X 3

Ninth Quarterly Progress Report

September 27 through December 28, 1987

NIH Contract N01-NS-5-2396

Speech Processors for Auditory Prostheses

Prepared by

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

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

CONTENTS

I. Introduction

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

1. Evaluation of alternative processing strategies in tests with a patient implanted with an initial version of a multichannel auditory prosthesis developed by the 3M Company of St. Paul, Minnesota;
2. Development of fitting procedures for application of "interleaved pulses" processors (see QPR 6, this project) in standard audiologic clinics;
3. Development of software for on-line feature analysis of results from tests of vowel and consonant confusions;
4. Continued psychophysical and speech perception studies with patient MH;
5. Presentation of project results in one talk at the 18th Annual Neural Prosthesis Workshop, in four invited lectures at the 9th Annual Conference of the Engineering in Medicine and Biology Society, and in one invited lecture at the 3M Company in St. Paul; and
6. Continued collaboration with the UCSF team in the development of the speech processor and transcutaneous transmission system for a next-generation auditory prosthesis.

In this report we will present data from our studies with patients implanted with the four-channel UCSF/Storz cochlear prosthesis. These data provide direct comparisons of the performance levels obtained with each patient's compressed analog (CA) processor with the performance levels obtained with the interleaved pulses (IP) processor developed in the early stages of this project. Descriptions of the activities indicated in points 1-4 and 6 above will be presented in future reports.

II. Studies with Patients Implanted with the UCSF/Storz Cochlear Prosthesis

In early February, 1987, we began an intensive series of studies with six patients implanted with the four-channel UCSF/Storz cochlear prosthesis. The main purpose of these studies was to compare in tests with the same patients the performance of the compressed analog (CA) processor of the UCSF/Storz prosthesis with the performance of the interleaved pulses (IP) processor developed in this project. We were motivated by previous observations from studies with two patients fitted with percutaneous cables for direct access to all 16 contacts in the UCSF/Storz electrode array (Wilson et al., 1987 and in press). In particular, results from these previous studies demonstrated that (a) different processing strategies can produce widely different outcomes for individual patients, (b) IP processors are far superior to the tested alternative processors for at least two patients with psychophysical signs of poor nerve survival, and (c) the performance of IP processors strongly depends on the number of available stimulation channels. With these observations in mind, we were most interested in comparing the IP and CA processing strategies in extensive tests with a larger population of patients. We wondered, for example, how the IP processor would perform for successful users of the CA processor, and whether the potential advantages of the IP processors could be realized for patients with four or fewer channels of stimulation.

A detailed report of our initial findings is presented here as Appendix 2. This report also describes considerations in the design of speech processors for cochlear prostheses, with an emphasis on considerations that relate to nerve survival. Data on the comparisons of CA and IP processors are presented on pages 13 through 25 and Figs. 6 through 11 of the appendix. These data are a subset of those collected in the studies with the patients implanted with the UCSF/Storz prosthesis. The additional data will be presented in future progress reports, as will the results from follow-up studies we have scheduled for three of the original six patients.

III. Plans for the Next Quarter

Our plans for the next quarter include the following:

1. Conduct follow-up studies in St. Paul with two patients implanted with an initial version of the multichannel cochlear prosthesis developed by the 3M Company;
2. Conduct follow-up studies at Duke with two patients (HE and MC2 of Appendix 2) implanted with the four-channel UCSF/Storz cochlear prosthesis;
3. Continue ongoing psychophysical and speech perception studies with patient MH (a Duke patient fitted with a percutaneous cable for direct access to her implanted UCSF/Storz electrode array), with an emphasis on investigations relating to pitch and loudness coding with cochlear implants;
4. Complete two manuscripts in preparation, one titled "Ensemble models of neural discharge patterns evoked by intracochlear electrical stimulation. I. Simple model of responses to transient stimuli" and the other titled "Evaluation of two channel, 'Breeuwer/Plomp' processors for cochlear implants";
5. Present project results at the Mayo Clinic Symposium in Audiology, February 19-20, 1988 (the lectures from this symposium will be televised to more than 50 participating clinics and universities via satellite transmission);
6. Attend the 11th Annual Midwinter Research Meeting for the Association for Research in Otolaryngology, January 31 to February 4, 1988; and
7. Continue our collaboration with UCSF to develop the speech processor and transcutaneous transmission system for a next-generation cochlear prosthesis.

IV. References

Wilson, B.S., C.C. Finley, M.W. White and D.T. Lawson: Comparisons of processing strategies for multichannel auditory prostheses. In Proc. Ninth Ann. Conf. Engineering in Medicine and Biology Soc., IEEE Press (CH2513-0/87/0000), 1987, pp. 1908-1910.

Wilson, B.S., C.C. Finley, J.C. Farmer, Jr., D.T. Lawson, B.A. Weber, R.D. Wolford, P.D. Kenan, M.W. White, M.M. Merzenich and R.A. Schindler: Comparative studies of speech processing strategies for cochlear implants, Laryngoscope, in press.

Appendix 1

Summary of Reporting Activity for the Period of September 27 through December 28, 1987, NIH Contract N01-NS-5-2396

The following publications and presentations were made in the present reporting period:

Wilson, B.S.: Speech processors for auditory prostheses. Presented at the 18th Annual Neural Prosthesis Workshop, Bethesda, MD, Oct. 28-30, 1987.

Wilson, B.S.: Review of RTI research on coding strategies for cochlear prostheses. Invited lecture presented at the 3M Company in St. Paul, MN, Nov. 12, 1987.

Finley, C.C., B.S. Wilson and M.W. White: A finite-element model of bipolar field patterns in the electrically stimulated cochlea -- A two dimensional approximation. In Proc. Ninth Ann. Conf. Engineering in Medicine and Biology Soc., IEEE Press (CH2513-0/87/0000), 1987, pp. 1901-1903.

White, M.W., C.C. Finley and B.S. Wilson: Electrical stimulation model of the auditory nerve: Stochastic response characteristics. In Proc. Ninth Ann. Conf. Engineering in Medicine and Biology Soc., IEEE Press (CH2513-0/87/0000), 1987, pp. 1906-1907.

Wilson, B.S., C.C. Finley, M.W. White and D.T. Lawson: Comparisons of processing strategies for multichannel auditory prostheses. In Proc. Ninth Ann. Conf. Engineering in Medicine and Biology Soc., IEEE Press (CH2513-0/87/0000), 1987, pp. 1908-1910.

Finley, C.C., B.S Wilson and M.W. White: Models of afferent neurons in the electrically stimulated ear. Invited paper presented at the Ninth Ann. Conf. Engineering in Medicine and Biology Soc., Boston, MA, Nov. 13-16, 1987.

In addition to the above, the following two invited papers were submitted for publication:

Wilson, B.S. (moderator), L.J. Dent, N. Dillier, D.K. Eddington, I.J. Hochmair-Desoyer, B.E. Pfingst, J. Patrick, W. Sürth and J. Walliker (panelists): Round table discussion on speech coding. To be published as a chapter in P. Banfai (Ed.), Cochlear Implants 1987, Springer-Verlag.

Wilson, B.S., C.C. Finley, D.T. Lawson and R.D. Wolford: Speech processors for cochlear prostheses. Paper submitted for publication in the special issue on "Emerging Electromedical Systems," Proc. IEEE, September, 1988.

Appendix 2

Speech Processors for Cochlear Prostheses

Blake S. Wilson[1,2], Charles C. Finley[1],
Dewey T. Lawson[1] and Robert D. Wolford[2]

[1]Neuroscience Program Office, Research Triangle Institute,
Research Triangle Park, NC 27709.

[2]Division of Otolaryngology and Center for Speech and Hearing
Disorders, Department of Surgery, Duke University Medical Center,
Durham, NC 27710.

Reprint or editing correspondence should be sent to:

Blake S. Wilson
Neuroscience Program Office
Research Triangle Institute
Research Triangle Park, NC 27709

Telephone: (919) 541-6974

ABSTRACT

This paper reviews considerations in the design of speech processors for cochlear prostheses. Background material is presented on aspects of normal hearing, patterns of nerve survival in the deaf ear, elements of speech essential for intelligibility, and the psychophysics of direct electrical stimulation of the cochlea. Then, to illustrate issues of processor design, two fundamentally different processing strategies are described in terms of the information that they can convey and in terms of how they might perform under various conditions of nerve survival. A summary of clinical tests comparing these strategies in eight implant patients follows. Key findings from the comparison studies are that (a) one processor is clearly superior for patients with psychophysical signs of poor nerve survival, (b) the other processor may be superior for patients with signs of good survival, and (c) different processing strategies can produce widely different outcomes for individual patients. The implications of these findings are discussed with an emphasis on interpretations relating to nerve survival. Future directions in the further development of speech processors for cochlear prostheses are outlined in a final section.

INTRODUCTION

The design of speech processors for cochlear prostheses is a multifaceted activity. At the most basic level such processors must extract or preserve from speech those parameters that are essential for intelligibility and then encode those parameters for electrical stimulation of the auditory nerve. Areas of knowledge necessary to the informed design of speech processors for cochlear prostheses include electrical engineering and the speech and hearing sciences. The range of options for processor design is quite large -- a latitude reflected in the many approaches that have been taken in the design of clinically-applied devices [1]-[4].

A remarkable finding from evaluations of these approaches is that each of several distinctly different processing strategies can produce high levels of speech perception in some patients. Unfortunately the converse is also true, in that poor levels of performance are found for other patients using the same strategies. A likely contributor to this disparity in performance levels is variation in nerve survival among the inner ears of different implant patients. In particular, a person blessed with excellent nerve survival may be able to make good use of a wide variety of inputs, including those provided by various contemporary cochlear prostheses, while a person with poor nerve survival may not be able to utilize information from many of these same inputs.

Our work in developing speech processors has focused on determining optimal strategies for implant patients with varying degrees of nerve survival. This work has led to the development of a new class of processing strategies that has provided improved performance for patients with signs of poor nerve survival, and has demonstrated that access to a variety of alternative strategies is required to obtain the best results across the entire patient population. In this paper we will summarize some of our work illustrating current issues in processor design. We hope the examples we present will help to (a) emphasize the potential importance of variations in nerve survival as a consideration in the design and application of speech processors and (b) provide the reader with a useful introduction to problems and opportunities in the further development of these devices.

BACKGROUND

Aspects of Normal Hearing

A highly simplified diagram of the peripheral auditory system is shown in Fig. 1. In normal hearing, sound waves in air cause vibrations of the eardrum that are conveyed to the inner ear via the three small bones of the middle ear. These middle ear bones act as a mechanical transformer to match impedances between sound transmission in air and sound transmission in the fluid-filled inner ear. Sound vibrations in the fluids displace a flexible membrane suspended in the spiral-shaped inner ear (or cochlea). Gradations in the stiffness and width of this membrane result in changes in frequency response as a function of position along the length of the cochlear spiral. The maximum displacements in response to high frequency stimuli occur at the basal positions, near the middle ear end of the cochlea, while maximum displacements in response to low frequency stimuli occur at apical positions, near the apex of the spiral. Approximately 3000 rows of sensory hair cells line the top of the basilar membrane, the hairs of these cells bending in response to shearing displacements with respect to the overlying tectorial membrane. This bending of the hairs in turn releases chemical transmitter substance at the base of the hair cells. Sufficient concentrations of transmitter substance will cause adjacent neurons to fire, signaling the presence of excitation at some particular site in the inner ear. Each neuron consists of a peripheral dendrite, a ganglion cell body, and a central axon. There are approximately 30,000 neurons in the normal human ear. These neurons, collectively known as the auditory or cochlear nerve, connect with structures in the brain that shape and interpret this input for such specific purposes as pitch discrimination, sound localization, and speech perception.

Patterns of Nerve Survival

Disruptions in any of these links between sound waves in air and auditory nerve discharge patterns can produce hearing impairment or deafness. One cause of hearing impairment, for example, is the gradual immobilization of middle ear components by bone growth, greatly reducing the efficiency of sound transmission to the inner ear. Fortunately, such deficits in hearing often can be overcome with conventional hearing aids or surgical procedures.

Another possible cause of hearing impairment is loss of sensory hair cells or auditory neurons. The hair cells are particularly fragile and can be damaged by exposure to loud sounds, by a wide range of drug treatments, by congenital disorders and by certain diseases. Moreover, damage to the hair cells and their supporting structures can lead to subsequent degeneration of the adjacent dendrites and ganglion cells. If large numbers of sensory hair cells or auditory neurons are destroyed, the connection between brain and vibrating structures is severed and the person with such a loss is rendered profoundly deaf. Fortunately, present evidence indicates that loss of sensory hair cells is a far more common cause of deafness than loss of auditory neurons [5]. In such cases direct electrical stimulation can excite the remaining neurons in ways that convey useful information to the brain. The idea of bypassing the entire vibratory path and the hair cells is the founding concept of cochlear prostheses as a treatment of deafness.

Although loss of sensory hair cells may be the principal cause of deafness, it is often accompanied by at least some loss of auditory neurons. This situation is depicted in Fig. 2. In the left panel is the pristine set of hair cells, dendrites, ganglion cells and central axons characteristic of normal hearing, while in the right panel the sparse and uneven survival of these elements typifies the deaf ear. A fundamental problem faced by designers of cochlear prostheses is the variability in neural survival patterns among deaf ears. This problem is particularly difficult and significant in the design of speech processors for use with implanted arrays of spatially-selective electrodes. In such systems a key objective is to exploit the "place" coding for frequency of stimulation in the normal auditory system. That is, when high frequency inputs are detected by the speech processor then appropriately-encoded stimuli are delivered to electrodes near the basal end of the cochlea, while low frequency inputs result in electrical stimulation near the apical end. The electrode array used in the majority of our studies has eight pairs of closely-spaced bipolar contacts [6]. The positions of the pairs span locations that in the normal ear correspond to a range of place frequencies from about 800 Hz to about 6000 Hz. In an implanted ear with excellent nerve survival near these positions, eight distinct sectors of excitation can be realized with a high degree of independence [7], [8]. This sector-by-sector control of nerve activity would be expected to provide a good representation of the spectral content of an acoustic input signal through appropriate adjustment of stimulus intensities at each electrode channel. However, in the more typical deaf ear, with less than complete nerve survival, such a representation of frequencies as a place code obviously will be distorted in areas of neural loss. The alternatives for dealing with areas of loss are (a) to leave a "hole" in the frequency-place map in such areas or (b) to raise stimulus levels to recruit more distant but still viable neurons. In either case the positions of excitation no longer simply reflect the frequency content of the acoustic input signal.

Another, probably more serious, difficulty posed by poor nerve survival is interaction among stimulated channels. Because the loudness of auditory sensations is likely to depend in large part on the total number of activated auditory neurons [9], one might expect that relatively high levels of electrical stimulation would be required in regions of poor nerve survival to evoke comfortably-loud percepts. In systems with multichannel electrode arrays these high levels of stimulation will almost certainly produce substantial overlaps among sectors stimulated by different electrodes [10]-[12]. The resulting excitation of common subpopulations of neurons -- the problem of channel interactions -- can severely limit the performance of multichannel prostheses.

Elements of Speech

A simple but useful model of speech production is illustrated in Fig. 3. This "source-filter" model recognizes the first-order independence of the excitation of the vocal tract and its response. Unvoiced components of speech are produced with a source of broadband turbulent noise. This noise is generated either by forcing air through a narrow constriction (for production of unvoiced fricatives like "s") or by building air pressure behind an obstruction and suddenly releasing this pressure when the obstruction is removed (for production of plosives like "t"). The spectral characteristics of this broadband noise then are modified by transmission through the vocal tract and by radiation of sound at the lips.

In contrast, voiced components of speech are produced by "ringing" the vocal tract with puffs of air released through the vibrating folds of the glottis. The shape of the vocal tract, in terms of tongue, lip, jaw and velum position, contributes resonances that influence sound transmission in the tract. The transfer function of the vocal tract is the "filter" of the source-filter model, and the broad spectral peaks in this transfer function are called "formants." The frequencies of the first two formants convey sufficient information for recognition of vowels and distinctions among other voiced sounds.

A third class of speech sounds is produced by the simultaneous combination of periodic glottal puffs and aperiodic noise sources. Two major categories of these sounds are voiced plosives and voiced fricatives. The addition of voicing can change a "t" sound to a "d" sound or an "s" sound to a "z" sound.

Experiments with models of the type shown in Fig. 3 have demonstrated that small sets of parameters can provide adequate specification for the production of intelligible speech [13], [14]. In general, the parameters must convey information on the source of the sound (i.e., periodic, aperiodic or mixed) and on the transfer function of the vocal tract. For voiced speech sounds this transfer function can be adequately specified by the first two or three formants, while for unvoiced speech sounds an indication of overall spectral shape is sufficient. Updating such parameters every 5 to 10 ms allows the production of intelligible speech. In fact, the parameters specifying the transfer function of the vocal tract may be quantized along rather coarse scales. The information rate required to transmit these parametric data can be as low as 1000 bits/s, which is far less than the 30,000 bits/s required for ordinary voice transmission over a typical telephone channel [13]. With such crude representations of speech being adequate to preserve intelligibility for listeners with normal hearing, one might expect that the same representations would be sufficient to convey intelligible speech through a cochlear prosthesis. The difficulty is in presenting the parameters in such a way that an implant patient can

(a) independently perceive each parameter over its full range and (b) discriminate a minimum number of steps within this range.

Implant Psychophysics

At least two mechanisms are likely to underlie the perception of pitch in listeners with normal hearing [9], [15]. One is the "place code" reviewed above. The other is "volley code," based on phase locking among auditory nerve responses for stimulus frequencies below 5 kHz. That is, each neuron of an ensemble of neurons preferentially responds at a particular phase of a sinusoidal stimulus. Even though each neuron does not respond to every period of the stimulus (except at very low stimulus frequencies), the summed ensemble volley of responses that reaches the brain contains intervals reflecting the frequency of stimulation. A slight jitter limits the ability of neurons to phase lock and "smears" the volley code representation above 4-5 kHz. In normal hearing both the place of maximal vibration along the basilar membrane and the rate of vibration at that point probably are conveyed to the brain for stimulus frequencies below about 5 kHz. At higher frequencies the place code may be the sole mechanism for pitch perception.

Both the place and volley codes appear to have at least some salience for implant patients. For a given electrode site and loudness level, perceived pitch follows the frequency of sinusoids, or the rate at which pulses are delivered, up to a "pitch saturation" limit, typically about 300 Hz [16]. Although a few exceptional patients can discriminate different frequencies of stimulation up to 1000-2000 Hz [17], the great majority of patients cannot discriminate frequencies above the 300 Hz limit.

As might be expected on the basis of our previous description of normal hearing, distinct tonal sensations can be evoked by varying the site of stimulation along the cochlear spiral. In general, stimulation at the basal end of the cochlea elicits a sharp or high-pitched percept and stimulation at the apical end elicits a dull or low-pitched percept. However, this ranking of pitches is not robust in all patients [16], [18]. In addition, some patients exhibit a nonmonotonic relation between perceived pitch and the position of stimulation along the cochlear spiral. A probable cause for both these deficiencies is poor or patchy nerve survival. Indeed, Pfingst and coworkers have demonstrated a strong correlation between the ability of their test monkeys to rank pitch according to electrode position and the histological picture of nerve survival obtained after the monkeys were sacrificed [19]. In all cases the ability to rank electrodes was severely degraded in regions of poor nerve survival.

Another major dimension along which electrical stimuli are perceived by implant patients is loudness. In general, loudness varies with the intensities and waveforms of electrical stimuli. Low frequency sinusoids around 100 Hz have the lowest thresholds and greatest dynamic ranges (between threshold and uncomfortably loud percepts) of all stimulus waveforms [16], [18]. The dynamic range for 100 Hz sinusoids can be as large as 40 dB. In contrast, the dynamic ranges for high frequency sinusoids (e.g., 500 Hz and higher) and short-duration pulses (e.g., 200 μsec/phase) are much narrower. A typical range for short-duration pulses, for example, would be 10-15 dB. All of these dynamic ranges for electrically-evoked hearing are substantially lower than the ranges for normal hearing, which are 100 dB or greater for a wide variety of acoustic stimuli.

PROCESSING STRATEGIES

In the remainder of this paper we will explore in some detail the implications of nerve survival for processor design and performance. In this section we will describe two different processing strategies in terms of the output signals they deliver to a multichannel electrode array. One of these strategies implicitly assumes good nerve survival in the implanted ear whereas the other strategy is designed specifically to avoid the deleterious effects of channel interactions in ears with poor nerve survival.

Compressed Analog Processors

A widely-applied processing strategy for multichannel cochlear prostheses is illustrated in Fig. 4. We call it the compressed analog (CA) strategy because the basic functions of the processor are to compress wide-dynamic-range speech input signals onto the narrow dynamic range available for electrical stimulation of the ear, and then to filter the compressed signal into individual frequency bands for presentation to each electrode's sector of the cochlea. Typical waveforms of such a processor are shown in the figure. The top trace in each panel is the input signal, which in this case is the word "BOUGHT." The other waveforms in each panel are the filtered output signals for 4 channels of intracochlear stimulation. The bottom left panel shows an expanded display of waveforms during the initial part of the vowel in BOUGHT, and the bottom right panel shows an expanded display of waveforms during the final "T." The lower panels in Fig. 4 thus exemplify differences in waveforms for voiced and unvoiced intervals of speech.

In the voiced interval the relatively large outputs of channels 1 and 2 reflect the low-frequency formant content of the vowel, and in the unvoiced interval the relatively large outputs of channels 3 and 4 reflect the high-frequency noise content of the "T." In addition, the clear periodicity in the waveforms of channels 1 and 2 reflects the fundamental frequency of the vowel during the voiced interval, and the lack of periodicity in the outputs of all channels reflects the noise-like quality of the "T" during the unvoiced interval. Features represented in this way are perceived to varying degrees by different implant patients. A principal concern is that simultaneous stimulation with continuous waveforms results in summation of the electric fields from the individual electrodes. This summation can exacerbate interactions between channels, especially for patients who require high stimulation levels. Summation of stimuli between channels also depends on the phase relationships among the waveforms. Because these relationships are not controlled in a CA processor, representation of the speech spectrum may be further distorted by continuously changing patterns of channel interaction. Therefore, one might expect that CA processors would work best for patients with low stimulation thresholds and good isolation between channels, i.e., for patients with relatively good nerve survival and electrode placement.

Interleaved Pulses Processors

The problem of channel interactions is addressed in the processor of Fig. 5 through the use of nonsimultaneous stimuli, brief pulses delivered sequentially to the different channels. There is no overlap between the pulses so that direct summation of electric fields produced by different electrodes is avoided. The energies of frequency components in speech are represented by the amplitudes of the pulses, and distinctions between voiced and unvoiced segments of speech are represented by the timing of cycles of stimulation across the electrode array. In this particular processor stimulus cycles are timed to begin in synchrony with the detected fundamental frequency for voiced speech sounds and at randomly spaced intervals for unvoiced speech sounds. The periodicity of stimulus cycles for a voiced speech sound can be seen in the lower left panel while the randomly spaced stimulus cycles indicating an unvoiced speech sound are illustrated in the lower right panel.

Because simultaneous stimulation of channels is avoided in this "interleaved pulses" (IP) processor, one might expect that its use could improve performance for patients with severe channel interactions.

EVALUATION STUDIES

We have studied relative levels of performance with the CA and IP processing strategies in eight implant patients using a broad range of speech tests. All patients had been implanted with the multichannel electrode array developed at the University of California at San Francisco (UCSF) and subsequently manufactured by Storz Instrument Company of St. Louis. This array has eight pairs of bipolar electrodes, with a 2 mm spacing between pairs [6]. In ears with good nerve survival such an array would be expected to allow a high degree of spatial selectivity in the excitation of auditory neurons [7], [8].

From the outset an important aim of the evaluation studies was to compare alternative processing strategies in tests with individual implant patients. In this way we could provide hitherto unrealized controls for differences among patients in (a) electrode placement, (b) the patterns of nerve survival in the implanted ear, (c) the integrity of the brain pathways associated with hearing, and (d) cognitive skill and language acquisition. In addition, our comparisons of processing strategies with individual patients would share a single type of electrode array and uniformly administered tests.

The first two patients studied were fitted with percutaneous cables and the remaining six with the four-channel transcutaneous transmission system (TTS) of the UCSF/Storz prosthesis [20]. Use of the cable allows direct access to all 16 electrode contacts in the array (usually configured as eight bipolar pairs) and direct control over the current or voltage waveforms of the stimuli. In contrast, alternating bipolar electrode pairs are assigned to the four channels of the TTS, and the current and voltage waveforms are complex functions of the nonlinear impedances of the electrodes and the frequency bands of the TTS channels.

Because the cable allowed a much greater degree of stimulus control, many different processing strategies were evaluated in the studies with the first two patients (LP and MH). The performance of each strategy was measured with confusion matrix tests. The confusion matrix for vowels included the tokens "BOAT," "BEET," "BOUGHT," "BIT," and "BOOT," and the confusion matrix for consonants included the nonsense tokens "ATA," "ADA," "AKA," "ASA," "AZA," "ANA," "ALA," and "ATHA." The tokens in the vowel matrix were selected to measure the ability to perceive differences in the second formant (F2) of the vowels, and the tokens in the consonant matrix were selected to measure the ability to distinguish the nonvisible consonants that have the greatest frequency of occurrence in spoken English [21].

Among the processing strategies evaluated with these first two patients, the IP approach achieved the best results. To further assess the performance of this strategy vis à vis the standard CA strategy of the UCSF/Storz prosthesis, an extensive series of standard tests was designed for the remaining six patients (all of whom were fitted with the TTS). In addition to the vowel and consonant confusion tests just outlined this series included: all subtests of the Minimal Auditory Capabilities (MAC) battery [22]; the Diagnostic Discrimination Test (DDT) of consonant confusions [23]; connected discourse tracking (verbatim repetition by the patient of a text read by the investigator) with and without the prosthesis [24], [25]; and the IOWA test of medial consonant identification with lipreading cues [26]. In this paper we will briefly review the results from the vowel and consonant confusion tests for all patients, and from the MAC and IOWA tests for the six patients fitted with the TTS. Detailed reports of the evaluation studies are available elsewhere [27], [28], [29], [30].

Patient LP

The first patient (LP) had a most discouraging picture of psychophysical performance. He had extremely severe channel interactions and high thresholds for bipolar electrical stimulation. His case was further complicated by extraordinarily narrow dynamic ranges and lability of thresholds and loudness levels both within and between testing sessions. LP's psychophysical findings of severe channel interactions, high thresholds and narrow dynamic ranges were all consistent with a picture of very poor survival of neurons in his implanted ear [12], [19], [29], [31], [32], [33].

As might be expected, LP received no benefit from the CA processor used in the standard UCSF prosthesis. Indeed, he refused to describe any of the percepts produced with this processor as speechlike.

The first application of a 6-channel, IP processor immediately moved LP into the speech mode of auditory perception [27], [29]. Of the 11 speech tokens initially presented to LP using the IP processor, 7 were spontaneously recognized as the correct words or syllables. Although his performance declined when the number of channels was reduced from 6 to 4, formal tests with vowel confusion matrices indicated that LP could perform at a level well above chance even with a reduced 4-channel version of the IP processor [27]. A medical complication required surgical removal of LP's implant, ending our study of him with these very encouraging preliminary results.

Patient MH

With the second patient we were able to evaluate differences in processor performance in much greater detail. This patient also had psychophysical manifestations of poor nerve survival [28], [29]. The results presented in Fig. 6 show her levels of performance in formal tests of vowel and consonant recognition. The diagonally hatched bars show her performance with lipreading, and the cross hatched bars without. The chance levels of performance are indicated by the horizontal lines in each panel. Different processors are represented by different sets of bars. The characteristics of each processor are indicated in the labels at the bottom. For example, the leftmost set of bars shows the scores for a 4-channel CA processor. The remaining sets of bars show the scores for four variations of IP processors. These variations were produced by manipulating (a) the number of stimulation channels and (b) the way in which the beginnings of stimulus sequences were timed. In one approach stimulus sequences were timed to start in synchrony with the fundamental frequency for voiced speech sounds and at randomly-spaced intervals during unvoiced speech sounds. This constituted explicit coding of fundamental frequency and voice/unvoice distinctions. In the other approach stimulus sequences were timed to follow each other as rapidly as possible, providing no explicit coding of voicing information. In all, the results shown in Fig. 6 allow direct comparisons of (a) a 4-channel CA processor vs. a 4-channel IP processor; (b) 4- vs. 6-channel IP processors; and (c) IP processors with and without explicit coding of voicing information. The comparisons indicate that:

1. Performance is markedly improved when a 4-channel IP processor is used instead of a 4-channel CA processor;
2. Scores are <u>much</u> higher in all categories except vowel identification with lipreading (where scores are about the same) when a 6-channel IP processor is used instead of a 4-channel IP processor; and

3. Explicit coding of voicing information improves the performance of IP processors, particularly in the categories of vowel identification without lipreading (4-channel processor), consonant identification without lipreading (6-channel processor) and consonant identification with lipreading (both processors).

Patients Fitted with the Transcutaneous Transmission System (TTS)

The studies with the two cables patients demonstrated that (a) different processing strategies can produce widely different outcomes for individual patients and (b) IP processors are far superior to the tested alternative processors for at least two patients with psychophysical signs of poor nerve survival. With these observations in mind, we were most interested in comparing the IP and CA strategies in extensive tests with a larger population of patients. We wondered, for example, how the IP processor would perform for successful users of a CA processor, and whether the potential advantages of the IP processors could be realized in patients with four or fewer channels of stimulation.

Six patients fitted with the 4-channel UCSF/Storz TTS participated in the follow-up studies. Each patient was studied for a one-week period in which (a) basic psychophysical measures were obtained on thresholds and dynamic ranges for pulsatile stimuli, (b) a variety of IP processors (with different choices of processor parameters) was evaluated with tests of vowel and consonant confusions, and (c) the best of these IP processors was evaluated using a broad spectrum of speech tests. As mentioned before, the speech tests included all subtests of the Minimal Auditory Capabilities (MAC) battery [22]; the Diagnostic Discrimination Test (DDT) of consonant confusions [23]; connected discourse tracking [24], [25]; and the IOWA test of medial consonant identification with lipreading cues [26]. Comparison data for the CA processor were obtained from MAC, tracking and DDT tests administered on a previous occasion (within the 3 months preceding the IP tests) as part of the clinical trials of the UCSF/Storz prosthesis. Repeat tests with the CA processor were administered in all cases where a significant difference in scores was found between the two types of processor. Finally, the vowel and consonant confusion tests and the IOWA test were administered by us for both processors.

All six patients had had substantial experience with the CA processor at the time of our studies. We expected that the learning effects of such experience would significantly favor the CA processor in the comparisons (see refs [34] and [35] for further discussions on this point). In a typical case, experience with the CA processor would approximate one year of daily use, while experience with the real-time implementation of the IP processor would be between 15 and 30 minutes before formal testing. Therefore, the information provided by an IP processor would have to be immediately accessible to the patient in order for the results to be at all comparable to those obtained with the CA processor.

An additional factor weighing against the IP processor in these performance comparisons was the use of the 4-channel TTS. The principal limitations of that system for IP processors are (a) inadequate levels of voltage compliance for stimulation with short-duration pulses, (b) the small number of channels, and (c) lack of current control in the stimulus waveforms. Extensive measurements (with patient MH; see [28], [29]) of performance changes with parametric manipulations in IP processors indicate that good performance appears to depend on the following factors, in approximate order of decreasing importance:

1. Total number of channels (large increases in performance are found when the number of channels is increased from 2 to 4 and from 4 to 6);

2. Number of channels updated per stimulus cycle (performance in tests of consonant identification declines precipitously if this number falls below 4);
3. Total duration of each stimulus cycle (performance gets better as duration is decreased, and is markedly better when the duration is less than 4-5 msec);
4. Time between sequential pulses (performance improves as the time between pulses is increased, up to the point at which the total duration of the stimulus cycle begins to exceed 4-5 msec); and
5. Explicit coding of voicing information (performance is better with explicit coding of voicing information, and the percepts elicited with processors that use such coding generally are described as more natural and speechlike).

Notice that the small number of channels and limited voltage compliance of the TTS place severe restrictions on meeting criteria 1-4 above. Also, the lack of current control introduces distortions in stimulus waveforms that may make it more difficult to avoid channel interactions.

A further limitation in number of channels was posed by particular patients in this study. Half of them had fewer than four functional channels. The loss of one or two channels in each of these cases has been attributed to fluid or particulate contamination admitted to the connector assembly of the implanted portion of the TTS during surgery [20]. Although the problem of contamination has been solved by modifying the surgical procedure and the design of the connector assembly, most of these six patients were implanted before the problem was evident and before these revisions had been made.

The parameters selected for the IP processors of these six TTS patients are presented in Table I. The best fulfillments of the five criteria were obtained for patients HE and MC2. Each had the use of all four stimulus channels and the time between sequential pulses was a relatively long 0.5 msec for both patients. In addition, the stimulus cycle time for MC2 was almost within the 5.0 msec criterion for this parameter.

In contrast, relatively poor sets of parameters had to be used for the remaining patients. Patients MC1 and JM had only three useable channels and patient RC only two. The cycle times remained excessive for patients JM and ET, even with the times between pulses compromised down to only 0.1 msec.

The results of processor comparisons are presented in Figs. 7-11. Fig. 7 shows performance levels for the tests of vowel and consonant identifications. Solid black bars represent the scores for lipreading alone, dotted bars for the CA processor alone, and cross-hatched bars for the CA processor with lipreading. Vertically-lined bars represent the scores for the IP processor alone and horizontally-lined bars for the IP processor with lipreading. Chance levels of performance are indicated by the horizontal lines in each panel.

Performance in the tests of vowel identification is quite high for both processors. All patients have scores of 80% correct or better for both processors with lipreading and 68% or better without lipreading. The means of the results for the two processors are not significantly different for either condition.

In contrast to the apparent equivalence of processors for the vowel test, the scores for the two conditions of the consonant test are significantly higher for the IP processor. For the "with lipreading" condition the scores for MC1, HE, RC and MC2 are substantially higher with the IP processor, while the scores for JM and ET are slightly lower. All patients except JM have higher scores using the IP processor in the "processor alone" condition. Especially large differences are found for HE and ET for this test. The scores for JM are approximately the same for the two processors.

An indication of superior consonant recognition with the IP processor is also seen in the results for the IOWA test of medial consonant identification with lipreading cues. These results are presented in Fig. 8. Every patient except MC2 has substantially higher scores with the IP processor. However, the difference in the means of the results for the two processors just fails to attain statistical significance. It is important to note that our procedure for administering this test was designed to confer any benefits of learning on the CA processor. The order of testing was to measure performance first with the IP processor plus lipreading, then with lipreading alone, and finally with the CA processor plus lipreading.

The remaining results presented in Figs. 9-11 are those from the MAC battery. Results from the subtests of prosodic perception (timing of syllable boundaries, voice fundamental frequency and word stress) are shown in Fig. 9. These results demonstrate that, in general, these subtests are too easy for the patients and processors under study. The scores for the noise/voice (N/V) and spondee same/different (Sp S/D) tests are very high for all patients and for both processors. (Spondees are two syllable words like "cowboy" with equal emphasis on the syllables.) The accent test is a more sensitive indicator of performance. Scores for this test show clear differences among patients, but relatively modest differences between processors. Three patients did moderately better with the IP processor on this test, (RC, ET, and MC2) and two patients did moderately better with the CA processor (MC1 and JM). Finally, for the Question/Statement (Q/S) test one patient did much better with the CA processor (MC1), three patients did somewhat better with the IP processor (HE, JM and RC) and two patients obtained identical scores with the two processors (ET and MC2). None of the differences in the means of the results for the two processors is statistically significant among the prosodic subtests of the MAC battery.

Results from the phoneme and word subtests of the MAC battery also demonstrate a general equivalence of the processors for the conditions of our study. These results are presented in Fig. 10. Again, none of the differences in the means of the results for the two processors is statistically significant. However, patients HE and MC2 have somewhat higher or equivalent scores on all four tests (including vowel, initial consonant, final consonant, and four-choice spondee) with the IP processor, and patients MC1 and RC have somewhat higher or equivalent scores with the CA processor.

Finally, results from the subtests of open-set recognition are shown in Fig. 11. Once again, none of the differences in the means of the results for the two processors is statistically significant. Patient MC2, however, has higher scores for the IP processor for all four subtests of spondee recognition (Sp), recognition of monosyllabic words (NU6), recognition of CID sentences (CID), and recognition of words in context (WIC). In addition, patient HE has much higher scores with the IP processor for the tests of spondee recognition and of words in context. On the other hand, patient RC has higher scores with the CA processor on every open-set test, and patients MC1 and ET show generally superior performance with the CA processor for those tests.

Summary of Results

In the studies with the two cable patients, each of whom had a bleak psychophysical picture consistent with poor nerve survival, the IP strategy produced much better results than the CA strategy. We believe this improved performance with the IP processor is attributable in part to the sizable reduction in channel interactions afforded by the use of nonsimultaneous stimuli. The studies with the cable patients further demonstrated a very strong correspondence between number of stimulation channels and performance. In the studies with patients fitted with the TTS, the IP and CA processors were compared under conditions of substantial experience with the CA processor, generally high levels of performance with the CA processor, and severe restrictions imposed by the TTS for implementing optimized versions of the IP processor. Despite these limitations, two patients in this second series immediately had better performance with the IP processor (patients HE and MC2) and three had similar or slightly inferior levels of performance with this processor (patients MC1, JM and ET). Only one patient (RC) had clearly superior performance with the CA processor. This patient also happened to have the highest level of performance among the six studied patients with the CA processor, and afforded the poorest fulfillment of our IP processor fitting criteria.

These results suggest that (a) most patients are likely to obtain at least equivalent results if an IP processor is used instead of a CA processor; (b) patients with psychophysical signs of poor nerve survival are likely to obtain better results with an IP processor; and (c) use of a TTS designed to support an IP processor (e.g., a TTS with eight channels of current-controlled outputs) is likely to produce results that are better than those obtained with the limited TTS of the present studies.

DISCUSSION

A key finding of the studies reviewed in this paper is that substantial gains in speech perception can be made by selecting an appropriate processing strategy for each patient. In our two series of patients RC obtained clearly superior results with the CA processing strategy while LP and MH obtained clearly superior results with the IP processing strategy.

These results demonstrate that access to a variety of alternative strategies may be required for optimizing the outcomes across a population of patients.

The patient who obtained superior results with the CA processor (RC) had only two functional channels of intracochlear stimulation. This number of channels is certainly too few for even a gross representation of the speech spectrum with an IP processor. The relatively poor performance of the IP processor therefore could be attributed to a poor fulfillment of its fitting criteria.

An alternative explanation for the superior performance of the CA processor is that RC made especially good use of the information present in the compressed analog waveforms. Indeed, the impressive results obtained with RC (2 channels), MC1 (three channels), certain patients in the Vienna series (1 channel; see [17]) and certain patients in the Symbion series (4 monopolar channels with relatively poor isolation; see [36]) support the hypothesis that the major bearer of information in CA processors is the waveform itself. Although results from studies conducted at UCSF demonstrate that additional information can be provided with four channels of stimulation using the UCSF electrode array [37]-[39], this additional information is obviously not required for superb performance in some patients. Most likely, the best results are obtained for patients who have the greatest access to information in the CA waveform(s). These patients might include those with exceptional abilities to discriminate (a) frequencies up through the range of the first formant of speech [17], [36], [40]; (b) rapid temporal variations in the envelopes of speech and speechlike stimuli [41]; and (c) subtle waveshape changes produced by the addition of frequency components beyond the first formant [42], [43].

If this second interpretation is at least partly correct for RC's case, then patients with such special abilities might be best served with a CA processor. Optimal implementations of such a processor would provide any additional information the patient might be able to utilize in multiple channels of stimulation. The maximum number of useful channels is likely to be limited, however, by the severe interactions that can occur between closely spaced electrodes when simultaneous stimuli are used.

In contrast, patients with a psychophysical picture consistent with poor nerve survival (i.e., severe channel interactions, high thresholds, narrow dynamic ranges, and perhaps limited abilities to discriminate frequencies and other stimulus attributes) are likely to receive greater benefit from an IP processor. For two such patients in our initial series large increases in performance were obtained when the number of stimulus channels was increased from 2 to 4 and from 4 to 6. Paradoxically, then, multichannel implants may provide relatively greater benefits to patients with signs of poor nerve survival than to patients without these signs.

To summarize the comparisons made above, Table II lists characteristics of the CA and IP processing strategies. Briefly, the CA strategy may be superior for patients with good nerve survival because such patients may perceive substantial temporal and frequency information in analog waveforms and because the lower stimulus intensities required for these patients, along with survival of ganglion cells and dendrites over the active electrodes, can greatly minimize channel interactions produced by simultaneous stimulation. On the other hand, the IP strategy may be superior for patients with poor nerve survival because isolation between channels for such patients is tremendously improved with the use of nonsimultaneous stimuli.

Finally, we note that these comparisons between processors suggest possibilities for further improvements in performance. One such improvement might be made by combining the best features of the CA and IP approaches in "hybrid" strategies. For a good nerve survival case, for example, the main benefits of the CA strategy might be realized with a single channel of stimulation. This would leave the remaining channels for the representation of frequency components in speech above the first formant. The excellent results obtained in the present studies with all eight patients using the IP processor (especially patients HE, MC2 and MC1) indicate that interleaved stimuli are likely to enhance speech representation even for patients with good nerve survival (presumably through further reduction of interactions between adjacent channels). The combined use of the CA and IP strategies therefore might confer in an optimal way the benefits of waveform discrimination and multichannel stimulation to fortunate patients with good nerve survival. Similarly, for cases in which nerve survival is patchy, psychophysical or electrophysiological tests might be conducted to identify areas of good survival. A bipolar pair of electrodes adjacent to one of these areas could be selected for compressed analog stimulation. This electrode channel would have low threshold and suprathreshold stimulus levels relative to electrodes adjacent to poor survival areas. Such low levels might allow the remaining electrode channels to receive IP stimuli with only minor channel interactions.

Potential applications of hybrid processors, along with the choices posed by the existing CA and IP strategies, emphasize the need for flexibility in the fitting of speech processors to individual patients. We believe further significant advances in the development of speech processors for cochlear prostheses will result from (a) an improved understanding of the electrode-nerve interface, especially as it relates to the pattern of nerve survival, and (b) design and application of better psychophysical and electrophysiological tests to infer the pattern of survival in the implanted ear.

CONCLUSIONS

The major conclusions from the studies reviewed in this paper are the following:

1. Different processing strategies can produce widely different outcomes for individual implant patients;
2. Interleaved pulses (IP) processors are far superior to the tested alternative processors for at least two patients with psychophysical signs of poor nerve survival;
3. The performance of IP processors strongly depends on the selection of processor parameters;
4. Processors other than the IP processors can be superior for patients with psychophysical signs of good nerve survival and for patients who cannot be fit with an optimized IP processor;
5. One such processor is the compressed analog (CA) processor of the present UCSF/Storz cochlear prosthesis; and
6. Substantial gains in speech understanding can be made by (a) selecting the best type of speech processor for each patient and (b) using implanted and external hardware capable of supporting a wide range of different processing strategies.

ACKNOWLEDGMENTS

We thank the patients who participated in the evaluation studies for their dedicated effort and pioneering spirit. We are pleased to acknowledge the important scientific contributions of M.W. White, M.M. Merzenich, R.A. Schindler, F.T. Hambrecht, G.E. Loeb, D.K. Kessler, B.A. Weber, J.C. Farmer, Jr., P.D. Kenan, R.V. Shannon, M.W. Skinner and L. Vurek. This work was supported by NIH contracts N01-NS-3-2356 and N01-NS-5-2396, through the Neural Prosthesis Program.

REFERENCES

[1] J.B. Millar, Y.C. Tong and G.M. Clark, "Speech processing for cochlear implant prostheses," J. Speech Hear. Res., vol. 27, pp. 280-296, 1984.

[2] B.C.J. Moore, "Speech coding for cochlear implants," in Cochlear Implants, R.F. Gray, Ed. San Diego, CA: College-Hill Press, 1985, pp. 163-179.

[3] C.W. Parkins, "Cochlear prostheses," in Neurobiology of Hearing: The Cochlea, R.A. Altschuler, D.W. Hoffman and R.P. Bobbin, Eds. New York: Raven Press, 1986, pp. 455-473.

[4] B.E. Pfingst, "Stimulation and encoding strategies for cochlear prostheses," Otolaryngol. Clin. N. Amer., vol. 19, pp 219-235, 1986.

[5] R. Hinojosa and M. Marion, "Histopathology of profound sensorineural deafness," Ann. N. Y. Acad. Sci., vol. 405, pp. 459-484, 1983.

[6] G.E. Loeb, C.L. Byers, S.J. Rebscher, D.E. Casey, M.M. Fong, R.A. Schindler, R.F. Gray and M.M. Merzenich, "Design and fabrication of an experimental cochlear prosthesis," Med. & Biol. Eng. & Comput., vol. 21, pp. 241-254, 1983.

[7] M.M. Merzenich and M.W. White, "Cochlear implant: The interface problem," in Functional Electrical Stimulation: Applications in Neural Prostheses, F.T. Hambrecht and J.B. Reswick, Eds. New York: Marcel Dekker, 1977, pp. 321-346.

[8] C. van den Honert and P.H. Stypulkowski, "Single fiber mapping of spatial excitation patterns in the electrically stimulated auditory nerve," Hearing Res., vol. 29, pp. 195-206, 1987.

[9] J.O. Pickles, An Introduction to the Physiology of Hearing, London: Academic Press, 1982, ch. 9, pp. 275-277.

[10] R.V. Shannon, "Multichannel electrical stimulation of the auditory nerve in man. II. Channel interaction," Hearing Res., vol. 12, pp. 1-16, 1983.

[11] R.V. Shannon, "Loudness summation as a measure of channel interactions in a cochlear prosthesis," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven Press, 1985, pp. 323-334.

[12] M.W. White, M.M. Merzenich and J.N. Gardi, "Multichannel cochlear implants: Channel interactions and processor design," Arch. Otolarynol., vol. 110, pp. 493-501, 1984.

[13] J.L. Flanagan, Speech Analysis, Synthesis and Perception, Second Edition. Berlin: Springer-Verlag, 1972.

[14] D. O'Shaughnessy, Speech Communication: Human and Machine. Reading, MA: Addison-Wesley, 1987.

[15] B.C.J. Moore, An Introduction to the Psychology of Hearing, London: Academic Press, 1982, ch. 4, pp. 115-140.

[16] B.E. Pfingst, "Psychophysical data from cochlear implants: Relevance to strategies for rehabilitation," Seminars in Hearing, vol. 6, pp. 7-21, 1985.

[17] I.J. Hochmair-Desoyer and K. Burian, "Reimplantation of a molded scala tympani electrode: Impact on psychophysical and speech discrimination abilities," Ann. Otol. Rhinol. Laryngol., vol. 94, pp. 65-70, 1985.

[18] R.V. Shannon, "Multichannel electrical stimulation of the auditory nerve in man. I. Basic psychophysics," Hearing Res., vol. 11, pp. 157-189, 1983.

[19] B.E. Pfingst, I. Glass, F.A. Spelman and D. Sutton, "Psychophysical studies of cochlear implants in monkeys: Clinical implications," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven Press, 1985, pp. 305-321.

[20] R.A. Schindler, D.K. Kessler, S.J. Rebscher, J.L. Yanda and R.K. Jackler, "The UCSF/Storz multichannel cochlear implant: Patient results," Laryngoscope, vol.96, pp. 597-603, 1986.

[21] E.D. Schubert, "Some limitations on speech coding for implants," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven Press, 1985, pp. 269-276.

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

[23] C.B. Grether, "Psychoacoustic assessment of speech communication systems: The diagnostic discrimination test," Project Themis. Air Force Office of Scientific Research, contract number F44620-69-C-0033, Sept. 22, 1970.

[24] C.L. De Filippo and B.L. Scott, "A method for training and evaluating the reception of ongoing speech," J. Acoust. Soc. Am., vol. 63, pp. 1186-1192, 1978.

[25] E. Owens and C. Telleen, "Tracking as an aural rehabilitative process," J. Acad. Rehabil. Audiol., vol. 14, pp. 259-273, 1981.

[26] R.S. Tyler, J.P. Preece and M.W. Lowder, "The Iowa cochlear-implant test battery," Laboratory Report. University of Iowa at Iowa City, Department of Otolaryngology - Head and Neck Surgery, 1983.

[27] B.S. Wilson, C.C. Finley and D.T. Lawson, "Speech processors for auditory prostheses," Seventh Quarterly Progress Report, NIH project N01-NS-3-2356. Bethesda, MD: Neural Prosthesis Program, National Institutes of Health, 1985.

[28] B.S. Wilson, C.C. Finley and D.T. Lawson, "Speech processors for auditory prostheses," Second Quarterly Progress Report, NIH project N01-NS-5-2396. Bethesda, MD: Neural Prosthesis Program, National Institutes of Health, 1986.

[29] B.S. Wilson, C.C. Finley, J.C. Farmer Jr., D.T. Lawson, B.A. Weber, R.D. Wolford, P.D. Kenan, M.W. White, M.M. Merzenich and R.A. Schindler, "Comparative studies of speech processing strategies for cochlear implants," Laryngoscope, in press.

[30] B.S. Wilson, D.T. Lawson, C.C. Finley, R.D. Wolford, D.K. Kessler and R.A. Schindler, "Direct comparisons of analog and pulsatile coding strategies with six cochlear implant patients," in prep.

[31] M.M. Merzenich, P. Leake-Jones, M. Vivion, M. White and M. Silverman, "Development of multichannel electrodes for an auditory prosthesis," Fourth Quarterly Progress Report, NIH project N01-NS-7-2367. Bethesda, MD: Neural Prosthesis Program, National Institutes of Health, 1978.

[32] J.N. Gardi, "Human brain stem and middle latency responses to electrical stimulation: Preliminary observations," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven Press, 1985, pp. 351-363.

[33] B.E. Pfingst and D. Sutton, "Relation of cochlear implant function to histopathology in monkeys," Ann. N. Y. Acad. Sci., vol. 405, pp. 224-239, 1983.

[34] R.S. Tyler, J.P. Preese, C.R. Lansing, S.R. Otto and B.J. Gantz, "Previous experience as a confounding factor in comparing cochlear-implant processing schemes," J. Speech Hear. Res., vol. 29, pp. 282-287, 1986.

[35] R.C. Dowell, P.M. Seligman, P.J. Blamey and G.M. Clark, "Evaluation of a two-formant speech-processing strategy for a multichannel cochlear prosthesis," Ann. Otol., vol. 96 (Suppl. 128), pp. 132-134, 1987.

[36] D.K. Eddington, "Speech recognition in deaf subjects with multichannel intracochlear electrodes," Ann. N. Y. Acad. Sci., vol. 405, pp. 241-258, 1983.

[37] M.T. Ochs, M.W. White, M.M. Merzenich and E.D. Schubert, "Speech recognition in single- and multichannel cochlear implants," J. Acoust. Soc. Am., vol. 77 (Suppl. 1), p. S81, 1985.

[38] M. White, M. Ochs, M. Raggio, D. Morledge and M. Merzenich, "Formant discrimination in a multichannel cochlear prosthesis," J. Acoust. Soc. Am., vol. 77 (Suppl. 1), p. S81, 1985.

[39] R.A. Schindler, D.K. Kessler, S.J. Rebscher, R.K. Jackler and M.M. Merzenich, "Surgical considerations and hearing results with the UCSF/Storz cochlear implant," Laryngoscope, vol. 97, pp. 50-56, 1987.

[40] M.W. White, "Formant frequency discrimination and recognition in subjects implanted with intracochlear stimulating electrodes," Ann. N. Y. Acad. Sci., pp. 348-359, 1983.

[41] I.J. Hochmair-Desoyer, E.S. Hochmair and H.K. Stiglbrunner, "Psychoacoustic temporal processing and speech understanding in cochlear implant patients," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven press, 1985, pp. 291-304.

[42] R.A. Dobie and N. Dillier, "Some aspects of temporal coding for single-channel electrical stimulation of the cochlea," Hearing Res., vol. 18, pp. 41-55, 1985.

[43] E.S. Hochmair and I.J. Hochmair-Desoyer, "Aspects of sound signal processing using the Vienna intra- and extracochlear implants," in Cochlear Implants, R.A. Schindler and M.M. Merzenich, Eds. New York: Raven Press, 1985, pp. 101-110.

[44] M. Lawrence, "Factors in the continuing development of auditory prostheses," in Controversies in Otolaryngology, J.B. Snow, Ed. Philadelphia: W.B. Saunders, 1980, pp. 433-442.

TABLE I

Parameters of Real-Time IP Processors

patient	channels[1]	pulse widths/phase (all +/- biphasic)[2]	pulse sep.[2]	cycle time[2]
MC1	3/3	0.5	0.5	4.5
HE	4/4	0.5	0.5	6.0
JM	3/3	1.0	0.1	6.3
RC	2/2	0.5	0.1	2.2
ET	4/4	1.0 1.0 0.5 0.5	0.1	6.4
MC2	4/4	0.3 0.7 0.3 0.3	0.5	5.2

[1] Channels are specified in the form N1/N2, where N1 is the number of channels updated in each stimulus sequence and N2 is the total number of available channels.

[2] Times are in milliseconds.

TABLE II

Characteristics of Processors*

ANALOG	PULSATILE
continuous waveforms, presented simultaneously	non-simultaneous pulses
severe interactions between channels for patients with poor nerve survival	improved channel isolation, especially for patients with poor nerve survival
in some patients, continuous waveforms can provide good temporal and frequency information (F0, voice/unvoice distinctions, F1, possibly F2)	limited transmission of temporal and frequency information (F0, voice/unvoice distinctions)

---

*Symbols used in this Table are F0 for the fundamental frequency of voiced-speech sounds, F1 for the first formant frequency of speech, and F2 for the second formant frequency of speech.

FIGURE CAPTIONS

Fig. 1. Functional elements of the auditory system. (Not to scale.)

Fig. 2. Patterns of nerve and hair cell survival for normal hearing (left panel) and for a typical deaf ear (right panel). Illustration adapted from Lawrence [44].

Fig. 3. Source-filter model of speech production. A_v and A_u denote gain factors to control the levels of source outputs. Illustration redrawn from O'Shaughnessy [14].

Fig. 4. Waveforms of a compressed analog (CA) processor.

Fig. 5. Waveforms of an interleaved pulses (IP) processor.

Fig. 6. Results of vowel and consonant confusion tests for patient MH. Diagonally-hatched bars indicate results obtained with lipreading and cross-hatched bars indicate results obtained without lipreading. The table at the bottom of the figure indicates the type of processor used (abbreviations are CA for "compressed analog" and IP for "interleaved pulses"); the number of stimulation channels; whether voicing information was explicitly coded for the IP processors; and the overall percent-correct scores from the four test conditions for each processor. The horizontal line in each panel shows the level of chance performance for that test.

Fig. 7. Results of vowel and consonant confusion tests for six patients fitted with the 4-channel UCSF/Storz transcutaneous transmission system.

Fig. 8. Results from the IOWA test of consonant identifications with lipreading cues. The patients are those described in the caption of Fig. 7.

Fig. 9. Results from subtests of the Minimal Auditory Capabilities (MAC) battery designed to evaluate perception of prosodic elements in speech. Abbreviations for the subtests are Q/S for Question/ Statement; N/V for Noise/Voice; and Sp S/D for Spondee Same/Different. The patients are those described in the caption of Fig. 7.

Fig. 10. Results from subtests of the Minimal Auditory Capabilities (MAC) battery designed to evaluate perception of phonemes and words in speech. Abbreviations for the subtests are Init. Cons. for Initial Consonant; Final Cons. for Final Consonant; and Sp 4-Ch for Four-Choice Spondee. The patients are those described in the caption of Fig. 7.

Fig. 11. Results from subtests of the Minimal Auditory Capabilities (MAC) battery designed to measure open-set recognition of speech. Abbreviations for the subtests are Sp for Spondee recognition; NU6 for recognition of monosyllabic words from Northwestern University list six; CID for recognition of everyday sentences from lists prepared at the Central Institute for the Deaf; and WIC for recognition of Words in Context. The patients are those described in the caption of Fig. 7.

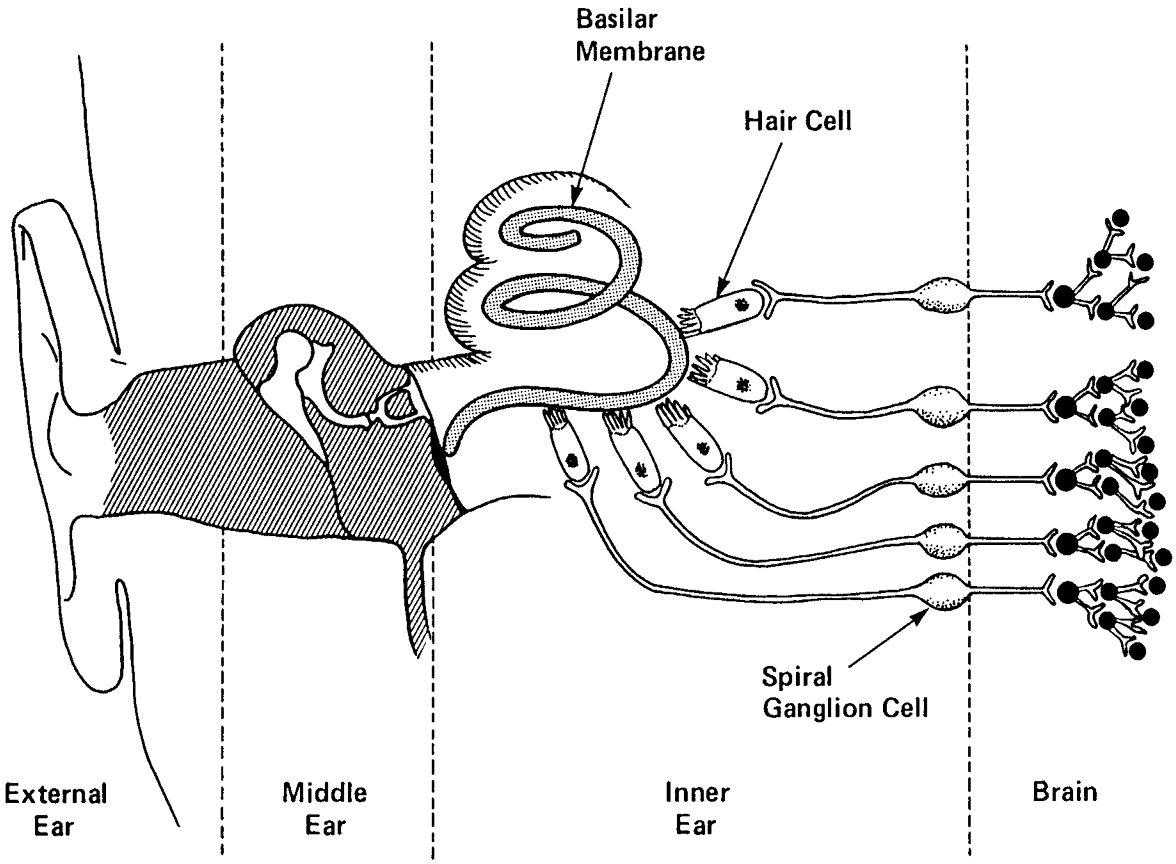

Basilar Membrane
Hair Cell
Spiral Ganglion Cell
External Ear
Middle Ear
Inner Ear
Brain

Fig. 1

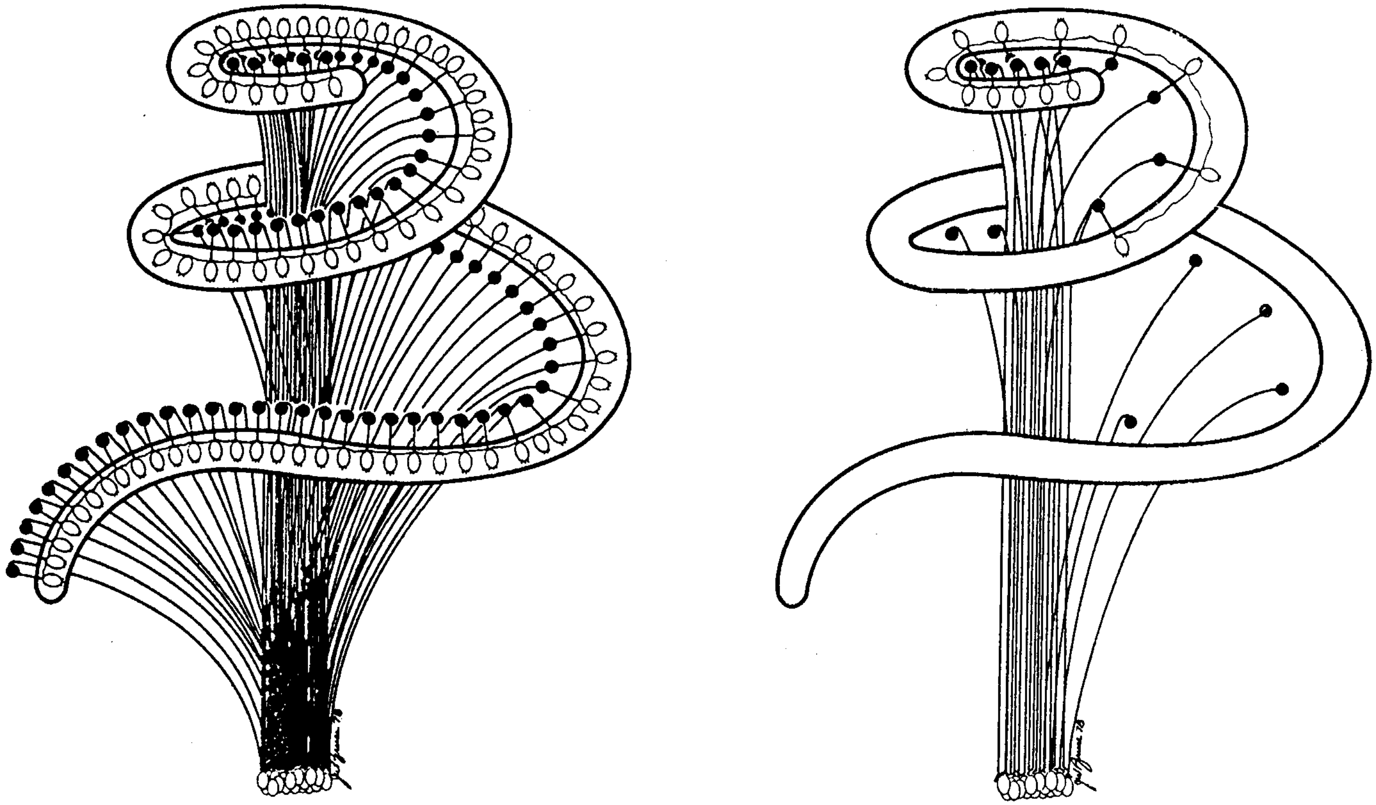


Fig. 2

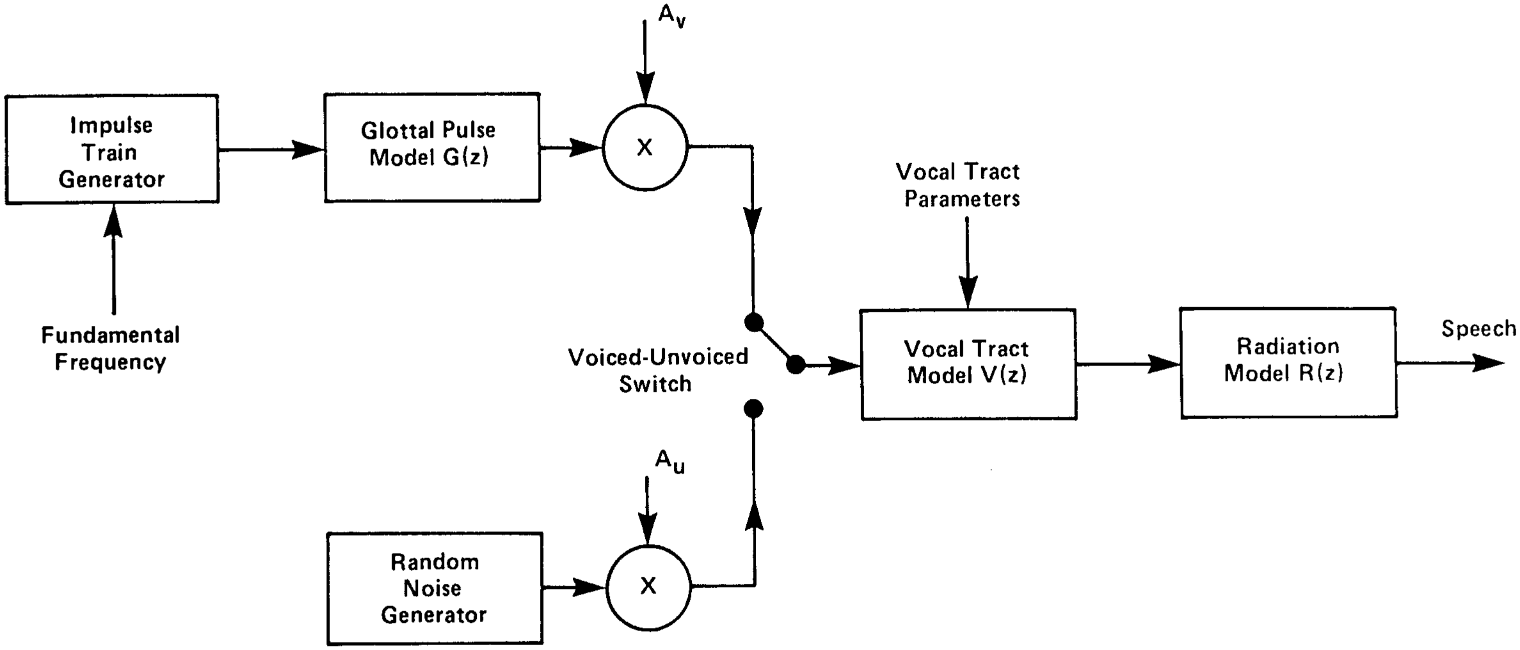
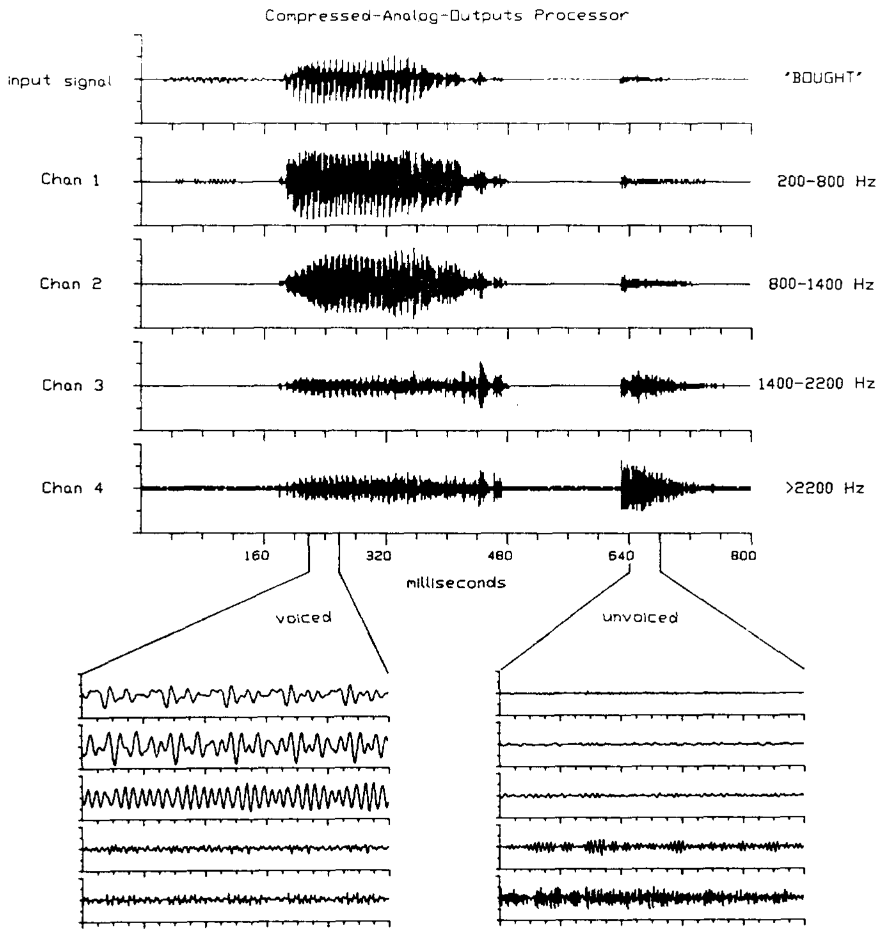

A_v
Impulse Train Generator
Glottal Pulse Model G(z)
X
Vocal Tract Parameters
Fundamental Frequency
Voiced-Unvoiced Switch
Vocal Tract Model V(z)
Radiation Model R(z)
Speech
A_u
Random Noise Generator
X

Fig. 3



Compressed-Analog-Outputs Processor

Fig. 4

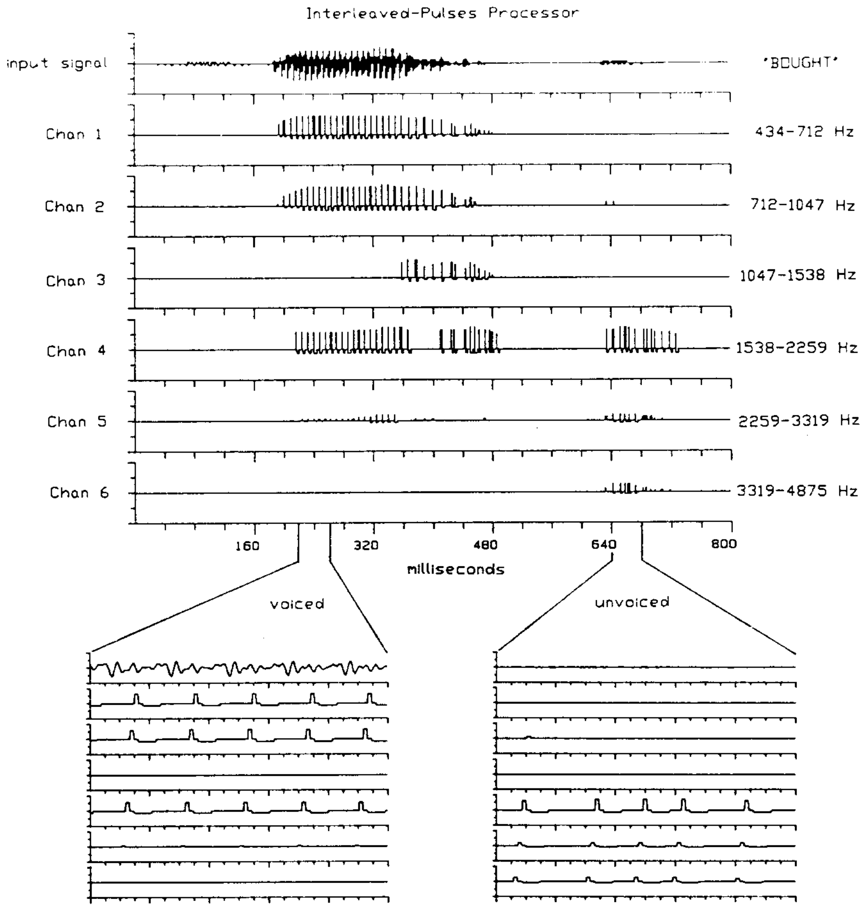

Interleaved-Pulses Processor
input signal
'BOUGHT'
Chan 1
434-712 Hz
Chan 2
712-1047 Hz
Chan 3
1047-1538 Hz
Chan 4
1538-2259 Hz
Chan 5
2259-3319 Hz
Chan 6
3319-4875 Hz
160
320
480
640
800
milliseconds
voiced
unvoiced

Fig. 5

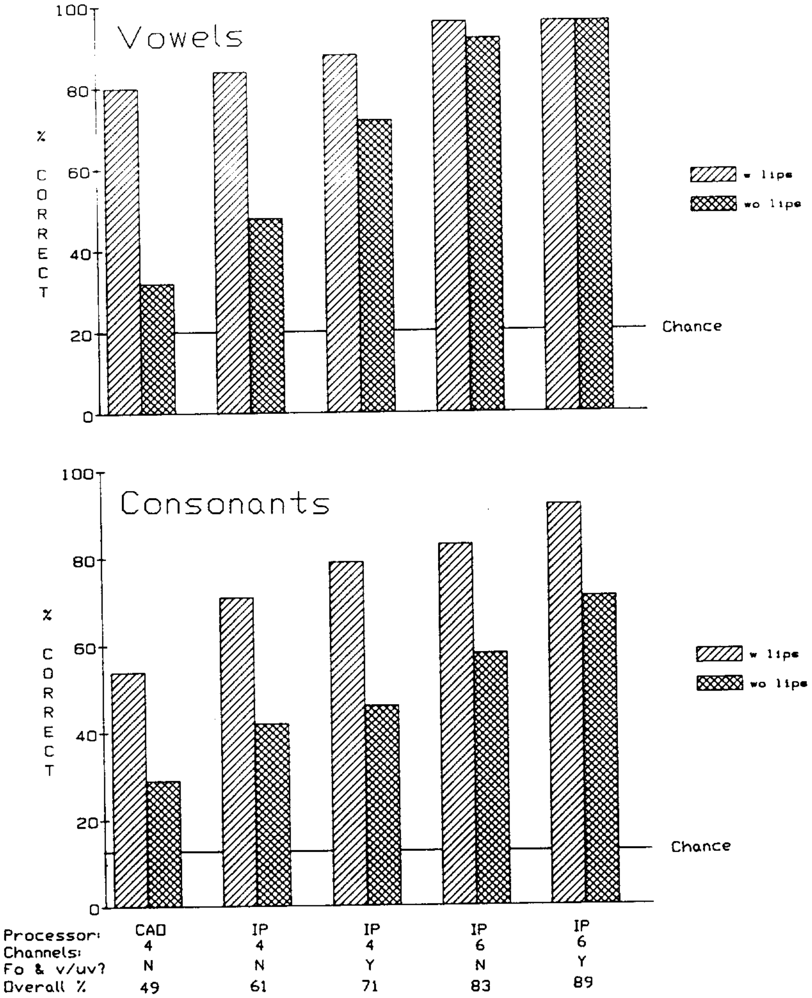

Vowels
Consonants
% CORRECT
100
80
60
40
20
0
▨ w lips
▩ wo lips
Chance
Processor: CAD IP IP IP IP
Channels: 4 4 4 6 6
Fo & v/uv? N N Y N Y
Overall % 49 61 71 83 89

Fig. 6

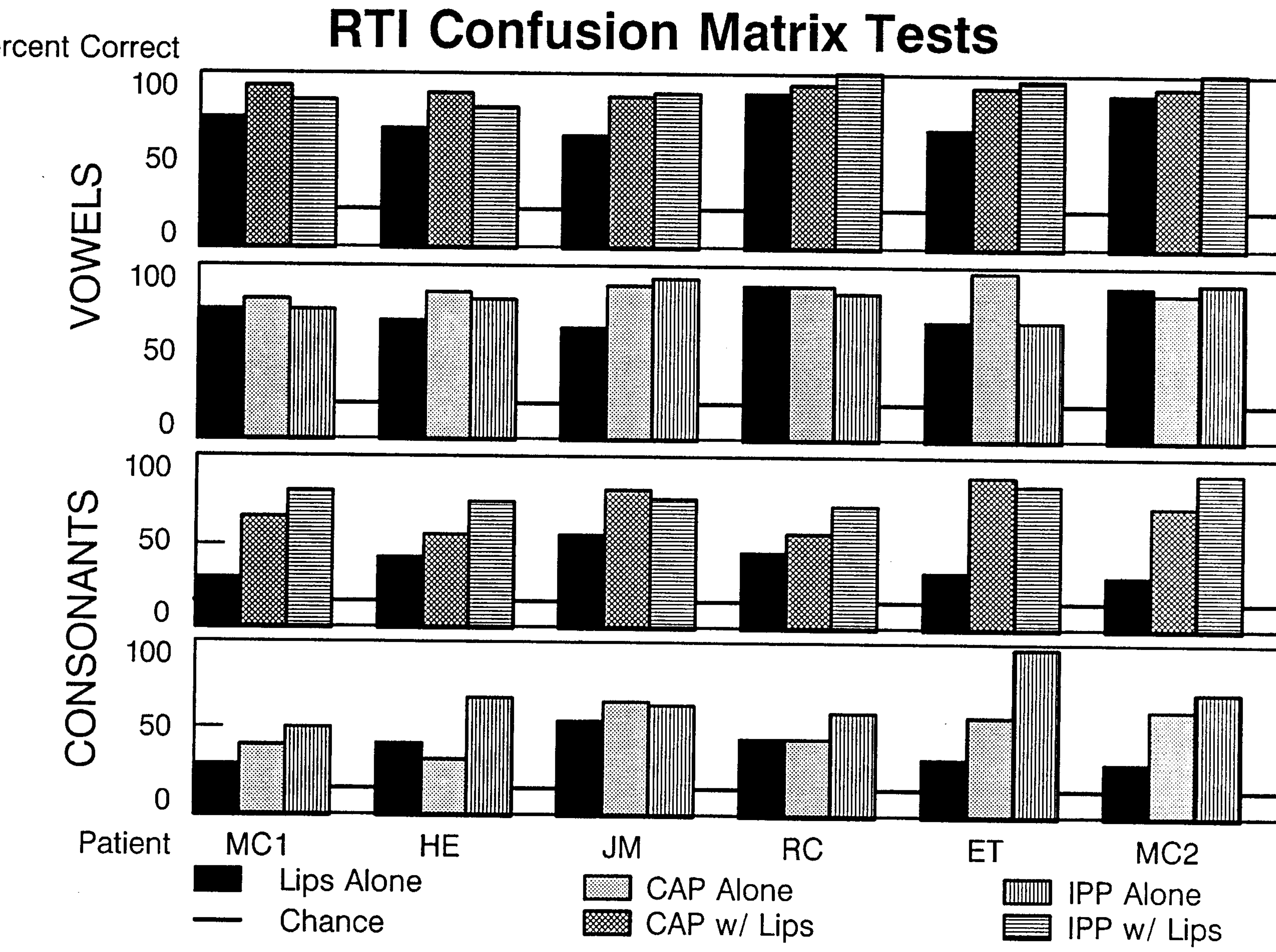

RTI Confusion Matrix Tests
rcent Correct
100
50
0
VOWELS
100
50
0
100
50
0
CONSONANTS
100
50
0
Patient
MC1
HE
JM
RC
ET
MC2
Lips Alone
Chance
CAP Alone
CAP w/ Lips
IPP Alone
IPP w/ Lips

Fig. 7

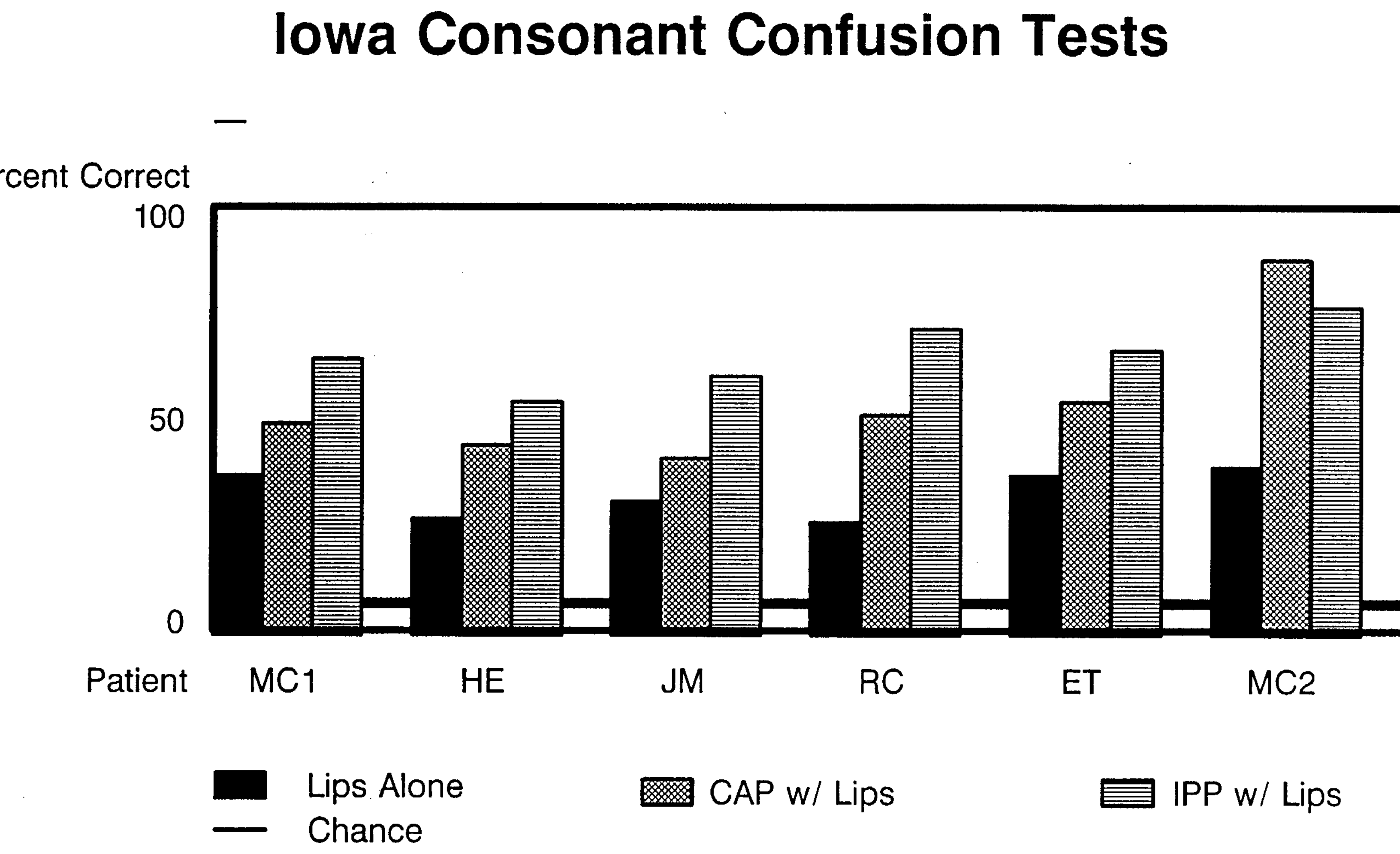

Iowa Consonant Confusion Tests
rcent Correct
100
50
0
Patient
MC1
HE
JM
RC
ET
MC2
Lips Alone
Chance
CAP w/ Lips
IPP w/ Lips

Fig. 8

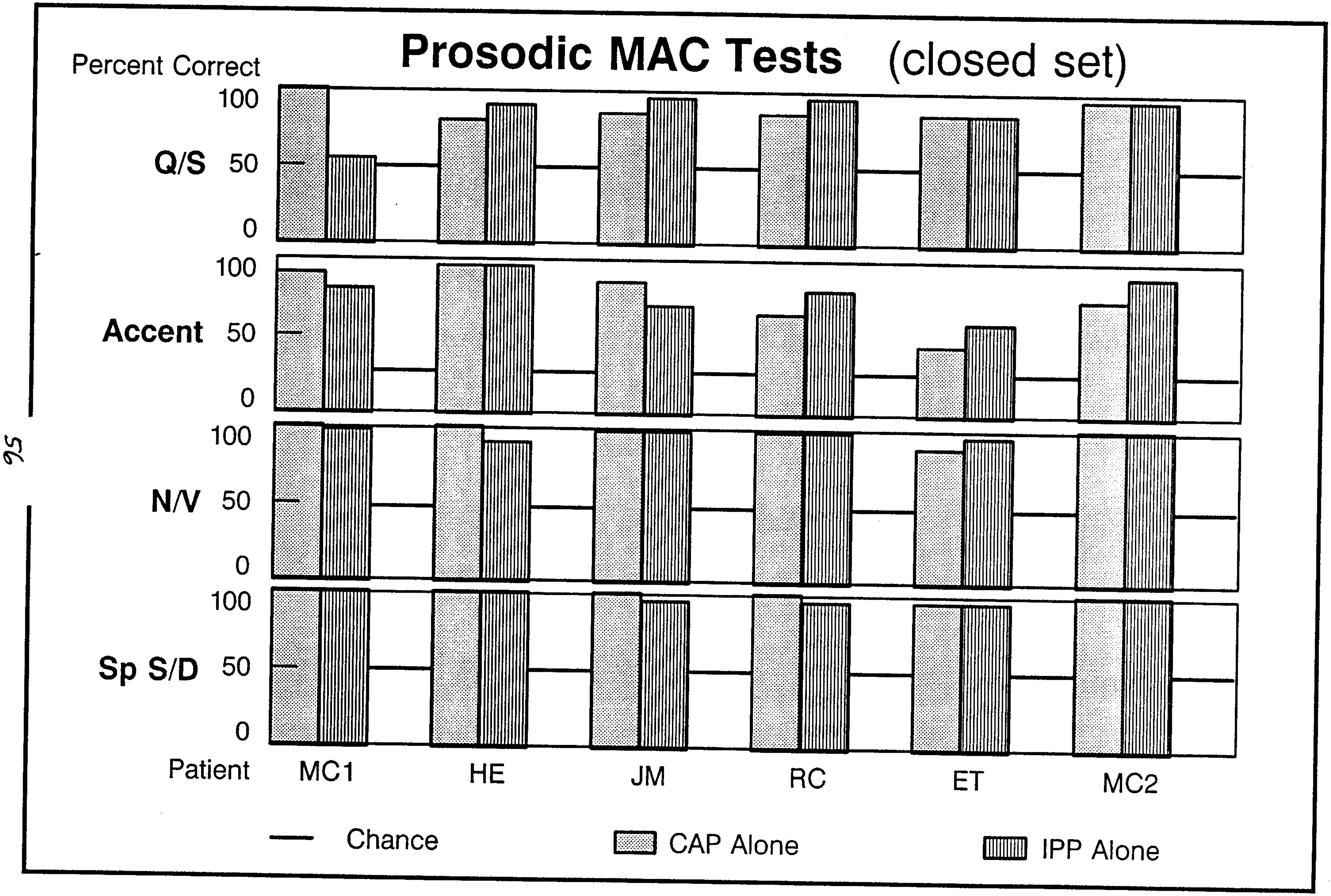

Prosodic MAC Tests (closed set)
Percent Correct
Q/S
Accent
N/V
Sp S/D
100
50
0
Patient
MC1
HE
JM
RC
ET
MC2
Chance
CAP Alone
IPP Alone

Fig. 9

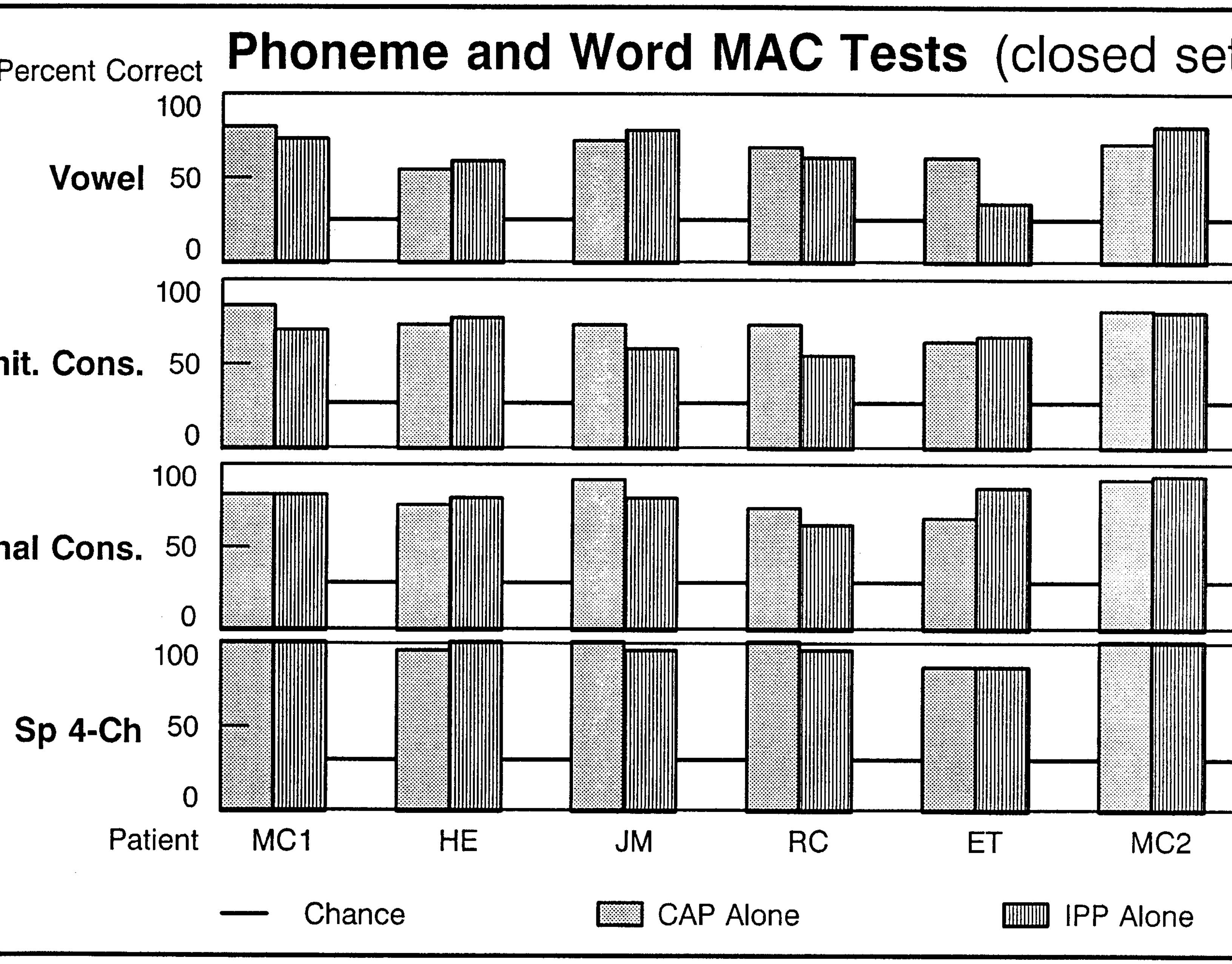

Phoneme and Word MAC Tests (closed se
Percent Correct
Vowel
nit. Cons.
nal Cons.
Sp 4-Ch
100
50
0
Patient
MC1
HE
JM
RC
ET
MC2
Chance
CAP Alone
IPP Alone

Fig. 10

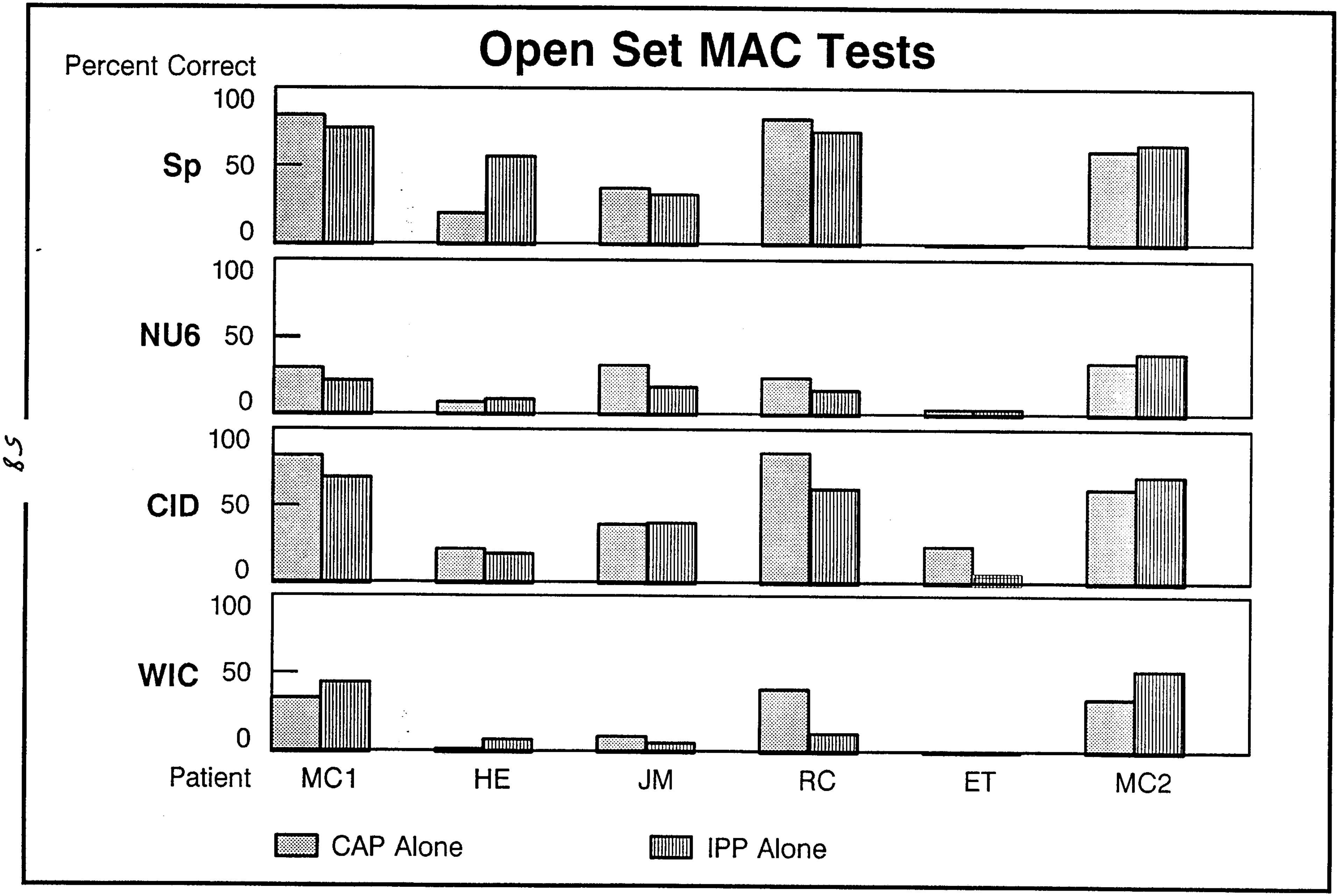

Open Set MAC Tests
Percent Correct
Sp
NU6
CID
WIC
100
50
0
Patient
MC1
HE
JM
RC
ET
MC2
CAP Alone
IPP Alone

Fig. 11