# Fourteenth Quarterly Progress Report

December 27, 1988 through March 26, 1989

NIH Contract N01-NS-5-2396

Speech Processors for Auditory Prostheses

Prepared by

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

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

CONTENTS

I. Introduction

The purpose of this project is to design and evaluate speech processors for 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. Initial studies with a patient (SW) implanted with the Nucleus cochlear prosthesis;

2. Evaluation of several processing strategies in studies with a patient (RB) implanted with the Symbion cochlear prosthesis;

3. Development of software for Multidimensional Scaling (MDS) analysis of confusion matrix data;

4. Design and construction of a general purpose interface in software for obtaining experimental access to the full capabilities of the Nucleus 22-electrode implant device;

5. Continued upgrade of the Cochlear Implant Laboratory at Duke, along the lines indicated in recent progress reports for this project;

6. Continued collaboration with the UCSF team on the development of the speech processor and transcutaneous transmission system for a next-generation cochlear prosthesis; and

7. Continued preparation of manuscripts for publication.

The main body of the present report is an analysis of the results from follow-up studies with two patients (MC1 and MC2) implanted with the UCSF/Storz cochlear prosthesis. Several processing strategies were evaluated with each of these patients, including hybrid strategies of the type proposed in the Twelfth Quarterly Progress Report for this project. Medial consonant and vowel tokens uttered by both male and female talkers were included and compared. New software tools for speech testing and for information transfer analyses were applied and evaluated.

## II. Binary Comparisons of Speech Processor Performance

### A. Introduction

Patients MC1 and MC2 were seen for follow-up studies at Duke University Medical Center (DUMC) in early November, 1988. MC2 was seen October 31 through November 5, and MC1 November 7 - 12. Both were patients implanted with the UCSF/Storz cochlear prosthesis and had been part of a group of six such patients studied by us earlier [9, 10, 11]. Previous studies with MC1 were conducted February 2 - 6, and those with MC2 March 23 - 27, 1987.

Of particular interest in these follow-up studies were the possible benefits of hybrid processor designs, strategies that combine a continuous compressed analog (CA) signal on one channel with interleaved pulses (IP) signals on all other available channels. Proposed on the basis of our earlier work comparing pure CA and IP strategies [10], hybrid processors may help some patients who derive substantial information from the waveform details of a single channel of analog stimulation but are unable to benefit from additional analog channels because of severe channel interactions. Use of IP stimuli on the other channels might offer substantial release from channel interactions in such cases, and make additional speech information available.

These studies also included medial consonant and vowel confusion tests with tokens uttered by both male and female talkers, allowing analysis of differences in processor performance across voice pitch. In the present report we shall emphasize binary comparisons of (1) differences in the utilization of similar processors by the two patients to obtain similar overall speech reception results, and (2) performance differences within individual strategies when processing male and female voices.

Both patients have the standard clinical UCSF/Storz implant and transcutaneous transmission system (TTS). MC2 uses the standard four-channel CA processor, while a MC1 uses a three-channel variation in which the signals normally sent to channels three and four are combined as channel three. The change was made seven months after implantation, necessitated by a failure mode of the implanted receiver system [12].

Onset of hearing loss for MC1 was at age 10, due to congenital lues, with profound deafness by age 55 and implantation eight years later in October, 1985. MC2 was 42 years of age at the onset of Meniere's syndrome, was profoundly deaf by age 47, and received the implant eleven years later in May, 1986. Both are among the best UCSF patients in terms of the level of speech perception they enjoy.

## B. Methods

### Processing Strategies

Four distinct types of processing strategies were compared in the course of these studies: (1) the CA strategy of each patient's clinical processor, (2) an IP processor utilizing all available stimulation channels, (3) a hybrid processor providing one channel of continuous CA stimulation and IP signals on all other available channels, and (4) a processor with only two channels of IP stimulation, based on the work of Breeuwer and Plomp [1, 13].

For each channel receiving pulsatile stimulation in any of these processors, the RMS energy in the corresponding frequency band of the input signal was coded as pulse amplitude within the available dynamic range for electrical stimulation--between perceived threshold of hearing and most-comfortable level (MCL). The available channels were stimulated cyclically in each case, the cycles beginning synchronously with voice pitch periods during voiced intervals. Unvoiced intervals were signaled explicitly in all these pulsatile processors, either by randomly varying the intervals between cycles ("Jittered") or by delivering cycles in immediate succession ("Max Rate"). Pulse widths for each channel were limited by the minimum that would allow the threshold-to-MCL perceptual scale to be spanned by the range of stimulus amplitudes available through the patients' TTS.

Balanced biphasic pulses were used for all these processors. While both patients had been studied in our laboratory before, threshold and MCL measurements were repeated in each case, for a variety of pulse widths.

For the full three- and four-channel IP processors, several different designs were evaluated for each patient with limited speech testing, and then one chosen to represent the IP type in the more extensive comparative tests.

Tests

In order to make better use of our limited testing time with these (and subsequent) patients, we sought to increase (1) the number of different vowel and consonant tokens used in confusion tests, (2) the number of presentations of each token, and (3) the speed of test administration. While the first two changes were desirable in order to increase the range and statistical significance of the tests, the third was necessary to preserve these tests as an effective tool for rapid screening of candidate designs during IP processor optimization.

The vowel token set was expanded from five to eight and the consonants from eight to fourteen for the present studies. [Based on experience with these two patients, our consonant token set has since been expanded further to sixteen.] The medial vowel tokens, in displayed matrix order, were the words "heed", "hawed", "head", "who'd", "hid", "hood", "hud", and "had". The medial consonant nonsense tokens, in a /aCa/ context, were m, n, f, v, s, z, sh, p, b, t, d, g, k, and j (/dʒ/).

Each of the vowel tokens was presented three times in each randomization, with three randomizations being administered in each experimental condition, for a total of nine presentations of each token. Each consonant token was presented five times in each of two randomizations, for a total of ten presentations in each condition.

In order to minimize the time required to evaluate an experimental condition with these expanded confusion tests, we sought both to reduce administrative time between tests, and to maximize the rate of token presentation and response within each test.

The first of these goals was achieved through the use of highly integrated test selection, administration, and recording software. The randomization and presentation of tokens, and the recording of all test parameters and responses were accomplished automatically, without experimenter intervention. Digital audio and video recordings (in any combination) of the tokens (in any randomized order) were presented from the laserdisc materials produced by the University of Iowa [3]. Four different recordings of each token were used--two with a male talker and two spoken by a female. Male and female talkers were not mixed within a test. Token recordings were selected from the material available on the videodisc, and their onset and offset addresses adjusted to eliminate or, failing that, to minimize the effects of extraneous visual cues. Additional goals in adjusting the onset and offset points of each token were uniform token length and uniform delay before onset of phonation within each token.

After each token was presented, a table of possible responses was displayed on the patient's video screen and the next token was cued up, pending keyboard entry of the patient's response. We anticipated that the most efficient technique for most patients would be for them to enter their own responses directly into the computer. As a backup procedure, the patient could identify a chosen response orally (by number, to avoid issues of interpretation of the patient's speech) and the experimenter enter it into the computer. Based on the present work with these two intelligent and experienced patients, we have adopted the latter procedure as the standard one in our laboratory. Any short-term gains in response speed by direct patient data entry are more than nullified by a longer-term decay of patient enthusiasm and intensity. Patients not only complete more tests per unit time when the one-way oral communication with the experimenter is included, they show much less evidence of fatigue at the end of each testing period. Advantages include greater flexibility in the patient's physical posture during the tests and greatly reduced demands for visual adjustments.

In addition to the tests of vowel and consonant confusions, as time permitted, we administered the Speech Pattern Contrast (SPAC) test from videodisc [7], the Minimal Auditory Capabilities (MAC) battery from audio tape, [5] and live voice connected discourse tracking [2, 6]. The SPAC tests were administered by software of our own design, following the standard sequences of presentation. As discussed above with respect to vowel and consonant confusion tests, patient responses in the SPAC were oral numbers, and the experimenter entered these into the computer. The pace of testing was limited only by the time taken by the patient to respond. The MAC battery and tracking tests were administered and scored by DUMC staff audiologist Robert Wolford.

## C. Results

Each confusion matrix test was archived automatically in the following form: a sequence number, date, and time stamp; switches and labels defining the experimental conditions (processor if any, whether or not visual cues were supplied, gender of talker, randomization number); and a list containing each response to each token presented.

Among the new software tools seeing first use in these studies was a program that assembles aggregate matrices. This program made it possible to sum the results from all tests satisfying specified ranges of sequence number, date/time, talker gender, processor, and/or visual cues, and then either to add or subtract the result from a target matrix. This made it quite easy to perform and view a wide variety of "cuts" through the multi-dimensional database. Since the database preserves each raw response and the aggregating program itself scores responses using the appropriate randomizations, it is possible also to check for learning-curve trends--or other systematic variations in performance--within as well as across individual tests.

Our confusion matrices were subjected to a variety of different types of analysis. Percent correct and raw confusion data were displayed for the experimenter during the administration of each test and printed out at the end. For the medial consonant tests, matrices also were displayed for the voicing feature and for place and manner of articulation (see Fig. 1). Such matrices are helpful in quickly identifying relative strengths and weaknesses of particular processing strategies. Symmetric off-diagonal patterns, for example, may offer an indication of information that is being conveyed effectively, but not yet correctly interpreted by the patient. Information transmission (IT) [4] and sequential information (SINFA) analyses, [8] were applied to the aggregate matrices to treat such questions more formally and in greater detail. A novel form of display was devised for the results of such analyses in an effort to recognize circumstances in which insignificant differences in the data lead to misleadingly large apparent differences in analysis output.

IOWA Test of Consonant Confusions          Patient: MC

Date: 881112     Time: 0943     Sequence: 110
Talker: Male     Randomization: 0
Hybrid processor alone

| | m | n | f | v | s | z | sh | p | b | t | d | g | k | j |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| m | 5 | | | | | | | | | | | | | |
| n | | 5 | | | | | | | | | | | | |
| f | | | 4 | 1 | | | | | | | | | | |
| v | | | 1 | 2 | 1 | 1 | | | | | | | | |
| s | | | | | 4 | | 1 | | | | | | | |
| z | | | | | 2 | 3 | | | | | | | | |
| sh | | | | | | | 5 | | | | | | | |
| p | | | | | | | | 5 | | | | | | |
| b | | | | | | | | | 5 | | | | | |
| t | | | | | | | | 1 | | | | | 4 | |
| d | | 1 | | | | | | | | | 4 | | | |
| g | | | | | | | | | | | 1 | 4 | | |
| k | | | | | | | | 2 | | 2 | | | 1 | |
| j | | | | | | | | | | 2 | | 1 | 1 | 1 |
| Heard: | 5 | 6 | 5 | 3 | 7 | 4 | 6 | 8 | 5 | 4 | 5 | 5 | 6 | 1 |
| Sent: | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |

Identification: 69%

| | uv | v |
|---|---|---|
| uv | 29 | 1 |
| v | 7 | 33 |
| Heard: | 36 | 34 |
| Sent: | 30 | 40 |

Voicing: 89%

Distribution of Errors
Voicing only: 18%
Place only: 55%
Manner only: 5%
Multiples: 23%

| | lab | l-d | alv | pal | vel |
|---|---|---|---|---|---|
| lab | 15 | | | | |
| l-d | | 8 | 2 | | |
| alv | 1 | | 22 | 1 | 6 |
| pal | | | | 5 | |
| vel | 2 | | 3 | | 5 |
| Heard: | 18 | 8 | 27 | 6 | 11 |
| Sent: | 15 | 10 | 30 | 5 | 10 |

Place: 79%

| | plo | fri | nas |
|---|---|---|---|
| plo | 29 | | 1 |
| fri | 4 | 26 | |
| nas | | | 10 |
| Heard: | 33 | 26 | 11 |
| Sent: | 30 | 30 | 10 |

Manner: 93%

Fig. 1. Example of data displayed for the experimenter during administration of medial consonant confusion tests and printed out at the conclusion of each test.

## Percent Correct Scores

Overall percent correct scores are displayed in Table 2 and Figure 2 for medial vowel and consonant confusion tests in a variety of experimental conditions and for both patients. Chance performance would be 12.5 % for vowel and 7.1 % for consonant identification.

Table 2. Speech Test Scores

(units of percent correct, unless otherwise noted)

| | | Confusions | | | | | | SPAC + | | | MAC Battery | | | | | | | | | | | | Tracking ++ | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | 8 Vows | | | 14 Cons | | | Overall | | | | | | Sp | | Int | Fin | Sp | Sp | | | | | |
| | | V | V+P | P | V | V+P | P | V | V+P | P | Q/S | Acc | N/V | S/D | Vow | Con | Con | 4Ch | Rec | NU6 | CID | WIC | V | V+P |
| **MC1** | | | | | | | | | | | | | | | | | | | | | | | | |
| CA | m | 51 | 99 | 100 | 32 | 80 | 39 | | | | | | | | | | | | | | | | | |
| | f | 35 | 99 | 95 | 30 | 86 | 63 | 39 | 66 | 49 | | | | | | | | | | | | | | |
| | mf | 43 | 99 | 98 | 31 | 83 | 51 | | | | | | | | | | | | | | | | | |
| | | | | | | | | | * | | | | | | | | | | | | | | | |
| H-CA | m | 51 | 96 | 82 | 32 | 85 | 60 | 39 | 49 | | | | | | | | | | | | | | | |
| Hybrid | m | 51 | 100 | 90 | 32 | 84 | 67 | | | | 100 | 90 | 100 | 100 | 72 | 70 | 83 | 95 | 64 | 30 | 78 | 28 | 34 | 68 |
| | f | 35 | 100 | 89 | 30 | 88 | 67 | 39 | 75 | 50 | | | | | | | | | | | | | | |
| | mf | 43 | 100 | 90 | 31 | 86 | 67 | | | | | | | | | | | | | | | | | |
| IP | m | 51 | 90 | 54 | 32 | 94 | 59 | | | | 95 | 90 | 100 | 100 | 77 | 73 | 71 | 100 | 84 | 32 | 76 | 34 | 34 | 80 |
| | f | 35 | 82 | 78 | 30 | 86 | 47 | 39 | | 48 | | | | | | | | | | | | | | |
| | mf | 43 | 86 | 66 | 31 | 90 | 53 | | | | | | | | | | | | | | | | | |
| BP | m | 51 | 86 | 49 | 32 | 87 | 48 | | | | 100 | 80 | 98 | 100 | 67 | 69 | 71 | 95 | 52 | 16 | 51 | 20 | 34 | 70 |
| | f | 35 | 85 | 58 | 30 | 88 | 48 | 39 | 74 | | | | | | | | | | | | | | | |
| | mf | 43 | 86 | 54 | 31 | 88 | 48 | | | | | | | | | | | | | | | | | |
| **MC2** | | | | | | | | | | | | | | | | | | | | | | | | |
| CA | m | | | 76 | 45 | 95 | 75 | | * | | 100 | 85 | 100 | 100 | 80 | 84 | 81 | 100 | 72 | 28 | 95 | 62 | | |
| | f | | | 92 | 38 | 100 | 76 | 60 | 85 | 59 | | | | | | | | | | | | | | |
| | mf | | | 84 | 42 | 98 | 76 | | | | | | | | | | | | | | | | | |
| Hybrid | m | | | 57 | 45 | 87 | 54 | | | | | | | | | | | | 48 | 12 | 84 | 50 | | |
| | f | | | 82 | | | | 60 | | 46 | | | | | | | | | | | | | | |
| | mf | | | 70 | | | | | | | | | | | | | | | | | | | | |
| IP | m | | | 73 | 45 | 86 | 43 | | | | 100 | 90 | 100 | 100 | 87 | 86 | 85 | 95 | 80 | 32 | 97 | 72 | 27 | 84 |
| | f | | | 75 | 38 | 92 | 48 | 60 | | 49 | | | | | | | | | | | | | | |
| | mf | | | 74 | | | | | | | | | | | | | | | | | | | | |
| BP | m | | 90 | 74 | 45 | 89 | 55 | | | | 95 | 85 | 98 | 100 | 68 | 89 | 81 | 100 | 48 | 32 | 91 | 50 | 27 | 77 |
| | f | | 90 | 56 | | | | 60 | 82 | | | | | | | | | | | | | | | |
| | mf | | 90 | 65 | | | | | | | | | | | | | | | | | | | | |

* — only one battery administered for this condition

+ — [% wrt chance]    ++ — [words-per-minute]

Abbreviations: V—vision only, P—processor only, V+P—both; m—male talker, f—female, mf—both
H-CA—pulsatile channels alone from hybrid processor
Q/S—Question/Statement, Acc—Accent, N/V—Noise/Voice, Sp S/D—Spondee same/diff.
Vow—Vowels, Int Con—Initial consonants, Fin Con—Final consonants
Sp 4ch—Spondee 4-choice, Sp Rec—Spondee recognition, NU6—monosyllabic word recog.
CID—CID sentences, WIC—Words in context

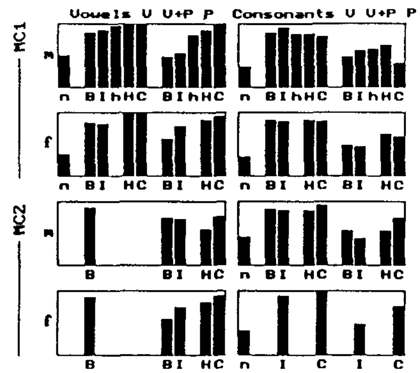


Fig. 2. Summary of percent correct scores for medial vowel and consonant confusion tests with patients MC1 and MC2. Male and female talkers are indicated by "m" and "f" respectively and each box contains vision alone, vision-plus-processor, and processor alone results in that order. Codes for the various processors are: n--none, B--BP (two-channel pulsatile, based on the work of Breeuwer and Plomp), I--IP , h--the pulsatile channels only a hybrid processor (labeled "H-CA" in Table 2), H--full hybrid processor, and C--CA. Full scale corresponds to 100% correct.

Information Transmission Analysis

To evaluate the patterns of confusions under the various conditions studied, aggregate matrices were subjected to the information transmission (IT) analysis described by Miller and Nicely [4]. In such an analysis the "relative transinformation" is calculated for selected articulatory or acoustic features of the phonemes in the identification tests. The relative transinformation score for each feature, expressed here as percent information transfer, indicates how well that feature was transmitted to the subjects. The consonant features selected for the present study were voicing, nasality, place of articulation, duration, frication, envelope cues, and visual (viseme) cues. For the vowel confusion matrices the chosen analysis features were first formant, second formant, duration, and viseme.

Each phoneme included in the matrix tests is assigned a classification under each of the analysis features. Then the phoneme identification results may be used to calculate how well each feature is being transmitted in each experimental condition.

## Sequential Information Analysis

While IT analysis can indicate how much information regarding the selected features is conveyed by the various processing strategies, interpretation of such results is complicated by redundancies in the associations between phonemes and feature classifications. Because of such overlapping information, IT analysis cannot determine the extent to which each separate feature is used by a subject in identifying phonemes.

A procedure developed by Wang and Bilger [8] removes the redundancies, and produces a model of how the selected features might combine to account for the observed phoneme identification performance. Called SINFA, for Sequential Information Analysis, this procedure begins with the unconditional relative information transfer results of a standard IT analysis. The feature with the highest such score then is identified as the most salient and is held constant while a second conditional IT analysis is performed across the remaining features. This process continues, with the most salient remaining feature being identified at each stage and held constant for all subsequent iterations, until all features have been identified or until the remaining features can account for less than 1% of the total received information.

The final output of SINFA thus consists of (1) a sequence of the features deemed most salient at each successive iteration, and (2) the relative contribution of each feature, in the same order, to the pattern of observed phoneme identification performance.

We have developed a way of displaying IT and SINFA results graphically that facilitates judgments as to the significance of differences across the various experimental conditions. An a anlysis of summary consonant and vowel confusion data for all four types of processor is displayed thusly in Figure 3. Figures 5 and 6 compare performance of each type of processor for the two different patients. Figures 9 and 10 make similar binary comparisons for male and female talkers.

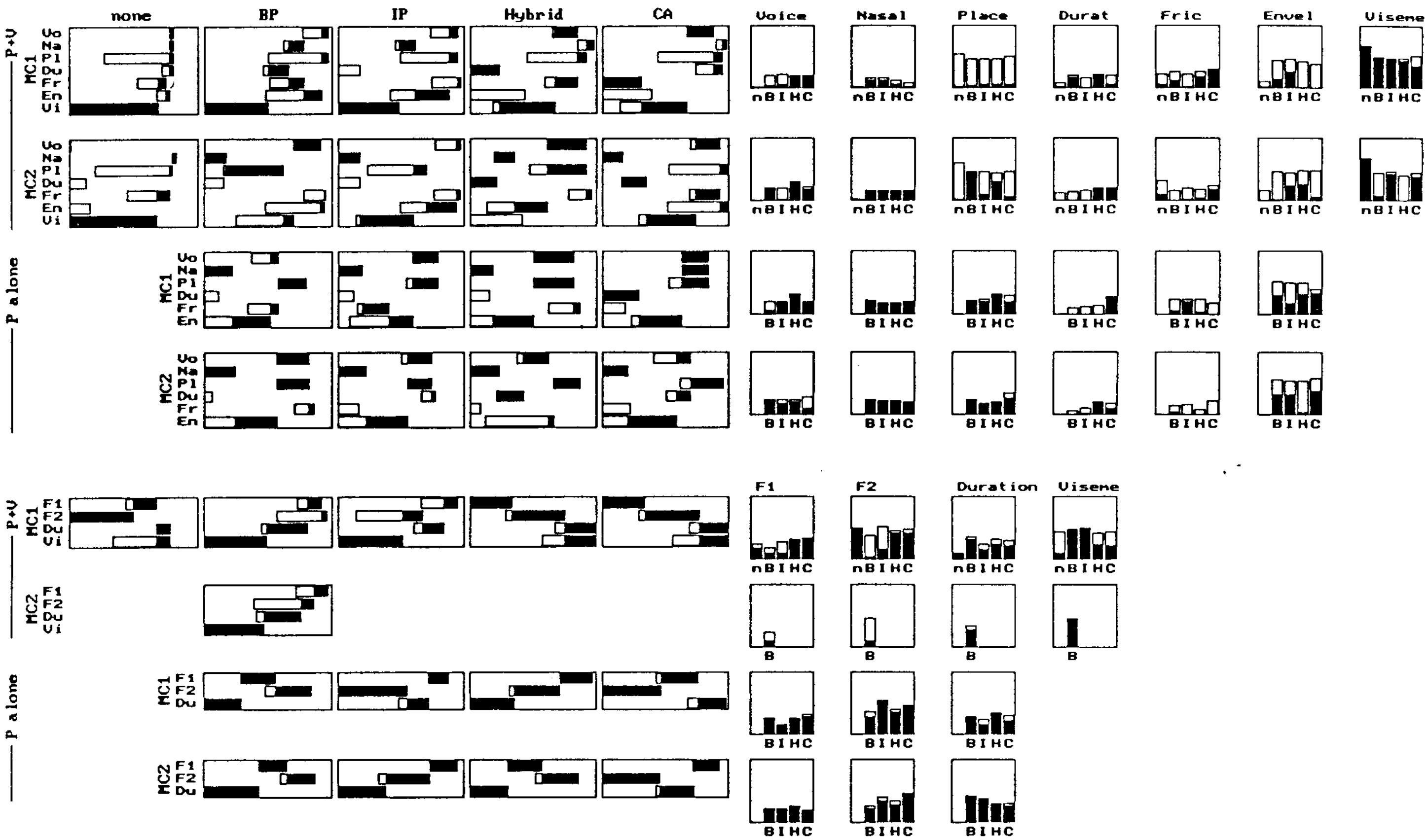


Fig. 3. Summary information transfer and SINFA analysis results for patients MC1 and MC2 in a variety of experimental conditions. The top four rows are processor analyses of medial consonant tests, bottom four of vowel data. Abbreviated processor labels are the same as in Fig. 2. See text for explanation. All scales are 100% of information received.

## Interpretation of IT and SINFA Results Displays

In the left half of such figures, each box contains data for several transmitted features in a single experimental condition. Within a box, the length of each horizontal bar represents a fraction of the total received information, with the full width of the box corresponding to 100 percent.

The solid bars indicate, from left to right in each box, the order in which features were identified by SINFA as most salient at each stage of the analysis. The length of each solid bar shows the additional fraction of the total received information accounted for by that feature at that point. Thus the right edge of the rightmost solid bar in each box indicates the aggregate proportion of the received information that is explained by all identified features. This proportion, in turn, indicates how completely the identified features can account for the subjects' observed judgments. Solid bars that overlap along the horizontal scale represent multiple features judged equally salient at that SINFA analysis stage. (The assignments across phonemes for two or more remaining features can become identical in the later iterations of a SINFA analysis.)

The open bars represent, on the same scale, the full or additional extent of the information accounted for by each feature in the unconditional analysis. Thus features selected by SINFA may have open extensions to their solid bars, indicating redundant information that did not increase the total received information in the sequence identified by SINFA. Finally, the unconditional information potential of features not selected by SINFA are displayed as open bars at the left edge of the box.

The same bars also are displayed vertically in the right half of such figures, where each box compares data for transmission of a single feature across the several different experimental conditions. The solid portion of each bar again represents the received information attributed to that feature in the particular order of saliency determined by the SINFA analysis, while open bars (or extensions) represent information identified in the unconditional analysis but not used by SINFA in accounting for the total information received. The full vertical scale of each box is 100 percent of received information. The horizontal order of the bars within each box in the right half of these figures is the same as the horizontal order of the experimental condition boxes in the left half, while the horizontal order of the feature boxes in the right half is the same as the vertical order of the bars within each box in the left half.

SPAC Scores

All Speech Pattern Contrast (SPAC) test materials presently available on videodisc [7] are with a female talker. SPAC results are available, then, for only one of the two binary comparisons--MC1 *vis a vis* MC2--being presented in this report. In each condition, the SPAC battery was administered twice and the results averaged and expressed as percentage with respect to chance. [Since chance performance in every case was 50 %, that value was subtracted from the raw scores and the result multiplied by two. Averaging the results from two administrations of the battery meant that a minimum of 24 trials were represented in each measurement.] The quantities we display and the number of underlying trials in each condition are: overall score (240), suprasegmentals (48), segmentals (192), vowels overall (48), consonants overall (144), initial consonants (72), final consonants (72), consonant place (48), voicing (48), and continuance (48). A full set of SPAC data was obtained for these two patients with IP, Hybrid, and CA processing strategies without visual cues and for BP processors with lipreading. Data for other conditions were obtained as time permitted. An overall summary of SPAC data is presented in Figure 4, and the appropriate binary comparison subset presented in Figure 7. The form of graphical presentation is closely analogous to that used in the right halves of IT and SINFA plots.

MAC Battery Scores

As may be seen in Table 2, the high levels of performance by both MC1 and MC2 on the closed-set parts of the Minimal Auditory Capabilities battery make those subtests relatively insensitive to the speech processing differences under investigation. Thus we will focus on the four open-set subtests of the MAC battery, to wit: spondee recognition, recognition of monosyllabic words from the NU #6 list, CID everyday sentence recognition, and recognition

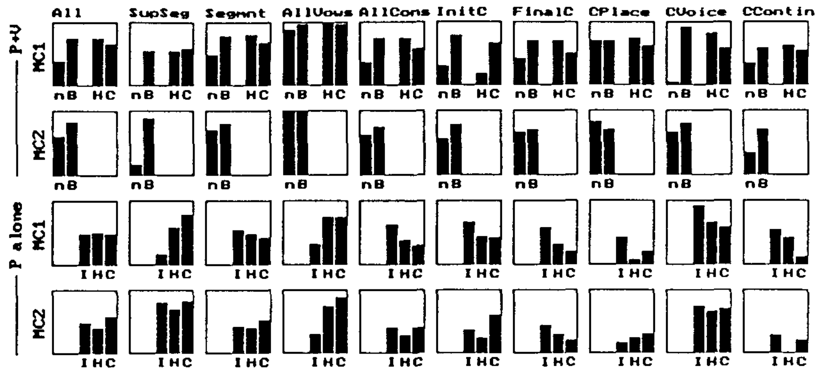


Fig. 4. Summary of SPAC results for patients MC1 and MC2 in a variety of experimental conditions. Abbreviated processor labels are the same as in Fig. 2. Percentages with respect to chance are graphed on a chance-to-100% scale for the quantities: overall score, suprasegmental features, segmental features, vowels overall, consonants overall, initial consonants, final consonants, consonant place, consonant voicing, and consonant continuance.

of Words in Context [5]. (The extent to which the first of these is really an "open-set" test for such experienced subjects is, of course, questionable.) Results from these tests, all with male talker and without visual cues, are included in Figure 8. The results are shown as percent correct identification scores.

Speech Tracking Scores

Tests of connected discourse tracking [2, 6] were administered to each patient with IP and BP processors and without processor. MC1 was tested with her Hybrid processor as well. Results, averaged across four test paragraphs and in words-per-minute units, are included in Table 2. Testing time limitations prevented our obtaining contemporaneous data under comparable conditions for the patients' CA processors.

## D. Discussion

### Comparisons among Processing Strategies

A major feature of work done for the present project has been the ability to compare a variety of different speech processing strategies in each patient studied. Four distinct types of processor were evaluated formally in the most recent studies with both MC1 and MC2. Fig. 2 summarizes the overall percent correct scores for the medial consonant and vowel confusion tests administered with each type. The order of display of the processor types, as will be the case throughout this report, is: BP, IP, Hybrid, and CA. This order corresponds generally to increasing stimulus density (increasing number of pulsatile channels, the substitution of first one, and then several continuous analog channels) and may be expected, in some patients, to correspond to increasing likelihood of channel interactions.

This happens also, for MC1, to be the order of increasing overall performance at vowel discrimination. For consonants MC1's overall performance depended strongly on whether visual cues were available and whether the voice being processed was male or female. With visual cues, performance for all four processors was roughly equivalent, with IP doing somewhat better for the male talker. Without lipreading, pulsatile stimuli seemed especially helpful to MC1 for the male talker, but at least one analog channel seemed important when the female voice was being processed. For MC2 without visual cues, the CA processor performed relatively better for consonants, while pulsatile processors were of relatively more help for vowels at male voice pitch.

Thus a cursory examination of these vowel and consonant identification scores immediately raises questions about the effects of voice pitch on processor performance and about differences in the uses made of similar processor features by different patients. Both issues will be considered in subsequent sections.

In these comparisons among processing strategies, of course, it is important to keep in mind that results may be biased by the extensive experience of both patients with one of the processor types, the CA strategy of their everyday prostheses.

The recent studies with MC1 and MC2 marked our first trials with hybrid processors. The data for consonant identification without visual cues include indications both of hybrids providing benefits in common with purely pulsatile strategies (MC1, male talker) and of their sharing strengths with purely analog processors (MC1, female talker). There also is evidence, however, of hybrid processors performing no better than two-channel BP pulsatile strategies (MC2, male talker) and little better than the two-channel pulsatile part of the same hybrid alone (MC1, male talker, labeled "h" in Fig. 2 and "H-CA" in Table 2.)

Beyond differences in raw percent correct scores, the vowel and consonant identification data may be examined in terms of information transmission analysis and SINFA. Fig. 3 is a summary of such analyses across the four processor types with male and female talker data combined. Such displays make a great deal of detailed information available in a form that, with practice, facilitates the discovery of significant patterns. In this report we shall note only a few example phenomena in Fig. 3, and thereafter restrict our discussion to binary comparison versions of such displays.

As one would expect, in those experimental conditions including visual cues ("P+V" with any or no processor) the viseme features for both vowels and consonants are well represented, as is consonant place and, to a lesser extent, frication. Adding any of the four processors to the visual cues significantly improves the transmission of consonant envelope, voicing, and nasality features. Among the vowel data, we note that MC1 finds the analog processors (H, CA) especially helpful with first formant information, and uses the pulsatile IP strategy more effectively to obtain second formant cues. MC2 makes good use of vowel duration cues from the pulsatile IP and BP processors. In fact, the BP processor conveys vowel duration information well in all four experimental conditions.

In interpreting information transmission results it is important to consider both unconditional IT and SINFA. There may be substantial duplication in the unconditional information available in two or more features, so that in SINFA terms access to both is little better than access to either. This occurs, for instance, with MC1's CA processor in the absence of visual cues for consonant voicing, nasality, and place features (see column "CA" of the third row in Fig. 3). Substantial differences in the sequence of features identified by SINFA as most salient, on the other hand, may merely reflect insignificant differences among features with essentially equivalent saliency (compare SINFA sequencing of the consonant place, envelope, and viseme features across all MC2's processors when used in conjunction with lipreading). Finally, consider the fraction of received information SINFA is able to account for given our choice of features for analysis. The fraction is quite large for all conditions that include a processor plus visual cues, and for vowel data, except with the BP processor. For consonants without visual cues, the fraction seems to depend more strongly on processor type than on patient, except for the case of the CA processor.

Other useful tools in assessing differences among processors include the SPAC tests. SPAC results for the most recent studies with MC1 and MC2 are summarized in Fig. 4. For patients who perform at the levels of MC1 and MC2, these tests tend to be more useful when administered without visual cues. However, the limited data we have with vision included ("P+V") do indicate the potential of the BP processor as an adjunct to lipreading in patients with only two independently-stimulable channels. These results do not--in the single patient MC1--demonstrate any advantage of a hybrid processor over a BP when used with visual cues.

Using the SPAC tests without visual cues, we have compared the performance of IP, hybrid, and CA processors in both patients. For MC1, the overall SPAC results are roughly equivalent, and the hybrid's performance is generally intermediate in the constituent tests (consonant place is the only exception), the other two processors' scores varying widely. For MC2, the hybrid yields the poorest results overall, notably so in consonant continuance. Similar patterns of relative processor performance across the two patients are observed for vowels and final consonants.

## Comparisons of the Two Patients' Use of Similar Strategies

While a great strength of our normal approach to speech processor evaluation is comparison of different strategies in the same patient, it is important also to assess the degree to which different patients may use the same processor information in different ways. Such inquiries are greatly complicated by the many types and levels of expected differences--physiological, intellectual, and psychological--among patients. The ability to look for patterns that are both characteristic of a particular patient and manifested consistently by that patient across a number of different processing strategies offers some advantage in this regard.

Some intriguing differences in the two patients' overall vowel and consonant identification score patterns have already been noted. We now approach a more detailed assessment of such differences using the analytic tools we have developed.

Figs. 5 and 6 compare IT and SINFA results for the two patients with and without the availability of visual cues, respectively. The displays are essentially the same as in Fig. 3, except that only two conditions are compared in each row, conditions differing only in the patient's identity. Since these results are in terms of percent of the total information received by the patient, we have included--at the right end of each row--the corresponding overall identification scores for each patient, as an indication of the level of performance achieved in each case.

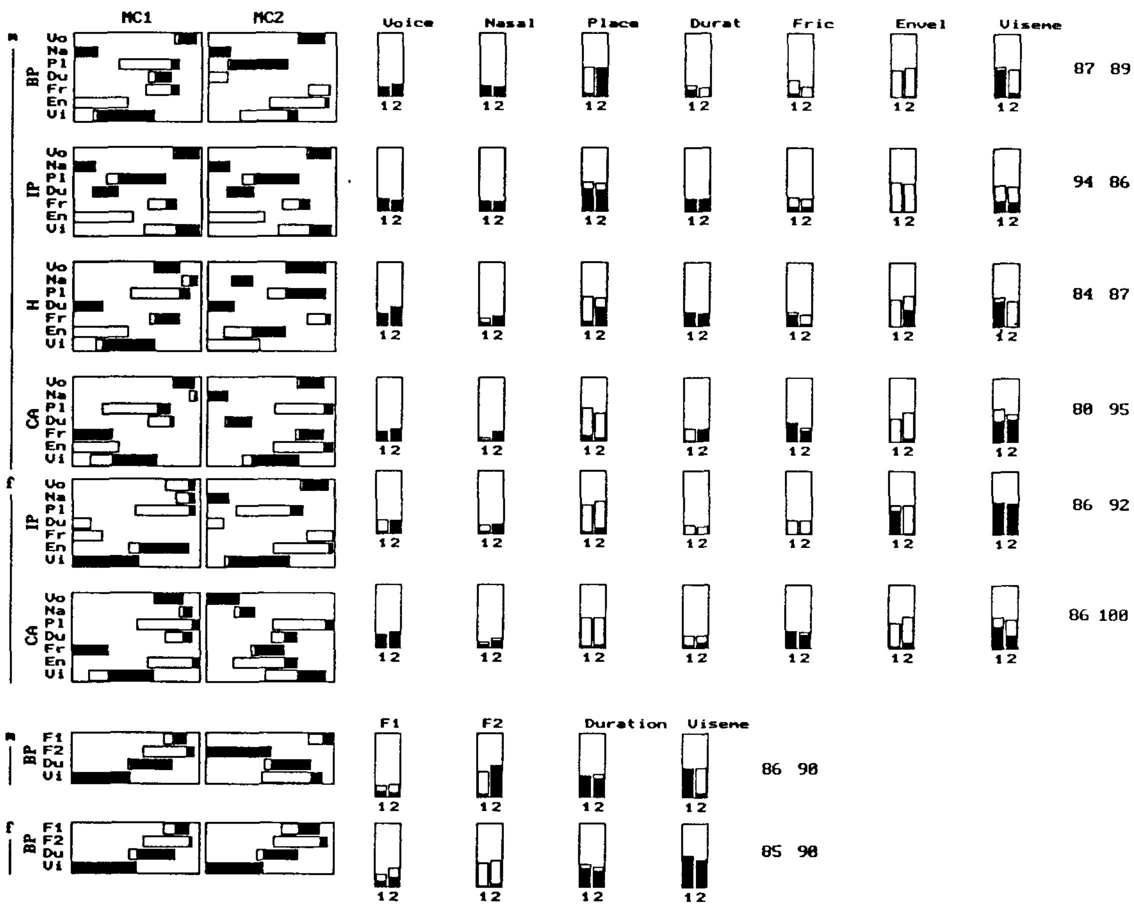


Fig. 5. Binary comparisons of the utilization of similar speech processors in conjunction with visual cues by two patients--MC1 ("1") and MC2 ("2"). Information transmission and sequential information (SINFA) analyses. Scale is 100% of information received. Numbers to right show overall percent correct scores for the raw confusion test results underlying each row's analyses. The top six rows are consonant data, the bottom two vowel.

When visual cues are available (Fig. 5), there seem to be only isolated instances of information transmission differences between the patients that are supported both by unconditional IT and by SINFA analyses. Examples--all for consonants and male talker--include nasality and frication via the CA type processor, voicing and frication via the hybrids, and frication via the BP.

For the male talker, MC1's best result with IP and MC2's with CA show similar unconditional patterns and similar, unexceptional degrees of redundancy in the SINFA sequence, as do MC2's best results with CA and a female talker. SINFA accounts for 94 % or more of the total received information in every case.

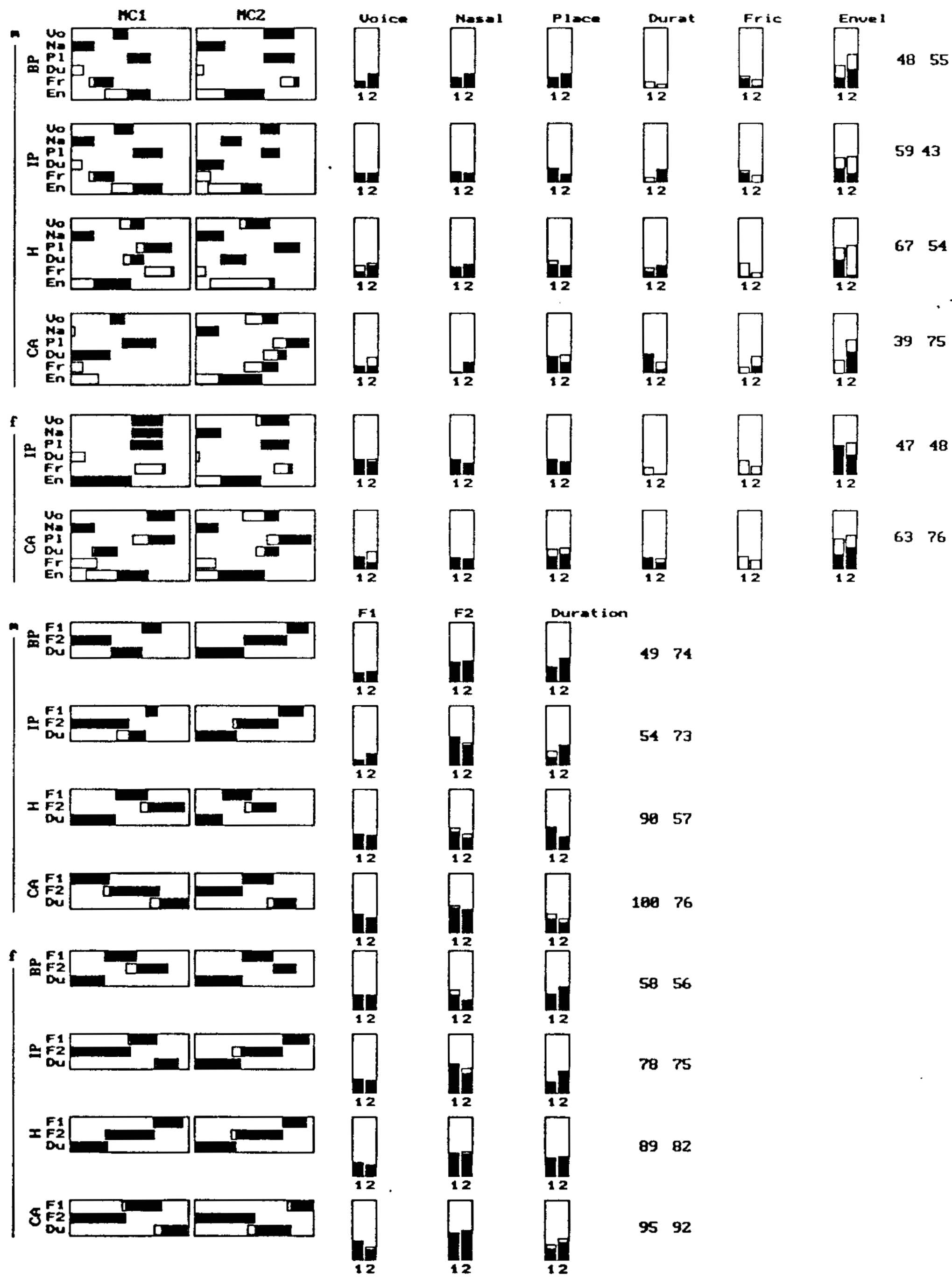

Fig. 6. Binary comparisons of the utilization of similar speech processors without visual cues by patients MC1 and MC2. Quantities are as described in Figs. 3 and 5. The top six rows are consonant data, the bottom eight vowel.

Processor-alone performance data, on the other hand (Fig. 6) display a number of strong patterns. The use of their CA processors by the two patients for consonants are strikingly different for a male talker, but not for female voice. In terms of our chosen set of features, SINFA accounts for 95 % of the information received by MC2 but only 71 % of that received by MC1 for the male talker. For female voice the difference is only 10 %. Some strengths seem available with both pulsatile and analog strategies (MC2's marked access to envelope information via CA and also BP, where the fraction of information accounted for by SINFA also differs by more than 20 %). Other differences appear to be associated with certain processor types (MC1's better access to frication cues via all processors that include pulsatile stimuli).

MC2's better performance with BP is associated with better IT scores for four of the six features, while MC1's advantage with a hybrid may lie in better fricative information, the particular sequence identified by SINFA notwithstanding.

Among the vowel data there is some suggestion of patient differences that extend across the range of processor types: MC2's use of duration information, for instance, and MC1's use of second formant cues also seem to correlate with overall performance. For vowels it is the hybrid processor with male talker that poses the largest difference in the fraction of received information accounted for--96 % for MC1 but only 68 % for MC2. The other threeprocessors all have differences in the 15 to 20 % range for male voice, with MC1's performance better accounted for with H and CA and MC2's with BP and IP. The largest difference with female talker is less than 5 %.

SPAC scores for the two patients are compared in Fig. 7. Note that several different levels of test are represented here, with some scores being constituents of others. For BP processors with visual cues, MC2's better overall score reflects superior perception of suprasegmental features, with

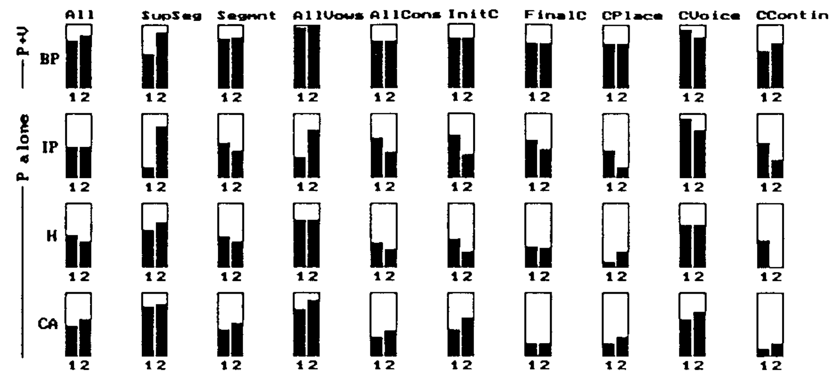


Fig. 7. Binary comparisons of the utilization of similar speech processors by patients MC1 ("1") and MC2 ("2"). Quantities are Speech Pattern Contrast (SPAC) test results as described in Fig. 4. Scale is chance-to-100%. Use of the BP processor was evaluated in conjunction with visual cues, while the other three strategies were evaluated processor alone.

differences in the use of consonant voicing and continuance balancing out between the patients. Overall scores for IP strategies without lipreading are equivalent, but with quite different underlying patterns, MC2 doing much better with suprasegmental and vowel information and MC1 showing superior results on every consonant test. In comparing SPAC performance with a hybrid processor alone, we find notable differences in access to consonant place and continuance, with MC1's better segmental results outweighing MC2's better suprasegmental showing in the overall score. For CA processors without visual cues MC2 achieved the higher scores in every case. Finally, we note that processors including pulsatile stimuli seemed to aid both patients with final consonant identification.

When it comes to comparison of the two patients in terms of the "open-set" subtests of the MAC battery, summarized in Fig. 8, it is convenient to relate performance with the other two experimental processor types to the clearly superior support offered both patients by IP. The decreased pulsatile information offered by BP strategies, then, causes more difficulty for MC2 than MC1, especially in CID sentence recognition. The loss, of course, corresponds to one half of MC2's four IP channels and only one third of MC1's

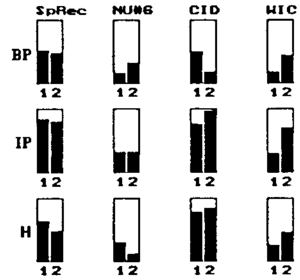


Fig. 8. Binary comparisons of the utilization of similar speech processors by patients MC1 ("1") and MC2 ("2"). Quantities graphed are percent correct scores for the open set subtests of the Minimal Auditory Capabilities (MAC) battery: spondee recognition, recognition of monosyllabic words from the NU #6 list, CID everyday sentence recognition, and recognition of words in context. Processors alone, with no visual cues.

three. In the hybrid strategies, however, substituting an analog signal for one of the IP channels also causes more damage to MC2's performance than to MC1's--this time most notably in monosyllabic word recognition.

Comparisons of Processor Performance for Male and Female Voices

Turning now to a more controlled binary comparison, one made within the data from each patient, we consider differences in the processing of male and female voices. Again, we begin by examining, in Fig. 9, IT and SINFA results for the various processors used in conjunction with visual cues. With visual cues available, there are no striking differences, in terms of our chosen set of analysis features, in the processing of male vs. female speech by our four processor types. The largest overall identification score difference was for IP-processed consonants and patient MC1. At first glance, the SINFA display for that condition may seem to indicate that a broader range of features was utilized in interpreting the male voice, but comparison of the unconditional IT data on the right suggests that the large SINFA sequence differences were triggered by rather subtle parameter differences. The fraction of total

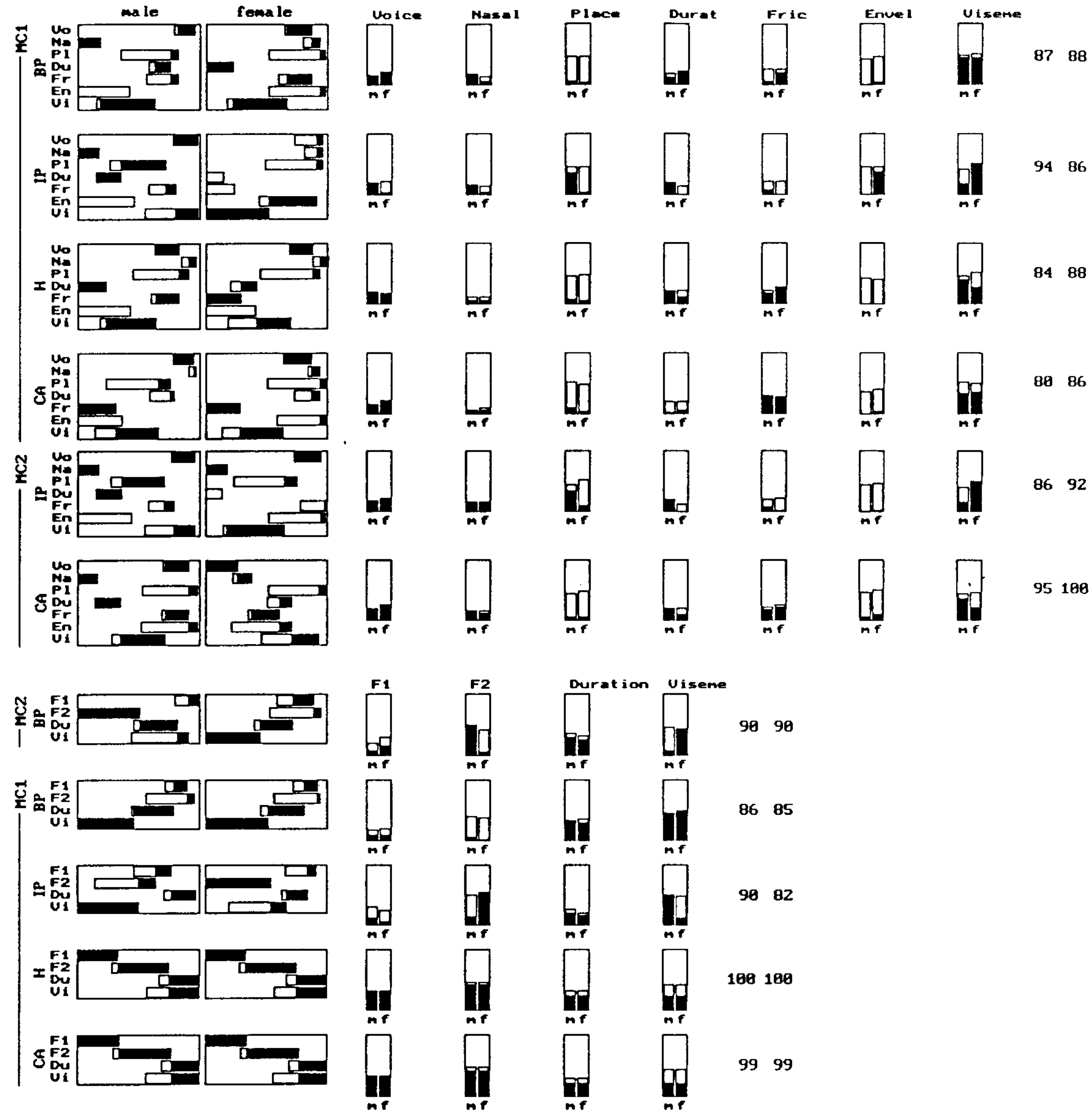


Fig. 9. Binary comparisons of the performance of various speech processors in conjunction with visual cues for male ("m") and female ("f") voices in tests with patients MC1 and MC2. Information transmission and sequential information (SINFA) analyses. Scale is 100% of information received. Numbers to the right show overall percent correct scores for the raw confusion test results underlying each row's analyses. The top six rows are consonant data, the bottom five vowel. Note that the patient order has been reversed for the vowel data in this figure, to facilitate comparison of BP processor data.

received information accounted for by SINFA for the conditions of Fig. 9 is greater than 94 % in every case except one (MC1, vowels, IP, female voice) where the fraction was 91 %.

When visual cues were not available, male-female voice differences became much more obvious. The IT and SINFA data are shown in Fig. 10.

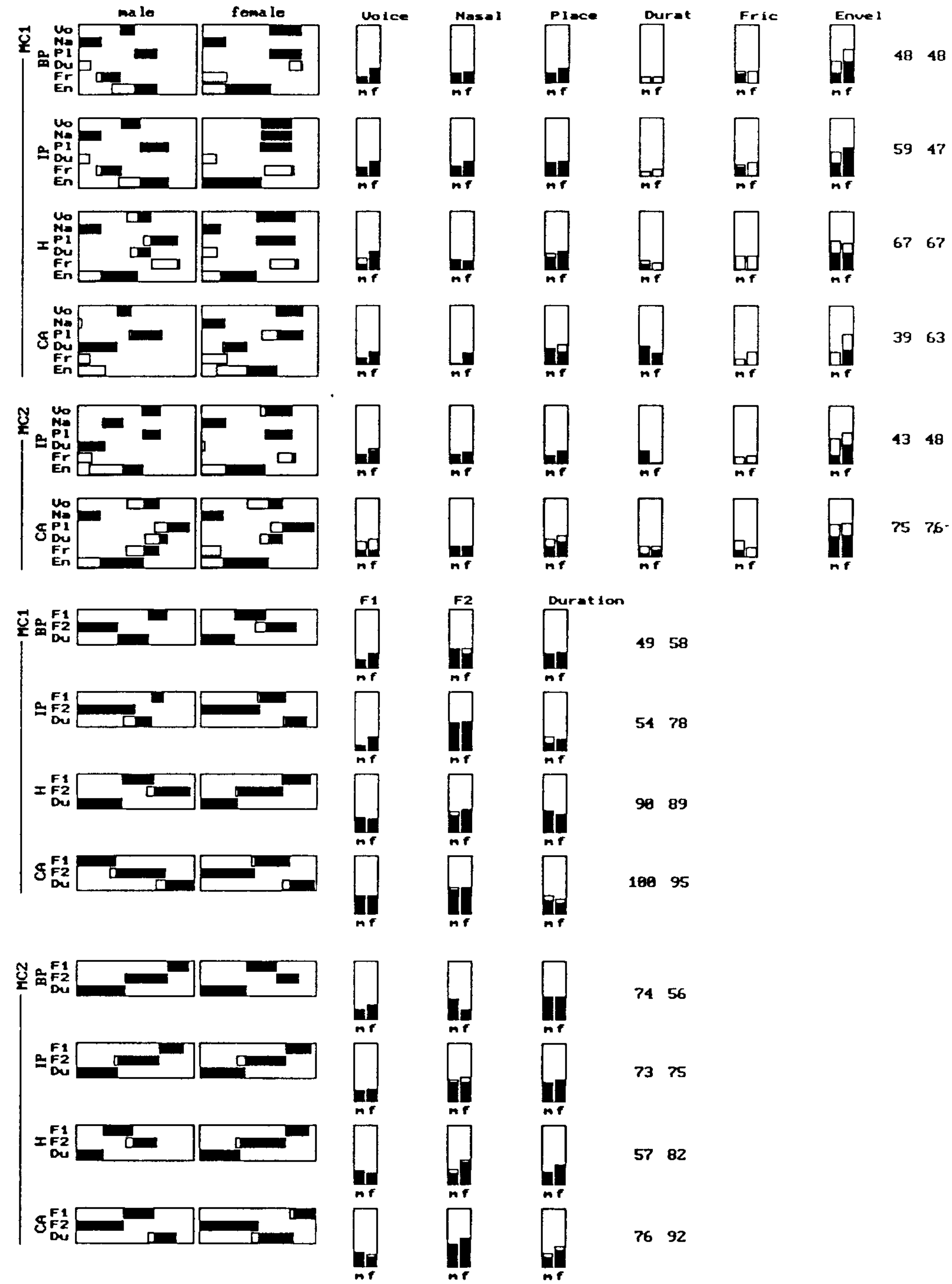


Fig. 10. Binary comparisons of the performance of various speech processors without visual cues for male ("m") and female ("f") voices in tests with patients MC1 and MC2. Quantities are as described in Figs. 3 and 9. The top six rows are consonant data, the bottom eight vowel.

For consonants, the voicing, place, and envelope features generally were transmited better when the female talker's voice was being processed, including cases for which overall identification scores and/or fraction of received information accounted for were closely matched. There is little difference in either quantity for MC1's hybrid and MC2's CA processors. For MC1's BP we see the same identification scores with different accounted-for fractions while for her IP processor there is a significant identification score difference with about the same fraction of received information accounted for in each case. The nasality, duration, and frication features each show strong male-female differences in one or more processors, but no patterns across different processor types.

Among the vowel data, first formant information differences underly the overall advantages in identification score for female voice with both MC1's pulsatile processors, accompanied by various differences in fraction of received information accounted for. The male voice superiority in vowel identification with MC2's BP processor is based on better second formant information. With MC2's hybrid and CA processors, second formant and vowel duration cues combine to produce large advantages for vowel identification with a female talker. For these two cases the total fraction of received information accounted for by SINFA is 68 % (male) and 94 % (female) for the hybrid and 85 % (male) and 100 % (female) for CA.

## E. Conclusions

### Hybrid Processor Prospects

The theoretical basis for expecting hybrid processors to be especially effective in certain special circumstances remains. In a patient who was (1) able to obtain a lot of information from a single channel of analog waveform and (2) plagued by severe simultaneous channel interactions, a hybrid processor might hold real promise.

In our most recent studies with patients MC1 and MC2, however, we have seen that such hybrid strategies hold no particular promise for a wider patient population. While we have obtained some data indicating that hybrid processors can combine identifiable attributes of both pulsatile and analog strategies, we have found that alternative pure pulsatile processors with the same number of channels allow equal or better performance in these two patients.

### Variations in Processor Performance with Voice Pitch

In the absence of visual cues, different patterns of information transfer emerge, depending on whether a strategy is processing speech from a male or a female talker. These patterns merit further study for the informed design of processors that are more robust in the face of voice pitch variations. The next steps beyond the preliminary studies outlined in this report will involve (1) investigation of alternative sets of categories for information transmission analysis and (2) obtaining sufficient data to allow statistical analysis of detected difference patterns. In particular, statistical correlations between various information transfer parameters and overall performance scores should be investigated.

## Patterns in Processor Utilization by Different Patients

There is evidence suggesting that the information made available by similar processing strategies may be used in distinctly different ways by different patients in achieving similar levels of overall speech perception. Efforts to pursue this line of inquiry will be complicated by all the unknown variables that accompany any study comparing results in different patients. Our ability to search for patient-specific patterns that persist across a range of different processing strategies may allow more progress than could otherwise be expected in this direction. The potential implications for cochlear prosthesis fitting in general make this topic worthy of further attention.

## Analysis Tools

Information transmission analysis, both in its unconditional and SINFA forms, consititues a valuable tool for future diagnostic and analytic work in support of speech processor design. The graphic displays of such data introduced in this report, especially when enhanced by common color coding of each analysis feature, allow rapid and effective examination of complex bodies of data, with access to enough information to assess the significance of observed patterns. Along with multidimensional scaling techniques and a better knowledge of the correlations between analytic quantities and the best practical measures of overall performance, IT and SINFA will find increasing use in the evaluation and custom design of speech processors for auditory prostheses.

## F. References

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

[2] De Filippo CL, Scott BL (1978) A method for training and evaluating the reception of ongoing speech. J. Acoust. Soc. Am. 63: 1186-1192

[3] Iowa Audiovisual Speech Perception Laser Videodisc, Tyler RS, Department of Otolaryngology--Head and Neck Surgery, The University of Iowa Hospitals.

[4] Miller GA, Nicely PE (1955) An analysis of perceptual confusions among some English consonants. J. Acoust. Soc. Am. 27: 338-352.

[5] Owens E, Kessler Dk, Raggio M, Schubert ED (1985) Analysis and revision of the Minimal Auditory Capabilities (MAC) battery. Ear Hear. 6: 280-287

[6] Owens E, Raggio M (1987) The UCSF tracking procedure for evaluation and training of spech reception by hearing-impaired adults. J. Speech & Hear. Disorders 52: 120-128

[7] SPAC Test Version II Laser Disc Recording, Boothroyd A, The Graduate Center of the City University of New York.

[8] Wang MD, Bilger RC (1973) Consonant confusions in noise: A study of perceptual features. J. Acoust. Soc. Am. 54: 1248-1266.

[9] Wilson BS, Finley CC, Lawson DT, Wolford RD (1988) Speech Processors for Cochlear Prostheses. Proc. IEEE 76: 1143-1154

[10] Wilson BS, Finley CC, Lawson DT (1989) Representations of Speech Features with Cochlear Implants. To appear in J.M. Miller and F.A. Spelman (eds.) Models of the Electrically Stimulated Cochlea, Springer-Verlag. [Manuscript included in 12th Qtr. Progress Report, Contract No. N01-NS-3-2356, NINCDS, NIH]

[11] Wilson BS, Finley CC, Lawson DT, Wolford RD (1989) Direct comparisons of analog and pulsatile coding strategies with six cochlear implant patients. In preparation.

[12] Wilson BS, Schindler RA, Finley CC, Kessler DK. Lawson DT, Wolford RD (1988) Present status and future enhancements of the UCSF cochlear prosthesis. In: Banfai P (ed) Cochlear implant: Current situation. Erkelenz, West Germany: Rudolf Bermann Gmbh, pp. 395-427.

[13] Wilson BS, Finley CC, Lawson DT (1987) Eighth quarterly progress report, NIH Contract N01-NS-5-2396.

## III. Plans for the Next Quarter

The next quarter will be the first for a new contract (N01-NS-9-2401). Our plans for the upcoming quarter include the following:

1. Patient studies with HP and MP. HP, implanted with the Nucleus device, will visit us July 5 - 14 for studies in which Dr. Bryan Pfingst and Dr. Margo Skinner will participate. Dr. Don Eddington will participate in the July 17 - 22 studies with MP, a patient implanted with the Symbion device.

2. Continued analysis of data from studies with patients RB and SW in the last quarter.

3. Preparation for, and participation in, the Engineering Conference on "Implantable Auditory Prostheses," to be held July 30 through August 4.

4. Continued preparation of manuscripts for publication.

APPENDIX

Summary of Reporting Activity for the Period of

December 27, 1988 through March 26, 1989

NIH Contract N01-Ns-5-2396

The following invited paper was submitted for publication:

Finley, CC: A finite-element model of radial bipolar field patterns in the electrically stimulated cochlea -- Two and three dimensional approximations and tissue parameter sensitivities. In Proc. Eleventh Ann. Conf. Engineering in Medicine and Biology Soc., IEEE Press, in press.

A preprint is attached.

A FINITE-ELEMENT MODEL OF RADIAL BIPOLAR FIELD PATTERNS IN THE ELECTRICALLY STIMULATED COCHLEA - TWO AND THREE DIMENSIONAL APPROXIMATIONS AND TISSUE PARAMETER SENSITIVITIES

Charles C. Finley

Neuroscience Program, Research Triangle Institute, Research Triangle Park, NC 27709
Div. of Otolaryngology, Dept. of Surgery, Duke University Medical School, Durham, NC

ABSTRACT

Field patterns within the electrically stimulated cochlea are described by a three-dimensional finite-element model. A radially-oriented bipolar electrode pair in an insulating carrier is modeled in scala tympani. Results are compared with an earlier two-dimensional model and field pattern sensitivities for selected tissue resistivities are explored.

INTRODUCTION

Previous work using finite-element models of intracochlear field patterns relied on a two-dimensional sheet description of a cochlear cross section with a bipolar electrode in scala tympani [1]. This earlier model had two distinct limitations. First, the aspect ratios for many triangular elements were not optimal. Second, the two-dimensional geometry made the implicit assumption that all regions of the cochlear cross section extended to infinity both above and below the plane of the cross section. This was a useful approximation for the soft tissue structures but was a significant distortion of the electrode surfaces. This paper presents estimates of field patterns produced by a more realistic three-dimensional, finite-element model of a cochlea implanted with a radially-oriented, bipolar electrode pair. Results from this model using other electrode configurations have been published previously [2].

THE THREE-DIMENSIONAL FINITE-ELEMENT MODEL

In this model a cross section of the cochlea is projected linearly along an axis perpendicular to the plane of the section, thus producing a short, straight segment of the cochlea. A radially-oriented bipolar electrode pair, mounted in a carrier insulator is located in scala tympani. An enlarged view of the central region of a layer from the model is shown in Figure 1. Each finite element is shrunk 15% geometrically for emphasis. Each layer of the model contains 304 nodes defining 204 five- and six-sided solid elements. Twelve layers of varying thicknesses are included in the complete model for a total of 1976 nodes and 2448 solid elements. Results for the two electrode configurations shown in Figure 2 are presented in this paper. One configuration (a) assumes the electrodes extend the full length (5.2mm) of the longitudinal axis of the model, whereas the other configuration (b) is a discrete focal pair spanning only 200 microns of the longitudinal axis. An

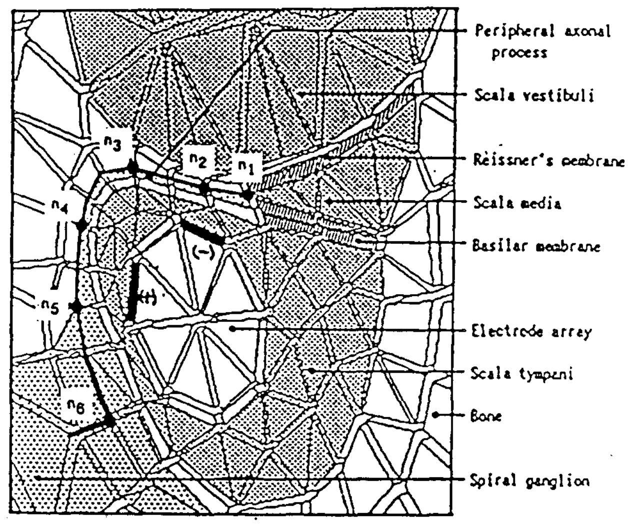


Figure 1. Single layer of FE model.

electrode configuration is specified by setting the resistivities of the surface plate elements of the electrode carrier to represent either insulators or electrode conductors. Node potentials at nodes bordering the conductive electrode surfaces are fixed arbitrarily at +100 and -100 mvolts prior to computation. Element resistivities (in ohm-cm) are defined regionally to characterize the electrodes(0.1), the carrier insulator($10^9$), the endolymph(60), the perilymph(70), Reisner's membrane(60480), basilar membrane(1800), the anisotropic neural tissue of the peripheral axon leading down from the habenula to the spiral ganglion(300 axial; 1500 transaxial), the spiral ganglion itself(300) and bone(630). The model sensitivity to tissue resistivity specifications is studied by

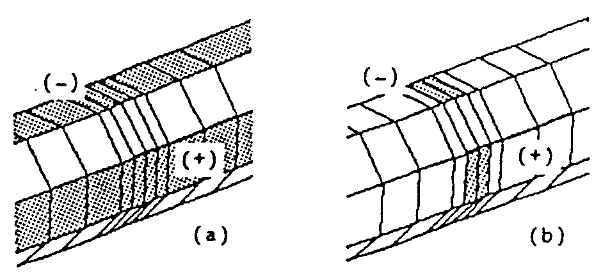


Figure 2. Electrode configurations (see text).

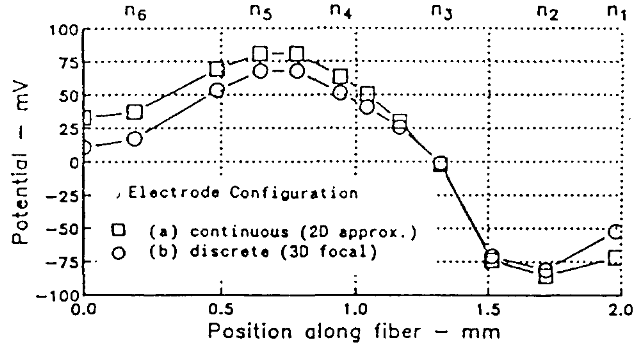


Figure 3. Field patterns for electrodes (a) and (b).

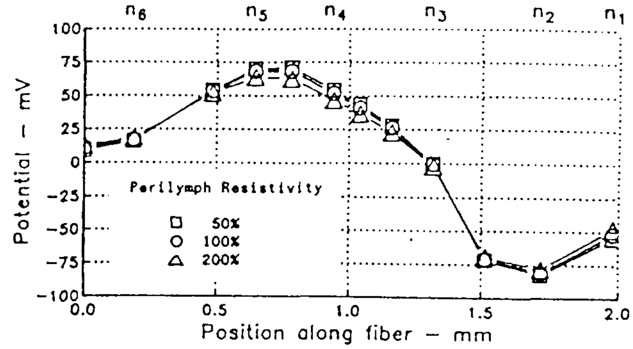


Figure 4. Perilymph effects on field patterns.

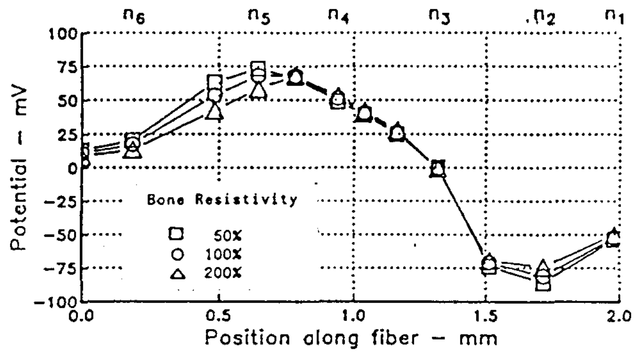


Figure 5. Bone effects on field patterns.

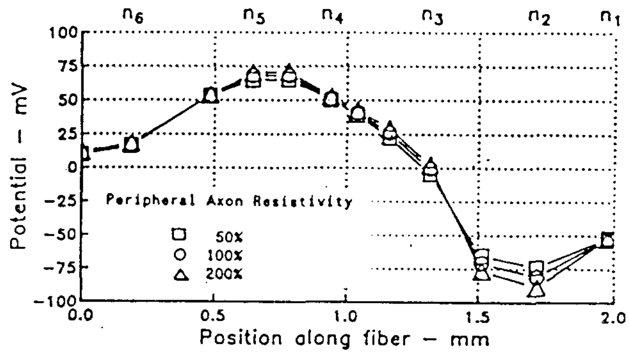


Figure 6. Peripheral axonal process effects on field patterns.

independently setting the resistivities of perilymph, bone and the peripheral axon to 50% and 200% of their standard values listed above. The outputs of the model are the potential patterns that occur in the vicinity of the neural elements for fixed potentials applied to the electrodes. For this paper, potentials along a single neural locus, located directly above the center of the electrode structure, are reported. This locus lies on the modiolar side of the peripheral axonal processes and extends from the center of the ganglion to the habenula as indicated in Figure 1. Further details of the model construction can be found in [2].

## RESULTS

Figure 3 shows the neural element potential levels for both electrodes described in Figure 2. Note that the electrode configuration that approximates the implicit assumptions of the previous two-dimensional model produces larger peak levels and steeper gradients in the region between the electrode pairs. However, at distant neural locus positions, the spatial gradients are significantly smaller.

Figures 4, 5 and 6 show the effects on neural locus potential levels for 50% and 200% manipulations of perilymph, bone and peripheral axonal process resistivities, respectively. In all cases, alterations of the potential levels are modest and the lower resistivity condition produces a more broadly spreading electrode field. Perilymph resistivity mostly affects potential levels between the electrodes, since the largest volume of perilymph in the vicinity of the electrode lies between the electrodes. Bone resistivity mostly affects potential levels in the ganglion region, due to the greater bone mass located near the ganglion in the modiolus. Only very thin bony elements surround the peripheral neural processes. Neural tissue resistivity has a slight impact on field patterns near the habenula in this model.

## CONCLUSIONS

Based on this more realistic three-dimensional model, the conclusions from previous work [1] are confirmed and restated. For a closely-placed, radially-oriented, bipolar scala tympani electrode:

- o first order effects on potential patterns are due to electrode geometry;
- o second order effects on potential patterns are due to tissue impedance effects.

## REFERENCES


1. Finley C, Wilson B and White M (1987): A finite-element model of bipolar field patterns in the electrically stimulated cochlea: A two-dimensional approximation. IEEE/9th Conf. EMBS, 1901-1903.

2. Finley C, Wilson B and White M (1989): Models of neural responsiveness to electrical stimulation. To be published as chapter 5 in Models of the Electrically Stimulated Cochlea, (Miller JM and Spelman FA, eds.), Springer-Verlag.


## ACKNOWLEDGMENT


This work was supported by contracts N01-NS-3-2356 and N01-NS-5-2396 from the Neural Prosthesis Program of NINCDS and a Professional Development Award from Research Triangle Institute.