X Original sent to printing

B

Sixth Quarterly Progress Report

August 1 through October 31, 1990

NIH project N01-DC-9-2401

"Speech Processors for Auditory Prostheses"

Prepared by

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

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

CONTENTS

I. Introduction

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

1. Evaluation of procedures used in our laboratory to determine processor parameters.
2. Further studies with Ineraid patient SR2, to evaluate parametric variations of the continuous interleaved sampler (CIS) and other processors.
3. Breadboarding and checkout of a prototype portable speech processor, based on the Motorola DSP 56001, for evaluation of the CIS strategy in field trials.
4. Presentation of project results at the *21st Neural Prosthesis Workshop*, held in Bethesda, MD, October 17-19.
5. Continued preparation of manuscripts for publication.

In this report we describe the evolution and refinement of procedures we use to determine optimal parameters for a particular processing strategy and patient (point 1 above). Examples include the fitting of single-channel processors for two patients with the auditory brainstem implant.

Complete descriptions of our work with multichannel subject SR2 (point 2), and of the development of the 56001-based portable processor (point 3), will be presented in future reports.

II. Parametric Variations and the Fitting of Speech Processors for Single-Channel Brainstem Prostheses.

Introduction

Repeatedly, our studies with multichannel cochlear implant patients have demonstrated that small parameter value changes in the fitting of pulsatile processors can make substantial differences in performance on tests of speech perception. It has become clear, both in the interest of better fittings in our own studies and in the interest of eventual effective application of the fruits of such research in clinical settings, that systematic ways of assessing the effects of parametric variations need to be developed.

The duration of a typical individual patient study in our laboratory has been only about five days. Such severe limitations on time with each patient have forced us to seek economies in the fitting process rather than do extensive, thorough evaluations of all--or even a significant number--of the possible parametric variations. Fitting procedure economies also will be necessary, of course, if the great potential of individually-tailored pulsatile processors is to be realized in clinical settings.

In the course of our patient studies to date, then, fitting strategies already have evolved to some extent in response to time pressures. With multichannel prostheses, our general approach has come to be the initial audition and performance testing of a variety of qualitatively different processing strategies until one is identified that seems to hold the greatest promise for the individual patient. Then, having adopted that basic strategic approach as a provisional standard, we typically evaluate a series of related processors, each representing the variation of a single parameter with respect to that constant reference processor. Once it is known whether each such variation produces an increment or decrement in performance, combinations of promising modifications can be applied simultaneously and the resulting processors evaluated in search of cooperative improvements in speech perception performance. When an individual variation has seemed to offer a really compelling improvement, however, in view of our limited time with each patient, an exception to this formal procedure has been allowed and a new provisional standard adopted for subsequent single-parameter manipulations. In some cases, depending on the remaining time, we have gone back immediately to repeat earlier variations with respect to the new standard.

This approach has been generally effective, and seemed reasonably sound and efficient in our hands. We have remained mindful, however, that there is no reason to assume that one can predict in any simple way, based on the effects of individual changes, what the effect of combinations of such changes will be. The changes are not independent of each other or functionally orthogonal in their measurable effects. A thorough parametric study is needed both to evaluate the efficacy of the techniques that we have been using, and also to consider what the prospects are for being able to establish safe, efficient, and reliable procedures for the fitting of such processors in the clinic.

We report here the first two stages of an increasingly systematic and thorough study of controlled parametric variations in ***single channel processor*** design and assessment. Compared to the extremely large number of possible channel-by-channel parameter manipulations for multichannel processors, the magnitude of the task involved in a systematic exploration of parametric variations for a single channel processor--while still a formidable and time consuming task--seems relatively tractable.

Subjects

Each of the two patients studied was restricted to the use of a single stimulating electrode. In both cases the electrodes were implanted at the surface of the ventral cochlear nucleus rather than in scala tympani of the cochlea. As background to these studies, we will present some data illustrating the systematic approaches that have evolved in our fittings of multichannel cochlear prosthesis patients. Thereafter the present report will deal exclusively with data from the patients with single-channel brainstem implants. The studies with these latter patients were conducted at Duke University Medical Center, in collaboration with DUMC audiologist Robert Wolford and with Dr. Robert V. Shannon of House Ear Institute (HEI).

Deafness in both patients was associated with the removal of VIIIth cranial nerve tumors that were part of clinical pictures of Type II neurofibromatosis (NF). In the case of the first patient, HB-1, a male aged 25 with a history of multiple neurofibromas from a relatively early age, electrodes were implanted during the second of two surgical procedures to remove bilateral acoustic tumors, interrupting the VIIIth cranial nerve on both sides. The case of the second patient, HB-2, a 40-year-old female, involved relatively late onset of NF Type II signs and no indication of neoplasms in addition to her bilateral acoustic tumors. Her electrodes were implanted during an initial surgery to remove the larger of the two tumors, severing the VIIIth nerve on one side. At that time a second surgical procedure already was planned to visualize the much smaller contralateral tumor, to remove it if possible without significant hazard to the adjacent intact nerve, and to decompress it if removal posed unacceptable risk [The latter course proved necessary.] Each patient received two active electrodes, but in both cases only one was found to allow stimulation of usable, purely auditory percepts. Indifferent electrodes were located in the *m. temporalis*. The implants for both patients had been performed in Los Angeles by HEI surgeons.

normal
residual
hearing
on
one side
(see
below)

Thus, HB-1 had a single available channel for electrical stimulation of the cochlear nucleus on one side and was totally deaf because of severed VIIIth cranial nerves bilaterally. HB-2 had normal hearing on one side and a single channel for electrical stimulation of auditory percepts via the cochlear nucleus on the other. The patients also differed markedly in experience and ability with an auditory prosthesis prior to our studies. HB-1 had several months of experience with a wearable processor resembling those used for 3M/House single-channel cochlear prostheses. HB-2 had no experience with a wearable processor. Testing of HB-2 in the HEI laboratories had been limited to the measurement of psychophysical functions. Thus the present study represented the first fitting of any speech processor for that subject. With normal hearing contralaterally, of course, HB-2 also lacked any experience as a profoundly deaf person, including any familiarity with standard speech perception test methods or experience relying on lipreading.

The hearing acuity of HB-2 was such that, even with an earmold used in conjunction with a circumaural sound barrier, quiet speech in the laboratory was easily understood through hearing alone. Special care was taken during all tests with this particular patient to eliminate conversation as well as any background noise that might have distracted her.

Patients HB-1 and HB-2 have been referred to, respectively, as ABI 12 and ABI 17 in published studies by HEI staff.

Methods

A very useful tool both for evaluating present laboratory techniques of speech processor optimization and also for assessing the prospects of various strategies for clinical fitting, would be a systematic mapping of a multidimensional parametric space in terms of one or more indices of speech comprehension performance. Having such a map would allow *post facto* consideration of the paths through that parametric space taken by various fitting strategies. Such strategies might include varying parameters singly with respect to a fixed provisional standard processor, or allowing a provisional standard to evolve by the immediate incorporation of beneficial variations as they were identified.

Not only might we test such alternative fitting strategies beginning with the same starting processor, but also assess how each strategy would fare depending on the choice of initial settings, and how sensitive the eventual outcome and the course of the selection process would be to that choice. [Appendix A to this report is a preliminary examination of several strategies in terms of two assumed performance index models in two dimensions.]

The results of such a study could be very informative as to the prospects for development of intelligent clinical fitting systems. Other results of such a study might include the identification of some strategies of choice, some starting conditions that might be particularly efficient, or some time-saving preliminary studies to be undertaken before beginning a more systematic parametric evaluation (in order, for instance, to avoid focusing on some minor local maximum of performance or adopting a highly inappropriate provisional standard).

w.r.t. what?

Tests

Choosing a test instrument for the rapid ranking of the performance of alternative speech processor configurations is not the same as choosing an instrument for documenting a patient's level of speech perception, nor is it clear *ab initio* that any single set of recorded speech samples can be suitable for both purposes. If the same recorded materials were suitable for both, of course, the efficiency of the fitting process would be enhanced in that the testing already done with a processor during its selection as the optimum configuration would constitute a head start on the documentation of its performance.

We have come to rely on consonant identification tests both as our primary tool for rapid choices among processors and as one element, along with the open set portions of the Minimal Auditory Capabilities battery, in documenting the levels of speech perception achieved with selected processors. [Correlation coefficients between overall information transmission in our consonant identification tests and scores for the four open set MAC subtests have been found to be as follows: Spondee recognition .86, CID sentence keyword recognition .70, SPIN sentence final word recognition .83, and NU-6 monosyllabic word recognition .79. The CID test correlation is significant at $p < .005$ while each of the other correlations is significant at $p < .001$. For more details see Wilson *et al.* 1990b.]

Early in our studies of alternative speech processors for cochlear prostheses we used a set of only eight consonants, along with a set of five vowels. We produced digital recordings of those thirteen speech tokens, so that the stimuli appropriate to candidate processor designs could be calculated offline for subsequent testing. Patients were allowed multiple repetitions of each token before identifying it by pressing a key on a computer terminal. Computer screen feedback as to the correctness of each response was optionally available. When visual cues were required for evaluating the effectiveness of processors used in conjunction with lipreading, one of the investigators mimed the tokens in synchrony with the calculated stimuli.

The size of the token sets was kept so small in order to minimize the time required to obtain an indication of relative processor merits. A videotaped fourteen-consonant test [Tyler *et al.* 1983] was used once a best processor was identified, for purposes of documenting patient performance. Our subjects found its pace excruciatingly slow.

When similar consonant identification materials became available on videodisc [Tyler *et al.* 1987], we produced a multi-randomization test version with the fourteen consonant set and a corresponding test with a set of eight vowels. Only a single presentation of each token was allowed before each patient response and feedback was eliminated as an option. At about the same time we completed the development of programmable laboratory speech processors capable of operating in real time, with no need for calculations in advance. We anticipated, on grounds of efficiency, that direct entry of responses by the patient into the computer would continue to be our consonant test method of choice. Because a number of patients had visual as well as hearing impairment, however, an alternative method was devised in which the patient responded vocally and the experimenter entered the response into the computer. To minimize the possibility of misinterpretation of a response by an investigator, the patient's vocal response was a numeric label, rather than a repetition of the presented token itself. A table of the speech tokens and the corresponding numeric labels was displayed to the patient between presentations on the same video monitor used to display the optional lipreading cues.

We were surprised to discover that in general our patients preferred the vocal response method. Moreover, they completed significantly more tests per unit time with that method and withstood longer periods of testing each day before showing signs of fatigue. We attribute these advantages to several factors, including the ability to adopt a variety of comfortable postures during testing, reduced visual strain in the absence of any need for the patient to redirect his or her glance or refocus between tokens, and the maintenance of direct human contact and communication throughout testing rather than the patients' being "turned over to the computer" for extended periods. The pace of testing with this method was such an improvement that we increased the consonant token set to sixteen for all our testing, including rapid processor screening. [Our original eight-consonant tests typically required six to eight minutes for three presentations of each token, or about 17.5 seconds per response. The videotape test with five presentations of each of 14 consonants required about 20 seconds per response. Our current 16-consonant tests, with five presentations of each token, typically are completed in ten minutes, or about 7.5 seconds per response.]

why?

While we have continued to administer vowel identification tests, we regard them as much less helpful in optimizing processor performance and will concentrate only on consonant tests in this report. [In contrast to the situation described above for the consonant identification tests, there is no significant correlation between overall information transfer in our vowel identification tests and any of the open set subtests of the MAC battery. The corresponding correlation coefficients are: Spondee recognition .32, CID sentence keywords .27, SPIN sentence final words .27, and NU-6 word recognition .24. For more details see Wilson, *et al.* 1990b.] The token set used for all the tests referred to in the sections that follow was edited from the Iowa videodisc [Tyler *et al.* 1987] to include two different presentations each by male and female talkers of each of the sixteen consonants in an /aCa/ context. The consonants were m, n, f, v, s, z, sh, th [voiced], p, b, t, d, k, g, j, and l. In editing the tokens from the Iowa videodisc we sought (1) uniform delays between playback onset and the beginning of speech sound (2) uniform overall length of playback, and (3) absence of obvious extraneous clues, either visual [eye blinks and glances, preparatory lip and head movements, *et c.*] or audible [background noise, preparatory lip smack, *et c.*]. Obtaining two exemplars for each consonant proved impossible under criterion (3), so we resorted to introducing a controlled number of

[illegible] space

extraneous visual events in a pattern designed to minimize possible contamination of the results.

While the consonant identification tests were being administered, the experimenters had access to several ongoing displays in the form of matrices. One was a sixteen-by-sixteen confusion matrix that could be color coded in a variety of ways to signal different classes of errors. Other displays included smaller matrices indicating the ability of the processor being evaluated to support distinctions between voiced and unvoiced consonant sounds, among various places of articulation, and among various manners of articulation. The vocal response method also allowed the experimenters more insight into changes in patients' alertness and fatigue, and a strong impression of different ways in which individual patients approached and dealt with these tests.

Some patients obviously were quite analytical in their approach to consonant identification tests. They would reason their way to a response when they were not sure, striving to hone their skills with each new processor and recognize ever more subtle distinctions among the tokens. Such an approach might raise concerns in that there would not be sufficient time to apply it in the "real world" situation of continuous connected speech and that scores obtained from such a patient would not correlate with the ability to perceive such speech.

Other patients, on the other hand, preferred simply to "say what they heard" and guess with apparent abandon when uncertain. A possible concern with this approach might be one of prolonged learning curves early in a patient's experience with a new processor.

While all of the patients we have studied shared a strong motivation to improve their levels of speech perception, some were much more overtly competitive in their approach to the tests while others behaved in a much more neutral manner, accepting our assertion that it was really the consonant *mis*identifications that helped us learn how to improve the prostheses. Early in our investigations with cochlear prostheses, providing feedback to certain highly competitive patients greatly increased their level of engagement in the tests and their rate of test completion.

Our general impression in the face of all these variations in patients' approaches to the consonant identification tests, was that they provided a reliable basis for ranking and optimizing speech processors for the whole range of patients. As suggested above, however, there was concern at times as to the relative uncertainties in the results for different patients and for single patients as a function of length of experience with each processor. These issues are addressed further at a later point in this report.

Background: Multi-Channel Parameters and an Example of one Fitting Strategy

The potential number of significant adjustable parameters for the kinds of multichannel cochlear prostheses we have been studying is quite large. As an indication of that fact, we present here a list of some of the parameters that we have varied systematically in the course of fitting various speech processors to individual patients. In many cases values of these parameters could be assigned individually for each channel of a multi-channel prosthesis.

Multichannel Processor Parameters

number of pulsatile channels utilized

number of pulsatile channels stimulated in each cycle

number of continuous analog channels utilized

frequency band associated with each channel

input signal equalization:
  - *slope*
  - *frequency setting*

envelope estimates for each channel:
  - *halfwave vs. fullwave rectification*
  - *frequency response of integrating filter*

order of pulsatile channel stimulation within each cycle:
  - *base-to-apex*
  - *apex-to-base*
  - *random*
  - *staggered*
  - *custom, to minimize specific channel interaction(s)*

electrode configuration:
  - *monopolar*
  - *bipolar*
    - *longitudinal*
    - *radial*

tonotopic assignment of channels to electrodes

pulses for each pulsatile channel:
  - *type:*
    - *balanced biphasic*
    - *asymmetric, but charge-balanced*
  - *leading phase:*
    - *anodic*
    - *cathodic*
  - *pulse duration per phase*
  - *delay between pulses within a stimulus cycle*
  - *additional delay between stimulus cycles*

dynamic range of each channel:
  - *loudness percept corresponding to maximum stimulation level*
  - *shape of amplitude mapping function*

explicit voiced/unvoiced coding in pulsatile stimulation cycle timing:
  - *pitch rate/max rate*
  - *pitch rate/jittered rate*
  - *none (constant rate)*

We shall not, in this report, discuss the neurophysiological models that provide the bases for structuring choices within this sea of options. As a simple example of rapid assessment of the effects of varying such parameters in the fitting of a speech processor, however, we shall examine briefly part of an actual multichannel fitting sequence from one of our studies. The processor being optimized was of the interleaved pulses (IP) type [Wilson *et al.* 1989], and the three parameters being varied were the number of channels being stimulated, channel order within each stimulus cycle, and the use of cycle initiation timing to signal the distinction between voiced and unvoiced segments of speech [Wilson *et al.* 1990a,c]. Each overall percent correct score is based on ten presentations of each of sixteen consonants in /aCa/ context, with sound alone, by either a male or female talker. The logical sequence of the fitting is outlined in Figure 1.

The initial choice of a provisional standard configuration was six channels stimulated in base-to-apex order in cycles timed to provide a rough, imperfect indication of the voiced/unvoiced distinction. Values of many of the other parameters from the list above had

Initial standard processor			Overall Percent Correct		Conclusions
channel update order	number of channels	v/uv coding	Male Talker	Female Talker	
base-to-apex	6	rough	94	69	

(differences wrt standard shown in **boldface**)

order of comparisons

channel update order	number of channels	v/uv coding			
apex-to-base	6	rough	91		worse
base-to-apex	**4**	rough	69	63	worse
base-to-apex	6	**none**	92	74	better for female voice
base-to-apex	6	**improved**	82		worse
random	6	**improved**	90	69	combination not as bad
random	**4**	**improved**	81	69	only worse for male voice
staggered	6	**improved**	94	71	another candidate male/female compromise

Figure 1.

been chosen on the basis of previous experience with similar processors and the results of a brief psychophysical assessment of the particular patient.

Each of the first three steps in the optimization sequence involved changing the value of a single parameter with respect to the same initial configuration. Reversing the channel order within each stimulus cycle and reducing the number of channels from six to four both produced decrements in performance on the consonant recognition tests. When the rough approximation to explicit voiced/unvoiced coding was eliminated, though [by having cycles succeed each other at a constant rate whether or not the speech had a recognizable voice pitch] a substantial improvement in recognition of the female talker's consonants accompanied a marginal decrease in scores for the male talker. This led to a more careful implementation of voiced/unvoiced coding, with stimulus cycles initiated at the voice pitch rate when voicing was present and at a fixed higher rate when no voice pitch could be detected. When that modification alone was made to the initial standard processor the result was substantial damage to the recognizability of the male talker's consonants.

Since a demonstrably imperfect indicator of voicing had been part of a rather good processor, it was decided to explore use of the improved voicing detection in combination with other parameter variations. Randomizing the channel order within each cycle was found to allow the use of the improved voicing indicator with much less damage to male consonant understanding, and when a four-channel version of that combination was tried only the scores for the male talker were reduced.

Clearly, the recognition scores were quite sensitive to channel stimulation order within each cycle and the randomized ordering seemed to work better than monotonic sequences in either direction. On the assumption that reducing instances of sequential stimulation of adjacent electrodes was the beneficial feature of the random ordering, we tried a version that combined six channels, the improved voicing indication, and a "staggered" channel order, *to wit* 6-3-5-2-4-1. This processor preserved an overall base-to-apex progression but never stimulated adjacent electrodes sequentially. [This design is consistent with a neurophysiological model in which non-simultaneous interactions between adjacent channels could reduce the saliency of place-of-stimulation information.] The resulting male talker score was equal to that of the initial standard, with the score for the female talker intermediate between that of the standard and the best yet observed.

Had there been time, it clearly would have been desirable to explore other parametric variations of this processor. Limits on access to the patient, however, dictated that a choice be made at this point, and the decision came down to one between two configurations offering slightly different tradeoffs in male/female consonant scores: 92/74 or 94/71.

Is this politically volatile?

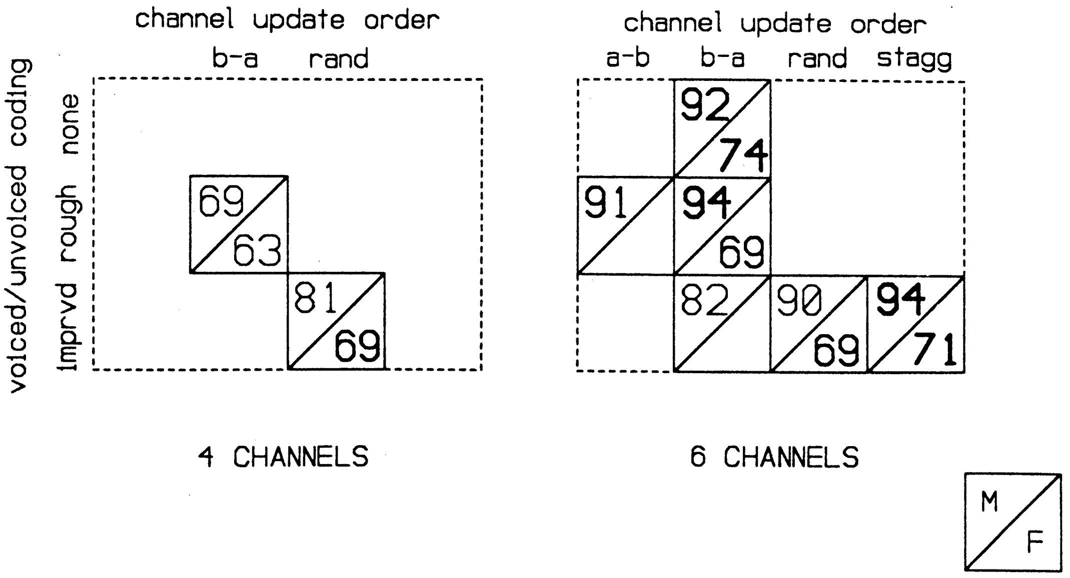

channel update order
b-a
rand
voiced/unvoiced coding
none
rough
imprvd
69
63
81
69
4 CHANNELS
channel update order
a-b
b-a
rand
stagg
92
74
91
94
69
82
90
69
94
71
6 CHANNELS
M
F

Figure 2.

Figure 2 displays the same overall percent correct scores in cells within a parametric space. Each varied parameter is represented as a spatial dimension--channel order along the x direction, voiced/unvoiced coding along y, and separate grids for each number of channels. These grids, when stacked as overlays, correspond to different values along the z axis. The highest scores are highlighted by successively bolder typefaces. While in general some of these parametric dimensions might be expected to correspond simply to different values of classic scalar parameters [*e.g.* pulse duration or pulse separation], our present example highlights other possibilities--rigidly bounded, coarsely quantized, and/or qualitative variations in processor design. [An absolute limit of six channels was imposed in the present example, for instance, because the patient possessed only six electrodes. While additional channel update orderings were certainly possible, none was expected to differ significantly from these four models. While many other modes of *conveying* a voiced/unvoiced distinction could be proposed, no still better way of *detecting* the distinction within these test materials was (or is) known.]

need to know if this is significant Δ

Note how little of this limited three-dimensional parametric space was explored in the course of our fitting strategy--half the cells in one two-dimensional plane and only two cells that differ in the third parametric dimension. Notice also the apparent separation of our two most promising candidate processors (92/74 and 94/71) in the parametric space, suggesting the existence of at least two distinct local maxima in processor performance. [The relative positions of random and staggered ordering along the x direction are, to be sure, arbitrary. The second performance index model examined in Appendix B is one possible topology consistent with the 6-channel plane of these data.]

The data of Figure 2 are repeated at the upper left in Figure 3, along with similar parametric topologies for three other indications of processor performance. These summaries of the three other matrix displays available to the experimenters during test administration are based on the same consonant identification tests. There are certainly no significant distinctions across this parametric space in terms of correctly identifying whether voicing is present in a consonant. In contrast, the percent correct place of articulation and manner of articulation score topologies do correspond with the overall percent correct landscape.

In the context of all these provisional judgments of one processor's supporting speech recognition significantly better than another, it would be nice to know the statistical uncertainties in each score. In our screening situation, with strong pressures to minimize testing time per configuration, there are many factors that might affect such uncertainties, *e.g.* the statistical significance of the number of presentations of each token in the particular condition, the consistency of performance characteristic of the individual patient, and the ongoing rate of learning of the patient with the particular candidate processor.

Also sources of distortion or incorrect settings of AGC, dynamic range etc

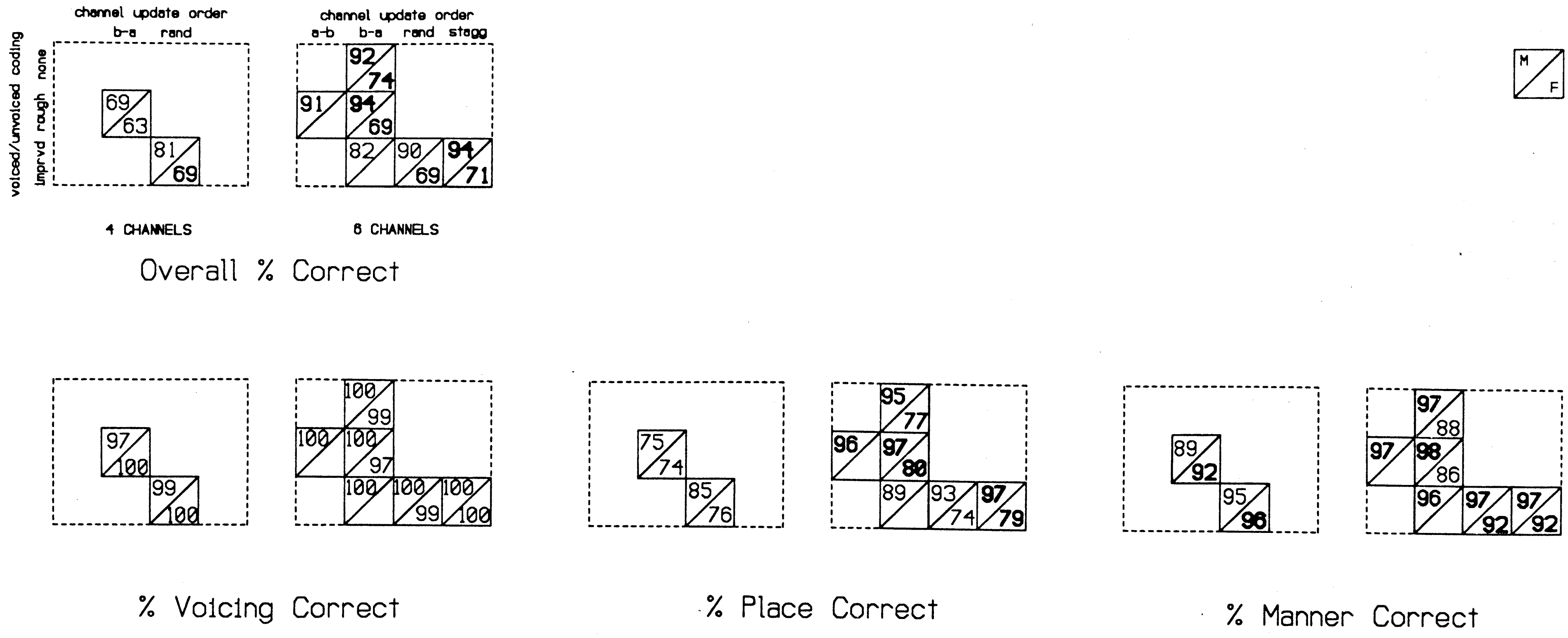

channel update order
b-a
rand
voiced/unvoiced coding
imprvd
rough
none
69
63
81
69
4 CHANNELS
channel update order
a-b
b-a
rand
stagg
92
74
91
94
69
82
90
69
94
71
8 CHANNELS
Overall % Correct
M
F
97
100
99
100
100
99
100
100
97
100
100
99
100
100
% Voicing Correct
75
74
85
76
95
77
96
97
80
89
93
74
97
79
% Place Correct
89
92
95
96
97
88
97
98
86
96
97
92
97
92
% Manner Correct

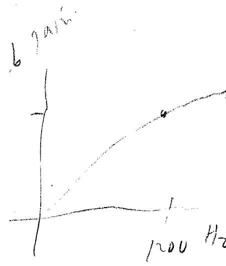
Figure 3.

Ten presentations of each of the sixteen consonant tokens were represented in each score we quoted in this example. This consisted of two administrations of an 80-token test (five presentations of each consonant) with different randomizations. Within each test all 80 tokens were randomized, rendering statistical or time-dependent analysis on any finer scale impossible. As our fitting procedures evolved to require both more and finer distinctions, we decided to revise our consonant identification test randomizations so that the nth presentation of every token would have taken place before the (n + 1)th presentation of any token occurred. This meant that each 80-token test also could be analyzed as a time sequence of five separate and complete presentations of the full 16-consonant set. Each 16-consonant subtest then could be analyzed statistically as a separate observation of performance, and learning effects could be evaluated on a much finer time scale.

Studies with Subject HB-1: An Example of an "Evolving Standard" Strategy

Both in the fitting strategy and in the nature of the parameters being varied, our studies with single channel brainstem patient HB-1 complement the methodological picture begun above in our background example of a multichannel cochlear prosthesis fitting. The strategy was one of a continuously evolving standard processor, into which beneficial parametric changes were adopted as soon as they were discovered, with all future comparisons being with respect to the evolving standard rather than to a common reference processor. All five varied parameters in the studies with HB-1 represent continuous scalar quantities: two each with units of time and frequency and one along the dimension of a perceptual loudness scale. All the consonant identification tests we will be discussing for HB-1 were conducted with sound alone, *i.e.* no visual cues were provided. All scores presented in the remainder of this report, for both patients, will be for a male talker only.

The logical sequence of the fitting procedure for HB-1 is outlined in Figure 4. The initial condition was a processor with 110 microsecond per phase balanced biphasic pulses separated by 330 microseconds. The speech signal input to the processor was equalized by a 6 dB per octave pre-emphasis, with a 1200 Hz 3-dB reference point for the high-pass filter involved. An amplitude envelope was derived through rectification and low pass filtering of the equalized speech signal. The low pass filter had a cutoff frequency of 400 Hz and acted to "smooth" or integrate the rectified signal. The resulting time-varying envelope was mapped onto a range of stimulus pulse amplitudes corresponding to a range of 0 to 5 on a scale of perceived loudness (with 0 as threshold, 5 as most comfortable, and 10 as upper limit for comfortable listening).



The first variation tested was a reduction in pulse separation from 330 to 55 microseconds. [The comparison made at each point in the logical sequence is indicated by a pair of parametric values in the appropriate column of Figure 4, separated by a slash. These compared values corresponded to the two overall percent correct scores at the right, in the same

logical step	pulse separation	smoothing frequency	range mapped	pulse duration	EQ limit	overall percent correct	conclusions
Initial Condition	330 us	400 Hz ↓	5 ↓	110 us/ph ↓	1200 Hz ↓	?->	
Reduce separation	330/55					50/64	reduced separation better
Raise/lower smoothing frequency	55 ↓	400/800 400/200				64/59 — 76) see p. 19 64/60	
Eliminate separation	55/0 55/0	800 200				59/61 60/58	zero separation almost as good if coupled with 800 Hz smoothing
Adjust dynamic range	0 ↓	800 ↓	5/3 5/7 7/8			61/64 61/64 64/59	0-7 perceptual range per channel best
Reduce duration			7 ↓	110/40		64/69	shorter better
Adjust EQ limit				40 ↓	1200/850 1200/1700	69/51 69/64	either alternative worse

Figure 4.

- It is not at all clear why ∆ in pulse separation would ∆ results unless indirectly through ∆ threshold.

order. Vertical arrows indicate parameters that did not change from comparison to comparison.] The reduced separation time produced substantially better consonant identification scores and, thus, was immediately incorporated into the evolving processor design.

In the next two comparisons, either raising or lowering the smoothing frequency produced a comparable decrease in consonant identification. Further trials in both these conditions but with pulse separation further reduced to zero found that the combination of zero separation and an 800 Hz envelope response limit gained a significant amount of potential bandwidth for information conveyance with a relatively modest reduction in consonant identification scores (from 64 to 61 percent overall). Thus zero pulse separation and the 800 Hz cutoff became part of the evolving optimized processor. ←?

? 59 (see fig 4)

The next parameter to be varied was the perceptual dynamic range onto which the smoothed energy amplitude envelope was to be mapped. Both expanding and contracting this range by two perceptual scale units produced comparable improvements in consonant identification. Since wider dynamic range seemed the more desirable, all other things being equal, the next trial variation was a further expansion, locating the apparent limit of beneficial evolution along this parametric dimension at a 0-to-7 perceptual range.

sounds like we are on a slant surface

Since reduction of pulse separation had been beneficial, and its elimination--when combined with providing more high frequency envelope structure--had caused little damage to performance, a logical next step was to further increase the stimulus pulse rate by reducing the pulse duration per phase. A reduction from 110 microseconds to 40 resulted in a consonant recognition score improvement from 64 percent to 69. The final variations attempted in the limited time available with patient HB-1 were a decrease and an increase in equalization filter limiting frequency, both of which were found to degrade processor performance on the consonant tests.

Thus the processor that emerged as the product of this fitting procedure incorporated changes in four of the five parameters varied. The accompanying improvement in overall consonant recognition scores was from 50 to 69 percent correct without visual cues.

was initial setting we tried at end of experiment

An understanding of the relative range of parametric space explored in this fitting sequence may be gained from the diagram in the upper part of Figure 5. This diagram is similar in many respects to that introduced in Figure 3, except that (1) only overall percent correct scores are displayed, (2) only male talker data are tabulated, and (3) five parametric dimensions are represented rather than just three. In Figure 5 each of the dashed rectangles is a plane in the same two parameters--smoothing frequency limit (x) and perceived loudness range (y)--but at different levels in the other three parametric dimensions--EQ limit frequency, pulse duration, and pulse separation. In other words the rectangles are overlays that can be

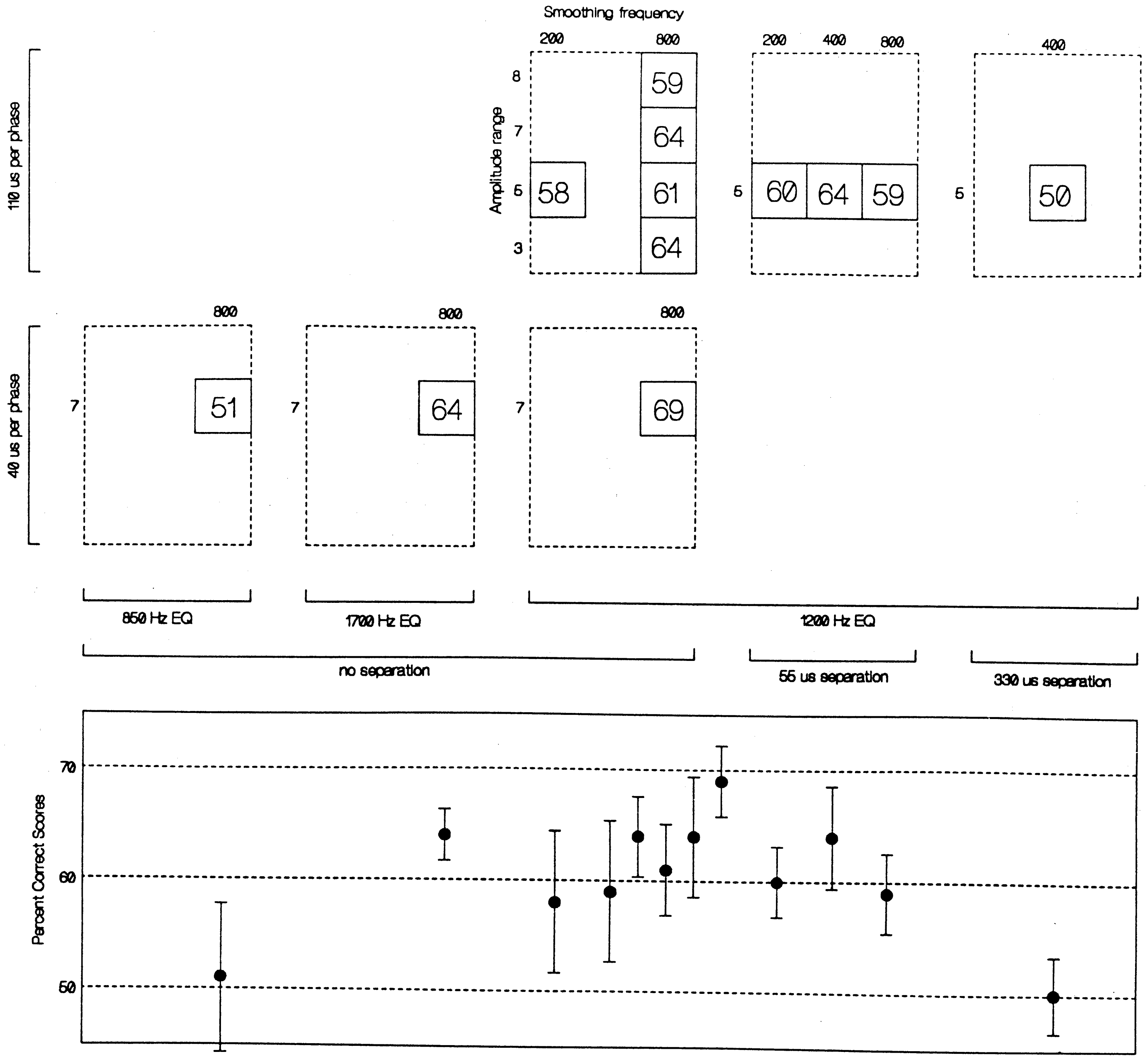

Smoothing frequency
200
800
200
400
800
400
Amplitude range
8
7
5
3
5
5
59
64
58
61
64
60
64
59
50
110 us per phase
40 us per phase
800
800
800
7
7
7
51
64
69
850 Hz EQ
1700 Hz EQ
1200 Hz EQ
no separation
55 us separation
330 us separation
Percent Correct Scores
70
60
50

Figure 5.

assembled in various ways depending on which third parameter is to be represented along the z axis.

Again, the relatively small region of parametric space explored by the fitting sequence is obvious. Another useful way of comparing this diagram to that of Figure 2 is to look for linear explorations along a single parametric dimension. In Figure 2 there were two such sequences of three adjacent conditions and one of two explored along the x and y directions, and two sequences of two adjacent conditions each along z. In the case of HB-1's fitting we have single sequences of three and four adjacent conditions respectively in the x and y dimensions. Stacking the overlays in the upper row, with pulse separation as the z dimension, demonstrates three distinct sequences of two adjacent conditions each in that direction. Overlaying the three planes in the second row makes the EQ limit frequency the z-axis parameter and highlights a three adjacent condition exploration along that dimension. Finally, overlaying the two planes in the third column shows a two-condition comparison along a pulse duration dimension. All told, this fitting sequence embodied 11 comparisons of processors differing only by a single step in a single parameter. This compares with a total of 7 such comparisons in Figure 2.

The presentation of consonant tokens in groups of sixteen, each a complete set as outlined at the end of the background section, allowed us in this case to characterize the relative uncertainties in these overall percent correct scores. If we consider each administration of a complete set of sixteen tokens as a single independent observation, the mean of all such observations available will be the overall percent correct score as before, but now with the possibility of calculating the standard deviation of that mean. Such standard deviations are displayed at the bottom of Figure 5 as error bars on points corresponding to the mean scores. The points are grouped under their respective cells on the parametric space diagram, with multiple cells from top to bottom in the same column being represented by a sequence of points from left to right under that column. Each point here represents from five to fifteen presentations of each of the sixteen consonants. The error bars indicating standard deviation of each mean vary considerably, depending also on the consistency of the patient's performance with each new processor. Clearly, there is substantial uncertainty involved in some of the pairwise comparisons in this fitting procedure. On the other hand, seven of the remaining eleven processor configurations show no overlap of standard deviation ranges with the highest scoring one, and only one other processor's error bar overlaps the best mean value.

While the choice of an optimal processor from among those explored seems clear in this case, note that a wide range of possible overall percent correct topologies could be consistent with these limited data. Some of those possibilities would provide benign environments for simple and efficient fitting strategies, while others would not.

Studies with Subject HB-2: A Pilot Systematic Study of a Parametric Space

Studies with the second brainstem implant patient, HB-2, represented our first attempt to explore a region of parametric space systematically. Our objective was to explore the region surrounding an overall percent correct maximum, rather than follow a specific fitting strategy. Thus, there is no logical sequence of tests to display in this case and we move immediately to Figure 6, a description analogous to that of Figure 5 for the previous patient. The parametric spaces of these two figures are more closely related to each other than to our original multichannel fitting example. In both single-channel cases a total of five parametric dimensions were explored. Three of these dimensions were the same for both patients: pulse duration, pulse separation, and smoothing frequency. Variations in the range of perceived loudness over which pulse amplitudes were mapped in HB-1's case were replaced in the studies of HB-2 by parametric variations in the mapping law that determined stimulus amplitudes between perceptual threshold and most comfortable loudness level. Finally the equalization frequency limit explored at the end of HB-1's fitting was replaced by a binary parameter representing the choice between full- and halfwave rectification of the signal from which time dependent energy estimates were derived. In the case of this implant patient the speech processor was being optimized as a single channel aid for use in conjunction with lipreading. Accordingly, all the consonant identification tests were conducted with visual cues.

The mean scores and their standard deviations are displayed below the parametric space diagram as points with error bars, just as in Figure 5. While the number of presentations of each consonant was constant throughout the central (102 microsecond duration) row of the diagram, it was allowed to vary elsewhere in the studies, constituting an additional source of variation in the magnitude of the error bars. While the error bar display indicates a level of discrimination among processors that is much improved overall with respect to Figure 5 [only three other processors' mean score error bars overlap at all with that of the highest], there still are specific comparisons of mean scores that would benefit from smaller uncertainties. Based on our experience with the study of HB-2 and the use of standard deviation of the mean calculations, the consonant recognition test programs have been modified for future studies to keep the experimenters constantly informed as to the statistical uncertainty, based not only on the current sequence of token sets but also on any previous testing done under the same conditions. [Examples of the modified experimenter console displays are included in Appendix B.] This should allow a more efficient use of limited patient access time, achieving statistically valid distinctions among candidate processors but not performing unnecessary tests merely to fulfill an arbitrary quota of presentations.

The increased thoroughness with which this parametric space was explored is evident in Figure 6. If we count the sequences of adjacent processors examined along each parametric axis as was done for the HB-1 case, we find for the smoothing frequency parameter in the x direction three sequences of three cells each and one of two, for a total of seven single-

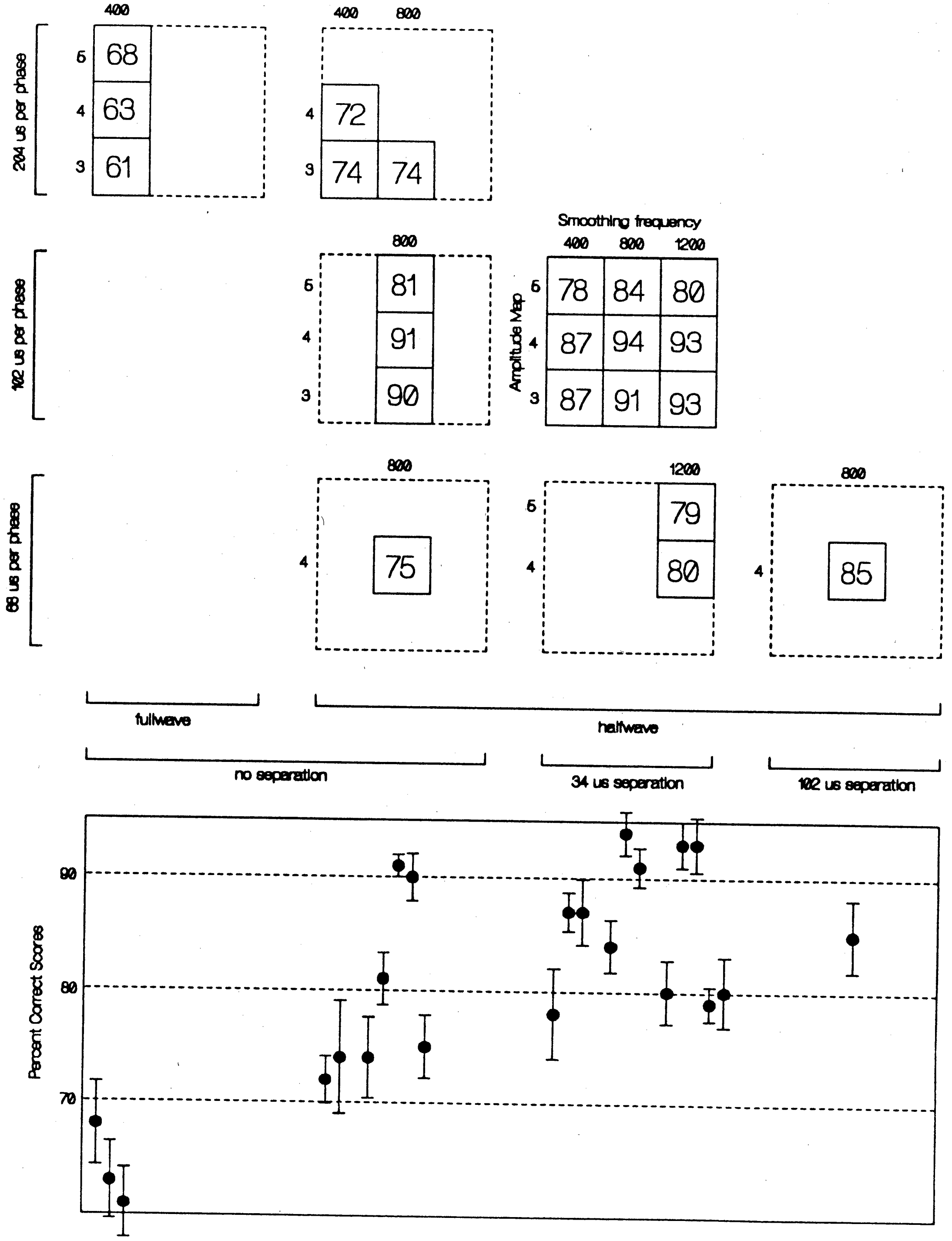

204 us per phase
102 us per phase
68 us per phase
400
800
1200
5
4
3
68
63
61
72
74
74
81
91
90
Smoothing frequency
Amplitude Map
78
84
80
87
94
93
87
91
93
75
79
80
85
fullwave
halfwave
no separation
34 us separation
102 us separation
Percent Correct Scores
90
80
70

Figure 6.

difference comparisons. Along the y axis, for the mapping law parameter, we explored five sequences of three processors and two of two, for a total of 12 comparisons. Combining the two overlays in the top row produces a pair of single-difference comparisons involving the full- and halfwave rectification distinction. Stacking overlays within a single column makes pulse duration the z axis. Here four adjacent processor pairs contributed another four single-difference comparisons. Along the pulse separation dimension we explored three two-cell sequences (middle row, second and third columns of overlays). If we include the comparison of 102 microsecond separation to none in the third row, second and fourth columns of overlays, the grand total of single-difference comparisons in these studies comes to 29. (A fully explored three-dimensional cube three cells long in each dimension would provide 54 such comparisons.)

In general terms the overall percent correct topology is very well behaved over the explored regions of this parametric space. There is no evidence of multiple local maxima or other features that might be expected to complicate the design of fitting strategies for efficient convergence on an optimal processor configuration. Within this (admittedly, still very limited) *terra cognita*, none of our previously discussed strategies would have been led astray. And, insofar as we can judge from the region explored, this would have been true regardless of the choice of starting condition. This finding is reassuring in that it supports the validity of the fitting strategies that have evolved in the course of our previous research. It also indicates the appropriateness of proceeding to a still more extensive and thorough level of parametric exploration. [A provisional coding form for routinely recording various speech test results in terms of each processor's location in parametric space is included in Appendix B.]

A brief inspection of Figure 6 will indicate that a search for the optimal processor design should focus on the center row of overlays, the processors that share a 102 microsecond-per-phase pulse duration. Those overall percent correct score cells are reproduced in row (1) of Figure 7, except that the zero separation overlay has been placed on the right rather than the left. As was the case in Figure 3, the highest scores have been highlighted with progressively bolder typefaces and the same pattern of cells has been repeated as necessary to facilitate comparison of a variety of data across the same group of processors. A set of labels for the cells has been included on row (1) to allow more economical descriptions of subsets in the discussion that follows.

In seeking to identify the optimal processor, a strong pattern in the overall percent correct scores (row (1) of Figure 7) supports elimination of the map 5 row of cells (g-j) and the 400 Hz smoothing column (e-g). Row (2) of Figure 7 displays scores for the percent of responses that were consistent with each of three characteristics of the presented consonant, the same quantities included in Figure 3 for our multichannel example fitting. While the percent voicing correct scores display the same pattern as the overall scores, the situation for percent place correct and percent manner correct is not so clear cut, especially for cells h and j.

(1)

Smoothing Frequency

Amplitude Map	400	800	1200	800
5	78	84	80	81
4	87	**94**	**93**	**91**
3	87	**91**	**93**	90

34 us--separation--none

Overall % Correct

g	h	i	j
f	a	b	k
e	d	c	l

(2)

87	89	91	90
95	**97**	**97**	**99**
93	**97**	**99**	**98**

% Voicing Correct

96	**100**	96	96
97	**99**	97	95
96	**98**	**99**	96

% Place Correct

92	95	92	**98**
95	**100**	**100**	97
94	97	**99**	96

% Manner Correct

(3)

47	58	58	56
74	**81**	**83**	**89**
70	79	**89**	**89**

Voicing % IT

44	64	47	82
76	**100**	**93**	**100**
90	**90**	**93**	**100**

Nasality % IT

82	87	77	82
81	**89**	**91**	83
82	**88**	**91**	83

Place % IT

77	**85**	69	**85**
75	**95**	**95**	80
71	80	**87**	74

Frication % IT

67	74	72	77
100	80	**100**	**92**
90	71	81	87

Duration % IT

64	71	69	78
88	**91**	**92**	**94**
85	**91**	**94**	**94**

Envelope % IT

82	87	81	83
81	**90**	**92**	84
82	88	**91**	86

Viseme % IT

(4)

6	5	4	

m/b confusions

1	2	1	2

d/z confusions

3		2	4
4		1	
1	1	1	

l/n confusions

(5)

4	3	2	6
4	4	7	8
5	4	3	5

s/z/sh confusions

10	10	12	9
8	5	1	6
6	7	4	6

j confusions

Figure 7.

The two rows labeled (3) in Figure 7 contain percent information transfer results based on analyses of seven attributes across our sixteen consonant tokens [Miller and Nicely 1955, Wilson *et al.* 1988, Lawson *et al.* 1989]. Information transfer for the voicing attribute displays the same pattern quite strongly, as do the results for place and envelope. The pattern for the nasality attribute is less strong with respect to eliminating the e-g column, while the frication and viseme patterns are less convincing with respect to row g-j. But only in the case of duration information transfer for the e-g column is there a distinctly different pattern. [The percent transfer of duration information also is the only quantity displaying two distinct local maxima.]

The surviving candidates for optimal processor, in view of the strong pattern noted above are those corresponding to cells a-d, k, and l. Within this group and in terms of the information transfer data, the cell d processor performance is notably low for five of the seven attributes. The processors associated with cells k and l transmitted voicing, nasality, and envelope information particularly well, but were notably weak in the areas of place and frication. The processors of cells a-c emerge as the ones with consistently excellent performance.

To select among the remaining three candidate processors we turn to row (4) of Figure 7 and patterns of specific consonant confusions. The numbers displayed in these cases are the number of instances in which particular errors were made. Certain errors occurred only for a particular parameter value. Patient HB-2 had difficulty discriminating between the consonants m and b *only* when amplitude map 5 was used. A more complex error--confusion between the consonants d and z--occurred *only* in association with map 3, providing a basis for eliminating cell c as a candidate in the absence of any stronger indication. Another particularly interesting pattern of specific confusions was that between the consonants l and n.

We note that such highly specific row and column error patterns support the validity of fitting strategies based on the effects of variations in single parameters.

To make a final distinction between cells a and b in choosing an optimal processor for HB-2, we turn to row (5) of Figure 7 and two classes of consonant confusions that posed special problems for that patient. We note that the processor associated with cell b was the only one that really distinguished the consonant j from a number of other sounds. Competing with this observation is the fact that cell b is a particularly problematic one with regard to discriminating among the consonants s, z, and sh. [Notice, by the way, that the performance of the processor of cell k is worse than either a or b in both cases.] The patient's ability to hear differences among s, z, and sh in side-by-side comparisons using the processor of cell b, as well as that processor's high information transfer scores for voicing and frication, suggest that HB-2's discrimination among those consonants would improve after practice with that processor. When making such final choices among apparently equivalent processors, identification of specific consonants that a patient can distinguish in a/b comparisons, but not (yet) reliably identify in tests involving all 16 consonants, may help improve the initial choice of speech processors for long-term use.

We note that lines of reasoning like that just outlined for identifying the best candidate processor for this patient (1) appear suitable for employment in "intelligent" computer fitting systems and (2) might not occur to a clinician who fitted such devices only occasionally.

Comput might do better ☺

Conclusions and Outline of Future Studies

While our choice of a consonant identification test as *the* gauge of performance guiding speech processor fitting was somewhat arbitrary, it seems to have done rather well. Consonant identification tests can be administered rapidly and repeatedly. The statistical significance of averaged scores from repeated tests can be assessed. Further analysis of the results in terms of the transfer of various types of information is straightforward and helpful in resolving fitting choices. Performance on consonant identification tests correlates well with such assessments of speech perception as the open-set tests of the standard Minimal Auditory Capabilities battery.

With regard to the first two stages of our exploration of processor parametric spaces, the results thus far support the validity of the fitting strategies that have evolved in the course of our experience optimizing a variety of processor designs. No anomalies have been found in consonant identification score topologies that would seriously limit their usefulness as a fitting gauge. Using the improved statistical monitoring tools and experience gained from these studies, our next step is a finer grained, more precise, more thorough, and more extensive exploration of parametric space.

Acknowledgements

We thank the subjects of these studies for their generous contributions of time and interest.

References

Box GEP, Hunter JS: Multifactor experimental designs for exploring response surfaces. Ann Math Statist 1957,**28**:195-241.

Lawson DT, Wilson BS, Finley CC: Speech processors for auditory prostheses. Fourteenth Quarterly Progress Report, NIH project N01-NS-5-2396. Neural Prosthesis Program, National Institutes of Health, 1989.

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

Neuman AC, Levitt H, Mills R, Schwander T: An evaluation of three adaptive hearing aid selection strategies. J Acoust Soc Am 1987,**82**:1967-1976.

Tyler RS, Preece JP, Lowder MW: The Iowa cochlear-implant test battery. Laboratory Report. University of Iowa at Iowa City, Dept. of Otolaryngology - Head and Neck Surgery, 1983.

Tyler RS, Preece JP, Lowder MW: The Iowa audiovisual speech perception laser videodisc. Laser Videodisc and Laboratory Report. University of Iowa at Iowa City, Department of Otolaryngology -- Head and Neck Surgery, 1987.

Wilson BS, Finley CF, Lawson DT: Speech processors for auditory prostheses. Twelfth Quarterly Progress Report, NIH project N01-NS-5-2396. Neural Prosthesis Program, National Institutes of Health, 1988.

Wilson BS, Finley CF, Lawson DT: Speech processors for auditory prostheses. First Quarterly Progress Report, NIH project N01-DC-9-2401. Neural Prosthesis Program, National Institutes of Health, 1989.

Wilson BS, Finley CF, Lawson DT: Speech processors for auditory prostheses. Third Quarterly Progress Report, NIH project N01-DC-9-2401. Neural Prosthesis Program, National Institutes of Health, 1990a.

Wilson BS, Lawson DT, Finley CF: Speech processors for auditory prostheses. Fourth Quarterly Progress Report, NIH project N01-DC-9-2401. Neural Prosthesis Program, National Institutes of Health, 1990b.

Wilson BS, Finley CC, Lawson DT: Representations of speech features with cochlear implants. In Miller JM, Spelman FA (eds): *Cochlear Implants: Models of the Electrically Stimulated Ear.* New York: Springer-Verlag, 1990c, pp 339-376.

Appendix A

Preliminary Assessment
of Various Strategies for Parametric Variation
in Speech Processor Fitting

In this appendix we shall consider four distinct, specific strategies for varying parameters in search of an optimal result. Two of the strategies, as discussed in the body of this report, have been used for speech processor optimization in our patient studies. A third has been proposed [Neuman *et al.*, 1987] for parametric fitting of hearing aids. The remaining strategy provides an intermediate model helpful in understanding the others. While we shall be displaying only two-dimensional examples of these strategies, our goal will be identification of the most promising approach(es) for more than two parameters. We begin by stating the selection rules governing each strategy's path through the parametric space. We assume that each distinct configuration of parameter values can be evaluated by a test that will yield a performance index for the configuration--a consonant identification test score, for instance, for each particular speech processor.

1. **Evolving Standard** *(Each dimension explored fully, once, in turn)*. One parametric dimension is chosen for initial exploration. That parameter is varied one step in each direction, and additional steps in the appropriate direction, until a maximum in performance index has been located. The corresponding value of the varied parameter is fixed, and becomes a permanent part of an evolving standard. A second parametric dimension is chosen and the process is repeated, culminating in the fixing of the second parameter. This process continues through all the dimensions to be explored.

Assume
independence
and no
multiple local
minima or
maxima

This approach is capable of moving relatively large distances in parameter space in relatively short order. It allows hierarchical ranking of the available parametric dimensions in terms of anticipated importance, sensitivity, *et c.* It readily accommodates parameters that are quantized or have fixed limits on their ranges.

An implicit assumption of this approach is that the choice of value for one parameter will remain valid even in the face of potentially large changes in some other parameter later in the sequence. Reexamination of the local variation of previously determined parameters may be necessary after any large movement across the parametric space.

2. **Probe Each Dimension in Both Directions** *(with respect to each successive fixed standard).* Every parameter under investigation is varied one step in each direction with respect to a single, fixed standard configuration. The performance changes accompanying each of these single parameter variations are analyzed and a new standard is established in response to all these data. The new standard, displaced in the direction of the inferred multidimensional gradient, may well be an as yet untested configuration. In a two-dimensional case, for instance, improved performance having been observed for +x and -y steps the new standard would combine both changes and be a configuration that had *not* been tested before. For this reason an additional selection rule is needed, allowing diversion to the *tested* configuration with the best performance if change along the inferred gradient in fact leads to inferior performance.

This approach is highly local in parameter space. Even though many configurations may have been tested in preparation for a change, the new standard will not have moved far along any parametric dimension. The path taken by this approach, on the other hand, is independent of the order of parameters varied and less likely than the evolving standard strategy (strategy 1 above) to ignore local structure in the performance index.

3. **Probe Each Dimension in a Fixed Single Direction** *(with respect to each successive fixed standard).* In an effort to be more efficient than strategy 2 this approach, while still varying *every* parameter under consideration with respect to a given standard configuration, initially does so in only *one* direction along each dimension. In the expectation that many of the gradients inferred from this procedure would not differ from those indicated by a bidirectional probe, this should reduce the number of configurations requiring tests. Again, a special selection rule seems appropriate: whenever the performance index declines in all the chosen parameter directions further tests should be made to rule out a local maximum or saddle point.

The path through parameter space taken by this approach, while independent of the order in which parameters are varied at each step, is highly dependent on the choice of *which* direction to probe along each dimension. This approach is local in the same sense as strategy 2 but, to the extent that the number of tests is reduced, may allow a wider search for the same amount of testing.

4. **Probe Each Dimension in a Chosen Single Direction** *(with respect to each successive fixed standard).* The direction along each dimension is chosen according to the "Modified Simplex" procedure of Neuman, *et al.* [1987]. This approach may be thought of as an adaptive version of strategy 3 above. The combination of directions probed along each dimension with respect to any standard configuration is determined by the gradients inferred from the results with respect to the previous standard. The "modified" nomenclature refers to the deci-

sion to probe unit distances along each orthogonal dimension, rather than probe to the vertices of an n-dimensional regular simplex [Box and Hunter, 1957].

This approach might be expected to be more efficient than strategy 3 by virtue of an intelligently chosen set of directions for parameter variations at each juncture, rather than an arbitrary one. The *initial* choice of probe directions along each parametric dimension, in the absence of the kind of information guiding subsequent choices, remains arbitrary.

In Figure A.1 we summarize the operation of each of these models in a two-dimensional parametric space for which there is a single, symmetric maximum in the performance index. We divide the parametric space into 25 cells and label the dimensions x and y. The performance maximum is associated with the parameter values of the center cell. The performance index of any cell outside the displayed grid is assumed to be less than that of any adjacent cell within the grid. For each of the four strategies and for each of the 25 cells as a starting point, we have summarized (1) the paths through the parametric space taken by the evolving or successive fixed standards and (2) the number of configurations (cells) that must be tested. The paths are indicated by lines beginning near the center of each cell and ending just inside the center cell in the top row of grids. The number of configurations that must be tested in each case is indicated in the cell corresponding to the starting point in the lower row of grids. The median, average, and maximum number of tests among the 25 cases for each strategy are indicated at the bottom of the figure. There are dual displays for three of the strategies, corresponding to different arbitrary choices.

The first two columns of Figure A.1 show the operation of strategy 1 for initial exploration along the x and y dimensions respectively. Clearly the path taken by this strategy through the parametric space is, for many starting points, highly dependent on that choice. Depending on the starting point, five to eleven cells must be tested in either case in order to locate the central optimal condition. A ninety degree rotation takes the one pattern into the other.

In the third column we see that strategy 2 is independent of any arbitrary choice, has a ninety-degree rotational symmetry, and on average requires that more cells be tested in the course of locating the performance index's maximum.

The results of strategy 3, displayed in the fourth and fifth columns, depend on the arbitrary choice of which directions to probe along each dimension. Patterns for two of the four possible choices are shown and are related by a 180-degree rotation. Some very complicated paths occur in the quadrant opposite the chosen directions for this strategy, but on average the efficiency in locating the central maximum is slightly better than that of strategy 1.

Evolving Standard

(each dimension explored fully in turn)

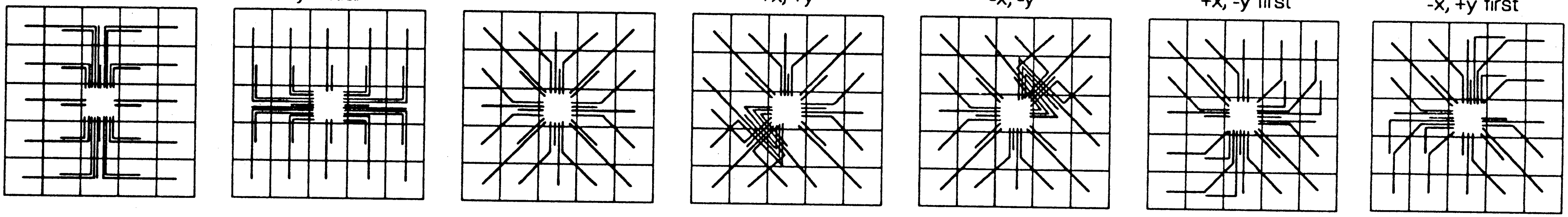
x initial | y initial

Probe each Dimension Locally wrt Successive Fixed Standards

Probe each dimension in both directions

Probe each dimension in only one direction: +x, +y | -x, -y

"Modified Simplex" Sequence: +x, -y first | -x, +y first



x initial

11	10	9	10	11
10	9	8	9	10
7	6	5	6	7
10	9	8	9	10
11	10	9	10	11

y initial

11	10	7	10	11
10	9	6	9	10
9	8	5	8	9
10	9	6	9	10
11	10	7	10	11

Probe each dimension in both directions

11	11	11	11	11
11	8	8	8	11
11	8	5	8	11
11	8	8	8	11
11	11	11	11	11

+x, +y

9	10	11	11	11
9	7	8	8	11
8	7	5	8	11
10	6	7	7	10
9	10	8	9	9

-x, -y

9	9	8	10	9
10	7	7	6	10
11	8	5	7	8
11	8	8	7	9
11	11	11	10	9

+x, -y first

9	9	8	8	11
9	6	6	8	10
8	6	5	8	10
8	8	8	8	11
11	10	10	11	11

-x, +y first

11	11	10	10	11
11	8	8	8	8
10	8	5	6	8
10	8	6	6	9
11	8	8	9	9

	x initial	y initial	in both directions	+x, +y	-x, -y	+x, -y first	-x, +y first
Median	9	9	11	9	9	8	8
Average	9.0	9.0	9.8	8.8	8.8	8.7	8.7
Maximum	11	11	11	11	11	11	11

Figure A.1

The final two columns in Figure A.1 represent the "modified simplex" approach, strategy 4. The striking differences in path in this case are due entirely to the choice of probe directions for the very first tests. Again, two of the four possible choices are shown, and again on average the number of cells that must be tested is slightly less than for strategy 1. Note that substitution of an initial probe in both directions along each dimension instead of an arbitrary choice would give strategy 3 the same path symmetry as strategy 2, but would also increase the number of tests required in many cases.

Having acquainted ourselves in some detail with the systematics of these four strategies operating with a very well-behaved performance index, we now turn to Figure A.2 and a more complex topology. In this figure the columns all correspond to exactly the same strategies, but the performance index now has two local maxima--middle of second column and second from bottom in the fourth column in each grid. As mentioned in the body of this report, we have chosen this example to be one of the performance topologies consistent with the fragmental data of Figure 2. We have, in effect, added rows of cells above and below, and a column of cells to the right of the grids in that example, and supplied plausible but arbitrary performance index values for all the untested conditions. Again, any cell outside the model grid is assumed to have a lower performance index than any adjacent cell within.

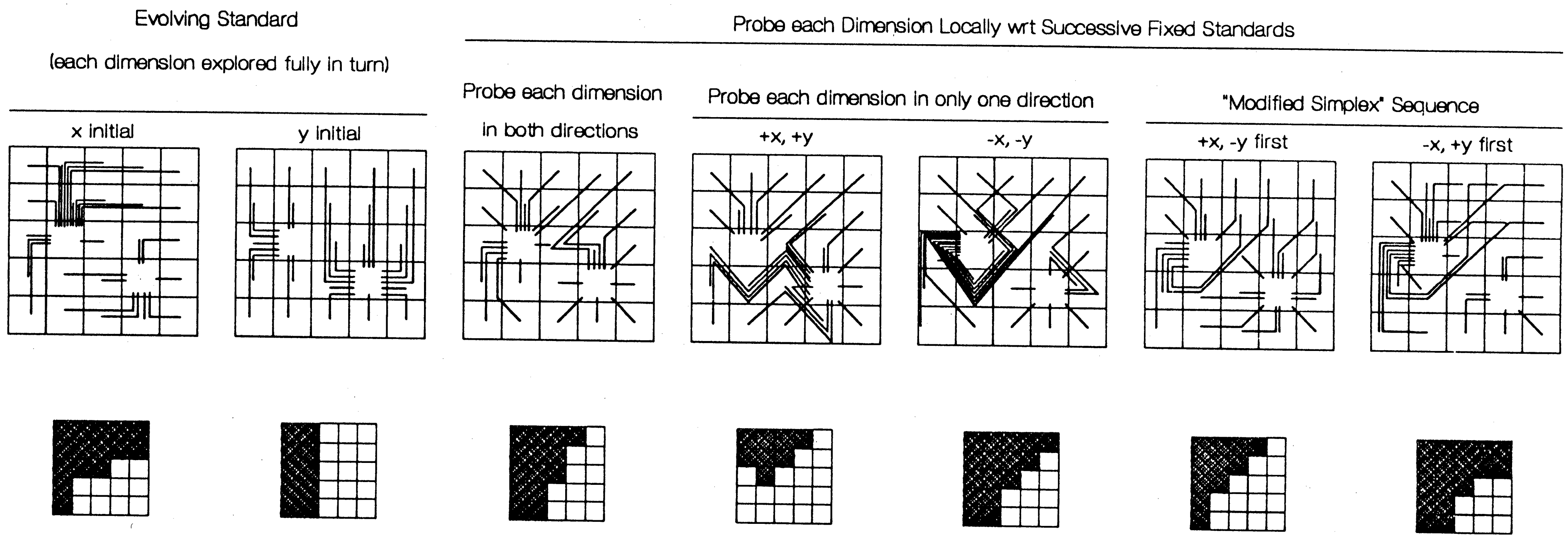
The paths in the upper row of grids now may lead to *either* of the two optimal conditions. The small grids in the new middle row indicate how the various starting points are divided by each strategy between the two outcomes. The same division is indicated in the bottom row by bolder lines between cells.

At first glance, one notices that the dual-maximum topology has broken the symmetries within the strategy 1, 3, and 4 pairs; that some very complicated paths have emerged; and that there is considerable variation in the average number of tests required and the association of starting points with particular maxima.

For strategy 1, the two choices of initial dimension to explore result in quite different average numbers of configurations tested (8.7 and 8.0) and in assignments of maxima to starting points (15/10 diagonal and 10/15 along x).

In the third column of grids in Figure A.2 we see that strategy 2 continues to require significantly more testing to locate an optimal cell. The zig-zags near the grid center in two paths correspond to invocations of the additional selection rule needed by this strategy for such circumstances. Otherwise, this strategy also is characterized by simple, straightforward paths.

The average number of conditions tested has become quite high for both variants of strategy 3 as well. This is due mainly to some very long, complicated paths in response to the dual local performance maxima. Note also that the worst case choice of starting point for this



10	9	10	11	12
9	8	9	10	11
6	5	7	8	9
10	7	6	5	8
11	10	9	8	9

10	7	11	8	11
8	6	10	7	10
8	5	9	6	9
8	6	8	5	8
10	7	9	6	8

11	11	11	11	14
8	8	8	11	11
8	5	8	8	8
10	8	8	5	8
13	13	8	8	8

10	11	11	11	11
7	8	8	10	11
11	5	7	7	8
11	9	6	5	7
14	10	6	9	7

12	8	13	9	16
7	10	6	13	10
8	5	10	7	6
10	9	8	5	9
12	12	8	8	7

9	8	8	15	11
6	6	12	8	8
6	5	6	6	8
9	8	6	5	8
12	8	8	8	8

11	10	10	16	11
8	8	8	8	15
8	5	6	8	8
8	8	8	5	6
10	10	8	6	6

Median	9	8	8	9	8	8	8
Average	8.7	8.0	9.2	8.8	9.1	8.1	8.6
Maximum	12	11	14	14	16	15	16

Figure A2

strategy now requires that 16 configurations be performance tested, and that the two variants result in 8/17 and 16/9 divisions of starting points between the two optimal outcomes.

A few very long and complicated paths also characterize the operation of strategy 4 for this more complicated performance topology. While the average number of tests required by the two variants spans essentially the same range as that covered by strategy 1, the worst cases are substantially worse, corresponding to those long, tortuous paths. It appears that such problems would not be eliminated by probing in all directions before the initial step, as was suggested for the simpler topology.

Our tentative overall assessment of these strategies is that number 1--the evolving standard--will best fulfill the needs of multidimensional parametric fittings of speech processors for cochlear and cochlear nucleus implants, perhaps with judicious augmentation by strategy 2 . The theoretical efficiency of strategy 1 compares well with the others, including the "modified simplex" approach, and its paths are least distorted by complex topologies. More importantly, strategy 1 better accommodates the limits and idiosyncrasies of some of the real speech processor parameters, such as inherently bounded or qualitative variations. And, most important of all, strategy 1 allows informed, hierarchical use of parameters, in terms of such characteristics as anticipated importance and sensitivity, degree of independence from other parameters, *et c.*

So long as clinical experience supports basing design decisions on the relative performance of speech processors differing only in the value of a single parameter, strategy 1 appears to be the approach of choice for efficient processor fitting.

Appendix B

New Tools for Speech Processor Fitting

This appendix includes two examples of tools developed for use in the fitting of speech processors for future patient studies in our laboratory. Both are direct outgrowths of the findings discussed in this report.

Figure B.1a is a screen image from the investigator's console at the conclusion of one of our 16-consonant identification tests. Most of the information displayed is the same as for the 14-consonant test screen discussed in a previous report [Lawson, *et al.* 1989]. In the upper right hand corner of this screen is a new feature, however, one allowing ongoing knowledge of the significance of measured differences in consonant identification. The five separate 16-consonant sets that comprise the 80-presentation test are listed on the display as they occur, and characterized in terms of: (1) an overall percent correct score observed for each set alone, with its uncertainty displayed as the standard deviation of any such single observation, and (2) a cumulative overall percent correct score to that point, with uncertainty displayed in terms of the standard deviation of the mean.

Figure B.1b is a screen image for exactly the same test results, but in the circumstance of there being a previous test in the archives for the exact same conditions. Such previous data is detected automatically at the beginning of each test. Its existence is signaled by the numbering of the 16-consonant sets in the current test [in this example beginning with set 6, reflecting one previous such test's contributing an initial five sets]. The standard deviations of each single observation and of the means reflect these previous data, as do the cumulative overall scores at the end of each set. This information will allow investigators to judge when significance in the distinctions among candidate processors can be achieved through acquiring further data for a given condition.

Figure B.2 is a coding form developed to assist investigators in recognizing patterns in test result topologies over multidimensional parametric spaces. The form is designed to be flexible and convenient for use in the context of decision making during processor fitting. The specific parameters charted, of course, are subject to revision. A separate copy of the form can be used for each type of result to be charted.

```
Medial Consonant Confusion Test:  Sequence  280
Male Talker,     MP31m3 processor
    m n f v s zshth p b t d g k j l
  m|2 3
  n|4 1
  f|    4   1
  v|      2       3
  s|        4   1
  z|          5
 sh|        1   4
 th|      1   1   3
  p|                5
  b|                  3   2
  t|                1   2     2
  d|                  2   3
  g|                      1 4
  k|                    1     4
  j|                            5
  l|                              5
Hear 6 4 4 3 6 6 5 6 6 5 3 6 4 6 5 5
Sent 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5
      Identification: 70%

    uv  v
 uv|30
  v|   50

   30 50
   30 50
Voicing:100%

     lab den alv pal vel
 lab| 10       5
 den|     13   2
 alv|  7      25   1   2
 pal|          1   4
 vel|          2       8

      17  13  35   5  10
      15  15  35   5  10
        Place: 75%

1    75% set         75% cumul
2    75% ±  0.0      75% ±  0.0
3    50% ± 14.4      67% ±  8.3
4    69% ± 11.8      67% ±  5.9
5    81% ± 12.0      70% ±  5.4

Distribution of Errors
    Voicing only:    0%
    Place only:     83%
    Manner only:     0%
    Multiple:        0%

     plo fri nas
 plo| 30
 fri|     35
 nas|         15

      30  35  15
      30  35  15
   Manner:100%
```

Figure B.1a

```
Medial Consonant Confusion Test:  Sequence  280      6   75% ± 12.3   67% ±  5.0
Male Talker,    MP31m3 processor                     7   75% ± 11.7   68% ±  4.4
     m n f v s zshth p b t d g k j l                 8   50% ± 12.5   66% ±  4.4
                                                     9   69% ± 11.7   66% ±  3.9
                                           uv  v    10   81% ± 12.1   68% ±  3.8
   m|2 3                             |
   n|4 1                             |       uv|30   |
   f|    4   1                       |        v|   50|  Distribution of Errors
   v|      2       3                 |                    Voicing only:    0%
   s|        4   1                   |          30 50     Place only:     83%
   z|          5                     |          30 50     Manner only:     0%
  sh|        1   4                   |       Voicing:100% Multiple:        0%
  th|      1   1   3                 |
   p|                5               |          lab den alv pal vel
   b|                  3   2         |
   t|                1   2     2     |      lab  10       5                plo fri nas
   d|                  2   3         |      den      13   2
   g|                      1 4       |      alv   7      25   1   2    plo  30
   k|                    1     4     |      pal            1   4       fri      35
   j|                            5   |      vel            2       8   nas          15
   l|                              5 |
                                             17  13  35   5  10        30  35  15
Hear 6 4 4 3 6 6 5 6 6 5 3 6 4 6 5 5         15  15  35   5  10        30  35  15
Sent 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5 5           Place: 75%            Manner:100%
        Identification: 70%
```

Figure B.1b

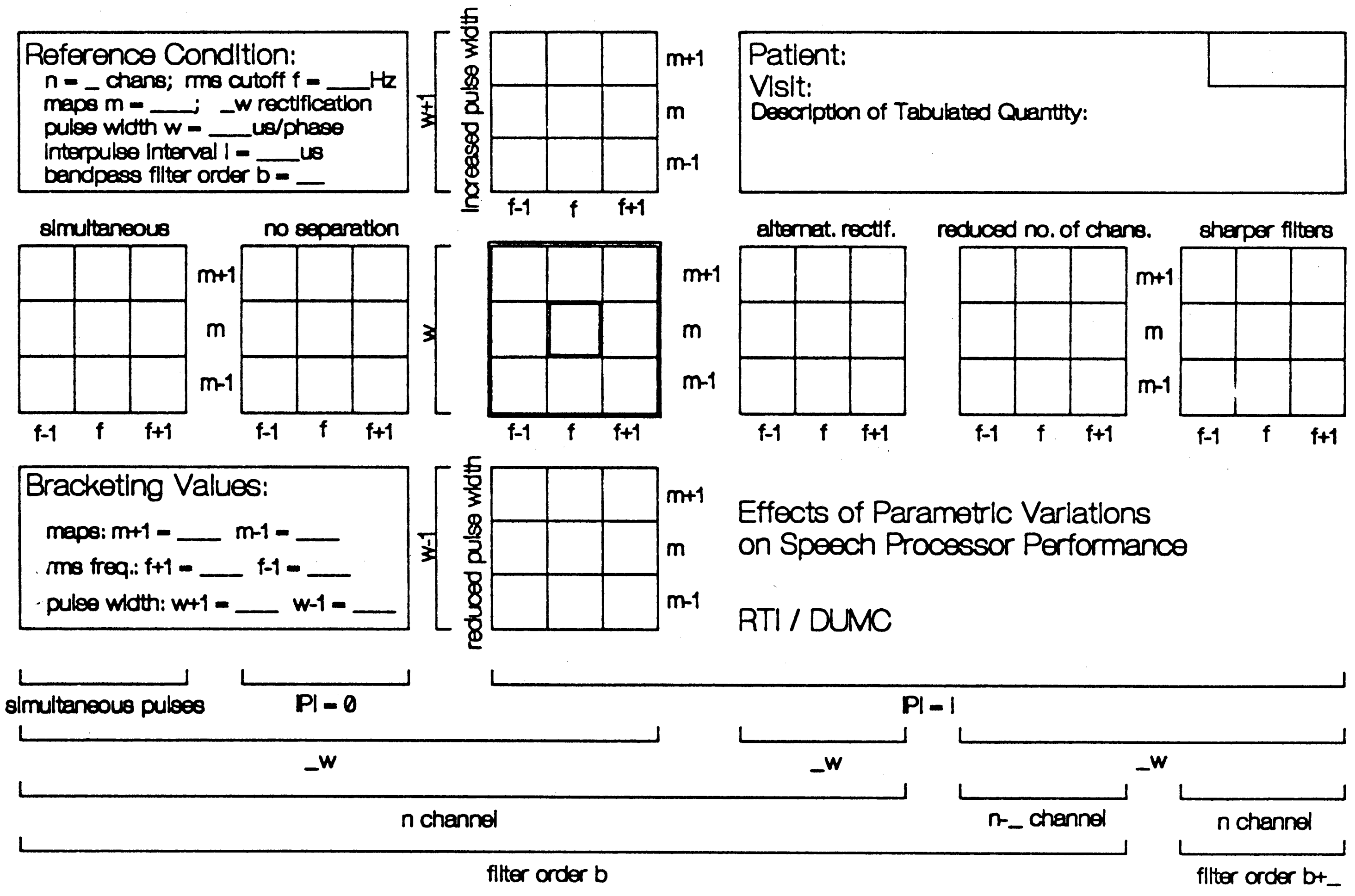

Reference Condition:
n = _ chans; rms cutoff f = ____Hz
maps m = ____; _w rectification
pulse width w = ____us/phase
interpulse interval I = ____us
bandpass filter order b = __
w+1
increased pulse width
m+1
m
m-1
f-1
f
f+1
Patient:
Visit:
Description of Tabulated Quantity:
simultaneous
no separation
w
alternat. rectif.
reduced no. of chans.
sharper filters
Bracketing Values:
maps: m+1 = ____ m-1 = ____
rms freq.: f+1 = ____ f-1 = ____
pulse width: w+1 = ____ w-1 = ____
w-1
reduced pulse width
Effects of Parametric Variations
on Speech Processor Performance
RTI / DUMC
simultaneous pulses
IPI = 0
IPI = I
_w
_w
_w
n channel
n-_ channel
n channel
filter order b
filter order b+_

Figure B.2

III. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Presentation of project results at the *Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, to be held in Philadelphia, PA, November 1-4 (Finley: "Radial Bipolar Electrode Placement in the Scala Tympani. Effects on Neural Potential Profiles and Longitudinal Spread of Excitation")
2. Continued development of a portable processor based on the DSP 56001 device.
3. Continued preparation of manuscripts for publication.
4. Initiation of a new series of studies with 6 or 7 subjects who have relatively low levels of performance with the Ineraid prosthesis. This series follows up on our encouraging results from one such subject, as described in *QPR 5* for this project.

Appendix 1

Summary of Reporting Activity for the Period of August 1 through October 31, 1990

NIH Contract N01-DC-9-2401

The following presentation was made in the last quarter of project work:

Wilson, B.S.: Speech Processors for Auditory Prostheses. Presented at the *21st Neural Prosthesis Workshop*, Bethesda, MD, October 17-19, 1990.