First Quarterly Progress Report

August 1 through October 31, 1992

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

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

CONTENTS

I. Introduction

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

1. Design and preliminary evaluation of a new *virtual channel interleaved sampling* (VCIS) processing strategy for multichannel cochlear implants. Studies were conducted with Ineraid subject SR2.
2. Design and construction of a new output board for our laboratory processor system, to provide the simultaneous stimuli needed by VCIS processors.
3. Coding of a new monitor system to support specification and subsequent implementation of VCIS processors.
4. Initial studies with Ineraid subject SR12, to evaluate a variety of processing strategies with a new subject.
5. Continued analysis of data from prior and current studies, to evaluate effects of single parameter changes on the performance of *continuous interleaved sampling* (CIS) processors.
6. Presentation of project results at the *96th Meeting of the American Academy of Otolaryngology -- Head & Neck Surgery* (Washington, D.C., Sept. 13), the *1992 Digital Signal Processing Workshop* (Utica, IL, Sept. 13-16), the *First European Symposium on Paediatric Cochlear Implantation* (Nottingham, England, Sept. 24-27), and the *Neural Prosthesis Workshop* (Bethesda, MD, Oct. 13-15).
7. Continued preparation of manuscripts for publication.

In this report we present results from the preliminary evaluation of VCIS processors (point 1 above). Work related to points 4 and 5 will be described in future reports.

II. Virtual Channel Interleaved Sampling (VCIS) Processors

Recent studies in our laboratory have focused on development of *continuous interleaved sampling* (CIS) processors (e.g., see Wilson et al., 1991). A block diagram and waveforms for CIS processors are presented for reference in Figs. 1 and 2, respectively. Each channel includes a bandpass filter and an envelope detector. The amplitudes of stimulus pulses are determined with a logarithmic or power-law transformation of the envelope signal. The corner (or cutoff) frequencies of the bandpass filters span the frequency range from 350 to 5500 Hz, evenly spaced along a logarithmic scale. In typical implementations five or six channels are used, with 400 Hz lowpass smoothing filters in the envelope detectors. The information represented by the implant thus consists of envelope variations, below 400 Hz, in each of five or six frequency bands of speech.

Design of VCIS Processors

A possible refinement and extension of the CIS approach is illustrated in Fig. 3. Here adjacent electrodes may be stimulated simultaneously to shift the perceived pitch in any direction with respect to the corresponding single-electrode percepts. Studies with implant subject SR2 indicate that perceived pitch can be manipulated through various choices of simultaneous and nonsimultaneous conditions. If, for instance, the apicalmost electrode of his Ineraid electrode array (electrode 1) is stimulated alone, a low pitch is reported. If the next electrode in the array (electrode 2) is stimulated alone, a higher pitch is reported. An intermediate pitch can be produced by stimulating the two electrodes together with identical, in-phase pulses. Finally, by reversing the phase of one of the simultaneous pulses pitch percepts outside the interval represented by the single-electrode percepts can be obtained. The availability of pitches other than those elicited with stimulation of single electrodes may provide additional channels of stimulation. We call these additional channels "virtual channels," and processors that use them *virtual channel interleaved sampling* (VCIS) processors.

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

Subject and Tests

Ineraid subject SR2 has participated in an extensive series of studies to evaluate effects of parametric changes in CIS processors and more recently to evaluate several implementations of VCIS processors. Results from his first tests with CIS processors are presented in earlier reports from our group (Wilson et al., 1990; Wilson et al., 1991) and are summarized here in Table 1 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.

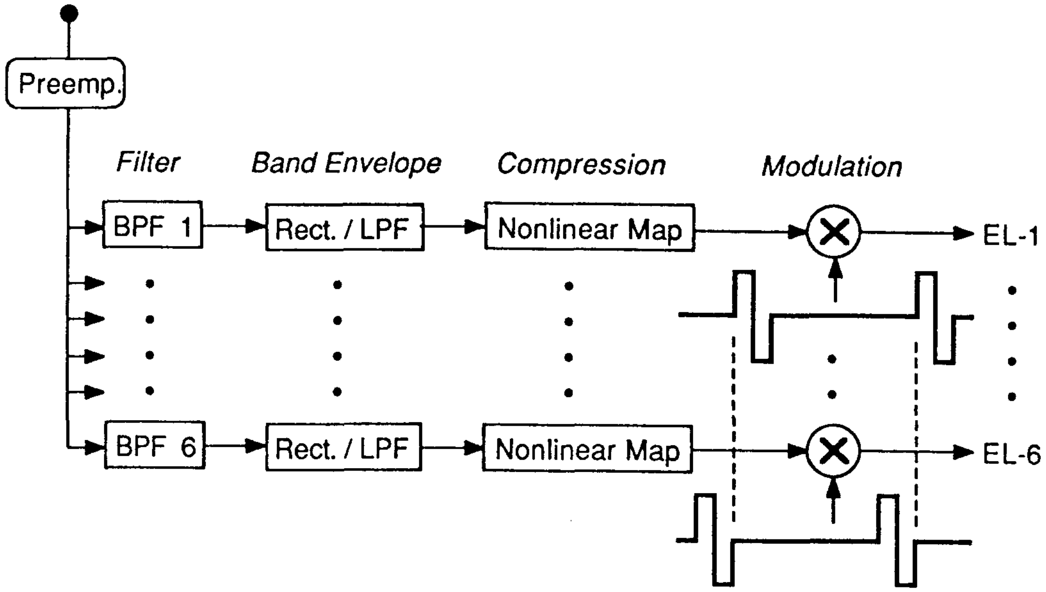

Figure 1. Block diagram of a CIS processor. A preemphasis filter (Preemp.) is used to attenuate strong low frequency components in speech that otherwise might mask important high frequency components. The preemphasis filter is followed by six channels of processing. Each channel includes stages of bandpass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

Results

These tests and others have been used to evaluate the subsequent implementations of CIS and VCIS processors. Results from a refined implementation of a CIS processor, using parameters somewhat different from those of the original implementation, are presented in the second column of numeric entries in Table 1. Results for a VCIS processor are presented in column 3. This processor used in-phase pulses to produce pitches between each pair of electrodes. Thus, six channels corresponded to conditions involving stimulation of one electrode alone, while five additional channels corresponded to conditions involving in-phase paired stimulation. As a precaution against possible learning or familiarization effects, different lists of words and sentences were used in each of the CID, SPIN and NU-6 tests for the different processors. Also, the NU-6 test was repeated for the "refined CIS" processor using another new list of words. The additional test listed in Table 1 involved identification of 24 consonants (/b, d, f, g, dʒ, h, j, k, l, m, n, ŋ, p, r, s, ʃ, t, tʃ, ð, θ, v, w, z, ʒ /) in an /a/-consonant-/a/ context, with each of the 24 played in block-randomized order 10 times for a male

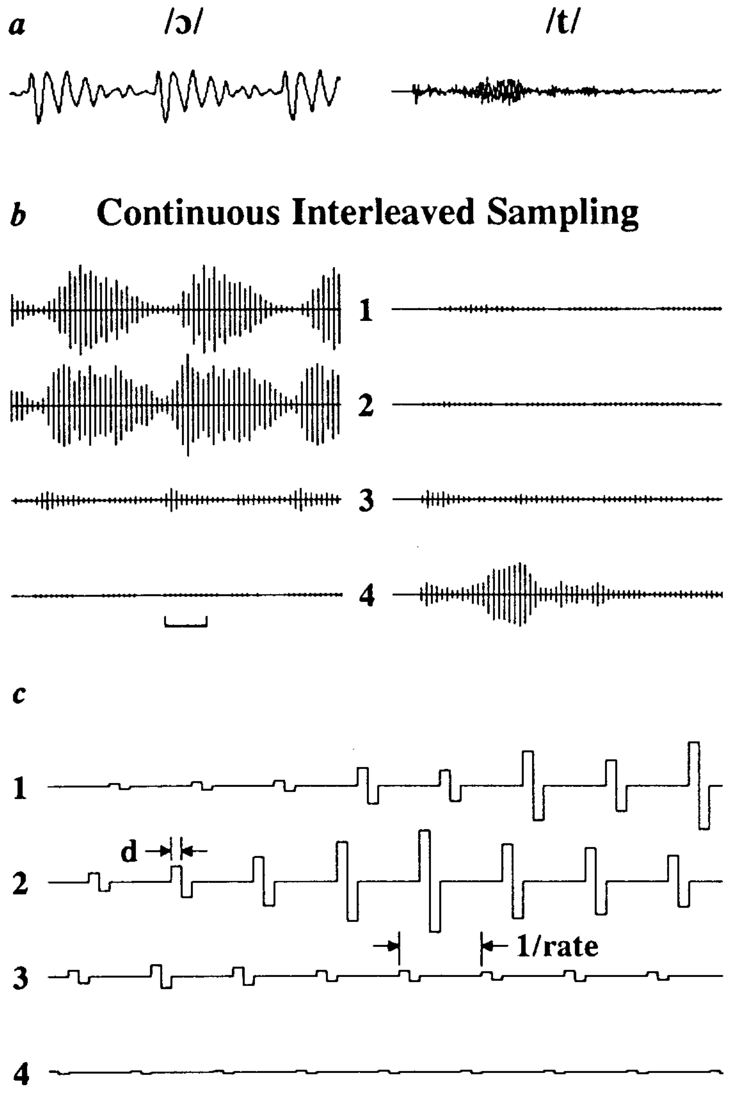

Figure 2. Waveforms of a four-channel CIS processor. *a*, Preemphasized (6 dB/octave attenuation below 1.2 kHz) speech inputs. Inputs corresponding to a voiced speech sound ('aw') and an unvoiced speech sound ('t') are shown in the left and right columns, respectively. The duration of each trace is 25.4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

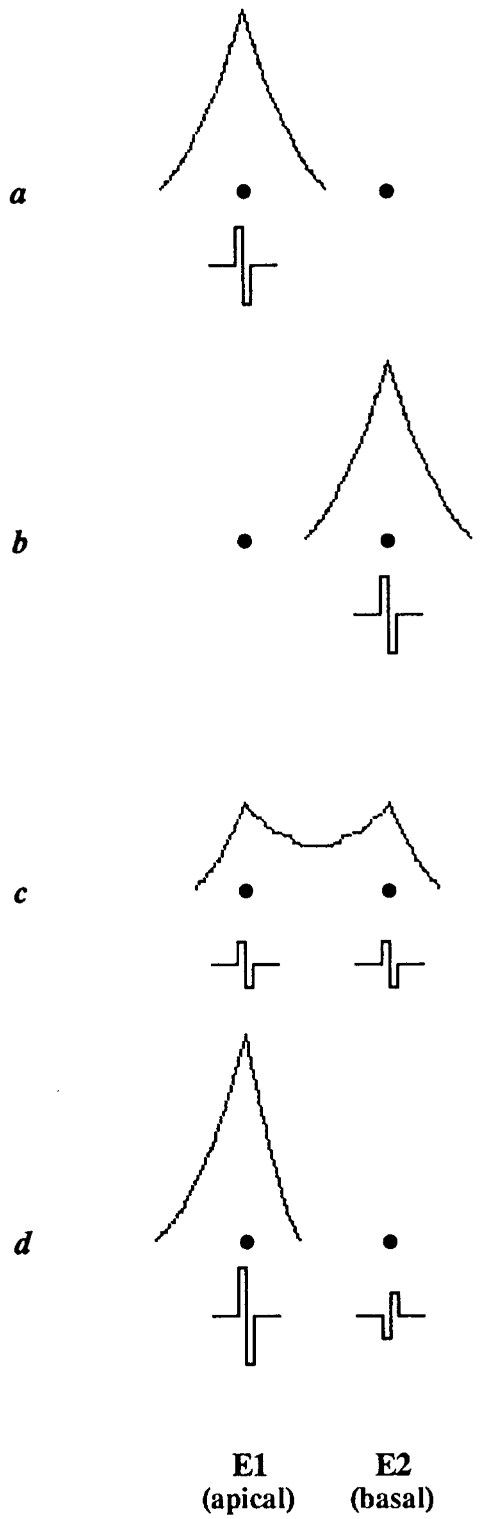

Figure 3. Schematic illustrations of neural responses for various conditions of stimulation. The top curve in each panel sketches a hypothetical profile of the number of neurons responding to the pulse or pulses at the bottom of the panel. The positions of two adjacent electrodes are indicated by the dots in each case. Profiles for stimulation of either electrode alone are presented in *a* and *b*, and profiles for paired stimulation in *c* and *d*. Implant subject SR2, listening to these different stimuli, can rank them according to their distinct pitches. Otherwise the percepts are similar, in that percepts arising from paired-pulse conditions and those arising from single-pulse conditions seem to differ only in pitch. For the tested subject (using the Ineraid electrode array), stimulation of electrode 1 alone (*a*) produced a low pitch, while stimulation of electrode 2 alone (*b*) produced a higher pitch. Simultaneous stimulation of both electrodes, with identical pulses having approximately half the amplitude of the single-pulse conditions (*c*), produced an intermediate pitch. Pairing stimulation of electrode 1 with a reversed-polarity pulse on electrode 2 (*d*) produced the lowest pitch among the illustrated conditions. Various lower pitches could be produced by manipulating the ratio of the pulse 2 to pulse 1 amplitudes over the range of 0.2 to 0.8. A ratio of 1.0 produced a pitch higher than that elicited by stimulation of electrode 1 alone. Similarly, pitches higher than that elicited by stimulation of electrode 6 alone (the basalmost electrode) could be produced by presenting a reversed-phase pulse (of lower amplitude) on electrode 5. Additional discriminable pitches between electrodes also might be produced by manipulating the amplitude ratio of in-phase pulses, but this possibility was not tested. The pulse width used for the listening tests was 33 μs/phase.

TABLE 1. 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.

speaker and 10 times for a female speaker. Multiple exemplars of each token were drawn from laser videodisc recordings under computer control (Lawson et al., in press). As with the other tests, the medial consonant tokens were presented in a sound-alone condition, with no feedback as to correct or incorrect responses.

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

Discussion

These results indicate quite high levels of speech recognition. The NU-6 score of 80% correct falls in the middle of the range of scores obtained by listeners with mild-to-moderate hearing losses (Bess and Townsend, 1977; Dubno and Dirks, 1982); the score of 92% correct falls in the range of scores obtained by listeners with mild hearing losses (Dubno et al., 1984); and it is not unusual for a subject with normal hearing to score 98% correct (e.g., Davis and Silverman, 1978; Frank and Craig, 1984).

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

While it seems unlikely that many patients will be able to enjoy the high levels of performance obtained by SR2 (e.g., see Wilson et al., in press), the present results demonstrate what is possible with electrical stimulation of the auditory system, using only six monopolar electrodes. In fact, the scores reported here are highly consistent with scores reported in the early literature on analysis/synthesis systems (sometimes called "vocoder" systems) for the reduced bandwidth transmission of speech. In those experiments speech test scores for listeners with normal hearing began to asymptote at high levels when 6 to 10 bandpass channels were used in an acoustic representation of envelope signals (Flanagan, 1972). The present results suggest that most or all of the information contained in such envelope signals can be transmitted across the electrode-nerve interface of a cochlear implant, and that quite high levels of speech reception can be supported with 6 or 11 channels.

Although the indications from preliminary implementations of VCIS processors are encouraging, much work remains to evaluate fully the potential of the VCIS approach. For example, only one combination of virtual channels was involved in the processor described here. As suggested in Fig. 3, this combination, using identical in-phase pulses on adjacent electrodes, might produce substantial overlaps in neural excitation fields across single-electrode and two-electrode channels (compare the hypothetical fields sketched for conditions *a*, *b* and *c* in Fig. 3). Other combinations may reduce such overlaps. Also, conditions utilizing reversed-phase pulses (condition *d* in Fig. 3) may prove useful in VCIS processors, as may simultaneous pulsatile stimulation of more than two electrodes.

Acknowledgements

We thank subject SR2 for his generous contribution of time. We also note that modeling studies, conducted under the auspices of Project IV in NIH Program Project Grant P01-DC00036, have played an important role in the formulation of VCIS processors. We expect that continued modeling studies may offer further insights into the fine control of the shapes and extents of neural excitation fields in the electrically stimulated cochlea.

References

Bess FH, Townsend TH (1977). Word discrimination for listeners with flat sensorineural hearing losses. *J. Speech Hear. Disorders* 42: 232-237.

Dorman MF, Dankowski K, McCandless G (1990). Longitudinal changes in word recognition by patients who use the Ineraid cochlear implant. *Ear Hear.* 11: 455-459.

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.

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

Dubno JR, Dirks DD, Morgan DE (1984). Effects of age and mild hearing loss on speech recognition in noise. *J. Acoust. Soc. Am.* 76: 87-96.

Davis H, Silverman SR (1978). *Hearing and Deafness*. Holt, Rinehart and Winston, New York.

Flanagan JL (1972). *Speech Analysis, Synthesis and Perception* (2nd Edition). Springer-Verlag, Berlin.

Frank T, Craig CH (1984). Comparison of the Auditec and Rintelmann recordings of the NU-6. *J. Speech Hear. Disorders* 49: 267-271.

Lawson DT, Wilson BS, Finley CC (in press). New processing strategies for multichannel cochlear prostheses. *Prog. Brain Res.*

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 (1990). *Speech processors for auditory prostheses*. Fourth Quarterly Progress Report, NIH project N01-DC-9-2401.

Wilson BS, Lawson DT, Finley CC, Wolford RD (in press). Importance of patient and processor variables in determining outcomes with cochlear implants. *J. Speech Hear. Res.*

III. Plans for the Next Quarter

Our plans for the next quarter include further studies with VCIS processors. We have scheduled two additional visits for subject SR2, and expect to evaluate various VCIS processors using reduced numbers of electrodes and different types of virtual channels. The results should provide guidance on the best way or ways to construct virtual channels.

Appendix 1

Summary of Reporting Activity for the Period of

August 1 through October 31, 1992

NIH Project N01-DC-2-2401

Reporting activity for the last quarter included the following:

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

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

Wilson BS: Processing strategies for multichannel cochlear implants. Invited lecture presented at the *First European Symposium on Paediatric Cochlear Implantation*, Nottingham, England, Sept. 24-27, 1992.

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

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

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