B X

Second Quarterly Progress Report

December 27, 1985 through March 27, 1986

NIH Contract N01-NS-5-2396

Speech Processors for Auditory Prostheses

Prepared by

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

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 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. Work in the present quarter included the following:

1. Tasks related to our second implant operation at Duke University Medical Center (DUMC) on February 22;

2. Subsequent testing of this implant patient, who was fitted with a percutaneous cable for extensive psychophysical and speech-perception studies;

3. Development and testing of new computer programs to support and extend such studies; and

4. Presentation of project results at the <u>9th Annual Meeting of the Association for Research in Otolaryngology</u>.

In this report we will describe the initial tests of speech perception conducted with this second patient (MH) at DUMC. In all, we have evaluated 13 major classes of speech-processing strategies, with an average of 4 variations per class for optimization of within-class parameters. The results to date indicate substantial and significant differences in performance with different speech-processing strategies; the best strategies provide world-record levels of performance on confusion-matrix material while other strategies provide only poor-to-moderate performance with the same tests. Detailed description of the psychophysical tests is deferred for now, but will be presented in a future report.

II. Description of the Patient

MH lost her hearing at age 25 to otosclerosis. She was 51 at the time of her first implant operation, to install her electrode array and percutaneous cable. When the dissection for opening the cochlea was completed, we discovered that the basal-most 4-5 mm of the scala tympani was obliterated with bone. This bone had to be drilled down for insertion of the electrode array. Therefore, the two basal-most pairs of electrodes were probably more distant from the target neural tissue than in other patients implanted with the UCSF/Storz electrode array. Once drilled, the bone did not further impede the insertion of the electrode array. The array was inserted to a depth of approximately 25 mm, and subsequent X-ray studies of the sinus cavities (at a later date) demonstrated that the implant followed the spiral course of the scala tympani and was in perfect position.

After the electrode array was inserted the electrical impedances of each electrode pair and each single electrode were measured. All impedances fell within the normal ranges for bipolar and monopolar configurations except for monopolar electrode 10 and for bipolar electrode pair 9-10. The "mushroom cap" of electrode 10 came off during manipulation of the array for insertion, so only the stub beneath the cap was available for passage of current in the impedance test. Because the decrease in surface area would greatly increase the current and charge densities for use of electrode 10 in normal applications of the implant, this electrode was never used in subsequent psychophysical and speech-testing studies (pair 9-10 is in the middle of the electrode array, where the pair numbers are ordered from apex to base; i.e., pair 1-2 is the apical-most pair and pair 15-16 is the basal-most pair). Finally, recovery from the surgery was uneventful and testing with the percutaneous cable was begun after a two-week period for healing.

III. Outline of Tested Speech Processors and Brief Overview of Principal Findings

An intensive series of tests was begun with patient MH in early March, 1986. A battery of psychophysical tests was first conducted to measure (a) thresholds for various stimuli and electrode-coupling configurations; (b) maximum comfortable loudnesses (MCLs) for most of the stimuli and coupling configurations used in the threshold tests; (c) the extent and nature of interactions between channels, including temporal channel interactions; (d) temporal integration for most of the bipolar pairs in the electrode array; (e) compensation of temporal integration and temporal channel interactions with "inverse-filtered" stimuli; (f) channel discrimination; (g) frequency discrimination; and (h) loudness and loudness matching for various stimuli and electrode-coupling configurations. Briefly, the results from these tests indicated that MH had a poor-to-moderate pattern of nerve survival in her implanted ear. Thresholds to bipolar stimulation were highly heterogeneous across the bipolar pairs of the electrode array, and thresholds to bipolar stimulation were much higher than thresholds to monopolar stimulation. In addition, interactions between most channels were severe, with good isolation found for only one-third of the possible channel combinations. Finally, the dynamic ranges for pulsatile and sinusoidal stimuli were generally narrow, although not as narrow as those found for patient LP at UCSF (see QPR 7, NIH project N01-NS-2356).

Despite these negative findings, however, patient MH was able to rank her channels of bipolar-pair stimulation along a tonotopic order from base to apex, and she had frequency difference limens that fell within the normal ranges for cochlear-implant patients. Also, her recollection of auditory sensations was excellent. She could describe her electrically-evoked auditory percepts in close analogy with everyday sounds she remembered. In all, then, the results of the preliminary psychophysical tests with MH indicated a generally poor (but heterogeneous) pattern of peripheral nerve survival coupled with an intact central auditory system. Although these findings were far from ideal, the psychophysical results for MH were more favorable than those of the last two patients we had studied. Specifically, patient LP had extremely-narrow dynamic ranges and other manifestations of very sparse nerve survival, and patient SG had poor nerve survival along with an apparently-compromised central auditory system (see QPR 1, NIH project N01-NS-5-2396). Complete expositions of the psychophysical tests conducted with these three patients (and one previous patient we studied at UCSF) will be presented in future quarterly reports.

Evaluation of alternative speech-processing strategies was begun in mid-March, immediately after the preliminary psychophysical studies. In all, 13 major classes of processing strategies were tested, with an average of 4 processors in each class for optimization of within-class parameters. A complete log of all tested strategies and results is presented in Appendix 2. In the remainder of this section we will present our method for comparing the performance of different processing strategies, describe the major classes of processing strategies tested to date, and review the major findings.

The performance of each processing strategy was measured with confusion-matrix tests. The confusion matrix for vowels included the tokens "BOAT," "BEET," "BOUGHT," "BIT," and "BOOT," and the confusion matrix for consonants included the nonsense tokens "ATA," "ADA," "AKA," "ASA," "AZA," "ANA," "ALA," and "ATHA." All processing strategies were implemented as computer simulations with the software of our "block-diagram compiler" (see QPR 4, NIH project N01-NS-2356). The processed speech tokens were presented to the patient through our hardware interface (QPR 2, NIH project N01-NS-2356), which provided an electrically-safe means of transmitting stimuli to the electrode array at bandwidths of up to 20 kHz on all eight channels. The presentation of each processed token was accompanied by a display of response options on a computer console used by the patient. When she responded, her response was used to update a matrix display on the investigator's computer console (not seen by the patient), and the next token for presentation was drawn from a randomized list. At the end of the test we would usually give the patient her score of overall % correct and an indication of the principal confusions she made during the test. No feedback was given during the test itself.

Four tests were given for each processing strategy: vowel recognition with lipreading; vowel recognition without lipreading; consonant recognition with lipreading; and consonant recognition without lipreading. Lipreading information was provided by miming tokens in synchrony with stimulus presentations. Only one of us (CCF) presented lipreading information for all tests. Finally, presentations of processed tokens were usually repeated at regular intervals until the patient responded. Although there is some evidence from other groups that repetition of test tokens can increase scores (particularly for tests not using nonsense tokens, such as tests of spondee recognition), we did not find statistically-significant differences in the scores of several tests of consonant recognition for single- and multiple-trial conditions. Unless otherwise indicated, the data presented in the remainder of this section were obtained from tests using repeated presentations.

The main findings of our initial tests of speech perception are summarized in Table 1. The scores tabulated are those for the processor with the highest overall %-correct figure for all processor variations within each major processor class. Before describing the rationale and results for each processor class, we will note a few general features of the data in Table 1. First, high scores are consistently found for the tests of vowel perception with lipreading. These scores are consistent with the fact that MH did well in a single test we administered to measure her performance with lipreading alone; she got 23/25 correct, a score that is not significantly different from most of the scores listed in Table 1 for vowel recognition with lipreading. Therefore, the scores for this test condition are not a particularly-sensitive indicator of processor performance, and MH "tops out" on vowels with lipreading.

Next, we note that scores on the tests of consonant recognition with lipreading are a <u>much</u> more sensitive indicator of processor performance. In two tests of consonant recognition with lipreading only (administered on different dates), MH got 14/24 and 11/24 correct. Scores even slightly greater than 14/24 correct in Table 1 for consonant recognition with lipreading indicate significant improvements over lipreading alone (e.g., the score of 16/24 correct for the last processor listed on the first page of Table 1 is significantly higher than the best score of the lipreading-only tests).

Table 1

Best Results from Various Classes of Speech Processors, Patient MH

Processor class	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips	total % correct
Compressed analog outputs, 4 channels	20/25	8/25	13/24	7/24	49
Breeuwer/Plomp Processor, 2 channels	22/25	10/25	18/24	9/24	60
Interleaved-pulses processor, 2 of 6 channels on at a time	24/25	22/25	14/24	9/24	70
same as above, but with analog "base-band" signal	24/25	16/25	16/24	5/24	62
Interleaved-pulses processor, all 6 channels on at a time	24/25	23/25	20/24	14/24	83
4-channel, interleaved-pulses processor, all 4 channels on at a time	21/25	12/25	17/24	10/24	61
Interleaved-pulses processor, all 6 channels on at a time, with "noise-biasing" signal delivered to 9/Ref	23/25	18/25	18/24	10/24	70
Interleaved-pulses processor, all 6 channels on at a time, with baseband channel of inter-leaved pitch pulses, Ch 2 off	20/25	10/25	16/24	8/24	55

*Chance is 5/25, with lipreading alone she got [illegible]/25

**Chance is 3/24, with lipreading alone, she got [illegible]/24

Best Results from Various Classes of Speech Processors, Patient MH, continued, p.2

Processor class	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips	% correct
Interleaved-pulses processor, with round-robin updates proceeding from base to apex	25/25	24/25	21/24	12/24	84
same as above, except that update cycles are timed to begin at pitch periods for voiced-speech sounds and at the max rate for unvoiced-speech sounds	24/25	20/25	20/24	12/24	78
4-channel version of above			17/24	14/24	N/A
Interleaved-pulses processor, all 6 channels on at a time, with round-robin updates proceeding from base to apex. In addition, the update cycles are timed to begin at pitch periods for voiced-speech sounds and at random intervals ("jittered") for unvoiced-speech sounds.	24/25	16/25	22/24	19/24	83
4-channel version of above	22/25	18/25	19/24	11/24	71

*Chance is 5/25
**Chance is 3/24

Third, the chance levels of performance are 5/25 "correct" for the vowel tests and 3/24 "correct" for the consonant tests. All but one of the scores shown on Table 1 are significantly above chance. The exception is the score of 5/24 for consonant recognition without lips, for the fourth class of processing strategies listed in the table.

Finally, we note that test/retest reliability was good for MH. When we retested a processor that produced low scores on a previous occasion MH would always again obtain low scores, and when we retested a processor that produced high scores on a previous occasion MH would always again obtain high scores. Also, MH's anecdotal remarks were stable across repeated tests of a single processor. When a "good" processor was retested MH would immediately identify it as such, usually in terms like "this is a good processor," "this processor sounds natural and like speech I remember," "this processor doesn't sound simulated," or "this processor is very clear." In contrast, a retest of a processor that produced low scores on a previous occasion would elicit comments like "this is a lousy processor," "this processor sounds like a man in a barrel," "the speaker sounds like he is talking through the telephone with a handkerchief or towel over the mouthpiece," or "this processor is not as clear as some you have tried." MH's anecdotal remarks were always consistent with her test scores on confusion-matrix material.

Perhaps the most-striking feature of the data in Table 1 is the large differences in performance found across processing strategies. To our knowledge, these data and the results obtained with our last patient at UCSF (patient LP, QPR 7, NIH project N01-NS-2356) are the first demonstrations of such differences in individual implant patients. That is, for the first time many fundamentally-different processing strategies have been evaluated in tests with controls for (a) differences in neural survival patterns among patients; (b) differences in cognitive ability among patients; and (c) differences in testing procedures among laboratories. In the two previous studies we know about in which such controls were implemented, only 3 (Eddington, 1980) or 5 (Simmons _et al_., 1986) strategies were evaluated, and only one variation was tested for each strategy.

We will now review the rationale and results for each class of speech processor listed in Table 1.

Compressed analog outputs, 4 channels

The first class of processor listed in Table 1 is the compressed analog outputs strategy, which emulates in software all the functions of the present UCSF/Storz speech processor. All tested variations of this processor used four output channels. The variations were in coupling configuration (including delivery to the best-isolated radial bipolar pairs in the electrode array, staggered pairs in the electrode array, and the apical four pairs in the electrode array), compression ratio (ratios of 2:1 and 3:1 were tested) and "tuning" procedure. The differences in tuning procedure were in the way the gains of the individual channel outputs were adjusted before the tests of vowel and consonant recognition. In the "RTI tuning procedure" the gains of individual channel outputs were not adjusted (i.e., the output level of each bandpass channel reflected the energy in that band). The overall level of the outputs was adjusted with a "master gain" control on the stimulus-isolation unit to provide moderate loudnesses for the tokens of the vowel confusion matrix. In the "UCSF tuning procedure" speech sounds were used to adjust the individual channel gains until all speech features of a small set of features were clearly audible. For example, the gain of the basal-most channel (Channel 4) would be increased in this procedure until an "s" sound was audible.

The best results for this class of processor were obtained with the use of staggered pairs in the electrode array, a compression ratio of 3:1, and the UCSF tuning procedure. In general these manipulations improved the consonant-recognition scores and slightly degraded the vowel-recognition scores obtained with other variations. Inasmuch as consonant recognition is more important for understanding connected speech than is vowel recognition, this configuration of the processor might be considered to be most-appropriate for daily use.

Clearly, the scores for this processor are low. The scores for tests with lipreading plus speech processor are about the same as the scores obtained for lipreading alone; the scores for the conditions without lipreading are among the lowest listed in Table 1. Although the "without lipreading" scores are significantly higher than chance, they are also consistent with a poor overall result for patient MH compared to other patients using the UCSF/Storz speech processor. That is, many patients in the UCSF series have much higher scores than MH on similar tests of vowel and consonant recognition (see, e.g., Schindler et al., 1986). The low scores obtained for MH probably reflect the relatively-poor status of her implanted ear. Specifically, her interactions between simultaneously-stimulated channels are generally severe, and her thresholds for bipolar-pair stimulation are higher than those for the great majority of other patients in the UCSF series.

Breeuwer/Plomp processor, 2 channels

Evaluation of the second processor listed in Table 1, the "Breeuwer/Plomp processor," was inspired by our desire to (a) reduce the spatial and temporal bandwidths of information presented to the electrode array and (b) increase MH's aided lipreading scores, particularly for consonants. The design of the processing strategy was based on the stunning results recently reported by Breeuwer and Plomp for lipreading supplemented with simple representations of the acoustic speech signal (Breeuwer and Plomp, 1984). These investigators measured the number of correctly perceived syllables in short Dutch sentences presented to 18 listeners with normal hearing. The test conditions included lipreading only, lipreading plus acoustic supplement, and acoustic supplement alone. The acoustic supplements included representations of the sound-pressure levels in one or two frequency bands of the speech signal. The bands were centered at 500, 1600 or 3160 Hz, and their widths were either 1 or 1/3 octave. The acoustic supplement signal was derived by sensing the RMS energy in the output of each bandpass filter and then amplitude modulating a sinusoid at the center frequency of the filter with the RMS level. For the two-filter conditions, the two amplitude-modulated sinusoids were summed to provide a combined output signal.

The best results reported by Breeuwer and Plomp are nothing short of astounding. They found that when lipreading was supplemented with two 1-octave bands at 500 and 3160 Hz, the number of correctly-perceived syllables jumped from 22.8% correct for lipreading only to 86.7% correct for lipreading plus processed speech supplement. The number of correctly-perceived syllables for the supplement alone was 26.7% correct. The score for lipreading plus acoustic supplement is fully consistent with substantial open-set recognition of speech. Breeuwer and Plomp suggest that the excellent results obtained with such a simple acoustic supplement may be explained by (a) the possibility that the perceived ratio between high-band and low-band energies codes well the voice/unvoice boundaries in connected discourse, which are not visible on the lips, and (b) the possibility that the perceived overall amplitude of the combined filter outputs codes well the temporal dynamics of speech, which also are not visible on the lips.

Our application of the processing strategy described by Breeuwer and Plomp consisted of mapping the outputs of the two filters onto the dynamic ranges of two of MH's channels of radial-pair bipolar stimulation. The tested variations of this basic processing strategy included (a) delivery of "compressed analog outputs" of the two filters (before the RMS-detection circuitry) to two well-isolated pairs in the electrode array (pairs 1-2 and 5-6); (b) delivery of short-duration (0.2 or 0.5 msec), interleaved pulses to these two electrode channels, where the pulse intensities were derived from a logarithmic transformation of the RMS outputs for each channel; (c) delivery of simultaneous pulses to these two channels, where the pulse durations and intensities were the same as those used for variation b above; (d) delivery of the above-described pulsatile stimuli to less-well-isolated pairs in the electrode array (pairs 1-2 and 3-4); and (e) use of increased cutoff frequencies in the RMS smoothing filters to represent voicing and fundamental-frequency information in the "ripples" of RMS outputs. The logarithmic mapping law for deriving pulse intensities was of the form:

$$\text{pulse intensity} = A \times \log(\text{RMS level}) + k,$$

where the parameters "A" and "k" were determined for each channel according to the threshold and MCL for pulses on that channel. This is the same mapping function that was so successfully applied in our studies with patient LP at UCSF (QPR 7, NIH project N01-NS-2356). The use of interleaved pulses was also inspired by (a) our finding with LP that his perception of speech tokens improved tremendously when simultaneous-outputs strategies were abandoned in favor of interleaved-pulses strategies and (b) the results of subsequent psychophysical studies with patients SG and MH at Duke which demonstrated that very substantial release from channel interactions could be obtained with the use of interleaved pulses. Finally, in the last tested variation of the basic Breeuwer/Plomp processor (variation e above), we thought that the additional information of voicing and fundamental frequency might be conveyed if the cutoff frequencies of the RMS-smoothing filters were increased from 30 Hz to 10% of the lower cutoff frequencies of each bandpass filter (i.e., 36 Hz for the low band and 224 Hz for the high band). In this way the "average" RMS levels in each band would still be represented, but the now-present ripples in the RMS outputs (particularly for the high band) would also signal the times of glottal openings in voiced-speech sounds.

The best results for the Breeuwer/Plomp processor were obtained with pulsatile outputs and with the increased cutoff frequencies of the RMS-smoothing filters. The results were about the same for the conditions of simultaneous and non-simultaneous stimulation of well-isolated pairs in the electrode array; the scores were lowered when less-well-isolated pairs were used. In this last set of conditions, scores were lower for simultaneous stimulation. The scores for presentation of compressed analog outputs to the two selected channels were much lower than the scores for any tested variation using pulsatile outputs. Finally, we note that MH found the variation with the increased cutoff frequencies for the RMS-smoothing filters to sound much more "natural and speech like." She could correctly make the male/female distinction with this variation and could not with the other variations (the speaker was a male, and MH thought the other variations sounded like a high-pitched female with a monotone voice).

The scores obtained with the best variation of the Breeuwer/Plomp processing strategy are impressive. These scores are at least somewhat higher in every category when compared with the scores of the compressed analog outputs strategy. The largest increase is in the category of consonants with lipreading, as might be predicted from Breeuwer and Plomp's findings with normal-hearing subjects. All scores are significantly greater than chance, and the scores for the processor plus lipreading conditions indicate that this processor would provide substantial benefit as an adjunct to lipreading. Because the Breeuwer/Plomp processor transmits only two channels of information, it might be especially useful for patients who have very poor patterns of nerve survival (with concomitant channel interactions) or electrode arrays with a low degree of spatial selectivity (e.g., monopolar electrode arrays such as the one used in the Utah/MIT/Symbion device), or both.

Interleaved-pulses processor, 2 of 6 channels on at a time

The next major class of tested processing strategies is interleaved-pulses processors of the type we applied in our studies with patient LP (QPR 7, NIH project N01-NS-2356). As described in detail in our report of LP's case, input speech signals are first high-pass filtered (1st order, with a break frequency of 1200 Hz) in these processors to flatten the speech spectrum and diminish the otherwise overwhelming influence of the first formant (F1). The output of the high-pass filter is then fed to a bank of bandpass filters whose center frequencies span the combined range of F1 and F2, along a logarithmic scale. The RMS energy in each band is sensed by a full-wave rectifier and low-pass filter connected in series to each bandpass filter output. Next, a "post-processor" is programmed to scan the RMS outputs each time a pulse is to be delivered to the electrode array. The output of a filterbank channel is delivered to its assigned electrode(s) only if (a) it is one of the two channels with the greatest RMS energy for the present time frame and (b) the RMS energy is above a preset "noise threshold." Finally, the amplitudes of the pulse(s) delivered to the selected channel(s) in the electrode array are derived with a logarithmic mapping law for each channel, as previously described for the Breeuwer/Plomp processor.

As indicated in Table 1, the scores for the interleaved-pulses processor, 2 of 6 channels on at a time, are at least somewhat higher than the scores for the compressed analog outputs strategy in every category, and substantially higher in the category of vowel recognition without lips. The overall %-correct score is much higher than the overall %-correct scores for the compressed analog outputs processor and the Breeuwer/Plomp processor. MH immediately volunteered that the 6-channel interleaved pulses processor (with 2 channels on at a time) was much clearer and much more understandable than the previous processors we had tested; indeed, she told us that "you brought the man out of the barrel with this processor." The processor still sounded "simulated" to MH, but the speech tokens were much clearer than the tokens heard from the other processors. The score for vowel recognition without lips is nearly perfect and represents superb performance for a cochlear-implant patient. However, the scores for consonant recognition are not much different from those of the compressed analog outputs strategy, and the score for the "consonants with lipreading" category is significantly lower for the interleaved-pulses processor compared with the high score obtained with the Breeuwer/Plomp processor. Therefore, a major focus of our subsequent effort was to evaluate processing strategies designed to retain the excellent performance on vowels while increasing the levels of performance on consonants.

Finally, we note that the tested variations of the present and subsequent interleaved-pulses processors included at least some manipulations of one or more of the following parameters: pulse width; interpulse interval; pulse type (i.e., current- and duration-balanced biphasic pulses or charge-balanced, "monophasic-like" pulses); and method used to produce charge-balanced, "monophasic-like" pulses. The effects of these manipulations for all the interleaved-pulses processors will be described in the last subsection of this major section on speech-processing strategies.

Interleaved-pulses processor, 2 of 6 channels on at a time, with the addition of an analog "baseband" signal

Our first attempt at increasing the scores of consonant tests for interleaved-pulses processors was to add a compressed analog "baseband" signal. The baseband signal was delivered to the apical-most pair in the electrode array, and six channels of interleaved pulses, 2 channels on at a time, were delivered to the remaining six pairs of electrodes in the array (pair 9-10 was not used, as previously noted). In one variation of this hybrid "interleaved-pulses/compressed-analog-baseband-signal" processor, the baseband signal was derived in exactly the same way as the lowest-band signal was derived for the compressed-analog-outputs strategy. Specifically, the input speech signal was first compressed at a compression ratio of 3:1, then passed through a bandpass filter whose corner frequencies were at 200 and 800 Hz, and finally passed through a high-pass "charge filter" whose corner frequency was at 300 Hz. One of the features strongly represented in the outputs of the above-described baseband channel was the periodicities of voiced-speech sounds. We therefore thought that fundamental frequency and voice/unvoice boundaries might be better conveyed with the addition of the baseband signal. Also, Mark White's finding of F1 discrimination for patients using a single-channel, compressed-analog-outputs processor (White, 1983) indicated the possibility that at least some additional F1 information might be perceived by our patient when this particular baseband signal was applied.

In another variation of the hybrid "interleaved-pulses/compressed-analog-baseband-signal" processor the corner frequencies of the bandpass filter were extended from 200-to-800 Hz to 50-to-4000 Hz. This manipulation was expected to increase further the representation of voicing information and perhaps improve the representation of the overall speech envelope.

The results obtained with these two variations of the hybrid processor were quite different. The best overall %-correct score was obtained with the variation using the "wide-band" baseband signal, where the bandpass was from 50 to 4000 Hz. The scores for both variations indicated a generally-destructive effect of adding the baseband signal to an interleaved-pulses processor.

First, for the "wide-band" variation, there were large decreases in the scores for vowel and consonant discrimination without lips. The decrease for vowels was probably attributable to increased interactions between channels. That is, the release from temporal channel interactions attained with the use of interleaved, nonsimultaneous pulses was most likely degraded by the presence of the continuous, "analog" baseband signal. Therefore, the representation of steady-state formant frequencies across independent channels may have been much less salient when the analog baseband signal was presented along with six channels of interleaved pulses.

Next, the decrease in the score for consonants without lipreading was somewhat surprising, but in retrospect consistent with the above hypothesis of exacerbated channel interactions. Note, however, that the score for consonant discrimination <u>with</u> lipreading was slightly higher for the processor in which the baseband signal was used. We suggest that envelope information may have been better represented with the addition of the baseband signal, and that such an improved representation may have provided a useful adjunct to lipreading.

As mentioned above, the scores obtained with the hybrid processor using the "narrow-band" baseband signal were quite different than the scores just reviewed for the hybrid processor using the "wide-band" baseband signal. Specifically, the scores for the "narrow-band" version were: 22/25 for vowels with lipreading; 12/25 for vowels without lipreading; 17/24 for consonants with lipreading; and 8/24 for consonants without lipreading. In all, then, the scores for vowel discrimination were lowered and the scores for consonant discrimination were increased when the bandwidth of the baseband channel was changed from 50-to-4000 Hz to 200-to-800 Hz. These differences in results might be explained in terms of the waveshapes of baseband-channel outputs produced by the two variations of hybrid processors. In particular, the wide-band signal had much higher "peak factors" and therefore was perceived by our patient as much louder than the narrow-band signal when both signals had the same RMS energy (also see Shannon, 1983). To achieve the same loudness levels, then, the overall level of the narrow-band signal was the higher of the two. This higher level may have (a) further exacerbated channel interactions, lowering the scores for vowel recognition and (b) improved somewhat the representation of voicing and envelope information, increasing the scores for consonant recognition. The overall %-correct score for the "narrow-band" version of the hybrid processors was 60%, not much different from the 62% listed in Table 1 for the "wide-band" version. Both scores were significantly below the 70% score obtained for the interleaved-pulses processor without the addition of the baseband signal.

<u>Interleaved-pulses processor, all 6 channels on at a time</u>

The disappointing results for the hybrid processors described above led us to abandon the hybrid approach in favor of methods to improve the representation of speech with non-simultaneous stimuli. Our first step in this direction was to increase the number of channels allowed to deliver interleaved pulses to the electrode array on each update frame. Specifically, pulses were delivered to all six channels of a six-channel processor in a round-robin fashion. The amplitudes of the pulses were derived from a logarithmic transformation of the RMS energies in each channel bandpass, as previously described. Unlike the interleaved-pulses processor with 2 of 6 channels on at a time, the processor with 6 of 6 channels on at a time had no logic to select the two channels with the greatest RMS energies in each time frame. Therefore, the full spectrum of speech over the F1 and F2 range was represented with the "6-of-6-channels" processor.

In addition to manipulations in pulse width, interpulse interval, and method for producing charge-balanced "monophasic" pulses, the tested variations of the "6-of-6-channels" interleaved-pulses processor included two patterns of round-robin updates. The first pattern of channel updates was the channel order: 1, 2, 4, 6, 3 and 5, where channel 1 was the apical-most channel and channel 6 was the basal-most channel; the second pattern of channel updates was the channel order: 6, 5, 4, 3, 2, 1. The first pattern presented stimuli in an order designed to minimize channel interactions (for MH's particular ear) and the second pattern presented stimuli in a temporal order that proceeded from base to apex. The temporal order of the second pattern, of course, mimicked the order of stimulation imposed in normal hearing by the travelling-wave mechanics of the basilar membrane. The travel time for normal hearing was closely approximated by appropriate selections of pulse duration and interpulse interval for the electrical-stimulation case. Specifically, 0.5 msec pulses spaced 0.2 msec apart were used, for a total "travel time" of 4.2 msec from base to apex over the length of the electrode array. Of course, the power-function of increases in delay times for normal hearing was only crudely approximated by the linear increase in delay times implemented in our interleaved-pulses processor.

The results for these two variations of interleaved-pulses processors, with all 6 channels updated on each round-robin cycle, are listed in separate entries in Table 1. The variation with the update order designed to minimize channel interactions is listed as the fifth processor on the first page of Table 1; the variation with the update order designed to mimic delays imposed by the travelling wave in normal hearing is listed as the first processor on the second page of Table 1.

The best results for these two classes of interleaved-pulses processors were obtained with the "natural" order of stimulation from base to apex. MH immediately remarked that this processor was the very best we had tested and that the processed tokens sounded like natural speech she remembered. The scores for the alternate order of channel updates were only somewhat lower, however, and the difference between update orders seemed more qualitative than quantitative.

For both variations of channel updates, the results demonstrate substantial improvements in vowel and consonant recognition over all previously-tested processors. Excellent, world-record levels of performance are retained on vowels while large increases in consonant recognition are realized for both the with-lipreading and without-lipreading conditions. The overall %-correct scores of 83% and 84% are also much higher than the scores of previously-tested processors. Moreover, these results for the interleaved-pulses processors, 6 of 6 channels on at a time, are among the best ever reported for any cochlear-implant patient, using any processing strategy. The overall %-correct score is fully consistent with substantial open-set recognition of speech. As will be described in section V of this report, "Plans for the Next Quarter," we plan to conduct tests of open-set recognition with interleaved-pulses and other processors in the immediate future.

To provide a more-detailed picture of processor performance, confusion matrices from the first tested variation of "6-of-6-channels" interleaved-pulses processors are presented on the next two pages (this variation of the processor used the channel update order designed to minimize channel interactions). On the vowel test with lipreading MH once perceived "bit" as "bought," and on the vowel test without lipreading she once heard "boot" as "bought" and once heard "boat" as "bit." Results with the consonant tests were also excellent; she got 20/24 correct with lipreading and 14/24 correct without lipreading. Most confusions in the "with lipreading" condition were in cells adjacent to the diagonal, indicating that most errors were made with phonetically-similar tokens. In particular, MH once perceived "ATA" as "ADA," once perceived "AZA" as "ASA," and once perceived "ALA" as "ANA." This is <u>stellar</u> performance for a cochlear-implant patient. Finally, she had some of the same confusions for the "without lipreading" condition plus 4 confusions between "ASA" and "ATHA." These last two tokens have a high degree of acoustic similarity; indeed, we (CCF and BSW) have a hard time hearing the difference between them in our recorded presentations.

To conclude this subsection on "6-of-6-channels" interleaved-pulses processors, we note that we were initially surprised at the large increases in performance obtained when we went from the "2-channels-on-at-a-time" processor to the "6-channels-on-at-a-time" processor. Before testing the "6-channels-on-at-a-time" processor we were concerned that the picture of

Interleaved Pulses Strategy,
all 6 channels on at a time,
Patient MH

With Lipreading:

RESPONSE

STIM	beet	bought	boat	boot	bit
beet	5				
bought		5			
boat			5		
boot				5	
bit		1			4

24/25

Without Lipreading:

RESPONSE

STIM	beet	bought	boat	boot	bit
beet	5				
bought		5			
boat			4		1
boot		1		4	
bit					5

23/25

Interleaved Pulses Strategy,
all 6 channels on at a time,
Patient MH

With Lipreading:

RESPONSE

STIM	t	d	k	s	z	n	l	th
t	2	1						
d		2				1		
k			3					
s				3				
z				1	2			
n						3		
l						1	2	
th								3

20/24

Without Lipreading:

RESPONSE

STIM	t	d	k	s	z	n	l	th
t	2	1						
d		2				1		
k	1		2					
s				1				2
z					2			1
n					1	2		
l						1	2	
th				2				1

14/24

channel interactions would be very complex with such a "dense" representation of the speech signal. A possible advantage of the "2-channels-on-at-a-time" processor was that the time between sequential pulses could be a millisecond or greater while still keeping the stimulus frequency on any single channel at 250 Hz or higher. This was important because MH could hear the "modulation whistle" of lower stimulus frequencies and would mistakenly identify a male speaker as a woman or child with a "high-pitched, monotonous voice." Also, we wanted to keep the cycle update times across channels at or below 5 msec because significant changes in speech can occur in intervals as short as 5-10 msec. The "2-channels-on-at-a-time" processor met the above timing criteria with times between pulses that we knew (from psychophysical measurements) would provide good isolation between the stimulated channels and thereby preserve the representation of RMS energies in the two selected channels. In contrast, the time between pulses had to be greatly reduced (e.g., from 1.5 msec to 0.2 msec) for the "6-channels-on-at-a-time" processor in order to keep the cycle update times at or below 5.0 msec. Because the psychophysical studies of temporal channel interactions demonstrated that the release from interactions afforded by the use of nonsimultaneous stimuli was not as great at short separations (e.g., 0.2 msec) compared to long separations (e.g., 1.0 msec or greater), we expected that increased channel interactions might degrade the performance of interleaved-pulses processors as the number of channels in each update cycle was increased.

Apparently, from the results presented at the beginning of this subsection, the improved representation of RMS energies across all six bands far outweighs possible effects of increased channel interactions. Performance in tests of vowel recognition is as good or better with the "6-channels-on" processor compared with that of the "2-channels-on" processor, and performance in tests of consonant recognition is <u>much</u> better for the "6-channels-on" processor. The improvement in consonant recognition may have resulted from an improved representation of the complex spectra and temporal dynamics of consonants. That is, vowels have "steady-state" regions in which the frequencies of 2 or 3 formants adequately specify phonemic identity. In constrast, consonants are not so well-specified by a formant model and also generally have far-greater rates of temporal changes within each phonemic unit. Consonants are instead specified by a host of features including (a) voicing/frication; (b) amplitude envelope; (c) loci and shapes of broad spectral peaks (i.e., the spectra of most consonants do not have a well-defined "formant" structure); and (d) rapid formant transitions from a leading vowel into a following consonant or from a leading consonant into a following vowel. These features other than steady-state formants are probably best represented with rapid updates of information on all channels of a multichannel array. Therefore, a "low-density" representation such as that provided by our "2-channels-on-at-a-time" interleaved-pulses processor, or by the "1-or-2-channels-on-at-a-time" processor used in the Melbourne/Nucleus device, may be adequate for transmitting information on vowels but inadequate for transmitting complete information on consonants. Finally, we note again that the "6-channels-on" processor sounded much more natural and speech-like to MH compared to all other tested processors. This last improvement might also have been a result of the increased temporal and spectral resolution probably provided by the "6-channels-on" processor.

<u>4-channel, interleaved-pulses processor,</u>
<u>all 4 channels on at a time</u>

Encouraged by our success with the 6-channel processor described in the previous subsection, we decided to evaluate the effects of reducing the number of interleaved-pulses channels to 4. The represented bandpasses were identical to those used in the simulations of the compressed analog outputs strategy (the first strategy listed in Table 1). As is evident from the results presented in Table 1, performance with the 4-channel, interleaved-pulses processor was better than the performance obtained with the compressed-analog-outputs strategy, but substantially worse than the performance obtained with the 6-channel, interleaved-pulses processor. The precipitous decline in all scores when the number of channels was reduced from 6 to 4 certainly indicates the importance of the number of channels in a multichannel auditory prosthesis. Excellent performance was obtained with the 6-channel, interleaved-pulses processor while poor-to-moderate performance was obtained with the 4-channel, interleaved-pulses processor. The differences in all categories of vowel and consonant recognition were surprisingly large, and indicate the need for caution when comparing the performance of single-channel and multichannel auditory prostheses. Specifically, the advantages of multichannel devices may be far more evident when the number of channels used is at or above some critical minimum, and when a good processing strategy is applied for the multichannel device.

An additional comparison of interest in regard to the results obtained with the 4-channel, interleaved-pulses processor is that these results were not much better than the results obtained with the 4-channel, compressed-analog-outputs strategy. Possibly, the isolation between channels for the latter strategy might have been moderately good inasmuch as alternate pairs of electrodes were used for the stimulation channels. In such a case one could reasonably expect that the performance of the compressed-analog-outputs strategy would be comparable to the performance of the 4-channel, interleaved-pulses processor (where isolation is good). More likely, however, is the possibility that different kinds of information were provided by the two processors and that each representation was adequate for limited recognition of vowels and consonants. For example, a better representation of voicing, voice/unvoice boundaries and amplitude envelope may have been conveyed in the baseband of the compressed-analog-outputs processor, while a better representation of frequency information above the baseband frequencies may have been conveyed with the interleaved-pulses processor (through reduced channel interactions). These or other factors could have "traded-off" to produce similar results on the vowel and consonant tests. Support for this notion comes from the fact that these two processors sounded quite different to MH: the compressed-analog-outputs processor produced tokens that MH described as unclear or "like a man talking in a barrel," and the 4-channel, interleaved-pulses processor produced tokens that sounded "clear but simulated." Also, MH often made mistakes in distinguishing a male voice from a female voice with the interleaved-pulses processor, while she rarely made such mistakes with the compressed-analog-outputs processor. These differences in percepts strongly suggest that the representations of speech information at the nerve were different for the two processors.

6-channel, interleaved-pulses processor,
with the addition of a "noise-biasing" signal

Our next attempt at improving the performance of 6-channel, interleaved-pulses processors was to add a "noise-biasing" signal. This noise-biasing signal was delivered to monopolar electrode pair 9/Ref, and consisted of a 200-6000 Hz bandpass of white noise. The idea was to produce a degree of stochastic independence between adjacent neurons in the excitation fields of bipolar-pair electrodes by imparting different discharge histories to different neurons. Specifically, we expected that even slight differences in anatomical and physiological properties of cochlear neurons would cause these neurons to respond at least somewhat differently to a near-threshold noise stimulus, so long as populations of neurons did not "phase-lock" to low-frequency components in the noise. Therefore, frequency components below 200 Hz were eliminated, and the filtered noise signal was delivered in a monopolar configuration to ensure a pervasive (and nearly uniform) spread of effect over the length of the electrode array. If our assumptions underlying the generation of stochastically-independent, "spontaneous-like" activity were correct, then responses evoked by deterministic pulses would be modified by the ongoing activity in the nerve, produced by the noise-biasing signal. The deterministic stimuli were delivered to bipolar pairs of the electrode array in the manner previously described for the "6-channels-on," interleaved-pulses processor.

The hypothesis we wished to evaluate with the addition of the noise-biasing signal was the suggestion that stochastic independence between adjacent neurons allows a greater bandwidth of temporal information to be represented in the ensemble responses of the auditory nerve (this hypothesis is fully described in our 8th QPR for NIH project N01-NS-2356). Briefly, if neurons in the excitation field have different and stochastically-independent discharge histories, then the burden of information transfer might be shared between previously-stimulated neurons still in their refractory periods and fully-recovered neurons, now ready to respond to the present stimulus. Because not every neuron in the excitation field is stimulated on every major peak in the stimulus waveform (or, in the present case, on every pulse), higher frequencies of stimulation might be represented to the central auditory system in a "time-shared" arrangement. That is, responses of discrete subpopulations of neurons could signal the occurrences of major events in the stimulus waveform at higher rates in the ensemble response than could identical responses throghout the total population of neurons. This, of course, is similar to the main argument made in Wever's volley theory of pitch perception for normal hearing, where he suggests that frequencies as high as 3 to 4 kHz can be represented in the ensemble response even though individual auditory neurons can only respond to acoustic stimuli at rates of up to about 200 spikes per second.

As indicated in the scores for the seventh processor class listed in Table 1, the effects of adding the noise-biasing signal were generally negative. We suggest that the lowered scores may have resulted from one or more of the following:

1. Channel interactions may have been exacerbated by the addition of a simultaneous "background" stimulus, as with the processor in which an analog "baseband" signal was added;

2. The central auditory system may not be able to "decode" the high-frequency information represented in the ensemble response; and

3. Stochastic independence among adjacent neurons was not achieved with our stimulus scheme.

Among these possibilities, we consider the first and second to be most likely. Certainly, the idea of increased channel interactions is supported by the disappointing results obtained with the 6-channel, interleaved-pulses processor using an analog "baseband" signal. Although the scores are higher for the "noise-biasing" processor, these scores also reflect possible improvements in performance afforded by the use of the "6-of-6-channels," interleaved-pulses processor instead of the "2-of-6-channels," inteleaved-pulses processor. In any case, the scores drop for the "2-of-6-channels" processor when the analog baseband signal is added and the scores also drop for the "6-of-6-channels" processor when the noise-biasing signal is added. One explanation for these findings is that the frequency representation across channels is significantly degraded when any simultaneous, continuous signal is added to the interleaved pulses presented by the unaltered processor.

Next, we note that the results of recent experiments conducted by Burns and Viemeister (1981) support the second possibility listed above. Specifically, these investigators asked their normal-hearing subjects to indicate when they heard changes in pitch when the modulation frequency of sinusoidally-amplitude-modulated (SAM) noise was manipulated. If the normal auditory system could decode the "time-shared" temporal representation of SAM noise, then one would expect that these subjects would exhibit good frequency discrimination up to 3 or 4 kHz; otherwise the frequency limit of good discrimination might approximate the maximum discharge rate of single fibers in the auditory nerve (i.e., around 200 to 300 Hz). In fact, Burns and Viemeister find a "pitch saturation" for their subjects at around 300 Hz. This limit is essentially the same as that found for cochlear-implant patients, and supports the conclusion that the central auditory system cannot decode a "time-shared" temporal representation in the ensemble activity of the auditory nerve.

Finally, we note that our method for producing stochastic independence among cochlear neurons was only a "best guess" first approximation. We plan modeling studies in the near future to evaluate the efficacy of this approach and possibly to refine the parameters of the noise-biasing signal to achieve higher levels of stochastic independence.

Interleaved-pulses processors that have explicit representations of fundamental frequency and voice/unvoice boundaries

The final set of processing strategies evaluated in this reporting period involved explicit representations of fundamental frequency (F0) and voice/unvoice (v/uv) boundaries. These voicing parameters were extracted from each of the tokens of the vowel- and consonant-confusion tests with a semi-automated procedure in which the following were sensed: (a) peaks in the speech waveform above a preset "noise deadband; (b) the areas under these peaks from deadband crossing to deadband crossing; (c) the polarity of each peak; and (d) running estimates of wide-band (up to 5000 Hz), voice-band (40 - 350 Hz), low-band (364 - 707 Hz), and high-band (2235 - 4470 Hz) RMS energies. A "pitch pulse" was detected for a given peak if the peak was of the preselected polarity and if preselected thresholds for the area under the peak, absolute magnitude of the peak, the ratio of the absolute peak magnitude to the voice-band RMS energy, the wide-band RMS energy, voice-band RMS energy, and low-band RMS energy were all exceeded. An unvoiced interval was detected if the wide-band RMS energy exceeded a preselected threshold and if the ratio of high-band-to-low-band RMS energies exceeded another preselected threshold. Silence intervals were detected for all segments of the digitized tokens that had wide-band RMS energies below a preselected threshold.

The detected parameters of the automated part of the procedure described above were written to the disk for subsequent editing. In the editing part of the procedure we examined every detected pitch pulse along with the input waveform. All erroneous indications of pitch pulses were then deleted from the "F0 file," and the revised information was stored on the disk. The marking of v/uv boundaries and of silence intervals was also examined for accuracy and modified where appropriate. The edited files then contained essentially perfect extractions of the times of all individual pitch periods and of v/uv boundaries for all tokens in our consonant and vowel tests. Therefore, the performance for each processor class described below indicates results for an idealized extraction of F0 and v/uv boundaries. In general, accurate extraction of these parameters in a real-time processor (especially for F0) is a formidable task, and our intent here was to evalute the <u>potential</u> for explicit representations of voicing information.

In the first tested class of processors that provided explicit representations of voicing information, the times of pitch periods were indicated by presentations of pulses on a "baseband" channel. These pulses were interleaved with the pulses presented on higher channels, as previously described. The total set of stimuli sent to the electrode array consisted of round-robin updates of interleaved pulses to the upper channels and single pulses to the apical-most, "baseband" channel when a pitch period was to be signalled. In this way we hoped to retain the excellent levels of performance previously obtained with the "all-channels-on," interleaved-pulses processors while adding an explicit representation of voicing information. We further hoped (and expected) that the additional voicing information would help our patient make voice/unvoice distinctions for consonants (e.g., improve her ability to distinguish "ASA" from "AZA"). Finally, we also expected that the additional voicing information would improve the "naturalness" of the percepts and the ability to make man/woman/child distinctions.

Unfortunately, as indicated in the entries for the last processor class on page 1 of Table 1, the addition of explicit voicing information in the manner just described lowered the scores in every category of consonant and vowel recognition compared with the "6-of-6-channels," interleaved-pulses processor. One possible explantion for the lowered scores is that we had to turn off the lowest channel of the six channels of "high-band" interleaved pulses because MH reported a mild pain sensation deep in her implanted ear when relatively-intense stimuli were delivered to this channel (electrode pair 3-4). Therefore, RMS energies in the band from 434 to 712 Hz were not represented in the aggregate outputs from the processor. Inasmuch as this band can signal the location of F1 and the presence of voicing (from the ratio of high-band-to-low-band energies), it may be that the absence of this information more than offset any contributions that might have been provided in the baseband "pitch channel."

Another possible explanation for decreased performance may be in the way F0 information was represented. The interleaving of "pitch pulses" on the basband channel of course disrupted the regular timing of updates on the upper channels. This disruption might have manifested itself as a degraded representation of band energies in the channels above the baseband channel. In particular, the insertion of a pitch pulse within the normal sequence of round-robin updates could have changed the amount of temporal integration at neural membranes over the update cycles in an irregular way. This irregularity, in turn, might have produced distortions in the representations of band energies for the upper, "5-of-5," interleaved-pulses channels.

Although the results with the first class of processors using explicit representations of voicing were generally discouraging, MH did remark that the tokens produced by these processors sounded "very natural" and that she heard clearly the "deep male voice" of the speaker. Thus, some additional voicing information seemed to be making its way into MH's percepts. We therefore decided to pursue other ways in which explicit representations of F0 and v/uv boundaries might be better integrated into the basic "6-of-6-channels," inteleaved-pulses processor.

Our first step in this direction was to abandon the concept of a baseband channel in favor of a new procedure in which the cycles of round-robin updates were timed to begin at pitch periods for voiced-speech sounds. In this way all six channels of the basic interleaved-pulses processor would be updated in strict sequence at every pitch period. Further, no stimuli were delivered in the interval between the completion of the update cycle for the present pitch period and the start of the update cycle for the subsequent pitch period. Because a substantial amount of time elapsed in the "no-stimulus" interval (typically, this time was between 3 and 7 msec for the pulse conditions and speech test tokens used), the onsets of channel update cycles were clearly periodic, and this periodicity reflected exactly the periodicity of F0 for voiced speech.

Next, for unvoiced speech sounds, the timing of channel update cycles was controlled in one of two ways. In the first procedure updates were made at the maximum rate (i.e., with only the interpulse time between sequential update cycles), as in the unaltered "6-of-6-channels" processor. We expected that the high density of stimulation produced by maximum-rate update cycles might maintain the excellent performance of the unaltered interleaved-pulses processor. In particular, we thought that maximum-rate updates would keep the unvoiced consonants at clearly-audible levels and would convey the complex temporal dynamics of these phonemes.

The other procedure for controlling channel updates during unvoiced segments was to begin the update cycles at random intervals, where the minimum interval was equal to the interval for maximum-rate updates. Thus, for most sequential cycles of channel updates there was a variable "dead time" between the cycles. We thought this "jittered" representation of unvoiced intervals might produce a noiselike percept, and thereby improve distinctions between voiced and unvoiced sounds. Instantaneous frequencies for the jittered updates ranged from 50 to 300 Hz.

Results for the processors just described are listed as the last four classes of processors in Table 1 (see page 2). The major classes are (a) the 6-channel, interleaved-pulses processor in which the update cycles are timed to begin at pitch periods for voiced-speech sounds and at the maximum rate for unvoiced-speech sounds; (b) a 4-channel version of this processor; (c) the 6-channel, interleaved-pulses processor in which the update cycles are timed to begin at pitch periods for voiced-speech sounds and at random intervals ("jittered") for unvoiced-speech sounds; and (d) a 4-channel version of this processor. The results for the first of these processors are similar to the results obtained for the unaltered, "6-of-6-channels," interleaved-pulses processor (the first processor listed on page 2 of Table 1). The only significant difference is a slight degradation in vowel recognition without lips for the processor in which voicing information was explicitly represented. Possibly, the decreased density of channel updates during voiced segments may have reduced the amount of transmitted information on the spectral content of the vowels.

Although the scores for the processor in which voicing information was explicitly represented were somewhat lower than the scores for the unaltered interleaved-pulses processor, MH preferred the former. She said the processor with explicit coding of F0 and v/uv boundaries sounded more like a "natural human voice" to her, and that she could easily perceive the sex of the speaker when this processor was used.

As mentioned above, the next tested class of processors using explicit coding of voicing information was a 4-channel version of the 6-channel processor just described. Because of a scheduling problem, only the consonant tests were conducted with the 4-channel processor. The overall scores on these tests were not much different from the overall scores obtained with the 6-channel processor. However, the scores on consonant tests were clearly improved for 4-channel, interleaved-pulses processors by the addition of explicit coding of voicing information (i.e., the scores for consonant recognition with and without lips were 17/24 and 10/24 for the unaltered 4-channel processor, and 17/24 and 14/24 for the 4-channel processor with explicit coding of voicing information). In all, these findings suggest that explicit coding of voicing information may have a larger beneficial effect for 4-channel processors than for 6-channel processors.

The third tested class of interleaved-pulses processors with explicit coding of voicing information was the 6-channel processor with jittered cycle updates for unvoiced intervals. As shown in Table 1, this processor produced phenomenally-good results on tests of consonant recognition. The scores were 22/24 for consonant recognition with lips and 19/24 for consonant recognition without lips. Apparently, the jittered representation helped MH make v/uv and other distinctions that greatly improved her scores on the consonant tests. She also remarked that the tokens produced by this processor were "very clear," "natural," and "like speech I remember." Finally, the consonants in all tokens of the consonant-confusion matrix were clearly audible. This finding effectively eliminated our concern that maximum-rate updates might be required to make unvoiced consonants loud enough to identify.

Unfortunately, the outstanding performance in recognition of consonants was not matched by outstanding performance in recognition of vowels. In particular, the score of 16/25 for recognition of vowels without lipreading was substantially lower than the scores obtained with 6-channel, interleaved-pulses processors without explicit coding of voicing information (23/25 and 24/25) and somewhat lower than the score obtained with the other 6-channel processor that did use explicit coding of voicing information (20/25). Although we have no ready explanation for the difference in the scores for the two 6-channel processors that used explicit coding of voicing information, we believe the differences between the scores for the processors with and without explicit coding of voicing information may reflect differences in the "densities" of channel updates during voiced-speech segments. As mentioned before, the lower densities used in the processors with explict coding of voicing information may degrade the representation of RMS energies across the interleaved-pulses channels.

The final tested class of processor with explict coding of voicing information was the 4-channel version of the previously-described, 6-channel processor with "jittered" updates during unvoiced intervals. This 4-channel processor had the best performance by far of all 4-channel processors listed in Table 1. The improvement in consonant recognition may have resulted from a superior representation of v/uv boundaries, as mentioned before, and the improvement in vowel recognition may have resulted from a superior representation of F0. Vowel recognition might have been improved by superior coding of F0 because differences in F0 and F1 can reliably indicate vowel height (high vowels like /i/, /I/, /U/ and /u/ have F1-F0 differences of less than three critical bands while low vowels have F1-F0 differences that exceed three critical bands, see Syrdal and Gopal, 1986). Such improvements may not have been as evident for the 6-channel processors with explicit coding of voicing information inasmuch as these latter processors probably provide a higher-resolution picture of F1 and F2 than the 4-channel processors. This higher-resolution picture of F1 and F2 may have masked any presumably-smaller contribution provided by F1-F0 differences.

To conclude this section on processors with explicit coding of voicing information, we note that the overall performance of the best of these processors closely approximates the performance of the best interleaved-pulses processor without explicit coding of voicing information (83 and 84%, respectively). However, the processors with explicit coding of voicing information generally sound more natural to MH and allow her to make reliable male/female distinctions. Also, in the case of the 6-channel processor with jittered intervals during unvoiced segments, significant improvements are found in the scores for recognition of consonants. Because consonant recognition is much more important than vowel recognition for understanding connected speech, this processor might be regarded as superior to the 6-channel processor without explicit coding of voicing information. The tradeoff in making a choice between these two strategies for implementation in a portable, real-time processor is that the above-mentioned advantages of the processor with explicit coding of voicing information can only be obtained at the cost of incorporating an accurate F0 and v/uv detector in the portable unit. As mentioned before, accurate extraction of these voicing parameters in real time is not trivial, and such extraction generally requires the use of complex systems of software and hardware.

Note on waveshape manipulations for interleaved-pulses processors

As mentioned before, a novel aspect of many of the interleaved-pulses processors tested with patient MH (but not with LP) is that "monophasic-like" pulses were used. These pulses were formed either by (a) generating in software asymmetric, charge-balanced pulses with a short-duration, high-intensity leading phase and a long-duration, low-intensity following phase of opposite polarity or (b) high-pass filtering true monophasic pulses to produce charge-balanced stimuli. The second phase of charge-balanced asymmetric pulses produced in method (a) above was 3.0 msec or less in duration for all processors in which these pulses were used. Also, peak currents never exceeded 1.0 mA for any type of stimulus waveform. Finally, we note that some high-pass filtering was applied to all stimuli delivered to the electrode array in that the signal drivers in the stimulus-isolation unit were AC coupled and the output currents of the isolation unit were fed to the electrodes through coupling capacitors. The effective corner frequency for the stimulus-isolation unit was around 20 Hz.

Use of "monophasic-like" pulses, as opposed to current- and duration-balanced biphasic pulses, was based on findings in psychophysical tests with patients SG and MH that indicated (a) a lower threshold and greater dynamic range could usually be obtained with one polarity of an asymmetric but charge-balanced pulse for radial bipolar pair stimulation and (b) a further "release" from temporal channel interactions could be obtained if monophasic pulses were first high-pass filtered (at about 100 Hz, single pole) before interleaved delivery to two channels in the electrode array. These advantages of "monophasic" pulses were particularly evident for pulse durations of 0.5 msec or less; indeed, parametric studies with the interleaved-pulses processors demonstrated that superior results could be obtained with 0.3 to 0.5 msec pulse durations (the tested range across all interleaved-pulses processors was from 0.2 to 0.7 msec), 0.2 msec interpulse interval (the tested range was from 0.1 to 1.5 msec), and with high-pass filtering at 100 Hz. With some processors the effects of these manipulations were large, and we therefore consider the specification of pulse parameters to be an important element in the design of interleaved-pulses processors.

IV. Concluding Remarks

The broadest conclusion to be drawn from the results presented in this report is that manipulations in the processing strategy used in an auditory prosthesis can have huge effects on recognition of consonants and vowels. Our patient attains outstanding levels of recognition with certain processing strategies, and poor-to-moderate levels of recognition with others. Certainly, this basic finding demonstrates the importance of selection of an appropriate processing strategy for an individual implant patient. Because our studied population of patients is limited, we do not know at this time whether one processing strategy will emerge as superior for all patients. However, for patients LP and MH, processors that represented the RMS energies in five or six bandpasses with interleaved pulses provided much better performance than the other strategies we have evaluated. We note, though, that both these patients had psychophysical manifestations of poor (patient MH) or extremely-poor (patient LP) nerve survival. It may be that a completely different class of processors would work best for a more-fortunate patient with good nerve survival. For example, the excellent results from approximately half of the patients in the UCSF series strongly indicate that a compressed-analog-outputs strategy may be as good as or superior to an interleaved-pulses strategy for cases in which nerve survival is good. We are anxious to test such a patient to evaluate this hypothesis. In the interim, however, the major lessons to us are that (a) different processing strategies can produce widely-different outcomes for individual patients; (b) interleaved-pulses processors are far superior to other processors for at least two patients with poor nerve survival; (c) processors other ~~that~~ than the interleaved-pulses processors may be superior for patients with good nerve survival; and (d) therefore it is important not to have an "adopted religion" for a single strategy of speech processing for auditory prostheses.

V. Plans for the Next Quarter

The major activity of next quarter will be continued testing of patient MH. Our plans include (a) evaluation of multiple processors in single tests of consonant and vowel recognition, where the presentations of both processors and speech tokens are randomized; (b) evaluation of the processing strategy used in the Melbourne/Nucleus device; (c) investigation of effects of electrode coupling configuration on recognition performance (e.g., radial bipolar versus longitudinal bipolar or radial bipolar versus monopolar); and (d) evaluation of selected processors with the tests of the Minimal Auditory Capabilities battery (the "MAC" test) and of the consonant and vowel tests developed by the Iowa group. Completion of these studies will extend our present range of tested processing strategies and coupling configurations, and will also allow us to compare performances across processors and laboratories with the standard (but lengthy) MAC and IOWA tests.

In addition to the above activities, we also expect to begin design and construction of portable speech processors in the next quarter. The first of these processors will implement in real-time hardware a 6-channel, interleaved-pulses strategy for our present patient. Also, we plan to evaluate the possibility of including explicit representations of F0 and v/uv boundaries in this processor, as described in section III of this report. Our leading candidate for extraction of F0 and v/uv boundaries is use of the Average-Magnitude-Difference-Function (AMDF) algorithm. This algorithm has already been implemented by us in a portable, real-time processor (QPR 6, NIH project N01-NS-2356), and we hope that we will be able to integrate the hardware and software developed for the AMDF processor into the larger system of the basic 6-channel, interleaved-pulses processor.

Finally, our plans include presentations of project results in invited lectures at the Kresge Hearing Research Institute in Ann Arbor and at the Conference on Speech Recognition with Cochlear Implants in New York.

VI. Acknowledgments

Dr. Robert Schindler and Mr. Steve Rebscher of UCSF, and Mr. Steve Hutchison and Mr. Gary Keibel of Storz Instrument Company, provided their able assistance at the operations for our first two implant patients at Duke. We are delighted to express here our great appreciation for the effort and generous contributions of time each of these individuals made to our new program. Finally, we are also pleased to acknowledge the important contributions of Dr. Margaret Skinner of Washington University Medical Center, who assisted us in some of the psychophysical studies conducted with patient MH.

VII. References

Breeuwer, M. and Plomp, R., Speechreading supplemented with frequency-selective sound-pressure information, J. Acoust. Soc. Am., 76 (1984) 686-691.

Burns, E.M. and Viemeister, N.F., Played-again sam: Further observations on the pitch of amplitude-modulated noise, J. Acoust. Soc. Am., 70 (1981) 1655-1660.

Eddington, D.K., Speech discrimination in deaf subjects with cochlear implants, J. Acoust. Soc. Am., 68 (1980) 885-891.

Schindler, R.A., Kessler, D.K., Rebscher, S.J. and Yanda, J., The University of California, San Francisco/Storz cochlear implant program, Otolaryngologic Clin. N. Amer., in press.

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

Simmons, F.B., Dent, L.J. and van Compernolle, D., Comparison of different speech processing strategies on patients receiving the same implant, Ann. Otol., 95 (1986) 71-75.

Syrdal, A.K. and Gopal, H.S., A perceptual model of vowel recognition based on the auditory representation of American English vowels, J. Acoust. Soc. Am., 79 (1986) 1086-1100.

White, M.W., Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes, Ann. N.Y. Acad. Sci., 405 (1983) 348-359.

Appendix 1

Summary of Reporting Activity for the Period of
December 27, 1985 through March 27, 1986,
NIH Contract N01-NS-5-2396

The following major presentation was made in the present reporting period:

Wilson, BS and Finley, CC: Latency fields in electrically-evoked hearing. Presented at the 9th Annual Meeting of the Association for Research in Otolaryngology, St. Petersburg, Florida, February, 1986.

An abstract for this presentation is shown on the next page of this Appendix.

LATENCY FIELDS IN ELECTRICALLY-EVOKED HEARING.

*B.S. Wilson and C.C. Finley, Neuroscience Program Office, Research Triangle Institute, Research Triangle Park, N.C. 27709.

In this presentation we will describe models that predict the spatial and temporal patterns of neural responses produced by intracochlear electrical stimulation. The simplest of these models couples a mathematical description of the field patterns generated by intracochlear electrodes with a mathematical description of strength-duration curves for electrical stimulation. The mathematical description of the field patterns ranges in complexity from an exponential-falloff model to our spiral-plane, finite-difference model of the UCSF electrode array (8th ARO, p. 105, 1984); data for the strength-duration model are obtained from the measurements of Loeb et al. (NYAS, 405:123-136, 1983) and van den Honert and Stypulkowski (Hearing Res., 14:225-243, 1984). For the exponential-falloff description of the electric field patterns, a parabolic-like profile of latencies is predicted by the combined model for monophasic, rectangular pulses. That is, as distance from the stimulating electrode (or electrode pair) increases, the electric field falls off and neurons at these locations are stimulated further and further out along their strength-duration curves. Ultimately, the strength of the current field falls below threshold (for the duration of the pulse) and neurons at locations more distant than this point are not stimulated. Manipulations of pulse intensity and duration have large effects (even for constant charge) on the extent of the excitation field and on the synchronicity of discharge across this field. These effects are highly correlated with psychophysical measures of loudness and threshold. Finally, the patterns of neural responses predicted with the finite-difference model of the UCSF array are in general more complex than the patterns just described for the exponential-falloff model. In particular, large discontinuities in the latency fields are predicted for balanced biphasic pulses delivered to the offset radial pairs of the UCSF array. Possible perceptual correlates of these discontinuities and other features of the latency fields will be listed; and the latency fields predicted for electrically-evoked hearing will be compared with the latency fields found in normal hearing. In our concluding remarks, we will mention some ways in which the latency fields produced in electrically-evoked hearing might be made to approximate the latency fields of normal hearing.

Appendix 2

Log of Speech-Testing Studies,
Patient MH

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
300	LPC resid coding, 1 pulse/10 ms, into pair 1-2	23/25 (later test date)	7/25, 7/25 (later test date)	8/24 (later test date)	4/24 3/24
150	4 channel compressed analog outputs, delivered to the apical 4 pairs. Compression ratio = 2:1. RTI tuning procedure.	23/25, 21/25	4/25, 6/20		
150	same as above, but using the UCSF tuning procedures	19/19	9/20	13/24	6/24, 5/24
150	same as above, but stimuli were delivered to pairs 1-2, 5-6, 11-12, and 15-16	24/25	7/25	9/24	4/24
150	just using the Ch 1 output, delivered to the inputs of the apical 4 pairs		9/25 (4/5 for "BOUGHT")		
160	4 channel compressed analog outputs, as in DESIGN 150, but with a compression ratio of 3:1. Stimuli delivered to pairs 1-2, 5-6, 11-12 and 15-16. UCSF tuning procedure.	20/25	8/25	13/24	7/24
160	same as above, except that the apical 4 channels are used	22/25	2/25	14/24	5/24
8	interleaved pulses processor. 6 channels of mapped stimulation, 2 channels at a time.	20/25	11/25, 20/25	13/24, 12/24	3/24
83	Hybrid processor in which a compressed analog signal is delivered to pair 1-2 and interleaved pulses are delivered to the remaining pairs	22/25	12/25, 11/25	17/24	8/24

Summary of Speech Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
83	same as above, except that the baseband channel was switched off	24/25 (25/25)	19/25 (13/25)	16/24 (13/24)	5/24 (5/24)
		later test date, see next page			
83	using just the baseband channel (i.e., all interleaved-pulses channels were switched off)	20/25	11/25		
N/A	lipreading alone	23/25		14/24	
84	Hybrid processor identical to DESIGN 83, but with the baseband extended from 50 to 4000 Hz	24/25	16/25	16/24	5/24
85	Plomp processor with compressed analog outputs delivered to pairs 1-2 and 5-6.	24/25	5/25	13/24	3/24
86	Plomp processor with interleaved-pulses outputs delivered to pairs 1-2 and 5-6.	20/25	7/25	18/24	10/24
87	Plomp processor with interleaved-pulses outputs delivered to pairs 1-2 and 5-6, and with RMS smoothers set to 10% of lower cutoffs of the bandpass filters. .2ms/ph + pulses, 1.5 ms interleaving	22/25	10/25	18/24 (all errors in cells adjacent to the diagonal)	9/24
87.1	same as above, but with time between pulses set at .2 ms	22/25, 23/25 (later date)	14/25 6/25 (later date)	14/24 (later date)	6/24 (later date)
		scores generally lower on this test day			

Summary of Speech-Testing Results

SIGN	Description	VOWELS*		CONSONANTS**	
		w. lips	w.o. lips	w. lips	w.o. lips
88	Hybrid processor, identical to DESIGN 84, but with no "speech filter" before the input to the baseband filter	22/25	11/25	13/24	5/24
88	same as above, but with the baseband channel switched off (generally superior to the above on vowels; no degradation in the consonants) This is the same condition as DESIGN 83, at the top of the previous page.	25/25 scores generally lower on this test day	13/25	13/24	5/24
95	without the baseband; interleave time = .2 ms (variant of DESIGN 88, with .2 ms⊕ pulses, separated by .2 ms; 6 channels)	25/25	16/25 21/25 (retest)	15/24 18/24 (retest)	5/24 9/24
95	with baseband (performance ↓)	25/25	17/25	14/24	4/24
96	Plomp, with 1.5 ms interleaving		tokens flawed		
37	Plomp, with .2 ms interleaving	23/25	9/25 12/25 (retest)	12/24	9/24
87	same as above, but with stimuli delivered to pairs 1-2, 15-16			14/24	
01	4 channel interleaved pulses, to apical 4 pairs (note first appearance of pain percepts)	22/25	13/25	12/24	
03	"monophasic" pulses to staggered channels (1-2, 5-6, 11-12, 15-16)	20/25	8/25	14/24	6/24

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
96	Plomp, with 1.5 ms spacing between pulses	24/25	11/25 11/25 (retest)	13/24 14/24 (retest)	5/24
505	Interleaved pulses processor, 2 channels on at a time, 4 channels in all. Electrodes: 1-2, 5-6, 7-8, 15-16.	20/25	9/25	13/24	7/24 8/24 (retest)
N/A	lipreading alone			11/24	
87	Plomp with dense pulses (retest?)	25/25	11/25	14/24	5/24
88	6 channel interleaved pulses, 2 channels on at a time, .5 ms/ph +/- biphasic pulses, spaced .7 ms apart, (channels adjusted to avoid pain), baseband <u>on full bore</u> (not the usual condition!)	17/25	2/25		
88	same as above, but with baseband switched off and with Ch 2 switched off	23/25			
88	same as above, but with Ch 2 inputs directed to Ch 1	24/25			
100	standard "monophasic" 6-channel interleaved pulses processor, .2 ms pulses, 1.5 ms between, 2 pulses at a time. "Monophasic" pulses are balanced in software. Ch 2 is not stimulated.	24/25	22/25	14/24	9/24
101	same as above, except that all 6 channels can have an output on each update cycle	25/25	18/25	17/24	7/24

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
105	same as DESIGN 100, but with .5 ms $\oplus$ "mono" pulses, spaced 1.2 ms apart	25/25	20/25	13/24	6/24
111	same as DESIGN 101, except that the break frequency of the speech filter is moved from 1200 Hz to 300 Hz	25/25	17/25 14/25 (retest)	13/24	7/24
102	Plomp processor, simultaneous pulses, delivered to "noninteractive" pairs 1-2 and 5-6	23/25 20/25 (next day)	12/25 13/25 (next day)	14/24	12/24
114	4-channel strategy, all outputs on, with presentation order of 1,3,2,4. Channels: 1-2, 5-6, 7-8, 15-16. .2 ms $\oplus$ pulses, separated by .8 ms. Pulses are charge balanced in software.	18/25	8/25	13/24	8/24
112	same as DESIGN 101, except that the pulses are high-pass filtered at 100 Hz	24/25	15/25 18/25 (retest)	13/24	8/14 (test died)
113	same as DESIGN 112, except that .3 msec pulses are used, spaced .2 ms apart	24/25 25/25 (retest on another date)	23/25 20/25 (-----)	20/24 19/24 (-----)	14/24 13/24 (-----)
113.1	same as above, but with a "noise-biasing" signal applied to elecs 9/Ref. Noise signal: 200-6 kHz, presented at threshold level.	23/25	18/25	18/24	10/24
116	same as DESIGN 112, except that the pulses are .5 msec in duration, spaced .1 ms apart	23/25	19/25	20/24	10/24

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
117	same as DESIGN 116, except that the inverse-filter cutoff is increased to 200 Hz	25/25	17/25	17/24	12/24
122	same as DESIGN 113, except that software-compensated "monophasic" pulses are used instead of "inverse-filtered" pulses (master gain = 700)	25/25	20/25	16/24	14/24
122.1	same as above, except that master gain = 650	25/25	17/25	18/24	11/24
700	interleaved pulses + pitch-pulse baseband. Channel gains adjusted to avoid pain (generally disappointing)	20/25	8/25		
119	same as DESIGN 113, except that ITHRESH = 4 (<u>much</u> better)	25/25	21/25		
119	tests on 5/28, 5/30	24/25	20/25		
		25/25	18/25	15/24	12/24
		25/25 (repeat)	19/25	20/24	
700	Ch 2 off.	20/25	10/25		
	Ch 2 off, Ch 1 at 500.			16/24	8/24
702	same as DESIGN 119, except that round-robin updates are from base to apex (i.e., 6-5-4-3-2-1) rather than from apex to base (i.e., 1,2,4,6,3,5) ("very natural, best processor yet")	23/25 25/25	24/25	21/24	12/24 10/24 (1 presentation)
701	same as DESIGN 119, except that linear mapping laws are used for the channel outputs instead of the "standard" logarithmic mapping laws	could not map		16/24	8/24

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
800	6-channel round-robin processor like DESIGN 702, except that the round-robin sequences are timed to begin at pitch periods for voiced-speech sounds. Max-rate outputs are used during unvoiced intervals as in DESIGN 702.	24/25 24/25 (repeat)	18/25 20/25	20/24 20/24	12/24 12/24
801	same as DESIGN 800, except that the interval between sequential pulses is increased from .2 ms to .4 ms	22/25	17/25	18/24	8/24
802	same as DESIGN 800, except that the cutoff frequencies of all RMS-smoothing filters are set at 25 Hz	24/25	22/25	19/24	11/24
805	same as DESIGN 800, except that jittered-rate outputs are used during unvoiced intervals	24/25 21/25	16/25	22/24	12/24 19/24
806	same as DESIGN 805, except that a <u>4-channel</u> output is used. Channels are 1-2, 5-6, 7-8, 13-14.	22/25	18/25	19/24	11/24
807	same as DESIGN 806, except that the channel assignments are changed to 1-2, 7-8, 11-12, 15-16			17/24	10/24
803	same as DESIGN 800, except that a <u>4-channel</u> output is used. Channels are 1-2, 5-6, 7-8, 13-14.			17/24	14/24

Summary of Speech-Testing Results

DESIGN	Description	VOWELS* w. lips	VOWELS* w.o. lips	CONSONANTS** w. lips	CONSONANTS** w.o. lips
804	same as DESIGN 803, except that the channel assignments are changed to 1-2, 7-8, 11-12, 15-16			18/24	10/24 9/24 (repeat)

*Chance is 5/25
**Chance is 3/24