitch, whereas stimulation of electrode 2 alone (*b*) produced a higher pitch. Simultaneous stimulation of both electrodes, with identical pulses having approximately half the amplitude of the single-pulse conditions (*c*), produced an intermediate pitch. Pairing stimulation of electrode 1 with a reversed-polarity pulse on electrode 2 (*d*) produced the lowest pitch among the illustrated conditions. Various lower pitches could be produced by manipulating the ratio of the pulse 2 to pulse 1 amplitudes over the range of 0.2 to 0.8. A ratio of 1.0 produced a pitch higher than that elicited by stimulation of electrode 1 alone. Similarly, pitches higher than that elicited by stimulation of electrode 6 alone (the basalmost electrode in the Ineraid array) could be produced by presenting a reversed-phase pulse (of lower amplitude) on electrode 5. Additional pitches between electrodes could be produced by constructing triads of pulses, as illustrated in *e* and *f*.

Condition *g* in Fig. 1 suggests a way in which the width of a neural excitation field might be reduced without altering the centroid or peak of the field, by supplying reversed-phase pulses on either side of a primary pulse. Subject SR2 reported that the pitch percept of this condition was indistinguishable from that of condition *b*.

Stimuli used for the informal listening tests consisted of 50 ms bursts of 33 μs/phase pulses, presented at 1364 pulses per second (pps). The duration and rate of the pulses are typical of those used in VCIS processors. The stimuli were presented at current amplitudes corresponding to most comfortable loudness (MCL) for each condition.

Although the informal listening tests indicated apparent differences in pitch percepts across conditions, we wanted to evaluate the anecdotal reports in a formal test. In this way we could rank the conditions in an order of ascending pitch and estimate the magnitude and reliability of pitch differences between conditions producing adjacent pitches.

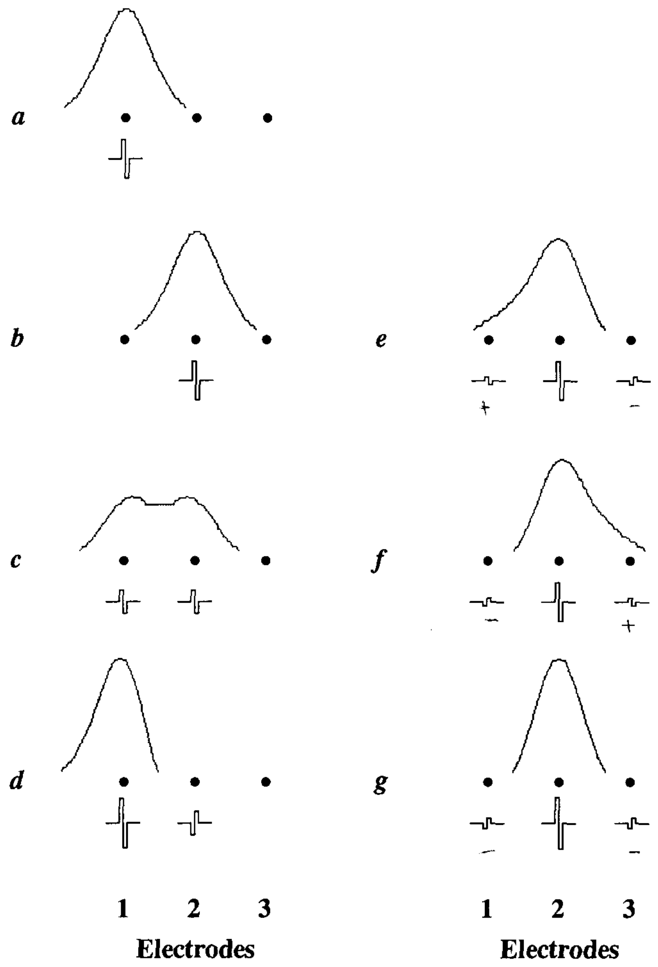

Fig. 1. Conditions of single-electrode and multiple-electrode (virtual channel) stimulation. See text for a description of the different types of information presented in each panel.

Methods

An identification experiment was conducted, with each of the seven stimulus conditions in Fig. 1 presented 30 times in a randomized order. The subject was asked to nominate a pitch from 1 to 7 following each trail, with a nomination of 1 corresponding to the lowest pitch among the conditions and a nomination of 7 corresponding to the highest pitch among conditions. Condition *d* was presented prior to the formal test as an example of a low pitch among conditions, and condition *f* was presented prior to the formal test as an example of a high pitch among conditions. No feedback was provided during the formal test. As before, all conditions were presented at MCL, with at least an approximate balance of loudness across conditions (MCL judgments are highly repeatable for this subject). The amplitude ratio of pulses for condition *d* was 0.5 and the amplitude ratio of flanking pulses to the central pulse for conditions *e* through *g* was 0.2. Identical amplitudes were used for the two pulses of condition *c*.

Results

The results are presented in Fig. 2, which shows the ranking of conditions in the leftmost column and histograms of pitch nominations in the adjacent column. The mean and standard deviation of the pitch nominations for each condition are shown in the remaining columns.

Note that the overall ranking of conditions is consistent with SR2's anecdotal reports. In particular, the mean of pitch nominations is lowest for condition *d*; the mean for condition *c* lies between the means for conditions *a* and *b*; the means for conditions *e* and *f* are lower than and higher than the mean for condition *b*, respectively; and the mean of condition *g* is adjacent to the mean of condition *b*. Note also that the variance of pitch nominations is not reduced with the use of reversed-phase flanking pulses on either side of a principal pulse (compare standard deviations for conditions *g* and *b*).

To evaluate the significance of differences among conditions, a one-way analysis of the variance (ANOVA) was conducted using the pitch nomination data. The result indicated that at least some of the means were different from each other ($F[6,203] = 87.5$; $p < .00000001$). Following the ANOVA, a *post hoc* analysis was conducted with Tukey's HSD statistic. The HSD (the "honestly significant difference" between any pair of means) for the $p = .05$ level was 0.78, and the HSD for the $p = .01$ level was 0.91. Table 1 shows a matrix of the differences in means among the tested conditions. Note that half of the adjacent conditions (*d* and *a*, *a* and *c*, and *b* and *f*) are statistically different, at $p = .01$, according to this (conservative) analysis. All of the means separated by one position in the ranking are different (compare conditions *c* and *g*, *e* and *b*, and *g* and *f*). A final observation from Table 1 is that the means of conditions *c* and *e*, *e* and *g*, and *g* and *b* are similar, i.e., the differences in means for these conditions are not statistically significant.

Discussion

These results indicate that percepts elicited by virtual channels may be separable from those elicited by single-electrode stimulation at adjacent positions in the pitch ranking order. Consistent with SR2's anecdotal reports, percepts elicited by conditions *a*, *c* and *b* form a continuum, with distinct increases in pitch for the order *a*, *c*, *b*. The difference in means between *a* and *c* was 1.06, and the difference in means between *c* and *b* was 1.50.

Similarly, pitch may be shifted with triads of pulses. Condition *e* produces a significantly lower pitch than condition *b*, and condition *f* produces a significantly higher pitch than condition *b*. The difference in means between *e* and *b* was 0.80, and the difference between *b* and *f* was 1.57.

As noted above, conditions *g* and *b* produced similar pitches, again consistent with SR2's anecdotal reports. However, use of reversed-phase flanking pulses in condition *g* did not appear to reduce the variance of pitch judgments, as might have been expected from a sharpening of the neural excitation field.

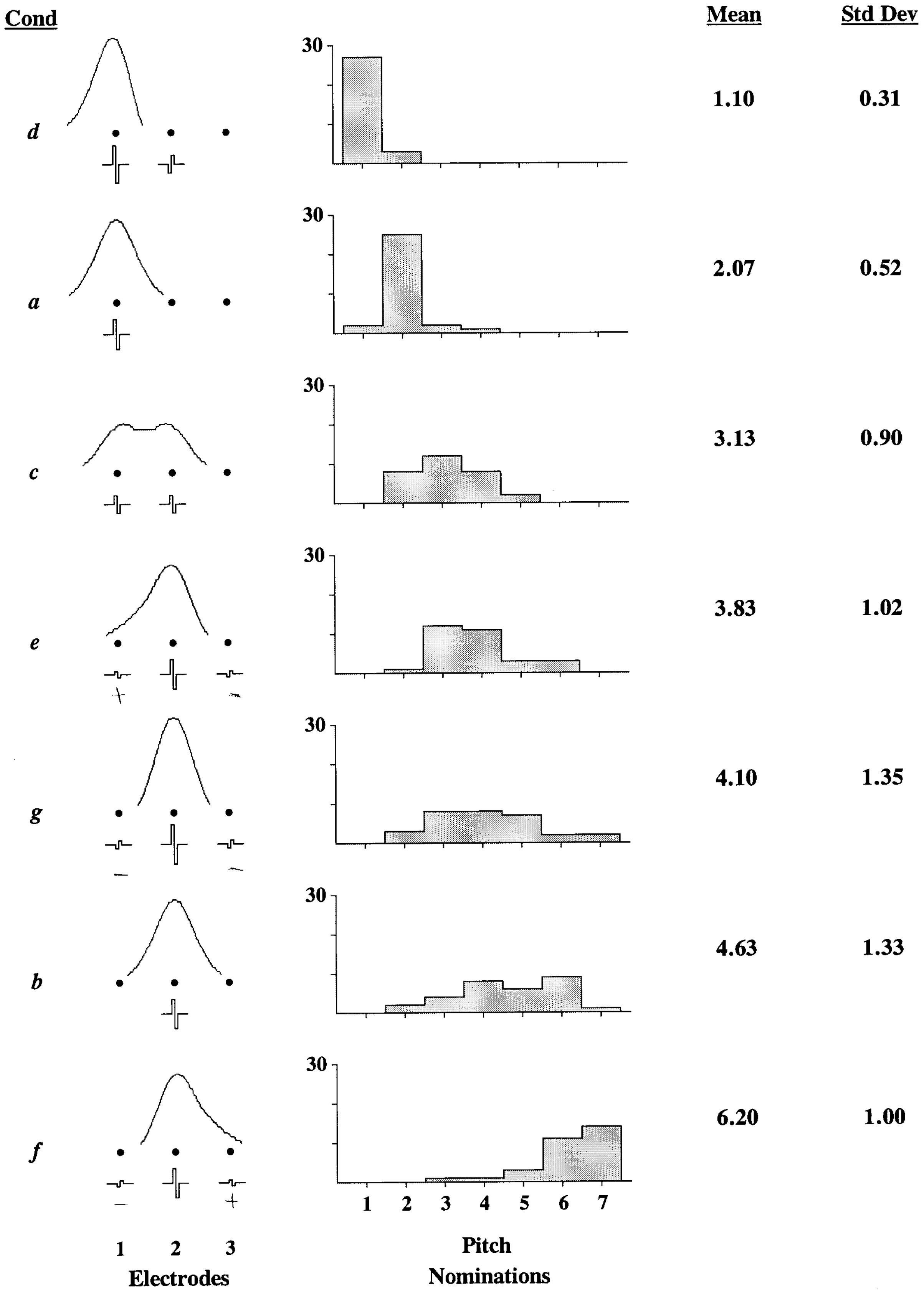

Fig. 2. Results from identification experiment. See text for details.

TABLE 1. Differences in means of pitch nominations among virtual channel conditions (see Fig. 2). Differences greater than or equal to 0.78 are significant at the p = .05 level, and differences greater than or equal to 0.91 are significant at the p = .01 level.

	d	*a*	*c*	*e*	*g*	*b*
a	0.97					
c	2.03	1.06				
e	2.73	1.76	0.70			
g	3.00	2.03	0.97	0.27		
b	3.53	2.56	1.50	0.80	0.53	
f	5.10	4.13	3.07	2.37	2.10	1.57

In general, the present results demonstrate the feasibility of the VCIS approach at a basic psychophysical level. A challenge for future work is to examine possibilities for constructing virtual channels that will (a) have a greater degree of perceptual independence from adjacent channels and (b) produce equal steps in pitch across channels in a VCIS processor. Use of conditions with a small variance in pitch judgments, for instance, may allow the addition of independent channels. Also, fine adjustment of flanking pulse amplitudes may be helpful in shifting pitches so that equal steps across channels can be approximated. We plan to evaluate several such possibilities in the next quarter, using identification measures like those presented in this report, along with measures of d' and cumulated d' (see, e.g., Braida and Durlach, 1972; Tong and Clark, 1985) to map perceptual distances across conditions.

Acknowledgement

We thank subject SR2 for his generous contribution of time.

References

Braida LD, Durlach NI (1972). Intensity perception II. Resolution in one interval paradigms. *J. Acoust. Soc. Am.* 51: 483-502.

Tong YC, Clark GM (1985). Absolute identification of electric pulse rates and electrode positions by cochlear implant patients. *J. Acoust. Soc. Am.* 77: 1881-1888.

III. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Continued studies with VCIS processors, including additional speech reception studies with Ineraid subject SR10 and additional psychophysical studies with Ineraid subject SR2.
2. Continued studies on the representation of complex tones by CIS and VCIS processors, with Ineraid subject SR2.
3. Continued studies on representations of frequency cues with CA, CIS and VCIS processors, in collaboration with Michael Dorman.
4. Presentation of project results in several invited lectures at the *1993 Conference on Implantable Auditory Prostheses*, to be held in Smithfield, RI, July 11-15.
5. Continued preparation of manuscripts for publication.

Appendix 1

Summary of Reporting Activity for the Period of

February 1 through April 30, 1993

NIH Project N01-DC-2-2401

Reporting activity for the last quarter included publication of one paper, acceptance of another, and presentations of invited lectures and comments at several conferences and symposia. Citations are listed below and one the papers is reproduced in Appendix 2.

Papers

Wilson BS, Lawson DT, Finley CC, Wolford RD: Importance of patient and processor variables in determining outcomes with cochlear implants. *Journal of Speech and Hearing Research* 36: 373-379, 1993.

Wilson BS, Lawson DT, Zerbi M, Finley CC: Recent developments with the CIS strategies. To appear in the Proceedings for the *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7, 1993 (10 pp).

Presentations

Wilson BS: Representations of envelope information with CIS and VCIS processors. Invited comments at a "mini symposium" on envelope representations with cochlear implants, House Ear Institute, Los Angeles, CA, February 25-28, 1993.

Wilson BS: Optimizing performance with new speech processing strategies. Invited lecture presented at the *1993 Cherry Blossom Conference: Current and New Applications in Hearing and Equilibrium*, American Academy of Otolaryngology -- Head & Neck Surgery, Washington, D.C., April 2, 1993.

Wilson BS: Recent developments with the CIS strategies. Invited lecture presented at the *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7, 1993.

Wilson BS and Dillier N: Co-Chairs, session on speech coding. *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7, 1993.

Appendix 2

Preprint of "Recent developments with the CIS strategies," to appear in the Proceedings for the *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7, 1993.

Recent Developments with the CIS Strategies

Blake S. Wilson, Dewey T. Lawson, Mariangeli Zerbi and Charles C. Finley
Neuroscience Program, Research Triangle Institute
PO Box 12194, Research Triangle Park, NC 27709, USA

Abstract - **Studies with continuous interleaved sampling (CIS) and related processing strategies for multichannel cochlear implants are reviewed, with an emphasis on recent studies. The studies include (a) within-subject comparisons of CIS and compressed analog (CA) processors; (b) evaluations of performance in noise for CIS and CA processors; (c) evaluations of effects of changes in parameters for CIS processors, such as changes in the number of bandpass and electrode channels; and (d) design and preliminary evaluation of one of the related strategies, virtual channel interleaved sampling (VCIS).**

I. INTRODUCTION

The design of continuous interleaved sampling (CIS) processors for electrical stimulation of the auditory nerve or auditory brainstem is illustrated in Figs. 1 and 2. Each CIS channel includes a bandpass filter and an envelope detector. The amplitudes of stimulus pulses are determined with a logarithmic or power-law transformation of the envelope signal. The corner (or cutoff) frequencies of the bandpass filters span the frequency range from 350 to 5500 Hz, evenly spaced along a logarithmic scale. In typical implementations five or six channels are used, with 400 Hz lowpass smoothing filters in the envelope detectors. The information represented by the implant thus consists of envelope variations, below 400 Hz, in each of five or six frequency bands of speech.

II. COMPARISONS OF CIS AND CA PROCESSORS

Performances of CIS and compressed analog (CA) processors have been compared in tests with 11 users of the Ineraid cochlear prosthesis. Seven of the subjects were selected for their high levels of speech recognition with the Ineraid CA processor and four were selected for their relatively poor performances with that processor. The "high performance" subjects were representative of the best results, in terms of speech recognition scores, obtained with any commercially-available implant system [1]. The purpose of this section is to provide a summary of results for both sets of subjects.

A. Processors

Waveforms of a CA processor are presented in Fig. 3. A microphone input signal is compressed or restricted to the narrow dynamic range of electrically-evoked hearing using an automatic gain control. The resulting signal then is filtered into four contiguous bands for presentation to each of four electrodes. As shown in the figure, information about speech sounds is contained in the relative stimulus amplitudes among the four electrode channels and in the temporal details of the waveforms for each channel.

A concern associated with this method of presenting information is that substantial parts of it may not be perceived by implant patients. For example, most patients cannot perceive frequency changes in stimulus waveforms above about 300 Hz (see, e.g., [2]). Thus, many of the temporal details present in CA stimuli are not likely to be accessible to the typical user.

In addition, the simultaneous presentation of stimuli may produce significant interactions among channels through vector summation of the electric fields from each electrode [3], [4]. The resulting degradation of channel independence would be expected to reduce the salience of channel-related cues. That is, the neural response to stimuli from one electrode may be significantly distorted, or even counteracted, by coincident stimuli from other electrodes.

The CIS approach addresses the problem of such channel interactions through the use of interleaved nonsimultaneous stimuli (Fig. 2*c*). Trains of balanced biphasic pulses are delivered to each electrode with temporal offsets that eliminate any overlap across channels. The amplitudes of the pulses are derived from the envelopes of bandpass filter outputs. In contrast to the four-channel clinical CA processors, five or six bandpass filters (and channels of stimulation) generally have been used in CIS systems to take advantage of additional implanted electrodes and reduced interactions among channels.

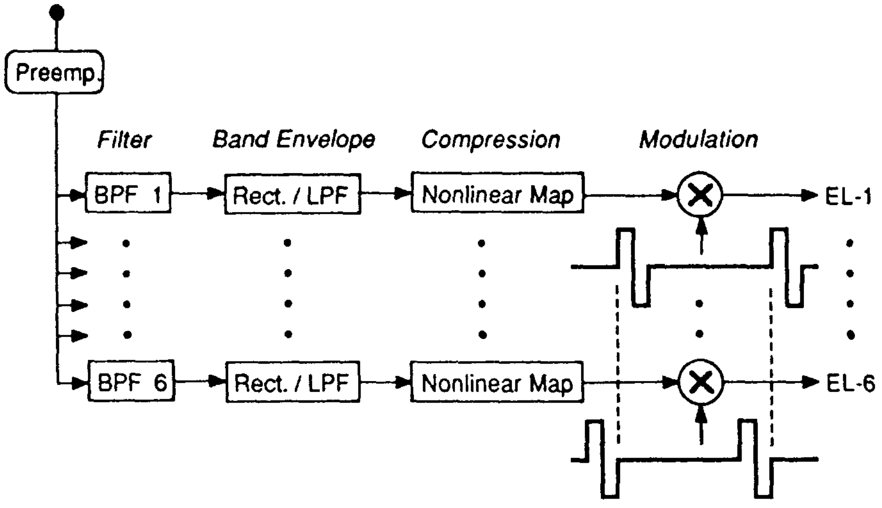

Figure 1. Block diagram of a CIS processor. A preemphasis filter (Preemp.) attenuates strong low frequency components in the speech input signal that otherwise might mask important high frequency components. The preemphasis filter is followed by six channels of processing. Each channel includes stages of bandpass filtering (BPF), envelope detection, compression, and modulation. The envelope detector consists of a rectifier (Rect.) followed by a lowpass filter (LPF). Carrier waveforms for two of the modulators are shown immediately below the two corresponding multiplier blocks. The outputs are directed to electrodes EL-1 through EL-6.

An important feature of the CIS approach is a relatively high rate of stimulation on each channel. Other pulsatile strategies present sequences of interleaved pulses across electrodes at a rate equal to the estimated fundamental frequency during voiced speech and at a jittered or fixed (often higher) rate during unvoiced speech [5]-[7]. Rates of stimulation on any one channel rarely have exceeded 300 pulses per second (pps). In contrast, CIS processors generally use brief pulses and minimal delays, so that rapid variations in speech can be tracked by pulse amplitude variations. The rate of stimulation on each channel usually exceeds 500 pps and is constant during both voiced and unvoiced intervals.

B. Methods

Each subject has been studied for a one-week period during 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 all subjects. In contrast, experience with CIS processors was limited to no more than several hours before formal testing. All comparisons within this eleven-subject study are on the basis of a single week of CIS optimization. In subsequent visits by some of the same subjects a potential for significant further optimization has been demonstrated.

1) Tests: The comparison tests included open-set 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 "high predictability" sentences from the Speech Perception in Noise (SPIN) test (presented in our studies without noise). All tests were conducted with hearing alone, using single presentations of recorded material, and without feedback as to correct or incorrect responses.

2) Processor parameters: Each subject's own clinical device was used for the tests with the CA processor. As mentioned above, selection of parameters for the CIS processor was guided by preliminary tests of consonant identification. The standard four channels of stimulation were used for the clinical CA processors [8], whereas five or six channels were used for the CIS processors. All CIS processors for the "high performance" subjects, SR2-8, had pulse durations of 102 μs/phase or less, zero delay between the sequential pulses on different channels, pulse rates of 817 pps or higher on each channel, and a cutoff frequency for the lowpass filters of 400 Hz or higher. The best processor for subject SR1 also fit this description, except that a delay of 172 μs was interposed between sequential 34 μs/phase pulses. The best processors for subjects SR9-11 used long-duration pulses (167 μs/phase),

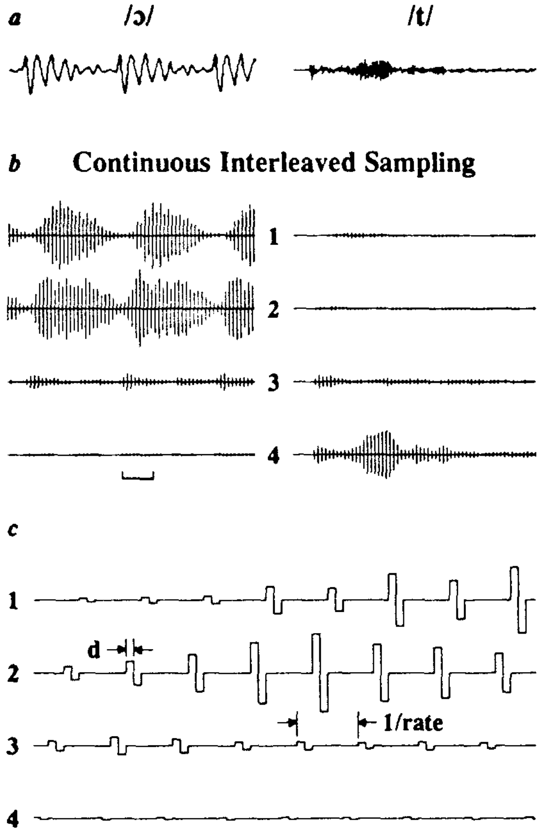

Figure 2. Waveforms of a four-channel CIS processor. *a*, Preemphasized (6 dB/octave attenuation below 1.2 kHz) speech inputs. Inputs corresponding to a voiced speech sound ('aw') and an unvoiced speech sound ('t') are shown in the left and right columns, respectively. *b*, Stimulus waveforms. The waveforms are numbered by channel, with channel 1 delivering its output to the apicalmost electrode. The pulse amplitudes in the illustration reflect the envelope of the bandpass output for each channel. In actual implementations the range of pulse amplitudes is compressed using a logarithmic or power-law transformation of the envelope signal. *c*, Expanded display of CIS waveforms (from the bracketed interval in *b*). Pulse duration per phase ('d') and the period between pulses on each channel ('1/rate') are indicated. The sequence of stimulated channels is 4-3-2-1. [The duration of each trace is 25.4 ms in *a* and *b*, and 3.3 ms in *c*.]

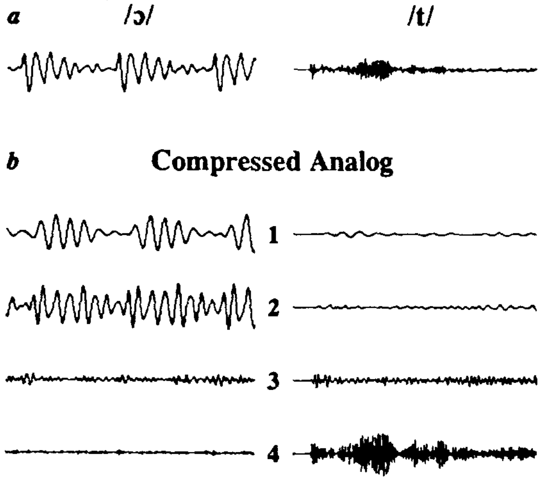

Figure 3. Waveforms of a four-channel CA processor. *a*, Preemphasized speech inputs. *b*, Stimulus waveforms.

paired with a relatively low rate of stimulation on each channel (500 pps) and a relatively low cutoff frequency for the lowpass filters (200 Hz).

3) Evaluation of practice and learning effects: Because the tests with the CA processor preceded those with the selected CIS processor for each subject, we were concerned that practice or learning effects might favor the latter in comparisons of the two strategies. To evaluate this possibility, the CID and NU-6 tests were repeated with the CIS processor for five of the "high performance" subjects (subjects SR3, SR4 and SR6-8), using a different recorded speaker and new lists of words and sentences. Practice or learning effects would be demonstrated by significant differences in the test/retest scores. However, no such differences were found ($p > 0.6$ for paired *t* comparisons of the CID scores; $p > 0.2$ for the NU-6 scores), and the scores from the first and second tests were averaged for all subsequent analyses.

C. Results

The results from one-week studies of each of the 11 subjects are presented in Table 1 and Fig. 4. CA and CIS scores for each of the "high performance" subjects are connected by the thinner lines near the top of each panel in Fig. 4, and scores for the four "low performance" subjects are connected by the thicker lines closer to the bottom of each panel. We note that low-performance subject SR1 had participated in an earlier study not involving CIS processors [6]. Results from his first week of testing with CIS processors are presented here. This is also true of high-performance subject SR2, who has returned to the laboratory for many additional studies with various implementations of CIS processors (see, e.g., [9]). In those subsequent tests SR2 has achieved even higher scores using a variety of six-channel CIS processors, with NU-6 percentages ranging from the high 80s to the mid 90s.

As is evident from the figure, scores for all 11 subjects are improved with the use of a CIS processor. The average scores across subjects increased from 57 to 80% correct on the spondee test ($p < 0.002$), from 62 to 84% correct on the CID test ($p < 0.005$), from 34 to 65% correct on the SPIN test ($p < 0.001$), and from 30 to 47% correct on the NU-6 test ($p < 0.0005$). Note that the range of difficulty among our four tests provides sensitivity to performance differences across the rather wide range of absolute performance represented in this eleven-subject study.

Perhaps the most encouraging of these results are the improvements for the four low-performance subjects. SR1, for instance, achieved scores with the CIS processor that would have qualified him for membership in the high performance group (with the clinical CA processor). Similarly, SR10 achieved relatively high scores with the CIS processor. 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 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.

Note that while these gains for SR10 are large, they are not atypical of results for the other subjects. His improvements follow the pattern of the other subjects, i.e., generally large gains in the scores of tests that are not limited by ceiling effects. The distinctive aspect of SR10's results is that he enjoys such gains even though he started at or near zero on all four tests. Thus, the relative improvements for SR10 are larger than those for any other subject in the series.

D. Discussion

The findings presented above demonstrate that use of CIS processors can produce large and immediate gains in speech recognition for a wide range of implant patients. Indeed, the sensitivity of some of the administered tests has been limited by ceiling (saturation) effects: five of the seven "high performance" subjects scored 96% or higher for the spondee test using CIS processors; all seven scored 95% or higher for the CID test; and five scored 92% or higher for the SPIN test. Scores for the NU-6 test, while not approaching the ceiling, still were quite high. The 80% score achieved by two of the subjects corresponds to the middle of the range of scores obtained by people with mild-to-moderate hearing losses when taking the same test [10], [11].

The improvements are especially striking when one considers the large disparity in experience with the two types of processor. At the time of our tests each subject had 1 to 5 years of daily experience with the CA processor, but only several hours over a few days with CIS. In previous studies involving within-subjects comparisons, such differences in experience have strongly favored the processor with the greatest duration of use [12]-[14].

Factors contributing to the performance of CIS processors might include (a) reduction in channel interactions through the use of nonsimultaneous stimuli, (b) use of five or six channels instead of four, (c) representation of rapid envelope variations through the use relatively high pulse rates, (d) preservation of amplitude cues with channel-by-channel compression, and (e) the shape of the compression function.

III. EVALUATIONS OF PERFORMANCE IN NOISE FOR CIS AND CA PROCESSORS

In recent studies we have begun systematic investigation of various aspects of CIS performance. These studies

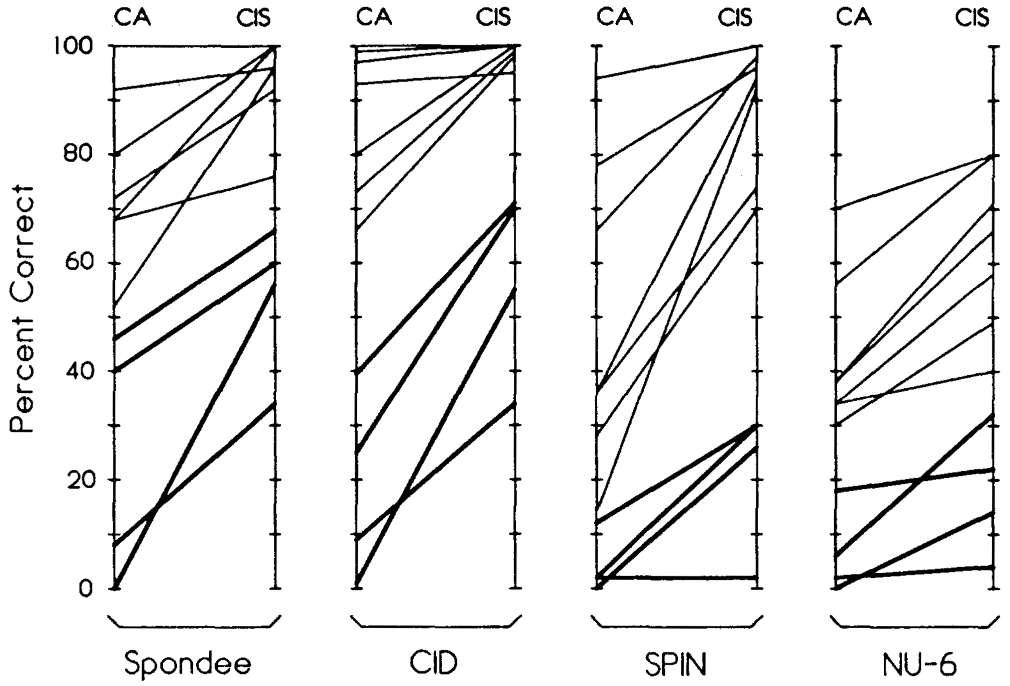

Figure 4. Speech recognition scores for CA and CIS processors. A line connects the CA and CIS scores for each subject. Light lines correspond to the seven subjects selected for their excellent performance with the clinical CA processor, while the heavier lines correspond to the four subjects selected for relatively poor performance.

Table 1
INDIVIDUAL RESULTS FROM THE OPEN-SET TESTS

	Spondee		CID		SPIN		NU-6	
Subject	CA	CIS	CA	CIS	CA	CIS	CA	CIS
SR2	92	96	100	100	78	96	56	80
SR3	52	96	66	98	14	92	34	58
SR4	68	76	93	95	28	70	34	40
SR5	100	100	97	100	94	100	70	80
SR6	72	92	73	99	36	74	30	49
SR7	80	100	99	100	66	98	38	71
SR8	68	100	80	100	36	94	38	66
SR1	40	60	25	70	2	30	6	32
SR9	8	34	9	34	2	2	2	4
SR10	0	56	1	55	0	26	0	14
SR11	46	66	40	71	12	30	18	22

include evaluations of CIS performance (a) under conditions of noise interference, (b) across numbers of channels, and (c) across other manipulations in CIS parameters. The purpose of this section and the next is to present preliminary results from studies (a) and (b).

A. Methods

Both studies were conducted with subject SR2. Because this subject's high scores had limited the sensitivity of our medial consonant and other tests in evaluations of CIS processors with a variety of new parameter sets, we decided to increase the difficulty of the consonant test by increasing the number of consonants from 16 to 24 (the set of 24 includes /b, d, f, g, dʒ, h, j, k, l, m, n, ŋ, p, r, s, ʃ, t, tʃ, ð, θ, v, w, z, ʒ /). Since this change, however, SR2 has achieved scores of 100% correct with the male speaker using several different implementations of CIS processors and scores as high as 97% correct for the female speaker.

While the sensitivity of even the 24 consonant test now is inadequate to distinguish among the better implementations of CIS processors for subject SR2, the test remains well suited to studies exploring decrements in performance due to reduced numbers of channels or increasing amounts of noise.

In the 24 consonant test, multiple exemplars of each consonant were presented in an /a/-consonant-/a/ context from laser videodisc recordings of male and female speakers [15]. A single block of trials consisted of five randomized presentations of each consonant by a single speaker. The tests were conducted with hearing alone and without feedback as to correct or incorrect responses.

B. Results

Results from the study to evaluate effects of noise interference are presented in Fig. 5. Here we show performances of CIS and CA strategies in noise without any special provisions for noise reduction. A six-channel CIS implementation was used with the following parameters: 33 μs/phase pulses, 2525 pps rate of stimulation on each channel, staggered update order, 12th order bandpass filters, fullwave rectifiers, and 400 Hz lowpass filters (1st order).

Consonant identification first was measured under quiet conditions, and then progressively greater amounts of multitalker speech babble were added to the primary speech signal. 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.

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. This is especially encouraging inasmuch as the CA processor in the Ineraid device has been identified as the most resistant to the deleterious effects of noise among several tested implant systems [16], [17].

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 (e.g., [18], [19]). 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.

IV. EVALUATION OF EFFECTS OF CHANGES IN THE NUMBER OF CHANNELS FOR CIS PROCESSORS

A. Methods

This study involved implementation and subsequent evaluation of CIS processors using different numbers of channels. Each n-channel processor used the n apicalmost electrodes and filtered the same total frequency range into n bands of equal width on a logarithmic scale. 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 with fewer than six channels to maintain this constant rate). In addition, each processor used 6th order bandpass filters, fullwave rectifiers, and 400 Hz lowpass filters (1st 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 other six-channel processors using a "staggered" order of channel updates (6-3-5-2-4-1; see [9]). Also, processors using fewer 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 this particular study was to evaluate effects of changes in the number of channels, while maintaining a consistency in all other CIS parameters.

B. Results

Results from the channel-number study are presented in Fig. 6. Note that the scores decline monotonically, for both male and female speakers, with reductions in the number of CIS channels. Also note that ceiling effects may distort

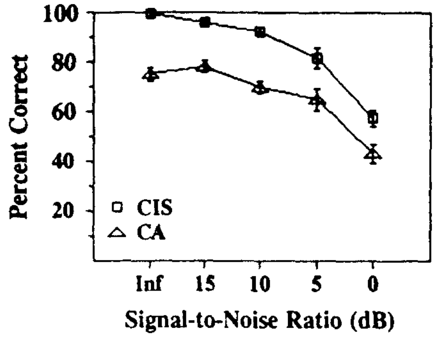

Figure 5. Percent correct medial consonant identification scores for CIS and CA processors as a function of signal-to-noise ratio (SNR). The SNR of "Inf" refers to presentation of the signal without any accompanying noise. Five presentations of each of 24 consonants by the male speaker were used in the consonant identification tests for each processor at each SNR. 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 averages of these scores (from five randomized sets) for the CIS processor, and the triangles show the averages for the CA processor. Standard errors of the mean are indicated by the vertical bars.

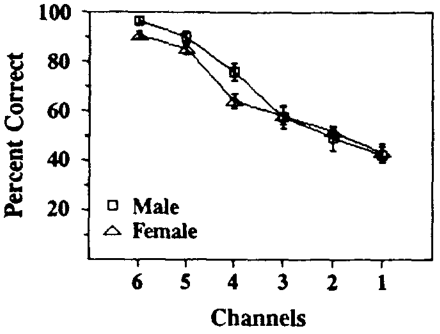

Figure 6. Percent correct medial consonant identification scores for CIS processors using different numbers of channels. Five presentations of each of 24 consonants by the male speaker, and five presentations of each consonant by the female speaker, were used in the tests with each processor. The square symbols show average percent correct scores for the male speaker, and the triangles show the averages for the female speaker. Standard errors of the mean are indicated by the vertical bars.

the slope and extent of increases from 5 channels to 6. Nevertheless there is a clear pattern of increases in consonant identification with increases in the number of channels. Obviously, further improvements in performance might be obtained by increasing the number of channels, either by adding electrodes or through some other means, such as the method described in the following section.

V. DESIGN AND PRELIMINARY EVALUATION OF VCIS PROCESSORS

A possible refinement and extension of the CIS approach is illustrated in Fig. 7. Here adjacent electrodes may be stimulated simultaneously to shift the perceived pitch in any direction with respect to the corresponding single-electrode percepts. Studies with implant subject SR2 indicate that perceived pitch can be manipulated through various choices of simultaneous and single electrode conditions. If, for instance, the apicalmost electrode of his Ineraid electrode array (electrode 1) is stimulated alone, a low pitch is reported. If the next electrode in the array (electrode 2) is stimulated alone, a higher pitch is reported. An intermediate pitch can be produced by stimulating the two electrodes together with identical, in-phase pulses. Finally, by reversing the phase of one of the simultaneous pulses pitch percepts outside the interval represented by the single-electrode percepts can be obtained. The availability of pitches other than those elicited with stimulation of single electrodes may provide additional channels of stimulation. We call these additional channels "virtual channels," and processors that use them virtual channel interleaved sampling (VCIS) processors.

An important feature shared by CIS and VCIS processors is interleaving of stimulus pulses. Pulses, or simultaneous combinations of pulses, are presented for each channel in a nonoverlapping sequence, such as the one shown in Fig. 2*c* for a CIS processor. Without interleaving, electric fields from other electrodes would interact (i.e., sum) with the fields produced by the stimuli for any given channel, thereby reducing the independence among channels.

A. Subject and Tests

Subject SR2 has participated in an extensive series of studies to evaluate effects of parametric changes in CIS processors and more recently to evaluate several implementations of VCIS processors. Results from his first tests with CIS processors are presented in earlier reports from our group [1], [20] and are summarized here in Table 2 for reference. The tests included the spondee, CID, SPIN and NU-6 tests described above.

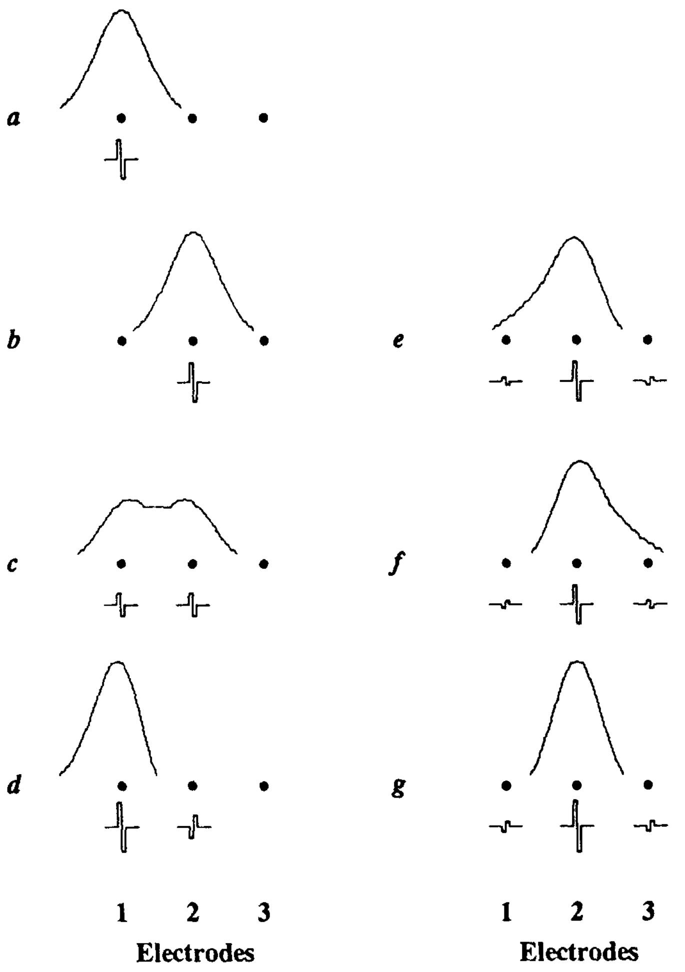

Figure 7. Schematic illustrations of neural responses for various conditions of stimulation. The top curve in each panel is a hypothetical sketch of the number of neural responses, as a function of position along the cochlea, for a given condition of stimulation. The condition of stimulation is indicated by the pulse waveform(s) below each dot, which represent the positions of three adjacent electrodes. Profiles of neural responses for stimulation of a single electrode are presented in *a* and *b*, and profiles for simultaneous stimulation of two electrodes in *c* and *d*. Implant subject SR2, listening to these different stimuli, balanced for loudness, can rank them according to their distinct pitches. Stimulation of apicalmost electrode 1 alone (*a*) produced a low pitch, while stimulation of electrode 2 alone (*b*) produced a higher pitch. Simultaneous stimulation of both electrodes, with identical pulses having approximately half the amplitude of the single-pulse conditions (*c*), produced an intermediate pitch. Pairing stimulation of electrode 1 with a reversed-polarity pulse on electrode 2 (*d*) produced the lowest pitch among the illustrated conditions. Various lower pitches could be produced by manipulating the ratio of the pulse 2 to pulse 1 amplitudes over the range of 0.2 to 0.8. A ratio of 1.0 produced a pitch higher than that elicited by stimulation of electrode 1 alone. Similarly, pitches higher than that elicited by stimulation of electrode 6 alone (the basalmost electrode) could be produced by presenting a reversed-phase pulse (of lower amplitude) on electrode 5. Additional discriminable pitches between electrodes could be produced by constructing triads of pulses, as illustrated in *e* and *f*. It may even be possible to reduce the width of an unshifted neural excitation field by supplying reversed-phase pulses on either side of a primary pulse, as illustrated in *g*. Subject SR2 reported that the pitch percept of case *g* was indistinguishable from that of case *b*. The pulse width used for all the listening tests was 33 μs/phase.

B. Results

These tests and others have been used to evaluate the subsequent implementations of CIS and VCIS processors. Results from a refined implementation of a CIS processor, using parameters somewhat different from those of the original implementation, are presented in the second column of numeric entries in Table 2. Results for a VCIS processor are presented in column 3. This processor used in-phase pulses to produce pitches between each pair of electrodes. Thus, six channels corresponded to conditions involving stimulation of one electrode alone, while five additional channels corresponded to conditions involving in-phase paired stimulation. As a precaution against possible learning or familiarization effects, different lists of words and sentences were used in each of the CID, SPIN and NU-6 tests for the different processors. Also, the NU-6 test was repeated for the "refined CIS" processor using another new list of words. The additional test listed in Table 2 was the 24 consonants test, with each of the 24 played in block-randomized order 10 times for a male speaker and 10 times for a female speaker. As with the other tests, the medial consonant tokens were presented in a sound-alone condition, with no feedback as to correct or incorrect responses.

Scores for all three processors are quite high. Indeed, most of the scores are at or near the upper scale limits for each of the tests. The only exception is the NU-6 test, for the two implementations of CIS processors. The NU-6 scores indicate an improvement in performance with the refined CIS processor over the original implementation. The refined processor used a somewhat higher rate of stimulation on each channel (2500 *vs* 1515 pps), shorter pulses (33 *vs* 55 μs/phase), a higher corner frequency for the input equalization filter (1200 *vs* 600 Hz), sharper bandpass filters (12th *vs* 6th order), and a lower cutoff frequency for the lowpass filters in the envelope detectors (400 *vs* 800 Hz). Also, the refined processor was evaluated in the 10th week of testing various CIS and other processors with this subject, spread over a three-year period. Learning effects associated with this additional experience also may have contributed to his improved scores [12]-[14], [21].

Table 2

PERCENT CORRECT SCORES FROM SPEECH TESTS WITH 6-ELECTRODE PROCESSORS, SUBJECT SR2

Test	CIS	Refined CIS	VCIS[a]
Spondees	96	100	100
CID	100	100	100
SPIN	96	100	100
NU-6	80	90,94[b]	98
Consonants	*	98.1 ±0.7[c]	97.1 ±0.8

[a] VCIS and refined CIS processors both used 12th-order bandpass filters, fullwave rectifiers, 400 Hz lowpass filters (1st order), and 33 μs/phase pulses. The rate of stimulation for each channel was 1365 pps for the VCIS processor and 2500 pps for the refined CIS processor.

[b] Scores from two separate administrations of the NU-6 test; total phoneme score was 287/300.

[c] SEM of block percent-correct scores, combined male and female speaker data.

*The 24 consonant test was not conducted during this initial fitting and evaluation of a CIS processor.

With the 11-channel VCIS processor SR2 achieved a score of 98% correct on the NU-6 test, making only one phoneme error (149/150 phonemes). He obtained scores of 100% correct on all remaining open-set tests, and a score of 97% correct on the consonant test. The few errors on the consonant test included five /v/-/ð/ confusions; two /f/-/θ/ confusions; and single confusions between /n/-/m/, /f/-/s/, /s/-/θ/, /ð/-/z/, /ð/-/b/, /b/-/d/, and /g/-/d/.

C. Discussion

These results indicate quite high levels of speech recognition. The NU-6 score of 80% correct falls in the middle of the range of scores obtained by listeners with mild-to-moderate hearing losses [10], [11]; the scores of 90 and 94% correct fall in the range of scores obtained by listeners with mild hearing losses [22]; and it is not unusual for a subject with normal hearing to score only 98% correct (e.g., [23], [24]).

The attainment of such scores with a cochlear implant is noteworthy. The sensitivities of standard audiometric tests of open-set speech recognition are inadequate for this subject with the best of these processors. Obviously, tests of even greater difficulty (e.g., tests involving presentation of low context sentences in noise) will be required in the future to discriminate differences in speech reception under such circumstances.

While it seems unlikely that many patients will be able to enjoy the high levels of performance obtained by SR2 (e.g., see [25]), the present results demonstrate what is possible with electrical stimulation of the auditory system, using only six monopolar electrodes. In fact, the scores reported here are highly consistent with scores reported in the early literature on analysis/synthesis systems (sometimes called "vocoder" systems) for the reduced bandwidth transmission of speech. In those experiments speech test scores for listeners with normal hearing began to asymptote at high levels when 6 to 10 bandpass channels were used in acoustic representations of envelope signals [26]. The present results suggest that most or all of the information contained in such envelope signals can be transmitted across the electrode-nerve interface of a cochlear implant, and that quite high levels of speech reception can be supported with 6 or 11 channels.

Although the indications from preliminary implementations of VCIS processors are encouraging, much work remains to evaluate the potential of the VCIS approach. First, the scores for this particular subject all approach or hit the ceiling for all tests with both the VCIS and refined CIS processors. Use of more difficult tests, and/or studies of subjects with lower scores, will be required to measure with confidence any difference between the processors. Second, only one combination of virtual channels was involved in the VCIS processor described here. As suggested in Fig. 7, this combination, using identical in-phase pulses on adjacent electrodes, might produce substantial overlaps in neural excitation fields across single-electrode and two-electrode channels (compare the hypothetical fields sketched for conditions *a*, *b* and *c* in Fig. 7). Reversed-phase pulse combinations (condition *d* in Fig. 7) may reduce such overlaps, as may simultaneous pulsatile stimulation of more than two electrodes.

VI. ACKNOWLEDGEMENTS

We thank the subjects of the described studies for their enthusiastic participation. We also are pleased to acknowledge the important scientific contributions of many colleagues, including, for the studies reviewed here, Michael F. Dorman, Donald K. Eddington, William M. Rabinowitz, Robert V. Shannon and Robert D. Wolford. This work was supported by NIH projects N01-DC-9-2401, N01-DC-2-2401 and P01-DC00036.

VII. REFERENCES

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

[2] R. V. Shannon: Psychophysics. In *Cochlear Implants: Audiological Foundations*, R. S. Tyler (Ed.). San Diego, CA: Singular Publishing Group, 1993, pp. 357-388.

[3] M. W. White, M. M. Merzenich and J. N. Gardi: Multichannel cochlear implants: Channel interactions and processor design. *Arch Otolaryngol* 110: 493-501, 1984.

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

[5] G. M. Clark: The University of Melbourne-Nucleus multi-electrode cochlear implant. *Adv Oto-Rhino-Laryngol* 38: 1-189, 1987.

[6] B. S. Wilson, D. T. Lawson, C. C. Finley and R. D. Wolford: Coding strategies for multichannel cochlear prostheses. *Am J Otol* 12, Suppl. 1: 56-61, 1991.

[7] B. S. Wilson: Signal processing. In *Cochlear Implants: Audiological Foundations*, R. S. Tyler (Ed.). San Diego, CA: Singular Publishing Group, 1993, pp. 35-85.

[8] D. K. Eddington: Speech discrimination in deaf subjects with cochlear implants. *J Acoust Soc Am* 68: 885-891, 1980.

[9] D. T. Lawson, B. S. Wilson and C. C. Finley: New processing strategies for multichannel cochlear prostheses. *Prog Brain Res*, in press.

[10] F. H. Bess and T. H. Townsend: Word discrimination for listeners with flat sensorineural hearing losses. *J Speech Hear Disorders* 42: 232-237, 1977.

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

[12] R. C. Dowell, P. M. Seligman, P. J. Blamey and G. M. Clark: Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis. *Ann Otol Rhinol Laryngol* 96, Suppl. 128: 132-134, 1987.

[13] R. C. Dowell, A. M. Brown and D. J. Mecklenburg: Clinical assessment of implanted deaf adults. In *Cochlear Prostheses*, G. M. Clark, Y. C. Tong and J. F. Patrick (Eds.). Edinburgh: Churchill Livingstone, 1990, pp. 193-205.

[14] R. S. Tyler, J. P. Preece, C. R. Lansing, S. R. Otto and B. J. Gantz: Previous experience as a confounding factor in comparing cochlear-implant processing schemes. *J Speech Hear Res* 29: 282-287, 1986.

[15] R. S. Tyler, J. P. Preece and M. W. Lowder: *The Iowa audiovisual speech perception laser videodisc*. Laser Videodisc and Laboratory Report, Department of Otolaryngology -- Head and Neck Surgery, University of Iowa Hospitals and Clinics, Iowa City, Iowa, 1987.

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

[17] R. S. Tyler and N. Tye-Murray: Cochlear implant signal-processing strategies and patient perception of speech and environmental sounds. In *Cochlear Implants: A Practical Guide*, H. Cooper (Ed.). San Diego, CA: Singular Publishing Group, 1991, pp. 58-83.

[18] J. W. Hall, M. P. Haggard and M. A. Fernandes: Detection in noise by spectro-temporal pattern analysis. *J Acoust Soc Am* 76: 50-56, 1984.

[19] B. C. J. Moore: Comodulation masking release and across-channel masking. In *Audition, Speech and Language*, M. E. H. Schouten (Ed.). Berlin: Mouton de Gruyter, in press.

[20] B. S. Wilson, D. T. Lawson and C. C. Finley: *Speech processors for auditory prostheses*. Fourth Quarterly Progress Report, NIH project N01-DC-9-2401, 1990.

[21] M. F. Dorman, K. Dankowski and G. McCandless: Longitudinal changes in word recognition by patients who use the Ineraid cochlear implant. *Ear Hear* 11: 455-459, 1990.

[22] J. R. Dubno, D. D. Dirks and D. E. Morgan: Effects of age and mild hearing loss on speech recognition in noise. *J Acoust Soc Am* 76: 87-96, 1984.

[23] H. Davis and S. R. Silverman: *Hearing and Deafness*. New York: Holt, Rinehart and Winston, 1978.

[24] T. Frank and C. H. Craig: Comparison of the Auditec and Rintelmann recordings of the NU-6. *J Speech Hear Disorders* 49: 267-271, 1984.

[25] B. S. Wilson, D. T. Lawson, C. C. Finley and R. D. Wolford: Importance of patient and processor variables in determining outcomes with cochlear implants. *J Speech Hear Res*, spring issue, 1993.

[26] J. L. Flanagan: *Speech Analysis, Synthesis and Perception* (2nd Edition). Berlin: Springer-Verlag, 1972.