Final Report

August 1, 1992 through July 31, 1995

NIH Contract N01-DC-2-2401

Speech Processors for Auditory Prostheses

Prepared by

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

Center for Auditory Prosthesis Research
Research Triangle Institute
Research Triangle Park, NC 27709

CONTENTS

I. Introduction

The purpose of this project was to design and evaluate speech processors for implantable auditory prostheses. Ideally, such processors 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.

In the project just preceding the current project (NIH N01-DC-9-2401), we developed the *continuous interleaved sampling* (CIS) strategy for representing speech information with implants (e.g., Wilson et al., 1991). This strategy, or a close variation of it, now is used in each of three commercially available implant systems and soon will be included in a fourth system as well. High levels of speech recognition have been obtained with the CIS strategy in a number of research studies and with the three clinical systems just mentioned.

In more recent studies, in the current project (NIH N01-DC-2-2401), we have continued development and evaluation of CIS processors. The studies have included evaluation of effects of parameter changes, such as changes in pulse duration, pulse rate, and channel update order. The studies also have included evaluation of CIS performance for patients with low levels of speech recognition with their clinical devices. In broad terms, results from the parametric studies show that large gains in performance can be made through informed choices of parameter values and further that the best choices for some parameters are not the same for all patients. Substitution of a CIS processor for a clinical *compressed analog* (CA) processor produces large gains in speech recognition for most of the patients studied to date, including patients at the low end of the clinical performance scale. In fact, some patients who started at or near zero levels of open set speech recognition with their clinical devices demonstrated substantial open set recognition with CIS. The improvements have been immediate, i.e., measures of speech reception have been made after only a few hours of experience in the laboratory with various CIS processors. In contrast, each subject had had at least one year of daily experience with his or her CA processor at the time of our tests.

Additional activities in the current project have included:

- Design and evaluation of "virtual channel" and "sharpened field" CIS processors
- Recording of intracochlear evoked potentials in patients implanted with the Ineraid device or a percutaneous connector version of the Nucleus device, to assess temporal patterns of neural response for a wide range of stimuli
- Psychophysical measures of stimulus scaling and identification, for pulse trains with different pulse rates and for sinusoidally amplitude modulated (SAM) pulse trains with a variety of carrier rates and modulation frequencies
- Psychophysical measures of stimulus scaling and identification for stimuli delivered to different single electrodes in the Ineraid implant and for stimuli delivered simultaneously to various combinations of electrodes, as in "virtual channel" and "sharpened field" CIS processors
- Psychophysical measures of forward masking to assess the spatial patterns of neural excitation produced with single electrode and multiple electrode maskers in Ineraid patients, using the procedure of Lim et al. (1989)

- Electrophysiological and psychophysical measures of forward masking with the masker and the probe presented on the same electrode, to asses the time course of recovery of neural responsiveness following a masker and to evaluate possible differences in electrophysiological and psychophysical measures (electrophysiological measures were made with recordings of intracochlear evoked potentials)
- Parametric and control studies with CIS processors, to (a) evaluate effects of changes in number of channels, rate of stimulation, pulse duration, and channel update order, and (b) compare simultaneous *versus* nonsimultaneous stimulation
- Measures of complex tone perception with CIS and "virtual channel" CIS processors
- Initial studies with 4 of a series of seven patients implanted with an experimental version of the Nucleus device, one providing a percutaneous connector for direct electrical access to the 22 implanted electrodes (this study is made possible by a collaborative agreement with Cochlear Corporation and Duke University Medical Center)
- Preliminary evaluation of CIS processors with an adjustable gain at the output of each envelope detector to normalize speech intensities across channels
- Preliminary evaluation of CIS processors with a noninstantaneous compressor at the output of each envelope detector to mimic crudely the noninstantaneous compression that occurs in normal hearing at the synaptic junctions between inner hair cells (IHCs) and single fibers of the auditory nerve
- Development of strategies to improve the representation of IHC/synaptic function for inclusion in new speech processor designs
- Evaluation of upward and downward extensions of the frequency range spanned by CIS channels, in studies with two Nucleus percutaneous subjects and two Ineraid subjects
- Development of hardware and software for support of patient studies, such as a laboratory interface for control of the implanted receiver in the Clarion (Advanced Bionics) device, refinement of an existing interface for control of the implanted receiver in the Nucleus device and a new version of the Auditory Brainstem Implant (ABI) device, software for implementing the algorithms of the psychophysical tests listed above, software for display and analysis of results from the psychophysical tests, software and hardware for implementing new speech processing algorithms, software for new or refined tests of speech reception, hardware for new current sources capable of presenting short-duration (e.g., 10 μs/phase) pulses with high fidelity and minimal crosstalk among separate channels, and incorporation of a high-resolution A/D converter for the laboratory speech processor system
- Collaborative studies with Michael Dorman, to evaluate representations of frequency cues through CA, CIS and "virtual channel" CIS processors
- Collaborative studies with Bryan Pfingst, to investigate mechanisms of pitch perception with cochlear implants
- Formation of a partnership with the University of Innsbruck to develop a portable processor capable of implementing CIS strategies, for use in research studies by our group and others in the United States
- Initial evaluation of this portable processor in our laboratory, in studies with two Ineraid patients (the outputs of the processor were routed through the percutaneous connector of the Ineraid implant for stimulation)

- Formation of a partnership with the Massachusetts Eye & Ear Infirmary and Hôpital Cantonal Universitaire in Geneva, Switzerland, to develop an especially powerful and flexible portable processor for upcoming studies in the succeeding project, at little cost to that project or the current one (costs and tasks are distributed among the three institutions)
- Transfer of CIS processor technology to Med El GmbH in Innsbruck, Austria, Advanced Bionics in Sylmar, CA, Smith & Nephew Richards in Memphis, Tennessee and Cochlear Pty. Ltd. in Sydney, Australia, for use in next generation cochlear prostheses
- Development and presentation of recommendations for a low cost but effective cochlear prosthesis for widespread use in China, at the invitation of the Chinese Government
- Preliminary evaluation of a prototype of the system recommended for China in speech reception tests with an Ineraid subject
- Initiation of studies to evaluate possible learning effects with extended daily use of a portable CIS processor
- Measures of speech reception with a single channel variation of CIS processors, for direct comparison with electrophysiological and psychophysical measures made with the same patients and electrodes by Paul Abbas and Carolyn Brown
- Publication of five papers and two book chapters
- Presentation of 21 invited lectures at national and international conferences

In this report we present results from several of the studies and activities outlined above. We also list publications and presentations for the project, and briefly indicate directions for future research. Additional information may be found in the Quarterly Progress Reports for the project, as indicated in Table 1. Reprints of the publications are included as an Appendix to this report.

Table 1. Principal topics of the Quarterly Progress Reports for NIH Project N01-DC-2-2401.

Report	Topic(s)
1	Virtual channel interleaved sampling (VCIS) processors, initial studies
2	Single parameter variation studies for CIS processors
3	Identification of virtual channel conditions on the basis of pitch
4	Representation of complex tones by sound processors for implanted auditory prostheses
5	Transfer and dissemination of CIS processor technology; parametric and control studies with CIS processors
6	Further evaluation of VCIS processors with reduced-electrodes conditions and with additional subjects
7	Initial report on temporal representations with cochlear implants, including modeling studies, psychophysical and electrophysiological studies, and comparisons of model predictions and data
8	Further studies of complex tone perception by implant patients
9	Strategies for the repair of deficits in temporal representations with implants
10	A channel-specific tool for analysis of consonant confusion matrices
11	Recordings of intracochlear evoked potentials for sustained electrical stimuli

II. Importance of the Patient Variable in Determining Outcomes with Cochlear Implants

An interesting aspect of our prior comparisons of *compressed analog* (CA) and CIS processors in tests with 11 subjects was that the amount of improvement with CIS was approximately uniform across subjects (Wilson et al., 1993). Pearson correlations of the CA and CIS scores ranged from 0.87 to 0.92 for four tests of open-set speech recognition, with the highest correlation found for the one test (the NU-6 test) that did not exhibit obvious ceiling effects. Such high correlations indicate that a substantial portion of the variance in the results can be attributed to effects of the patient variable, i.e., a patient's score with one processor is strongly related to his or her score with the alternative processor. For the NU-6 test, for example, 85 percent of the variance in the results is explained by the patient variable. Thus, although we can improve outcomes with a change in the processing strategy, the patient variable has an even greater effect, at least for fittings limited to a few days. A patient who enjoys a high ranking with one processor will retain that ranking with another processor, whereas a patient who has a low ranking with one processor will also retain that ranking with another. Identification of the factor or factors that allow some patients to perform at high levels with a variety of processing strategies (and presumably with a variety of prosthesis systems) might lead to the development of improved prognostic tests for prospective implant patients and to a better understanding of the basic mechanisms underlying implant function.

III. Parametric Studies with CIS Processors

In recent studies we have evaluated effects produced by changes in parameters for CIS processors (e.g., see QPR 5; Lawson et al., 1993; Wilson et al., 1994). We have compared performances with processors using (a) different numbers of channels, (b) nonsimultaneous *versus* simultaneous stimulation across channels, (c) different pulse durations and rates, and (d) different update orders for the channels within each cycle of nonsimultaneous stimulation. Findings to date indicate that each of these manipulations can affect performance and that the best choices for some parameters vary from patient to patient. The findings may be summarized as follows:

- In studies with one subject, decreases in the number of CIS channels in steps of 1, from 6 channels to 1, produced significant decreases in percent correct scores on the 24 consonant test at each step, for both male and female speakers. Studies with other subjects also have demonstrated improvements in performance with the addition of CIS channels.
- Use of simultaneous stimulation produced large decrements in consonant identification for three tested subjects. This result was consistent with the idea that simultaneous stimulation might produce strong interactions among channels and thereby degrade the representation of channel-related cues.
- Effects of changes in pulse durations and rates have been studied with several subjects. Results for a subject in the "low CA" group (subjects with relatively poor speech reception performance with the clinical CA processor) indicated a maximum of performance on the 16 consonant test with a pulse rate of 833/s and a pulse duration of 33 μs/phase. Use of a higher (2525/s) or lower (500 or 417/s) rate produced a decrement in performance. Also, use of pulses with greater durations (at rates below 2525/s, where such pulses can be used in a 6-channel processor without temporal overlaps) produced scores about the same as or lower (sometimes substantially lower) than those obtained with 33 μs/phase pulses. Results for other subjects, in the "high CA" group (subjects with high levels of speech reception performance with the clinical CA processor), indicated small but significant increases in performance on either the 16 or 24 consonant test as the pulse rate was increased from 833 to 1365/s, and from 1365 to 2525/s, using 33 μs/phase pulses.
- In standard implementations of CIS processors a "staggered" order of channel updates has been used. This order imposed the maximum possible spatial separation between sequentially stimulated channels for each stimulus cycle (e.g., for a 6-channel processor an order of 6-3-5-2-4-1 has been used, where electrode 1 is the apicalmost electrode). In tests conducted near the beginning of our studies with CIS processors we compared the staggered order with a base-to-apex order. A base-to-apex order mimics the direction of the traveling wave of mechanical displacements along the basilar membrane found in normal hearing. We therefore anticipated a possible advantage to the base-to-apex order. The results, however, showed that the staggered order was as good as, or substantially better than, the base-to-apex order for the studied subjects. As a control, we decided in recent studies with three subjects to compare the staggered order with an apex-to-base order. We expected that the apex-to-base order might produce a decrement in performance, inasmuch as it (a) placed sequentially stimulated channels adjacent to each other, as in the base-to-apex order, and (b) was opposite to the direction of the traveling wave found in normal hearing. To our surprise, the apex-to-base order was clearly superior for two of the three subjects. Large increases in information transfer for the place of articulation feature were seen for both subjects. In addition, the apex-to-base order produced large gains in the transmission of voicing, duration, and envelope information

for one subject. Both subjects performing better with it expressed a strong preference for the apex-to-base order, and said that processors using that order sounded more natural, more intelligible, and lower in overall pitch than otherwise identical processors using the staggered update order. In contrast, results for the third subject did not demonstrate a clear difference in performance with the two orders.

IV. Importance of Processor Fitting

Findings presented in the prior section indicate that choices of pulse rate, pulse duration, and channel update order can have large effects on the performance of CIS processors. Such effects are illustrated further in Figure 1, which shows a history of improvements for one subject over the course of three successive visits to our laboratory. The subject used his Ineraid CA clinical device daily before and between visits to our laboratory.

In the first visit, this subject was fitted with a CIS processor using relatively long pulses (167 μs/phase), a relatively low pulse rate (500/s), and a staggered order of channel updates. As shown in the figure, application of this processor produced a quite large improvement over the clinical CA processor used in his daily life. The score for consonant identification improved from 25 to 56 percent correct. The scores for the open set tests also improved, as indicated here for the CID sentence and NU-6 word tests. These increases in performance were obtained after no more than several hours of experience with CIS processors, compared with more than a year of daily experience with the CA processor.

Seventeen months later the subject returned to the laboratory for a second visit, this time to evaluate effects of manipulations of pulse duration and pulse rate. The best combination of tested durations and rates for this subject was 33 μs/phase pulses presented at 833/s. We decided at the end of the visit to evaluate fully a CIS processor using these parameters. We also repeated the consonant test for the CA processor, to evaluate the possibilities that his performance with that processor had improved through additional daily experience, or that his performance on the consonant test had improved through additional practice and familiarity with the test.

The subject again obtained a score of 25 percent correct for the consonant test (combined male and female speakers) with the CA processor. Scores for the CIS processor were generally better than those obtained during the first visit, with a CIS processor using a much longer pulse duration and a lower pulse rate. The score for the consonant test improved from 56 to 79 percent correct; the score for the CID test stayed about the same, from 55 to 53 percent correct; and the score for the NU-6 test almost doubled, from 14 to 26 percent correct.

Six months later, during a third visit to our laboratory by this same subject, we compared CIS processors with different update orders, and found that an apex-to-base order produced an improvement over the staggered order. Based on this finding, we again decided at the end of the visit to evaluate a CIS processor with an apex-to-base update order in greater detail, and to repeat the consonant test for the CA processor.

The subject's performance on the consonant test with the CA processor improved, from 25 to 37 percent correct. Most of the improvement was for the female speaker. The improvement may have reflected better use of cues provided by the CA processor through additional experience, or increased familiarity with the female version of the consonant test (which during this third visit had been used more intensively than before).

Performance with the apex-to-base CIS processor was substantially better than performance with an

	Proc	Ch	Dur	Rate	Order	Date
░	CA	4				11/91
■	CIS	6	167	500	stag	11/91
▒	CIS	6	33	833	stag	4/93
▓	CIS	6	33	833	a-b	10/93

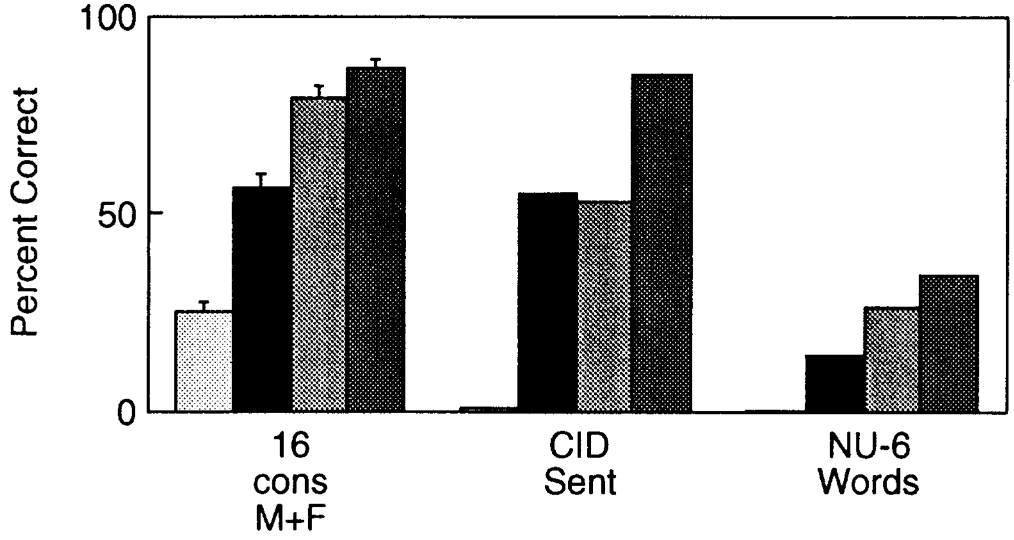


Figure 1. Improvements in speech processor performance for Ineraid subject SR10. The table above the bar chart specifies different processors. The leftmost column indicates the general strategy, CA or CIS. The remaining columns indicate, from left to right, the number of channels (Ch) in CA or CIS processors, the duration in μs per phase (Dur) of pulses used in CIS processors, the stimulus rate in pulses per second for channel in CIS processors (Rate), the order of channel updates used in CIS processors (either staggered [stag] or apex-to-base [a-b]), and the date of testing. The error bars show standard errors of the mean for the consonant test.

otherwise identical processor using the staggered update order. The consonant test score improved from 79 to 87 percent correct; the score for the CIS test from 53 to 85 percent correct; and the score for the NU-6 test from 26 to 34 percent correct.

Note that the improvements across three visits took this subject from zero or near zero levels of open set recognition with the CA processor to quite high levels with his final CIS processor. This example shows what is possible for at least one patient at the low end of the clinical performance spectrum, and indicates the importance of parameter choices for CIS processors.

V. "Virtual Channel" and "Sharpened Field" CIS Processors

One result from parametric studies with CIS processors is that speech reception performance generally increases with the number of distinct channels used, at least through 6 channels. An example of one such finding is presented in Figure 2. The study in this case involved tests of consonant identification with Ineraid subject SR2, using CIS processors with 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 of the Ineraid implant. 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 envelope filters (1st order). For consistency across conditions, 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.

The results show that scores on tests of consonant identification decline monotonically, for both male and female speakers, with reductions in the number of CIS channels. Also note that ceiling effects may distort 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.

Design of VCIS Processors

Findings such as the one just presented motivated us to explore ways of increasing the effective number of channels for patients using the Ineraid implant. This led to the development of "virtual channel" CIS (VCIS) processors, which use simultaneous stimulation of two or more electrodes to produce 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 Figure 3. 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, with the dots representing 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*.

In psychophysical tests (see QPR 3) Ineraid subject SR2 can rank conditions *a* through *f* according to their distinct pitches. Stimulation of apicalmost electrode 1 alone (condition *a*) produces a low pitch, whereas stimulation of electrode 2 alone (condition *b*) produces a higher pitch. Simultaneous stimulation of both electrodes, with identical pulses having approximately half the amplitude of the single-pulse conditions (condition *c*), produces an intermediate pitch. Pairing stimulation of electrode 1 with a reversed-polarity pulse on electrode 2 (condition *d*) produces the lowest pitch among the illustrated conditions. Similarly, pitches higher than that elicited with stimulation of electrode 6 alone (the basalmost electrode in the Ineraid array) can be produced by presenting a reversed-phase pulse (of

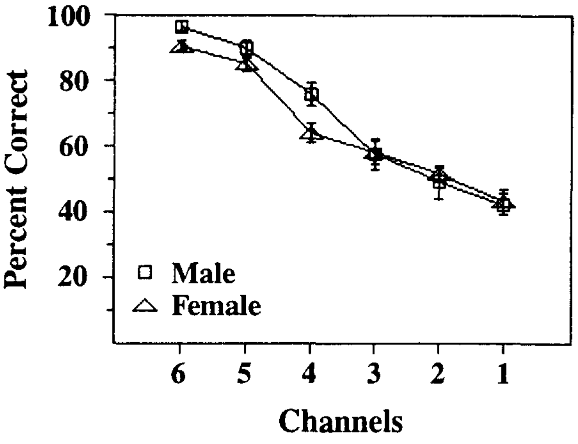


Figure 2. Performance of CIS processors as a function of the number of channels used. Percent correct scores from tests of consonant identification with Ineraid subject SR2 are shown. Five presentations of each of 24 consonants by a recorded male speaker, and five presentations of each consonant by a recorded 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 averages of these scores (from five randomized sets) 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.

lower amplitude) on electrode 5. Additional pitches between electrodes can be produced by constructing triads of pulses, as illustrated in panels *e* and *f*. The pitch produced with the stimulus of condition *e* is lower than that elicited with stimulation of electrode 2 alone (condition *b*), whereas the pitch produced with the stimulus of condition *f* is higher than that elicited with stimulation of electrode 2 alone.

Condition *g* in Figure 3 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-polarity pulses on either side of a principal pulse. Subject SR2 reports that the pitch percept of this condition is indistinguishable from that of condition *b*. We note that this general type of "sharpened field" stimulation also has been described by Townshend and coworkers (1987) and by Jolly, Spelman and Pfingst (1994).

Ineraid subjects SR10 and SR13 also have participated in psychophysical studies to evaluate perceptual differences among virtual channel and single-electrode stimuli. Subject SR10 was tested with conditions *a* through *c* and subject SR13 with conditions *a* through *d*. The results were the same as those indicated above for SR2.

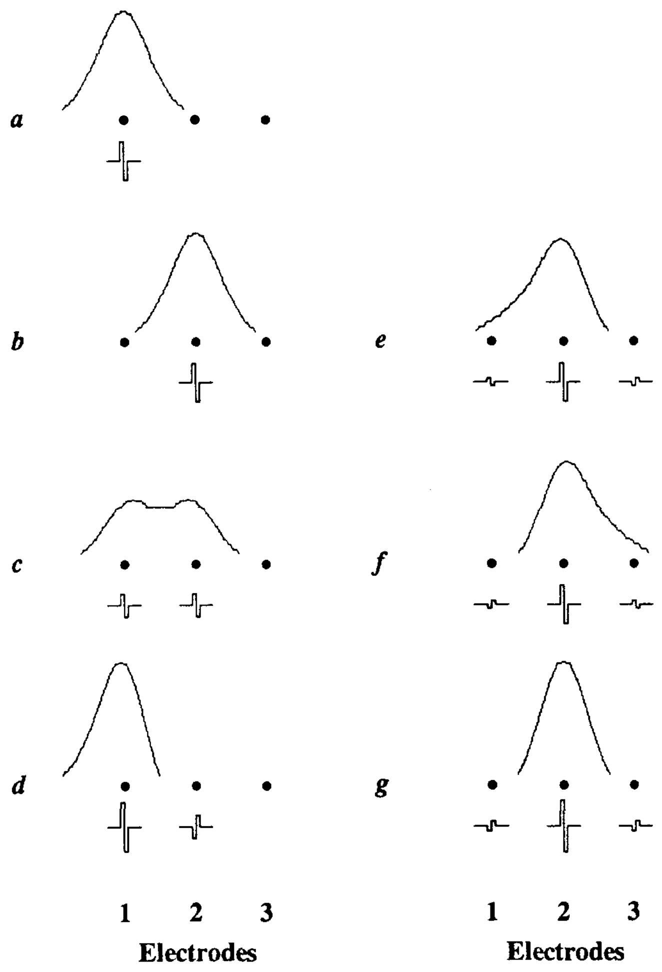


Figure 3. Conditions of single-electrode and multiple-electrode (virtual channel) stimulation. See text for a description of the different types of information presented in each panel.

Initial CIS/VCIS Comparison

One implementation of a VCIS processor is illustrated in Figure 4. The virtual channels use identical in-phase pulses presented simultaneously on adjacent electrodes. These channels are combined with six single-electrode channels to form an 11 channel processor. As in standard CIS processors, the stimulus for each of the channels is presented in a nonoverlapping sequence. Without such interleaving of stimuli, electric fields from other electrodes would interact (i.e., sum) with the fields produced by the stimulus for any given channel, thereby reducing the independence among channels.

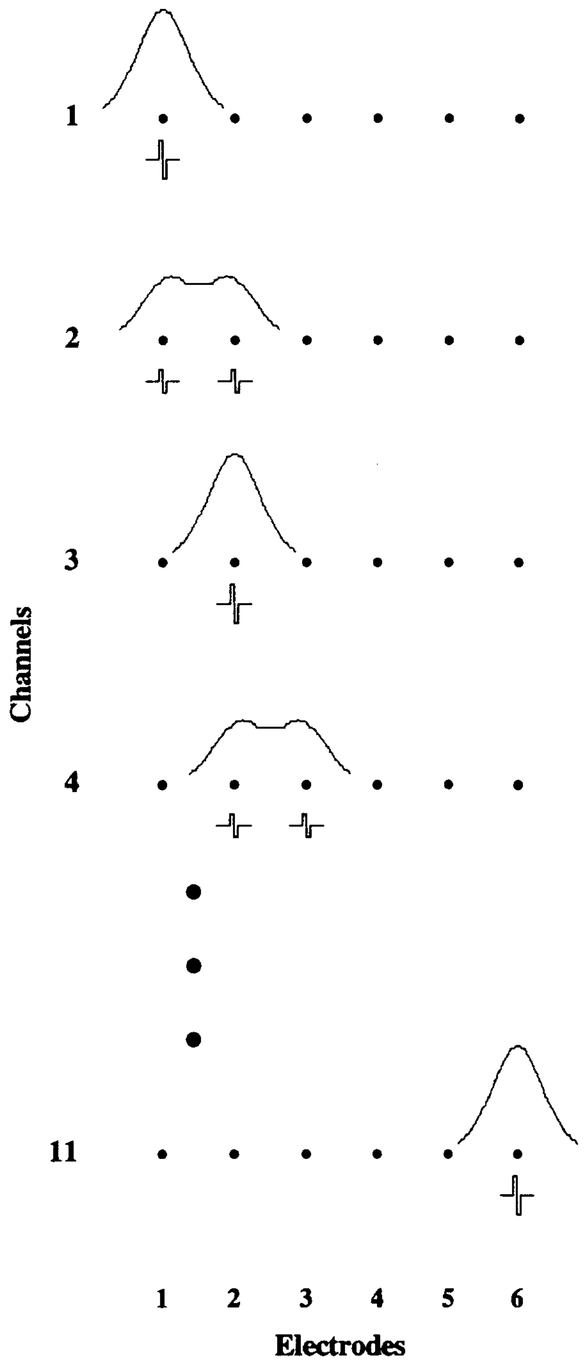


Figure 4. Construction of an 11 channel VCIS processor.

Table 2. Percent correct scores from speech tests with 6-electrode processors, subject SR2. The processors included a 6-channel CIS processor, a refined 6-channel CIS processor, and an 11-channel VCIS processor.

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.

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

An 11 channel VCIS processor of the type illustrated in Figure 4 has been compared with 6 channel CIS processors in initial tests with subject SR2 (Wilson et al., 1994 and QPR 1). This subject has participated in an extensive series of studies to evaluate effects of parametric changes in CIS processors and more recently to evaluate implementations of VCIS processors. Results from his first tests with CIS processors (e.g., Wilson et al., 1991) are summarized in Table 2 for reference. The tests included open-set recognition of 25 two-syllable words (spondees), 100 key words in the Central Institute for the Deaf (CID) sentences of everyday speech, the final word in each of 50 "high predictability" sentences in the Speech Perception in Noise (SPIN) test (presented without noise in our studies), and 50 one-syllable words from Northwestern University Auditory Test 6 (NU-6). All tests were conducted with hearing alone and the test items were presented from standard recordings without feedback or repetition.

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 the 11 channel VCIS processor are presented in column 3. 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 involved identification of 24 consonants in an /a/-consonant-/a/ context. Each of the 24 was presented 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.

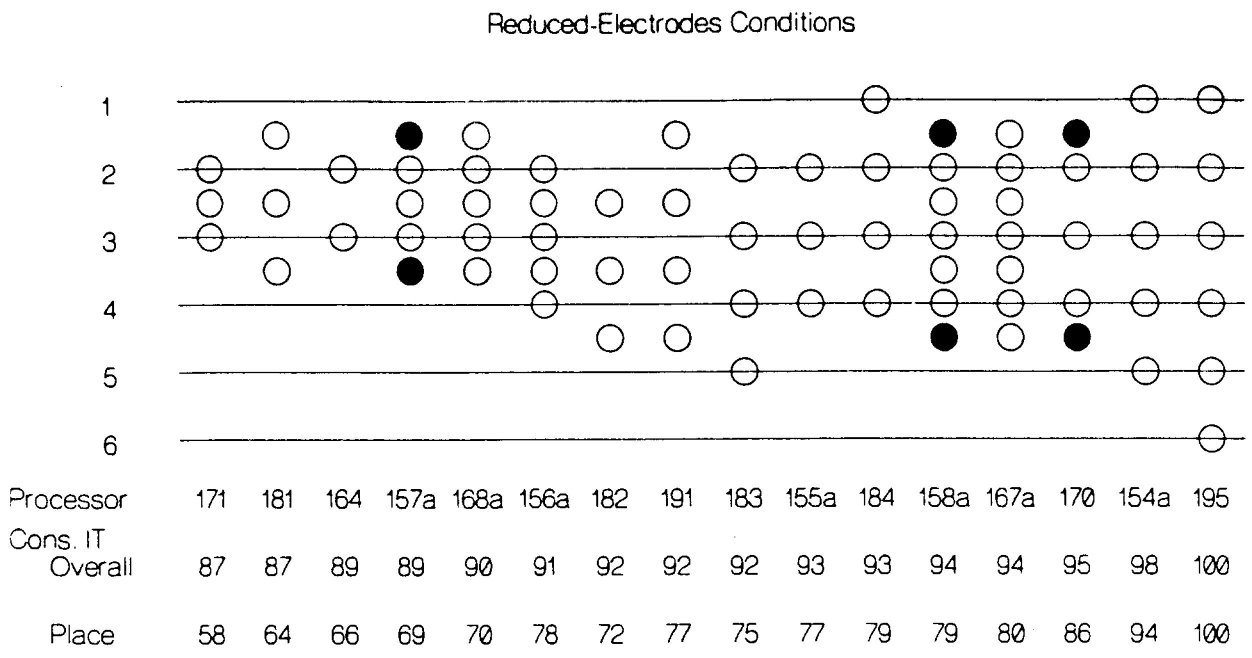
Scores for all three processors are quite high. Indeed, most of the scores are at or near the upper scale limits for each of the tests. The only exception is the NU-6 test, for the two implementations of CIS processors. The NU-6 scores indicate an improvement in performance with the refined CIS processor over the original implementation. The refined processor used a somewhat higher rate of stimulation on each channel (2500 *versus* 1515 pps), shorter pulses (33 *versus* 55 μs/phase), a higher corner frequency for the input equalization filter (1200 *versus* 600 Hz), sharper bandpass filters (12th *versus* 6th order), and a lower cutoff frequency for the lowpass filters in the envelope detectors (400 *versus* 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 or practice effects associated with this additional experience also may have contributed to his improved scores (Dorman et al., 1990; Dowell et al., 1987; Tyler et al., 1986).

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.

Evaluation of VCIS Processors with Reduced Numbers of Electrodes

Following the initial comparison of CIS and VCIS processors, we decided to evaluate a variety of CIS and VCIS processors with reduced numbers of electrodes (see QPR 6). The principal motivation for these additional studies was to reduce test scores to a range in which they would be more sensitive to differences among processors. In addition, we were interested in evaluating the potential benefit of virtual channels for patients with a limited number of usable electrodes.

The conditions and results of the additional studies, again with subject SR2, are presented in Figure 5. The horizontal lines indicate the positions of six physical electrodes. An open circle on one of the lines indicates a channel of stimulation with a single electrode. An open circle between lines indicates a virtual channel formed by stimulation of adjacent electrodes with identical in-phase pulses (corresponding to condition *c* in Figure 3). A closed circle indicates a virtual channel formed by presentation of a principal pulse at one electrode paired with simultaneous presentation of a reversed-polarity half-amplitude pulse on an adjacent electrode (corresponding to condition *d* in Figure 3). As an example, the leftmost condition in Figure 5 is that of a three-channel processor using two single-electrode channels and one virtual channel. Electrodes 2 and 3 are used for the single-electrode channels, and electrodes 2 and 3 are stimulated together with identical in-phase pulses for the virtual channel. The next condition in Figure 5 also has three channels, but in this case each of the channels is a virtual channel formed with identical in-phase pulses. The fourth condition uses the three channels of the first condition along with two additional virtual channels formed with reversed-polarity pulses. The apicalmost virtual channel is produced by simultaneous stimulation of electrode 2 with a principal pulse and electrode 3 with a reversed-polarity pulse at half the amplitude of the principal pulse. Similarly, the basalmost virtual channel is produced with a principal pulse on electrode 3 and a reversed-polarity pulse on electrode 2. Note that this condition is designed to convey five channels of information with



Processor	171	181	164	157a	168a	156a	182	191	183	155a	184	158a	167a	170	154a	195
Cons. IT Overall	87	87	89	89	90	91	92	92	92	93	93	94	94	95	98	100
Place	58	64	66	69	70	78	72	77	75	77	79	79	80	86	94	100

Figure 5. Conditions and results of reduced-electrodes study. See text for explanation of symbols denoting conditions at the top of the figure. Results from tests of consonant identification are presented at the bottom, and include overall information transfer (Cons. IT Overall) and information transfer for the place of articulation feature (Cons. IT Place).

only two electrodes.

The processors for each of the conditions in Figure 5 were evaluated with the consonant test. Each of the 24 consonants was presented at least 10 times with the male speaker for each of the conditions. The results are presented in terms of overall information transfer (Cons. IT Overall) and the information transfer score for the place of articulation feature (Cons. IT Place). The results and conditions are arranged in order of increasing scores for overall information transfer.

The ranking of conditions from left to right seems to indicate improvements in scores with increases in the total distance along the cochlea spanned by the channels. The addition of virtual channels *per se* does not appear to improve scores, even when that increases the total number of channels. Improvements in scores can be produced by increasing cochlear distance either with single-electrode channels or with virtual channels.

To evaluate these impressions, we conducted a stepwise linear regression analysis of the data, where

Table 3. Independent variables used in the stepwise linear regression analyses of results from the reduced-electrodes study.

Variable	Description
nchans	number of channels
Pivc	fraction interior virtual channels
eacr	exterior virtual channels, reversed phase
eacin	exterior virtual channels, in phase
dist	total cochlear distance swept by the channels
center	center of cochlear distance
space	average spacing between channels
ratio	number of channels / number of electrodes
vc	presence of virtual channels

the dependent variable was either overall information transfer or information transfer for the place of articulation feature. The independent variables included those listed in Table 3.

The regression analyses indicated that the single variable of cochlear distance accounted for 83.3% of the variance in the overall IT scores ($p < .000002$) and for 84.1% of the variance in the place IT scores ($p < .000002$). No other variable accounted for a significant portion of the variance for either dependent variable. The regression equations were:

$$\text{Overall IT} = 3.1*\text{dist} + 84.2$$

$$\text{Place IT} = 9.6*\text{dist} + 51.8$$

The addition of interior virtual channels does not figure in the regression analyses. It may be that intermediate pitch percepts already are produced by CIS processors, using single-electrode channels. That is, even though stimuli are presented nonsimultaneously, intermediate pitches might be produced between adjacent electrodes. A critical question relates to the time over which the central auditory system integrates inputs to make inferences about pitch. Recent studies by Colette McKay and Hugh McDermott of the Melbourne team (personal communication to Wilson, 1993; McDermott and McKay, 1994; McKay et al., 1994) suggest that this interval is at least 400 μs, which is much longer than the time between sequential pulses in typical implementations of CIS processors.

CIS/VCIS Comparisons with Additional Subjects

In parallel with the reduced-electrodes study, we also evaluated full VCIS processors with additional subjects (QPR 6). The results are presented in Figure 6. The initial findings for SR2 are repeated in the top panel. Scores for subjects SR10 and SR13 are shown in the middle and bottom panels, respectively. The consonant tests for SR2 included 24 consonants with both male and female speakers, whereas the tests for SR10 and SR13 included 16 consonants. Tests with both male and female speakers were used for SR10, whereas only the male speaker was used for SR13.

The processors implemented for SR13 used the five apical electrodes only, because stimulation of basalmost electrode 6 elicited a somatic sensation at levels just above auditory threshold. Each subject was fitted with an 11 channel VCIS processor. The 11 channel processors for SR2 and SR10 used interior virtual channels, as illustrated in Figure 4. The processor for SR13 used four interior virtual channels (corresponding to the positions between the five available electrodes) and two exterior virtual channels, formed with a principal pulse on the apicalmost or basalmost electrode paired with a half-amplitude reversed-polarity pulse on the adjacent electrode (i.e., on electrode 2 for a principal pulse on apicalmost electrode 1, or on electrode 4 for a principal pulse on basalmost electrode 5).

In general, the results do not demonstrate any advantage for VCIS processors. Scores from the consonant tests are not statistically different for any of the subjects. Results from the open-set tests are mixed, with somewhat better scores for VCIS in some cases (e.g., the NU-6 test for SR2 and the CID test for SR10) and for CIS in others (e.g., the NU-6 test for SR10). The only large difference between processors is in the NU-6 scores for SR13, where the score for the CIS processor is clearly better than the score for the VCIS processor.

Although the speech reception scores for the two types of processor were similar, each of the subjects expressed a preference for the VCIS processor. Each of them said that the VCIS processor sounded more natural and seemed more intelligible than the CIS processor. SR2 and SR13 also compared the two processors for listening to music, and both said the VCIS processor produced a richer and more natural sound than the CIS processor.

Evaluation of "Sharpened Field" CIS Processors

Studies to evaluate sharpened field processors have been conducted with Ineraid subjects SR2 and SR10. Tests of consonant and vowel identification were used. As indicated above, sharpened field processors use reversed polarity pulses on either side of a principal pulse to alter the shape of the neural response field. Presumably, such a manipulation would sharpen the field over the electrode used for the principal pulse.

Preliminary analysis of the results from these rather extensive studies suggests that identification of vowels may be slightly improved with sharpening, if the amplitude of the reversed polarity pulses is not greater than 20 percent of the amplitude of the principal pulse. In addition, consonant identification appears to be slightly improved under some conditions with small (less 12 percent amplitude) flanking pulses. Larger amplitudes of the flanking pulses produce decrements in performance for consonants. Additional studies may be conducted when our analysis of the present results is completed. For now, however, it appears that any positive effects of field sharpening are likely to be small.

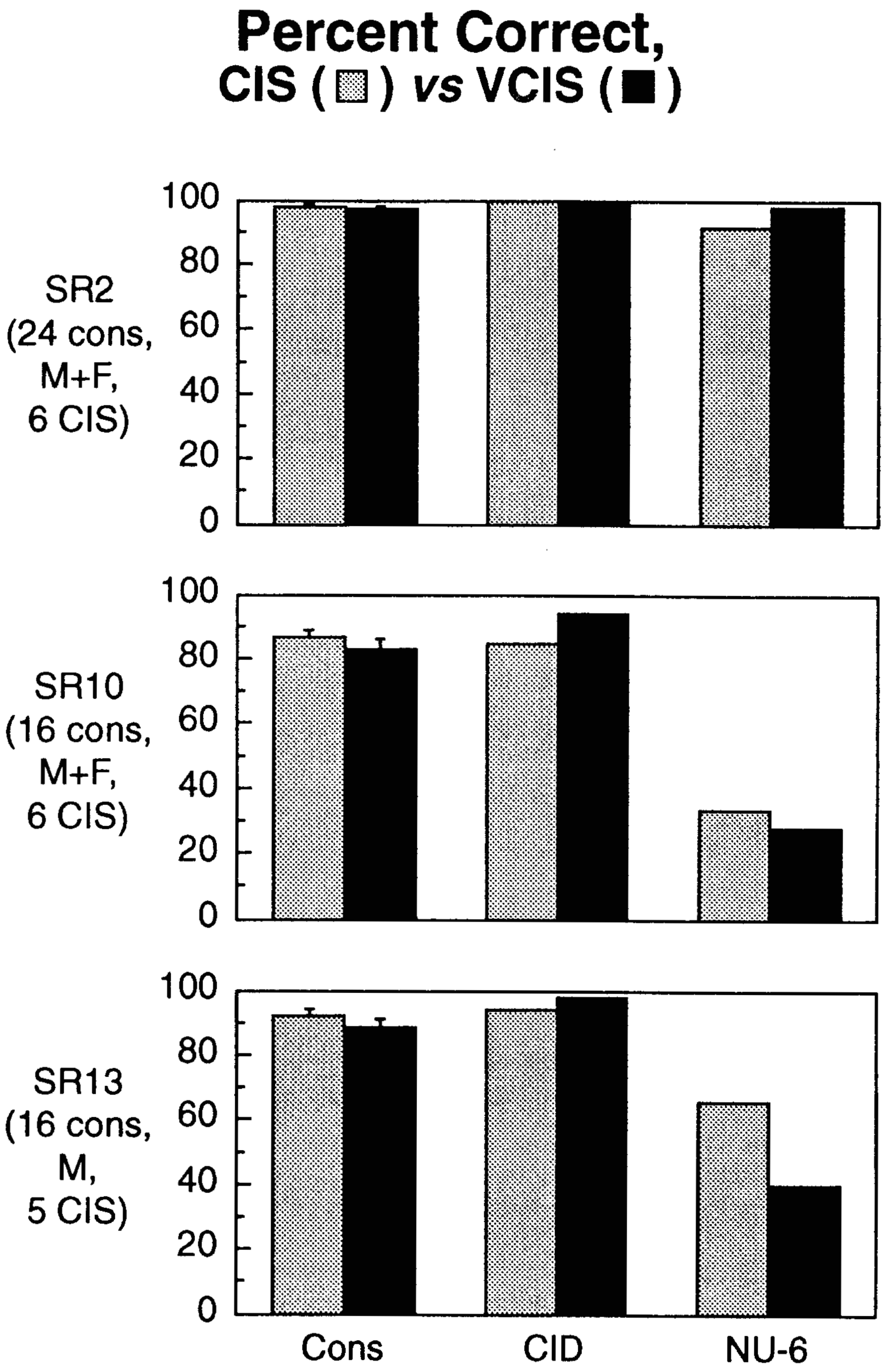


Figure 6. Comparisons of CIS and VCIS processor performance for three subjects. Tests included identification of consonants in an /a/-consonant-/a/ context (Cons), recognition of key words in the Central Institute for the Deaf sentences of everyday speech (CID), and recognition of monosyllabic words from Northwestern University Auditory Test 6 (NU-6).

Discussion

Present implementations of VCIS processors offer no obvious advantage over CIS despite some very encouraging initial results, and despite the preference for VCIS expressed by all three subjects. It is possible that we have not selected the best tests to demonstrate a difference between processors, as suggested by the anecdotal remarks of the subjects. For example, a difficult test of vowel identification might demonstrate a difference between processors with sparse *versus* dense spatial representations (e.g., CIS *versus* VCIS, where each of the processors use the same number of electrodes). Also, tests of speech reception in noise, such as multitalker speech babble, sometimes can demonstrate differences in processor performance that are not demonstrated by tests of speech reception in quiet conditions (e.g., see Skinner et al., 1994).

We are not quite ready to discontinue our studies with VCIS processors. Evaluation of VCIS with different tests seems warranted. Also, we note that alternative implementations of VCIS processors may be superior to the present implementations. For example, selective use of single-electrode and multiple-electrode channels may allow implementation of a processor with a relatively large number of channels (e.g., 7 or 8) with a high degree of perceptual distinctness among channels.

As indicated above, analysis of results from initial studies with sharpened field processors is not yet complete. Depending on the outcome of that analysis, we may conduct additional studies to answer questions raised or to pursue demonstrated possibilities for improving speech reception performance with the sharpened field approach.

VI. Nucleus Percutaneous Study

Studies with seven subjects using an experimental version of the Nucleus device were initiated in this project. Unlike the standard clinical device, the experimental device has a percutaneous connector, which provides direct electrical access to the implanted electrodes. Such access is required for evaluation of CIS processors using typical rates of stimulation (e.g., in excess of 400 pps on each channel of a six-channel processor) and for the recording of intracochlear evoked potentials.

The studies are a result of a cooperative agreement among Cochlear Corporation, Duke University Medical Center (DUMC), and RTI. In broad terms, Cochlear Corporation is providing the devices and supporting clinical costs for the subjects. The surgeries and subsequent medical and audiological services are performed at DUMC by DUMC personnel. The research aspects of the studies are the responsibility of the RTI team, as supported by this project.

Five of the seven subjects have received their experimental implant systems to date. Research studies have begun with four of these five. A total of 21 two-week sessions is included in the studies (three sessions for each subject). Approximately half of those sessions have been completed.

The main purpose of the studies is to evaluate a wide range of processing strategies for patients using the Nucleus electrode array. The principal comparisons have been and will be among CIS processors using different numbers of channels and between various CIS processors and variations of the SPEAK processor used with the present Nucleus clinical device (see Skinner et al., 1994). Each subject has been or will be given a SPEAK processor for daily use away from the laboratory.

We expect to complete a set of core tests with each subject. They include:

1: Determine an optimal number of channels for a CIS processor, using standard parameters of 33 μs/phase pulses with the positive phase leading at the intracochlear electrodes, 833 pps on each channel, a staggered order of channel updates, a 200 Hz lowpass filter in each of the envelope detectors, a fullwave rectifier in each of the envelope detectors, the standard span of frequencies generally used with CIS processors (approximately 350 to 5600 Hz), and the standard preemphasis filter generally used with CIS processors (6 dB/octave attenuation below 1.2 kHz). Comparisons will include processors with 4, 6, 8 and 11 channels. Initial tests of electrode ranking abilities will be used to select the best (most discriminable) sets of electrodes for each condition. In addition, tests will be conducted with processors using the maximum number of discriminable channels and the maximum number of available physical channels (which may be less than 22, in that not all electrodes are available for some subjects).

2: Compare six channel CIS processors using different sets of electrodes. At least two sets will be included (preliminary results indicate that performance can be quite sensitive to choice of electrodes, even with similar or identical spacings between adjacent electrodes). All other parameters will be the same as those listed under point 1 above, i.e., the standard parameters.

3: Evaluate effects of manipulations in the standard parameters for the best of the six channel CIS processors identified in 1 and 2 above. Manipulations will include (a) substitution of an apex-to-base for the staggered update order; (b) use of a 400 Hz lowpass filter in the envelope detectors; (c) use of the opposite polarity of stimulus pulses; (d) use of an extended range of frequencies for the bandpass filters, e.g., 350 to 9000 Hz as opposed to the standard 350 to 5600 Hz; and (e) use of a higher pulse rate, e.g., 2525 pps on each channel.

4: Evaluate a processor that is identical to the best of the six channel CIS processors identified in 1 and 2 above, except that the rate of stimulation on each channel will be reduced to 250 pps.

5: Evaluate a processor that utilizes all of the physically available electrodes and selects the six corresponding channels with the greatest magnitudes of envelope signals for each cycle of stimulation. Cycles will be repeated at the rate of 250/s. All other parameters will be the same as those listed under point 1 above.

6: Evaluate a processor that is identical to the processor of point 5 above, except that the rate at which cycles are repeated will be increased to 833/s.

The above comparisons and evaluations should demonstrate:

- sensitivity of CIS performance to number of channels (point 1)
- sensitivity to electrode choice for a fixed number of channels (point 2)
- sensitivity to parameters other than channel number (point 3)
- sensitivity to reduction in rate of stimulation, to a rate approximating those used in the SPEAK processor (point 4)
- possible benefit of "roving" electrode channels, as in the SPEAK processor (point 5; this processor should provide performance similar to that of the clinical SPEAK processor used by the subjects)
- possible benefit of increasing rate of stimulation for a SPEAK-like processor (point 6)

At the end of each visit by each subject we will continue to evaluate the SPEAK processor and the best CIS processor identified up to that point using six or fewer channels. This will provide a basic comparison of SPEAK and CIS processors, under the condition of extensive daily experience with the SPEAK processor and limited laboratory experience with CIS. The tests include identification of consonants in an /a/-consonant-/a/ context and open set recognition of CNC words, NU-6 words, key words in CUNY sentences in quiet, and key words in different lists of CUNY sentences presented at a +10 dB speech-to-babble ratio. The repeated tests across sessions with the SPEAK processor will allow evaluation of possible learning effects with that processor and will provide contemporaneous comparisons with other processors.

We also plan to evaluate use of a bipolar-plus-one electrode coupling configuration with the SPEAK processor. Results from that evaluation will be compared with results obtained with the standard monopolar configuration for these subjects.

We expect to complete the Nucleus percutaneous study in the period of the next project, which has been approved for funding. Data from the study will be presented in progress reports for that project.

VII. Design for an Inexpensive but Effective Cochlear Implant System

The *1993 Zhengzhou International Conference on Cochlear Implants and Linguistics* was held in Zhengzhou, China, October 23-26, 1993. The Conference was organized by Min-Sheng Dong, M.D., Professor and Chief Surgeon of the Henan Medical University in Zhengzhou, and by Fan-Gang Zeng, Ph.D., Research Scientist at the House Ear Institute in Los Angeles, California. Approximately 130 physicians, scientists and engineers attended the Conference, including representatives from most centers in China involved with the development or clinical application of implant systems. A panel of five experts from the United States and Canada was invited to provide information on recent developments in implant design and performance. The panel also was invited to offer recommendations for the design of an inexpensive but effective implant system, that might be suitable for widespread use in China. The panelists included Gerald Loeb, Stephen Rebscher, Robert Shannon, Blake Wilson and Fan-Gang Zeng. Blake Wilson's participation in the conference was supported by the Commission of Science and Technology of Henan Province. Support for the subsequent evaluation of a prototype system, as described below, was provided by this project. Dewey Lawson and Mariangeli Zerbi conducted the evaluation studies.

The system recommended by the panel includes a speech processor, four pairs of transmitting and receiving coils, and an electrode array with four monopolar electrodes. All implanted components are passive, reducing to a minimum the complexity of manufacture and allowing high reliability. A transcutaneous link is used to minimize the possibility of infection, which in turn should minimize medical costs in maintaining device function. The electrode array has a mechanical memory to position the electrodes in close proximity to the inner modiolar wall of the scala tympani. A reference electrode is implanted in the temporalis muscle. A four-channel CIS strategy is used for the speech processor. The speech processor and transcutaneous link have been evaluated in preliminary tests with a patient implanted with the Ineraid electrode array and percutaneous connector. A prototype of the link, consisting of four pairs of transmitting and external receiving coils, was used, with the outputs of the receiving coils routed to the apical four electrodes of the Ineraid array via the percutaneous connector. The tests included identification of 24 consonants in an /a/-consonant-/a/ context. The subject scored $89 \pm 2\%$ correct with a standard laboratory implementation of the four-channel CIS processor using current-controlled stimuli and $92 \pm 1\%$ correct with the prototype system. These scores show that use of the coils does not degrade performance. Results from other studies indicate that many patients can achieve high levels of speech recognition with CIS processors and monopolar electrodes. The system recommended here includes those components along with (a) an electrode array that may improve placement of the intracochlear contacts and (b) a simple transcutaneous link that does not degrade performance.

The excellent performance obtained with the prototype system supports development of a commercial system that can be easily manufactured at low cost. Such a system may be especially appropriate for widespread use in countries like China, with large numbers of people with profound hearing loss and highly limited resources for procurement of medical devices.

VIII. Representation of Complex Tones by Sound Processors for Implanted Auditory Prostheses

Background

Crudely mimicking functions of normal acoustic hearing, multichannel sound processors for implanted auditory prostheses seek to convey sound spectra in two distinct ways -- place of stimulation and stimulation periodicity.

Place information is conveyed by selectively stimulating groups of neurons associated with different locations along the organ of Corti and, hence, different sensations of pitch. The effectiveness of this approach for conveying spectral information depends on identifying perceptually distinct channels of stimulation, which may correspond to stimulation with different physical electrodes or different combinations of electrodes capable of addressing distinct populations of neural elements. The number of such channels available to any individual patient may depend on the number of implanted electrodes, the site of implantation, and patient differences such as extent and pattern of neural survival. For efficient and unambiguous information transmission it is desirable that stimulation of each channel be as independent as possible. Non-simultaneous stimulation of channels in an effort to avoid vector summation of fields and resulting channel interactions has been shown to improve scores in speech recognition tests for many patients. In some cases, manipulation of the order and/or rate of stimulation further improves speech reception, perhaps by further increasing the independence of the available channels in patients subject to non-simultaneous interactions involving transient polarization of cell membranes.

Distinct frequency bands are the sources of spectral information to be conveyed as place of stimulation via the perceptually distinct channels. The design of appropriate sets of such bands includes choices of the number of distinct bands to analyze, the overall frequency range of the set of bands, the frequency range of each individual band, and the sharpness of the band edges (filter order).

Periodicity information is conveyed by temporal variations in the stimulus amplitudes of each channel. Upper limits on the frequency of such amplitude variations arise from at least two considerations. (1) If stimuli are being presented nonsimultaneously to reduce channel interactions, then the signal on any one channel will be a series of pulses occurring at some rate of stimulation. If the amplitudes of those pulses are modulated by signals including components at frequencies greater than one half the stimulation rate, then aliasing can occur. A manifestation of an insufficient sampling rate, aliasing will add anomalous low frequency components to the amplitude modulation of the stimulus pulses. Thus the highest frequency periodicity information conveyed within any one channel should not exceed one half the pulse stimulation rate of that channel, and if the rate of stimulation is reduced (*e.g.*, to reduce nonsimultaneous channel interactions) the maximum modulation frequency may have to be reduced accordingly. [We note, however, that in certain circumstances aliasing noise may help a patient recognize the presence of high frequency speech cues.] (2) While it has been reported that some patients with implanted auditory prostheses can detect differences in periodicity information up to 2000 Hz or so (Hochmair *et al.*, 1983), most cannot discriminate such differences above a few hundred Hz (*e.g.* Shannon, 1993). Providing additional stimulus modulation components at frequencies too high for a given patient to utilize may even reduce that patient's performance levels on speech recognition tests.

Thus, when deriving a temporal envelope signal to characterize variations in energy in the chosen frequency band for each channel, it is important to use a smoothing filter to impose an appropriate upper frequency limit on such variations -- one appropriate both to the individual patient's perceptual abilities and to the particular processor's rate of stimulation on each channel.

Traditional pitch perception studies using single pure tone stimuli can be administered to patients with multichannel sound processors and multi-electrode implanted arrays. The data from such studies reflect not only the functioning of basic place and periodicity mechanisms but also artifacts resulting from processor channel design. Either some pure tones will influence the signals in more than one channel because of adjacent bandpass filter overlap, or some pure tones will fall between filters and therefore not be represented as salient electrical stimuli. The former case is typical of processors in use today.

While some overlap between adjacent frequency bands may enhance place pitch discrimination performance for single pure tone stimuli (Dorman, 1993), such a design raises potential problems for periodicity information and for pitch perception of real-world complex tone stimuli. The relative phase generally will be uncontrolled between the envelope variations of a single pure tone as conveyed in two channels whose bands overlap its frequency. In many cases the pure tone periodicity will be a frequency too high to be conveyed as a modulated amplitude -- either because of the aliasing constraint or because of limits in the patient's perceptual abilities.

What information will be conveyed when a complex musical tone is analyzed by such a multichannel sound processor? Consider a class of complex tones, each composed of several pure-tone partials -- one or more fundamentals and/or various upper harmonics. When such a complex tone is input to a multichannel processor that has slightly overlapping frequency bands, each partial may affect the output in a single channel exclusively or (if the partial's frequency falls in a region of band overlap) the outputs in two adjacent channels. Correspondingly, a given processor channel may convey information for (1) a single partial exclusively, (2) a single partial that also affects the output of an adjacent channel, (3) more than one partial exclusively, (4) more than one partial that also affects the output of one or both the adjacent channels, or (5) some combination of exclusive and non-exclusive partials.

In contrast to the situation for single pure tones, complex tone inputs might result in uncontrolled and/or unnatural interactions among channels, confounding perceptions of pitch or musical intervals. On the other hand, multichannel processor responses to complex tones might provide additional information useful in pitch perception. As an example of the latter possibility, note that the presence of two partials exclusively in the same channel's frequency band might well produce beating at a frequency low enough to be conveyed to (and perceived by) the patient as channel envelope modulation. Such information conceivably could provide much more support for an accurate pitch interval determination than a pair of partials each of which was conveyed in a separate channel or under less controlled circumstances (*e.g.*, in a region of overlapping frequency bands).

Approach

In an initial set of studies, a variety of digitally synthesized complex tone stimuli were input to selected continuous interleaved sampling (CIS) processors which in turn stimulated a research subject's intracochlear electrode array. The data collected initially included anecdotal descriptions of the percepts elicited by individual stimuli, anecdotal descriptions of the perceived differences between members of various pairs of complex stimuli, and surveys of relative overall pitch judgments within such complex stimulus pairs.

Each **stimulus** was approximately 0.5 seconds in duration (22,000 samples at 44.1 kHz), including approximately 11 msec each of linear fade-in and fade-out (500 samples each). Each stimulus was presented from a digitally synthesized file of 16-bit samples, constructed by adding pure tone sinusoidal partials selected from the harmonics either of a single fundamental or of two fundamentals differing by a chosen musical interval. Single fundamentals were chosen from a four octave *equal tempered* scale ascending from 110 Hz. When two fundamentals were to be separated by a consonant pitch interval within a single stimulus, however, the frequency interval was made exact (*i.e. just intonation* was used within stimuli). Consistent with the spectral envelopes of typical musical tones, relative nth harmonic amplitudes proportional to 1/n were chosen to ensure relatively strong beat phenomena. [See Rossing, 1990].

For simplicity of analysis, we also required that each partial used in a stimulus fulfill an additional criterion with respect to the CIS processor for which it was intended. The criterion ensured that each partial be represented exclusively in a single processing channel -- falling at a frequency that put it within 1 dB of the maximum sensitivity of that channel's input passband, for instance, and at least 10 dB down in any adjacent band. In some cases a minimum of 20 dB of adjacent band rejection was imposed, further reducing the number of available partials. Tables were prepared of harmonic partials fulfilling the various combinations of criteria and experimental conditions.

Using those tables, we designed combinations of partials to test the efficacy and relative salience of various potential mechanisms for conveying subtleties of perceived pitch and timbre to cochlear implant patients. Examples of such mechanisms included harmonic consistency of partials between and within channels and beat rate patterns between and within channels. Some stimuli were designed with conflicting cues to assess their relative salience. Our design emphasis in the experiments constructed using such tables was to eliminate uncontrolled interaction between two bands due to partials that affect modulation envelopes in both, while preserving the possibility of two partials interacting within a band, so long as both partials affect that band's envelope exclusively.

We use various combinations of the stimuli assembled in these tables to study the mechanisms of complex tone pitch perception with existing speech processors designed for implanted auditory prostheses. We anticipate that this may lead to better ways of supporting complex tone pitch perception in future processor designs -- both for better voice pitch perception within speech and for improved access to music via implanted prostheses. While the nature of electrical stimulation in such prostheses allows only the most crude mimicking of the functions of normal hearing, it also provides the possibility of stimulus patterns unattainable in normal listeners. The study of perceptions arising from such "unnatural" stimuli may provide additional insights regarding CNS processing of auditory input. Finally, we anticipate that use of such carefully constructed complex tone stimuli may constitute a useful tool in the diagnosis of significant differences among implanted patients and optimization of processors for individual patients.

The research **subject** chosen for the pilot studies was Ineraid patient SR2. He was selected on the basis of (1) excellent performance with existing processor designs, already extensively studied in our laboratory and elsewhere, (2) exceptional analytic and descriptive abilities regarding his auditory percepts, (3) experience as a musician -- both before losing his normal hearing and recently with an analog clinical prosthesis, and (4) familiarity with some basic music theory. Percutaneous access is available to all six of the electrodes implanted in SR2's right cochlea. He is right handed.

The CIS **processors** used in the preliminary studies included a standard 6-channel design [number 163b] that had been used by the subject for a wide range of previous studies in our laboratory, an 11-channel virtual channel interleaved sampling (VCIS) processor [200b] also previously evaluated in our lab, and a six channel CIS variation [284] without the normal preemphasis. An additional processor [355] used in more recent complex tone studies was identical to 163b except for its bandpass filters, which were 24th order instead of 12th. [For a general description of the design of CIS processors see Wilson, *et al.,* 1991. VCIS designs are described in QPR 6.]

The **experimental conditions** explored in the preliminary pilot studies [QPR 4] included presentation of the two complex tones of a stimulus pair with and without a one second intervening delay. As the preemphasis filter typically included in CIS and VCIS processors effectively contributes a spectral weighting proportional to harmonic number over part of the represented frequency range (attenuating components below 1.2 kHz at 6 dB/octave), a condition essentially correcting for this effect was included among the pilot studies. The effect of order of presentation within each stimulus pair also was explored. There was no balancing of the overall loudness across stimuli, within or among pairs.

The statistical portions of our pilot studies were administered by interactive computer programs. Statistical surveys with any given set of stimuli were completed before any anecdotal studies with the same stimuli were begun, to avoid biasing the statistical results by any strategies or analytic categories acquired by the subject in the course of describing his percepts. The same complex tone stimulus pairs were used in such anecdotal comparisons, initially presented manually by an experimenter and occasionally augmented by additional comparison tones. Later studies following up on anecdotal reports were administered by computer program.

Harmonic Partials of a Single Fundamental

One category of tests was concerned with whether a patient might be able to make absolute pitch judgments on the basis of adjacent harmonics within complex tones of a single fundamental, and with differences in such an ability depending on whether the adjacent harmonics were conveyed via a common channel of stimulation or via separate channels. Wherever adjacent partials were presented to a common channel, the beat rate created in that channel was at the fundamental frequency of the complex tone -- whether or not the fundamental was actually present as a partial -- or some larger multiple of that frequency. Among the tests in this category were:

- ***Pitch salience of a single complex tone.*** Relative pitch judgments for pairs of complex tone stimuli with independent control of patterns of channel utilization and patterns of included harmonics.
- ***Perception of a missing fundamental.*** Investigation of the pitch perceived when various subsets of harmonically related upper partials were presented simultaneously, in the absence of the fundamental to which they were all related.
- ***Relative salience of conflicting cues.*** Relative pitch judgments for pairs of complex tone stimuli designed to convey conflicting information (*e.g.* rising fundamental combined with the use of a lower frequency channel).
- ***Perception of chromatic intervals between sequentially presented complex tones.*** Systematic exploration of intervals from the tritone to the octave, using comparable sets of harmonics and channels.
- ***Perception of a sequentially presented semitone interval as represented by various subsets of harmonics.***

Simultaneous Complex Tone Intervals (Harmonic Partials of Two Distinct Fundamentals)

Another category of tests was concerned with the perception of simultaneous complex tone intervals, *i.e.* complex tones whose partials include harmonics of two distinct fundamentals. Again we focused initially on the salience of cues resulting from adjacent partials and on the various place and periodicity roles played by such partials. In this case, however, the beat frequencies between pairs of partials conveyed in common channels varied widely, with certain patterns being characteristic of particular ratios between fundamental frequencies. Many of the adjacent partial beat frequencies were relatively small fractions of the lower fundamental and thus more accessible as periodicity information via electrical stimulation.

We began by examining the distinctive patterns among adjacent partials -- and the potential beat frequencies between them -- for the five most consonant intervals between complex tone fundamentals: octave, perfect fifth, perfect fourth, major third, and minor third. Going from the octave to the minor third in this sequence, one finds progressively larger numbers of distinct frequency differences between adjacent partials, including progressively lower potential beat frequencies between such partials. Beating between adjacent partials at such frequencies may well be salient to electrically-stimulated auditory implant patients even when the absolute frequencies of the partials are far too high to provide usable periodicity information.

Subsequent Studies

There have been six subsequent studies [QPR 8], two of which represented systematic explorations of perceptual categorizations that the subject volunteered anecdotally during the original pilot studies. A third was designed to exploit and further explore some of the subtle abilities demonstrated by the subject in the earlier work. The remaining studies were designed to probe the limits of specific previously observed abilities and effects.

- ***Identification of Constituent Tones.*** The subject was asked whether or not a candidate tone was made up entirely of partials contained in a more complex reference tone. The data were analyzed both (1) in terms of how accurately the subject's responses could be predicted using various hypotheses involving potential cues, and (2) in terms of how various structural attributes of stimuli affected the subject's accuracy.
- ***Processor Bandpass Filter Order Effects.*** This study was a test of the sensitivity of our complex tone data to traces of spectral components in processor channels adjacent to the ones for which they were intended. Constituent tones identification tests were repeated with a CIS processor based on 24th order, rather than 12th order, bandpass filters.
- ***Interval Consonance Judgments in a Nontraditional Context.*** A particular nontraditional musical scale [Mathews and Pierce, 1989] supplies many of the same structural cues present in traditional consonant intervals, but sounds quite different to people with normal hearing. This study investigated whether a subject using a CIS processor could detect a difference. It collected anecdotal descriptions of both sequential and simultaneous complex intervals.
- ***Systematic Examination of Perceptual Category Assignments: Single Stimuli.*** A number of descriptive terms volunteered frequently by the subject during earlier complex tone studies formed the basis for this more structured, automated interview regarding a wide range of stimulus tones. Analysis of these data included association of descriptive categories both with stimulus structures and with the use of other such categories. .
- ***Systematic Examination of Perceptual Category Assignments: Stimulus Pairs.*** The same tones described in the single stimulus interviews later were presented in pairs in another automated interview. Comparison judgments along eight different perceptual dimensions were obtained.
- ***Inconsistency Detection Thresholds.*** Complex tones typically consist of several harmonics of a common fundamental. In such cases there is a high degree of consistency among the partials, each one being an integer multiple of the fundamental frequency and any pairs of adjacent harmonics beating at that same frequency. This study was designed to determine roughly how large an inconsistency must be in order to alter a complex tone percept for a CIS processor user. The stimulus pairs perceptual category interview was repeated for pairs of complex tones with gradually increasing inconsistencies in beat rates and/or absolute frequencies.

Most of the **stimuli** used in these later studies were like those used earlier. One study additionally required the synthesis of partials based on fundamentals separated by an arbitrary frequency interval. In another, stimuli were restricted to odd harmonics only and to fundamentals along a highly unusual musical scale. The basic approaches to stimulus design and synthesis, however, have remained the same throughout all the studies to date.

While our complex tone studies continued to be restricted to work with a single **subject**, knowledge gained from other types of investigations with him added whole new dimensions to the unique set of advantages he offered such pilot research. In psychophysical experiments we obtained periodicity pitch saturation data for the same pulse configuration used in SR2's processors (33 ms per phase balanced biphasic). Direct intracochlear evoked potential (EP) measurements for similar stimuli indicated an accompanying onset of a failure of EP magnitudes to accurately represent each pulse's amplitude within a stimulus pulse train. We have demonstrated that an ensemble model of electrical neural stimulation can accurately predict SR2's EP responses to a wide range of stimulus patterns like those produced by his processors. Finally, SR2 recently has begun use of a six channel, 40 μs/phase CIS strategy on an everyday basis as part of a study by Eddington, *et al.*, at Massachusetts Eye and Ear Infirmary.

Summary of Principal Findings

Preliminary results from the early studies included indications of the importance of channel (place) cues, the importance of intrachannel beat frequencies as (temporal) cues, and the predictability of relative strengths among competing spectral cues in some cases. We noted two distinct patterns of changes in percept that occurred after extended initial comparisons of certain pairs of stimuli: (1) an irreversible change, after which the original percept could not again be found by the subject, and (2) the sudden emergence of an ambiguity, with the subject thereafter able to obtain either percept at will. Examples of complex tone stimuli were found for which the use of preemphasis filtering, the choice of pair presentation order, and the imposition of an interstimulus delay would, individually or in combination, dramatically affect the subject's percepts We observed, on occasion, a surprising ability of the subject to recognize musical intervals and accurately to match his (unmonitored) voice pitch to that of an electrically conveyed complex tone.

Among the results of the subsequent studies were:

- The presence of differences in more than one structural attribute of two stimuli (among fundamental frequency, harmonics, and channels involved) substantially increased the subject's accuracy in detecting a difference. This was especially true if one of the differing attributes was channels. In the absence of adjacent harmonic pairs, channel differences were more easily recognized by the subject than purely harmonic differences. For complex tones with relatively few partials, however, patterns in the included harmonics were at least as good a predictor of the subject's constituent tones identifications as patterns in the channels involved. The combination of both patterns was a better predictor than either alone.
- Among complex tones involving adjacent harmonic pairs in various contexts, both intrachannel and interchannel pairs appeared useful to the subject in recognizing differences in (implied) fundamental frequencies. The presence of intrachannel pairs generally increased the subject's accuracy identifying constituent tones but, in comparisons of complex tones based on the same fundamental, the presence of such a pair (beating at a frequency consistent with the fundamental) reduced the likelihood of perceiving simultaneous differences in both harmonic and channel patterns. The presence of interchannel pairs of adjacent harmonics was particularly helpful when the only structural attribute difference was fundamental frequency or harmonic pattern.
- The use of 24th order bandpass filters rather than 12th order made a significant difference in constituent tones identification tests when the candidate tones involved more than a single partial. Simple models predicting the subject's responses on the basis of stimulus structure generally were more accurate for the higher order filters. In some of the most complex comparisons investigated, the higher order filters also supported better accuracy in the subject's judgments themselves.

- Despite many similarities to the stimulation patterns produced by traditional consonant intervals between complex tones, the subject immediately recognized that consonant Bohlen-Pierce intervals were fundamentally different. With experience, he became willing to accept them as analogous.
- The subject reliably reported increases in overall pitch in complex tones with small harmonic inconsistencies in a beat frequency of 208 Hz, whether the simultaneous reference in another channel was a fixed 208 Hz beat or a single partial at 624 Hz. With a fixed 208 Hz reference beat in another channel, he reliably detected small inconsistencies in 624 Hz and 1040 Hz single partials.
- Complex tones described by the subject as "dissonant" were likely to exhibit intrachannel beating between adjacent partials separated by roughly 200 Hz. Complex stimuli described as "single pure tones" included intrachannel beating at any of a wide range of beat frequencies, but seldom around 200 Hz. Stimuli described as "pleasant combinations of tones" usually had four or more harmonics -- whether of a single or different fundamentals -- and included both interchannel and intrachannel pairs.
- Well defined statistical patterns relate several of the subject's perceptual categorizations to structural attributes of the complex tone stimuli they describe.

Implications for Speech Processor Design

Our findings indicate that envelope fluctuations caused both by single harmonic modulation of a channel and by beating between intrachannel pairs of adjacent partials can usefully influence judgments as to the pitch and timbre of musical tones. We have observed this to be true over a range of channel modulation frequencies wider than that fully supported by the CIS processors SR2 has used to date. Combined with a knowledge of the relationship between rate and perceived pitch for this subject and with the results of modeling calculations validated by intracochlear evoked potential measurements, these findings strongly suggest that SR2's perception of speech might be significantly improved by CIS processors utilizing still higher pulse rates and appropriate envelope filtering. More subtle differences in the inputs to a CIS processor can be heard and interpreted if ways can be found to convey them unambiguously via processor channel signals to the eighth nerve.

The present subject is limited to a maximum of six CIS channels. While the pattern of channels stimulated in a given instance clearly is a powerful source of information as to the quality of a sound, intrachannel beats also have proven important. Indications are that increasing the number of perceptually distinct CIS channels would provide more detailed information through patterns of channels stimulated. Beyond some point, however, increasing the number of channels might reduce a user's access to other spectral information. An increased number of increasingly narrow channel passbands would inevitably reduce the number and frequency range of intrachannel beats, and increase the fraction of partials conveyed, perhaps less helpfully, in more than one channel. There may be a very fine balance to be struck between spatial and temporal modes of conveying information via cochlear implants in order to optimize perception of spectral information, and this may be true quite aside from questions of electrode design, electrode position, and the limits of neural spatial resolution.

We have identified some significant abilities and determined some limits imposed on those abilities by present CIS processors. Among other things, our complex tone studies will serve as benchmarks for assessing future improvements in processing strategies. Our now extensive database of stimuli and percepts will make such assessments much more efficient and precise than has been possible to date.

Implications for Perception of Musical Sounds by Implant Patients

In carefully controlled isolation, a wide range of potential spectral cues contained in the structure of complex musical tones can be conveyed to at least one user of a CIS processor. While that fact alone does not indicate that such cues could be utilized in less controlled circumstances, we note that the abilities demonstrated in these studies begin to offer an explanation for the richness and subtlety of SR2's descriptions of what he hears when highly complex recordings of musical performances are input to various CIS processors.

A Musical Instrument for Users of Cochlear Implants

In our experience, a majority of cochlear implant users find listening to music via their speech processors profoundly disappointing if not unpleasant. This seems especially true for those patients who had the most musical experience and/or training before the loss of normal hearing. (See also Gfeller and Lansing, 1991.)

In view of the growing number of people relying on CIS processors for hearing and understanding speech, research like that described in this report may find applications not only in improved speech processing strategies but also in musical instruments. We already know enough to contemplate the design of music synthesizers optimized for producing inputs to specific cochlear implant processors. Thus far, of course, our work on complex tone representation and perception via cochlear implants has been limited to a single research subject and we cannot be certain of the generality of our findings. One thing we have seen in that subject, however, we know to be true of others: a strong desire to experience music again with something of its remembered consonance, detail, and subtlety.

Such a synthesizer, generating an analog signal expressly for direct electrical input to a CIS processor, would produce music that need never exist in the form of mechanical vibrations or pressure waves. In the hands of an otherwise profoundly deaf musician on an everyday basis, the musical potential of a reasonably flexible synthesizer design could be explored much more fully and quickly than in a laboratory. With little effort, researchers could look over the shoulder of the musician as he or she synthesized auditory percepts ever closer to remembered musical instrument sounds. Analysis by synthesis is a well established technique, and the knowledge to be gained from such an exercise in this context is obvious. Most musicians would likely find more satisfaction, however, in creating satisfying new timbres without acoustical precedent, and compositions that exploit them. The opportunity of listening to the signals produced by such synthesis notwithstanding, the hearing researcher inevitably will feel handicapped by the inability ever to really know what the musician heard or intended. It is important to remember that such a difficulty is not at all unusual in music, especially in the case of compositions and performance practice from days before the existence of acoustic recording technology. It will be an interesting challenge to see both how well "acoustic" music can be conveyed to the user of a cochlear implant and how well the subtleties of her/his musical creations can be conveyed to people with "normal hearing" and other users of implants.

IX. Temporal Representations with Cochlear Implants

A focus of recent work in our laboratory is the measurement and improvement, if possible, of temporal representations with cochlear implants. By temporal representations we mean the temporal patterns of responses evoked in the auditory nerve by electrical stimuli and the fidelity with which those patterns represent the time waveforms of the stimuli.

Studies of temporal representations have included (a) development of a model to predict the population response of the auditory nerve to electrical stimulation, as described in QPR 7; (b) recordings of intracochlear evoked potentials for a wide range of stimuli, as described in QPRs 7 and 11; (c) measurement of psychophysical responses to some of the same or similar stimuli, as mentioned in QPR 7; and (d) development of strategies for the repair of demonstrated deficits in temporal representations, as described in QPR 9. Work to develop the population model was supported under a separate NIH project, Project IV of NIH P01 DC00036. The remaining activities were supported by the present project.

Recordings of intracochlear evoked potentials (EPs) have been made with implant patients having percutaneous connectors, i.e., patients with either the Ineraid device or an experimental version of the Nucleus device that includes a percutaneous connector. Potentials are measured at unstimulated electrodes in the implant array. An example is presented in Figure 7 for trains of identical pulses presented to electrode 3 in the implant array of Ineraid subject SR2. Potentials following the pulses were recorded differentially between intracochlear electrode 4 and the ipsilateral mastoid. Body potential was measured with a reference electrode at the wrist. Additional details on the recording technique are presented in the caption for Figure 7 and in QPR 7.

Stimuli used for EP studies have included

- Trains of identical pulses, with various pulse rates, pulse amplitudes, and burst durations.
- Pairs of pulses, including identical pulses with a wide range of interpulse intervals and pulses with fixed interpulse intervals and a wide range of amplitudes for the first pulse.
- Sinusoidally amplitude modulated (SAM) pulse trains, with various carrier rates, modulation frequencies, modulation depths, and burst durations.
- Single pulse probe following pulse train and SAM pulse train maskers, with wide ranges of parameters for the maskers. The interval between the offset of the maskers and the onset of the probe pulse was varied over six logarithmic steps from 2.5 to 80 ms.
- Pulse trains with either linear or exponential onset ramps of pulse amplitudes.
- Pulse trains with other manipulations in pulse amplitudes.
- The pulsatile outputs of a single-channel *continuous sampling* (CS) speech processor.

Responses to Pulse Trains and SAM Pulse Trains

An example of one subject's patterns of responses to pulse trains and SAM pulse trains is presented in Figure 8. This figure shows the normalized magnitudes of the evoked potentials (measured as the

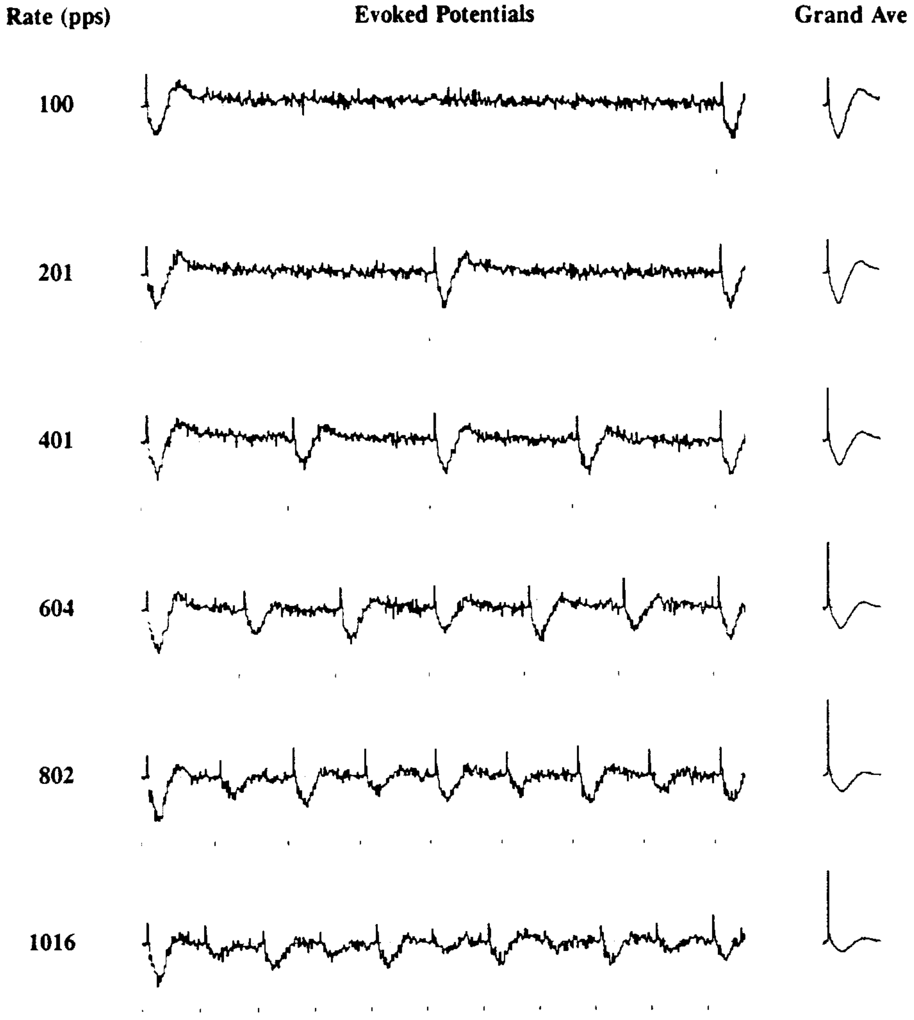


Figure 7. Recordings of intracochlear evoked potentials for Ineraid subject SR2. Stimuli included 16.4 μs/phase pulses presented at the indicated rates to electrode 3 in the Ineraid array (the array has 6 intracochlear electrodes, with electrode 1 the most apical). Pulse amplitude was 750 μA for all illustrated conditions. This amplitude produced a most comfortable loudness (MCL) percept for the pulse rate of 1016/s. Lower pulse rates at the same amplitude produced lower loudnesses. The times of pulse presentations are indicated in the figure with the short vertical lines beneath each EP trace. Potentials were recorded differentially between intracochlear electrode 4 and the ipsilateral mastoid. Body potential was measured with a reference electrode at the wrist. A blanker circuit was used during pulse presentations to reduce the magnitude of pulse artifacts in the recordings. In addition, responses from 200 sweeps of pulses with the positive phase leading were added to responses from 200 sweeps of pulses with the negative phase leading. As illustrated, these two procedures reduced pulse artifacts to a low level (residual artifacts are seen in the "spikes" preceding each EP). The average of EPs following each pulse for a given condition is shown in the right column, under the heading of "Grand Ave." The horizontal dotted lines in the EP columns indicate zero potential. Note that EPs fail to follow pulses with equal magnitudes for rates of stimulation above 201 pps.

difference in amplitudes of the first negative peak and the first positive peak in the EP waveforms, see QPR 7 for additional details) following each stimulus pulse for the entire 1000 ms of each record. The initial and final 100 ms of the records are shown in Figures 9 and 10, respectively. In addition to the EP magnitudes, normalized amplitudes of the stimulus pulses are indicated in Figures 9 and 10.

For reference, this subject (SR2) enjoys high levels of speech recognition with his implant. He obtains percent correct NU-6 word scores in the high 90s for a variety of CIS processors used in conjunction with his Ineraid electrode array.

Figure 8 shows a decrement in response over time for trains of identical pulses. This is most evident for the 401 pps stimulus, where the magnitudes of EPs following each pulse continue to decline out to about 300 or 400 ms.

Although small, decrements in response over time also can be observed for the 100 pps stimulus. This is most easily seen by comparing the upper panels of Figures 9 and 10.

In contrast to the slow decrement in response observed for the 401 and 100 pps stimuli, an alternating pattern of response is observed for the 1016 pps stimulus. The difference between the relatively large EP magnitudes for odd-numbered pulses and the relatively small EP magnitudes for even-numbered pulses declines over time. Ultimately, the EP magnitudes for odd- and even-numbered pulses become indistinguishable, as may be seen in the middle panel of Figure 10. The average of EP magnitudes for sequential 1016 pps pulses also declines over time. Most or all of the decrements appear to occur within the first 300 to 400 ms, as with the 100 and 401 pps stimuli.

Patterns of responses to SAM pulse trains reflect to some extent the modulation of pulse amplitudes. For the 101.6 Hz modulation, responses are in synchrony with the periodicity of the modulation waveform. A detectable response is observed for the fifth stimulus pulse in each cycle of 10 pulses. The normalized amplitude of this pulse is 0.905. A similar or larger response is observed for the sixth pulse in each cycle. The normalized amplitude of this pulse is 1.0. A diminished or no detectable response is observed for the subsequent, lower amplitude pulses in each cycle.

Details in the patterns of response change over the first several cycles. A just detectable response is elicited by the fourth pulse (normalized amplitude of 0.655) in the first cycle (Figure 9).

In addition, the responses to pulses five and six change somewhat over the first several cycles. EP magnitudes are quite similar for those pulses in the first cycle. In subsequent cycles the response to pulse five is progressively diminished over the first 200 ms of the stimulus while the response to pulse six is slightly augmented for cycles two and three and then maintained for the remainder of the record (Figures 9 and 10).

Although minor variations are observed in responses from cycle to cycle, the peak magnitudes of EPs across cycles are approximately uniform for the full duration of the 101.6 Hz SAM stimulus.

For the 406.4 Hz modulation condition, a relatively large response is elicited by stimulus pulse two,

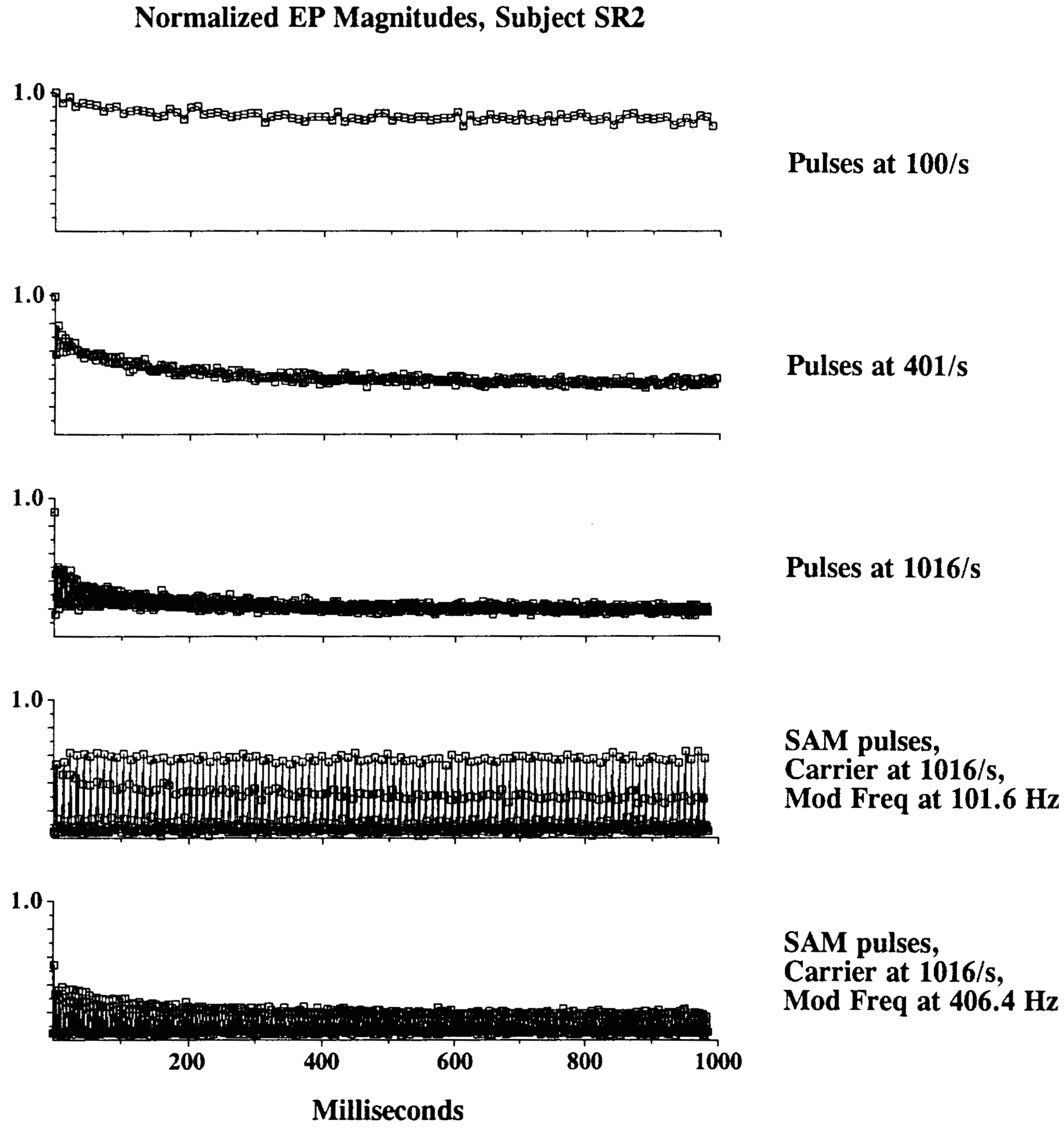
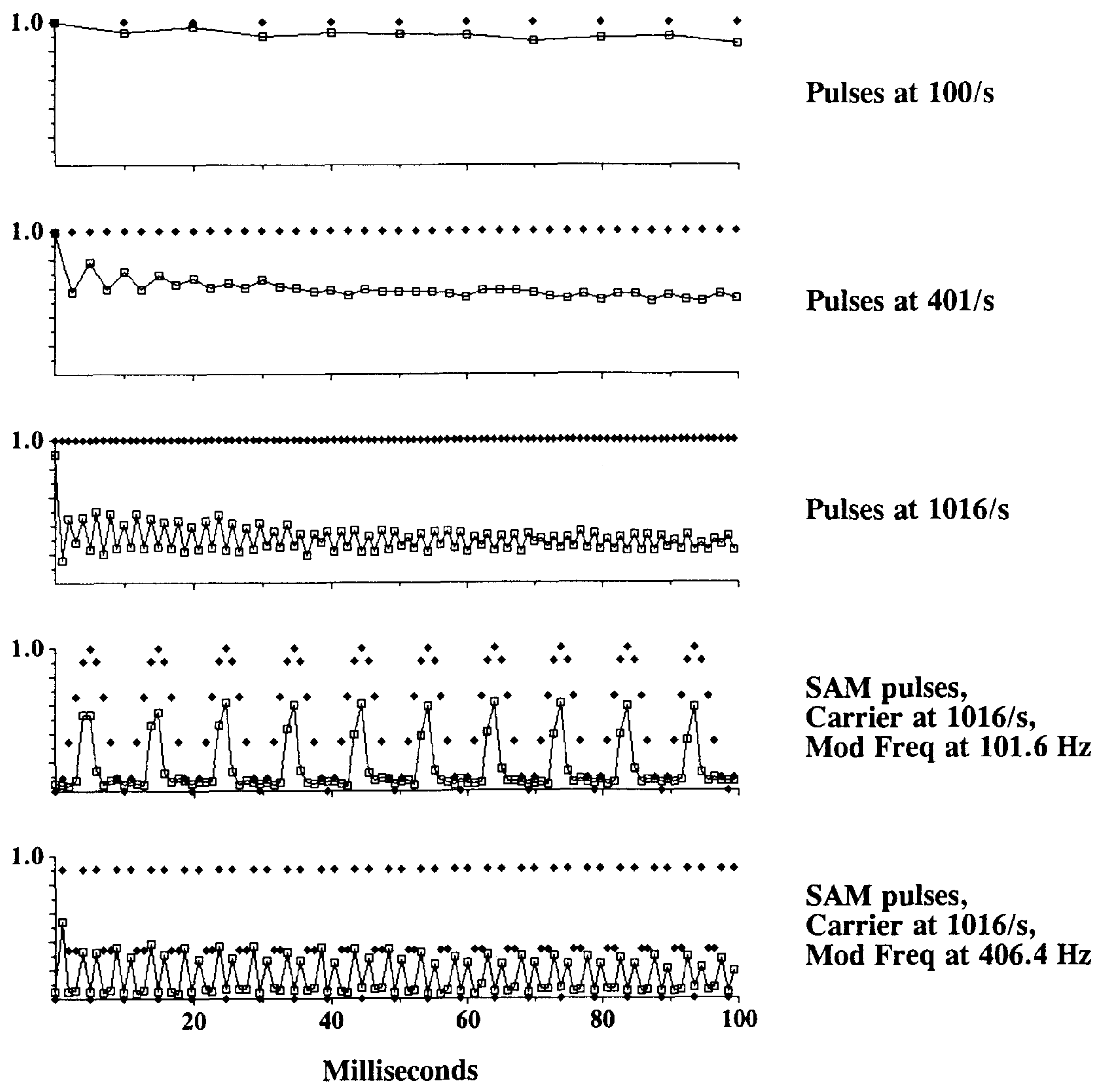


Figure 8. Normalized magnitudes of evoked potentials for Ineraid subject SR2. Stimuli were delivered to electrode 3 of the implant with reference to a remote electrode in the temporalis muscle. Potentials were recorded differentially between electrode 4 and an external electrode at the ipsilateral mastoid. The top three panels show EP magnitudes for trains of identical pulses with the indicated rates of pulse presentations within the trains. The bottom two panels show EP magnitudes for sinusoidally amplitude modulated (SAM) pulse trains with the indicated carrier rate and modulation frequencies. The level of the carrier, and the amplitude of the pulses in the trains of identical pulses, was 290 μA. The duration of all pulses was 32.8 μs/phase. The condition involving the presentation of identical pulses at 1016/s produced a most comfortable loudness (MCL) percept. Lower loudnesses were produced for all other conditions. The maximum EP magnitude across these five conditions for this subject was 90.2 μV.

Normalized EP Magnitudes and Stimulus Pulse Amplitudes, Subject SR2



Pulses at 100/s

Pulses at 401/s

Pulses at 1016/s

SAM pulses,
Carrier at 1016/s,
Mod Freq at 101.6 Hz

SAM pulses,
Carrier at 1016/s,
Mod Freq at 406.4 Hz

Figure 9. First 100 ms of the records shown in Figure 8. Open squares show normalized EP magnitudes and filled diamonds show normalized pulse amplitudes.

Normalized EP Magnitudes and Stimulus Pulse Amplitudes, Subject SR2

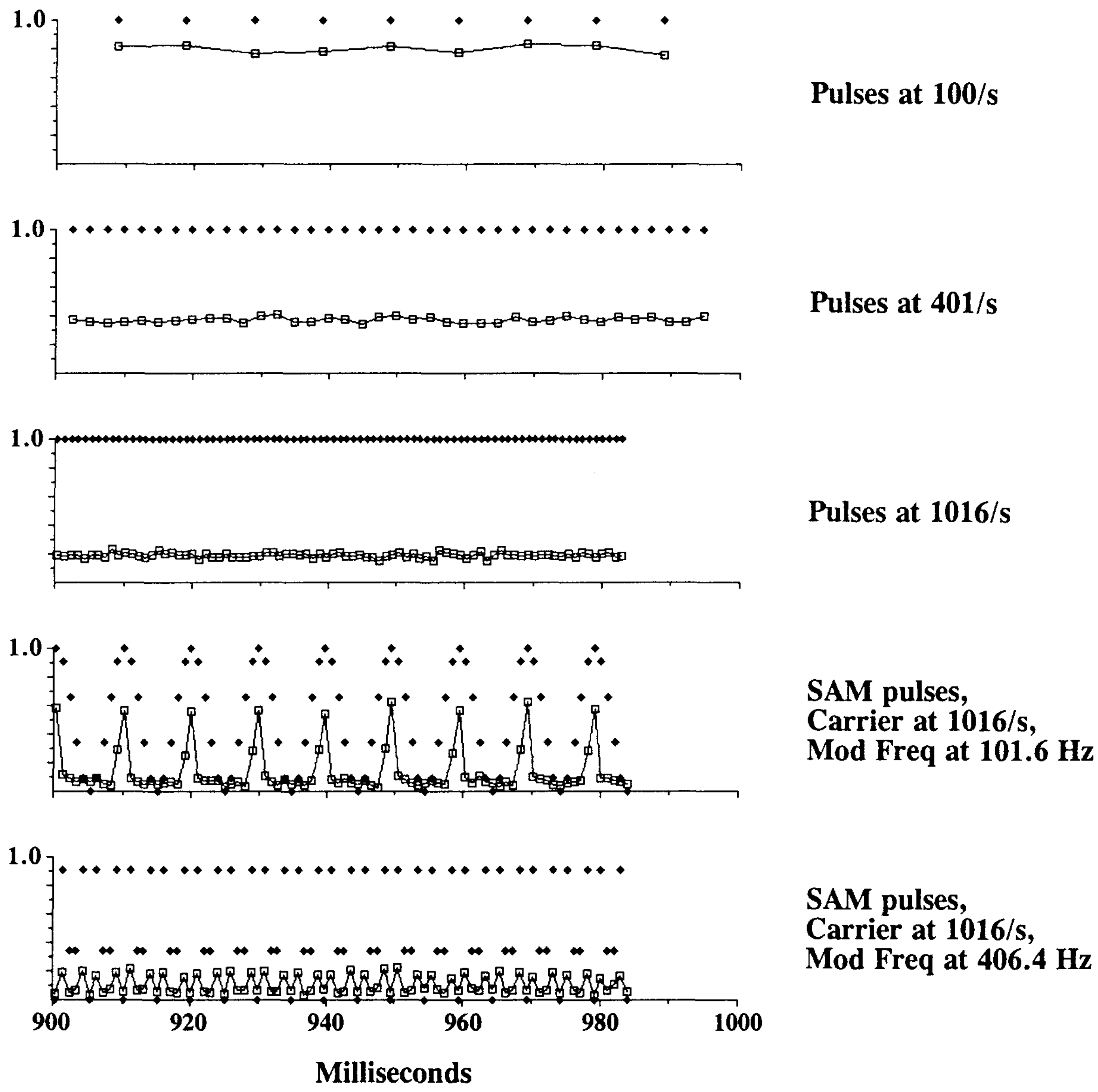


Figure 10. Final 100 ms of the records shown in Figure 8. Open squares show normalized EP magnitudes and filled diamonds show normalized pulse amplitudes.

and little or no response by pulses three and four (bottom panel of Figure 9). Pulse two is the first non-zero pulse, with a normalized amplitude of 0.905. Pulses three and four have normalized amplitudes of 0.345. This pattern of pulse amplitudes -- zero, 0.905, 0.345 and 0.345 -- is repeated across cycles for 406.4 Hz modulation of a 1016 pps carrier.

Peak magnitudes of EPs decline over time with 406.4 Hz modulation. The response to pulse two is greater than the response to all subsequent pulses with a normalized amplitude of 0.905. Also, the responses to those subsequent pulses exhibit a slow decrease in magnitude thereafter. The time course of this slow component appears to be similar to that of responses to identical pulses presented at 401/s, quite near the modulation frequency of the present condition. However, the magnitude of the decrement appears to be somewhat less with SAM pulse trains than with identical pulses presented at the modulation frequency (compare second and fifth panels in Figure 8).

The overall response with 406.4 Hz modulation is somewhat lower than the overall response with 101.6 Hz modulation. This may reflect the difference in the peak amplitude of stimulus pulses for the two conditions. For the 406.4 Hz condition that amplitude was 0.905, whereas for the 101.6 Hz condition it was 1.0.

In general, these responses to SAM pulse trains reflect features of the modulation waveform. Unlike responses to identical pulses at the carrier rate, the responses to SAM pulse trains are sustained at a relatively constant level over one second of stimulation with the 101.6 Hz modulation frequency. At the higher modulation frequency a slow decrement is observed in the response over time, similar to but not quite as great as the decrement observed for identical pulses presented at the modulation frequency.

Results Across Subjects

Results for SR3, using 1000 ms pulse trains and SAM pulse trains as above, were quite similar to those just described for SR2 (see QPR 11 for details). Recordings of responses to pulse trains have been obtained in studies with five additional subjects (SR10, SR14, NP1, NP2 and NP4). Burst durations always included 200 ms and, for some of the subjects, shorter durations. Pulse rates typically have included 100 to 1000 pps, in steps of 100 pps. Stimuli have been presented at levels corresponding to a "most comfortable loudness" judgment for each subject and electrode. Recordings of responses to SAM pulse trains have been obtained in studies with one additional subject (SR10). The burst duration for these latter studies was 200 ms. Carrier rates for the SAM pulse trains have included 1000, 500 and 250 pps. Modulation frequencies have included 50, 100, 150 and 200 Hz for the two lower carrier rates and those frequencies along with 300 and 400 Hz for the 1000 pps carrier rate.

In general and as described above for subject SR2, responses to trains of pulses show approximately equal magnitudes of EPs across pulses for low rates (e.g., 100 pps). At higher rates an alternating pattern of response is observed, with a large EP in response to the first pulse, a diminished EP to the second pulse, a partially recovered response to the third pulse, and so on. Also, a gradual reduction in the average response across sequential pulses is observed over the duration of 200 ms bursts for rates at and above 200 pps. The alternating response first appears at different rates for different subjects and for different electrodes within a subject. The rates were between 200 and 600 pps, for the tested subjects and electrodes. It may be, as suggested in QPR 7, that the "pitch saturation limit" observed with implant patients is related to the inability of the auditory nerve to reflect fully the stimulus waveform for rates above 200-600 pps. In addition, differences among subjects and electrodes in the rate at which the alternating response first appears may reflect differences in the functional status of the population of auditory neurons excited by each electrode.

Responses to SAM pulse trains demonstrate various distortions and ambiguities in the neural representation of the modulation waveform, depending on the carrier rate and modulation frequency. At low carrier rates, the patterns and magnitudes of EPs reflect features of both the modulation and carrier waveforms. At high carrier rates, EPs reflect primarily features of the modulation waveform. The inability of the nerve to follow constant amplitude pulses at relatively high rates apparently does not impair its ability to follow relatively low modulation frequencies in SAM pulse trains with high carrier rates. In general, higher carrier rates allow the faithful representation of higher modulation frequencies. Distortions and ambiguities appear when the modulation frequency is greater than 20 to 30 percent of the carrier rate. These results support the use of relatively high carrier rates in speech processors that use modulated pulse trains as stimuli.

Responses to the Pulsatile Outputs of Single-Channel Speech Processor

To examine the neural representation of more complex stimuli with implants, we also have recorded evoked potentials in response to the pulsatile outputs of a single-channel speech processor. The processor was a single-channel variation of CIS processors, which we call a *continuous sampling* (CS) processor, since interleaving is not relevant when there is only one channel. The CS processor uses the same front end as CIS processors, with a pre-emphasis filter (attenuation of 6 dB/octave below 1.2 kHz) and the same envelope detector and mapping function as in each CIS channel. The bank of bandpass filters is omitted in the CS processor, so the input to the envelope detector is the broad band speech signal, as modified somewhat by the pre-emphasis filter.

An example of the presented stimuli and recorded responses for one of the tokens in our consonant test, /asa/, is presented in Figure 11. The stimuli were delivered to electrode 3 in the Ineraid implant of subject SR2, with reference to a remote electrode in the temporalis muscle. Recordings were made differentially between the adjacent electrode 4 and an external electrode at the ipsilateral mastoid. Stimulus pulses were 32.8 μs/phase in duration, and were presented at the rate of 824/s. Normalized amplitudes of the pulses are shown in the top panel of Figure 11 and normalized magnitudes of the EPs following each pulse are shown in the bottom panel. The initial /a/ occurs during the first 140 ms of the records, the /s/ in the interval from about 190 to 330 ms, and the final /a/ in the interval from about 350 to 640 ms.

In broad terms, the neural response reflects the relatively deep modulations of pulse amplitudes during the /a/ segments and the relatively small differences among pulse amplitudes during the /s/ segment. Peak magnitudes of the response are greatest during the /a/ segments, where the pulse amplitudes reach peak levels.

The pattern of response to the temporal fine structure of stimulus pulses during the initial 100 ms of the

Stimuli and Responses for a Processed Speech Token (/asa/)

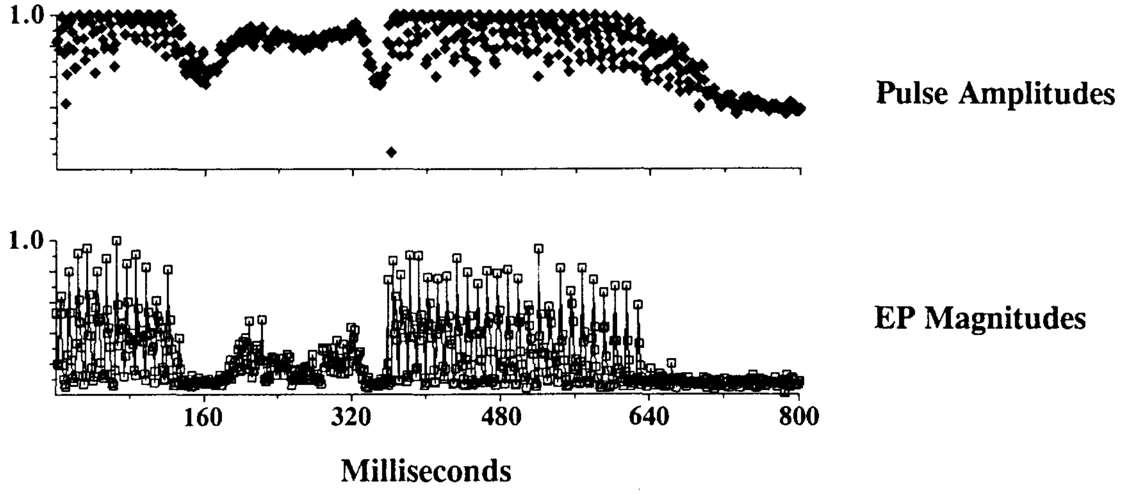


Figure 11. Normalized pulse amplitudes (top panel) and normalized EP magnitudes (bottom panel) for a processed speech token. A continuous sampling (CS) processor was used to process the speech token /asa/. Neural responses to the output of the processor were recorded for Ineraid subject SR2, as described in the text.

records is shown in Figure 12. Here, normalized pulse amplitudes and normalized EP magnitudes are plotted in the same panel to facilitate comparisons.

Although the pattern of responses reflect the fundamental frequency of the vowel, with peaks in the response at the first intense stimulus pulse in each (approximately 10 ms) period, other features in the stimulus are not represented. In periods two through eight, for instance, a series of three or more pulses with identical or nearly identical amplitudes is presented at the beginnings of the periods. The neural response to the first pulse in each of these periods is large, as noted before. However, the response to the second pulse is much smaller in all cases. Responses to subsequent pulses show an alternating pattern, much like the one observed before for identical pulses presented at the rate of 1016/s for this subject (middle panel of Figure 9). Thus, the pattern of response to these subsequent pulses in each period appear to reflect primarily properties of the auditory nerve, as opposed to the pattern of stimulation (and intended pattern of response).

The overall level of response during the /s/ segment appears to be depressed in relation to the pulse amplitudes. Figure 13 shows, however, that EP magnitudes for pulses of the same amplitudes are quite similar for the /a/ and /s/ segments. If any fatigue or accommodation occurs over the course of the /s/, it must be quite small.

Stimuli and Responses for a Processed Speech Token (/a/)

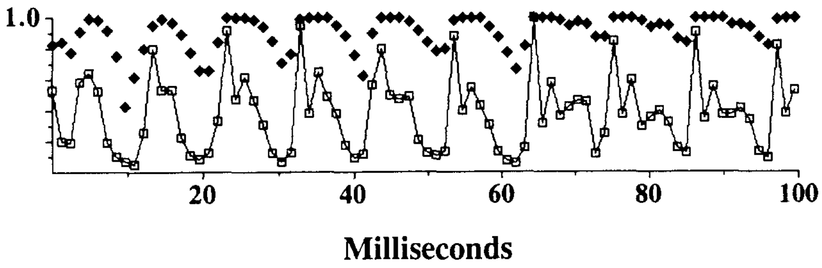


Figure 12. First 100 ms of the records shown in Figure 11. Open squares show normalized EP magnitudes and filled triangles show normalized pulse amplitudes.

Discussion

Patterns of response to processed speech stimuli appear much as would be predicted from the patterns of response to simpler stimuli. For presentation of sequential pulses with identical or nearly identical amplitudes, one could predict that, for pulses rates approximating 1000/s, a large EP would be elicited by the first pulse in the series, a much smaller EP by the second pulse, a partial recovery of the EP for the third pulse, and so on. Such patterns are observed during the vocalic segments of the example speech stimulus.

Also, responses to SAM pulse trains with high modulation depths (80 or 100 percent; see QPR 11) show a "peaking" of the neural response to a particular phase of the modulation waveform. Details other than the timing of that phase are represented poorly if at all in the response. In the example shown for the processed speech token, modulation depths approximate 40 percent or greater during the vocalic segments. As might be predicted from the results from studies with SAM pulse trains, the patterns of response to the processed speech stimulus show strong peaking in the responses, with peaks occurring at the fundamental frequency of the vocalic segments. Other details in the stimulus are not represented.

The fidelity of neural following to variations in pulse amplitudes might be improved through a change in the mapping function, e.g., a more compressive mapping function as suggested in QPR 11. Such a compression would reduce depths of modulation which can (perhaps somewhat surprisingly) improve the representation of the modulation waveform (see patterns of responses for a 20 percent depth of modulation, as presented in Figures 9, 10, 12 and 13 of QPR 11). Also, successful application of any of the repair strategies outlined in QPR 9 for this project might be helpful. An observation for now is that temporal representations with implants are crude and highly limited, even for the best patients. Performance of these devices might be improved substantially with an amelioration or repair of such defects.

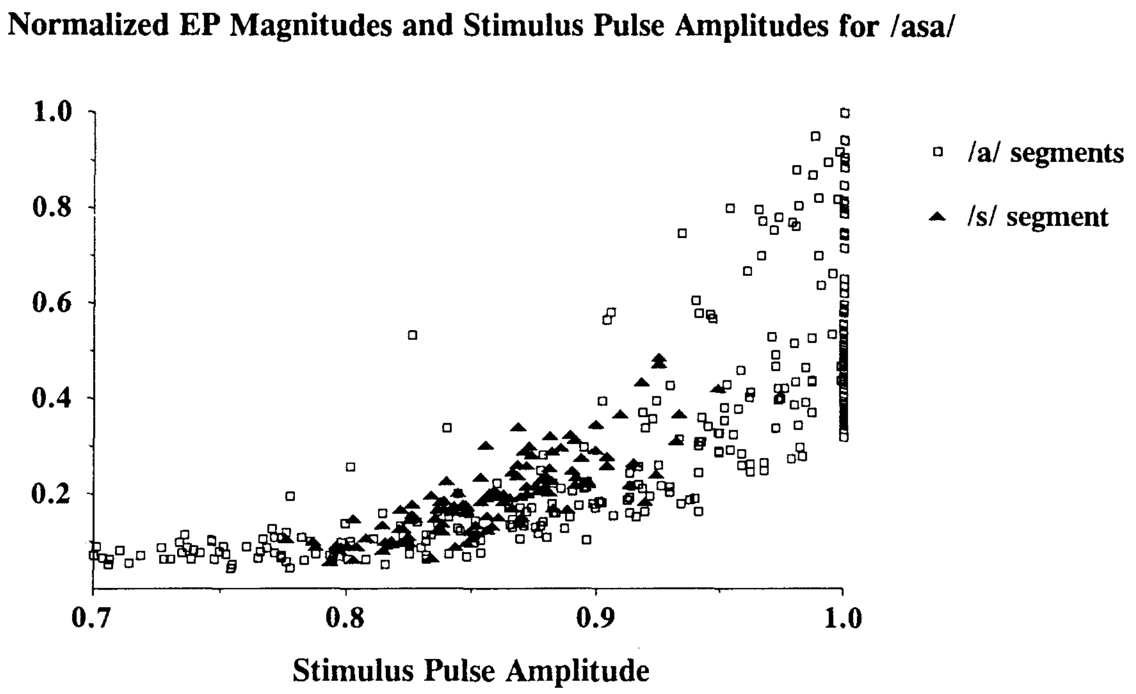


Figure 13. Scatter plot of normalized EP magnitudes versus normalized stimulus magnitudes for the speech token /asa/. Open squares show EP and stimulus magnitudes during the /a/ segments, and filled triangles show those magnitudes during the /s/ segment. The EP magnitudes are derived from recordings of neural responses in studies with Ineraid subject SR2 and correspond to the data also presented in Figures 11 and 12.

X. Record of Reporting Activity for NIH Project N01-DC-2-2401

We have maintained a high level of reporting activity throughout the current project. This activity, for the period of August 1, 1992 through July 31, 1995, includes 11 Quarterly Progress Reports, this Final Report, 7 papers, and 26 invited lectures. In addition, Wilson served as the Chair or Co-Chair for sessions at three international conferences. In the great majority of cases all expenses were reimbursed for the invited lectures, allowing us to present project results at no cost to the project.

Papers

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

Wilson BS, Lawson DT, Finley CC, Wolford RD (1993) Importance of patient and processor variables in determining outcomes with cochlear implants. J. Speech Hear. Res. 36: 373-379.

Lawson DT, Wilson BS, Finley CC (1993) New processing strategies for multichannel cochlear prostheses. Prog. Brain Res. 97: 313-321.

Wilson BS, Finley CC, Lawson DT, Wolford RD, Zerbi M (1993) Design and evaluation of a continuous interleaved sampling (CIS) processing strategy for multichannel cochlear implants. J. Rehab. Res. Devel. 30: 110-116.

Wilson BS, Lawson DT, Zerbi M, Finley CC (1994) Recent developments with the CIS strategies. In IJ Hochmair-Desoyer and ES Hochmair (Eds.), *Advances in Cochlear Implants*, Manz, Vienna, pp. 103-112.

Wilson BS, Lawson DT, Zerbi M, Finley CC, Wolford RD (1995, in press) New processing strategies in cochlear implantation. Am. J. Otol., vol. 16 (September issue).

Wilson BS, Lawson DT, Zerbi M (1995, in press) Advances in coding strategies for cochlear implants. Adv. Otolaryngol. Head Neck Surg., vol. 9 (publication scheduled for August).

Chaired Sessions

Wilson BS (1992) Chair, session on Audiological Assessment and Device Programming. *First European Symposium on Paediatric Cochlear Implantation*, Nottingham, England, Sept. 24-27.

Wilson BS, Dillier N (1993) Co-Chairs, session on Speech Coding. *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7.

Wilson BS (Chair), Cazals Y, Dillier N, Mac Leod P, McDermott H, Pelizzone M (Panelists) (1995) Round Table on Sound Signal Processing. *IIIrd International Congress on Cochlear Implant,* Paris, France, April 27-29.

Invited Presentations

Wilson BS (1992) Speech processing for auditory prostheses. Lecture for the course on "Current Status of Multichannel Cochlear Implants," *96th Meeting of the American Academy of Otolaryngology -- Head & Neck Surgery*, Washington, D.C., Sept. 13.

Wilson BS (1992) Processing strategies for multichannel cochlear implants. *First European Symposium on Paediatric Cochlear Implantation*, Nottingham, England, Sept. 24-27.

Wilson BS (1992) Panelist, round table on Programming. *First European Symposium on Paediatric Cochlear Implantation*, Nottingham, England, Sept. 24-27.

Wilson BS (1992) Speech processors for auditory prostheses. *Neural Prosthesis Workshop*, Bethesda, MD, Oct. 13-15.

Wilson BS (1993) Representations of envelope information with CIS and VCIS processors. *Mini Symposium on Envelope Representations with Cochlear Implants*, House Ear Institute, Los Angeles, CA, Feb. 25-28.

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

Wilson BS (1993) Recent developments with the CIS strategies. *Third International Cochlear Implant Conference*, Innsbruck, Austria, April 4-7.

Wilson BS, Lawson DT, Zerbi M, Finley CC (1993) CIS and "virtual channel" CIS (VCIS) processors. *1993 Conference on Implantable Auditory Prostheses*, Smithfield, RI, July 11-15.

Lawson DT (1993) Representation of complex tones by CIS processors. *1993 Conference on Implantable Auditory Prostheses*, Smithfield, RI, July 11-15.

Wilson BS, Lawson DT, Zerbi M, Finley CC, Wolford RD (1993) New processing strategies in cochlear implantation. Lecture for the special session on "Basic Science Update," *American Neurotology Society*, Minneapolis, MN, Oct. 1.

Wilson BS (1993) Speech processors for auditory prostheses. *Neural Prosthesis Workshop*, Bethesda, MD, Oct. 13-15.

Wilson BS (1993) Introduction to speech processor design and testing. *1993 Zhengzhou International Symposium on Electrical Cochlear Hearing and Linguistics*, Zhengzhou, China, Oct. 23-26.

Wilson BS (1993) New processing strategies for cochlear prostheses. *1993 Zhengzhou International Symposium on Electrical Cochlear Hearing and Linguistics*, Zhengzhou, China, Oct. 23-26.

Wilson BS (1993) Further studies with CIS and related processors. *1993 Zhengzhou International Symposium on Electrical Cochlear Hearing and Linguistics*, Zhengzhou, China, Oct. 23-26.

Lawson DT (1993) Cochlear implants: Current research. *5th Annual Audiology Videoconference*, Mayo and Sheldon Reese Foundations, Jacksonville, FL, Nov. 6.

Wilson BS (1994) Review of speech processor studies. Guest Lecture, University of Iowa, Department of Otolaryngology -- Head & Neck Surgery, Iowa City, IA, Jan. 18.

Wilson BS (1994) Progress in speech processor design. *Fifth Symposium on Cochlear Implants in Children*, New York, NY, Feb. 4.

Wilson BS (1994) Review of speech processor studies. Guest Lecture, Indiana University School of Medicine, Department of Otolaryngology -- Head & Neck Surgery, Indianapolis, IN, March 9.

Lawson DT (1994) Current research on cochlear implants. *Annual Carolinas Audiology Conference*, Asheville, NC, April 1.

Wilson BS (1994) Progress in the development of speech processors for cochlear prostheses. Lecture for the special session on "Electro-Auditory Prostheses," *127th Meeting of the Acoustical Society of America*, Cambridge, MA, June 8. [Abstract published in J. Acoust. Soc. Am. 95: 2905.]

Wilson BS (1994) Cochlear modeling studies. *Neural Prosthesis Workshop*, Bethesda, MD, Oct. 18-21.

Wilson BS (1994) Speech processors for auditory prostheses. *Neural Prosthesis Workshop*, Bethesda, MD, Oct. 18-21.

Wilson BS, Lawson DT, Zerbi M, Finley CC (1995) New developments in speech processors. Lecture for the "Workshop on Auditory Prosthetics," *Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg, FL, Feb. 5-9.

Lawson DT (1995) Design and performance of speech processors for cochlear prostheses. *Otolaryngology Grand Rounds,* Duke University Medical Center, Feb. 15.

Wilson BS (1995) Future directions in speech processing. *CIS Workshop* (sponsored by Med El GmbH and held in conjunction with the *IIIrd International Congress on Cochlear Implant*), Paris, France, April 26.

Wilson BS (1995) Continuous Interleaved Sampling and related strategies. *NIH Consensus Development Conference on Cochlear Implants in Adults and Children*, May 15-17.

Additional Presentations

Zerbi M, Wilson BS, Finley CC, Lawson DT (1992) A flexible speech processor for cochlear implant research. *1992 Digital Signal Processing Workshop*, Utica, IL, Sept. 13-16.

Loeb GE, Shannon RV, Wilson BS, Zeng F-G, Rebscher S (1994) Design for an inexpensive but effective cochlear implant system: Recommendations of an expert panel from the *1993 Zhengzhou International Symposium on Electrical Cochlear Hearing and Linguistics*. *International Cochlear Implant, Speech and Hearing Symposium*, Melbourne, Australia, Oct. 23-28 (The order of authors for this paper was determined by a random number generator).

Finley CC, Wilson BS (1995) Responses of the auditory nerve to repetitive electrical stimuli as demonstrated with recordings of intracochlear evoked potentials. Poster presentation, *Midwinter Meeting of the Association for Research in Otolaryngology*, St. Petersburg, FL, Feb. 5-9.

Wilson BS, Finley CC (1995) Temporal representations with cochlear implants. *IIIrd International Congress on Cochlear Implant,* Paris, France, April 27-29.

XI. Suggestions for Future Research

Large gains in open-set speech recognition have been produced for implant patients with the advent of new processing strategies (NIH Consensus Statement, 1995). Although the present results are most encouraging, many important questions remain unanswered, and many possibilities for further improvement remain unexplored. These questions and possibilities are being addressed in our laboratory, under the auspices of a new project with the Neural Prosthesis Program, and elsewhere. Particularly promising lines of investigation include the following:

- Continued evaluation of parameter choices for CIS processors in studies with additional subjects, including detailed evaluation of tradeoffs among pulse duration, pulse rate, interval between sequential pulses, and cutoff frequency of the lowpass filters in the envelope detectors.
- Development of procedures to make the demonstrated potential benefit of parameter optimization available to implanted patients in clinical settings.
- Identification of the factor or factors underlying patient variability.
- Continued evaluation of possible learning effects with extended use of CIS processors.
- Continued evaluation of "virtual channel" and "sharpened field" CIS processors, in studies with additional subjects and tests.
- Completion of studies in progress with subjects using a percutaneous-connector version of the Nucleus device, to evaluate various implementations of the CIS and SPEAK processing strategies in conjunction with the Nucleus electrode array.
- Continued development of new processing strategies, which may produce further increases in performance.
- Continued development of highly-flexible portable processors to support field studies with CIS and other strategies.
- Development and evaluation of techniques to reduce the deleterious effects of environmental noise on processor performance.
- Design and evaluation of strategies to exploit possible improvements in sound localization and speech reception in noise presented by bilateral implants.
- Continued studies with the Auditory Brainstem Implant, especially if patients implanted with any of the new multichannel penetrating electrodes become available for such studies.
- Continued recordings of intracochlear evoked potentials in implant subjects, to measure temporal patterns of neural population responses for a wide variety of stimuli.
- Continued development and evaluation of strategies to ameliorate or repair demonstrated deficits in temporal representations with cochlear implants.
- Continued studies to measure perception of complex stimuli other than speech sounds, such as complex tones.
- Development of new test materials, to provide sensitive measures of performance for subjects who now, with certain speech processors and implant devices, encounter ceiling effects with all of our standard tests.

XII. 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, Paul Abbas, Carolyn Brown, Michael Dorman, Donald Eddington, James Patrick, Bryan Pfingst, William Rabinowitz, Robert Shannon, Richard Tyler and Robert Wolford.

XIII. References

Dorman MF (1993) Speech recognition and frequency discrimination with the Ineraid and CIS cochlear prostheses. J. Acoust. Soc. Am. 94: 1888 (abstract).

Dorman MF, Dankowski K, McCandless G (1990) Longitudinal changes in word recognition by patients who use the Ineraid cochlear implant. Ear Hear. 11: 445-459.

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

Gfeller K, Lansing CR (1991) Melodic, rhythmic, and timbral perception of adult cochlear implant users. J. Speech Hear. Res. 34: 916-920.

Hochmair-Desoyer IJ, Hochmair ES, Burian K, Stiglbrunner HK (1983) Percepts from the Vienna cochlear prosthesis. Ann. N.Y. Acad. Sci. 405: 295-306.

Jolly CN, Spelman FA, Pfingst BE (1994) Focusing potentials in the cochlea: Modeling and experimental data. In: *Abstracts of the Seventeenth Midwinter Research Meeting, Association for Research in Otolaryngology*, St. Petersburg Beach, FL, February 6-10, p. 162.

Lawson DT, Wilson BS, Finley CC (1993) New processing strategies for multichannel cochlear prostheses. Prog. Brain Res. 97: 313-321.

Lim HH, Tong YC, Clark GM (1989) Forward masking patterns produced by intracochlear electrical stimulation of one and two electrode pairs in the human cochlea. J. Acoust. Soc. Am. 86: 971-980.

Mathews MV, Pierce JR (1989) The Bohlen-Pierce Scale. In Mathews MV, Pierce JR (Eds.), *Current Directions in Computer Music Research*. MIT Press, Cambridge, MA, pp. 165-173.

McDermott HJ, McKay CM (1994) Pitch ranking with nonsimultaneous dual-electrode electrical stimulation of the cochlea. J. Acoust. Soc. Am. 96: 155-162.

McKay CM, McDermott HJ, Clark GM (1994) Pitch percepts associated with amplitude-modulated current pulse trains in cochlear implantees. J. Acoust. Soc. Am. 96: 2664-2673.

NIH Consensus Statement (1995) Cochlear implants in adults and children. Office of Medical Applications of Research, National Institutes of Health, Bethesda, MD.

Rossing TD (1990) *The Science of Sound, 2nd Edition*. Addison-Wesley.

Shannon RV (1993) Psychophysics. In Tyler RS (Ed.), *Cochlear Implants: Audiological Foundations*. Singular Publishing Group, San Diego, CA, pp. 357-388.

Skinner MW, Clark GM, Whitford LA, Seligman PM, Staller SJ, Shipp DB, Shallop JK, Everingham C, Menapace CM, Arndt PL, Antogenelli T, Brimacombe JA, Pijl S, Daniels P, George CR, McDermott HJ, Beiter AL (1994) Evaluation of a new spectral peak (SPEAK) coding strategy for the Nucleus 22 channel cochlear implant system. Am. J. Otol. 15, Suppl. 2: 15-27.

Townshend B, Cotter N, Van Campernolle D, White RL (1987) Pitch perception by cochlear implant patients. J. Acoust. Soc. Am. 87: 106-115.

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

Wilson BS, Finley CC, Lawson DT, Wolford RD, Eddington DK, Rabinowitz WM (1991) Better speech recognition with cochlear implants. Nature 352: 236-238.

Wilson BS, Lawson DT, Finley CC, Wolford RD (1993) Importance of patient and processor variables in determining outcomes with cochlear implants. J. Speech Hear. Res. 36: 373-379.

Wilson BS, Lawson DT, Zerbi M, Finley CC (1994) Recent developments with the CIS strategies. In IJ Hochmair-Desoyer and ES Hochmair (Eds.), *Advances in Cochlear Implants*. Manz, Vienna, pp. 103-112.

Appendix

Reprints of Published Articles for NIH Project N01-DC-2-2401

Return to Bib file

USJH