Second Quarterly Progress Report

November 1, 1992 through January 31, 1993

NIH Contract N01-DC-2-2401

Speech Processors for Auditory Prostheses

Prepared by

Dewey T. Lawson, Blake S. Wilson and Mariangeli Zerbi

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

CONTENTS

I. Introduction

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

1. Continued evaluation of a new *virtual channel interleaved sampling* (VCIS) processing strategy for multichannel cochlear implants (see QPR 1, this project). Studies were conducted with Ineraid subject SR2.
2. Conduct of collaborative studies with Michael Dorman to evaluate representations of frequency cues through *compressed analog* (CA), *continuous interleaved sampling* (CIS) and VCIS processors.
3. Continued analysis of data from prior and current studies, to evaluate effects of single parameter variations on the performance of CIS processors.
4. Transfer of CIS processor technology to the University of Innsbruck, in support of their efforts to utilize the technology in a next-generation multichannel prosthesis.
5. Completion of negotiations with Cochlear Corporation for a cooperative study involving six patients who will have percutaneous access to an implanted Nucleus electrode array. Evaluation of CIS and other processing strategies with these subjects is scheduled to begin in the summer of 1993.
6. Continued preparation of manuscripts for publication.

In this report we present results from studies to evaluate effects of changes in pulse durations, pulse rates, and envelope smoothing filters used in CIS processors (examples of evaluations under point 3 above). Work related to points 1 and 2, along with results obtained in studies of other parameters of CIS processors, will be described in future reports.

Single Parameter Variation Studies for CIS Processors

Introduction

The principal objective of work under this NIH contract is the identification and development of improved speech processing strategies for implanted auditory prostheses. Accordingly, we evaluate the potential of promising new processing strategies in studies with patients who have had a wide range of results with their clinical devices. Such studies typically involve repeated trials with different settings of each strategy's adjustable parameters in attempts to approach an optimal fitting in the limited time available with each subject. Methods of rapid optimization, now desirable for greater research productivity, will become a necessity for widespread clinical use of such processing strategies.

Our approach has emphasized comparisons of patient performance with processors differing only in the value of a single parameter. We hope to detect patterns in the effects of such single-parameter manipulations and to assess the robustness of those patterns across variations in the settings of other parameters. [See Lawson *et al.* 1990.]

The list of parameters with a demonstrated ability to produce significant changes in patient performance is a long and growing one. In our most recent patient studies with continuous interleaved sampling (CIS) processors, for example, the following parameters routinely have been varied:

number of pulsatile channels utilized

mode of stimulation on each channel
  (single *vs.* multiple electrodes per channel)

frequency band associated with each channel
  band frequency ranges
  bandpass filter order

envelope estimates for each channel
  halfwave vs. fullwave rectification
  smoothing low-pass filter cutoff frequency
  smoothing filter order

sequence of channel stimulation within each cycle
  (apex-to-base, base-to-apex, staggered, random, etc.)

stimulation pulses
  pulse timing
    duration per phase
    delay between successive pulses
  leading phase polarity
    (anodic *vs.* cathodic)

dynamic range of each channel
  shape of stimulation amplitude mapping function
  loudness percept corresponding to maximum amplitude

With the number of significant parameters so large, and access to each research patient so limited, a thorough study of the parametric space of such a multichannel processor is inconceivable. As previously reported [Lawson *et al.*, 1990] studies with single-channel processors for stimulation of auditory brainstem electrodes afforded us the opportunity for a manageable pilot study of the effects of single-parameter variations. Our next step was to identify significant but tractable parametric subspaces of multichannel processors for systematic exploration.

Studies

We now report results from studies of two such two-dimensional parametric subspaces within the CIS family of processors. One is a plane involving variations in pulse rate and envelope smoothing filter cutoff frequency, the other a plane representing variations in pulse rate and pulse duration.

The study of the effect of pulse rate and pulse envelope smoothing filter frequency on speech intelligibility is related to the effects of aliasing; to the advantages, if any, of pulse rates higher than those required to avoid aliasing for a given envelope upper frequency limit; and to the potential benefits of providing information on higher frequency envelope fluctuations. Aliasing effects are a concern whenever a processor attempts to convey band envelope variations at frequencies above one half the stimulation pulse rate on each channel--the result being the addition of low rate fluctuations to the output pulse envelopes.

The study comparing variations in pulse rate and pulse durations arises from our observation of two distinct patterns in patient performance. As just noted, the use of shorter pulses and higher pulse rates allows representation of higher frequencies in the modulation waveform for each channel without aliasing. In addition, the dynamic range of electrical stimulation--from threshold to most comfortable loudness (MCL)--is a strong function of pulse rate and a weaker function of pulse duration. Large increases in dynamic range generally have been observed with increases in pulse rates from about 400 pps to 2500 pps. Smaller increases often (but not always) have been observed with increases in pulse duration for a fixed rate of stimulation--from roughly 50 μs/phase to about 200 μs/phase. For several subjects, however, we have observed that the salience of channel ranking can decline with decreases in pulse duration below 100 μs/phase. Such subjects might well benefit from the use of broad duration pulses (100 μs/phase or greater) to preserve channel cues, even though sacrificing some of the dynamic range (and higher frequency envelope variations) obtainable with shorter pulses and higher rates of stimulation. Another possible advantage of relatively low rates of stimulation is reduction of channel interactions. Allowing time between pulses on sequential channels can reduce the temporal integration component of channel interactions. Thus use of time delays between short duration pulses in the stimulation sequence across electrodes may reduce interactions. Alternatively, use of long duration pulses with no time delay also might reduce temporal interactions, by ensuring a relatively long interval between the excitatory phases of successive pulses.

Subject

Having identified two relatively manageable subspaces for systematic investigation of the effects of single-parameter manipulations, we elected to begin the studies with subject SR2.

A user of the Ineraid clinical device, subject SR2 offered percutaneous access to six intracochlear electrodes. He had achieved very high levels of performance with a variety of different speech processing strategies, suggesting excellent and uniform survival of stimulable neural elements. He had a great deal of experience with studies of speech reception, both in our laboratory and at the Massachusetts Eye and Ear Infirmary. His performance in speech reception tests was known to be highly reproducible over a wide range of intervals between repeated tests. He was known to adapt very rapidly to new processors, requiring only a few minutes' experience with a new strategy before productive testing. In short, subject SR2 seemed to offer the best prospects for rapidly and reliably assessing a large number of processor conditions with minimal danger of the results being biased by some patient-specific pattern of neural survival, or by some defect in auditory processing central to the cochlear nerve.

Tests

The performance evaluation instrument chosen for these studies was a medial consonant identification test. We chose to use 24 English consonants presented in an /aCa/ context edited from a University of Iowa videodisc recoding [Tyler *et al.* 1987]. The 24 consonants were: /b, d, f, g, dʒ, h, j, k, l, m, n, ŋ, p, r, s, ʃ, t, tʃ, ð, θ, v, w, z, ʒ/. In editing the tokens from the 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 audible cues [background noises, preparatory lip smack, *et c.*]. The tokens were edited solely for sound-alone use, and the resulting videodisc segments would not be appropriate for use with video images. Three distinct exemplars each were found for the male and female talkers for each of the 24 consonants. The tokens were presented a single time with no feedback provided as to the correctness of responses. The order of presentation and choice of exemplar were randomized within complete sets of 24 tokens, with five such sets making up each administration of the test. Male and female voices were not mixed, but evaluated separately. Presenting the tokens in complete sets allowed ongoing assessment of statistical uncertainties in the scores [Lawson, *et al.*, 1990].

Among the advantages of such consonant identification tests for our studies is the availability of information transmission (IT) analyses of the raw results, offering the possibility of deeper insight into the nature and potential consequences of various differences in performance level [Wilson *et al.*, 1991]. Other advantages are that these tests can be administered rapidly and repeatedly, over many processor conditions and within short time intervals, without a subject's increasing familiarity with the material contaminating the results. Finally, we have documented high correlations between results on such consonant identification tests and scores on standard open set speech tests [correlation coefficients of 0.79 and 0.83 for a similar 16-consonant identification test and the two most difficult open-set MAC battery tests, with significance at $p < .001$; see Wilson, *et al.*, 1990]. Such open set tests themselves would not be appropriate for the present study because of the time required to administer them and the inability to repeat them after brief intervals without concerns as to increasing subject familiarity with the materials.

Methods

The use of medial consonant identification tests with subject SR2, however, did present some special difficulties. Even with sound-alone presentation and the use of 24 consonants, SR2 achieves identification scores with many 6-channel CIS processors that are biased by ceiling effects. For this reason, we chose to more intensively study four-channel processor variations in parametric regions that had been chosen on the basis of less sensitive results with six-channel designs. Previous studies with subject SR2, of processors differing only in number of channels, indicated that four-channel processors would provide enough errors in a 24-consonant identification test to ensure some measurement sensitivity [Lawson, *et al.*, 1992; Wilson, *et al.*, 1992].

There was an inevitable tradeoff between the amount of testing in each experimental condition and the number of different conditions that could be evaluated in the time available. Thus during the studies we often faced decisions with regard to the competing goals of minimum statistical uncertainty in the assessment of performance in each condition and optimal characterization of the overall performance topology of the parametric space. The need for such decisions was intensified when, early in the studies, striking differences appeared between results for male and female voices. In making such decisions we considered that, where statistical uncertainties might remain greater than we would have preferred for some particular conditions, inclusion of the results from neighboring cells in the parametric plane could increase confidence in the absolute performance levels while the finer scale coverage of the parametric space preserved the possibility of detecting any local topological features. The number of presentations of each token in each condition may be found in Appendix 2 to this QPR.

Figures 1 - 8 display results from the study of variations in pulse rate and the envelope smoothing filter cutoff frequency, while Figures 2 - 16 display similar results from the study of variations in pulse duration and pulse rate. In both cases data are gathered in separate figures for overall percent correct consonant identification; overall percent information transmission; and percent information transmission for each of the following speech features: voicing, envelope, frication, place of articulation, duration, and nasality. In each figure results for the current systematic study of four-channel processors are shown to the left and earlier less sensitive data for similar six-channel processors to the right. Male voice data are shown at the top in each case and female voice results below. Cells containing overall percent correct scores with error bars overlapping those of the highest score are shaded. In information transmission figures, cells containing scores within an arbitrary five percent of the highest score are shaded.

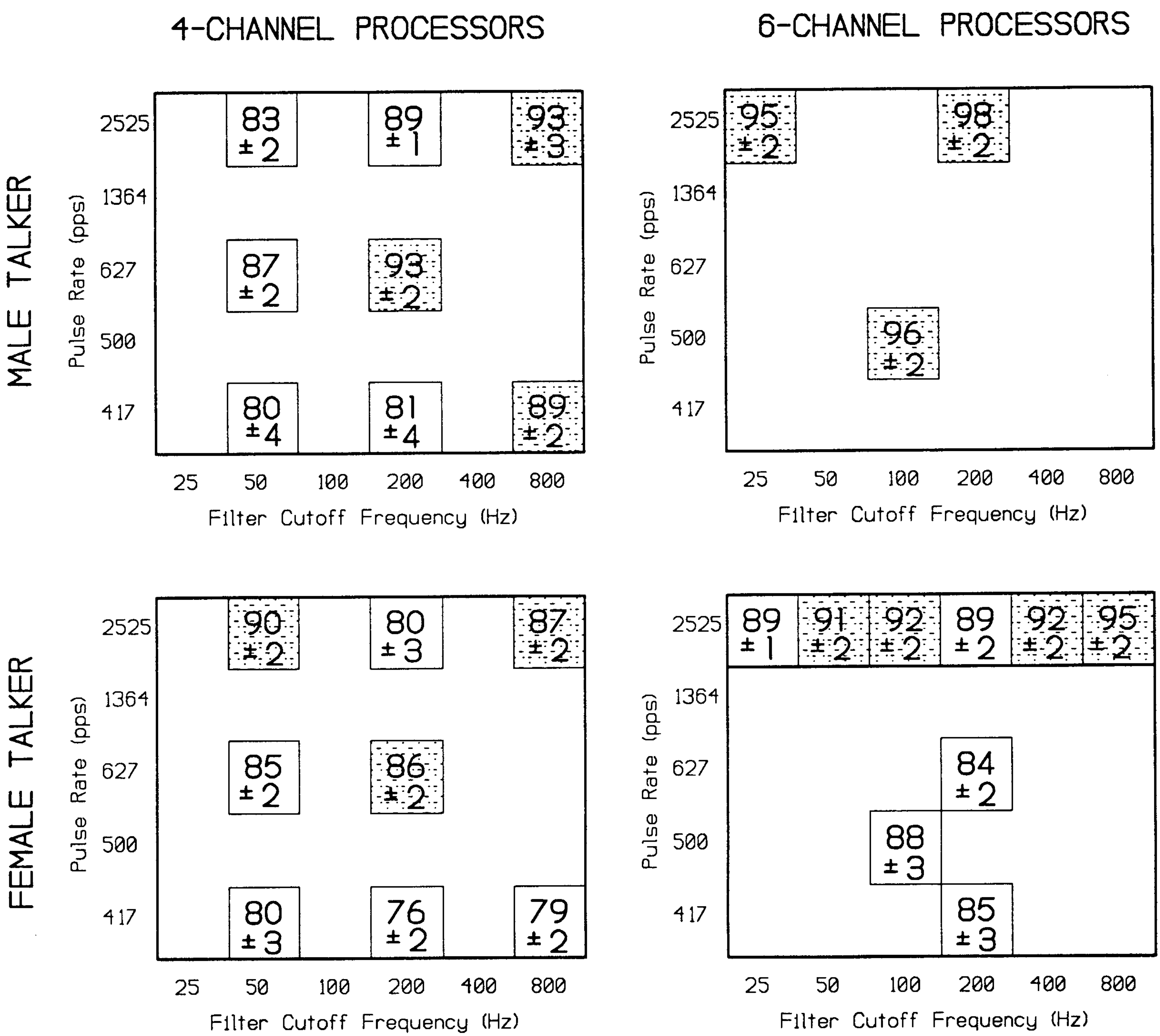
Study of Pulse Rate and Smoothing Filter Cutoff Variations

Characteristics that were common to all the processors of this study included: logarithmically equal channel bands spanning 350 to 5500 Hz with sixth-order bandpass filters, full-wave rectification and fourth-order smoothing filters for the envelope detectors, single-electrode channel stimulation, 33 μs/phase pulse duration with anodic phase leading, a power law amplitude function from perceived threshold to MCL that closely approximated logarithmic mapping on each channel, and no automatic gain control (AGC). The six-channel processors stimulated all six of the subject's electrodes in staggered order (6-3-5-2-4-1), while the four-channel processors stimulated apicalmost electrodes 1 through 4 in the order (3-1-4-2), there being no four-channel sequence that can avoid any sequential stimulation of adjacent electrodes.

The Nyquist criterion that the pulse rate in pps exceed twice the low pass smoothing filter cutoff frequency in Hz, is violated in the 800 Hz/417 pps condition and only marginally achieved in the 200 Hz/417 pps condition. All other cases in this study satisfy the criterion. Thus the conditions along the main diagonal from lower left to upper right correspond to processors in which pulse rate should be sufficient to avoid aliasing. Conditions above that diagonal provide still higher pulse rates, without exploiting them to convey higher frequency envelope information. The two conditions below that diagonal should involve aliasing effects--in the one case relatively modest ones and in the other case quite severe.

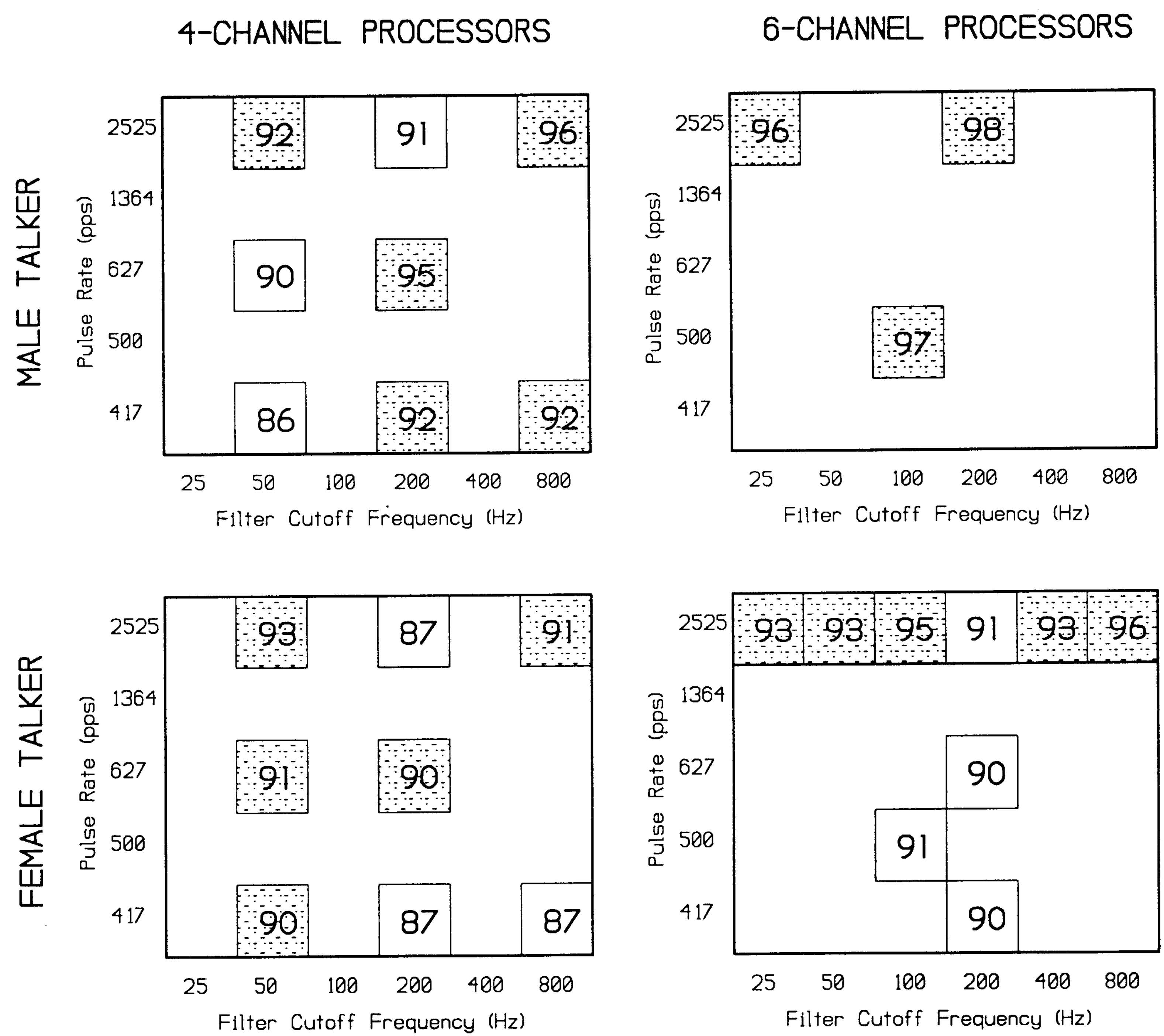
Turning to the performance data displayed in Figures 1 through 8, we concentrate first on single parameter variations in the four-channel processor results. There is a strong pattern of better understanding of the male talker with the use of higher smoothing filter cutoff frequencies. Among the individual information transmission features only duration shows no evidence of such a pattern. With the female voice, on the other hand, there is a weaker pattern of overall IT and several features' favoring lower cutoff frequencies. For single parameter variation along the rate of stimulation axis there is a general pattern of improved performance at higher rates--at least for the increase from 417 to 627 pps-with both male and female voices.

In terms of the highest performance region of this parametric plane, the combination of high rate and high cutoff frequency performs well for male voice, while high rate and low cutoff frequency is more weakly indicated for the best results with a female talker.



Percent Correct Identification

Figure 1



Percent Overall Information Transmission

Figure 2

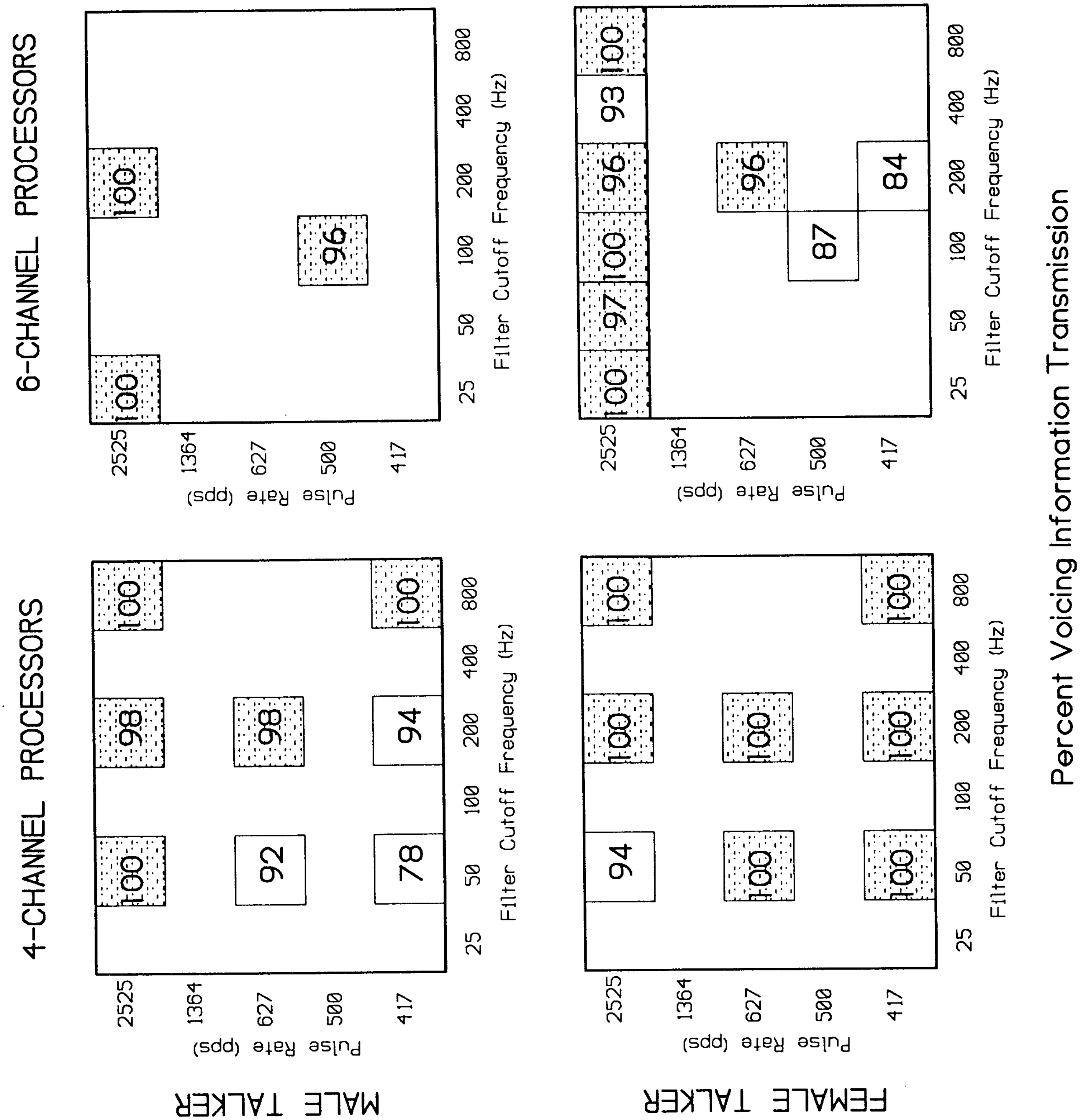

4-CHANNEL PROCESSORS
6-CHANNEL PROCESSORS
MALE TALKER
FEMALE TALKER
Pulse Rate (pps)
Filter Cutoff Frequency (Hz)
Percent Voicing Information Transmission

Figure 3