Original text

B

Ninth Quarterly Progress Report

May 1 through July 31, 1991

NIH Contract N01-DC-9-2401

Speech Processors for Auditory Prostheses

Prepared by

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

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

CONTENTS

I. Introduction

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

1. Evaluation of the prototype for the portable processor described in QPR 7 for this project, in tests with Ineraid subject SR2.
2. Evaluation of the new MiniMed processor and transcutaneous transmission system (TTS) in tests with subject SR2 (the outputs of the receiver in the TTS were routed to the percutaneous connector of the subject's implant system).
3. A broad range of studies with subject SR2, including (a) evaluation of an automatic gain control used in conjunction with *continuous interleaved sampling* (CIS) processors, (b) investigation of various parameter changes in CIS processors, (c) preliminary evaluation of a new variant of CIS processors in which the order of pulse presentations is randomized for each succeeding sequence of stimulation across the electrode array, (d) preliminary psychophysical studies of modulation perception, and (e) measurement of thresholds, most comfortable loudnesses, and dynamic ranges for pulse trains using a variety of pulse durations and rates.
4. Design of multilayer circuit boards for reducing the size, and increasing the reliability and reproducibility, of the prototype portable processor (the present prototype is implemented using wire wrap technology).
5. Completion of software interfaces for the laboratory DSP56001 system, to allow rapid specification of parameters for speech processing strategies.
6. Development of DSP56001 software for new types of processors (e.g., hybrids of the CIS and *peak picker* strategies; new variations of CIS processors).
7. Presentation of project results in invited lectures at the *121st Meeting of the Acoustical Society of America*, held in Baltimore, MD, April 29 to May 3 (Wilson), and at the *1991 Conference on Implantable Auditory Prostheses*, held in Pacific Grove, CA, June 2-7 (Wilson and Finley).
8. Publication of a paper in **Nature**, on "Better Speech Recognition with Cochlear Implants."
9. Continued preparation of manuscripts for publication.

In this report we describe work related to points 1, 2, 3a, 3d and 3e above. In addition, a description of software interfaces for the DSP56001 system (point 5) is presented in Appendix 3. Work related to points 3b, 3c, 4 and 6 will be presented in future reports.

II. Evaluation of the Prototype for a Portable Processor

The prototype for a highly programmable and powerful portable processor has been described in QPR 7 for this project. The system is based on two DSP56001 processors, which share external random access memory (RAM) on a common bus. One DSP56001 serves as the master processor and the other DSP56001 serves as its slave. Both processors may compute simultaneously at full speed (32 MHz) within the limits of separate on-chip memories. For less demanding applications the slave processor can be powered down or even unplugged to conserve battery power.

The present prototype is constructed using wire wrap technology on a relatively large board. As a prelude to reducing the size of the prototype (using multilayer printed circuit boards), we wanted to evaluate the prototype in tests with one of our better subjects. In this way we could be sure that the performance of the prototype, in terms of speech recognition scores, matched that of our laboratory system before we made a large investment in the design of the printed circuit boards.

Tests

Tests were conducted to compare speech reception scores obtained previously with a CIS processor using our TMS320-based laboratory system with scores obtained with two similar CIS processors using the DSP56001-based prototype. The tests included identification of 16 consonants (/b, d, f, g, dʒ, k, l, m, n, p, s, ʃ, t, ð, v, z/) in an /a/-consonant-/a/ context and the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery [Owens et al., 1985].

In the consonant tests multiple exemplars of the tokens were played from laser videodisc recordings of a male speaker [Tyler et al., 1987; Lawson et al., 1989]. A single block of trials consisted of five randomized presentations of each consonant.

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

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

All tests were conducted with hearing alone, without feedback as to correct or incorrect responses. Results for the consonant identification test were expressed as percent information transfer for various articulatory and acoustic features [Miller and Nicely, 1955], and results for the remaining tests were expressed as the percentage of correct responses.

Processors

Comparisons were made among three parametrically similar CIS processors (see Appendix 2 for a brief description of the CIS strategy). The processors included MP28m3, implemented with the TMS320-based laboratory system, and MP2 and MP6A, implemented with the DSP56001-based prototype. All processors used 33 μs/phase pulses, presented at the rate of 2525 pps on each channel. In addition, all processors used 6 channels of stimulation, a fullwave rectifier and 400 Hz lowpass filter in the envelope detector for each channel, and a staggered order of channel updates (6-3-5-2-4-1). Differences among these otherwise similar processors included the orders of the lowpass and bandpass filters for each channel. Processors MP28m3 and MP6A used first-order lowpass filters, and processor MP2 used fourth-order filters. Processors MP28m3 and MP2 used sixth-order bandpass filters, and processor MP6A used twelfth-order filters. Automatic gain control was not used in any of the processors. Processor MP28m3 was evaluated in the summer of 1990 and the remaining processors in the summer of 1991.

Results

Speech test scores for the three processors are presented in Fig. 1. The scores demonstrate high levels of performance with each processor. Indeed, most scores either approach or encounter the 100% ceiling. Among the open-set tests, for instance, only the NU-6 scores drop below 96% correct. We note that novel lists of words and sentences were used for the NU-6 and CID tests with each of the processors.

While performances among these processors are quite similar, processor MP6A appears to provide somewhat better transmission of consonant place information, and somewhat higher open-set scores, than the other processors. Processors MP28m3 and MP2 appear to provide essentially identical consonant and open-set performances. More difficult tests (or many more trials for the consonant test) would be required to demonstrate significant differences between these processors, if indeed such differences exist.

From these results and the subject's anecdotal appraisals we conclude that processor implementations with the prototype are at least as good as those with the TMS320-based laboratory system. Possible advantages of the prototype implementations include a 24-bit word length instead of a 16-bit word length and a higher bandwidth of the current source(s) used for stimulation (see QPR 7 for details). For the foreseeable future we plan on using DSP56001-based systems for our laboratory studies as well as for our portable processor implementations. This will allow evaluation of a broader range of processing strategies, through the greater speed and precision of computations with the DSP56001, and will provide program compatibility with the "DSP engine" of the portable processor.

In view of the present findings of essential similarity between the prototype and TMS320-based laboratory systems, we have designed multilayer printed circuit boards for reducing the size, and increasing the reliability and reproducibility of the prototype. The boards will be used in constructing wearable processors for daily use outside of the laboratory. Such use will allow evaluation of possible learning effects with CIS and other processors.

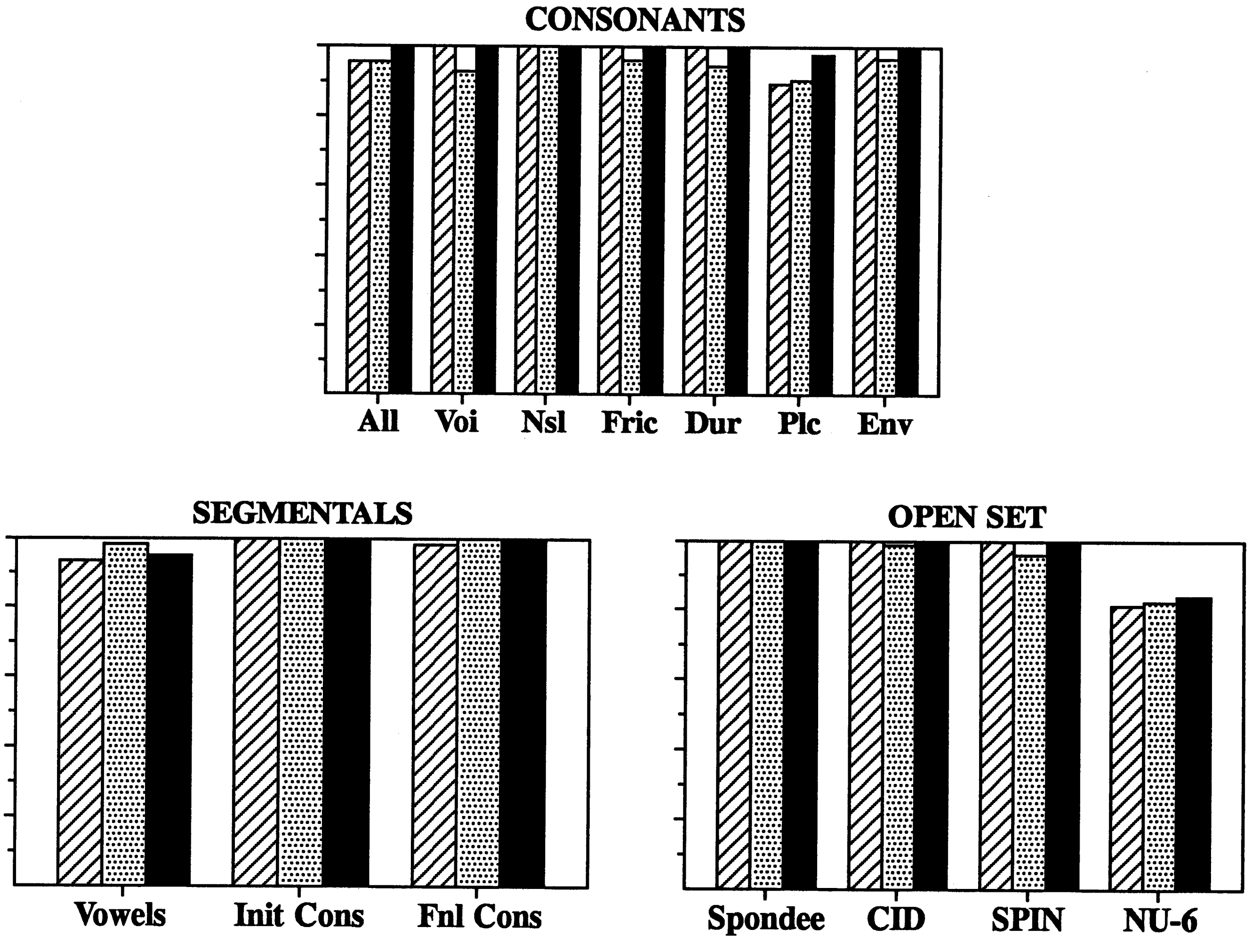

FIG. 1. Comparisons of speech test scores for CIS processors implemented with the TMS320-based laboratory system and with the prototype of the portable processor, based on the DSP56001. The striped bars show results from processor MP28m3, implemented with the TMS320-based laboratory system, and the stippled and solid bars show results from processors MP2 and MP6A, respectively, implemented with the DSP56001-based prototype. The top panel shows relative information transfer for articulatory and acoustic features of consonants. The features include overall transmission (All), voicing (Voi), nasality (Nsl), frication (Fric), duration (Dur), place of articulation (Plc), and envelope cues (Env). Full scale corresponds to 100% information transfer. Ten presentations of each consonant by a male speaker were used in the tests with processors MP28m3 and MP2, and five presentations by the same speaker were used in the test with processor MP6A. The bottom panels show scores from the segmental and open-set tests of the Minimal Auditory Capabilities (MAC) battery. See text for abbreviations. Full scale corresponds to 100% correct. Data are from tests with Ineraid subject SR2.

III. Evaluation of Components in the MiniMed Cochlear Prosthesis

MiniMed Technologies, Inc. recently has introduced a new cochlear prosthesis. The prosthesis includes a programmable speech processor, a high-bandwidth transcutaneous transmission system (TTS), and the UCSF electrode array. The speech processor is capable of implementing versions of the *compressed analog*, *interleaved pulses*, and CIS strategies. Further, the coupling configuration of the electrode array can be altered under external control. Stimuli may be delivered to as many as 16 monopolar electrodes, 8 "offset radial" bipolar electrodes, 8 longitudinal bipolar electrodes, or to combinations of monopolar and offset radial electrodes.

We have evaluated the speech processor and TTS in preliminary studies with Ineraid subject SR2. The outputs in the receiver of the TTS were routed through the percutaneous connector of SR2's implant system. The speech processor was programmed to implement a six-channel CIS processor with the following parameters: pulse duration of 74 μs/phase, pulse rate of 1126 pps on each channel, pulses delivered in an apex-to-base order (1-2-3-4-5-6), 6th-order bandpass filters, halfwave rectifiers and 400 Hz lowpass filters (1st order) in the envelope detectors, and monopolar electrode coupling. In addition, the MiniMed processor used an automatic gain control (AGC) in an initial stage of processing, so that the signal presented to subsequent stages would have an approximately uniform level across a 400 ms integrating window (AGC attack was less than 1 ms; AGC release was 400 ms).

This processor was compared to several other CIS processors, using our laboratory system and the DSP56001 prototype. The other processors included MP28m3 and MP2, which have been described in section II of this report. These processors are representative of typical CIS implementations. A final processor, MP15, was designed to mimic closely the parameters used in the MiniMed processor implementation. In particular, all parameters listed above for the MiniMed processor were used except the one for pulse duration. The duration used in MP15 was 67 μs/phase, instead of 74 μs/phase, to allow rapid implementation of MP15 on our DSP56001 prototype processor without having to recompute all the filter coefficients for a new master clock frequency (filter coefficients were available for clock intervals of 33 μs, 67 μs, and 100 μs). Like the MiniMed processor, MP15 used an AGC with rapid attack (< 1 ms) and slow release (400 ms) times.

Results from the comparison of the MiniMed processor implementation with processor MP28m3 are presented in Fig. 2. Our full battery of tests was administered for each of the processors, including use of both male and female speakers for the tests of consonant and vowel identification (vowels included /i, ɔ, ɛ, u, I, U, ʌ, æ/, presented in an /h/-vowel-/d/ context, using the Iowa laser videodisc materials). In addition, the CID and NU-6 tests were repeated for both processors using novel lists of sentences and words and a different recorded male speaker. Results from these repeated tests were either identical or quite similar to the results from the initial tests. The scores presented in Fig. 2 are averages of the two test administrations for each processor.

As is evident from Fig. 2, results obtained with the MiniMed processor were similar to the results obtained previously with processor MP28m3. Processor MP28m3 appears to provide somewhat greater transmission of consonant duration information, while the MiniMed processor appears to provide somewhat greater transmission of first formant information for vowels. Scores for all other

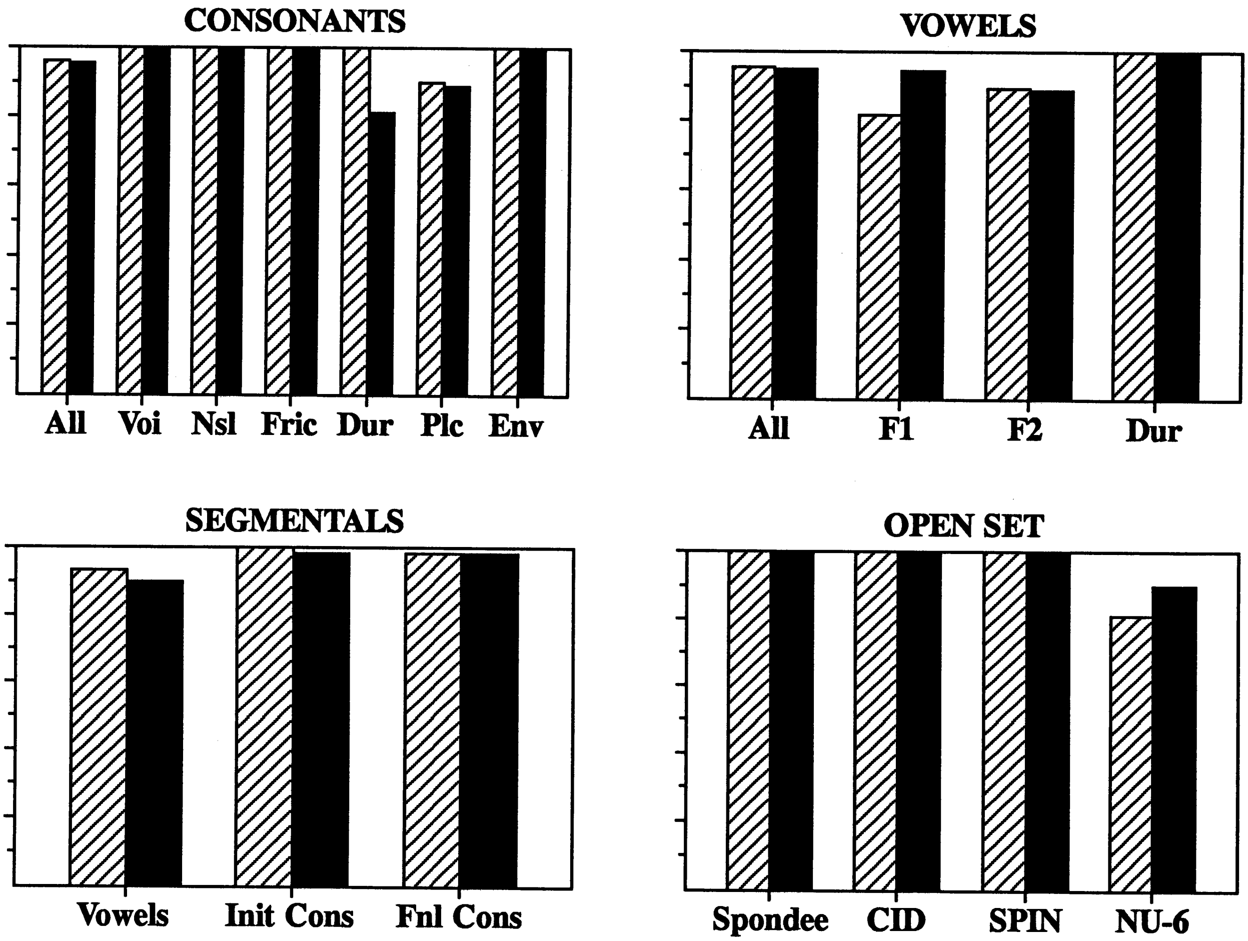

FIG. 2. Comparison of speech test scores for laboratory processor MP28m3 (striped bars) and the MiniMed implementation of a six-channel CIS processor (solid bars). Twenty presentations of each of 16 consonants were used in the consonant identification test for processor MP28m3, and ten presentations of each consonant were used for the MIniMed processor. For the vowel identification tests 18 presentations of each of 8 vowels were used for both processors. Presentations for both the consonant and vowel tests were equally divided between male and female speakers for both processors.

consonant and vowel features are either identical or essentially identical for the two processors. Similarly, scores for the segmental tests, of vowel and consonant recognition, are quite close for the two processors. Open-set scores for the spondee, CID and SPIN tests are 100% correct for both processors. Finally, the MiniMed processor appears to provide somewhat higher scores for the NU-6 test of monosyllabic word recognition. The average NU-6 score for the MiniMed processor was 90% correct, while the average score for processor MP28m3 was 81% correct.

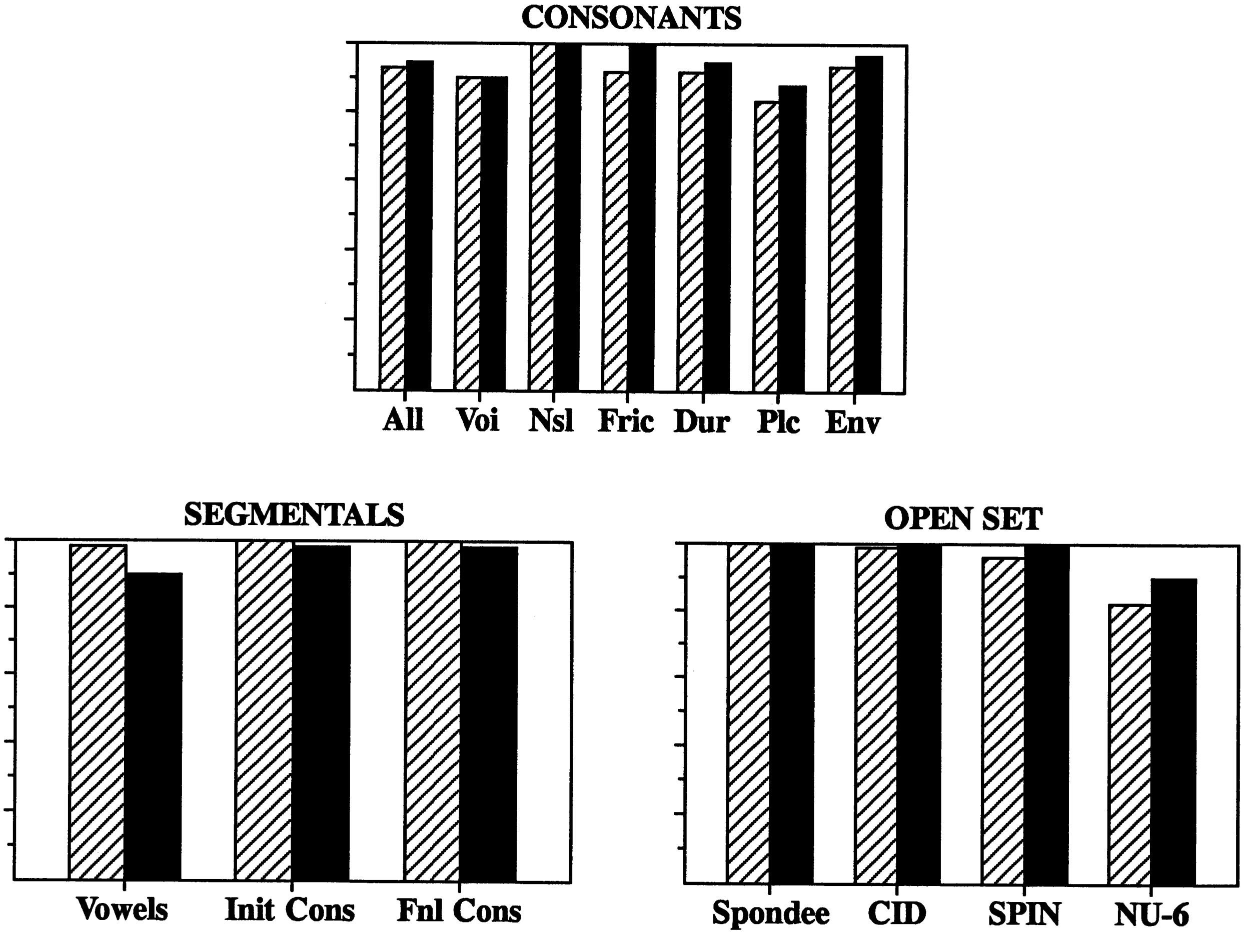

FIG. 3. Comparison of speech test scores for laboratory processor MP2 (striped bars) and the MiniMed implementation of a six-channel CIS processor (solid bars). Twenty presentations of each of 24 consonants were used in the consonant identification test for processor MP2, and ten presentations of each consonant were used for the MiniMed processor. The presentations for each processor were equally divided between the male and female speakers.

The similarity of results for the MiniMed processor implementation and MP28m3 demonstrates that the MiniMed implementation is capable of supporting the same high levels of speech recognition obtained with one of the better implementations of the CIS strategy using our laboratory system. Subject SR2 mentioned that the MiniMed processor was highly intelligible and further that use of the AGC brought the voices of all speakers in the room up to comfortable loudnesses, which he liked.

Results from the comparison of the MiniMed processor implementation with processor MP2 are presented in Fig. 3. For this comparison the consonant identification test was expanded to include 24

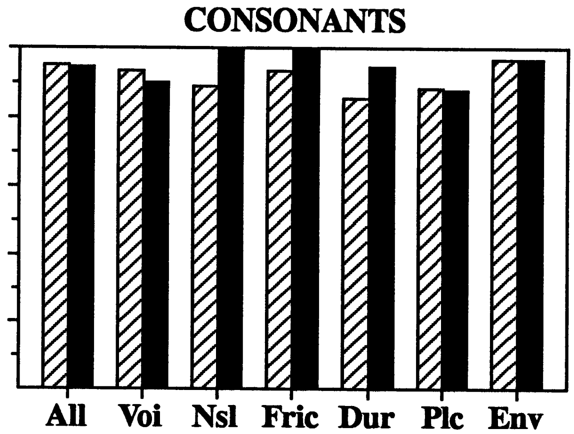

FIG. 4. Comparison of speech test scores for laboratory processor MP15 (striped bars) and the MiniMed implementation of a six-channel CIS processor (solid bars). Five presentations of each of 24 consonants by the male speaker, and five presentations of each consonant by the female speaker, were used in the tests with each processor.

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/), again using both male and female speakers. In addition, the comparison included the segmental and open-set tests of the MAC battery. As with the previous comparisons (Figs. 1 and 2), the CID and NU-6 tests were repeated using novel lists and a different recorded speaker. Scores from the repeated tests again were identical or essentially identical to the scores from the initial tests, and the results shown in Fig. 3 are the averages of the two tests for each processor.

Once again a strong similarity of results is seen for the comparison of the MiniMed processor and processor MP2. The MiniMed processor appears to have a slight advantage in the transmission of frication information (upper panel, Fig. 3), and processor MP2 appears to support somewhat better recognition of vowels, as demonstrated by the difference in scores for the segmental vowel test (lower left panel, Fig. 3). Scores for the spondee and CID tests are at or near the 100% ceiling for both processors. Scores for the SPIN and NU-6 tests appear to be slightly better with the MiniMed processor.

The final comparison involved the use a laboratory processor designed to mimic closely the CIS parameters and AGC of the MiniMed processor. The comparison included tests of consonant identification, using all 24 consonants and both male and female speakers.

Results from these tests are presented in Fig. 4. Essentially identical scores are found for overall transmission, voicing, place of articulation, and envelope cues. Scores for the remaining features of nasality, frication, and duration information appear to be somewhat higher with the

MiniMed processor. All scores approach or hit the ceilings of 100% information transmission. These are quite high scores, especially for multiple speakers and a large number of consonants.

All three comparisons between the MiniMed and laboratory processors demonstrated the efficacy of the MiniMed processor, at least for subject SR2. We note that SR2 is among the best implant patients in the world in terms of his speech recognition abilities, that his test performance remains quite high across ranges of CIS processor parameters that would produce large performance differences in other patients, and that the high scores shown in Figs. 2-4 should not be anticipated for more typical subjects.

IV. Evaluation of Automatic Gain Control

Two of the comparisons presented in the last section were between processors that did and did not use an automatic gain control (AGC). These included the comparison of the MiniMed processor with processor MP28m3 (Fig. 2) and the comparison of the MiniMed processor with processor MP2 (Fig. 3). In both comparisons the MiniMed processor used an AGC and the laboratory processor did not.

While similar results were obtained for the two processors in each of these comparisons, suggesting little or no effect of the AGC on speech reception performance with our recorded test materials, other differences in the processors could have masked an improvement produced by the AGC, or could have compensated for a deleterious effect produced by the AGC. Therefore, we wanted to compare processors that were identical in all respects except for the presence or absence of an AGC.

Two such comparisons were made using the designs of processors MP6A, described in section II, and MP15, described in section III. In particular, processors MP6 and MP15 (with AGC) were compared with processors MP6A and MP15N (without AGC). Otherwise MP6A and MP6 were identical, as were MP15N and MP15.

Results from the first comparison are presented in Fig. 5. The tests included identification of 24 consonants by the male speaker, and the segmental and open-set tests of the MAC battery. As with prior comparisons, the CID and NU-6 tests were repeated for both processors using novel lists and a different recorded speaker. Results from the repeated measures again were identical or essentially identical to the results from the initial tests. The scores presented in Fig. 5 are averages of the two test administrations for each processor.

All scores except those for the NU-6 test approach or hit the 100% ceilings for processors MP6 (with AGC) and MP6A (without AGC). In addition, the NU-6 scores are quite close for the two processors. In terms of these tests, then, use of the AGC does not produce either an improvement or decrement in speech recognition. More difficult tests would be required to demonstrate an effect of the AGC on the recognition of recorded material, if indeed such an effect exists.

Results from the second comparison, presented in Fig. 6, support the same conclusion. Here processors MP15 and MP15N were evaluated with tests of consonant identification, using all 24 consonants and both male and female speakers. Again, scores obtained with the processor using an AGC (MP15) were quite similar to the scores obtained with the processor not using an AGC (MP15N). The scores for voicing and nasality appear to be somewhat better for the processor without an AGC, while the score for place of articulation appears to be somewhat better for the processor with the AGC. All other scores are essentially identical for the two processors.

These findings indicate that use of an AGC does not degrade CIS performance, as measured by the described tests. This outcome further suggests that an AGC can be used in a portable processor, to equalize loudnesses of speakers at different distances from the microphone, without compromising the intelligibility of speech percepts.

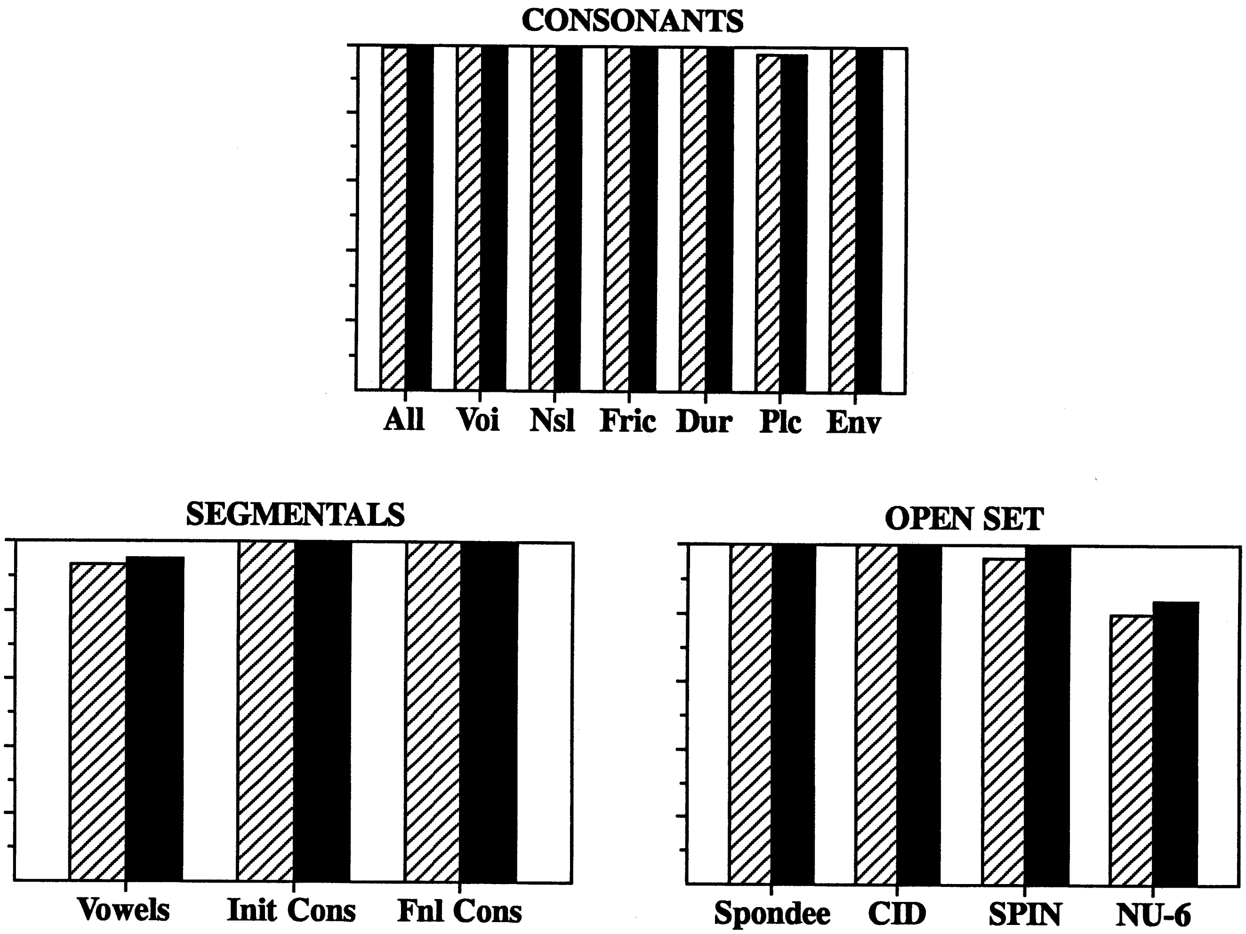

FIG. 5. Comparison of speech test scores for a CIS processor using an AGC (MP6, striped bars) and an otherwise identical processor not using an AGC (MP6A, solid bars). The AGC used in processor MP6 had a rapid attack time of less than 1 ms, and a slow release time of 400 ms. Five presentations of each of 24 consonants by the male speaker were used in the consonant identification tests for both processors.

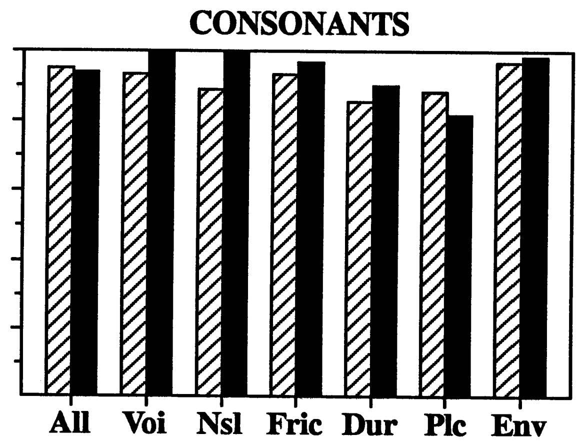

FIG. 6. Comparison of speech test scores for another CIS processor using an AGC (MP15, striped bars) and an otherwise identical processor not using an AGC (MP15N, solid bars). The AGC used in processor MP15 was identical to the one used in processor MP6 (Fig. 5). Five presentations of each of 24 consonants by the male speaker, and five presentations of each consonant by the female speaker, were used in the consonant identification tests for both processors.

V. Preliminary Studies of Modulation Perception

Perception of features in modulated pulse trains has obvious relevance to the design of the CIS, *interleaved pulses* [Wilson et al. 1991b], *Multipeak* [Dowell et al., 1991; Patrick and Clark, 1991; Skinner et al., 1991], and *spectral maxima sound processor* [McKay et al., 1991; McKay and McDermott, 1991] strategies. However, relatively few studies of modulation perception have been conducted using electrical stimuli. To date only Temporal Modulation Transfer Functions (TMTFs) have been measured [Shannon, 1986; 1989; 1991], and results have been similar to the TMTFs of listeners with normal hearing, for analogous acoustic stimuli.

We have conducted pilot studies with Ineraid subject SR2 to investigate other aspects of modulation perception. In these studies high frequency pulse trains (e.g., 2000 pps) were modulated by sinusoids at lower frequencies. The depth of modulation could be varied from 0% percent (no modulation) to 100% (full modulation). Among the preliminary findings were the following:

- Differences in modulation frequency could be detected up to about 600 Hz on one tested channel (electrode 1) and up to about 300 Hz on another (electrode 4). These apparent limits of "pitch saturation" were lower than those for detection of changes in rates of pulsatile stimulation for this same subject on the same channels (in excess of 800 pps for both channels).
- The percepts elicited by modulated pulse trains took on a rough quality when the increasing modulation frequency approached 1/4 of the carrier frequency.
- A continuous whistle was heard when the carrier frequency fell below a critical level of about 800 pps.
- Discrimination of changes in modulation frequency appeared to be worse than discrimination of frequency changes for unmodulated stimuli (e.g., trains of pulses presented at low repetition frequencies).
- In the range of modulation frequencies below the cutoff of the TMTF (e.g., below 150-200 Hz), modulation depths of at least 10% were required before the modulation could be heard without concurrent perception of the carrier (this is different from the same/different task used in measuring the TMTF).
- Further increases in the depth of modulation produced either *no* changes or quite subtle changes in percepts, including loudness, quality, etc.
- Peak stimulus levels corresponding to threshold and most comfortable loudness (MCL) were nearly constant across modulation depth, except for a reduction of both at 0% modulation.

These apparent features of modulation perception, if confirmed in formal studies with additional subjects, have important implications for the design of CIS and other pulsatile processors. We plan to begin such formal studies in the late winter or early spring of 1992.

VI. Measures of Dynamic Range for a Variety of Pulse Durations and Rates

The fitting of CIS processors requires measures of threshold and most comfortable loudness (MCL) for pulsatile stimulation on each electrode channel [Wilson et al., 1991a; QPR 4, this project]. We have made these measures for a variety of pulse durations and rates for all subjects in our Ineraid series. A consistent and somewhat surprising finding is that dynamic range, from threshold to MCL, increases with increases in pulse rate.

A representative set of measures is presented in Fig. 7. For each pulse duration thresholds are shown with the open symbols and MCLs with the closed symbols, for the indicated rates of pulsatile stimulation. The data are from tests with subject SR2, using the most apical electrode in his implant (electrode 1).

In all cases, threshold declines monotonically with increases in pulse rate. The rate of decline approximates the 3 dB/octave figure reported by Shannon for rates in excess of 100 pps and for pulse durations of 400 μs/phase or less [Shannon, 1985].

In contrast to the 3 dB/octave rate of decline for threshold, stimulus levels required for attaining MCL exhibit a much shallower falloff across pulse repetition frequency, for all tested pulse durations. Therefore, the dynamic range between threshold and MCL grows with increases in pulse rate.

These increases in dynamic range (DR) are illustrated in Table 1 and Fig. 8. Both show the calculated DRs for the stimulus conditions of Fig. 7. Notice that the increases in DR from low rates of stimulation to high rates of stimulation can be quite large. For example, a DR of 10.7 dB is found for 33 μs/phase pulses presented at 150 pps, while a DR of 18.3 dB is found for the same pulses presented at 5000 pps.

Finally, Fig. 9 shows DRs for all six electrode channels in SR2's implant, for 100 μs/phase pulses presented at 100 and 800 pps. As with the most apical channel, large increases in DR are observed for channels 2-6 when pulses are presented at the higher rate. The average increase across channels is 4.9 dB, which is a substantial fraction of the total DR at 100 pps.

These results indicate an advantage in the use of relatively high rates of stimulation for pulsatile processors in that DRs will be greater with the higher rates. However, other factors also may be important in choosing optimal stimulus parameters. For several of the subjects in our Ineraid series, for instance, we have observed that the salience of channel ranking can decline with decreases in pulse widths below 100 μs/phase. A favorable tradeoff for such subjects might involve the use of broad duration pulses (e.g., 100 μs/phase or greater) to preserve channel cues, while sacrificing some of the potential DR, that could be obtained with shorter pulses and higher rates of stimulation. We plan to evaluate such tradeoffs in much greater detail in future studies.

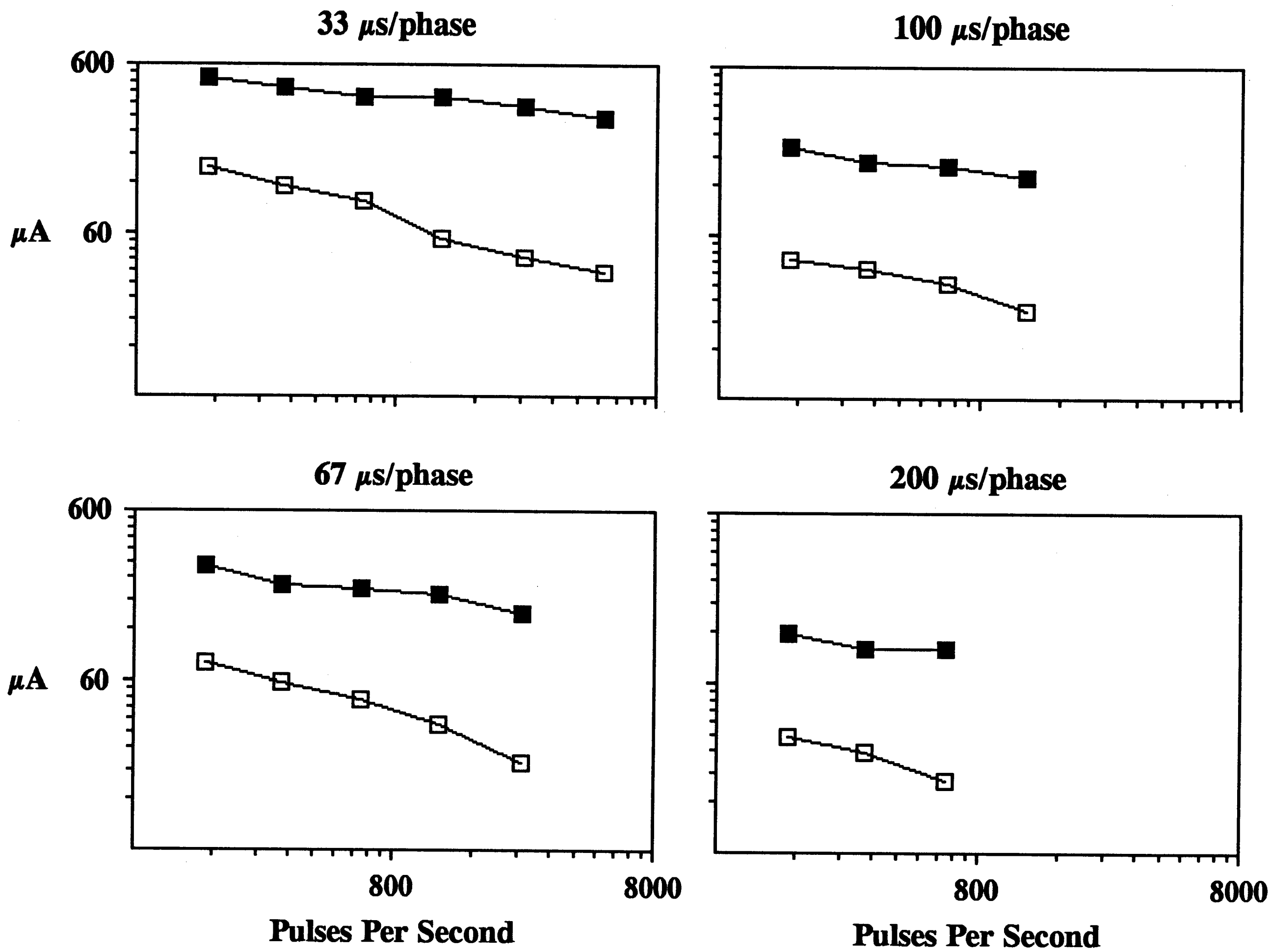

FIG. 7. Stimulus levels corresponding to threshold (open symbols) and most comfortable loudness (closed symbols) for a variety of pulse durations and rates. Pulses were delivered in 50 ms bursts. Data are from measures on electrode 1 for subject SR2.

TABLE 1. Dynamic ranges from threshold to most comfortable loudness for a variety of pulse durations and rates. Data are from measures on electrode 1 for subject SR2.

	Pulse Duration (μs/phase)			
Rate (pps)	33	67	100	200
5000	18.3			
2500	17.9	17.5		
1200	16.8	15.2	16.1	
600	12.2	13.0	14.2	15.6
300	11.8	11.5	12.8	12.3
150	10.7	11.5	13.5	12.2

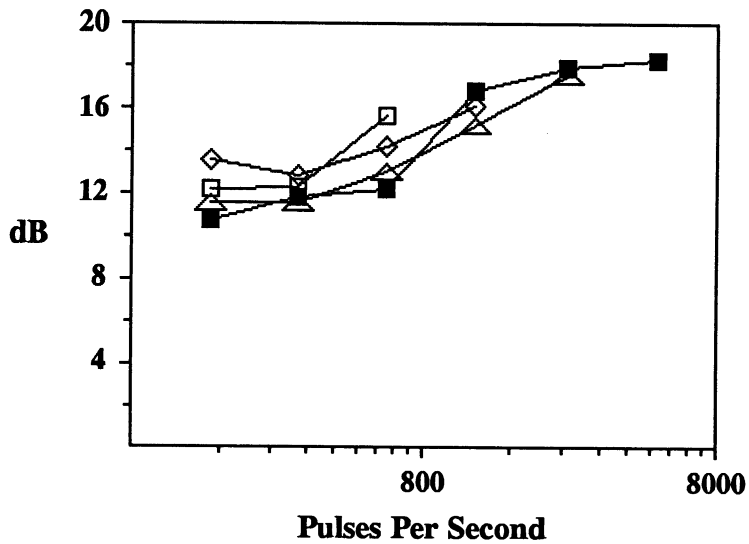

FIG. 8. Dynamic ranges (DRs) from threshold to most comfortable loudness for a variety of pulse durations and rates. Different symbols show the DRs for different pulse durations (closed squares for 33 μs/phase, triangles for 67 μs/phase, diamonds for 100 μs/phase, and open squares for 200 μs/phase). Data are from measures on electrode 1 for subject SR2.

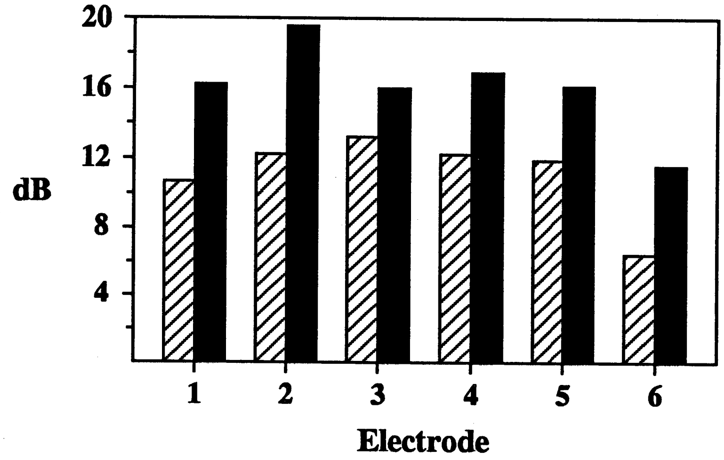

FIG. 9. Dynamic ranges (DRs) from threshold to most comfortable loudness (MCL) for the six intracochlear electrodes in the Ineraid implant. The striped bars show DRs for 100 μs/phase pulses presented at 100 pps, and the solid bars show DRs for those pulses presented at 800 pps. Data are from measures with subject SR2.

VII. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Initiation of studies with Ineraid subject SR9, to evaluate CIS processors with a patient who has relatively poor results using her clinical *compressed analog* (CA) processor. (This continues a new series of studies with patients who have poor clinical performances. Results for the first subject in the series were presented in QPR 5 for this project. A total of approximately six subjects is planned for the series.)
2. Further studies with subject SR2, to complete work begun in the present quarter. The studies will include further exploration of parametric spaces for CIS processors, evaluation of new processor structures such as the hybrid *peak picker*/CIS strategy, and measures of consonant identification for the CA and CIS processors in the presence of multitalker speech babble.
3. Presentation of project results in invited lectures at the *International Symposium on Natural and Artificial Control of Hearing and Balance*, to be held in Rheinfelden, Switzerland, September 4-8, the *Annual Meeting of the American Neurotology Society*, to be held in Kansas City, MO, September 21, and the *Neural Prosthesis Workshop*, to be held in Bethesda, MD, October 22-24.
4. Continued preparation of manuscripts for publication.

VIII. References

Dowell, R.C., Dawson, P.W., Dettman, S.J., Shepherd, R.K., Whitford, L.A., Seligman, P.M, and Clark, G.M. (1991). "Multichannel cochlear implantation in children: A summary of current work at the University of Melbourne," Am. J. Otol. **12**, Suppl. 1, 137-143.

Lawson, D.T., Wilson, B.S., and Finley, C.C. (1989). "Speech processors for auditory prostheses," Fourteenth Q. Prog. Rep., NIH project N01-NS-5-2396.

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

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

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

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

Patrick, J.F., and Clark, G.M. (1991). "The Nucleus 22-channel cochlear implant system," Ear Hear. **12**, Suppl. 1, 3S-9S.

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

Shannon, R.V. (1985). "Threshold and loudness functions for pulsatile stimulation of cochlear implants," Hear. Res. **18**, 135-143.

Shannon, R.V. (1986). "Temporal processing in cochlear implants," in *Sensorineural Hearing Loss: Mechanisms, Diagnosis, and Treatment*, edited by M.J. Collins, T.J. Glattke, and L.A. Harker (University of Iowa Press, Iowa City, IA), pp. 323-334.

Shannon, R.V. (1989). "Detection of modulation in pulse trains by patients with cochlear implants," J. Acoust. Soc. Am. **85**, Suppl. 1, s68.

Shannon, R.V. (1991). "Temporal modulation transfer functions in patients with cochlear implants," J. Acoust. Soc. Am., in press.

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

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

Wilson, B.S., Lawson, D.T., Finley, C.C., and Wolford, R.D. (1991b). "Coding strategies for multichannel cochlear prostheses," Am. J. Otol. **12**, Suppl. 1, 56-61.

Appendix 1

Summary of Reporting Activity for the Period of

May 1 through July 31, 1991

NIH Project N01-DC-9-2401

Reporting activity for the last quarter included three invited lectures and publication of a short article in **Nature**, listed below. The published abstract for the *ASA* presentation is reproduced on the next page, and a reprint of the **Nature** article is presented in Appendix 2.

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

Wilson, BS (1991). General Chair. *1991 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, June 2-7.

Wilson, BS (1991). New levels of speech recognition with cochlear implants. *1991 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, June 2-7.

Finley, CC (1991). Models of potential distributions for various types and placements of electrodes. *1991 Conference on Implantable Auditory Prostheses*, Pacific Grove, CA, June 2-7.

Wilson, BS (1991). Strategies for representing speech with cochlear implants. In the special session on speech perception and hearing handicap, *121st Meeting of the Acoustical Society of America*, Baltimore, MD, April 29 to May 3. [Abstract published in **J. Acoust. Soc. Am.**, 89, Part 2: 1957.]

8:30

6SP2. Interference reduction for the hearing impaired. Patrick M. Zurek (MIT, Res. Lab. of Electron., Rm. 36-730, Cambridge, MA 02139)

Despite continuing technical improvements to hearing aids, user dissatisfaction with the benefit provided remains high. The most frequent source of complaint concerns the interfering effect of environmental background noise on speech reception. This paper will review the factors thought to be responsible for this increased susceptibility to interference as well as approaches that are being taken to improve noisy speech reception through hearing aids. Two of the factors that often contribute to poor speech reception are decreased audibility of speech sounds and loss of binaurality. Techniques that have shown promise for reducing interference include adapting the frequency response of single-microphone hearing aids to minimize spread of masking, and fixed and adaptive beamforming using microphone arrays. A summary will be given of the benefits provided by these techniques and their dependencies on acoustic conditions. [Work supported by NIH.]

8:55

6SP3. Auditory psychophysical performance without a cochlea. Robert V. Shannon (House Ear Inst., 2100 W. Third St., Los Angeles, CA 90057)

Auditory psychophysical performance has been measured using electrical stimulation of the remaining VIII nerve and of the cochlear nucleus in deaf patients. Psychophysical measures of temporal envelope processing show relatively unimpaired performance in these patients compared to normal hearing, and speech discrimination scores indicate that speech information relating to temporal envelopes can be effectively transmitted and received. This finding also indicates that the cochlea and VIII nerve may play relatively little role in the following tasks: detection of gaps, detection of modulation, recovery from adaptation, nonspectral pitch discrimination, and duration discrimination. Intensity perception is impaired with electrical stimulation: dynamic range of usable loudness is only 10 to 20 dB and intensity discrimination experiments indicate only 10 to 40 discriminable intensity levels. Frequency resolution is completely absent in electrical stimulation and can be only crudely reconstructed by multiple electrodes stimulating discrete neural segments. Psychophysical experiments on multiple electrodes indicate that patients can perceive and discriminate complex dynamic patterns of electrical activity changing in both stimulation frequency and electrode location. The combination of temporal envelope formation and the coarse "frequency" resolution provided by multiple electrodes is adequate to convey a surprising amount of speech information. [Work supported by NIH.]

9:20

6SP4. Strategies for representing speech with multichannel cochlear implants. Blake S. Wilson (Neuroscience Prog., Res. Triangle Inst., Research Triangle Park, NC 27709 and Div. of Otolaryngol., Duke Univ. Med. Ctr., Durham, NC 27710)

Various strategies for representing speech information with multichannel cochlear implants will be described, including *compressed analog* (CA), *interleaved pulses* (IP), and *continuous interleaved sampling* (CIS) strategies. Results obtained in within-subject comparisons of strategies will be reviewed. In general, these comparisons have demonstrated large differences among strategies. Recent studies with the CIS strategy recorded large individual improvements and established a new standard of open-set speech recognition among seven subjects chosen for high levels of performance with their CA processors. The CIS strategy presents brief pulses in immediate succession across electrode channels, with the pulse amplitudes for each channel reflecting the envelope of the energy in a corresponding frequency band. The high rate of stimulation on each channel is designed to improve the representation of temporal events in speech, while the use of nonsimultaneous pulses is designed to increase the salience of channel cues through elimination of current summation between channels. [Work supported by NIH, through the Neural Prosthesis Program.]

6 THURSDAY AM

9:45

6SP5. Can we really understand speech through the skin? Mary Joe Osberger (Dept. of Otolaryngol., Indiana Univ. School of Medicine, Riley Hospital A56, Indianapolis, IN 46220)

The purpose of this presentation is to (1) present longitudinal data on the speech perception abilities of profoundly hearing-impaired adults and children who use wearable vibrotactile aids, and (2) raise issues relevant to developing improved tactile devices. Data collected to date reveal that most of the profoundly hearing-impaired subjects in this study perceived speech better when using a seven-channel vibrotactile aid than when they used a two-channel device. Even with the seven-channel instrument, the highest levels of performance were limited largely to discrimination and identification of speech features (segmental and suprasegmental) and enhanced speechreading. Whereas these are clinically significant findings, it is not clear if these results reflect perception of linguistically relevant units of speech or merely perception of acoustic events. Understanding words in sentences without visual clues would provide the most convincing evidence

Appendix 2

"Better Speech Recognition with Cochlear Implants,"
reproduced from Nature, vol. 352, pp. 236-238.

5. Jarvik, E. *Basic Structure and Evolution of Vertebrates* Vol. I (Academic, London, 1980).
6. Milner, A. R. In *The Phylogeny and Classification of Tetrapods. Vol. I: Amphibia, Reptiles, Birds* (ed. Benton, M. J.) 59-102 (Clarendon, Oxford, 1988).
7. Sushkin, P. P. *Trav. Inst. pal. Acad. Sci. USSR* **5**, 43-91 (1936).
8. Bystrow, A. P. *Acta zool.* **19**, 209-295 (1938).
9. Grande, L. & Bemis, W. E. *Memoir 1 J. vert. Paleont.* **11** (1991).
10. Campbell, K. S. W. & Barwick, R. E. *Geol. Mag.* **125**, 207-227 (1989).
11. Allis, E. P. *J. Anat.* **56**, 189-294 (1922).
12. Lebedev, O. A. *Dokl. Akad. Nauk. SSSR* **278**, 1407-1473 (1984).
13. Duellman, W. E. & Trueb, L. *Biology of Amphibians* (McGraw-Hill, New York, 1986).
14. Carroll, R. L. *Vertebrate Palaeontology and Evolution* (Freeman, New York, 1988).
15. Edwards, J. L. *Am. Zool.* **29**, 235-254 (1989).

ACKNOWLEDGEMENTS. We thank S. E. Bendix-Almgreen and the staff of the Greenland Geological Survey for organizing the expedition to collect the material and for permission to work on the specimens, and S. M. Finney for preparing the material.

Better speech recognition with cochlear implants

Blake S. Wilson*†, Charles C. Finley*†, Dewey T. Lawson*, Robert D. Wolford‡, Donald K. Eddington§ & William M. Rabinowitz§

* Neuroscience Program, Research Triangle Institute, Research Triangle Park, North Carolina 27709, USA
† Division of Otolaryngology and ‡ Center for Speech and Hearing Disorders, Duke University Medical Center, Durham, North Carolina 27710, USA
§ Research Laboratory of Electronics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, Cochlear Implant Research Laboratory, Massachusetts Eye and Ear Infirmary, Boston, Massachusetts 02114, and Department of Otology and Laryngology, Harvard Medical School, Boston, Massachusetts 02115, USA

HIGH levels of speech recognition have been achieved with a new sound processing strategy for multielectrode cochlear implants. A cochlear implant system consists of one or more implanted electrodes for direct electrical activation of the auditory nerve, an external speech processor that transforms a microphone input into stimuli for each electrode, and a transcutaneous (rf-link) or percutaneous (direct) connection between the processor and the electrodes. We report here the comparison of the new strategy and a standard clinical processor. The standard compressed analogue (CA) processor[1,2] presented analogue waveforms simultaneously to all electrodes, whereas the new continuous interleaved sampling (CIS) strategy presented brief pulses to each electrode in a nonoverlapping sequence. Seven experienced implant users, selected for their excellent performance with the CA processor, participated as subjects. The new strategy produced large improvements in the scores of speech reception tests for all subjects. These results have important implications for the treatment of deafness and for minimal representations of speech at the auditory periphery.

Features of the two strategies are illustrated in Figs 1 and 2. Both strategies use multiple monopolar electrodes to mimic crudely the tonotopic or 'place' coding of frequencies in the normal cochlea. In particular, high-frequency sounds are indicated by stimulating electrodes toward the base of the cochlea, whereas low-frequency sounds are indicated by stimulating electrodes closer to the apex.

In the CA strategy, microphone signals varying over a wide dynamic range are compressed to the narrow dynamic range of electrically evoked hearing[3,4] using an automatic gain control (AGC). The AGC output is filtered into four contiguous frequency bands for presentation to each of four electrodes. Information about speech sounds is contained in the relative stimulus amplitudes among the four electrode channels and in the temporal details of the waveforms for each channel (Fig. 2*b*). A concern associated with this method of presenting information is that only part of it can be perceived by implant patients[5]. For example, most patients cannot perceive frequency changes in stimulus waveforms above a 'pitch saturation limit' in the region of 400 Hz[4,6-9]. Thus, many of the temporal details present in CA stimuli are not accessible to the typical user. In addition, the simultaneous presentation of stimuli may produce significant interactions among channels through vector summation of the electric fields from each electrode[6,10,11]. The resulting degradation of channel independence would be expected to reduce the salience of channel-related cues.

The CIS strategy addresses the problem of channel interactions through the use of interleaved nonsimultaneous stimuli (Figs 1*b* and 2*c*). Trains of balanced biphasic pulses are delivered to each electrode with temporal offsets that eliminate any overlap across channels. The amplitudes of the pulses are derived from the envelopes of bandpass filter outputs. In contrast to the four-channel CA strategy, five or six bandpass filters (and channels of stimulation) generally are used in the CIS system to take advantage of additional implanted electrodes and reduced interactions among channels. The envelopes of the bandpass outputs are formed by rectification and lowpass

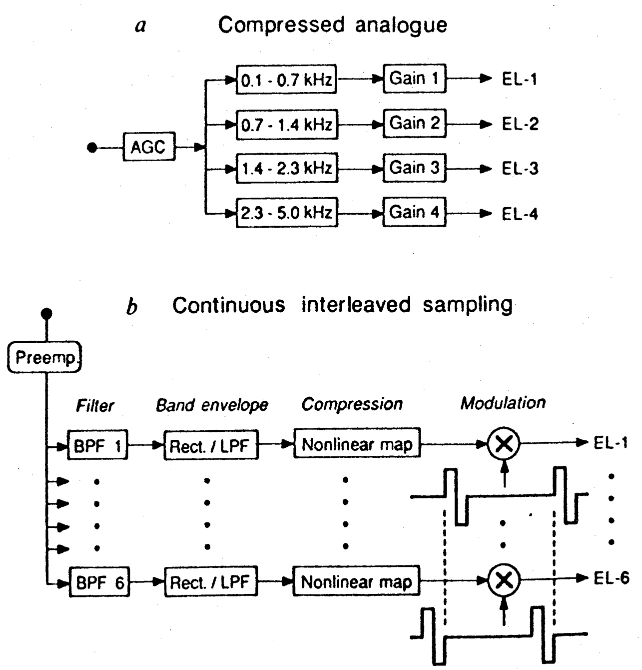

FIG. 1 Block diagrams of major processing steps in the CA and CIS strategies. *a*, The CA strategy uses a broadband AGC, followed by four channels of bandpass filtering (with the indicated frequencies) and adjustable gain controls. The outputs of the gain stages are connected to four intracochlear electrodes (EL-1 to EL-4). *b*, The CIS strategy uses a preemphasis filter (Preemp.) to attenuate strong low-frequency components in speech that otherwise might mask important high frequency components (high frequency emphasis is accomplished in the CA strategy by adjustment of the channel gain controls). The preemphasis filter is followed by five or 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.

TABLE 1 Individual and average scores from the tests of Fig. 3

Subject	Spondee CA	Spondee CIS	CID CA	CID CIS	SPIN CA	SPIN CIS	NU-6 CA	NU-6 CIS	Tracking CA	Tracking CIS
SR2	92	96	100	100	78	96	56	80	81	94
SR3	52	96	66	98	14	92	34	58	51	89
SR4	68	76	93	95	28	70	34	40	—	—
SR5	100	100	97	100	94	100	70	80	—	—
SR6	72	92	73	99	36	74	30	49	43	56
SR7	80	100	99	100	66	98	38	71	51	68
SR8	68	100	80	100	36	94	38	66	56	94
Average	76	94	87	99	50	89	43	63	56	80

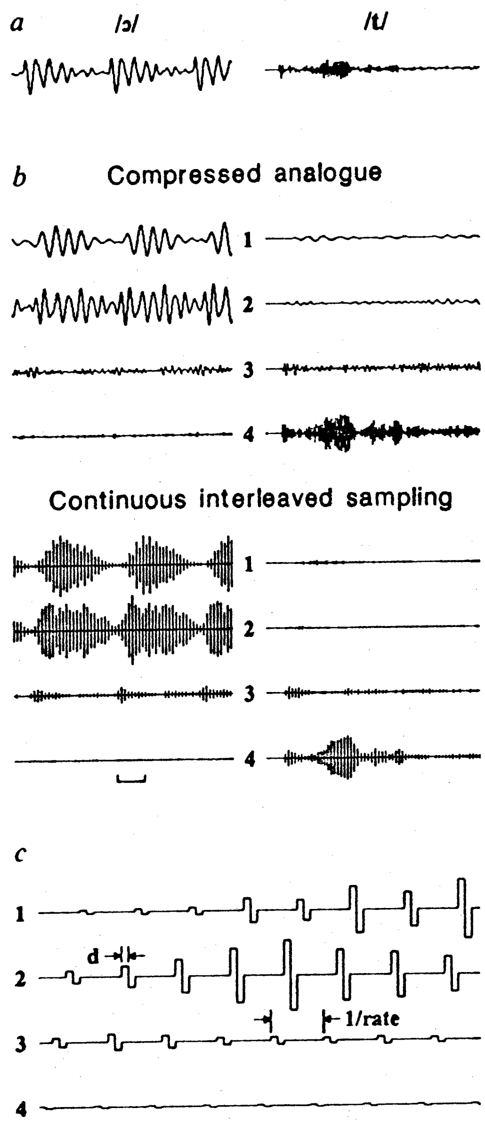

FIG. 2 Waveforms produced by simplified implementations of the CA and CIS strategies. *a*, Preemphasized (6 dB per octave attenuation below 1.2 kHz) speech inputs. Inputs corresponding to a voiced speech sound ('aw') and an unvoiced speech sound ('t') are shown in the left and right columns, respectively. The duration of each trace is 25.4 ms. *b*, Stimulus waveforms for the two processing strategies. The waveforms are numbered by channel, with channel 1 delivering its output to the apicalmost electrode. To facilitate comparisons between strategies, only four channels of CIS stimulation are illustrated here. In general, five or six channels have been used for that strategy. The pulse amplitudes 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 total duration of each trace is 3.3 ms.

filtering. Finally, the amplitude of each stimulus pulse is determined by a logarithmic or power-law transformation of the corresponding channel's envelope signal at that time. This transformation compresses the signal into the dynamic range appropriate for that channel.

A key feature of the CIS strategy is its relatively high rate of stimulation on each channel. Other pulsatile strategies present sequences of interleaved pulses across electrodes at a rate equal to the estimated fundamental frequency during voiced speech and at a jittered or fixed higher rate during unvoiced speech[12,13]. Rates of stimulation on any one channel rarely have exceeded 300 pulses per second (p.p.s.). In contrast, the CIS strategy generally uses brief pulses and minimal delays, so that rapid variations in speech can be tracked by pulse amplitude variations. The rate of stimulation on each channel usually exceeds 800 p.p.s. and is constant during both voiced and unvoiced intervals. The CIS strategy is designed to reduce channel interactions while still representing most or all of the temporal information that can be perceived by implant patients. The use of separate compression functions may further improve the temporal representation by exploiting the full dynamic range of each channel.

Comparisons of the CA and CIS strategies were made in tests with seven patients. All had excellent performance with their Ineraid clinical device[1,2], and were selected to be representative of the best patients using this or any other implant system[13-17]. The comparison tests included open-set recognition of words, sentences and paragraph material. Recognition of words and sentences was evaluated with four tests from the Minimal Auditory Capabilities (MAC) battery[18], and recognition of paragraph material was evaluated using connected discourse tracking[19,20]. The MAC tests included recognition of 50 one-syllable words from Northwestern University Auditory Test 6 (NU-6), 25 two-syllable words (spondees), 100 key words in the Central Institute for the Deaf (CID) sentences of everyday speech, and the final word in each of 50 sentences from the Speech Perception in Noise (SPIN) test (noise was not presented in the present study). All tests were conducted with hearing alone, and all tests except tracking used single presentations of recorded material with no feedback as to correct or incorrect responses. Each subject's own clinical device was used for the tests with the CA processor. Selection of parameters for the CIS processor was guided by preliminary tests of consonant identification[5,12]. The standard four channels of stimulation were used for the clinical CA processors[1,2], whereas five or six channels were used for the CIS processors. All CIS processors had pulse durations of $\leq 102\ \mu s$ per phase, pulse rates of ≥ 817 p.p.s., and a cutoff frequency for the lowpass filters of ≥ 400 Hz.

Results from the MAC and tracking tests are presented in Fig. 3 and Table 1. Both the high scores obtained with the CIS strategy and the substantial improvements made by each subject are impressive. Indeed, the sensitivity of several tests is limited for these subjects by ceiling (saturation) effects: five subjects scored $\geq 96\%$ for the spondee test using the CIS processor; all seven subjects scored $\geq 95\%$ for the CID test; and five subjects scored $\geq 92\%$ for the SPIN test. In addition, three of the tracking scores, of ≥ 89 words per min, approached those of normal

27

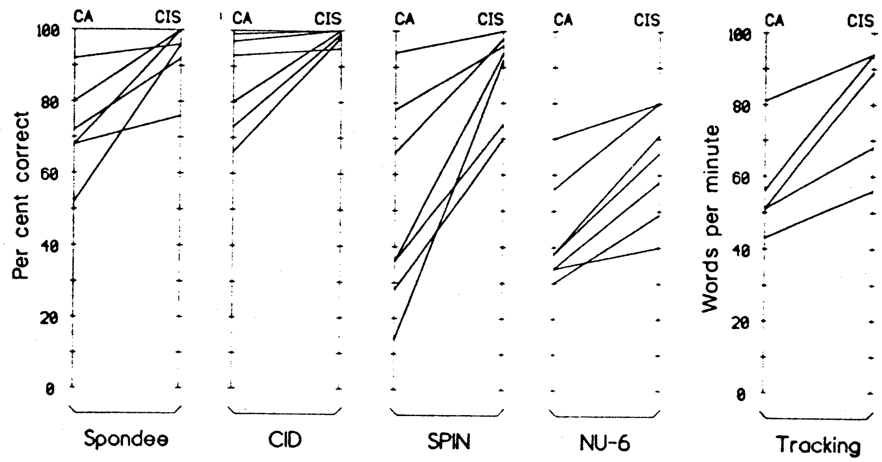

FIG. 3 Speech recognition scores for the CA and CIS processors. For every test, every subject scored higher, or repeated a score of 100% correct, using the CIS strategy. Paired t comparisons show that the increases across subjects are significant for the spondee ($P<0.05$), SPIN ($P<0.01$), NU-6 ($P<0.002$) and tracking ($P<0.02$) tests. In addition, a two-way analysis of the variance, using the five tests and two strategies as factors, demonstrates a highly significant difference between strategies ($F[1,56]=34.1$; $P<0.0000005$), with no significant interaction between factors ($F[4,56]=1.4$; $P>0.2$).

METHODS. In the MAC tests the subject's task was to recognize and report as many words as possible in the NU-6, spondee, CID, and SPIN lists. Because the tests with CA preceded those with CIS, we were concerned that practice or learning effects might favour the latter in comparisons of the two strategies. To evaluate this possibility, the CID and NU-6 tests were repeated with the CIS processor for five of the subjects, using a different recorded speaker and new lists of words and sentences. Practice or learning effects would be demonstrated by significant differences in the test/retest scores. But, no such differences were found ($P>0.6$ for paired t comparisons of the CID scores; $P>0.2$ for the NU-6 scores), and the scores from the first and second tests were averaged for all subsequent analyses. In the tracking test the subject's task was to repeat verbatim unfamiliar paragraphs read by a trained speaker[20]. When errors occurred, various techniques such as repeated presentation of phrases or words were used until verbatim repetition was achieved. The score was calculated by dividing the number of words in four paragraphs by the time required to complete those paragraphs. Because of schedule limitations, the tracking test was included for only five of the seven subjects.

hearing subjects[19-22]. Finally, the 80% score achieved by two of the subjects on the NU-6 test is in the range of scores obtained by people with mild-to-moderate hearing losses when taking the same test[23,24].

The improvements are even more striking considering the large disparity in experience with the two processors. At the time of our tests each subject had 2–5 years of daily experience with the CA processor, but only several hours of experience over a few days with CIS. In previous studies involving within-subject comparisons, such differences in experience have strongly favoured the processor with the greatest duration of use[25,26].

Together these results establish new levels of performance for cochlear implants. Factors contributing to this performance may include (1) reduction in channel interactions through the use of nonsimultaneous stimuli, (2) use of five or six channels instead of four, (3) representation of rapid envelope variations through the use of high pulse rates, (4) preservation of amplitude cues with channel-by-channel compression, and (5) the shape of the compression function. Studies are in progress to evaluate possibilities 1–3 and 5. Preliminary results suggest that all of these factors affect performance, and that factors 1 and 3 are especially important.

In addition to their obvious implications for the treatment of deafness, the present results demonstrate that a simplified representation of speech is able to support relatively high levels of open-set recognition. The information presented by the CIS strategy is limited to envelope variations in five or six bands, with the maximum frequency of the variations (lowpass filter cutoff) typically set at 400 Hz. No specific features of the speech input, such as the fundamental or formant frequencies of voiced sounds, are extracted explicitly. The present results complement existing studies of simplified acoustic representations of speech for normal hearing listeners[27-32] and provide further insight into the minimal cues necessary for speech understanding. □

Received 10 May; accepted 3 June 1991.

1. Eddington, D. K. *J. acoust. Soc. Am.* **68**, 885-891 (1980).
2. Eddington, D. K. *Ann. N. Y. Acad. Sci.* **405**, 241-258 (1983).
3. Pfingst, B. E. *Arch. Otolaryngol.* **110**, 140-144 (1984).
4. Shannon, R. V. *Hear. Res.* **11**, 157-189 (1983).
5. Wilson, B. S., Finley, C. C. & Lawson, D. T. in *Cochlear Implants: Models of the Electrically Stimulated Ear* (eds Miller, J. M. & Spelman, F. A.) 339-376 (Springer, New York, 1990).
6. Eddington, D. K., Dobelle, W. H., Brackmann, D. E., Mladejovsky, M. G. & Parkin, J. L. *Ann. Otol. Rhinol. Lar.* **87**, suppl. 53, 1-39 (1978).
7. Müller, C. G. *Ann. N.Y. Acad. Sci.* **405**, 412-420 (1983).
8. Simmons, F. B. *Arch. Otolaryngol.* **84**, 2-54 (1966).
9. Tong, Y. C., Blamey, P. J., Dowell, R. C. & Clark, G. M. *J. acoust. Soc. Am.* **74**, 73-80 (1983).
10. White, M. W., Merzenich, M. M. & Gardi, J. N. *Arch. Otolaryngol.* **110**, 493-501 (1984).
11. Wilson, B. S., Finley, C. C., Lawson, D. T. & Wolford, R. D. *Proc. IEEE* **76**, 1143-1154 (1988).
12. Wilson, B. S., Lawson, D. T., Finley, C. C. & Wolford, R. D. *Am. J. Otol.* **12**, suppl. 1, 56-61 (1991).
13. Clark, G. M. *et al. Adv. Otol.-Rhinol. Lar.* **38**, 1-189 (1987).
14. Dorman, M. F., Hannley, M. T., Dankowski, K., Smith, L. & McCandless, G. *Ear Hear.* **10**, 44-49 (1989).
15. Gantz, B. J. *et al. Laryngoscope* **98**, 1100-1106 (1988).
16. Schindler, R. A. & Kessler, D. K. *Am. J. Otol.* **8**, 247-255 (1987).
17. Tyler, R. S., Moore, B. C. J. & Kuk, F. K. *J. Speech Hear. Res.* **32**, 887-911 (1989).
18. Owens, E., Kessler, D. K., Raggio, M. W. & Schubert, E. D. *Ear Hear.* **6**, 280-287 (1985).
19. De Filippo, C. L. & Scott, B. L. *J. acoust. Soc. Am.* **63**, 1186-1192 (1978).
20. Owens, E. & Raggio, M. *J. Speech Hear. Disorders* **52**, 120-128 (1987).
21. Robbins, A. M., Osberger, M. J., Miyamoto, R. T., Kienle, M. L. & Myres, W. A. *J. Speech Hear. Res.* **28**, 565-578 (1985).
22. Sparks, D. W., Ardell, L. A., Bourgeois, M., Wiedmer, B. & Kuhl, P. K. *J. acoust. Soc. Am.* **65**, 810-815 (1979).
23. Bess, F. H. & Townsend, T. H. *J. Speech Hear. Disorders* **42**, 232-237 (1977).
24. Dubno, J. R. & Dirks, D. D. *J. Speech Hear. Res.* **25**, 135-141 (1982).
25. Dowell, R. C., Seligman, P. M., Blamey, P. J. & Clark, G. M. *Ann. Otol. Rhinol. Lar.* **96**, suppl. 128, 132-134 (1987).
26. Tyler, R. S., Preece, J. P., Lansing, C. R., Otto, S. R. & Gantz, B. J. *J. Speech Hear. Res.* **29**, 282-287 (1986).
27. Flanagan, J. L. *Speech Analysis, Synthesis and Perception* (Springer, Berlin, 1972).
28. Hill, F. J., McRae, L. P. & McClellan, R. P. *J. acoust. Soc. Am.* **44**, 13-18 (1968).
29. Remez, R. E., Rubin, P. E., Pisoni, D. B. & Carrell, T. D. *Science* **212**, 947-950 (1981).
30. Rosen, S. in *Cochlear Implant: Acquisitions and Controversies* (eds Fraysse, B. & Cochard, N.) 3-26 (Impasse La Caussade, Toulouse, 1989).
31. Rosen, S. M., Fourcin, A. J. & Moore, B. C. J. *Nature* **291**, 150-152 (1981).
32. Van Tasell, D. J., Soli, S. D., Kirby, V. M. & Widen, G. P. *J. acoust. Soc. Am.* **82**, 1152-1161 (1987).

ACKNOWLEDGEMENTS. We thank M. Dorman, R. Tyler, M. Lowder and K. Dankowski for their help in identifying subjects and B. Moore, C. van den Honert and B. Weber for their editorial comments. We also thank each of the subjects for their time and interest. This work was supported by the NIH.

Appendix 3

Real-Time Speech Processor Architecture for the RTI DSP56001 Platform

REAL-TIME SPEECH PROCESSOR ARCHITECTURE FOR THE DSP56001 PLATFORM

Processor software for the 56001 DSP platform is designed to be flexible, enabling rapid creation, verification and patient testing of many different speech processing strategies. The software is modular in nature with the function of individual modules determined by specification of system parameters. Parameter specification for these processors is made at several different software levels. The first, and architecturally the lowest, level is the processor algorithm main body, which could for example be a standard CIS processor, a CIS processor with RMS energy estimators, a peak-picker processor, a CIS processor with stationary noise subtraction, a simultaneous analog processor, or any other variation. The second level is the processor header file. In the header, parameters such as filter coefficient files, number of channels, and type of rectification are specified. The third level is a patient test text file which contains specific psychophysical test data and descriptions of stimulus temporal properties such as pulse phase durations, interpulse intervals and update orders. The top level of parameter control is the monitor program that runs in the controlling PC and is responsible for downloading, execution and monitoring of real-time code running on the 56001 DSP platform. At this level the operator may choose among various mapping law functions on a channel-by-channel basis. Figure 1 shows a flow graph of how a processor is configured.

This versatile software structure provides for the parametric manipulation of every aspect of a given class of processor architecture, yet it produces efficient, compact code capable of real-time processing speeds. It also provides reference files documenting the exact configuration of each processor tested. The following further illustrates the details of specifying the properties of a processor.

FIRST (LOWEST) LEVEL – SPEECH PROCESSOR ALGORITHM

This level contains the main body for each algorithm. A new algorithm often involves simply changing a macro or a function of an existing algorithm. Since small changes at this level can significantly affect memory allocations, register utilization, and efficient use of the intrinsic DSP pipeline architecture, it is to our advantage to work with a pool of basic optimized processors and make parametric changes at a higher level. Currently, the basic algorithms are:

1. CIS with standard energy estimators consisting of rectification followed by a low-pass smoothing filter;
2. Peak-picker;
3. Hybrid peak-picker/CIS;
4. CIS with true RMS energy estimators;
5. CIS implemented using two DSP56000 processors, setting the stage for expanded processor throughput for future designs;
6. CIS with pseudo stationary noise subtraction on a channel-by-channel basis.

Listing 1 is an example of the basic algorithm for the standard CIS processor. The modular program uses macros for coding a group of instructions that are repeated within the pool of algorithms.

SECOND LEVEL -- REAL-TIME PROCESSOR HEADER

The processor header is an assembly code file where variables and coefficient files are defined. Some of these variables are used for assembly control so that the same main program can be used while changing small parts of the assembled code itself. A header file is attached as listing 2.

The following variables or files can be changed easily in the header without the need for writing new code:

1. Number of channels;
2. Multiplexer addresses for the channels;
3. Monopolar or bipolar stimulation(defined by mux addresses above);
4. Full-wave or half-wave rectification for energy estimators;
5. Filter coefficient files to be included (these files are created using a commercial software package);
6. Orders of bandpass and lowpass filters (number of poles);

The following processor-specific variables are also included in the header:

7. Window width for the RMS algorithm;
8. Number of sample periods before looking for a peak in the peak picker processor;
9. Percentage of noise to subtract in the noise subtraction algorithm.

THIRD LEVEL - PATIENT TEST FILE

The patient test file is a text file written for each individual test. It lists variables that are used by the monitor program to calculate the compression tables and download parameters to the DSP56001 processor. This file also serves as documentation of the processor for keeping track of test conditions. Figure 2 is an example of such a file. The file specifies the following variables:

1. Processor file name -- DSP56001 load file;
2. Patient's thresholds, in microamperes for specific test pulse width and rate of stimulation;
3. Patient's corresponding MCLs in microamperes;
4. Maplaw factor -- factor to multiply the maplaw by before downloading compression tables into DSP memory;
5. Pulse width;
6. Channel order -- either a DSP56001 load file with random order or a particular short list in the file can be specified;
7. Dead time -- time between channel pulses.

As we expand the repertoire of processors and increase the number of variables which are frequently manipulated, the test file will expand. Existing code can be rearranged without much difficulty, so that more variables can be changed at the test file instead of the header. Yet, some variables must be changed at the Header, before assembling the processor, to facilitate generation of efficient, real-time code.

FOURTH (HIGHEST) LEVEL -- MONITOR PROGRAM ON CONTROLLING COMPUTER

The monitor program communicates from the controlling PC to the DSP processor board. The monitor program downloads the processor load file and other variables specified in the test file. It uses the threshold and MCL data specified in the test file to calculate the compression tables. It then downloads the mapping laws into the DSP56001 external memory as specified in the maplaw menu. The following options are available in the maplaw menu:

1. Define a different compression curve for each channel or the same for all channels;
2. Choose one of seven different constants which define the compression curve;
3. Define a different gain for each channel or same for all channels;
4. Define a "noise offset" for each channel -- instead of having the mapping law start immediately at psychophysical threshold, it will map the first n (noise offset) table values to zero and then continue with the maplaw at threshold.

Figure 3 shows the screen display for the maplaw menu. In the example, MP6 was the test file used, as is displayed at the right top corner of the screen. Map 1 was chosen for all the channels (only 6 channels were used) and gains have not been defined yet. The menu displays the map and gain for each channel since they can be specified separately.

Due to the complexity that the DSP56001 code can achieve, we have devised this structure so that minimum time is spent rewriting code and testing it. Not constrained by memory space at this time, we are able to write flexible software. In the future, this code could be optimized for space if necessary.

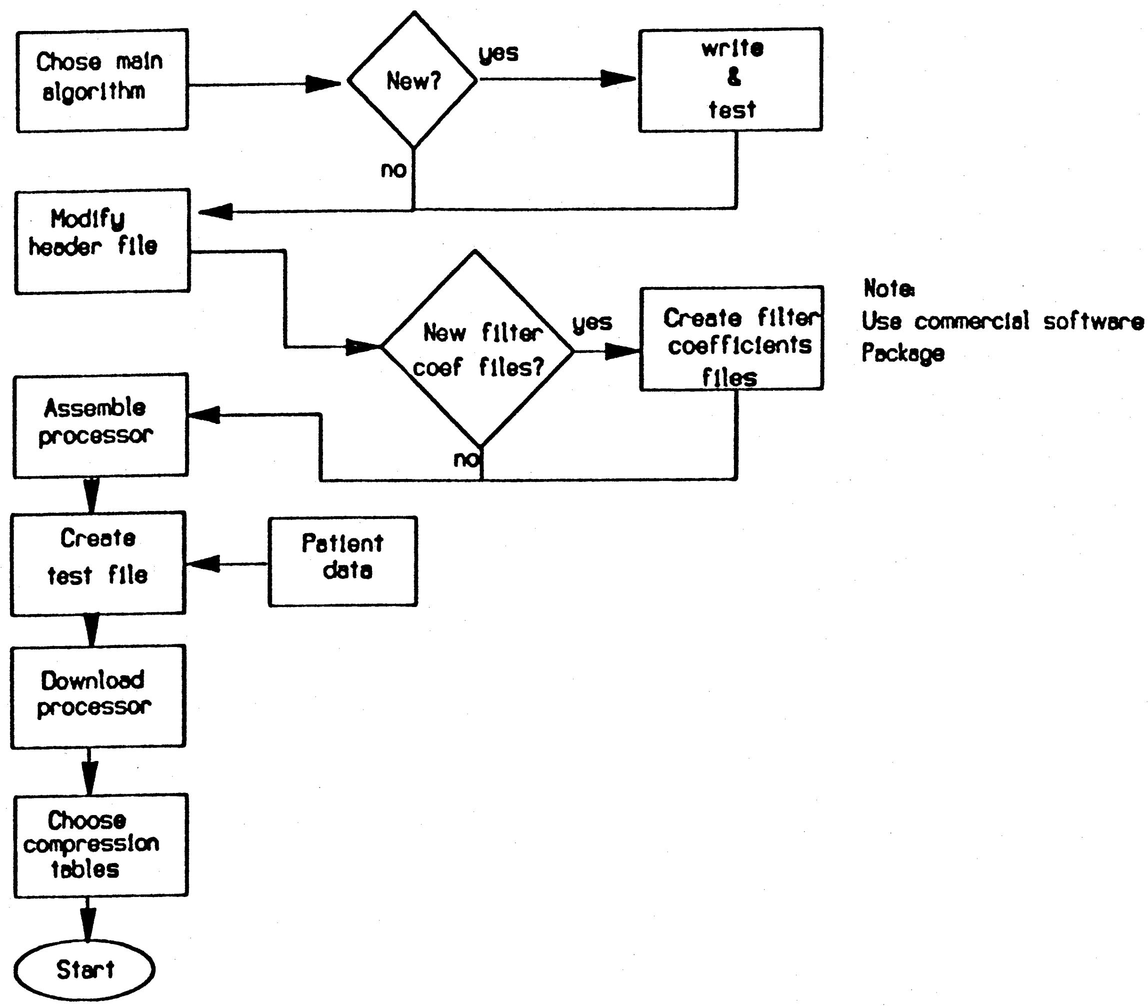

Fig. 1. **Manipulation of the Speech Processor Parameters.**

```
        *MP17
*
* 12th order bandpass filters, 400Hz, 1st order smoother FW rect
a:fb6t114.lod
*factor - double
1.0
*mcl from 1 to 7 - double
293.6
357.6
382.0
390.3
393.7
545.7
1000
*thresholds - double
39.6
34.2
40.5
48.4
48.9
83.5
0
*global gain - decimal integer
4095
*pulse width - integer
33
*flag: o for specified order
*      f for download random data file
*for random data file need to specify name of file and legth
f
a:randout1
8190
```

Fig. 2. **Example of a Patient Test File.**

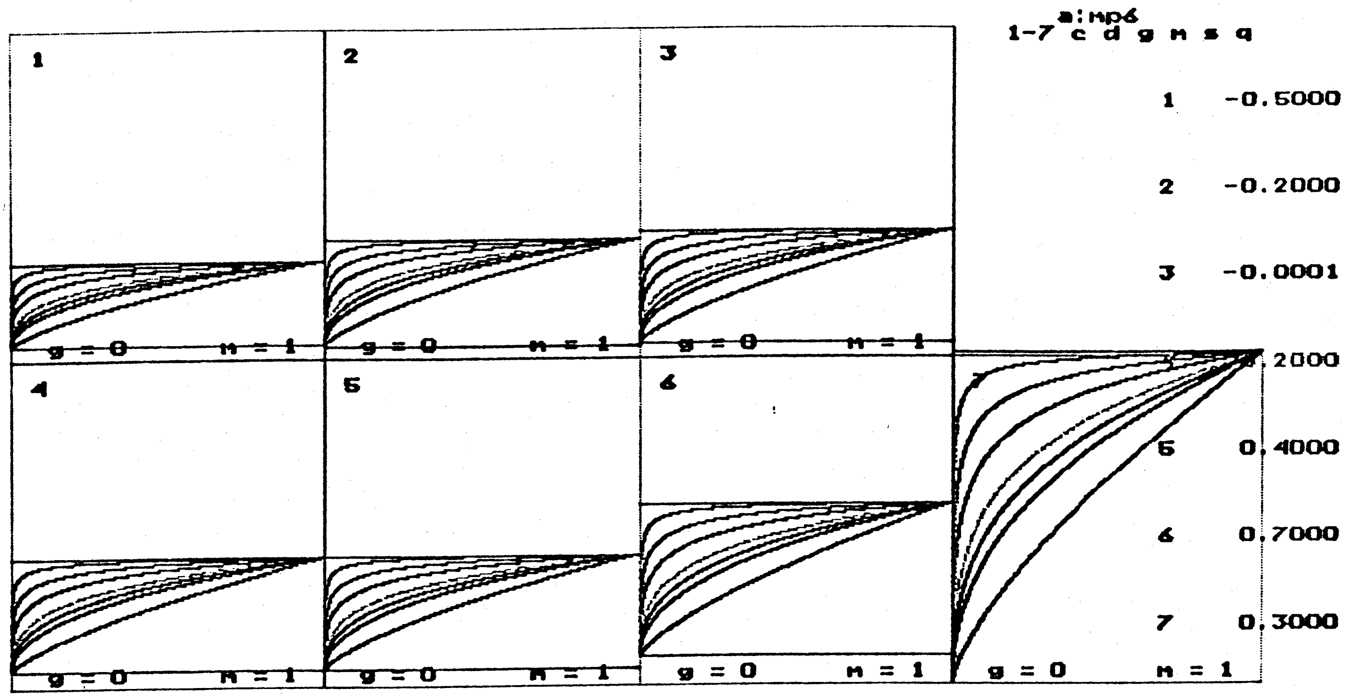

Fig. 3. The maplaw menu displays each channel's compression curves separately. The channel number is displayed at the left top corner of each box. The compression curve loaded is displayed on the right bottom corner of each box, and the output gain is at the left bottom corner. Horizontal lines indicate psychophysical threshold + MCL for each channel.

```
;************************************************************************
;
; FLTS13.ASM
;
;  by: M. Zerbi
;  Date: 6/11/91
;
; Speech processor 13 uses IRQB to sample the ADC.  It has 3 part
; interrupt for the SCI timer - CON, CNEG , and COFF.  It uses a
; list in memory to get to the next channel
;************************************************************************

FLTS13
        INCLUDE 'HEAD13.ASM'     ; defines memory addresses needed

        INCLUDE 'CASCDE13.ASM'   ; macro implements cascade filter

        INCLUDE 'MCHAN13.ASM'    ; macros implements channel outputs

        ORG     X:STATE1
ST1     DSM     TOTM             ; define modulo state1 storage of 1 starting
                                 ; at x:(0)

        ORG     X:STATE2
ST2     DSM     TOTM             ; define modulo state1 storage of 1 starting
                                 ; at x:(0)

        INCLUDE 'SBODY13.ASM'    ; main body - standard CIS
        END
```

Listing 1. Example of a Basic Algorithm

```
;***********************************************************************
;
; SBODY13.ASM
;
; Here is a list of the registers used and for what purpose
;       r0 - X mem addr for state1 calculations
;       m0 = addition of all NSEC - 1
;       r1 = X mem addr for state2 calculations
;       m1 = addition of all NSEC - 1
;       r2 = address for next jump in the interrupt routine
;       r3 = low pass filter input
;       n3 = offset to get table gains
;       r4 = Y mem addr to filter coefficients
;       m4 = 5 * (addition of all NSEC) - 1
;       r6 = Y mem addr for mapping tables
;       r5 = X mem addr where filter's output is saved
;       m5 = number of filters - 1
;       r7 = address to get next channel from the list
;       m7 = how long the list is
;
;***********************************************************************

START   EQU     $40                     ;origin for user program

    ORG     P:$000A                     ;IRQB vector space NOTE assumed that ADC
    JSR     SAMPINT                     ;jump to sample interrupt

    ORG     P:$0018                     ;SCI transmit data vector space
    JSR     TRANSMIT                    ;jump to transmit interrupt - put out pu

    ORG     P:START                     ;origin for user program

    ORI     #3,MR                               ;disable all interrupts
    MOVEP   #M_TIM,X:SCCR                       ;set sci timer values
    MOVEP   #M_IMASK,X:M_IPR                    ;enable sci transmit and IRQB in
    MOVEP   #$1202,X:SCR                        ;enable sci transmit interrupts
    BSET    #1,X:PCC                            ;enable port C pin for transmit

    MOVEP   #WAITREG,X:M_X                      ;set wait state for external mem

    MOVE    #ST1,R0                             ;point to filter state1
    MOVE    #ST2,R1                             ;point to filter state2
    MOVE    #COEF,R4                            ;point to filter coefficients
    MOVE    #5*TOTM-1,M4                        ;addressing modulo 5*nsec

    MOVE    #FTOUT,R5                           ;address where filter output wil
    MOVE    #TOTF+TOTC-1,M5                     ;mod5 = NUMBER OF FILTERS + NUM

    MOVE    #TOTM-1,M0
    MOVE    #TOTM-1,M1

    MOVE    #21,N3                      ;offset to get table gains
```

Listing 1. cont.

```
        CLR     A                       ;initialize internal state storage

        REP     #TOTM                   ;*    zero all x mem used for state1
        MOVE    A,X:(R0)+               ;*    and state2 calculations
        REP     #TOTM
        MOVE    A,X:(R1)+
        REP     #TOTF+TOTC
        MOVE    A,X:(R5)+               ;clear the filters outputs
        MOVE    A,X:FLTR                ;clear the interrupt flag
        MOVE    Y:(R4)+,Y0              ;a must be initially zero ,y0=b10/2

        BSET    #3,Y:SHORT              ;enable VREF for the dac
        MOVE    A,Y:DAC                 ;set dac to 0
        MOVEP   #$B,Y:SHORT             ;enable the mux and close switch
        MOVEP   #0,Y:AMUX               ;latch speech to ADC MUX

        IF  TEST==0
        MOVE    #CONA,B1
        MOVE    B1,X:$1
        MOVE    #CONB,B1
        MOVE    B1,X:$2
        MOVE    #CONC,B1
        MOVE    B1,X:$3
        MOVE    #COND,B1
        MOVE    B1,X:$4
        MOVE    #CONE,B1
        MOVE    B1,X:$5
        MOVE    #CONF,B1
        MOVE    B1,X:$6
        ENDIF

        MOVE    #ADRTBL,R7              ;start of table for channel addresses
        MOVE    #>$000F,X1
        MOVE    X:(R7)+,B
        AND     X1,B
        MOVE    B1,R2
        MOVE    X:TBL,M7                ;length of table
        MOVE    X:(R2),R2               ;get first address


        ANDI    #$fc,MR                 ;allow interrupts

WT
        WAIT
        JCLR    #0,X:FLTR,WT            ;wait till new sample available
                                                    ;put DSP to slee
        BCLR    #0,X:FLTR               ;clear the interrupt flag
        ORI     #$08,MR                 ;setting scaling mode - scale up

;
; FILTERS
;
; sample- highpass filter -- bandpass filter - rectifier - lowpass filter
;                          |- bandpass filter - rectifier - lowpass filter
```

Listing 1. cont.

```
;                                 ├ bandpass filter - rectifier - lowpass filt
;                               ├ bandpass filter - rectifier - lowpass filter
;                                 └ bandpass filter - rectifier - lowpass filt
;
; NOTE:  Bandpass filters done in DO LOOP is OK until you decide to use
; bandpass filter with different number of poles.  Then it would be easi
; to do it straight like the lowpass filters.  The low pass filter are d
; this way so you can have option of changing rectifiers.
;
    RDADC                              ;MACRO - manipulate sample data
    CASCDE    1,1                      ;MACRO - filter for one pole  highpass

    DO  #TOTC,ENDBP                    ;Do bandpass filter 1 through 6
    CASCDE  BPNSEC,1                   ;MACRO - BPn
    MOVE     X:FTOUT,Y1                ;move output of HP filter into y1
ENDBP
.
    MOVE         #CTAB1,R6
    MOVE         #FTOUT+1,R3
    DO           #TOTC,ENDLP
    LPFILTER                           ;low pass filter macro wich includes rec
                                           ;and output manipulation
ENDLP
    NOP
    JMP        WT

;***********************************************************************
; SAMPINT
; IRQB interrupt routine.
; The routine samples ADC
;
SAMPINT
    BSET         #0,X:FLTR            ;set flag that interrupt occured
    MOVEP        Y:ADC,X:SAMPLE       ;read ADC
    BCHG         #23,X:SAMPLE         ;change data to offset binary
    MOVEP        #SMPL,Y:CONV         ;dummy move to send cmd to 2nd adc
RTN RTI

;***********************************************************************
; TIMINT
;
; SCI interrupt routine.
; The routine sets the interrupt flag FLTR.
;
    IF           TEST==0                        ;when sending pulses to DAC
TRANSMIT
    ORI       #$08,MR                           ;set scaling mode
    JMP          (R2)                           ;jump to corresponding E channel

;
;         Channel n,m
; calling macros were first letter defines the channel it is
; outputing information to, the number identifies the electrode.
; The next channel it will jump to will be defined by list pointed to by
;
```

Listing 1. cont.

```
;

    MCHAN      A,1                  ;channel A, electrode 1
    MCHAN      B,2
    MCHAN      C,3
    MCHAN      D,4
    MCHAN      E,5
    MCHAN      F,6
    ZERODAC                         ;macro to turn off the channels

    ELSE

TRANSMIT
    MOVEP      #0,X:STX                    ;restart clock

    MOVE       X:DACOUT,R2                 ;get filter output
    MOVE       B1,X:SAVEB1
    MOVE       B0,X:SAVEB0
    MOVE       X0,X:SAVEX0
    MOVE       X:(R2),X0
    MOVE       X1,X:SAVEX1
    MOVE       X:TGAIN,X1
    MPY        X0,X1,B          X:SAVEX1,X1
    MOVE       B1,Y:DAC
    MOVE       X:SAVEB1,B
    MOVE       X:SAVEB0,B0
    MOVE       X:SAVEX0,X0
    RTI

    ENDIF
```

Listing 1. cont.

```
;*****************************************************************
; HEAD13.ASM
; Variables are defined here
;
; NOTE: variables that might likely change with the change of structure
; have ** at the begining of the comment
;
;*****************************************************************

; For variable pulse duration change M_TIM accordingly (for 32MHz)
; d=15us -> $6, d=30us -> $E, d=45usec -> $15, d=60usec -> $1D,
; d=75usec -> $24
M_TIM    EQU      $000E             ;** Set SCI timer for 30kHz

TEST     EQU      0                 ;** TEST=1 when you want to check out fi
                                    ; outputs and TEST=0 for normal pulse ge

CADD1     EQU     $0E               ;**channel 1 mux address which is  a  1
                                    ; & E15
CADD2    EQU      $1E               ;**ch 2 mux address  - E2 & E15
CADD3    EQU      $2E               ;**ch 3 mux address - E3 & E15
CADD4    EQU      $3E               ;**ch 4 mux address - E4 & E15
CADD5    EQU      $4E               ;**ch 5 mux address - E5 & E15
CADD6    EQU      $5E               ;**ch 6 mux address - E6 & E15

M_IPR    EQU      X:$FFFF           ; interrupt priority register
M_X      EQU      X:$FFFE           ; IO wait state register
SCCR     EQU      X:$FFF2           ; SCI interface clock control reg
SCR      EQU      X:$FFF0           ; SCI interface control register
PCC      EQU      X:$FFE1           ; Potr C control register
STX      EQU      X:$FFF4           ; SCI transmit register
M_IMASK  EQU      $C828             ; interrupt mask - Highest to lowest IPL
                                    ; host, IRQB (trigger mode = negative ed
WAITREG  EQU      $5555             ; waits states


ADC      EQU      Y:$FFC0           ; Y memory address for ADC
AMUX     EQU      Y:$FFC1           ; Y memory address for ADC Mux
CONV     EQU      Y:$FFC2           ; Y memory address for 2nd ADC convert c
SMPL     EQU      $010000           ; ADC sample command
DAC      EQU      Y:$FFC8           ; Y memory address for DAC write
SINK     EQU      Y:$FFCD           ; test address line to sink osc
MUX      EQU      Y:$FFC9           ; Y memory address for mux
SHORT    EQU      Y:$FFCC           ; Y memory address for shorting switch

STATE2   EQU      X:$80             ; X mem address for state2 - address imp
                                    ; modulo arithmetic
STATE1   EQU      X:$40             ; X mem address for state1
FTOUT    EQU      X:$C0             ; X mem address for filter output
TOTF     EQU      $D                ;** total number of filters including 6
TOTC     EQU      $6                ;**total number of channels
BPNSEC   EQU      $3                ;**nsec for the individual bandpass filt
```

Listing 2. Processor Header File

```
LPNSEC   EQU      $1                  ;**nsec for the individual lowpass filte
TOTM      EQU     TOTC*(BPNSEC+LPNSEC)+1  ;**  Sumation  of nsec of each f
FLTR     EQU      X:8                 ; interrupt flag in Y mem
SAMPLE   EQU      X:0                 ; where ADC input is stored in X mem
GAIN     EQU      X:9                 ; gain to multiply or divide by (asl or
TGAIN    EQU      X:$A                ; test gain for output to DAC
SAVEB0   EQU      X:$FB               ; save b0 acc during interrupt
SAVEB1   EQU      X:$FC               ; save b acc during test
SAVEX1   EQU      X:$FF               ; save a acc during test
SAVEX0   EQU      X:$FE
SAVEA    EQU      X:$FD
TPOS     EQU      X:$C                ;positve pulse width
TNEG     EQU      X:$D                ;negative pulse width


DACOUT   EQU      X:FTOUT+TOTF+TOTC+1      ; address for DAC output on test
CTAB1    EQU      Y:$C000             ; Y mem address for  channel 1 pulse mag

GAIN1    EQU      X:$D7               ;gain value for table output for channel

POSOUTA  EQU      X:FTOUT+TOTF-TOTC
NEGOUTA  EQU      X:FTOUT+TOTF-TOTC+1
POSOUTB  EQU      X:FTOUT+TOTF-TOTC+2
NEGOUTB  EQU      X:FTOUT+TOTF-TOTC+3
POSOUTC  EQU      X:FTOUT+TOTF-TOTC+4
NEGOUTC  EQU      X:FTOUT+TOTF-TOTC+5
POSOUTD  EQU      X:FTOUT+TOTF-TOTC+6
NEGOUTD  EQU      X:FTOUT+TOTF-TOTC+7
POSOUTE  EQU      X:FTOUT+TOTF-TOTC+8
NEGOUTE  EQU      X:FTOUT+TOTF-TOTC+9
POSOUTF  EQU      X:FTOUT+TOTF-TOTC+10
NEGOUTF  EQU      X:FTOUT+TOTF-TOTC+11

RECT     EQU      1                   ;**set to 0 for half-wave rectifier and
                                      ; full wave rectifier
ADRTBL   EQU      X:$A000             ; start of table for channel addresses
TBL      EQU      X:$E                ;table length

; Filters coeficient files - load in y internal memory
         ORG      Y:0
COEF
         INCLUDE 'LP12.ASM'
         INCLUDE 'BP1.ASM'
         INCLUDE 'BP2.ASM'
         INCLUDE 'BP3.ASM'
         INCLUDE 'BP4.ASM'
         INCLUDE 'BP5.ASM'
         INCLUDE 'BP6.ASM'
         INCLUDE 'LP2.ASM'
         INCLUDE 'LP2.ASM'
         INCLUDE 'LP2.ASM'
         INCLUDE 'LP2.ASM'
         INCLUDE 'LP2.ASM'
         INCLUDE 'LP2.ASM'
```

Listing 2. cont.