INFORMATION TO USERS

This was produced from a copy of a document sent to us for microfilming. While the most advanced technological means to photograph and reproduce this document have been used, the quality is heavily dependent upon the quality of the material submitted.

The following explanation of techniques is provided to help you understand markings or notations which may appear on this reproduction.

1. The sign or "target" for pages apparently lacking from the document photographed is "Missing Page(s)". If it was possible to obtain the missing page(s) or section, they are spliced into the film along with adjacent pages. This may have necessitated cutting through an image and duplicating adjacent pages to assure you of complete continuity.

2. When an image on the film is obliterated with a round black mark it is an indication that the film inspector noticed either blurred copy because of movement during exposure, or duplicate copy. Unless we meant to delete copyrighted materials that should not have been filmed, you will find a good image of the page in the adjacent frame.

3. When a map, drawing or chart, etc., is part of the material being photographed the photographer has followed a definite method in "sectioning" the material. It is customary to begin filming at the upper left hand corner of a large sheet and to continue from left to right in equal sections with small overlaps. If necessary, sectioning is continued again—beginning below the first row and continuing on until complete.

4. For any illustrations that cannot be reproduced satisfactorily by xerography, photographic prints can be purchased at additional cost and tipped into your xerographic copy. Requests can be made to our Dissertations Customer Services Department.

5. Some pages in any document may have indistinct print. In all cases we have filmed the best available copy.

University
Microfilms
International

300 N. ZEEB ROAD, ANN ARBOR, MI 48106
18 BEDFORD ROW, LONDON WC1R 4EJ, ENGLAND

Design Considerations of a Prosthesis for the Profoundly Deaf

By

Mark W. White
B.S. (University of Nebraska at Lincoln) 1971

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Engineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, BERKELEY

Approved:

Edward L. Keller, Ph.D.

Edwin R. Lewis

Michael M. Merzenich

Committee in Charge

DEGREE CONFERRED DECEMBER 9, 1978

Acknowledgements

I would like to thank the many individuals who have contributed, in so many ways, to the observations described in this manuscript. The high quality of research and cooperation among researchers in all areas of the auditory sciences has contributed immeasureably to the cochlear prosthesis project. One of the most pleasureable aspects of this project has been the study of the literature from several interrelated fields in neurophysiology and the auditory sciences.

In our laboratory, many individuals have contributed their ideas and labor to the cochlear prosthesis project. Mike Merzenich has contributed his enthusiasm, his optimism, and many very fruitful ideas. In particular, Mr. Merzenich's inferior colliculus excitation mapping technique has been very valuable; and his enthusiasm for the development of evoked response mapping techniques has been all-important in the development of these methods. Mike Merzenich also contributed his time and skill during the inferior colliculus mapping studies described in this manuscript. Pat Jones and Birgitta Bjorkroth, in cooperation with Mr. Merzenich, have been responsible for all anatomical preparations and the analysis of those preparations. Robin Michelson has contributed many ideas to this project. Since the inception of this project, Mr. Michelson has been a very motivating individual. Perhaps, the most important idea contributed in this project was the concept of the project

itself (i.e., the concept of a multielectrode cochlear prosthesis). Mr. Michelson has been instrumental in the development of the implantable multichannel array and the surgical techniques required for the insertion of such an array. Tom Poulter, Melvin Bartz, and Kent Taylor, in cooperation with Mr. Michelson and Mr. Merzenich, have been responsible for the fabrication of the implantable arrays. Earl Schubert and Shelia Walsh have contributed well designed testing techniques, particularly appropriate for the "non-mimicking" approach to stimulus processing. Also, Mr. Schubert and Ms. Walsh have expanded the rationale for the "non-mimicking" approach to stimulus processing. Rod Yip and Paul Knight have made a considerable contribution to the usefulness of the stimulus generation and data access programs. Paula Osofsky and Maxine Calnon have been very helpful in typing, on-line proofing, and assembling this manuscript. Marshall Fong has contributed his design and fabrication skills in the stimulus delivery system discussed in chapter 5.

TABLE OF CONTENTS

Chapter 1

Introduction

A number of the design considerations relevant to the design of a multielectrode cochlear prosthesis are discussed in this manuscript. Speech and stimulus processing options and methods of implementing these options for patient testing are the primary topics of this manuscript.

The primary goal of a multichannel cochlear prosthesis would be the restoration of speech intelligibility. Single electrode systems have been produced and tested (Bilger, 1977). These single channel systems have the advantage of simplicity and appear to aid the patient in environmental awareness, speech production, and lip reading. However, speech intelligibility scores indicate that the single electrode channel falls far short of conveying enough information for the understanding of speech.

Simmons (1966) and Bilger (1977) have reviewed the history of electrical stimulation within the auditory system. The more modern efforts have been discussed by Merzenich (1975, 1973), Simmons (1966), Hanson and Lauridsen (1975), Eddington, Mladejovsky, and Dobelle (1977).

Chapter 2 of this manuscript describes the methods used during the neurophysiological phase of the work described in this manuscript. The first part of chapter 3 is a presentation of the results of the neurophysiological experiments pursued at the

Coleman Memorial Laboratory. The second part of chapter 3 relates the physiological data to the problem of stimulus processing. Chapter 4 is a presentation of a modular software method for efficiently implementing the various stimulus processing options to be tested. The use of this software is illustrated in Chapter 4 - with an example of an analysis-synthesis spectrum channel vocoder. Chapter 5 is a description of one method to generate and control the stimuli that are delivered to the patient's electrodes.

Chapter 2

Physiological Methods and Procedures

Our research group has been testing electrodes that are placed within the scala tympani of the cochlea (Michelson, 1971). Figure 2.1 illustrates the location of the electrode implant within the animal's cochlea.

Electrode fabrication

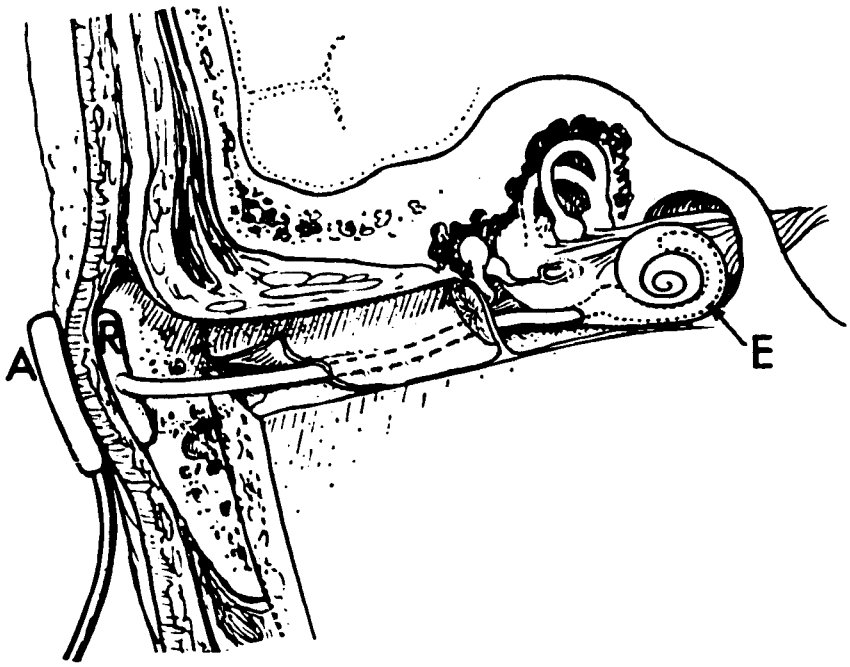
Figure 2.2 illustrates a typical scala tympani electrode implant. A cast of the scala tympani is first produced, using a low melting point Wood's metal. The molten metal is perfused into the cochlea via the round window, with the fresh cochlea immersed in warm water. The relatively undistorted cast of the scala tympani is then removed by dissection. (In a variant of this technique, an acrylic cast is made, but shrinkage is more significant in acrylic scala tympani casts.) The cast is then plated with a thin layer of copper; and a pivot drill is used to introduce positioning holes at precisely defined locations, re the surface detail of the scala tympani. The copper layer is not so thick as to obscure this fine surface detail. The Wood's metal is then removed in warm water, and the thin copper die scrubbed. It is then plated with an ultrathin layer of chrome. This chrome layer does not significantly reduce the internal dimensions of the die; it should eliminate the hazard of Woods' metal or copper contamination of the finished multielectrode array. The chrome plating is very carefully

Figure 2.1. A) Diagram illustrating the location of the scala tympani implant within the human cochlea.

The electrode implant (E) extends about 22-25 mm into the scala tympani in the basal coil; it is inserted via the round window. This diagram also illustrates the location (Mastoid cortex) of the receiver (R) and transmitter antenna (A) which may eventually be used by implanted patients (Gwellala, 1976). In animals, percutaneous electrode connectors (i.e., connectors that penetrate the skin) are used (Mooney, et al, 1977). Percutaneous connectors will be utilized during initial phases of patient testing.

B) Diagrammatic representation of the implant residing in cochlea (shown in cross section).

A



B

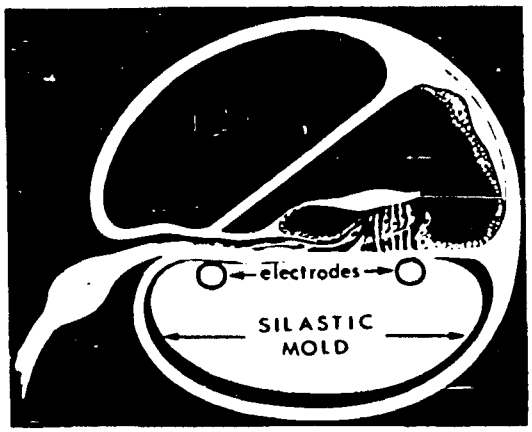


Figure 2.2. Photograph of a Michelson-type electrode (Michelson, 1971), constructed for implantation in the scala tympani of the cat.

The electrodes are embedded in a silastic carrier, molded to fit the most basal 11 mm of the scala. Leads are insulated up to the tip. The exposed flat circular tips are approximately 100μ in diameter; they face upward toward the basilar partition and bony spiral lamina. The Pt-Rh electrode contacts are accentuated by the open circles which are about 2.5 X the electrode's diameter. Other types of electrode tips have been studied (e.g. spheres of 300μ diameter, roughened wire tips, 30μ diameter wire tips). Animal implants have 2-8 exposed electrode tips arrayed across both the transverse and longitudinal dimensions of the scala.


b
a
c

checked, visually, using a dissecting microscope. The arrays are then constructed, using 3 mil or 1 mil Pt or Pt-Rh (90%-10%) wire. In one fabrication procedure, wires are fed into holes positioned along the bottom of the die, and then trimmed after removal of the completed array from the die. In a second variation, used just as successfully, the contact faces of the wires are dressed prior to introduction into the die. They are individually positioned on the bottom of the die using a special jig, and then tacked down with silastic adhesive. In this instance, there is more flexibility in the positioning of contact surfaces, and in the potential for modification of the electrode's surface area. For example, spheres have been formed at the tips of the wire. In both variations, a loop is made in the wire just at the surface of the die, to stabilize it in position. (Without this provision, the wire tip is pulled down into the silastic or it is freed from the silastic when the array is straightened to be inserted.) The die is then filled with medical grade silastic, to produce the finished electrode, a wire array mounted in a silastic carrier molded to fill the scala tympani. In the larger wire arrays produced as models of potential human electrodes, all of the wires can be fed into a central channel in this silastic carrier. This provides for further relief from distortion of the array attendant to straightening the array at the time of insertion.

Implantation procedure (surgical)

Adult normal cats were implanted with intracochlear stimulating

electrodes. Under sodium pentobarbital anesthesia and using sterile surgical techniques, arrays are implanted within the scala tympani. Almost all electrode implantation surgery has been conducted by trained otologic surgeons.

The surgical approach requires a post-auricular incision with elevation of a flap and the skin of the posterior canal wall. The drum is elevated and the round window is exposed. The electrode array is inserted through the round window into the scala tympani (Michelson, 1971). Electrode leads are brought out into a groove in the root of the zygoma and implantable connectors are attached to the cranium with stainless steel screws and acrylic. The implant from the root of the zygoma to the connector is held in place with medical grade methyl methacrylate.

In all animals, electrophysiological experiments began no earlier than two weeks after implantation, and in the great majority of cases, more than six weeks after implantation.

Electrophysiological methods.

Methods for determining certain features of electrically-generated neural discharge patterns have been developed. These methods are useful in determining where, along the longitudinal dimension of the cochlea, fibers are excited for given stimulus current amplitudes. These methods are useful in determining the longitudinal location of functionally excitable fibers. Threshold stimulus current amplitude as a function of longitudinal cochlear place is determined in one method. In this monograph, this function is commonly referred to as a "spatial excitation pattern" or a "spatial excitation map".

A) Excitation characterization using inferior colliculus recording

The method of mapping excitation patterns for implanted electrodes using physiological unit recording experiments in animals in the binaural, tonotopically-organized central nucleus of the inferior colliculus has been previously reported (Merzenich, et al, 1973; Schindler, et al, 1977). The spatial pattern of response for any electrode stimulating the auditory nerve of the cat can be determined. Advantage is taken of two basic organizational features of that auditory nucleus. First, there is a simple, highly ordered topographic representation of the auditory nerve array within this sturcture

(Merzenich and Reid, 1974). Second, neurons within a restricted portion of this nucleus receive excitatory binaural input (i.e., driving input from both cochleas). A neuron's input is from the same highly restricted area in both cochleas (Merzenich and Reid, 1974). After determining the "characteristic frequency" for a given colliculus neuron to sound stimulation of the unimplanted cochlea, one can determine the neuron's threshold of response to electrical stimulation delivered to the implanted ear. The "characteristic frequency" is that frequency at which the neuron is excited at the lowest sound pressure level. This characteristic frequency for the neuron is considered to correspond to a point along the basilar membrane from which the neuron receives its input (see Greenwood, 1961). It is possible to determine the electrical response threshold for any given electrode, as a function of location across the cochlear nerve array, by determining ipsilateral acoustic characteristic frequency and then determining contralateral threshold for electrical stimulation for a large series of neurons spanning nearly the entire length of the cochlea. Because of limitations in speaker output at highest frequencies, excitation maps were usually limited to the most apical 18-19 mm of the cat's cochlear sensory epithelium (22 mm long).

In these acute physiological recording experiments, implanted cats were anesthetized with sodium pentobarbital. The inferior colliculus was directly exposed by ablation of overlying occipital cortex. The

dorsal surface of the inferior colliculus was then photographed, and electrode penetrations sited by reference to the photograph and to the surface vasculature as viewed in a dissecting microscope (25X). Recording microelectrodes were introduced into the nucleus in the sagittal plane, angled 20-30 degrees caudally. Etched platinum-irridium glass-coated platinum-plated recording microelectrodes were introduced with the use of a Kopf hydraulic microdrive (controlled by a stepping motor) into a sealed chamber mounted on the skull.

A neuron or cluster of neurons within the central nucleus was monitored. If the responses were stable, the characteristic frequency (CF) was measured for ipsilateral sound stimulation. After determining CF, ipsilateral acoustic stimulation was discontinued and biphasic electrical stimulation was delivered to an electrode in the implanted, contralateral ear. Stimulus level was varied until the response threshold was determined. Threshold was determined for all or most of the electrode combinations. The recording electrode was advanced 20-200 micorns and the process repeated until neurons representing almost the entire length of the cochlea had been recorded. From such information, plots of electrical threshold vs. cochlear "place" were obtained.

Acoustic stimuli were generated using a dynamic acoustic driver powered by an oscillator and its attenuator. The system was closed, i.e., sound delivery tubes were sealed into the driver cans, and into the outer ear canal. In many experiments, a hardware system was used

to control timing, pulse or click or tonal frequency and stimulus envelopes of both sound and electrical stimuli. Delivery of stimuli in this system was under control of a "stimulus programmer", consisting of timers, preset counter, and appropriate logic necessary for control of stimulus duration, repetition rate and number and for generation of synchronizing marks signaling stimulus onset and offset, and marking cycle phase.

Computer generated acoustic stimuli were used when automatic acoustic driver frequency and/or amplitude compensation were desired. The computer system had the further advantage that complex acoustic stimuli were easily generated. The output of a 12-bit digital to analog converter was passed through a 9-pole, one dB. Chebychev lowpass filter (this filter contributed little delay distortion over the stimulus frequency range), and was then attenuated by programmable attenuators connected to the acoustic drivers. The D-A converter received numbers from a computer generated stimulus table at the rate of 82,230 samples per second.

In nearly all animal studies, electrical stimuli were computer generated in a similar manner. Two electrode channels were under independent stimulus control at any given time. Stimuli were generated using battery powered photo-coupled isolators (very similar to the stimulators discussed in chapter 5) with current as the controlled output variable. Reed relays under program control were used to switch the two isolators to the desired electrode combinations in multielectrode arrays. In the present configuration, any electrode elements can

be stimulated, in any sequence. Each isolator was connected to a programmable attenuator which, in turn, was driven by a 12-bit digital to analog converter. Each converter received numbers from a computer generated stimulus table at the rate of 41,115 samples per second.

During and after electrical or acoustic stimulus presentation, the output of an action potential level discrimination was recorded in computer memory using conventional techniques. Post-stimulus times for each discharge were recorded on disk, along with associated parameters of stimulation and recording. This information was later transferred to digital cassettes. Programs have been developed for the efficient retrieval of these data. During the experimental session, the most recent post stimulus time histograms were displayed, along with a graph of past and present discharge rates. A variety of analysis programs have been written for later data processing and a digital plotter is incorporated into this facility for high resolution graphics to assist in documentation of experimental results.

B) Excitation characterization using brain stem evoked response recording (BSER)

Recording of brain stem potentials at the scalp has been employed to study the excitation characteristics of implanted electrodes. Studies of these brain stem auditory potentials provide a convenient noninvasive method for assessing electrode function (Figure 2.3). In future patient studies, such techniques will be essential for the evaluation of electrode and nerve function.

Animal studies have now been conducted with many implanted

electrodes. One observation from these results is that the growth of magnitude of these potentials is a monotonically increasing function of the number of fibers excited. Inferior colliculus unit mapping studies in the same animals were used to estimate the number of fibers excited (Figure 2.4). By knowing intracochlear electrode element locations and by deriving BSER magnitude functions, excitation patterns can be roughly estimated. In these studies, the growth of scalp potential with stimulating current is plotted and compared with the scalp potential generated by whole nerve stimulation (generated by moderate to high level monopolar[1] stimulation). An estimate can be made of the fraction of auditory nerve fibers being stimulated.

Accuracy of excitation estimates made from evoked response recordings may be improved using a number of signal processing techniques.[2]

Other evoked response techniques may be developed and utilized for animal and patient testing in the future. A second procedure employing

---

1. Monopolar stimulation is defined here as electrical stimulation with at least one electrode located outside the cochlea.

2. Better excitation estimates using evoked response techniques (e.g. BSER magnitude comparisons) may require: a) An improved signal to noise ratio. Matched filtering techniques should be useful in addition to signal averaging. The frequency response of such a matched filter should be so shaped as to give a larger weight to those frequencies which exhibit the better signal to noise ratios (Childers and Durling, 1975). b) A determination of the best analysis time window for evoked response magnitude estimation. Those parts of the response time series that are most reliable in estimating the number of fibers should be emphasized. Also, any portion of the response time series that is corrupted by stimulus artifact should be deemphasized. c) A stimulus sequencing technique that avoids small, slowly changing response drifts. Even though

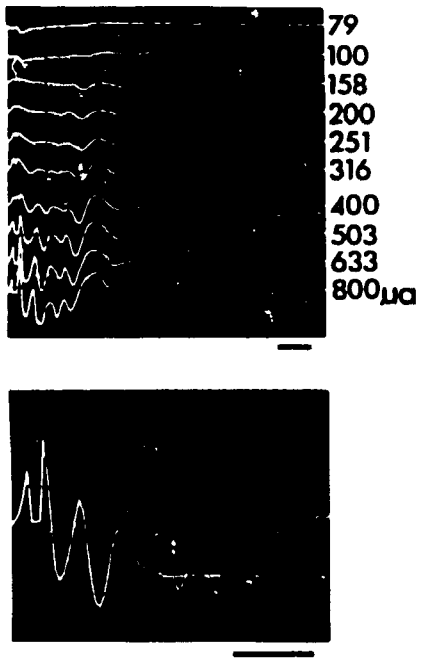
Footnotes continued:

most portions of the BSER are relatively stable, stimulus sequencing such as ABCDABCD.....(each letter represents a different stimulus) or pseudo-random sequencing may offer better results than sequencing AA...BB...CC...DD... d) A very effective artifact suppression technique: this will increase the usable portion of evoked response time series.

the brainstem evoked response should provide a more accurate estimate of the excitation pattern. In this experiment, two nearby channels are excited in rapid sequence. In this discussion "channel" is taken to mean "one transversely oriented bipolar electrode pair". Each channel's biphasic stimulus pulse does not temporally overlap the other channel's biphasic pulse. The entire stimulus sequence occurs within the nerve's absolute refractory period (less than 700 μsec., Geisler, 1968). The stimulus level of electrode 1 is systematically increased, while the second electrode, excited later, is maintained at a constant level just above the threshold. When excitation from the first stimulation spreads across the region at which the second (probe) electrode is situated, the probe electrode cannot contribute to the compound BSER, as the fibers normally excited by electrode 2 (the probe electrode) have been put into absolute refraction by the pulse at the first electrode. Thus, the first electrode's excitation has spread the interchannel distance over the nerve array. The experimental paridymn requires the comparison of two brainstem responses. They are the BSER evoked by stimulating only electrode 1 and the BSER evoked when both electrodes are stimulated in sequence. If the two responses are significantly different, then complete excitation overlap has not occurred. If the responses are identical(i.e., within statistical criteria), then complete excitation overlap has occurred. This technique may offer a very useful means of determining nerve and electrode performance in the implanted patient. For instance, patchy nerve survival may

Figure 2.3. Derivation of BSER magnitude vs. stimulus amplitude function.

Responses averaged for 500 trials are shown in each trace at the left. Stimulation current levels are shown at the right of each averaged response. Stimuli were 100 μsec. biphasic pulses. Bars at the bottom of this response series are 1 msec. in duration. Response magnitude (rms) is plotted as a function of stimulus current level at the right. Response magnitude was measured over a time window starting approximately 0.5 msec. after stimulus onset and ending 3.5 msec. after stimulus onset. Brainstem evoked response (BSER) recording has been employed extensively in deriving basic information about multielectrode performance, and in monitoring the physiological status of the auditory nerve array in long implanted cats.


79
100
158
200
251
316
400
503
633
800µa

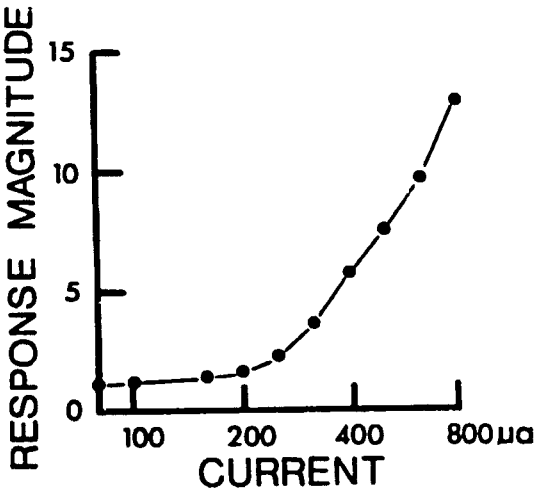
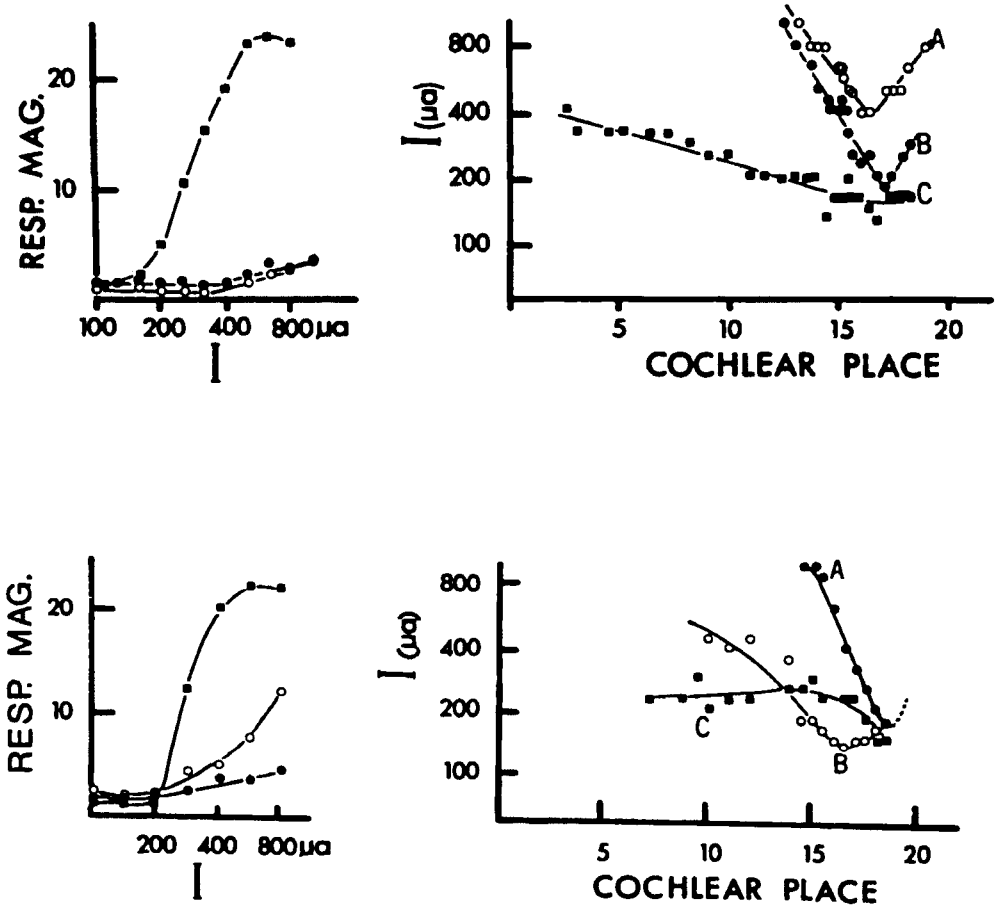

RESPONSE MAGNITUDE
15
10
5
0
100
200
400
800µa
CURRENT

Figure 2.4. On the left: Magnitudes of BSER potentials plotted as a function of stimulus level, from experiments conducted in two different implanted cats (upper and lower graphs). On the right: Excitation threshold current as a function of cochlear place, derived in an inferior colliculus electrode mapping study. In the experiment represented by the upper graphs, electrodes whose responses are represented by curves A and B were bipolar, with interelectrode separation of approximately 200 microns. Curve C was derived with monopolar stimulation, i.e., with stimulation between one of the electrodes in the "A" bipolar pair and a ground wire in the middle ear. In the lower graphs, curve A was derived with a normal bipolar electrode; curve B was derived with a bipolar electrode in a region in which the basilar membrane was perforated; curve C was derived with stimulation using a monopolar electrode. These and acoustic stimulus studies (Davis, 1976; Jewett and Williston, 1971; Thorton, 1975) reveal that the growth a magnitude of brainstem auditory potentials roughly reflects the increase in number of fibers excited in the auditory nerve array, as a function of stimulus level. The correspondence between the IC excitation map and the BSER intensity series data is best at response levels considerably higher than threshold. Stimulus artifact and noise obscure the lower level responses in the BSER recording. At low levels, the response magnitude primarily reflects the level of the stimulus artifact.


RESP. MAG.
20
10
100 200 400 800μa
I
I (μa)
800
400
200
100
A
B
C
5 10 15 20
COCHLEAR PLACE
RESP. MAG.
20
10
200 400 800μa
I
I (μa)
800
400
200
100
A
C
B
5 10 15 20
COCHLEAR PLACE

be such that two electrode pairs can only stimulate the same nerves and are not able to stimulate differential sectors of the nerve. Patchy nerve survival should be detectable by using this technique.

The second procedure allows one to determine the stimulus level at which the first electrode's excitation has spread the inter-channel distance over the nerve array. With knowledge of the distance between these "channels", with knowledge of the threshold current, and with the knowledge - from single unit colliculus studies - that straight line "threshold current (in dB) vs. cochlear place" relationships are a good approximation for bipolar electrodes (Figure 3.1, 3.2), a reasonably accurate "threshold current vs. cochlear place" profile may be estimated.

Possibly, middle and late evoked responses may be used to advantage in this technique of proper stimulus sequencing is used (see "Implementation of BSER experiments".).

Using the same stimulus paridymn, a psychophysical test procedure might be very useful. The patient could be queried as to whether the two stimulus presentations sounded the "same" or "different". This would be analogous to determining whether the two evoked responses were the "same" or "different".

Preliminary experiments indicate this BSER non-invasive excitation mapping method (defined here as the "interference method") will provide excellent, high resolution excitation maps. Appropriately combined with the "BSER response magnitude vs. stimulating current" function, this technique offers a very versatile method for mapping excitation patterns of numerous electrode array configurations.

Implementation of BSER experiments

Brainstem evoked response(BSER) studies require that one or two electrode channels be independently stimulated. During and after stimulus presentation, brainstem auditory potentials must be recorded in computer memory for subsequent signal averaging.

For animal experiments, stimuli were generated as described in the colliculus studies, using battery-powered photo-coupled isolators with current as the controlled output variable, reed relays under program control to switch between electrodes, and computer generated waveforms. There were very few restrictions on the type of stimuli that could be generated by this method. Sets of stimulus parameters and stimulus tables were stored on disk. Using these stored parameters, the system sequentially generated the many types of stimuli desired, and subsequently averaged and recorded the evoked responses.

After differential amplification and initial bandpass filtering (10 Hz. to 8 KHz.), responses were sampled by an analog to digital converter at 20,556 samples per second. After each trial, these samples were summed with the total average. During the experiments, the waveforms obtained from signal averaging were displayed on a monitor oscilloscope and were recorded on digital cassettes, along with associated stimulus parameters. Summing, differencing, auto- and cross-correlation programs were used to compare waveforms. For example, the auto-correlation program was used as a means of measuring the RMS magnitude of the response for specified time windows. Discrete linear phase filtering (i.e., filters with a constant delay for all frequencies)

techniques were also valuable in further increasing the signal to noise ratio without adding delay distortion.

Response comparisons (e.g. R_A-R_B) were often utilized. The stimulus trials were sequenced ABABAB..... rather than AA...BB... This technique avoids small response drifts in the relatively stable brainstem evoked response.

Electrode characterization using voltage vs. current measurements

The measurement of electrode voltage as a function of stimulating current is pertinent to electronic stimulator design. Also, the measurement may give some insight into the electrochemical effects occurring at the electrode-tissue interface. Two methods have been used for the electrode measurements.

One method, which was highly automated, recorded electrode voltage while the electrode current was controlled by the computer at a sampling rate of 82,230 samples per second. Almost any sequence of controlled current or voltage could be generated. Generally, sinusoidal and pulse waveforms of selected amplitudes and frequencies were presented sequentially while the voltage was being recorded digitally. Computer controlled reed relays allowed for automatic switching amoung electrode combinations. The waveforms were then recorded on a digital plotter and a digital cassette.

The other method consisted of a battery powered oscilloscope, sinusoidal oscillator, resistors, and a switch. This equipment was portable and was used for quick electrode checking during implantation surgery.

Cochlear histology

At the termination of each physiological experiment, the cochleas were exposed and directly perfused with buffered glutaraldehyde (or osmium). If it was crucial to identify precisely the location of the electrode, fixative was delivered through the oval window and a wide hole in the cochlear apex. The electrode leads were cut at the margin of the round window, and the cochlea removed as rapidly as possible and placed into fixative. If this information was not critical, and/or if optimal fixation of histological material was required, the electrode array was removed, and the cochlea then perfused through the round and oval windows. The cochleas were first examined in surface preparation; then semi-thin sections were examined with the light microscope. Also, critical areas were identified in surface preparations and semi-thin sections were examined using transmission electron microscopy.

for number sequence only

Chapter 3

Stimulus Processing Considerations

What stimulus processing procedures may be useful in a prosthesis for the profoundly deaf? There are a number of approaches to this problem. We can make reasonable guesses as to the most useful speech processing options on the basis of speech perception studies, physiology of the auditory nerve, directed electrophysiological studies, relevant psychoacoustic studies, and knowledge of the nature of the sensation generated with single channel prosthetic devices already implanted in patients. Two different approaches, that may be valuable in delivering speech information to the prosthesis recipient, have been pursued. In one approach, information is processed with the object of approximately mimicking the normal eighth-nerve input to the auditory nervous system generated acoustically by normal speech, or by speech resynthesized using channel vocoder. In the second approach, perceptually important information would be delivered with the stimulus processing employing the best information-bearing modes of stimulation. The best information-bearing modes of stimulation may in part be determined from basic psychoacoustic tests (e.g. pattern discrimination tests of rapidly changing intensity and time transistions). Recognizing the merits of both philosophies, the investigator may find that appropriate combinations of these two approaches may prove most fruitful. Table 3.1 offers a general outline of these stimulus processing approaches.

I. The Mimicking Approach

In the mimicking approach, the electrodes' defined excitatory

capabilities would be employed to effect the perceptually important aspects (especially relative to the perception of speech) of the neural firing patterns elicited during acoustic-to-neural transduction. In general, it is assumed that it is auditory nerve activity that is to be mimicked. There are at least two reasons for this assumption. Auditory nerve coding is better understood than is neural coding in other auditory centers; and all acoustic information that is transduced is conveyed by the acoustic nerve (i.e., no parallel afferent neural paths exist at this level).

In designing a processor, it is valuable to understand: A) What neural excitation patterns are generated by electrical-to-neural transduction; B) What features of the acoustic speech signal are important in understanding speech; C) What neural excitation patterns (both spatial and temporal) are generated by acoustic-to-neural transduction; D) What features of the excitation pattern generate the various percepts (both basic percepts, such as loudness, and speech-related percepts, such as intonation).

None of these areas can be considered to be well understood. However, some very useful information has been compiled by many investigators. Some of the more relevant aspects of this information will be summarized in the remainder of this chapter.

With this information, one could attempt to find spatial-temporal, electrically-generated excitation patterns that closely match patterns elicited by perceptually significant acoustic stimuli. A procedure that would closely match estimated electric excitation patterns with estimated

Table 3.1. Outline of Some Possible Approaches to Stimulus Processing for a Cochlear Prosthesis

I. The Mimicking Approach

The "mimicking approach" is a general heading for those processing techniques which are designed to reproduce certain feature(s) of normal hearing function. These various processing techniques include:

A. Approximate replication of spatial-temporal eighth-nerve excitation patterns. This might include:

1. Replication of those neural patterns generated by normal speech.

2. Replication of those neural patterns generated by processed speech. For example, one might attempt to replicate excitation patterns elicited by acoustic stimuli which are generated by an analysis-synthesis vocoder (e.g. a spectrum channel vocoder or a formant vocoder).

B. Approximate replication of the perceptually significant features of normal excitation patterns. The "perceptually significant features" may be estimated by applying:

1. Eighth-nerve coding theories. These theories are estimates of how stimulus parameters are represented in the eighth-nerve discharge pattern.

2. The literature relating to distinctive features of the acoustic speech signal, and information relating to the acoustic-to-neural transduction process.

3. Psychophysical testing techniques, applied to implanted subjects, to determine the perceptual significance of electrically generated excitation patterns.

C. Approximate replication of the percepts of a normal hearing individual. This version of the mimicking approach would involve the matching of those percepts generated by acoustic stimulation in the normal ear with those percepts generated by electrical stimulation. This would involve the development of a psychophysics of eighth-nerve electrical stimulation. To some degree, this psychophysical data could be supplemented by the extensive body of information developed in the psychophysics of acoustic stimulation.

II. The Non-Mimicking Approach

It may be more reasonable to derive perceptually important information from the acoustic speech signal and deliver it employing the best information-bearing modes of stimulation without attempting to mimick normal peripheral neural responses.

acoustic excitation patterns would be very useful. This is the essence of a strictly interpreted mimicking philosophy. However, exact pattern matches will almost certainly not be possible, even if the transduction mechanisms were better understood. Exact pattern matches would probably require one electrode pair for each acoustic nerve fiber. If the perceptually significant features of the acoustic-to-neural excitation patterns (i.e., the neural code) can be determined, it might be possible to develop better perceptual "matches" than would be possible with a strict excitation pattern matching paradigm. Probably, the most useful excitation pattern features are those features that best enable the subject to learn the useful discriminations.

Because exact excitation pattern matches are probably not possible, and because the state of the auditory nervous system may be considerably altered in the deaf person, a comprehensive patient testing and parameter adjustment procedure is likely to be of immeasurable value. Patient tests should indicate what aspects of the pattern matching procedure need to be refined. Test procedures may indicate that certain aspects of the mimicking technique should be discarded for other processing found empirically to be more useful. The procedure should indicate the effect of patient learning. A potential testing and "design by patient feedback" procedure is briefly summarized in the discussion relating to the second processing approach.

A. Neural excitation patterns generated by electrical stimulation

1. Spatial domain excitation characteristics

Techniques have been developed to determine where, along the longitudinal (i.e., axial) dimension of the cochlea, acoustic nerve

processes have been excited. These techniques are summarized in "Methods and Procedures". The spatial excitation patterns generated by selected electrode configurations have been summarized by Merzenich and White, 1977.

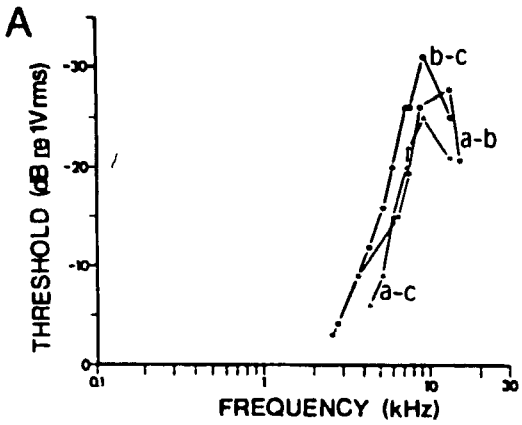
Spatial excitation patterns generated by radially oriented bipolar electrode pairs have been investigated. If both electrode points are within the cochlea the electrode is defined as "bipolar". Electrode point separations of 300 to 1500 microns have been investigated. Acoustic nerve fiber thresholds to electrical stimulation as a function of the fiber's origin along the longitudinal dimension have been estimated. The previously described inferior colliculus excitation mapping technique was used to estimate the function. Fibers nearest the electrode pair are excited at the lowest stimulus levels. Fiber threshold increases as the distance between the fiber's origin and the electrode pair increases. Suprathreshold stimuli stimulate approximately the same number of fibers originating on one side of the electrode pair, as those fibers originating on the other side. See Figure 3.1.

The low-frequency side of the "threshold vs. cochlear place" profile is shown for 15 bipolar electrode pairs in Figure 3.2. For bipolar electrode pairs judged to be "properly" positioned (the leftmost 12 curves, Figure 3.2) excitation threshold changed by about 10 to 25 dB. per octave (or about 10 to 25 dB. per 3 mm. sector of the basilar partition).

If bipolar electrodes are rotated far medial or lateral to this region, excitation profiles are not as sharp (rightmost three curves,

Figure 3.1. Response patterns for closely spaced, bipolar scala tympani electrodes.

The series of curves in A and B were derived from two different three-electrode arrays (inset) in two different cats. The electrode triad in both cats was centered at about the 10 KHz. location within the cochlea. Each curve represents the pattern of response generated by a different electrode combination of the triad (a-b, b-c, a-c). In the six examples shown, interelectrode spacing varied from about 400 to 1,100 um. For example, interelectrode spacing for the electrode illustrated in graph "B" was approximately 400 μ for a-c; 800 μ for a-b; and 1000 μ for c-b. Each point on the graphs represents the threshold to a 100 Hz. contralateral sinusoidal electrical stimulus (ordinate) for a given neuron or cluster of neurons within the central nucleus of the inferior colliculus, plotted as a function of the best frequency of that neuron to ipsilateral sound stimulation. The electrode shown in the inset is like those employed in these experiments. It is mounted in silastic which is molded to fill the scala tympani, with the contact surfaces of the electrodes facing upward toward the basilar membrane. The Pt-Rh electrode contacts are accentuated by the open circles, which are about 2.5 X their diameter. There is also an electrode pair in the more basal aspect of this particular scala tympani implant.


A
THRESHOLD (dB re 1V rms)
b-c
a-b
a-c
FREQUENCY (kHz)


a
b
c

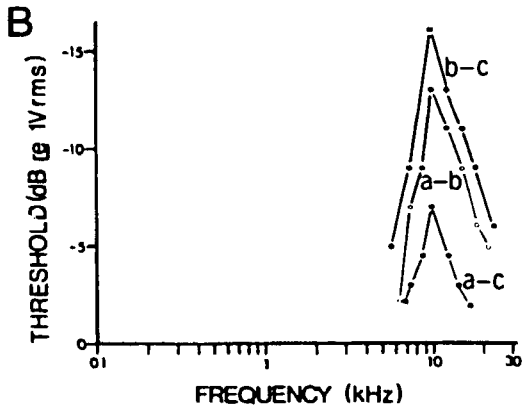

B
THRESHOLD (dB re 1V rms)
b-c
a-b
a-c
FREQUENCY (kHz)

Figure 3.2. Threshold plotted as a function of represented cochlear frequency, for the low frequency side of the restricted excitation region of 12 well-positioned intracochlear bipolar electrodes (left-most 12 curves) and 3 mispositioned intracochlear electrode pairs. Results are representative of all derived bipolar electrode maps. Interelectrode spacing varied from about 400 to about 1200 microns for illustrated examples. Divisions along the abscissa are octaves (or approx. 3mm sectors of the basilar partition of the cat).

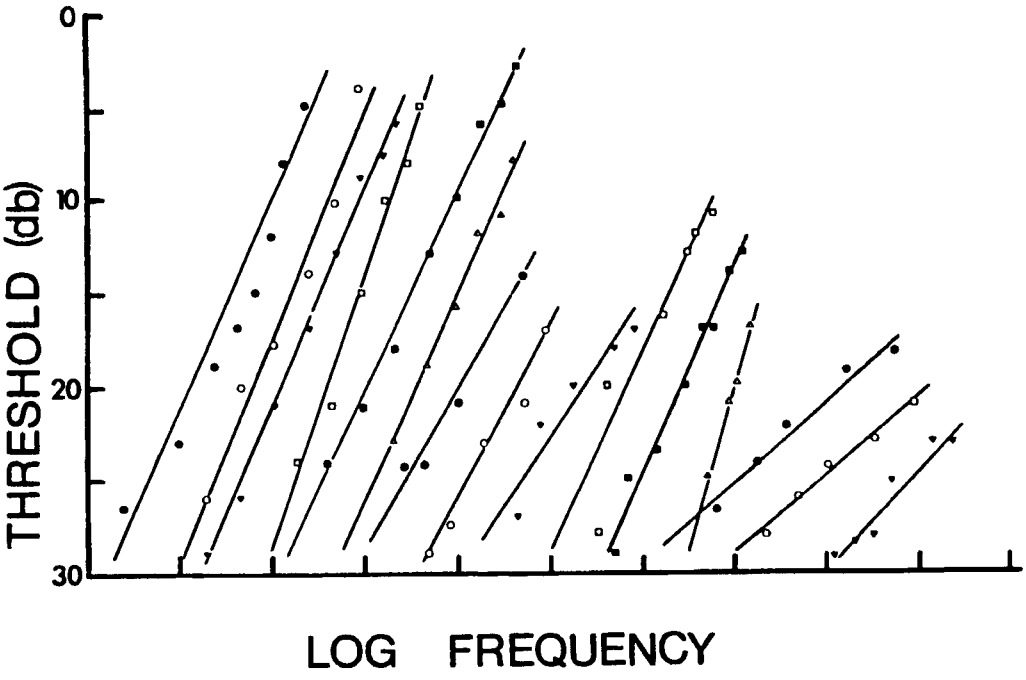

0
10
20
30
THRESHOLD (db)
LOG FREQUENCY

Figure 3.2). If the basilar membrane is perforated at the time of implantation, excitation with bipolar electrodes in the region of the perforation is inexplicably very broad. Three such electrode pairs have now been studied (e.g., Figure 2.4).

The shape of excitation curves, for longitudinally placed electrode pairs, has not been thoroughly studied. However, some data has been obtained for closely spaced pairs (Figure 3.1).

As contact separations are increased, the electrical threshold consistently decreases. Also, the increment in fibers excited, per unit increment in stimulating current, increases substantially as separation increases. Decreased threshold as contact separations increase is due, probably, to two factors. As electrode contact separation increases, the electric field "expands" spatially, such that higher current densities exist near the neural structures. The second factor was first described by Rushton (1927). Rushton found that a fiber's threshold decreased as the length of exposed fiber was increased. The exposed length of fiber was exposed to a constant current density. In our case, as contact separation increases, so does the effective length of nerve exposed to the electric field. As a consequence, neural threshold decreases.

Monopolar electrodes (with one of the two electrodes extracochlear; either on the promentory or on the scalp) are excellent examples of electrode pairs with large electrode separations. Monopolar electrodes exhibit low thresholds and very broad excitation curves (Figure 3.3). In some of these electrodes, an increment of only a few decibels from

threshold is required to stimulate nearly the whole acoustic nerve. Less than a decibel increase from threshold is required to excite a 4 mm. segment of nerve. Fine control of the number of neurons excited would be difficult with such a monopolar electrode. Implantable receiver-stimulators may be severly taxed to deliver the rather high amplitude resolution potentially required for such electrodes. However, coding requirements may not be so stringent. A fine control of the amount of nerve excited may not required.[1] However, even if fine control is not necessary, variations in fiber excitability as a function of longitudinal distance may cause multipeaked or other highly irregular and unpredictable excitation patterns.

Individual excitation patterns are altered when two or more electrode pairs are stimulated simultaneously (Figure 3.4 and 3.5). Electric field interactions can be avoided by using temporally non-overlapping, short duration, zero-net-charge stimuli (see Figure 3.6). Potentially, field interactions might move centers of excitation. Spatial excitation pattern shifting might lend itself well to a stimulus processor with formant tracking incorporated in the initial stages. Excitation modeling and inferior colliculus excitation mapping studies should better determine the detailed spatial excitation patterns obtained when two adjacent, radially oriented, electrode pairs are simultaneously stimulated.

Footnote:

1. Hansen and Lauridsen (1975), Chouard (1977), and Eddington (1977) were able to vary the sensation of loudness by modulating biphasic pulse rate. A pulse rate change probably does not significantly change the number of fibers excited, because these short duration biphasic pulses return unexcited fibers to their resting state.

Figure 3.3. Excitation patterns derived from the study of 4 typical monopolar scala tympani electrodes.

The approximate sites of the intracochlear electrodes are indicated by the numbers along the abscissa. Excitation with such electrodes has invariably been observed to be very broad, regardless of electrode size or location within the scala tympani. Also, the site of maximum excitation does not necessarily correspond with intracochlear electrode location. Stimuli were charge-balanced, biphasic (100 μsec. each phase) current pulses. Five decibels represents a 100 μamp. current level.

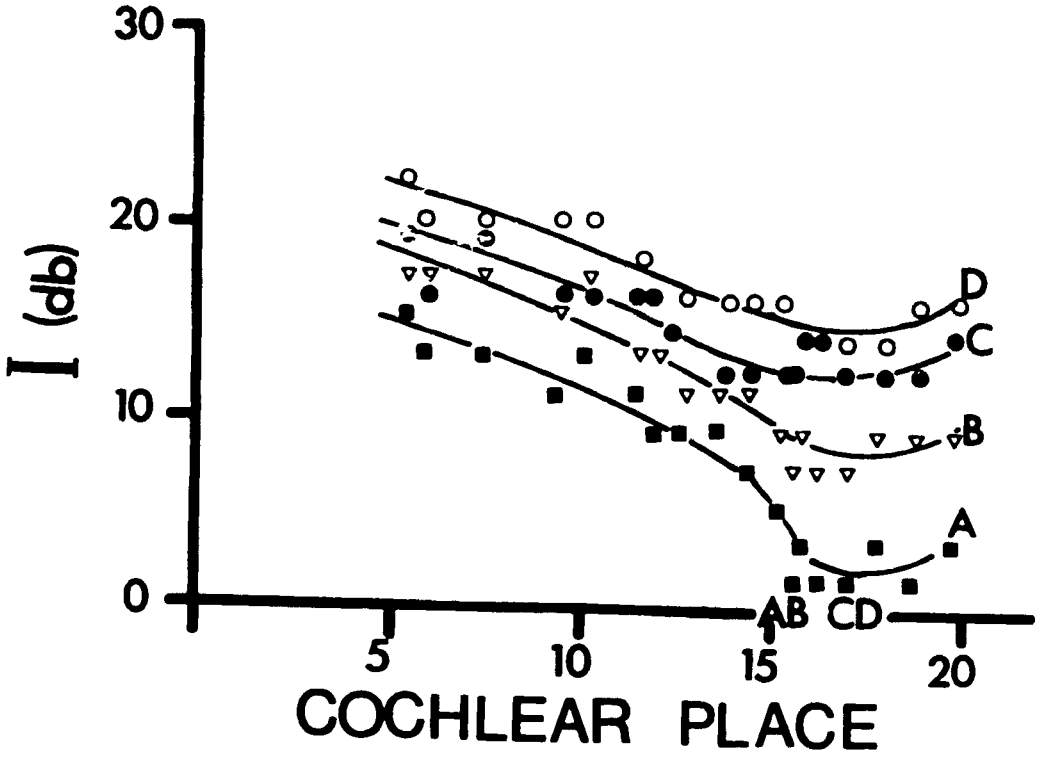
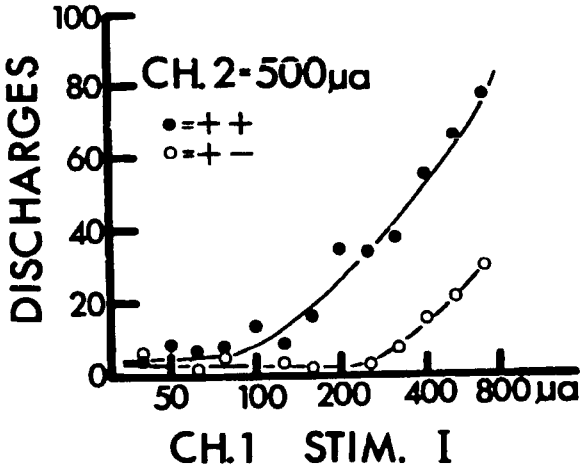

30
20
10
0
I (db)
D
C
B
A
AB
CD
5
10
15
20
COCHLEAR PLACE

Figure 3.4. This figure illustrates one class of experiments employed to study interelectrode interaction.

The basic conditions of the experiment are indicated schematically, at the right. Differences in responses of a neuron as a function of stimulus level when simultaneous biphasic pulses have the same phase (current adding: filled circles), as compared with opposite phases (current subtracting: open circles) provide a measure of the extent of field interation. This studied neuron derived its input from the cochlear partition in the region between two adjacent bipolar electrodes. Single channel, closely spaced, bipolar electrode threshold do not change appreciably (< 1 dB) when stimulus polarity is reversed.

FIELD INTERACTION



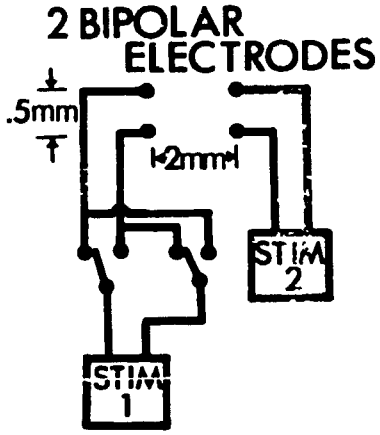
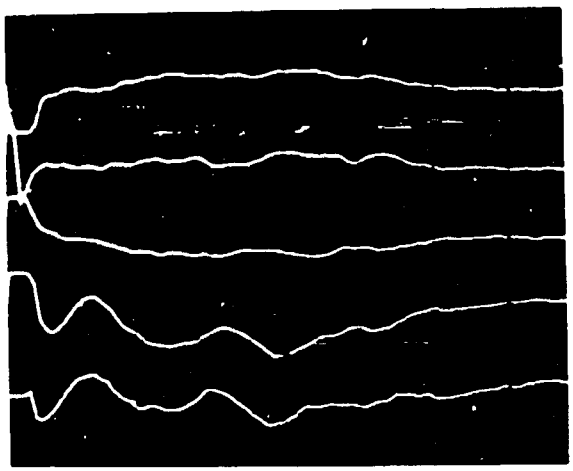


Figure 3.5. Interelectrode interaction, as revealed in far-field potential recording.

Electrodes 1 and 2 are bipolar pairs approximately 2 mm. apart. In this series both were being stimulated at current levels near threshold. The response to stimulation of "channel" 1 alone (2000 trials) is shown by the first trace, and by "channel" 2 by the 2nd trace. With stimuli delivered simultaneously out of phase, fields subtract, and the response in the third trace (below threshold) was derived. With stimuli delivered simultaneously in phase, the fields sum and a strong response is seen (4th trace). The difference (5th trace) between the third and fourth trace is a measure of the extent of interelectrode interaction.



1

2

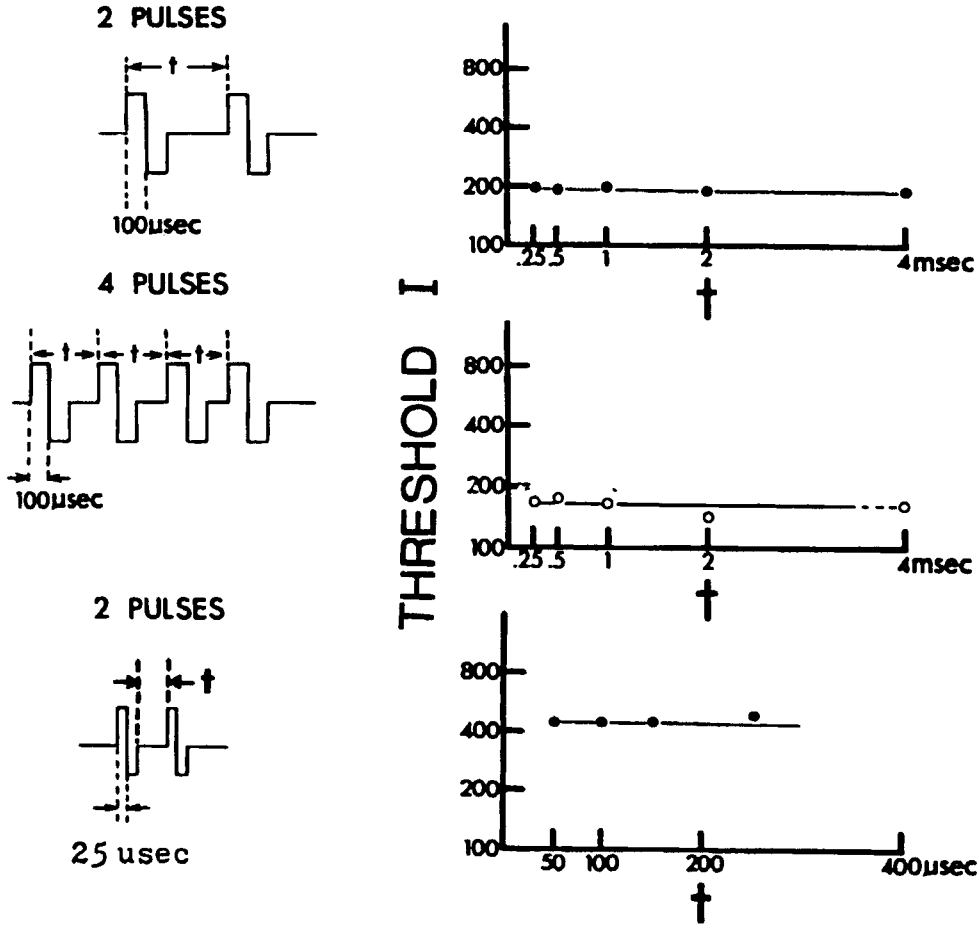
1+2 Fields subtracting

1+2 Fields summing

Difference

Figure 3.6. Interelectrode interactions can be circumvented by appropriate sequencing of charge-balanced biphasic stimuli.

The basic experiment is shown schematically for each graph, at the left. Inferior colliculus unit threshold is plotted as a function of interpulse interval for each stimulus condition, at the right. This and much similar data reveals that subthreshold biphasic stimuli presented as little as 50 μsec. apart in one channel cannot alter the threshold of excitation of an adjacent channel. This data is very compatible with Hill's model for neural excitation (Figure 3.7).


2 PULSES
t
100µsec
4 PULSES
t
t
t
100µsec
2 PULSES
t
25 usec
THRESHOLD I
800
400
200
100
.25
.5
1
2
4msec
800
400
200
100
.25
.5
1
2
4msec
800
400
200
100
50
100
200
400µsec

Spatial domain excitation models

Accurate and widely applicable theoretical electrical excitation models not only need to consider the rather complex impedance structure of the cochlea; but, in addition, need to consider the excitability of the neural processes as a function of the previously calculated potential distribution across the length of each neural element (McNeal, 1976). McNeal's model assumes that the electrical potential outside the neural process is determined only by: 1) the stimulus current, 2) tissue impedance distributions, and 3) electrode distribution and geometry. Also, it assumes that the electrical potential outside the neural process is not distorted by active processes within the nerve fibers. For the scala tympani implant, the impedance structure might be approximated by using concentric cylinders (or partially closed cylinders), analogous to the cochlear walls and the electrode carrier (if any). After the potential distribution has been determined, a neural model, such as the Frankenhaeuser-Huxley (1964) model (Figure 3.11) is used with anatomical data (Spoendlin; 1969, 1972) to determine the neural excitability due to the imposed potential distribution. The Frankenhaeuser-Huxley model is effective in both the temporal and spatial domains. A simpler, passive neural model (e.g. the passive version of the Frankenhaeuser-Huxley model) may be very useful in predicting spatial excitation patterns (McNeal, 1978).

2. Temporal domain excitation characteristics

What is the temporal pattern of neural discharge resulting from electrical stimulation? Hill (1939a) offers a simplified single fiber excitation model (Figure 3.7). Accommodative and passive membrane

impedance effects are simulated in this model. This lumped circuit model is very useful in estimating neural threshold for a wide range of stimuli. This model "predicts" the basic shape of the strength duration curve.

Merzenich (1973) and Kiang (1972) have experimentally determined the relationship between response threshold and sinusoidal electrical stimulus frequency. Their data are quite similar. The observed relationship is well modeled using Hill's model with t_m = 1 msec. and t_a = 4.5 msec. Hill, Katz, and Solandt (1936) have studied the excitation characteristics of nerve under the influence of sinusoidal current.

Incorporating an element that simulates membrane noise improves the model (Weiss, 1966). Kiang's figures illustrating "temporal discharge patterns in response to tone and sinusoidal current" (Figure 3.8) and "discharge rate as a function of stimulus current level" (Figure 3.9) are described, for the most part, by this model (Figure 3.10). Note that the widths of the peaks in the histograms of figure 3.2 are narrower for the electric stimulus than for the acoustic stimulus. System noise may be relatively greater in the case of acoustic-to-neural transduction than in electric-to-neural transduction. Noise-generated at the hair cell synapse may be the primary noise generator in normal acoustic-to-neural transduction (Schroeder and Hall, 1974).

Refractory effects have been added to this model. Geisler (1966,

Figure 3.7. Hill's lumped circuit neural excitation model.

This model is useful in predicting neural threshold. In this model, neural discharge occurs when the voltage across the threshold detector reaches a specified level. Accommodative effects and passive membrane impedance effects are simulated in this model.

Note that t_m is a function of electrode separation and placement (Hill, 1936a).

This model does not simulate refractory phenomenon. Consequently, excitability is not well predicted if the neuron recently discharged. For the acoustic nerve, refractory effects are negligible 2 msec. after discharge (Geisler, 1968).

This model does not consider the effect of lumping the spatially distributed membrane capacitance and resistance. For example, the finite intracellular resistance is not incorporated into this model. The more sophisticated Frankenhaeuser and Huxley (1964) model incorporates these features (Figure 3.11). As a consequence, the Frankenhaeuser and Huxley model more accurately represents spatial effects (e.g. electrode placement effects).

Variables and Constants:

$I(t)$ stimulus current.

c_m lumped membrane capacitance.

G_m lumped membrance conductance. $G_m = R_m$

R_a resistance used to simulate accommodation.

C_a capacitance used to simulate accommodation.

c_m membrane capacitance per unit area; typical value for frog sciatic nerve is 30.4 mmho/cm^2.

t_m membrane time constant = C_m/G_m. Using Merzenich's (1973) and Kiang's (1972) threshold vs. stimulus frequency data, t_m can be estimated. For widely spaced electrodes t_m is in the range of 1 to 2 msec.

t_a time constant for accommodation. Using the same data as for t_m, t_a can be estimated to be 3-6 msec.

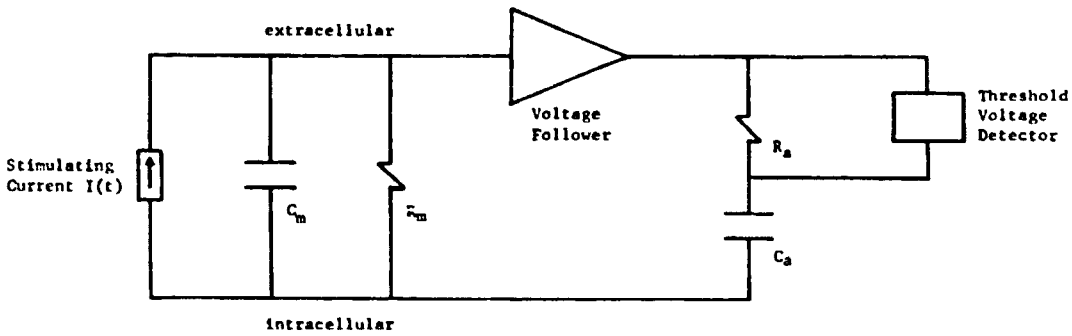

extracellular
Voltage
Follower
Threshold
Voltage
Detector
Stimulating
Current I(t)
C_m
r_m
R_a
C_a
intracellular

Figure 3.8. Temporal discharge patterns in response to tone and sinusoidal current.

The column on the left contains PST histograms of responses to a tone (0.2KHz., 67 dB SPL) for 6 units of different characteristic frequency (CF) in a normal cat. Zero time in these histograms is referred to the positive zero crossings of the voltage into the earphone. Upward deflection in the tracing at the bottom corresponds to rarefaction at the tympanic membrane. The column on the right contains PST histograms of responses to a sinusoidal current (0.2 KHz., 0.25 mamp. p-p) for 6 units in a cat treated with neomycin. (This drug causes hair cell destruction, but little ganglion cell loss.) Since these units did not respond to sound, no CF could be determined. Zero time in these histograms is referred to the positive zero crossings of the input to the stimulating electrodes. Upward deflection in the tracing at the bottom corresponds to negativity at the cotton wick electrode on the round window. The other electrode made of stainless steel, was placed on the periosteum over the apex of the cochlea. (From Kiang and Moxon, 1972).

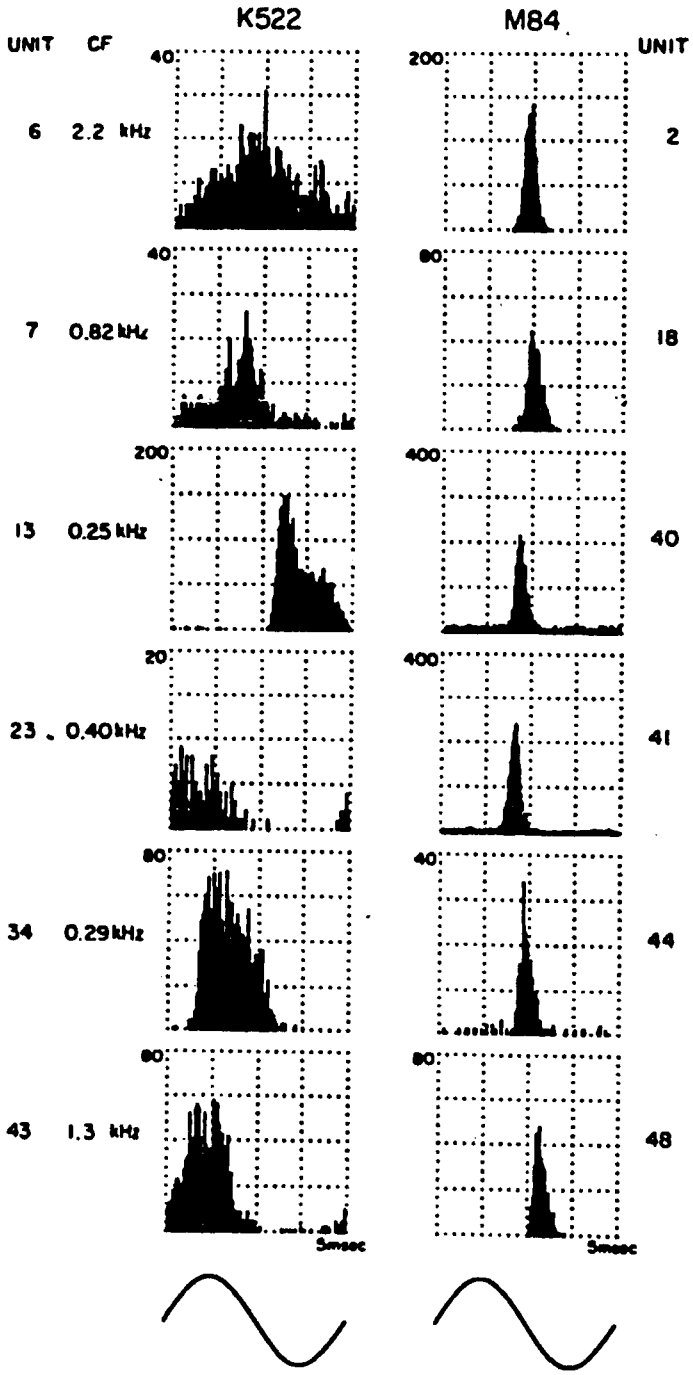

K522
M84
UNIT
CF
UNIT
6
2.2 kHz
2
7
0.82 kHz
18
13
0.25 kHz
40
23
0.40 kHz
41
34
0.29 kHz
44
43
1.3 kHz
48
5msec
5msec

Figure 3.9. Discharge rate in a single auditory nerve fiber as a function of stimulus level for both tone and sinusoidal current.

For each point shown, discharge rate was determined from the number of spikes that occurred during 20 seconds of exposure to a stimulus held at constant level. The horizontal for the two kinds of stimuli were aligned by equating the levels of tone and current at the respective thresholds. Frequency of both tone and current was 370 Hz. (CF). The tendons of the middle ear muscles were cut in this cat, to minimize electromechanical effects. (From Kiang and Moxon, 1972).

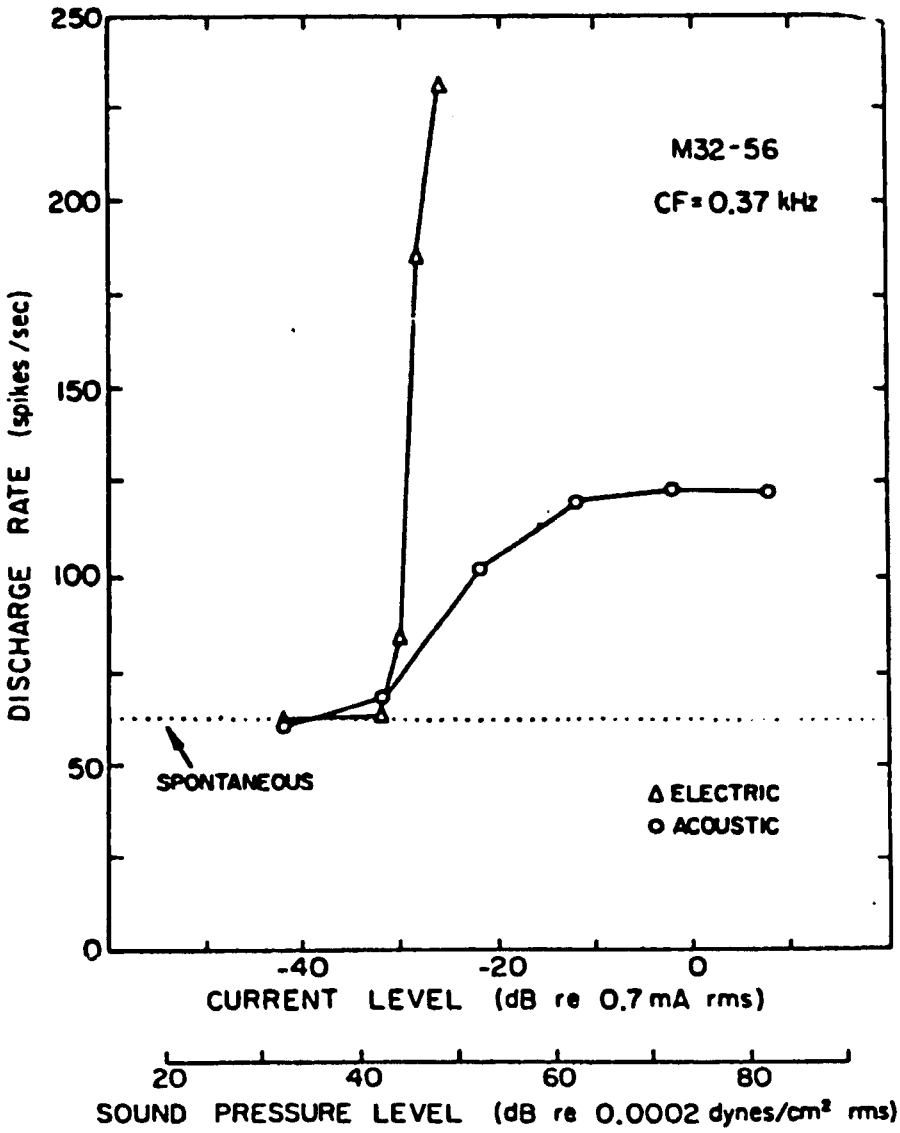

250
200
150
100
50
0
DISCHARGE RATE (spikes/sec)
M32-56
CF = 0.37 kHz
SPONTANEOUS
Δ ELECTRIC
o ACOUSTIC
-40
-20
0
CURRENT LEVEL (dB re 0.7 mA rms)
20
40
60
80
SOUND PRESSURE LEVEL (dB re 0.0002 dynes/cm² rms)

Figure 3.10. Single fiber electrical excitation model.

This model incorporates Hill's model. In addition, the neuron's stochastic behavior and refractory effects are simulated.

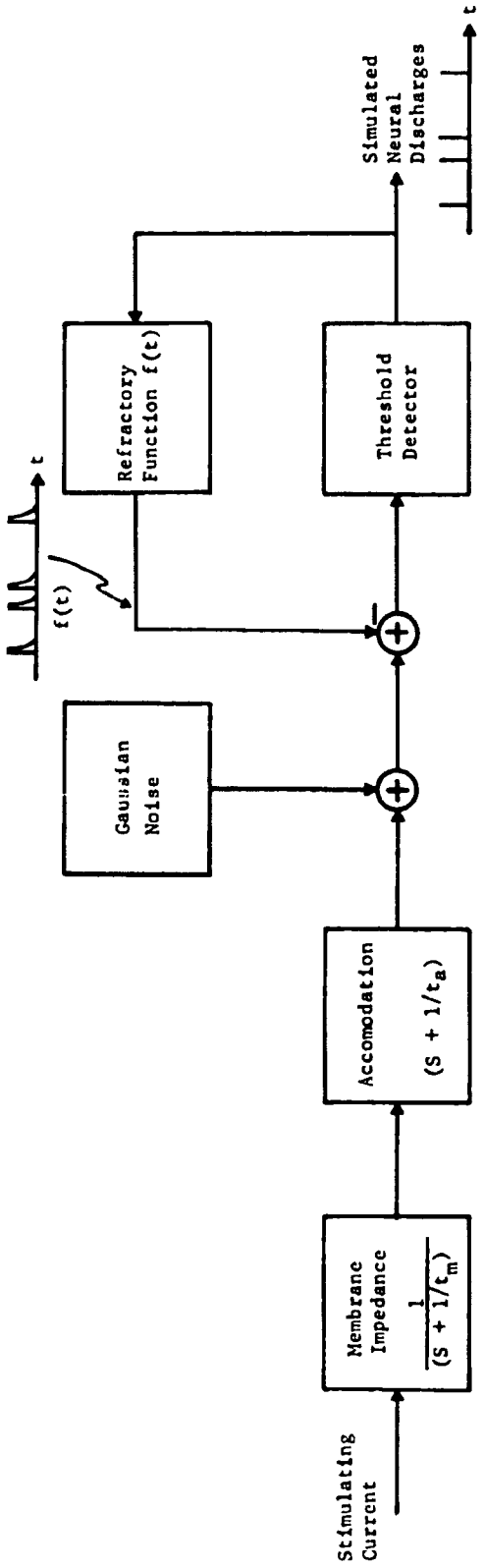
The probability of a neuron's discharge to a brief electric pulse is a sigmoid function of the pulse amplitude; indicating the presence of Guassian noise somewhere in the neural system (Verveen, 1959). An estimate of the power spectrum of ganglion cell "threshold noise", in a deafened cochlea, has not been attempted.

To account for refractory effects, Weiss (1966) used a simple exponential function with a 1 msec. time constant, to represent the change in threshold after discharge. Geisler (1968) divided the refractory period into an absolute refractory period (0.7 msec.) and a relative refractory period. His refractory function is expressed:

f(t) = infinity [for t between 0.0 and 0.7 msec.]

f(t) = exp(-t/tr) / (1-exp(-t/tr)) [for t greater than 0.7 msec.

where, t is the time after discharge onset and tr = 0.5 msec.]


Stimulating Current
Membrane Impedance
$\frac{1}{(S + 1/\tau_m)}$
Accomodation
$(S + 1/t_a)$
Gaussian Noise
f(t)
t
Refractory Function f(t)
Threshold Detector
Simulated Neural Discharges
t

1968) and Weiss (1966) use similar exponential functions to specify the decay of refractory effects after discharge. Geisler estimates that the refractory time constant is no more than 0.5 msec. for the auditory nerve. Geisler also incorporates an absolute refractory period of 0.7 msec. Both authors characterize acoustic nerve fibers as quickly resetting. The major part of the resetting process is generally completed within 2 msec. of discharge. Gray (1967), through an analysis of the conditional probability of spontaneous discharge, suggests that a fiber is not completely recovered until 20 msec. after discharge.

A more sophisticated deterministic model for myelinated nerve, from Franenhaeuser and Huxley (1964), is illustrated in figure 3.11. The spatial distribution of the nodes of Ranvier and the intracellular resistances separating these nodes are represented in this model. This model enables one to predict the spatial and temporal effects of various electrode placements. For instance, the changes in the stimulus strength-duration curve due to various electrode separations and placements can be simulated (McNeal, 1976). As a second example, the model predicts that the stimulus charge required for excitation gradually decreases as pulse duration decreases. This holds for durations as little as 10 μsec. The prediction agrees with empirical evidence for the myelinated nerve (Crago, 1974).

Additional electrophysiological data have been obtained from peripheral recording sites other than the acoustic nerve. Merzenich (1973) reports that the periods of low frequency sinsoidal stimuli were reproduced in the post-stimulus-time histograms for frequencies up

to 400 to 700 Hz. when 50 stimulus repetitions were averaged. This data is qualitatively similar to the periodic discharges obtained with periodic acoustic stimulation (Rose, 1967; Kiang, 1965).

In psychophysical tests of patients implanted with a single electrode channel, Bilger (1977) found that frequency difference limens (i.e., just noticeable differences) were normal for frequencies up to 250 Hz. Subjects exhibited a variety of performance levels above this frequency. Merzenich (1973) studied three such patients. Subjects could detect relatively small differences in stimulus frequency up to 300 to 600 Hz. One subject, for example, had difference limens of 2 Hz. at 100 Hz., 7 Hz. at 200 Hz., 9 Hz. at 500 Hz., and 60 Hz. at 900 Hz. Merzenich's subjects appeared to have no frequency discrimination above 1 KHz.

The results in the previous two paragraphs imply that neural discharge "periodicity", generated by electrical stimulation, is capable of conveying information to the patient for stimulus frequencies up to 500 to 1000 Hz.

Anecodotically, Michelson (1971, 1973) reports that some implanted (single channel) patients can discriminate between male and female speakers. At least one subject appeared able to discriminate among several speakers of the same sex. It should be emphasized that a number of cues could be used to discriminate amoung speakers; one of which is glottal pulse frequency. However, if this implication is correct, periodicity (at rates up to glottal pulse frequencies) information can be used for speech perception tasks.

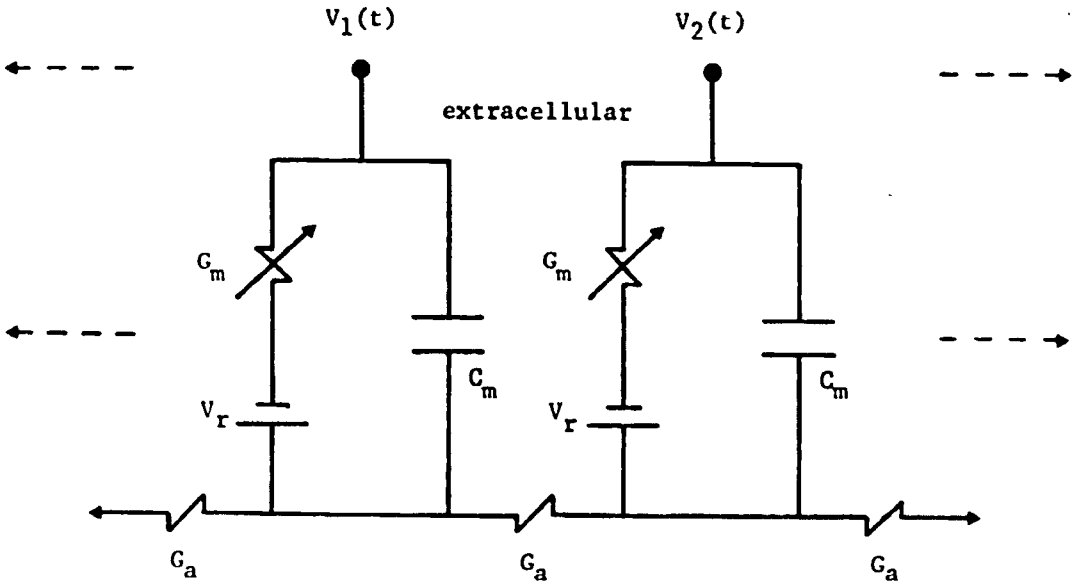
Figure 3.11. Frankenhaeuser and Huxley's nyelinated nerve model

This model assumes that the electrical potential outside the fiber is determined only by the stimulus current, passive tissue impedances, and electrode geometry. G_m can be considered constant for voltages less than 80 percent of the threshold voltage. In this case, the model reduces to the familiar passive core-conductor model (BeMent, 1969b; McNeal, 1978). Note that neural noise phenomena are not simulated with this model.

Variables and Constants:

C_m	Nodal membrane capacitance.
$G_m(t,V_m)$	Nodal membrane conductance, time and voltage dependant (Hodgkin and Huxley. 1952).
G_a	Internodal axoplasmic conductance.
V_r	Steady state resting potential.
$V_e(t)$	Extracellular nodal potential.
$V_i(t)$	Intracellular nodal potential.
$V_m(t)$	Membrane potential = $V_i - V_e - V_r$

For a more complete listing of variables, conductance equations, and the recent estimates of typical fiber parameters, see McNeal, 1976 .


$V_1(t)$
$V_2(t)$
extracellular
G_m
G_m
C_m
C_m
V_r
V_r
G_a
G_a
G_a
intracellular

3. Summary of Excitation Control Techniques

Bipolar electrode placement has a profound effect on the spatial excitation pattern. The threshold decreases and the spread of excitation, per increment of stimulating current, increases with increasing intracochlear contact separation. The "place" of excitation is dependent on the location of the bipolar electrode pair. In a 16 wire electrode array, 120 electrode pair combinations are available. Electrodes with more than two contacts offer even more combinations. However, some combinations with large separations will broadly excite the nerve and probably will allow little control of the spatial excitation pattern. With bipolar electrode pairs investigated, stimulus amplitude primarily changes the excitation "width" (i.e., the excitation pattern edges) but not the center (i.e., the 1st moment) of excitation. After determining electrical threshold as a function of cochlear place for a fixed electrode pair, one can control stimulus amplitude to effect any "width" of excitation desired.

If electrode pairs are simultaneously stimulated, excitation patterns cannot be predicted merely by spatially summing each channel's independent response. Temporally interlacing (i.e., the waveforms do not temporally overlap) charge-balanced, electrode-channel pulses will circumvent this problem. However, field interactions may be useful in increasing the reperatoire of excitation patterns that may be elicited with electrical stimulation. With such "multipolar electrode channels" (i.e., more than two electrode contacts comprise an electrode channel), appropriate electrode contact spacing anc array geometry may play a key role in the effective use of such

interactions.

Temporal firing patterns can be estimated using a model for each fiber (e.g. figures 3.10 or 3.11) and a model for the tissue's impedance. The models should reflect the variations among fibers. For example, fiber threshold is a function of the fiber's distance from the electrode.

With one electrode pair, it is fairly simple to mimick a normal firing pattern in any one particular fiber. The biphasic pulse amplitude is set substantially above the fiber's threshold. If the pulse width is short and the fiber is not in refraction, each biphasic pulse will generate one temporally coincident neural discharge. If the fiber's spontaneous activity is insignificant, this technique allows almost exact temporal control of discharges in any one particular fiber.

In a population of fibers, all fibers whose threshold is sufficiently low and are not in refraction will synchronously discharge with short duration biphasic pulses. Those few fibers whose threshold is marginal will not discharge on every biphasic pulse, but only on some of the biphasic pulses. If longer duration stimulus waveforms are used, tissue impedance effects should be considered; because stimulus waveforms might be significantly altered across the cochlea.

In one processing scheme, one would use a bank of bandpass filters to analyze speech. The output of each filter would be further processed and then used to excite a small population of

acoustic nerve fibers. One might attempt to electrically excite each small population of neurons such that the composite post stimulus time histogram of this small population of neurons is similar in shape to the envelope of a bandpass filtered version of the speech signal (Kiang and Moxon, 1972). If biphasic pulse stimulation waveforms are used, some combination of pulse amplitude and pulse rate modulation might be used to obtain the desired composite PST histogram. Pulse amplitude modulation would vary the number of fibers excited and pulse rate modulation would vary the neural discharge rate.

One could attempt to excite the nerve with more than low frequency envelope information (0-25 Hz.). Higher frequency envelope information (50-300 Hz.) might help to convey voicing information. Still higher frequency "phase-locking" information might be used by the nervous system (Evans, 1977; Moller, 1977). Electrical stimuli, phase-locked to the spectral peak(s) within the bandpass filter's passband, might be useful to the nervous system.

One variation of this stimulation method would involve the temporal interlacing (within each electrode channel) of a range of biphasic pulse amplitudes. The higher amplitude pulses would occur relatively infrequently compared to the total number of biphasic pulses generated. Such a stimulus sequence could generate relatively low discharge rates on the perimeter of each electrode channel's excitatory domain. At distances closer to the particular electrode channel, discharge rates would increase. Such a pattern would more closely mimick the excitation pattern generated by a low SPL acoustic tone. Channel interactions, proper stimulus amplitude control, and

electrode sequencing might be used to approximate the asymetrical spatial-temporal patterns observed, particularly when mimicking higher stimulus levels (see Figure 1 of Pfeiffer, 1975; and Figure 7 of Evans, 1975b). Spatial excitation pattern mapping, using BSER and psychophysical techniques, may be useful in determining the parameter values associated with this stimulation method.

Near each electrode channel, a small population of neurons will discharge almost synchronously if the biphasic pulse is relatively short and the nerves are not in refraction. Although this stimulation method may prove to be very useful, this type of stimulation does not generate the normal nerve's activity; in that normal acoustic nerve activity does not behave so synchronously during speech reception (Johnson, 1976). Other stimulation techniques may be useful if it is necessary to reduce the synchronization of the discharge patterns within each small population of nerves. Continuous waveforms and/or longer fixed-shaped sets of waveforms offer possible solutions. At present, it is not clear what stimulus waveform processing could be used to elicit composite temporal discharge patterns similar to those elicited from acoustic stimulation of a normal cochlea. Models of the nerve that reflect the excitability of the nerve before and during refraction would be useful in estimating effective stimulus processing methods (Fitzhugh, 1961; Hodgkin and Huxley, 1952). System identification techniques (Marmarelis, 1973a,b; Stark, 1968) could be applied directly to the study of nerve response to electrical stimulation.

Although it might not be possible to find an inverse to such a complicated function, reasonable approximations probably could be obtained. As a first approximation, an inverse filter could be estimated by finding the inverse filter of the linear filter function in Hill's model.

It may be useful to view each fiber as having a coupling coefficient associated with each electrode channel.[1] The coupling coefficients could represent the coupling between the electrode channel's stimulating waveform and the fiber. The coupling coefficients may be complex numbers due to tissue impedances. The coupling coefficients would decrease, as the distance between the fibers and electrode contacts increase. Each fiber would have a coupling coefficient associated with each electrode channel.

---

1. An electrode channel can be composed of two or more electrode contacts or poles. Controlling currents at these multiple poles is used to change the spatial and temporal aspects of the excitation patterns.

4. Effects of long-term implantation

Schindler (1974, 1976) and associates have conducted long-term implantation tests. Cats were implanted for up to 2.5 years. They were then sacrificed and their cochleas histopathologically examined. The great majority of ganglion cell processes in the organ of Corti survived long-term implantation. Hair cells directly above the scala tympani implant did not survive. Severe nerve degeneration was observed near perforations of the basilar membrane, or following damage to the bony endosterum.

In addition, a number of these long implanted, neomycin deafened cats were electrically stimulated. Recording from the inferior colliculus, the investigators verified that thresholds were not different from those obtained from recently implanted cats (Merzenich, Schindler, White, 1974).

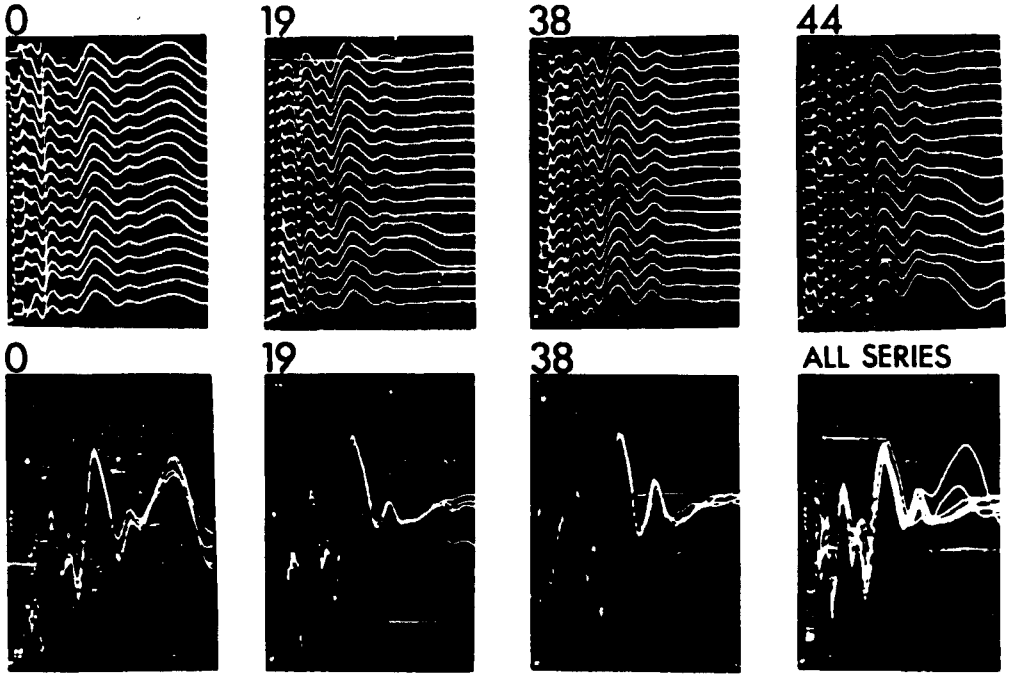
5. Effects of high-level and long-term stimulation

Brummer (1975, 1977), Dymond (1976), and Guyton and Hambrecht (1974) have reviewed the general topic of high level and long-term stimulation using noble metal electrodes.

The effects of high-level, continuous stimulation within the scala tympani for moderate periods have been examined (Merzenich and White, 1977). Figure 3.12 and Figure 3.13 illustrate the results of these "overstimulation" experiments. The general experimental approach employed in these studies is reviewed in the figure legend of figure 3.12. The responses to continuous, high-level electrical stimulation were continuously monitored; and threshold was measured at frequent

Figure 3.12. Representative case from study of functional-histopathological consequences of heavy continuous stimulation.

In this example the stimulation level was 1 at 100 μsec. per phase. Each series at the top represents brainstem evoked responses for 12,000 stimuli per trace. Each trace is the averaged response due to 10 charge-balanced, biphasic 100 μsec. pulses per second for 20 minutes. The number at the top of each series is the time (in hours) of the first stimulus in the 6-hour series, from onset of continuous stimulation. Below each series, the 2nd, 6th, 11th, and 16th traces are shown superimposed. In the superimposed traces labeled "ALL SERIES", the fifth trace from 8 different six-hour series are shown superimposed. The thresholds for brainstem responses were derived at 6-hour intervals and were found constant (upper graph). The amplitude of a given wave in the far-field potential response series varied significantly through the duration of stimulation, but showed no overall decline during more than 50 hours of continuous, heavy stimulation. The stimulus level of one milliampere at 100 μsec. per phase is probably a realistic upper limit for operational multichannel prostheses.


0
19
38
44
0
19
38
ALL SERIES

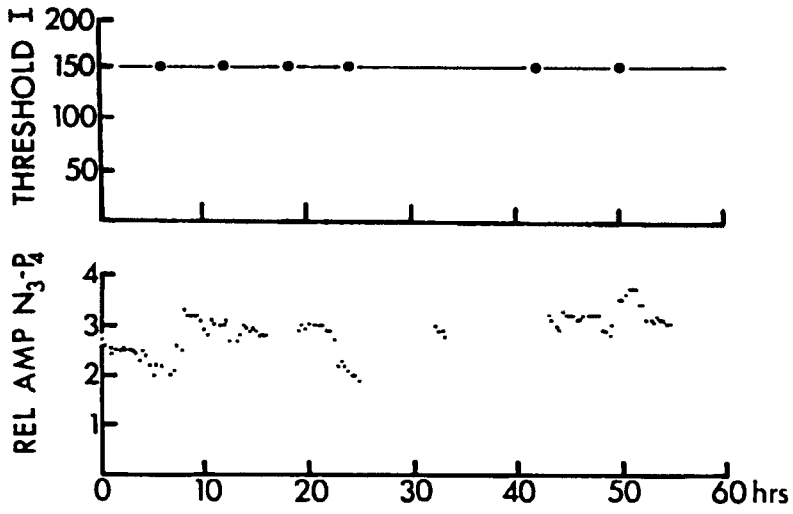
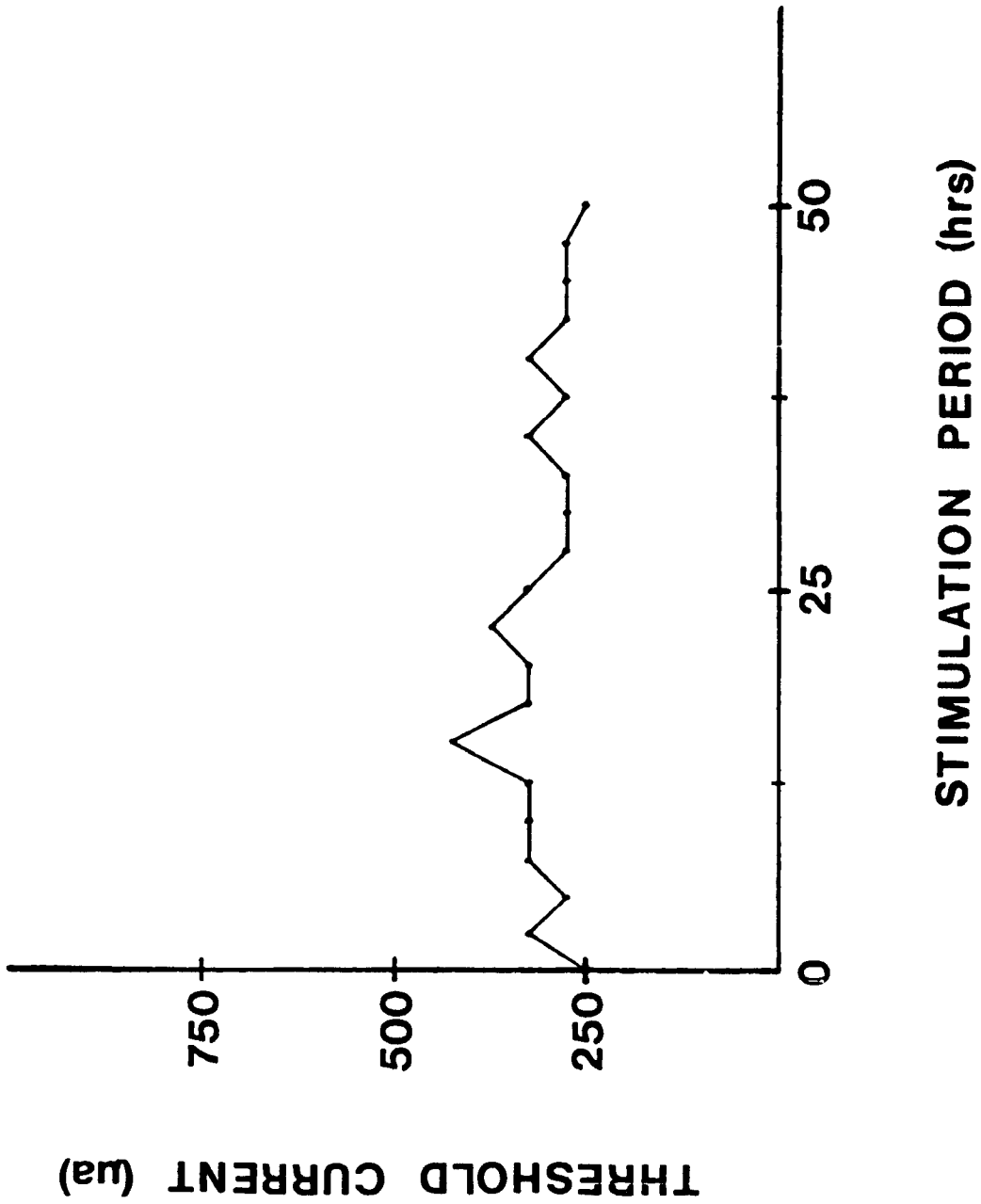

THRESHOLD I
200
150
100
50
REL AMP N3-P4
4
3
2
1
0
10
20
30
40
50
60 hrs

Figure 3.13. Threshold current as a function of high-level stimulus exposure time.

This cat was stimulated with 2 ma. balanced-change biphasic pulses (200 μsec., each phase) at 30 pps. Total exposure time was approximately 50 hours. The threshold to biphasic current pulses (100 μsec., each phase) was determined at 2.5-hour intervals throughout the 50-hour exposure. Measured threshold did drift, but no upward trend in threshold was noticeable.

Variations in threshold were more pronounced in this experiment than in the previously illustrated case. This electrode's BSER magnitude "grew" very slowly with increases in stimulus current. Consequently, small changes in stimulus artifact and biological noise perturbated the measured threshold more than in the previous illustration.


THRESHOLD CURRENT (ua)
750
500
250
0
25
50
STIMULATION PERIOD (hrs)

intervals. The charge and current stimulation levels were chosen to exceed those required for operation of multichannel cochlear implants. The exposure was two to three days in duration. The cats were sacrificed at approximately three weeks after exposure; their cochleas then were perfused, and the auditory nerve examined in detail. No damage to auditory nerve fibers in these two animals was observed. The electrophysiological evidence is in agreement with this finding.

Studies using progressively higher stimulus levels, pulse rates, and longer term continuous stimulation have been initiated in this laboratory in an effort to determine the upper limits of "safe" intracochlear stimulation (Walsh, 1978).

Although no adverse effects to neural tissue have been observed to date, there is an electrically-induced change in the connective tissue above the electrode.

6. Stimulation of neural structures not associated with the auditory system

Stimulation at levels above those required for operation of multielectrode nerve stimulation devices (up to 5ma, 100 μsec. per phase of the biphasic pulse) with bipolar electrodes (with electrode element separations up to 10 mm.) does not lead to obserable excitation of the facial nerve or muscles. With monopolar stimulation, facial nerve or direct muscle stimulation is commonly observed at stimulus levels as low as 200-500 μamp.

Vestibular responses, such as reflex eye movement, were not routinely monitored. Future studies should include such

monitoring, especially when the animal is being stimulated while fully alert.

7. Electrode voltage-current characteristics

The measurement of electrode voltage as a function of stimulating current is pertinent to electronic stimulator design. Also, the measurement may give some insight into the electrochemical effects occurring at the electrode-tissue interface.

Controlled current sinusoidal and biphasic pulse waveforms at selected amplitudes and frequencies are automatically delivered to electrodes in normal saline or *in vivo*. The electrode voltage as a function of time is recorded digitally and later plotted. The results of this testing are similar to those summarized by Dymond (1976). The effects of frequency dispersion were noted. Increases in electrode admittance were observed at the higher current levels. Larger surface area electrodes (e.g. 5-25 mil. diameter spheres) exhibited less of this effect at the same stimulus levels. One possible explanation is that fewer of the higher energy reactions are occurring in the larger surface area electrodes as a consequence of the lower current densities generated.

B. Perceptually significant features of the acoustic speech signal

A great deal of research has been directed at understanding the perceptually significant features of the acoustic speech signal. The reader is referred to the excellent reviews of this research (Liberman, et al, 1967; Flanagan, 1972; Broad and Shoup, 1975).

Most investigators consider speech recognition to involve several interactive stages of analysis. These stages might be ordered as follows: speech signal input, speech signal parameter extraction, phonetic and prosodic analysis, syntactic and symantic analysis, and final output. The interaction between these analysis stages appears to be very significant.

Initially, because of the humans' extremely adaptive perceptive capacities, as compared to those of current, automated speech recognition systems, stimulus processing algorithms probably will not be designed to substitute for the higher levels of the speech analysis "chain" such as phonetic analysis. Initially, the stimulus processing algorithms probably will be designed to effectively communicate parameters extracted from the speech waveform. In some cases, these parameters will be chosen to directly approximate or mimic the function of the outer, middle, and inner ear functions. In other cases, other choices of parameters and stimulation methods will be used because it is believed that the subject may readily learn to use such stimuli (House, Stevens, Sandel and Arnold, 1962). Initial differential discrimination tests will allow the investigator to efficiently "fine tune" the processing algorithms. These differential

tests will indicate which stimulation methods most effectively communicate signal characteristics (such as temporal changes in the signal) to the subject. The subject's ability to make useful absolute discriminations after a learning period will be used as an index to the effectiveness of these processing algorithms.

A reasonably complete list, (Broad and Shoup, 1975) of those parameters that most effectively represent the perceptually significant aspects of speech, is presented below. These parameters can be used to describe most of what we know about the acoustic speech wave and its relation to phonetic units.

The first four formant frequencies, their bandwidths, and their amplitudes.

The formants are the natural resonance modes of the vocal tract. A formant has a frequency of oscillation F_i and a time constant or, alternatively, a bandwidth B_i. The formant also has an amplitude A_i which has been shown by Fant to be largely redundant if the formant frequency and bandwidth are known. The formant amplitude is not totally redundant, however, and is included here for the sake of completeness.

Fundamental Frequency (F_o)

The fundamental frequency F_o is the same as the frequency of oscillation of the true vocal folds. Note that the formant frequencies have no necessary harmonic relation to the fundamental.

Energies of contiguous frequency bands

These frequency bands may be logarithmically spaced at third or

half-octave intervals or uniformly spaced at some interval. When reliable measurements of the formant parameters can be made, the spectral energies are largely redundant; however, there are many cases, especially during voiceless sounds produced by turbulence noise, when the formants are not measurable, or perhaps not useful, and more accessible parameters such as spectral energy are needed.

Voicing detection

Accurate voicing detection is very useful in automated speech recognition algorithms and most vocoders.

The number of zero crossings of the speech signal

This measure can be helpful for voicing detection, or for turbulence noise detection. It is somewhat redundant relative to the previously listed parameters.

The extreme value within 5-25 msec. segments of the speech wave divided by the rms value of the segment

This parameter could be used for locating bursts in the speech wave. These result from the releases of stop consonants.

The rms energy of each 5-25 msec. segment of speech

This parameter is simple to obtain and can be very useful. For example, it can be used to detect silence and gaps in the speech waveform.

Although this list is not complete, it represents all of the major features of the speech waveform that are required for the perception of the speech signal. However, it is not believed that information reguarding the speaker's voice quality would be ascertain-

able from these parameters.

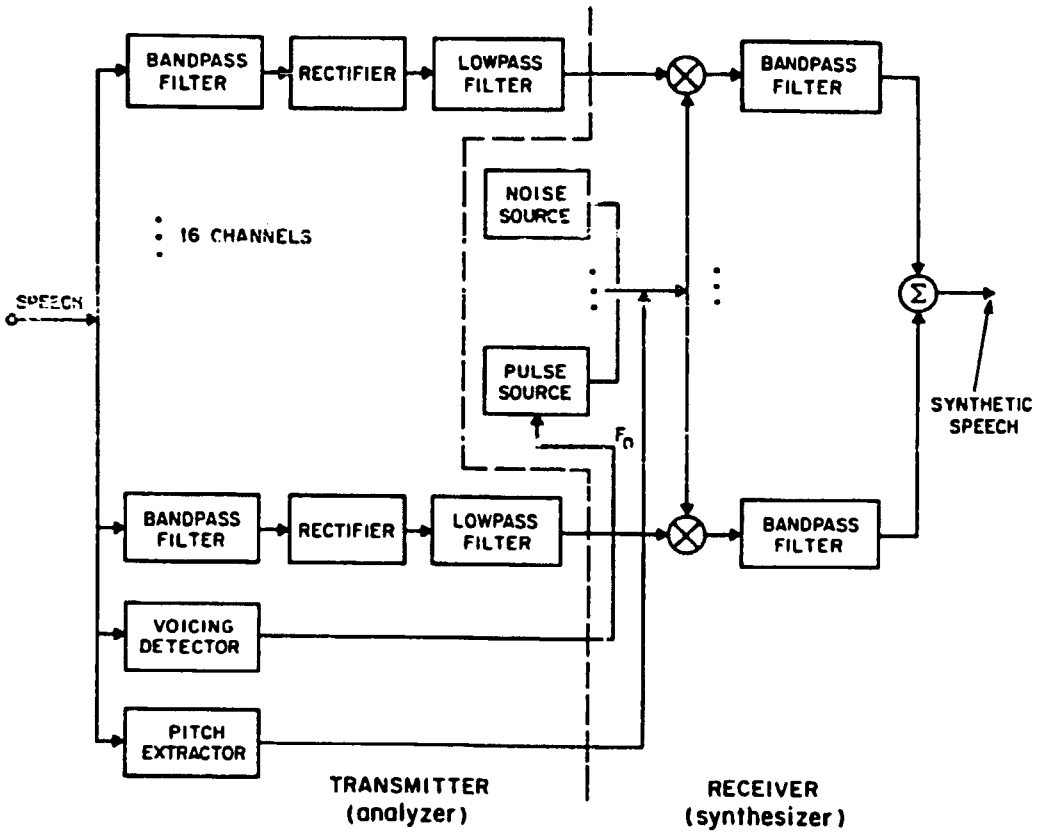
Vocoder studies and their relevance to the understanding of speech perception

Vocoder realizations (Flanagan, 1972; Dudley, 1939) offer insight into the perceptually significant features of the acoustic speech signal. Vocoder analysis stages generally reduce the information rate required to convey speech. Redundant and irrelevant information is discarded in the analysis stage. A number of vocoder analysis stages (e.g., the spectrum channel vocoder analysis stage) appear to be strikingly similar to signal processing models of the inner ear (Siebert, 1973, 1968; Weiss, 1966; Geisler, 1968). Because of the above mentioned characteristics, vocoder analysis stages may be useful in the initial processing stages of an auditory prosthesis.

The spectrum channel vocoder incorporates one important constraint of speech production and one of perception. The channel vocoder recognizes that the vocal excitation can be a broad spectrum, quasi-harmonic sound when the speech is voiced, or a broad spectrum, random signal when the speech is unvoiced. It also recognizes that perception, to a large degree, is dependent upon the preservation of the shape of the short-time amplitude spectrum. Figure 3.14 is a block diagram of a standard spectrum channel vocoder.

The excitation information is measured by the bottom two branches of the circuit. The excitation analysis circuitry is composed of a voicing detector and a fundamental frequency extractor. The voicing detector determines whether there is a significant periodic component

Figure 3.14. Block diagram of a standard spectrum channel vocoder. Both the analysis and synthesis stages are illustrated. The diagram is from Rabiner and Gold, 1975.


BANDPASS FILTER
RECTIFIER
LOWPASS FILTER
BANDPASS FILTER
NOISE SOURCE
16 CHANNELS
SPEECH
PULSE SOURCE
F_0
SYNTHETIC SPEECH
BANDPASS FILTER
RECTIFIER
LOWPASS FILTER
BANDPASS FILTER
VOICING DETECTOR
PITCH EXTRACTOR
TRANSMITTER (analyzer)
RECEIVER (synthesizer)

in the speech signal. If so, the detector signals the synthesizer's circuitry to generate a periodic signal to mimic the speech signal's excitation component. The frequency of the periodic excitation component is estimated by the pitch (fundamental frequency) extractor. The fundamental frequency signal is smoothed by a lowpass filter of approximately 25Hz. If the speech signal contains an insignificant periodic component, the voicing detector commands the synthesizer to generate a random signal to mimic the speech signal's aperiodic excitation component.

Eight to sixteen spectrum channels in the upper part of the circuit measure the short-time amplitude spectrum at eight to sixteen spectral bands. Each channel includes a bandpass filter, a rectifier, and a lowpass filter (25 Hz.).

The speech spectrum is reconstructed from the transmitted data at the synthesizer (receiver). Excitation, either from a pitch-modulated, constant average power pulse-generator, or from a broad-band noise generator, is applied to an identical set of bandpass filters. The outputs from the filters are amplitude modulated by the spectrum-defining signals. The amplitude is not modulated at a rate any greater than 25 Hz. A short-time spectrum, approximating that measured at the analyzer (transmitter), is recreated.

A ten channel vocoder requires a transmission bandwidth of approximately 275 Hz. (10 spectrum channels at 25 Hz. apiece plus a 25 Hz. channel for the fundamental frequency). The intelligibility of the synthesized speech can be maintained relatively high, with a

vocoder having as few as ten channels. Typical syllable intelligibility scores for a ten-channel vocoder are approximately 84 per cent. However, voice quality and naturalness are significantly altered in transmission by vocoder.

As previously noted, the outputs from the individual synthesizer filters are not modulated at a rate any greater than 25 Hz. It is doubtful that the cochlea can spatially resolve frequency components within this 25 Hz. band at cochlear places equivalent to 500 Hz. or higher in frequency (Ritzma, 1974; Flanagan, 1955). Consequently, only the temporal features of the firing patterns are probably used by the nervous system to extract the envelope components of the vocoder synthesized speech signal. As a consequence, it may not be necessary to implant more electrode channels than that number of channels required for a standard spectrum channel vocoder; if each electrode channel can communicate to the nerve this temporal information; and in turn, the more central nervous system can appropriately adapt to the differences between the normal and electrically elicited spatial-temporal firing patterns. The discussion above has neglected the voicing information contained in the vocoder synthesized speech signal. Although voicing information may be conveyed to a normal hearing person by either or both spatial and temporal mechanisms, temporal mechanisms appear sufficient to communicate this information to the implanted subject. Experiments with implanted subjects appear to verify that a single electrode channel is capable of communicating at least some of the voicing-fundamental frequency information

(Bilger, 1977; Merzenich, 1973; Michelson, 1971, 1973; also see end of the section entitled: "Temporal domain excitation characteristics").

If necessary, further reductions in the acoustic signal's redundancy can be obtained. The spectrum channel vocoder's channel signals are not completely independent. One channel vocoder variation, called a peak-picking vocoder, attempts to reduce this dependence (Peterson and Cooper, 1957; Flanagan, 1972). It operates by transmitting three to five or more channel signals which, at any instant, represent local maxima of the short time spectrum. The identities of the "picked" maximum channels and their amplitudes are signaled to a conventional multi-channel vocoder synthesizer. Thus, at any one time only a few channels of the synthesizer are activated. This version of a peak-picking channel vocoder is very similar to a formant vocoder with rather coarse formant frequency quantitization.

Formant vocoders generally utilize knowledge about the speech production mechanism to a greater extent than the other vocoder realizations. For instance, information about the filtering capacity of the human vocal tract is utilized to determine the number of formants and the number of poles required to model the tract for formant trajectory calculations. Also, formant trajectories are known to be reasonably continuous functions because of the continuity in the motion of the vocal tract. These and other constraints are used in formant vocoders to reduce channel transmission bandwidth requirements without severely compromising the intelligibility of the speech. However, the formant vocoder does strip some useful information from the input signal (see "Energies of contiguous

frequency bands" above). As the channel bandwidth is reduced, more useful information is lost. Also, if the speech signal is in a noisy environment, the formant vocoder's performance is degraded more than the channel vocoder's performance in a similar environment. Formant vocoders depend heavily on the single-speaker-low-noise model for formant tracking. If extra-speaker spectral components are present, the formant extraction algorithm often fails by inadvertently tracking a "phantom formant". The spectrum channel vocoder is more "robust" in a noisy environment. Spectral information from both the noise source and speaker summate in a manner more similar to that observed in a normal unaltered hearing environment. The listener's very capable perceptive mechanism is more able to differentiate between the two sources when a spectrum channel vocoder is used.

A modified version of the formant or spectrum channel vocoder processing may be useful. Inter-speaker variability is a major component of the variability in extracted parameters (Broad and Shoup, 1975). It may be useful to reduce this inter-speaker variability. For instance, the channel vocoder's spectral estimates or the formant vocoder's estimated formant frequencies could be normalized in order to diminish the inter-speaker variation of these parameters. An average of the speaker's acoustic characteristics could be used to normalize the speaker's vocal tract filtering parameters.

The spectrum channel vocoder and the formant vocoder analysis stage suggest stimulus processing models for a cochlear prosthesis.

The spectrum channel vocoder stimulus processing might involve stimulating n electrode channels; each electrode channel might represent one vocoder analysis channel. Each analysis channel's output magnitude might be conveyed by changing stimulus amplitude and/or by changing stimulus timing (e.g. frequency) and/or by changing stimulating electrodes and/or by utilizing electric field interactions between three of more electrode contacts. These techniques interact with each other to affect both the spatial and temporal aspects of the excitation pattern.

Peak-picking and formant vocoders offer a very similar analogy to the channel vocoder analogy. These analysis stages might be used to extract the formant frequencies and then disable excitation on all electrode channels that do not represent these formant frequencies. Those channels that most closely represent the formant frequencies would be treated much as those in the spectrum channel vocoder analogy. Spectral magnitude at those frequencies might be conveyed by changing stimulus amplitude and/or by changing stimulus timing (e.g. frequency) and/or by changing stimulating electrodes and/or by utilizing electrode electric field interactions between three or more electrode contacts.

The formant vocoder analogy reduces the number of electrode channels that are stimulated during any 5-25 msec. segment of speech. This could be advantageous if the subject has difficulty in assimilating more information than these few channels. However, if the subject is capable of benefitting from the additional information

presented on the other electrode channels, a spectrum channel vocoder analogy may be more appropriate. Perhaps the peak-picking vocoder analogy offers a middle ground between the channel and formant vocoder analogs. The number of local maxima that are picked can be specified by the investigator. Therefore, the number of active channels can be adjusted to determine subject performance as a function of the number of active channels.

Further reduction in the information transfer rate may be required. It may only be possible to communicate a very small amount of information over a fixed amount of time. As a consequence, only the most useful information extracted from the speech waveform should be communicated. The second formant is perhaps one of the most important low data rate parameters that could be conveyed (Liberman, et al, 1967). Lip reading and other non-acoustic input should also be considered in determining what parameter subset would be most useful. For instance, all those parameters that represent movements of the speech production mechanism, that are not easily seen by the naked eye, would be very useful.

Neural excitation patterns generated by acoustic-to-neural transduction.

Acoustic nerve excitation patterns elicited during acoustic stimulation are difficult to predict. A great deal of basic research has been directed to describe and understand this transduction process (Kiang, 1965; Sach, et al, 1968, 1974; Evans, et al, 1971, 1972, 1974, 1975; Rose, 1973; Geisler, et al, 1974; Goblick and Pfeiffer, 1969; Pfeiffer and Kim, 1975; Smith and Zwislocki, 1975).

Eighth-nerve fiber responses to speech stimuli were recorded by Kiang and Moxon (1972). This is, perhaps, one of the more relevant papers because speech stimuli were used to elicit neural responses. They observed that eighth-nerve fibers with similar characteristic frequencies had similar response patterns. The major difference between responses was the difference in each fiber's spontaneous rate. When the same speech stimulus was filtered by a 1/3 octave bandpass filter, centered at the fiber's characteristic frequency (CF), the envelope of the filter's output was similar to the envelope of the poststimulus time histogram of the fiber. With speech stimuli, at low stimulus levels, the gross patterns of responses in auditory nerve fibers are determined primarily by the tuning properties of these fibers and the spectral properties of the speech stimulus. Responses of fibers, with different CF, differ considerably at the lower stimulus levels; but are more alike at higher stimulus levels. At higher stimulus levels, spectral components to which the eighth-nerve fibers are less

sensitive can become sufficiently large to elicit responses. The tuning curves of eighth-nerve fibers are generally very selective for frequencies near the CF at stimulus levels from threshold to 30-40 dB above threshold. At higher stimulus levels, the tuning curve's low frequency slope becomes considerably less steep. Pfeiffer and Kim (1975) report that spatial excitation spreads markedly at high stimulus levels for single tone stimuli. The broadening of the tuning curve at higher stimulus levels, combined with response behavior similar to that of an automatic gain control (Seibert, 1973), may explain why fibers with different CF's respond more alike as the speech stimulus level is increased.

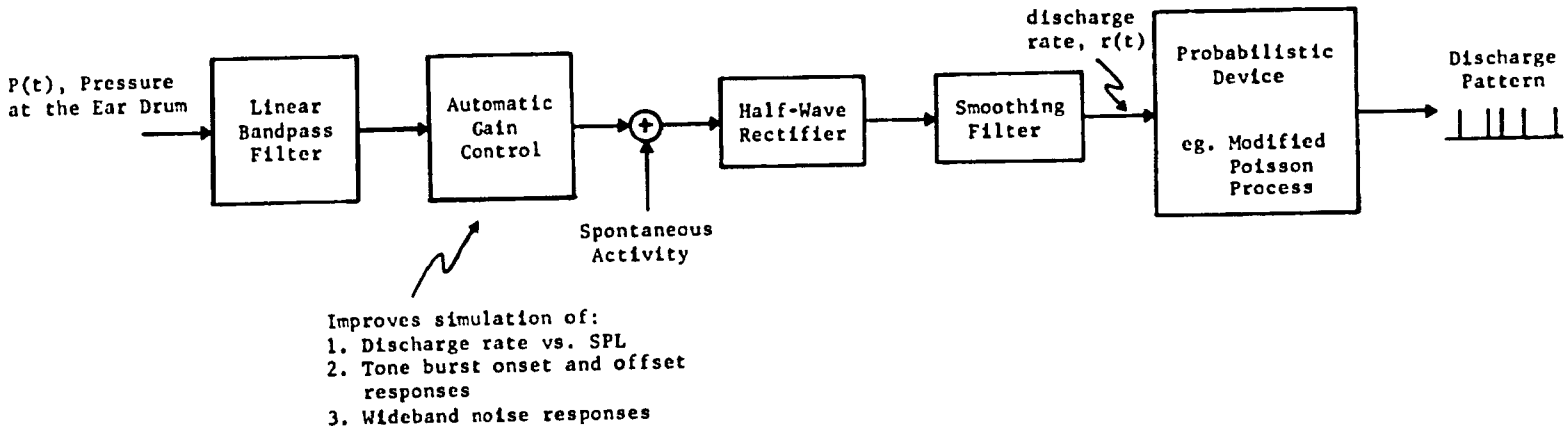
A number of acoustic-waveform-to-eighth-nerve-activity models have been proposed (Siebert, 1973, 1968; Weiss, 1966; Geisler, 1968; Pfeiffer, 1970; Schroeder and Hall, 1974). None of these models were conceived to simulate all of the phenomena observed at the eighth nerve. However, the models help to summarize many of the more significant phenomena.

Figure 3.15 illustrates a phenomenological model by Siebert (1973). This model simulates many of the significant aspects of the acoustic-to-neural transduction for steady sine waves and clicks over limited intensity ranges. The frequency response of the linear narrow-band filter, represents the typical tuning curve of individual eighth-nerve fibers. The smoothing filter has a short time constant (approximately -3 dB at 200-500 Hz; -6 dB/octave) so that, combined with the rectifier, it acts as an envelope detector at high frequencies

Figure 3.15. Siebert's simplified model for auditory firing patterns (after Siebert, 1973).

One version of Siebert's
single fiber model

(best for SPL's less than 70dB)



but follows the details of the rectified waveform at lower frequencies. The rather high cutoff frequency of the smoothing filter indicates that voicing information and other speech information is available in the temporal domain at the eighth-nerve level. Frequency components as high as 4-5 Khz can be detected in the poststimulus time histograms of eighth-nerve fibers. Synaptic mechanisms may set this limit. Evans (1977) and Moller (1977) speculate that this high frequency temporal information might be used to resolve complex spectral information. Fibers may be preferentially "phase-locked" to the spectral peak(s) within the fiber's passband. Thus, fibers with different CF's would contain temporal information on different spectral peaks. The cochlea's filtering action would enable the fibers to carry this high frequency temporal information with little masking from spectral components outside the fiber's spectral domain.

The probabilistic device could be a Poisson generator modified to account for the neuron's refractory period or a model neuron such as those proposed by Schroeder and Hall (1974), Geisler (1966, 1968), and Weiss (1966).

By inserting an automatic gain control at the output of the narrow-band filter, a wider range of stimulus intensities and other classes of stimuli, such as tone bursts and wide-band noise, are more accurately simulated.

The Siebert model does not simulate non-linear interactions between the responses to two click stimuli spaced several milliseconds apart (Goblick and Pfeiffer, 1969). It seems as though the second

click stimulus becomes less effective as the delay between the clicks decreases, qualitatively as if an automatic gain control operated on the input to the narrow-band filter (Siebert, 1973).

The Siebert model does not simulate the phenomenon of two-tone suppression of single cochlear-nerve fibers (Sachs and Kaing, 1968). Pfeiffer (1970) proposed a model for this phenomenon. The proposed model is similar to Siebert's simplified model. The narrow-band bandpass filter is replaced with two linear bandpass filters in cascade; with a memoryless non-linearity inserted between these two bandpass filters. The non-linear function is $f(x) = f\ x^{1/3}$. Responses of the model are similar to those of eighth-nerve to two-tone stimulation. Certain combinations of inputs suppress the output. The frequency and stimulus level combinations that produce suppression are strikingly similar to those frequencies and stimulus levels that produce suppression in eighth-nerve responses (Green, 1976). A $2f_1-f_2$ major cross-product term and an impulse response, that has similarities to measured click responses, are presented as further evidence for the model's merits. Another proposed phenomenological model incorporates a gain control after Siebert's narrow-band bandpass filter. The gain control receives gain-controlling inputs from spectral magnitude extractors tuned to the suppressive bands.

Siebert's model does not attempt to model non-linear combination tone responses (Goldstein and Kiang, 1968). The non-linear summation of two or more tones generates intermodulation distortion products. Some of these spectral components appear to be propagated in the

cochlea in a manner similar to that of normal acoustic stimuli. The distortion products, that are propagated, appear to elicit eighth-nerve responses similar to those generated by normal acoustic stimulation.

These non-linear phenomena appear to complicate phenomenological models of eighth-nerve responses. However, even responses to complex stimuli, such as speech, can be approximated by Siebert's model (Kiang, 1972). Indeed, many of the response properties are well simulated by Siebert's model or other models similar to Siebert's.

Perceptually important features of eighth-nerve firing patterns

How are parameters of the acoustic stimulus represented in the firing patterns of the eighth nerve? For example, how is stimulus intensity or the frequency spectrum of the stimulus represented in the firing patterns?

In the case of a single tone, fiber discharge rates saturate or nearly saturate (Sach and Abbas, 1974) at moderate stimulus intensities, in those fibers originating in the region of maximum sensitivity to the stimulus tone. At lower stimulus levels, discharge rates, in the region of maximum sensitivity to the stimulus tone, may convey a considerable amount of information. Most theories, dealing with single tone frequency and intensity discrimination, stress the importance of changes in the edges of the excitation pattern during moderate and high level stimulation (Green, 1976). Excitation pattern edges may be the most sensitive cues to changes in the single tone stimulus, at moderate and high stimulus levels. The excitation pattern is generally represented by a two-dimensional ploy of discharge rate versus basilar membrane "place". As the tone's intensity is increased, the edges of the excitation pattern would tend to move in opposite directions; thereby indicating an increase in the number of fibers excited. As the tone's frequency is increased, the edges of the excitation pattern would tend to move in the same direction (i.e., toward the base).

How are stimulus amplitudes and frequencies conveyed when more

complex stimuli are presented? For example, it is not clear what excitation pattern features are used to detect small changes in white noise amplitude over a range of 100 dB. Weber's law (dI/I=k) is applicable, except near the threshold of hearing. Only a 2% change in amplitude is necessary for detection of the change (Miller, 1947). As another example, a sine wave is presented in a noise containing a spectral notch such that on either side of the signal's frequency there is more noise energy than signal energy. Changes in the sinusoidal signal's amplitude are easily detectable.

Over wide dynamic ranges, it is not clear what excitation pattern cues might be used to resolve numerous spectral components in a spectrally complex stimulus. Non-linear effects such as two-tone suppression may be illustrative of effects similar to classic "lateral inhibition". Lateral inhibition might tend to bring fibers out of their saturated region and into their active region (Ruggero, 1973). However, there is considerable question as to whether such suppressive effects are substantial enough to account for the observed frequency resolution over wide dynamic ranges (Evans, 1977).

Evans lists several ways in which frequency resolution may be obtained over such a wide dynamic range: "It is possible that these cells (i.e., eighth-nerve fibers) are sensitive to differences in cochlear fiber discharge rate, too small to be detected by single fiber recording. Alternatively, it is possible that these cells receive their input from the fibers reported (Nomoto, et al, 1964;

Sachs and Abbas, 1974) to have a dynamic range in excess of 60 dB. Another possibility is that the cochlear fibers transmit their "place" information by means of the fine time structure of their discharge patterns. For signals up to 5KHz, the range to "phase-locking" extends from below discharge rate threshold to above saturation (Rose, et al, 1971). We do not know, however, whether the signal level can be coded in these terms over this wide range."

In reference to a model very similar to Siebert's model, DeBoer (1973) states: "Each nerve fiber transmits information that is specific to it and there is little interaction. The specificity of coding is strongly related to the spectrum of the stimulus. As a matter of fact, the "principle of specific coding" is an approximation, but it can in this purified and abstract form serve well to provide a basis for understanding auditory processing. It is to be noted that the principle of specific coding does not only cover the fact that the activity of each nerve fiber (i.e., the rate of its action potentials) is determined by the input spectrum filtered by the appropriate filter. It includes the property that the timing of action potentials is governed by the waveform of the filtered stimulus as well." It also includes the mutual correlation and fine-grain temporal differences between activities of neighboring fibers. This latter property is particularly well illustrated by a three-dimensional cochlear excitation pattern where "longitudinal fiber location" and "time" are the independent variables, and "firing rate" or "time of action potential occurance" is the dependent variable (see Figure 11; DeBoer, 1973).

If the fine time structure of neural firing patterns is perceptually significant, electrical stimuli should contain components designed to convey this information. As Evans and DeBoer suggest, this "phase-locking" information might be used by the nervous system to help determine the spectral content of the stimulus.

If the fine-grain temporal differences between activities of closely neighboring fibers is perceptually significant in speech perception, precisely controlled electrical field interactions combined with accurate control of stimulus timing and stimulus amplitude parameters might be essential. Accurate non-invasive excitation mapping techniques combined with spatial and temporal excitation pattern models would be invaluable in the proper implementation of such a processing goal.

Psychophysical data relevant to electrode placement considerations

Electrode contact separations may be an important parameter in a cochlear prosthetic device. For example, if electrode contacts are too far apart, transitions in formant frequencies may not be as easily conveyed to the subject. This might be inferred from the studies of Divenyi and Hirsch (1974) and Bregman and Cambell (1971). Using rapid sequences of tones, these investigators discovered that discriminations were better if all tones in the sequence were within a 1 - 2 critical band frequency range. It should be noted that these studies did not use speech as a stimulus; therefore the conclusions may not be indicative of the nervous system's responses to speech stimuli.

Other questions arise relative to electrode placement. For example, should the electrodes be placed only at "places" normally stimulated by speech, or should some electrodes be placed more toward the base? Spreading the electrode array across the entire cochlea could allow differential access to more functioning nerve fibers. On the other hand, speech recognition might be severly hampered if the central nervous system cannot adapt to this change.

II. The "non-mimicking" approach or "information theory" approach.

The proponents of the non-mimicking approach argue that it may be more reasonable to derive perceptually important information from the acoustic speech signal and deliver it employing the best information bearing modes of stimulation without attempting to mimick normal peripheral neural responses.

Rationale for the "non-mimicking" approach.

One argument for such an approach is from purely practical grounds. The time for patient testing may be severely limited and thus not allow the investigators sufficient time to mimick the normal eighth-nerve excitation patterns. Indeed, given any amount of time, it may not be possible to sufficiently mimick the perceptually significant features of the normal ear's eighth-nerve discharge patterns. The non-mimicking approach may offer a relatively efficient and straightforward testing and design protocol.

As the characteristics of acoustic stimuli are adjusted to demonstrate greater apparent similarity to speech, there is a deterioration of performance during learning, except when the stimuli are actual speech signals (House, et al, 1962). In their study, stimuli, seemingly very similar to speech, were not learned as quickly as those stimuli which had relatively clear-cut differences along the "natural" or "primitive" dimensions (e.g. loudness, pitch, and duration). Therefore, one might contend, that unless a prosthesis is able to almost perfectly mimick the perceptually significant features of the excitation patterns elicited

during speech, it might be wiser to use a different approach. Results show that performance during learning is better when each stimulus is encoded into several physical dimensions than when the stimuli lie along a unidimensional continum[1] (House, et al, 1962).

Furthermore, it appears that the human is able to adapt to many alterations in the speech waveform (Licklider, 1946) and speech spectrum. For example, speech perception can be relearned even if the speech spectrum has been inverted (Blesser, 1969). Speech recognition is easily accomplished over a wide range of speakers; even though the formant frequencies vary widely across speakers for the same word (Ladefoged and Broadbent, 1957; Broad and Shoup, 1975). People adapt to speakers with strong accents; and, with further practice, people can learn foreign languages.

However, with the exception of the Blesser experiment, the same type of speech production mechanism is involved in these examples. Consequently, certain features of the acoustic speech signal and eight-nerve discharge pattern are invariant in these examples. Even, in the Blesser experiment, one could argue that some features in the eighth-nerve discharge pattern are similar to those patterns generated by the human's speech production apparatus.

---

1. Multidimensional scaling analysis techniques might be used to advantage in the study of implanted subjects. These techniques allow the investigator to determine: (1) in which and in how many dimensional spaces stimuli are perceived, (2) whether all dimensions contributed equally to the perception of stimuli or whether some dimensions contributed more to perception than do others, and (3) whether there are differences in individual subjects' ranking of perceptual parameters (Danhauer and Singh, 1975).

Proposed test and design protocol for the "non-mimicking" approach

In the non-mimicking approach, one must derive perceptually important information from the acoustic speech signal and deliver it employing the best information-bearing modes of stimulation. The speech processing system can be conceptualized as a two stage process (Figure 3.16). The "speech analysis stage" would extract perceptually important information from the speech signal; and the "stimulus processing stage" would deliver this perceptually important information to the nervous system. The stimulus processing stage would employ the best information-bearing modes of stimulation in delivering this information to the nervous system.

The stimulus processing stage may be conceptualized as a mapping process. That is, the stimulus processing stage would map the speech analysis stage's outputs into the currents flowing through the stimulating electrodes.

The best information-bearing modes of stimulation could be determined, in part, by differential discrimination tests of changing (i.e., articulatory rate changes) stimulus amplitude, stimulus waveform, stimulus frequency, and changing spatial excitation pattern[1] (Schubert and Walsh, 1976). Difference limens could be determined over the useful range of parameter transition rates. These difference limens would constitute an initial estimate of the relative information-bearing capacity of the various modes of stimulation. Discrimination testing should span the entire range of transition rates, where potentially useful differential

discriminations are demonstrated by the subject. Zero and near zero transition rates should be tested in order to estimate the subject's "static" stimulus discrimination ability.[2]

Similar tests using two or more channels[3] could be used to determine the effects of other channels on the discrimination of stimulus differences occurring only in one channel. Also, discrimination of stimulus differences occurring in both of two stimulated channels could be investigated.

The differential discrimination tests (and possibly absolute discrimination tests of simple one or more channel, stimulus patterns) could be used to design the initial stimulus stage.

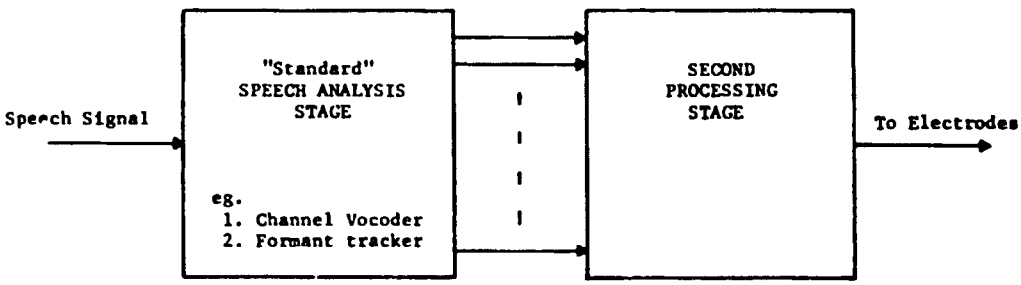
The reductions in information rate afforded by standard speech analysis stages is well documented (Flanagan, 1972; Rabiner and

---

1. Changes in spatial excitation patterns, generated by one electrode channel[3], are elicited by changing one or more of the following stimulus parameters:
   a) Total stimulus amplitude.
   b) Certain features of the stimulus waveform. For example, changes in biphasic pulse width change the spatial excitation pattern.
   c) Relative amplitudes of the currents generated at the channel's multiple electrode poles. This allows one to manipulate electrical field interactions, the positions of current sources and sinks, and the temporal summation of eighth-nerve membrane charge.
2. Cole and Scott (1974) propose essentially three types of cues to speech. One set is essentially invariant, another transitional, and the envelope of the speech waveform helps to integrate syllables into higher order units such as words and phrases.
3. An electrode channel can be composed of two or more electrode poles. Controlling currents at these multiple poles is used to change the spatial and temporal excitation pattern. Spatial patterns generated by a given channel are controlled as outlined in footnote 1 of this page.

Figure 3.16. Block diagram of a generalized speech processing system.

The processing task is divided between a "speech analysis stage", which extracts perceptually significant parameters from speech, and a "stimulus processing stage", which employs the best information-bearing modes of stimulation in delivering this parametric information to the nervous system.


Speech Signal
"Standard"
SPEECH ANALYSIS
STAGE
eg.
1. Channel Vocoder
2. Formant tracker
SECOND
PROCESSING
STAGE
To Electrodes

Schafer, 1977; Markel and Gray, 1977). Most analysis stage design parameters have been well investigated and might require only minor changes from their initial settings. Analysis stage bandwidth and dynamic range requirements for speech transmission have been investigated. By comparing these requirements with the person's discrimination abilities one could design the initial stimulus processing stage and choose a close to conventional speech analysis stage.

Because of the ease of implementation, and relatively good bandwidth performance, a slightly modified voice excited channel vocoder analysis stage may be a good choice. Also, the channel vocoder divides the speech spectrum in a manner similar to that dictated by the mimicking approach, and in a manner similar to the longitudinal electrode positions. Thus, this choice has some appeal relative to the mimicking approach.

Also, peak-picking and formant tracking analysis stages with baseband filters to convey voicing information are good choices. Other speech parameters (e.g. see the list of parameters in the section titled: "perceptually significant features of the acoustic speech signal") may be useful in conveying speech information that is less readily available to the patient. These parameters could be extracted and included in the analysis stage's set of outputs.

After the initial processor design, speech test material (e.g. the DRT, Voiers, 1973; the SNST, Nye and Gaitenby, 1974) would be processed and used for patient testing. For example, the

Diagnostic Rhyme Test (DRT) categorizes the subject's confusions in differentiating between two words. The test categorizes these confusions according to the acoustic feature(s) nomally used to differentiate between the two words.

Plots of the speech analysis stage's output signals and stimulus parameters versus time, for each speech segment, should be available for interpretation of test results. These interpretations would guide additional single and multichannel tests (as previously described). These tests would use stimulus patterns similar to the features of speech that were confused. The interpretations from the tests could be used to refine both the speech analysis and stimulus processing stages.

It might require a considerable amount of subject learning to utilize such a processing system. In this laboratory, some of the testing material has been designed to measure the rate of learning.

Appropriate combinations of the two approaches (i.e., the mimicking and information theory approaches) may prove most fruitful. An investigation of a judiciously chosen set of models should offer some insight into which aspects of coding are relatively unflexible and which aspects are significantly influenced by experience.

Summary of some possible approaches to stimulus processing

Mimicking approaches

The "mimicking approach" is a general title for those processing techniques which are designed to reproduce certain feature(s) of normal hearing function.

For example, one might attempt to replicate the eighth-nerve excitation patterns. One could attempt to replicate those neural patterns generated by the normal acoustic speech waveform; or one could attempt to replicate those excitation patterns elicited by acoustic stimuli which are generated by an analysis-synthesis vocoder, such as a spectrum channel vocoder or a peak-picking vocoder.

One might attempt to replicate only the perceptually significant features of normal excitation patterns. The "perceptually significant features" may be estimated by applying eighth-nerve coding theories, the literature relating to distinctive features of the acoustic speech signal, and information relating to the acoustic-to-neural transduction process. In addition, psychophysical testing techniques, applied to implanted subjects to determine the perceptual significance of electrically generated excitation patterns, would be an invaluable tool in such a design.

One might attempt to replicate the percepts of a normal hearing individual. This version of the mimicking approach would involve the matching to those percepts generated by acoustic stimulation in the normal ear with those percepts generated by electrical stimulation. This version would involve the development of a psychophysics of

eighth-nerve electrical stimulation. To some degree, this psychophysical data could be supplemented with the extensive body of information developed in the psychophysics of acoustic stimulation.

Non-mimicking approach

In the "non-mimicking approach" one would derive perceptually important information from the acoustic speech signal and deliver it, employing the best information-bearing modes of stimulation without attempting to mimick normal peripheral neural responses.

Appropriate combinations of these approaches may prove most fruitful. A thorough patient testing and evaluation program should be invaluable in the design and evaluation of such a prosthesis.

For number sequence only.

Chapter 4

Speech and Stimulus Processing Software

This chapter contains a summary of a modular software system developed to efficiently implement the various stimulus processing options to be tested. The use of this software is illustrated with the implementation of an analysis-synthesis spectrum channel vocoder. Such a vocoder is an excellent example because its analysis stage is one of the speech processing stages proposed for an initial stage of the speech processor. Also, within this vocoder, filter banks, automatic gain controls, spectral flatteners, and envelope detectors are illustrated. In addition, the illustrated channel vocoder may be useful in simulating certain aspects of the implanted person's perceptions.

Formant tracking software has been obtained from several sources (Markel and Gray, 1977; Robinson, 1968). This formant tracking software has been well documented (also see: Rabiner and Schafer, 1977) and consequently is not described in this document.

Vocoder implementation

A number of computer programs and subroutines have been written to implement a spectrum channel vocoder. Various versions of the spectrum channel vocoder can be implemented with the same basic programs. These versions are: 1) A standard channel vocoder which contains a fundamental frequency (Fo) extractor and a voice/unvoiced (V/UV) decision making device (see Figure 4.1). 2) A voice excited channel vocoder. In this device, excitation information is transmitted in a subband of the original speech (see Figure 4.2). 3) A modified version of the standard channel vocoder in which the excitation component is always noise. According to Rabiner (1977), this vocoder generates whisper-like speech. 4) A modified version in which the excitation component is always periodic, in which Fo is experimentally controlled. The dashed lines in figure 4.2 indicate the modifications necessary to implement versions 3 and 4.

Also, the software implementation incorporates the following features:

1) Channel bandwidths may be unequal as well as having unequally spaced center frequencies. Any number of channels may be specified.
2) Filter coefficients and other processing parameters are easily changed.
3) Each synthesis channel filter may be of differing bandwidth and center frequency than the corresponding analysis channel filter. This may allow the vocoder to simulate electrode placements at various basilar membrane locations other than those corresponding to the center

frequencies of the analysis channels.

4) The program may process long durations of continuous speech, only limited by secondary storage. However, the vocoder program can execute in a 24K word, 16 bit minicomputer.

5) Processing speed is very reasonable due to the efficient vector arithmetic, FFT, and convolution routines. These routines are written in assembler and utilize the hardware floating point processor. Vocoder execution times range from 20 to 100 times the duration of the speech segment. Execution time is a function of filter lengths, memory available, sampling rates, number of channels, and other factors.

6) Simulation of hardware analog processing systems should be relatively simple, because most applicable analog infinite impulse response filters are easily simulated using finite impulse response filters of sufficient impulse duration.

If a hardware implementation of zero delay distortion FIR filters is desired, hybrid techniques may be desirable. Charge coupled devices (CCD) may offer one solution. State-variable analog filters with Bessel-like filter characteristics offer another potential solution.

Software modules

Such routines as a finite impulse response (FIR) filter that down samples, an automatic gain control, an interpolating filter, an envelope detector, filter coefficient calculating routines, and a signal processing interpreter have been written. Table 4.1 contains a list and brief description of these higher level routines. These reentrant FORTRAN routines call simpler reentrant lower level routines

Figure 4.1. Block diagram of a standard spectrum channel vocoder.

Essentially, the spectrum channel vocoder divides the speech signal into two components. The vocoder is an application of the assumed source-system independence model.

The semantically important vocal tract configurational changes are represented by the slowly varying spectrum channel envelope amplitudes. The vocal tract filter is analogous to the "system" part of this model. Each analysis channel's output bandwidth need only be 25 Hz.

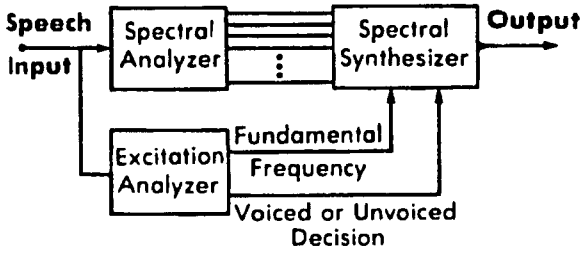
The excitation source's signal is useful semantically in determining a voiced or unvoiced state. For semantic purposes, the excitation signal can be rather crudely represented and reconstructed. Little bandwidth is required.

In the typical vocoder, 8 to 16 lowpass filtered, envelope signals and the voiced-unvoiced and fundamental frequency (Fo) signals are used to reconstruct the speech signal. During synthesis, the excitation signal is created from either a pulse generator, whose frequency is controlled by the Fo signal, or a random noise generator. The voiced-unvoiced signal controls a switch used to specify which source signal will excite a bank of bandpass filters. In some vocoder versions a mixer controls the relative proportion of noise and periodic excitation.

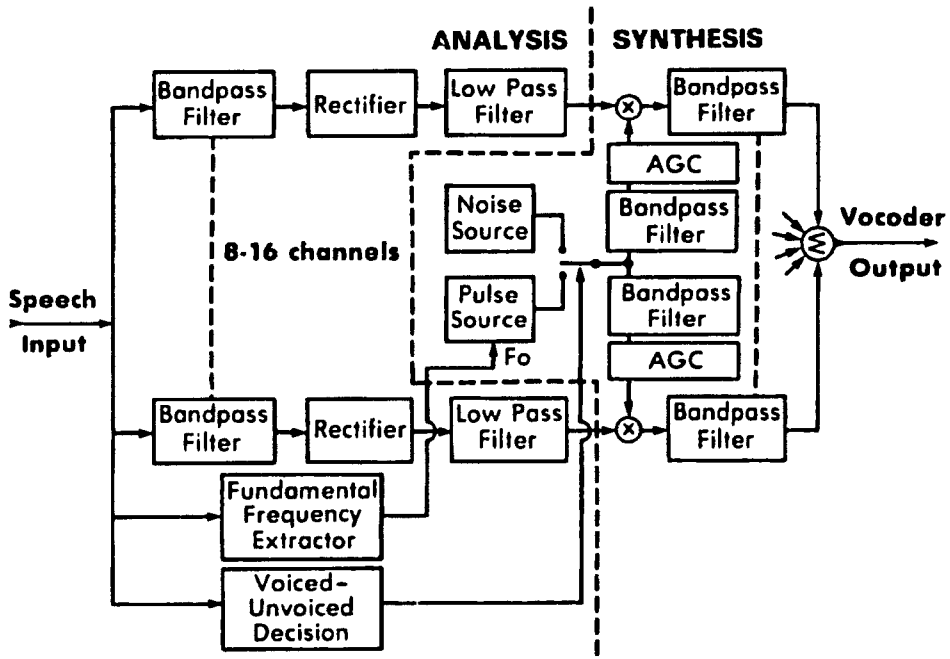
In more modern spectrum channel vocoders, the source signal's spectrum is grossly flattened. This technique eliminates slowly varying fluctuations in the simulated excitation signal's spectral components (Rabiner and Gold, 1975). During synthesis, each channel's bandpass-AGC combination performs the function of grossly flattening the excitation signal's spectrum.

The lowpassed spectrum envelope signals (from the spectral analysis stage) modulate the outputs of the bandpass-AGC spectral flattening stage. This modulation controls the energy in each of the frequency bands. To the first order, the modulation is analogous to the vocal tract's filtering of the source's excitation signal.

The synthetic speech signal is obtained by summing these modulated outputs.



SIMPLIFIED DIAGRAM



STANDARD CHANNEL VOCODER

Figure 4.2. Block diagram of a voice excited channel vocoder.

The voice excited channel vocoder is very similar to the standard channel vocoder. However, excitation information is extracted in a different manner.

A subband (i.e., baseband) of the speech signal is used to convey the voicing information. A subband from 100 to 500 Hz., or even up to 900 Hz., is used.

On synthesis, this baseband signal is spectrally expanded and flattened to derive a gross estimate of the glottal source signal. The baseband's periodicity, or the absence of it, is maintained in the spectral expansion and flattening process.

The lowpassed spectrum envelope signals (from the spectral analysis stage) modulate the outputs of the bandpass-AGC spectral flattening stage. This modulation controls the energy in each of the frequency bands. The modulation is analogous to the vocal tract's filtering of the source's excitation signal.

The synthetic speech signal is obtained by summing these modulated outputs.

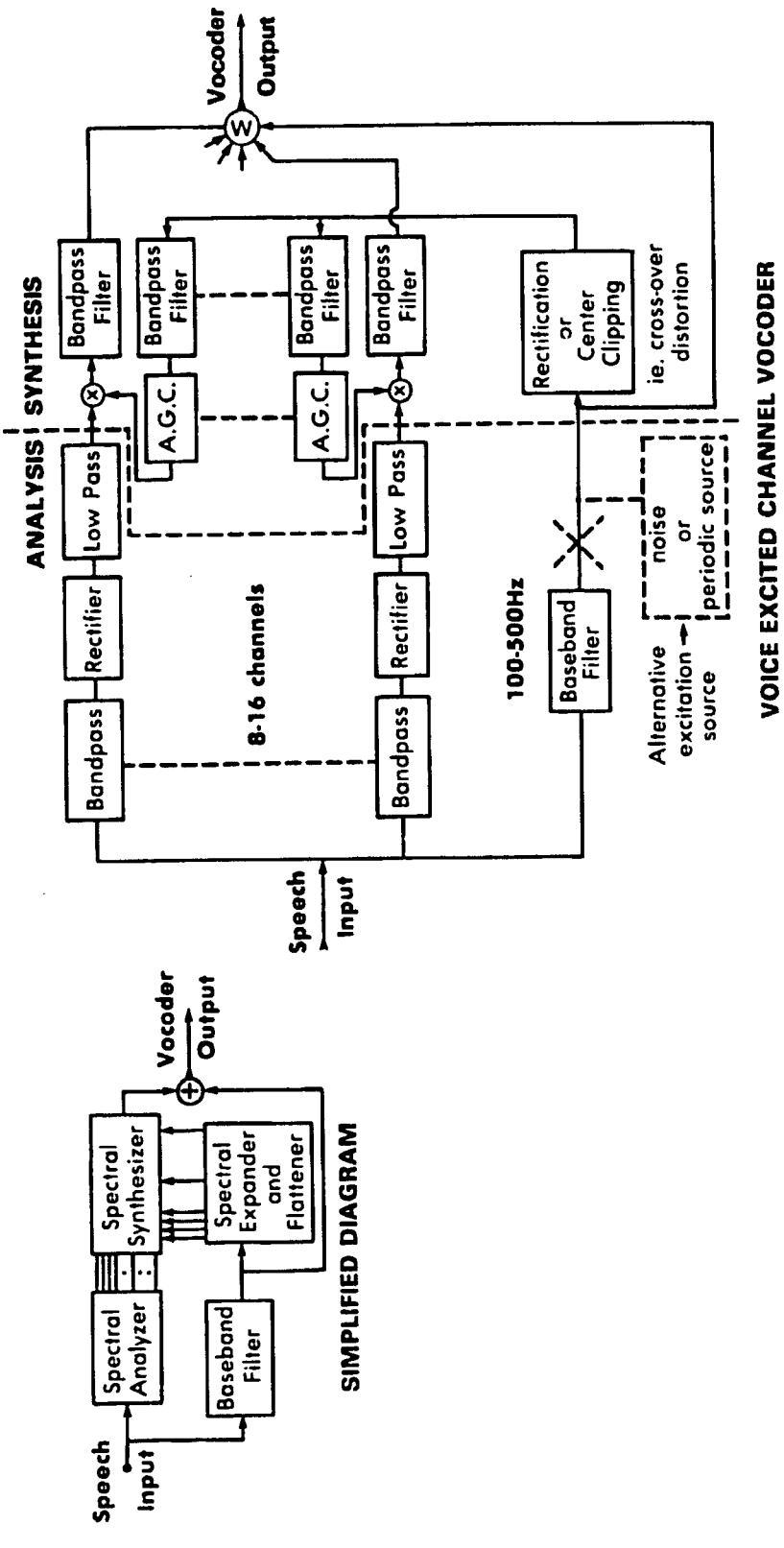

Vocoder
Output
ANALYSIS
SYNTHESIS
Bandpass Filter
A.G.C.
Low Pass
Rectifier
Bandpass
8-16 channels
100-300Hz
Baseband Filter
Rectification or Center Clipping
ie. cross-over distortion
Alternative excitation source
noise or periodic source
Speech Input
VOICE EXCITED CHANNEL VOCODER
Spectral Synthesizer
Spectral Expander and Flattener
Spectral Analyzer
SIMPLIFIED DIAGRAM

Table 4.1. A listing and brief description of the higher level speech and signal processing routines.

1.	AGC	This subroutine performs the function of an automatic gain control.
2.	ANADEC	This subroutine performs a magnitude estimate of a spectral segment.
3.	BASEBAND	This main program generates a vocoder baseband file, a spectrally expanded version of the baseband file, and a file that is a duplicate of the input file except for a compensatory delay. This delay compensates for the baseband filter's delay.
4.	CCLIP	This subroutine "center clips" a vector. For a segment of speech, the largest magnitude sample (referred to as MAX) is determined. All sample magnitudes below some specified fraction of this maximum (i.e. FRACTION* MAX) are set to zero. All other samples are reduced in magnitude by this same value, FRACTION*MAX. This routine simulates the effect of cross-over distortion and is used to spectrally expand and somewhat flatten a bandlimited time series. Generally, the input signal's periodicity, or lack of it, is maintained in the routine's generated output.
5.	CHANNEL	This routine performs the spectral channel analysis-synthesis operations required for a single channel of a spectrum channel vocoder. This subroutine simulates

Table 4.1. (continued)

various versions of standard or voice excited spectrum channel vocoders.

6. FILTCOEFF This program generates low pass filter coefficients for finite impulse response filter routines.

7. FIR This subroutine performs a convolution of two vectors. This routine can also "down-sample" (i.e. decimate) to greatly improve processing speed.

8. INTERPOLATE This routine performs the function of interpolation including the filtering required for such "up-sampling".

9. LPFILTER This routine is a finite impulse response filter routine with equal input and output sampling rates assumed. It uses a decimation-interpolation technique to improve processing speed (Rabiner and Schafer, 1977).

10. PREPROCESS This main program converts the raw speech files from fixed point to floating point format and then grossly flattens the speech spectrum by linear filtering.

11. SUBTEST This program is a simple interpreter which executes stored programs. It was written for the dual purpose of testing the various signal processing routines and of "breadboard-testing" various signal processing schemes.

12. TESTGEN This routine generates test signals for the purpose of testing the signal processing routines. Gaussian noise and amplitude modulated tones are examples of the test signals generated.

13. TRACE — This program plots a given vector located on a disk file.

14. VOCODER — This is the main calling program for performing spectrum vocoder analysis-synthesis of a speech signal.

15. VOCODERFILT — This routine efficiently generates the many filter coefficient files necessary for the vocoder simulation. A minimum of the experimenter's time is necessary for specifying the filter parameters.

(e.g. vector arithmetic routines, a convolution routine, a vector maximum routine). Many lower level routines are written in assembler and utilize the hardware floating point processor. In addition, each assembler routine has a counterpart written in FORTRAN that executes at approximately one sixtieth the speed. These FORTRAN routines are used to document, verify, and "back-up" their assembler counterparts. Table 4.2 contains a list and brief description of the lower level subroutines. These more basic routines perform the same function as the extremely fast hardware array processors presently on the market. Thus, if such an array processor is acquired, only trivial software modifications would be required before such a device could be utilized.

The signal processing modules have the following general features:

1) The routines are reentrant. If an algorithm is to be called upon from numerous parts of a program, only one copy of that routine need be stored in memory.

2) The routines can efficiently "piecewise process" an arbitrarily long time series. State-save scalars, vectors, and their stack pointers are utilized to implement this feature. One segment of speech at a time is processed. All information required for processing the next contiguous segment is stored in state-save vectors. These variables are stored on the run-time stack. Therefore, reentrance is preserved.

3) Input and output routines have been wirtten to sequentially access segments of time series vectors stored in random access format on disk.

4) The routines self-initialize. When the initialization flag is set,

Table 4.2. A list and brief description of the lower level speech and signal processing routines

1.	ASINE	This is a support routine for the SINE subroutine. This routine generates a vector of a sinusoidal time series. Two versions exist; one in ASSEMBLER and one in FORTRAN.
2.	BPCOEFF	This routine calculates the coefficients for a bandpass filter by using the windowing technique. The versatile Kaiser window function is used.
3.	CADD	Adds a constant to a vector. ASSEMBLER and FORTRAN versions.
4.	CLAMP	This routine modifies the values of elements of a vector if the values are above or below specified maximum and minimum values.
5.	CMULT	This routine multiplies a constant times a vector. Both ASSEMBLER and FORTRAN versions.
6.	CONTRACT	This routine takes the real part of a complex vector.
7.	DECIMATE	This routine performs the function of time series decimation. That is, the output vector contains only every "n"th element of the input vector.
8.	DELAY	This routine simulates a delay of an integral number of samples.
9.	ERFIX	An error reporting routine that allows for recovery by the operator.
10.	ERRPT	This routine reports an error without suspending

Table 4.2. (continued)

		normal execution of the program.
11.	EXPAND	This routine places real values in the proper location for discrete fourier transformation.
12.	FADR	This routine merely calculates and returns the absolute core address of the first argument in the FORTRAN call. Written in ASSEMBLER.
13.	FFT	Performs a fast fourier transform. This routine also does the necessary data formating when the input and output formats are either real or complex. The support library "DFT.LB" was written in ASSEMBLER by the manufacturer of our computer.
14.	FIRATIO	This routine determines the set of decimation ratios appropriate for a given bandpass filter. Any decimation ratio that would produce significant aliasing is rejected by this routine.
15.	FNAME	This subroutine converts an integer number into an ASCII character string, which may be used as a disk file name.
16.	FRMDISK	Specialized disk file routine for vector input.
17.	HALF	This routine simulates an ideal half wave rectifier.
18.	HAMCOEFF	This routine generates Hamming window coefficients of any specified vector length.
19.	HANNCOEFF	This routine generates Hanning window coefficients.
20.	HARD	This routine simulates a hard limiter.
21.	HCOMDENOM	Support subroutine for the vector input and output

Table 4.2. (continued)

		routines. This routine keeps track of the location of current data to be accessed and updates associated pointers.
22.	HICDIV	Determines the highest common denominator.
23.	INO	Calculates Bessel function values for the Kaiser window routine.
24.	INSERT	This routine allows the experimenter to view and alter the contents of a core-resident vector.
25.	ITOR	Converts a 16-bit, fixed-point vector to floating-point, 32-bit format.
26.	KAISER	This routine generates a Kaiser window of a specified duration and frequency transition. Also, the stop band ripple is calculated and passed back to the calling routine.
27.	LOGV	This routine takes the log of a vector.
28.	LP	This routine generates samples of an ideal low pass filter's impulse response.
29.	MAXV	Finds the maximum element magnitude in a vector.
30.	MSINE	This routine multiplies an input vector by a sinusoidal time series.
31.	MULSUM	This is an extremely fast vector routine for the purpose of calculating time series correlations and convolutions. ASSEMBLER and FORTRAN versions have been written.
32.	MWRD	This routine supports disk file reading.

Table 4.2. (continued)

33.	MWWRT	This routine supports disk file writing.
34.	NUMSPLIT	This routine splits a floating point number into its mantissa and exponent.
35.	POWER	Generates a power vector from a complex input vector.
36.	PRNT	This routine prints a header, the scale factor, the number of data points in the vector, and then prints the vector's element values, five values to a line.
37.	RAND	Generates a single random number in integer format. Both ASSEMBLER and FORTRAN versions.
38.	RMOVE	This does a vector move starting at the "end" of each vector and progressing downward in core. This routine allows "right shifts" of data within one vector. ASSEMBLER and FORTRAN versions.
39.	RTOI	Converts a 32-bit floating point vector to a 16-bit integer format vector.
40.	SCOPE	This routine plots a core-resident vector on the plotter.
41.	SETGAIN	This routine allows the user to set any linear filter's gain at a specified frequency.
42.	SINE	Generates a vector of sinusoidal values.
43.	STIMFILTER	A program that filters evoked response data. The user can specify data tapering and editing before filtering. The filter is a finite impulse response filter with coefficients calculated by multiplying a Kaiser window by the impulse response of an "ideal" bandpass filter.

Table 4.2. (continued)

44.	TAPER	This routine multiples a vector by a cosine shaped window. The function modifies 10 percent of the data at both ends of the vector.
45.	TODISK	Specialized vector-to-disk output routine.
46.	VABS	This routine finds the absolute value of each element of vector. Both ASSEMBLER and FORTRAN versions.
47.	VADD	Routine adds two vectors. ASSEMBLER and FORTRAN versions.
48.	VDIV	Division of two vectors. ASSEMBLER and FORTRAN versions.
49.	VGAUSS	This routine generates a vector of normally distributed random numbers. Uses the central limit theorem technique to convert a uniform distribution to a normal distribution.
50.	VMAG	Generates a magnitude vector from a complex input vector.
51.	VMOVE	Vector move routine. Both ASSEMBLER and FORTRAN versions.
52.	VMULT	Vector multiply routine. Both ASSEMBLER and FORTRAN versions.
53.	VRAND	This routine generates a vector of uniformly distributed random numbers.
54.	VSQRT	This routine finds the square root of each element of a vector.
55.	VSUB	Vector subtraction routine.
56.	ZEROV	Zeros a vector. Both ASSEMBLER and FORTRAN versions.

all variables that require initialization are initialized. Initialized variables are stored on the run-time stack.. Therefore, reentrance is preserved.

5) Some of the higher level FORTRAN programs dynamically allocate storage. At execution, the programs utilize all the available storage. Thus, the program can execute with very little memory at the expense of processing speed; or it can execute at a higher speed when more memory is available.

6) 32-bit floating point arithmetic is used. This floating point representation greatly reduces the "noise" generated during processing, especially when processing is recursive (Rabiner and Gold, 1975). Inadvertent 16-bit fixed point overflow or underflow is eliminated. Programming is also simplified because scaling techniques required for fixed point or block floating point format are not necessary. The price paid is the doubling of intermediate storage requirements. Because of the hardware floating point processor, little loss in speed is incurred.

7) The hierarchial subroutine system uses only a few common routines (e.g. vector arithmetic and FFT routines) that do the bulk of the time consuming processing. These often-used routines are the only routines that required optimization.

Program organization

The voice excited vocoder configuration diagramed in figure 4.2 was implemented in software. The diagram in figure 4.3, as well as detailed block diagram "blow-ups" in figures 4.4, 4.5, and 4.6,

Figure 4.3. Block diagram of a software simulation of a channel vocoder.

Recorded test material is digitized by a 12-bit analog to digital converter. This fixed point representation is converted to floating point format and pre-emphasized to approximately flatten the signal's gross spectral content (Markel and Gray, 1976). The program named PREPROCESS is responsible for this processing. Program BASEBAND (also see Figure 3.6) generates three outputs from this preprocessed signal: 1) A lowpass filtered "subband" or "baseband" of the speech signal, containing speaker excitation information. 2) The undistorted speech signal appropriately delayed to compensate for the baseband filter's delay. 3) An output generated by spectrally expanding the baseband signal. The baseband's periodicity, or lack of it, is maintained in the generated output. This periodicity information is useful. For example, if the signal is periodic, voicing is present.

Subroutine CHANNEL uses outputs "2" and "3" to synthesize the vocoder channel signals that are summed with the baseband signal "1" to form the composite output of the channel vocoder (see Figure 4.4).

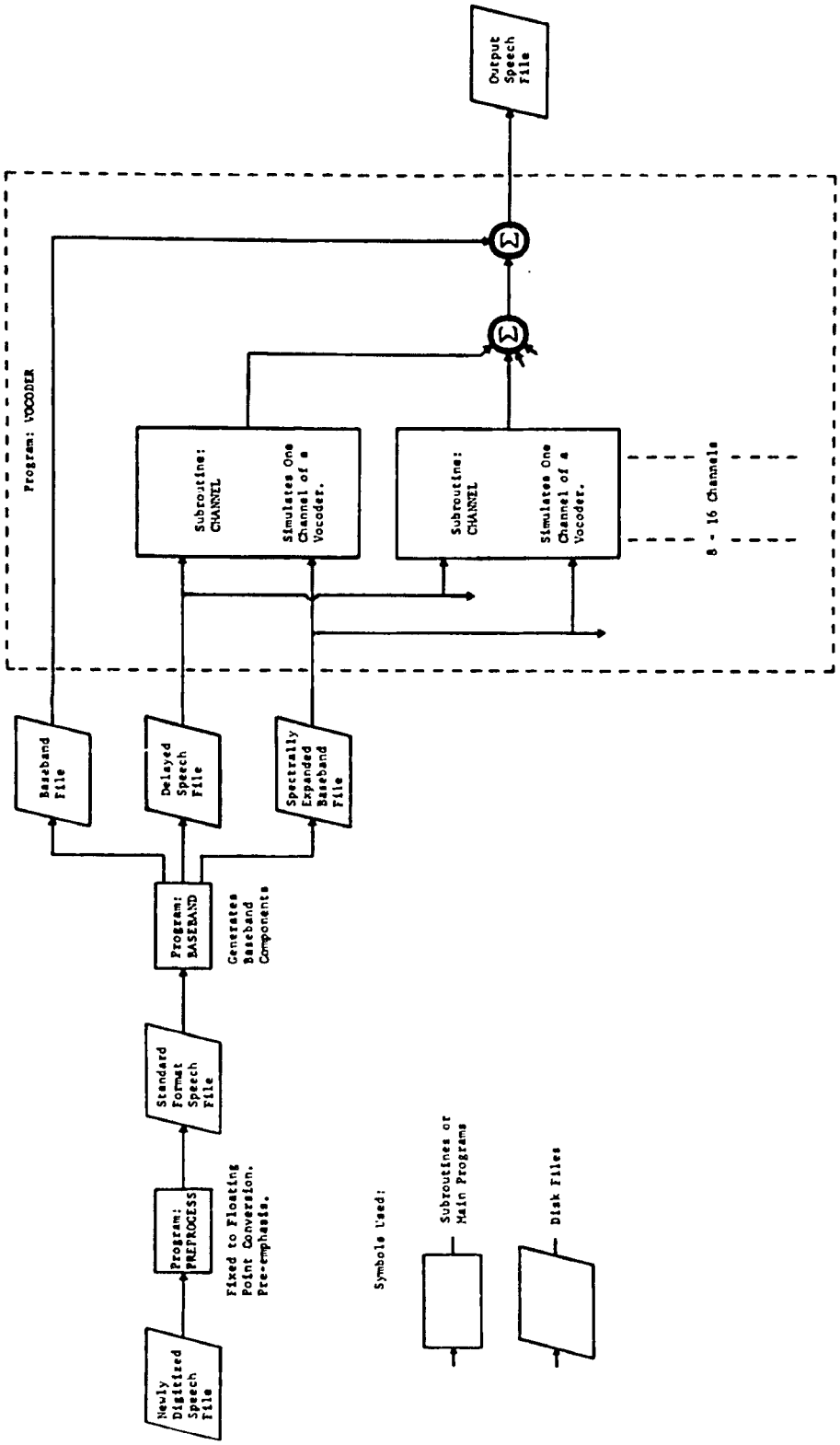

Newly Digitized Speech File
Program: PREPROCESS
Fixed to Floating Point Conversion, Pre-emphasis.
Standard Format Speech File
Program: BASEBAND
Generates Baseband Components
Baseband File
Delayed Speech File
Spectrally Expanded Baseband File
Program: VOCODER
Subroutine: CHANNEL
Simulates One Channel of a Vocoder.
Subroutine: CHANNEL
Simulates One Channel of a Vocoder.
8 - 16 Channels
Output Speech File
Symbols Used:
Subroutines or Main Programs
Disk Files

Figure 4.4. Block diagram of subroutine CHANNEL. This routine simulates one analysis-synthesis channel used in a spectrum channel vocoder realization.

Subroutine ANADEC generates an on-going magnitude estimate (i.e., envelope detection with envelope frequencies up to 25Hz.) of a spectral segment of the speech signal. Only the slowly varying component of the envelope is important to semantic content. ANADEC's output represents, to a first order, the effects of the vocal tract's time varying, filtering function on the excitation source's signal. Of lesser semantic importance, this subroutine's output also contains information regarding the excitation source's slowly varying properties.

During synthesis, subroutine ANADEC's output is used to modulate the appropriate spectral segment of the grossly estimated excitation signal. The resultant output, once interpolated, is the CHANNEL routine's output.

The excitation signal is approximated by spectrally expanding the excitation-rich baseband signal. This is one of the functions of program BASEBAND. The baseband's periodicity, or lack of it, is maintained in the expanded baseband version. This expanded spectrum is then flattened by a bank of bandpass filters andautomatic gain controls, in which each filter is in series with an automatic gain control. There is one bandpass-AGC combination in each channel. This spectral flattening does not destroy the periodicity information (i.e. voiced/unvoiced and glottal vibratory rate information), because the automatic gain control does not reduce amplitude variations occurring faster than 50-75 Hz. However, the AGC does greatly limit amplitude variations slower than 25 Hz. This insures that the effect of ANADEC's slowly varying modulation will not be corrupted by slowly varying components derived from the expanded baseband signal. All slowly varying information is carried in ANADEC's outputs. Some of this slowly varying information would be doubly represented if it were not for this spectral flattening (Rabiner and Gold, 1975).

Subroutine CHANNEL

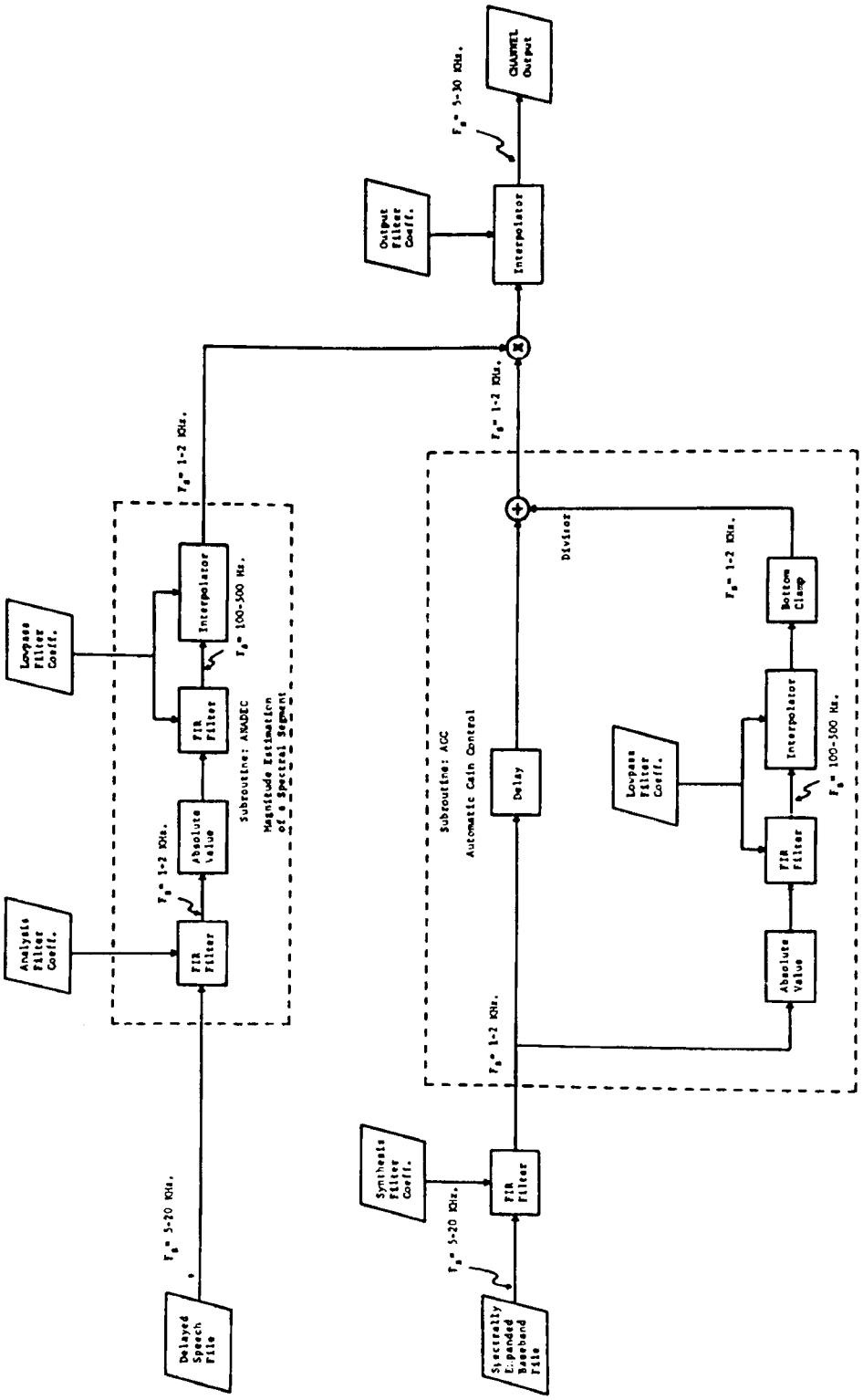


Figure 4.5. Block diagram of program PREPROCESS.

This program converts newly digitized speech material into 32-bit floating point format. The spectrum of the speech signal is grossly flattened by a linear filter. The filter equation is:
(present output sample value) = (present input sample value) - (.9)* (previous input sample value)
This initial filtering operation is commonly referred to as pre-emphasis filtering. Optionally the signal can be passed through an automatic gain control.

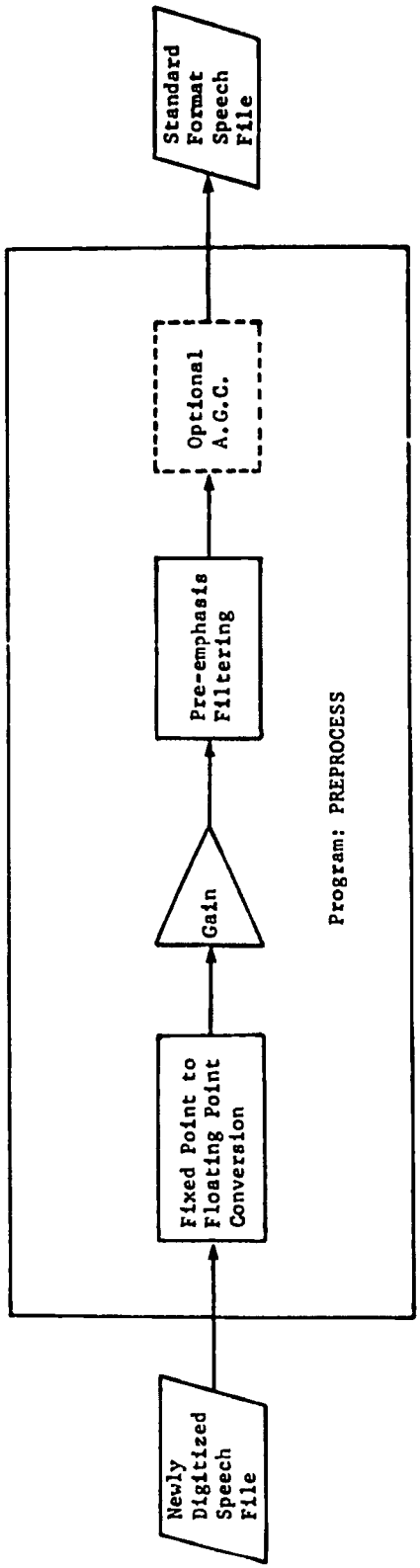

Newly Digitized Speech File
Fixed Point to Floating Point Conversion
Gain
Pre-emphasis Filtering
Optional A.G.C.
Standard Format Speech File
Program: PREPROCESS

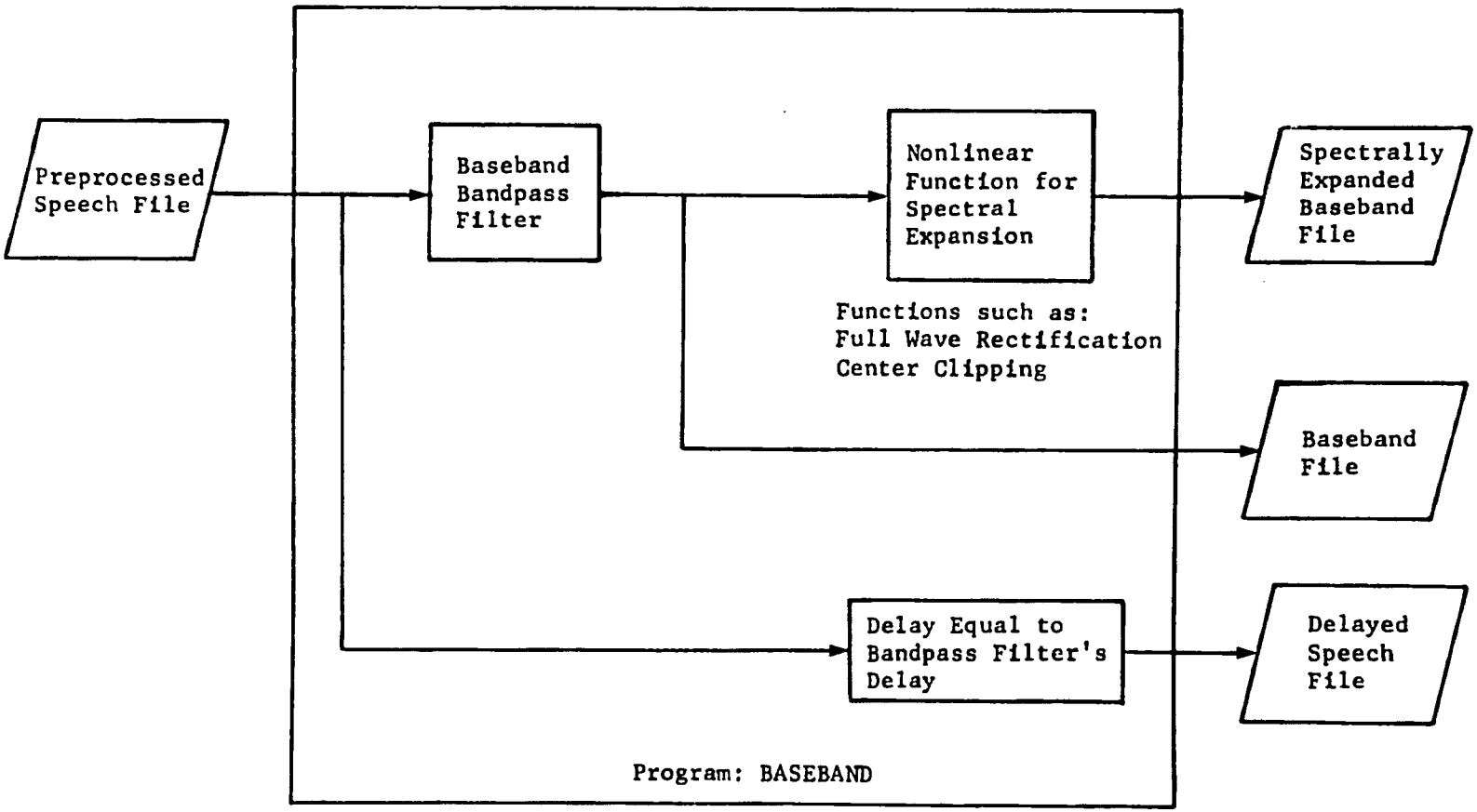
Figure 4.6. Block diagram of program BASEBAND.

This program generates three outputs from the preprocessed speech signal.

The baseband output is merely a low passed subband or baseband of the original speech signal. This baseband is rich in excitation information (e.g. voiced/unvoiced and glottal vibratory rate information).

Another output is generated by spectrally expanding the baseband signal. The expansion function is implemented with a center clipping routine (CCLIP) or an absolute value routine (VABS).

The third output is merely the input signal delayed to compensate for the baseband filter's delay.


Preprocessed Speech File
Baseband Bandpass Filter
Nonlinear Function for Spectral Expansion
Spectrally Expanded Baseband File
Functions such as:
Full Wave Rectification
Center Clipping
Baseband File
Delay Equal to Bandpass Filter's Delay
Delayed Speech File
Program: BASEBAND

summarize the basic structure of this software realization.

A number of details regarding this particular implementation will be discussed further in the following paragraphs.

Filter bank design

The filter banks are a group of finite impulse response (FIR) filters. One set of filter design routines generates linear phase filter coefficients. Other filter types are easily implemented. For example, filter coefficients that are similar to the estimated eighth nerve impulse response at selected basilar membrane locations may be useful for the implanted patient (Pfeiffer and Kim, 1975; Flanagan, 1972).

Linear phase filter techniques reduce filter bank design difficulties. Disturbing temporal effects arising when the bandpass filters have unequal time delays are circumvented using these filters.

Golden (1963, 1968) delves into the relatively cumbersome and conceptually unwieldly techniques for obtaining high fidelity filter banks using infinite impulse response (IIR) filters. Because of these factors, IIR filter banks are not as efficiently changed to accommodate varying requirements. Filter bank designs with equal bandwidth filters are well developed and are generally more computationally efficient than those made of unequal bandwidth filters. However, the speech production and perception mechanisms dictate just such unequal bandwidth filters for the most efficient speech signal representation (Flanagan, 1972; Rabiner and Gold, 1975; Kryter, 1962).

Linear phase FIR filter design techniques, in conjunction with

filter bank design methods using the window technique (Rabiner and Schafer, 1977), greatly speed the design of spectrally flat filter bank systems. Also, window filter design methods and interative approximation methods (McClellan, et al, 1973) allow great flexibility in realizing varied design specifications. With these techniques, filters with equal or unequal bandwidths may be distributed as desired across the spectrum of interest. One is assured of a nearly flat composite frequency response. For typical vocoder filter parameters, composite response ripple is less than a few tenths of a decibel.

For designing a single filter, the window design technique involves multiplying an "ideal" filter's (e.g. a rectangular-shaped passband) impulse response by a window function (see Figure 4.7). In the frequency domain, the ideal frequency response is convolved with the window's frequency response to yield a response with skirts of finite slope. In other words, the zero transition width of the ideal filter changes to a non-zero value. Also, the resultant passband and stopband frequency responses are not perfectly constant at all frequencies in the band. That is, some stopband and passband ripple is generated. This ripple is often called the "approximation error" and is greatest within the transition region.

Many types of windows have been developed for various purposes (Childers and Durling, 1975). The Kaiser (1963, 1966) window function offers great flexibility. With this window function, one can make a trade off between transition frequency width and peak approximation

Figure 4.7. Diagram illustrating the window technique for filter design.

Part "A" illustrates the time (left) and frequency (right) domain representations of an "ideal" filter's impulse response. For an "ideal" lowpass filter (i.e. rectangular passband) with a cutoff frequency of "Wp" and unity passband gain, the impulse response is:

$$IR(n) = \sin (W_p * n * T_0) / (n * T_0 * pi)$$

where T_0 is the sampling interval and pi is 3.1416...

Part "B" illustrates the time (left) and frequency (right) domain representations of a typical window function. For example, a Kaiser window might be used.

Equation of the Kaiser window function:

$$W(n) = I_0(B*SQRT(1-(2*n/(N-1))**2))/I_0(B)$$

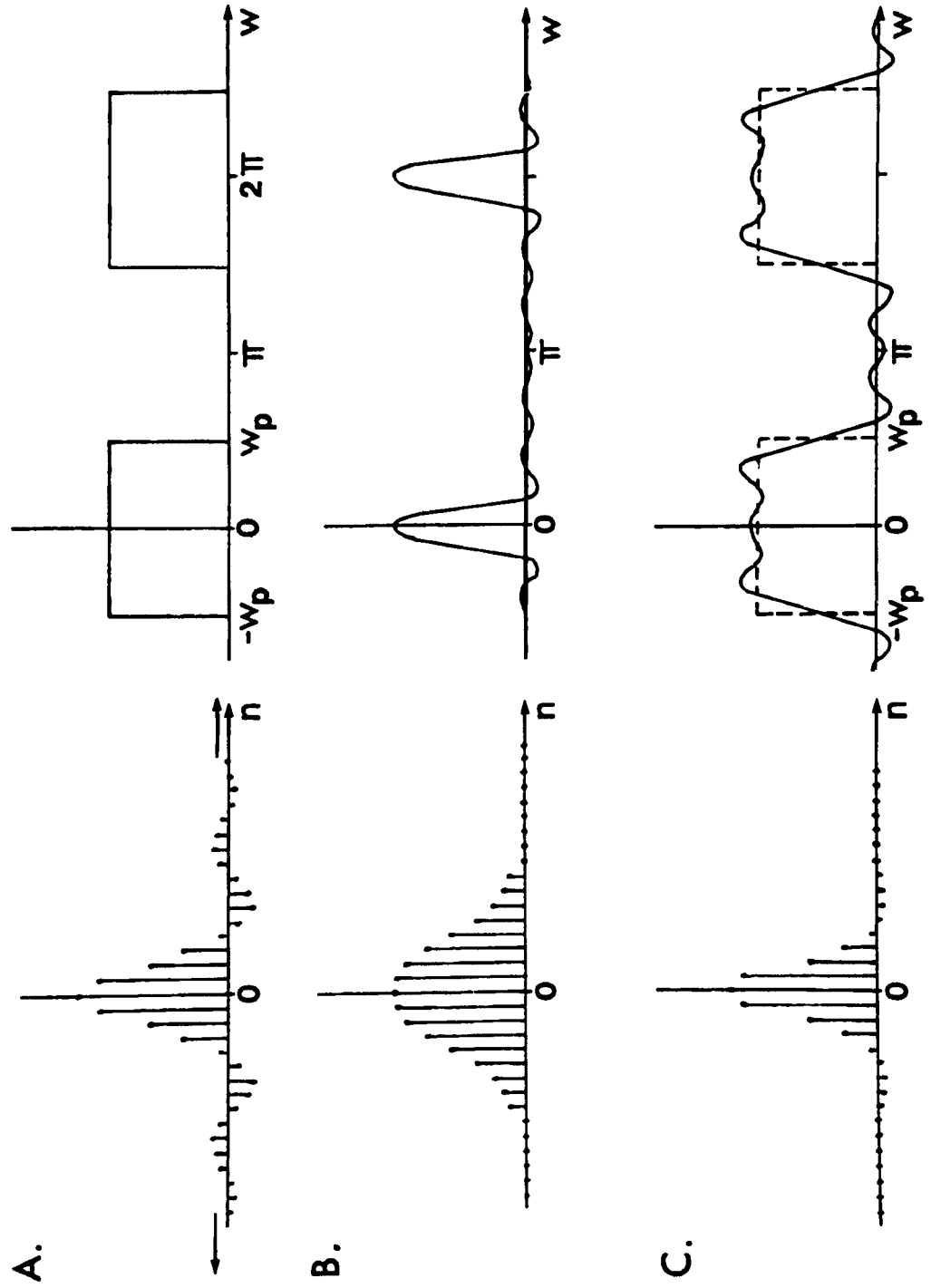
for n greater than or equal to -(N-1)/2 and

less than or equal to (N-1)/2

W(n) = 0 for all other values of n

where B is a constant that specifies the frequency response tradeoff between the peak sidelobe ripple and the "width" or energy of the main lobe. $I_0(x)$ is the modified zeroth order Bessel function.

Part "C" illustrates the result of time domain multiplication of the ideal filter's impulse response by the window function. The result in both the time and frequency domain is illustrated.


A.
B.
C.
n
0
w
$-w_p$
w_p
π
2π

error. As with most other windows, the transition region, delta W, is inversely proportional to window length, L.

A property of the window design method is very useful in the design of filter banks. The frequency response of the resultant filter is nearly antisymmetric about the point (Wp, .5). This is illustrated in figure 4.8. In designing a two-filter, filter bank with adjacent passbands, much of one filter's approximation error is cancelled in the filter bank's composite response by the other's nearly equal and opposite error.

If the same window is used for all filters, a bank of these filters should exhibit a very flat composite frequency response. Due to approximate antisymmetry at the juncture of these filters, filter approximation errors will be partially cancelled by other filters in the bank. At the very least, errors will not grossly summate in the composite response. Figure 4.9 is an example of such a filter bank response. This example emphasizes that filter bandwidths need not be equal; just that the transition frequency widths be equal. Thus, the window functions should be equal.

Finite impulse response filters

Finite impulse response filters offer a number of advantages in the speech processing field: 1) FIR filter bank design techniques are versatile. 2) Computational noise is not commulative. Recursive filters suffer from cummulative errors if input data and/or filter coefficients are represented with limited resolution. These cummulative errors occur because past filter outputs, with their errors, are used to calculate present and future filter outputs. 3) Computational speed is very comparable to recursive filters if decimation or FFT convolution techniques are used (Gold and Rader, 1969). 4) Linear phase filters are easily implemented.

Three basic methods are used for implementing a finite impulse response filter. They are: 1) Direct implementation using discrete convolution. 2) Discrete convolution using decimation and interpolation techniques (Rabiner and Schafer, 1977). The decimation technique reduces the filter's output sampling rate relative to the filter's input sampling rate. 3) The fast fourier transform convolution algorithm (Stockham, 1966). In this technique, time domain convolution is implemented by frequency domain multiplication of the transformed signal and the transformed filter's impulse response. The result of the multiplication is transformed back to the time domain.

The last two methods were primarily developed to improve computational efficiency. The FIR filters can be easily implemented using any one of these three methods. One method could easily be substituted

Figure 4.8. Frequency response of a window designed lowpass filter (after Rabiner and Schafer, 1977).

This diagram illustrates the approximate antisymmetry at the point (W_p,.5). W_p is the ideal lowpass filter's cutoff frequency. W_p is also the resultant filter's -6 dB cutoff frequency. Delta W (ΔW) is the frequency transition width. Sigma (δ) is the peak error ripple in the frequency response of the resultant filter.

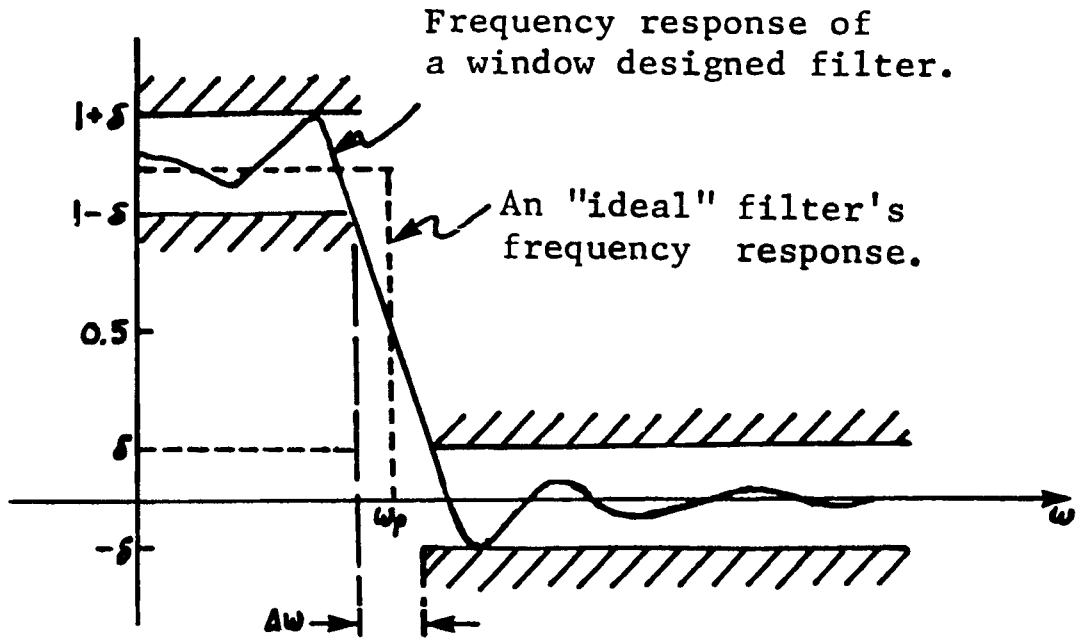
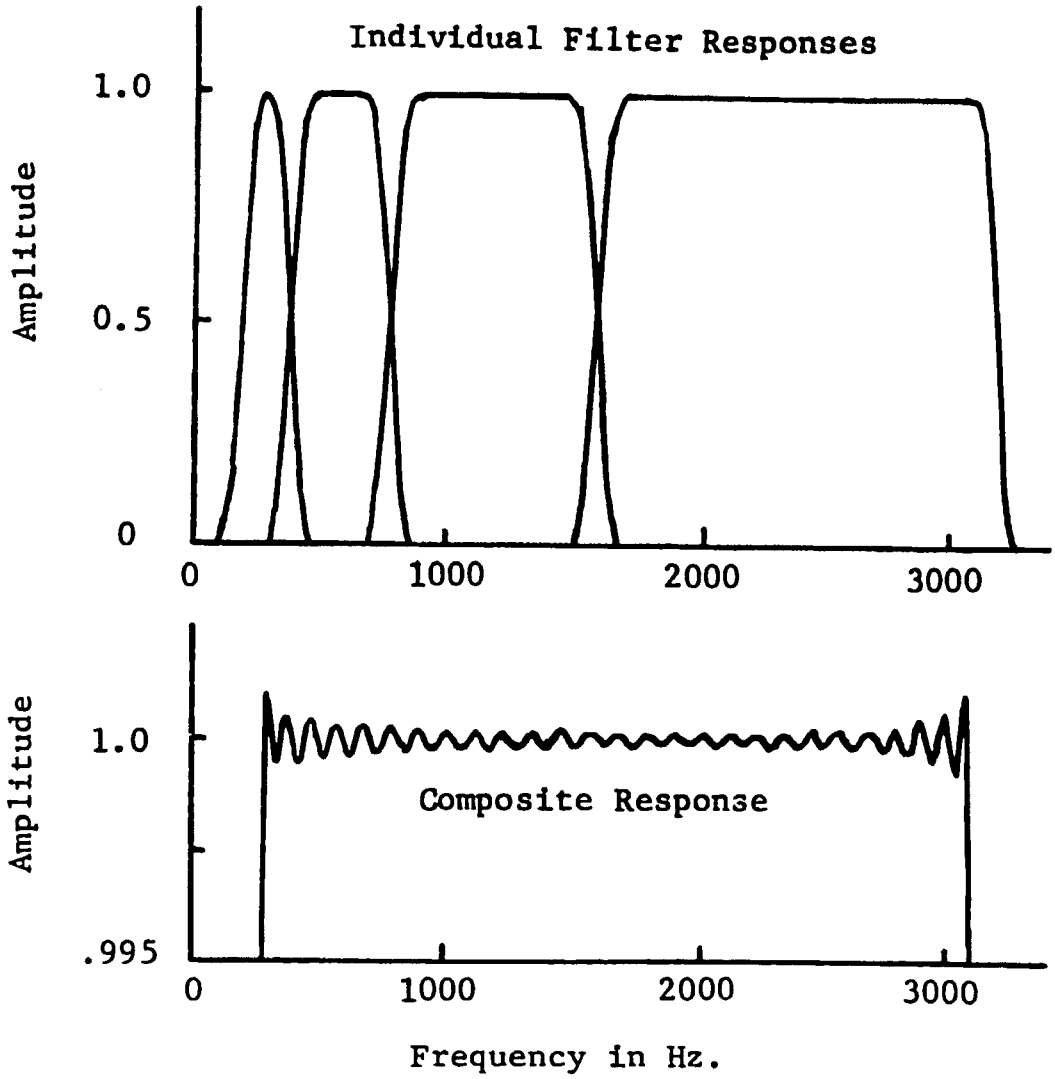

Frequency response of
a window designed filter.
An "ideal" filter's
frequency response.
1+δ
1-δ
0.5
δ
-δ
ωp
ω
Δω

Figure 4.9. Example of a filter bank design using the window design technique. Individual filter responses and the composite response are illustrated (after Rabiner and Schafer, 1977).


Individual Filter Responses
Amplitude
1.0
0.5
0
0
1000
2000
3000
Amplitude
1.0
.995
Composite Response
0
1000
2000
3000
Frequency in Hz.

for another; just by substituting one subroutine for another.

The decimation method was initially implemented. This method was chosen for the following reasons: 1) The sampling rate in intermediate processing stages is reduced. Therefore, processing speed is increased in the intermediate processing stages. Also, a lower sampling rate reduces primary and secondary storage requirements. 2) A hardware implementation of this method is more economically feasible than implementing the FFT method. The FFT method would require extremely fast and expensive FFT processor(s); whereas, relatively simple CCD hybrid convolving circuitry could be used to implement the second method. However, the expense of such FFT processors may be greatly reduced in the near future due to improved FFT algorithms (Kolba and Parks, 1977) and reduced hardware prices (Broderen, 1975; Noyce, 1977). 3) For vocoder filtering requirements, the processing speed is approximately equivalent to, or better than, that of an FFT implementation.

Stockham's somewhat conservative estimate of FFT filter processing time for each input sample is (Gold and Rader, 1969):

$$\text{Time} = 2*\text{Kfft}*(\log_2(N)+1)$$

This neglects auxiliary storage time requirements.

Where:

Kfft = (time required for an M point FFT) / (M * Log_2(M))

N = (the number of samples in the filter's impulse response)

For an assembler written vender supplied, FFT routine, Kfft is approximately 150 μsec when executed on a Data General NOVA 1220.

For a single stage filter, using the decimation technique, processing time per sample may be estimated as:

Time = Kmado * N / Dratio

Where:

Kmado is the multiply, add, and overhead time required per sample in the down-sampled convolution. For the author's assembly language routine, Kmado is 40μsec.

Dratio = (input sampling rate) / (output sampling rate)

This ratio is sometimes referred to as the decimation ratio.

Dratio, for most practical spectrum channel vocoder filters, is set between five and ten. Aliasing criteria set these limits. For the following analysis, Dratio will be conservatively chosen to be equal to five.

The decimation method is faster than the FFT method for filter lengths less than "N_e". Setting the two comutation time equations equal, one may solve for "N_e".

$$2 * Kfft * (\text{Log}_2(N_e) + 1) = Kmado * N_e / Dratio$$

Substituting practical values for the variables:

$$2 * (150\ \mu sec.) * (\text{Log}_2(N_e) + 1) +/ = (40\ \mu sec.) * N_e / 5.0$$

The solution to this equation indicates that N must be somewhat larger than 256 if the FFT realization is to be more efficient. Most vocoder filter impulse responses are on the order of 50 to 300 samples. Thus, the single stage decimating filter appears to be quite competitive with the FFT realization for moderately narrow bandwidth filters.

Decimation-Interpolation for the purpose of reducing computational and memory requirements.

Rabiner and Schafer (1978) offer a very good summary of the decimation-interpolation technique. Decimation involves "down-sampling" a time series; such that only every nth sample is saved for further processing. Interpolation is the inverse process; in which a time series of m samples is converted into a time series of (n) (m) samples. Savings in computation and memory can be obtained if a time series is first down-sampled, then processed, and then interpolated to obtain a time series at the input signal's sampling rate.

The constraints on this processing technique primarily involve aliasing considerations. The input time series should not be down-sampled so severely that the result will be significantly aliased. In addition, further processing of the down-sampled signal should not generate any spectral components that will be significantly aliased.

In determining the decimation-interpolation ratio, consideration should be given to the interpolator's output filter. Interpolation is a two stage process. First, the interpolator's input is expanded such that "n-1" zero value samples are placed between each input sample. Second, this newly created series is filtered to obtain the output time series. The intermediate time series spectrum contains "n" representations of the input spectrum. The output filter and decimation-interpolation ratio should be chosen such that only the desired spectral representation is passed to the output time series.

Spectral Flattening

Rabiner and Gold (1975) suggest a bank of bandpass filter-AGC combinations for gross spectral flattening of the excitation signal. With this technique, slowly varying fluctuations in the expanded baseband signal's spectral components cannot corrupt the synthesizer's output. All pertinent, slowly varying fluctuations are reflected in the spectrum analyzer's envelope signals.

The described vocoder implementation incorporates this recent suggestion. A comprehensive comparison of this spectral flattening method with other methods (see figure 4.10) has not been attempted. However, certain comparisons can be made. When two or more harmonics are present in the filter's passband, a hard limiting system probably will not have the fidelity of an AGC system. This follows from the effect of the hard limiter on two harmonics. The smaller amplitude harmonic is generally further suppressed; thereby causing it to contribute less to the vocoder channel's output than what it had originally contributed (Rabiner and Gold, 1975).

Choice of processing parameters and examples of speech processing

The analysis and synthesis bandpass filters have common design requirements. Although one might assume that sharp cutoff bandpass filters (e.g. elliptical, butterworth) would perform well; their long (50msec. or more) and "ringing" impulse response tends to temporally

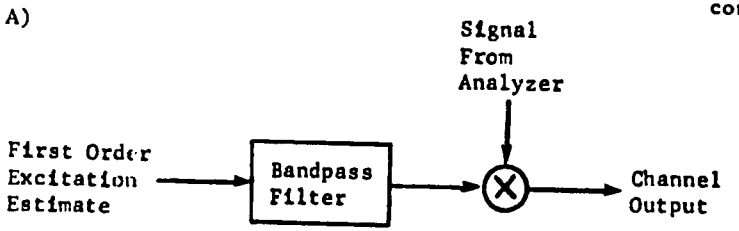
Figure 4.10. Three types of vocoder synthesis stages.

A) Vocoder synthesis stages without spectral flattening.

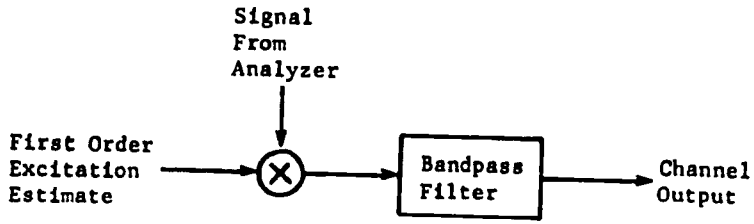
B) A vocoder synthesis channel which uses two bandpass filters and a hard limiter for spectral flattening.

C) A vocoder synthesis channel which uses a bandpass filter and AGC for spectral flattening. Since very little harmonic distortion is generated, the output bandpass filter is generally not necessary.

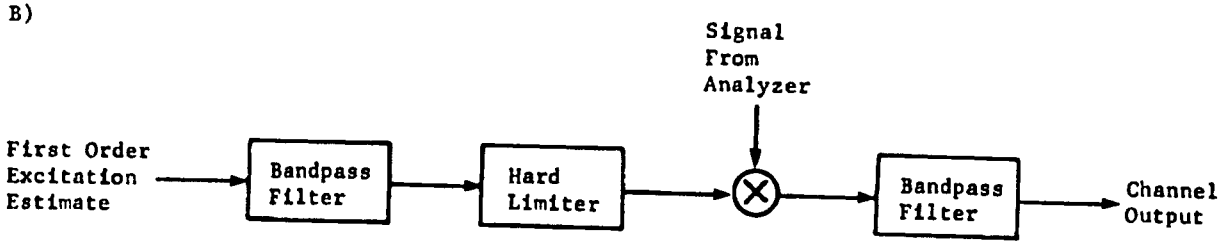
A)



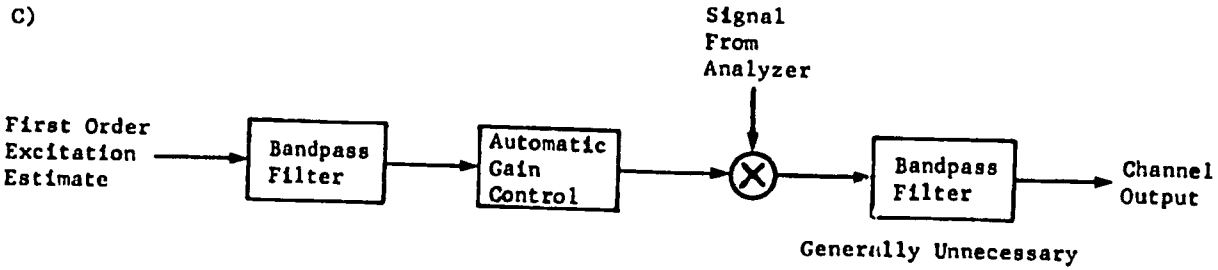
or this configuration:



B)



C)



Generally Unnecessary

"smear" and "echo" important ongoing changes in the speech signal (Golden, 1968). In speech processing, filters with relatively short impulse responses (10-30 msec.) give the best balance of temporal and spectral resolution.

FIR bandpass filters, with 10 to 30 msec. impulse responses and transition widths of 100 to 200 Hz. with corresponding stopband attenuations from 35 to 55 dB, are good candidates for vocoder bandpass filters. Bessel infinite impulse response filters, having similar characteristics, have been the filter of choice for hardware implementations of channel vocoders.

The analysis stage lowpass filter should attenuate voicing components (frequencies above 50-100 Hz.) but pass the slowly changing components (frequencies below 25 Hz.). This extremely small bandwidth represents a considerable reduction in information rate. The semantic information is primarily generated by the slowly moving vocal tract configuration. As with the bandpass filters, this lowpass filter should have very little ringing. This prevents disturbing reverberation effects. The Hanning-Tukey window, the Hamming window, and the Kaiser window function are possible candidates for such a FIR filter realization. Bessel or Lerner IIR filters are also candidates.

The AGC lowpass filter has similar requirements to that of the analysis lowpass filter. Little, or no, impulse response ringing (see figure 4.12) and a nearly unity passband gain(see figure 4.13) are important specifications for such a filter design. The AGC and the analysis lowpass filters' stopband and passband frequency specifications are similar.

Figure 4.11 indicates the system measurement locations used in these and subsequent illustrations. Figures 4.15, 4.17, and 4.19 illustrate spectrum channel vocoder processing of a speech signal. Figure 4.14 displays the two inputs to the vocoder channel. Table 4.3 contains a listing of the processing parameters used for the vocoder illustrated in figure 4.15. Figures 4.17 and 4.19 illustrate the effects of slight processing parameter modifications.

Figures 4.15 and 4.17 are essentially identical, even though the stopband attenuation of the analysis and synthesis bandpass filters has been changed from 34.7 dB to 53.1 dB.

Figures 4.15 and 4.19 illustrate response differences due to changing the lowpass filter parameters. The responses are very similar. However, the longer duration filter (i.e. Kaiser window convolved with itself in the time domain) tends to "smooth-out" the envelope detector's output. The higher frequency passband components have been somewhat attenuated. The Hanning filter's passband has a more constant passband gain. However, other filters (including other Kaiser designs) can be designed with flatter passbands, little ringing, and reasonable stopband attenuations.

Figure 4.20 presents time series data from a 16-channel, constant channel bandwidth, voice excited vocoder. Tables 4.4 through 4.6 summarize the parameters used for this vocoder simulation. This figure is presented to facilitate interchannel comparisons. For example, changes in formant frequencies can be visualized by comparing the envelope detector outputs.

Accurate vocoder performance evaluations and "fine tuning" of processing parameters for this vocoder and other variations require well designed listening tests. These tests are presently being implemented in Coleman Memorial Laboratory at the University of California, San Fransicso. Listening tests, such as the diagnostic rhyme test (DRT; Voiers, 1973) and the Syntactically Normal Sentence Test (SNST; Nye and Gaitenby, 1974), will be used to gain a better measure of the differences, to the listener, between sets of processing parameters. Some sets of parameters may allow the vocoder to acoustically simulate certain aspects of a cochlear prosthesis. Also, as mentioned previously, the spectrum channel vocoder analysis stage may be a very useful initial cochlear prosthesis processing stage.

for number sequence only

Figure 4.11. Block Diagram of One Analysis-Synthesis Channel Used in a Spectrum Channel Vocoder Realization.

This figure is composed of Figure 4.4 with markers added (e.g., A). These markers indicate where time series signals were measured. Subsequent figures will contain plots measured at these system measurement points.

Subroutine CHANNEL

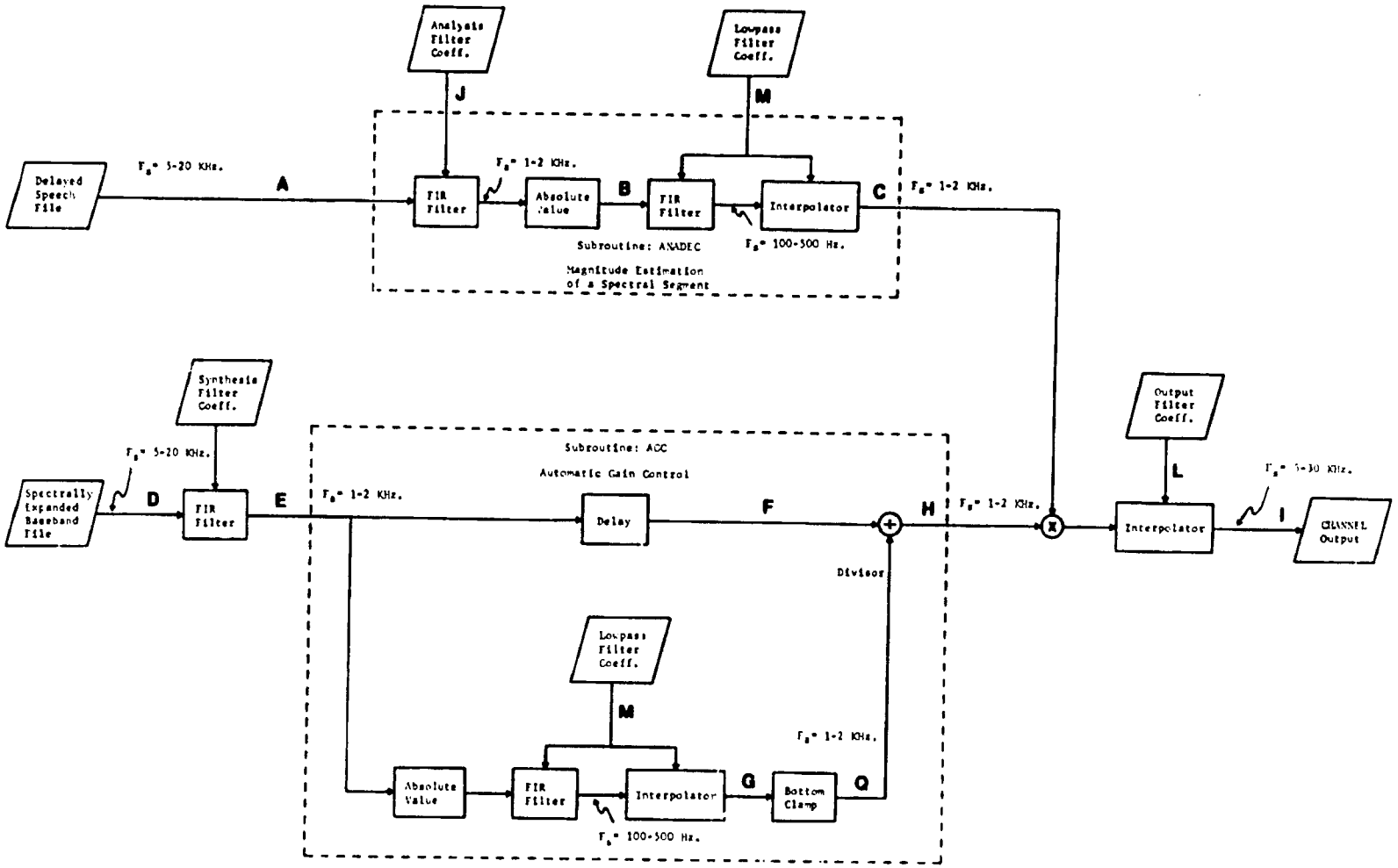


Figure 4.12. Plots of Time Series Data Derived from the AGC Section of A Vocoder. Refer to Figure 4.11 for the corresponding block diagram locations.

The input (location "D") to the synthesis bandpass filter was an amplitude modulated sinusoid. A 1200 Hz carrier was 100% modulated by a 15 Hz sinusoid. The filter had a 200 Hz bandwidth centered at 1200 Hz. The bandpass filter was set to decimate its output sampling rate by a factor of 6.

Curve 1. "E" is the bandpass filtered and decimated version of the input.

Curve 2. "G", the envelope detector's output, is superimposed on the detector's input, "E". Note where "G" becomes negative (see arrows).

Curve 3. "Q" is almost identical to "G" except that the signal has been "bottom clamped". The arrows indicate where bottom clamping has occurred on the time series.

Bottom clamping definition:
Q(n) = G(n) if G(n) is greater than some minimum vlaue, "RMIN".
Q(n) = RMIN if G(n) is less than or equal to "RMIN".
In this example, "RMIN" was set near zero.

Curve 4. "H" is the result of passing "E" through an automatic gain control with this set of lowpass filter and bottom clamp parameters. The Kaiser lowpass filter had a considerable amount of ringing in its impulse response. It also contained passband and stopband ripple. Filter gain was slightly above unity at some frequencies.

This example illustrates that the AGC's output can potentially go out of bounds, as the AGC's envelope detector output crosses zero. Arrows in curve 4 indicate where in the wave form this effect can be clearly seen. The bottom clamp helps, but the output discontinuity generated is not desirable.

This is a somewhat extreme example, because the input was chosen to have a very high modulation level (100%) and a relatively high modulation frequency (15 Hz). Later examples of normal speech better illustrate the normal requirements and performance of such an automatic gain control.

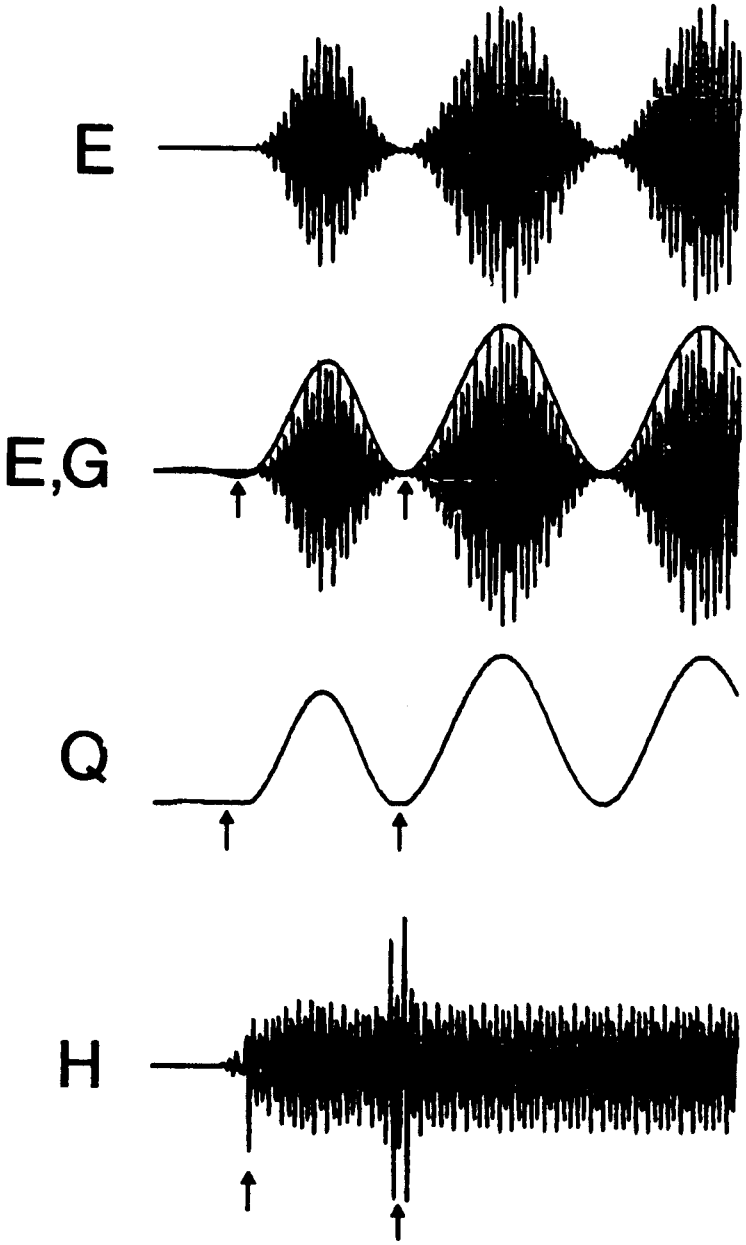

E
E,G
Q
H

Figure 4.13. Plots of time series data derived from the AGC section of a vocoder.

This illustration is similar to the previous one. The input signal is identical. However, the AGC's lowpass filter (Hanning window; approx. 25 msec. impulse response) impulse response never becomes negative. This insures that the envelope detector's output will not go negative. This eliminates the possibility that the AGC's output will go out of bounds.

However, all passband frequencies above D.C. are slightly attenuated. Indeed, the frequency response monotonically decreases. This lowpass filter's frequency response is not as "flat" as the previous example's lowpass response. Consequently, this AGC exhibits a greater "feedthrough" of the higher frequencies than the previous AGC. This is illustrated in the AGC's output, curve "H".

"E" is the bandpass filtered and decimated version of the input.

"G" is the envelope detector's output.

"H" is the result of passing "E" through the automatic gain control.

As previously stated, this is a somewhat extreme example, because the input was chosen to have a very high modulation level (100%) and a relatively high modulation frequency (15 Hz.). Later examples of normal speech better illustrated the normal requirements and performance of such an automatic gain control.

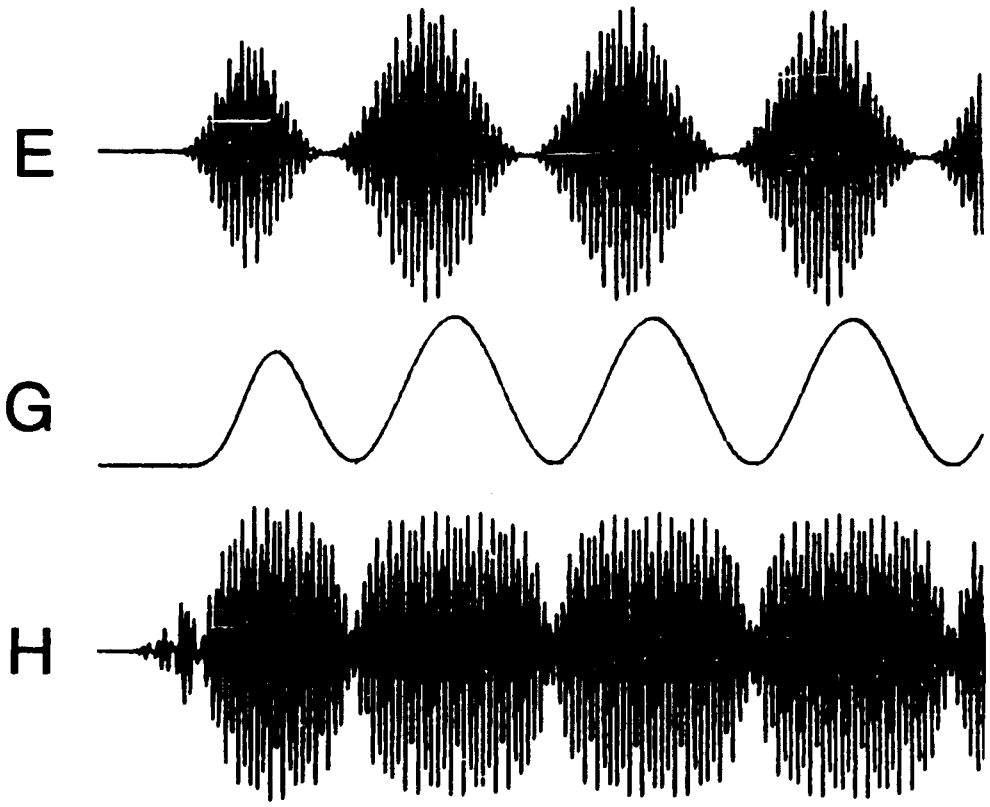

E
G
H

Figure 4.14. Plots of the two input signals"driving" the CHANNEL subroutine.

"A" is the delayed speech signal.

"D" is the spectrally expanded baseband signal.

These input signals are used as inputs for the subsequent vocoder performance illustrations.

The vertical scale factors have been selected to obtain reasonable interplot spacing and resolution. The input signal sampling rate is 10 KHz.

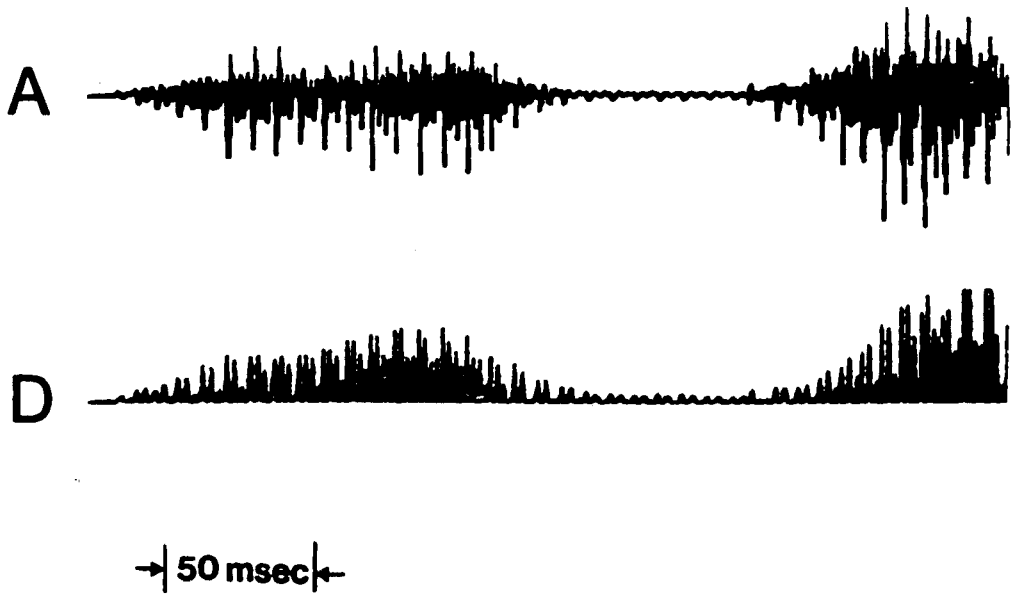

A
D
50 msec

Figure 4.15. Plots of time series data obtained at intermediate locations in the channel vocoder processing chain. Refer to Figure 4.11 for the corresponding system locations.

"A" is the delayed input signal.

"B" is the signal after it has been bandpass filtered and decimated by the analysis filter and then full wave rectified.

"C" represents the lowpassed version of "B". Notice that the voicing (i.e. glottal pulse) component has been eliminated by this filtering. "C" is the output of the analyzer's envelope detector.

"D" represents the spectrally expanded baseband input to the vocoder channel.

"E" is a bandpass filtered and decimated version of the signal "D". This synthesis stage filtering determines the channel's output frequency range (i.e. where on the basilar membrane the channel's excitation will occur).

"G" is an envelope estimate of "E". This signal will be used to automatically control the amplitude of "F".

"F" is merely a delayed version of "E". Its delay compensates for the lowpass filter's delay.

"H" is similar to "E", except that all slowly changing amplitude variations in "E" have been minimized by the Automatic Gain Control (AGC). Note that the voicing component is present and has not been attenuated by the AGC operations.

"I" represents the vocoder channel output. The analyzer's output ("C") modulates the excitation component ("H") to generate the decimated output signal. The signal is then interpolated to obtain the channel's output.

The vertical scale has been individually chosen for each curve; such that reasonably legible curves were generated.

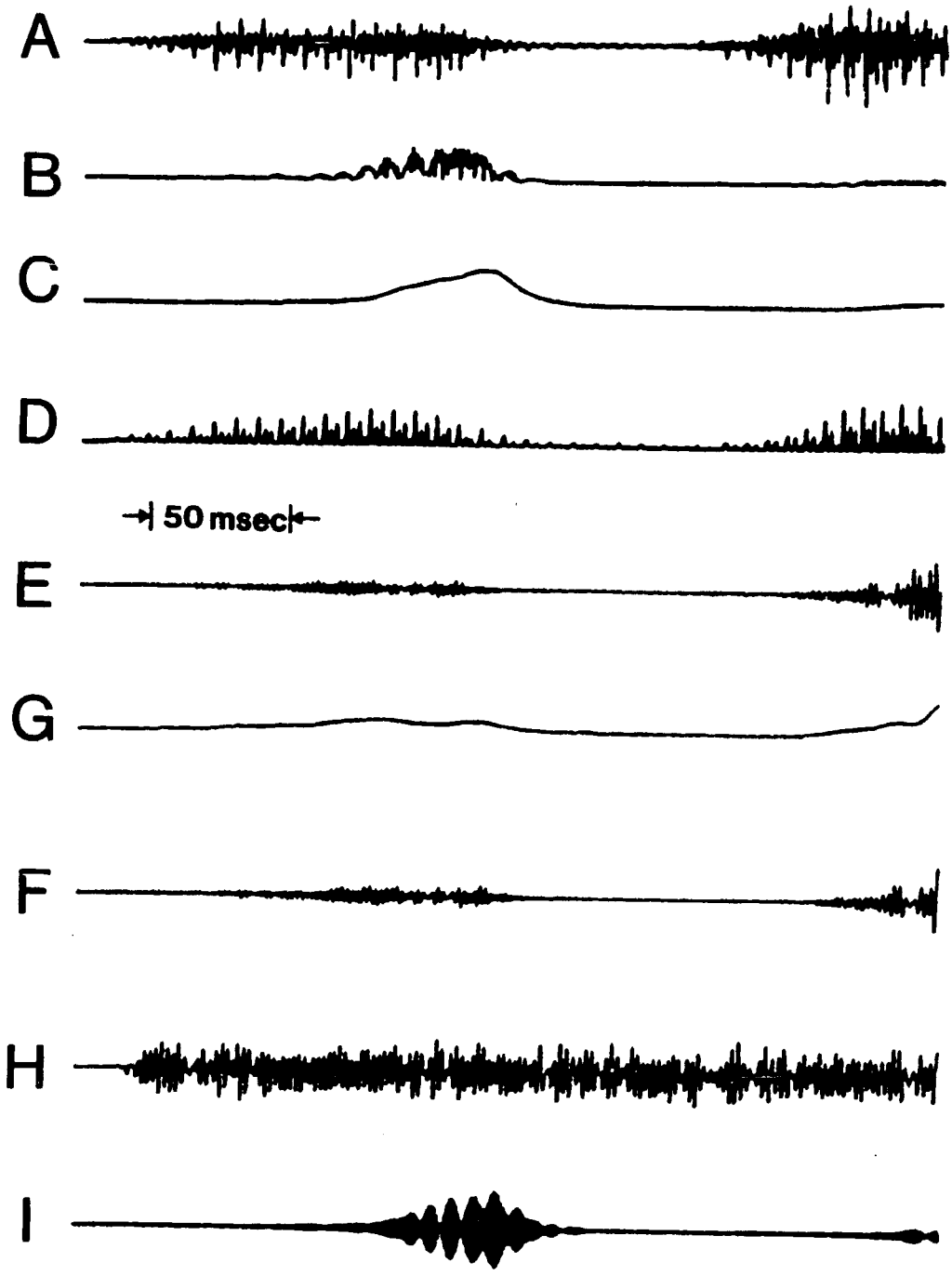

A
B
C
D
50 msec
E
G
F
H
I

Figure 4.16. Impulse response plots of the three bandpass filters used to generate the previous illustration.

"J" is the analysis filter's impulse response.

"K" is the synthesis filter's impulse response. In this case, "J" = "K".

"L" is the output interpolating filter's impulse response.

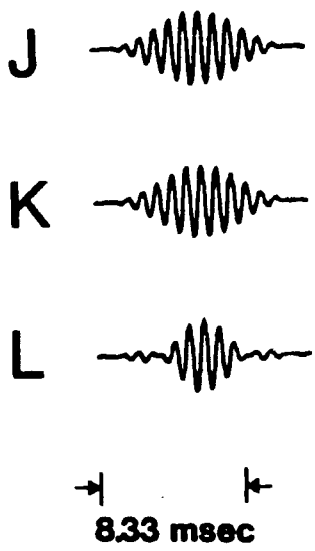

J
K
L
8.33 msec

Table 4.3. Processing parameters used to specify the vocoder channel illustrated in the previous illustration.

Input and output sampling rate = 10 KHz.

Analysis and synthesis bandpass filter characteristics:

Type of filter = Kaiser window design

Filter length = 121 pts. (i.e. a 12.1 msec. impulse response)

Transition width (delta F) = 200 Hz.

-6 dB frequencies = 1100, 1300 Hz.

Stopband ripple = passband ripple = 34.7 dB down

Sampling rate after bandpass filtering = 1.666 KHz.

(i.e. decimation ratio = 6)

Lowpass filter characteristics:

The analysis and AGC lowpass filters are identical.

Type of filter = Hanning window

Filter length = 55 pts. (i.e. 33 msec. impulse response at 1.666 KHz.)

Cutoff frequency = 39 Hz. (3 dB down)

42 dB down at 90 Hz.

Decimation ratio = 1

Output interpolating filter section:

Type of filter = Kaiser window design

Filter length = 121 pts

Transition width = 200 Hz.

-6 dB frequencies = 1000 Hz. and 1400 Hz.

Stopbands are above 1500 Hz. and below 900 Hz.

Table 4.3 continued:

Stopband ripple = passband ripple = 34.7 dB down

Sampling rate after interpolation = 10 KHz. (i.e. interpolation ration = 6)

Figure 4.17. Plots of time series data obtained at intermediate locations in the channel vocoder processing chain.

This illustration is similar to Figure 4.15. The processing parameters are identical except for those relating to the analysis, synthesis, and output bandpass filters. In this case, the filter lengths were increased to 185 samples. This new length generates a passband and stopband ripple of -53.1 dB as opposed to the previous illustration's -34.7 dB. Little change is noticeable in the various signals in the system. Only the added delay is noticeable.

"A" is the delayed input signal.

"B" is the signal after it has been bandpass filtered and decimated by the analysis filter and then full wave rectified.

"C" represents the lowpassed version of "B".

"D" represents the spectrally expanded baseband input to the vocoder channel.

"F" is a bandpass filtered, decimated, and delayed version of the expanded baseband signal.

"G" is an envelope estimate of "F".

"H" is similar to "F", except that all slowly changing amplitude variations have been minimized by the AGC.

"I" represents the vocoder channel output. The analysis stage output ("C") modulates the excitation component ("H") to generate the intermediate output signal. The signal is then interpolated to obtain the channel's output.

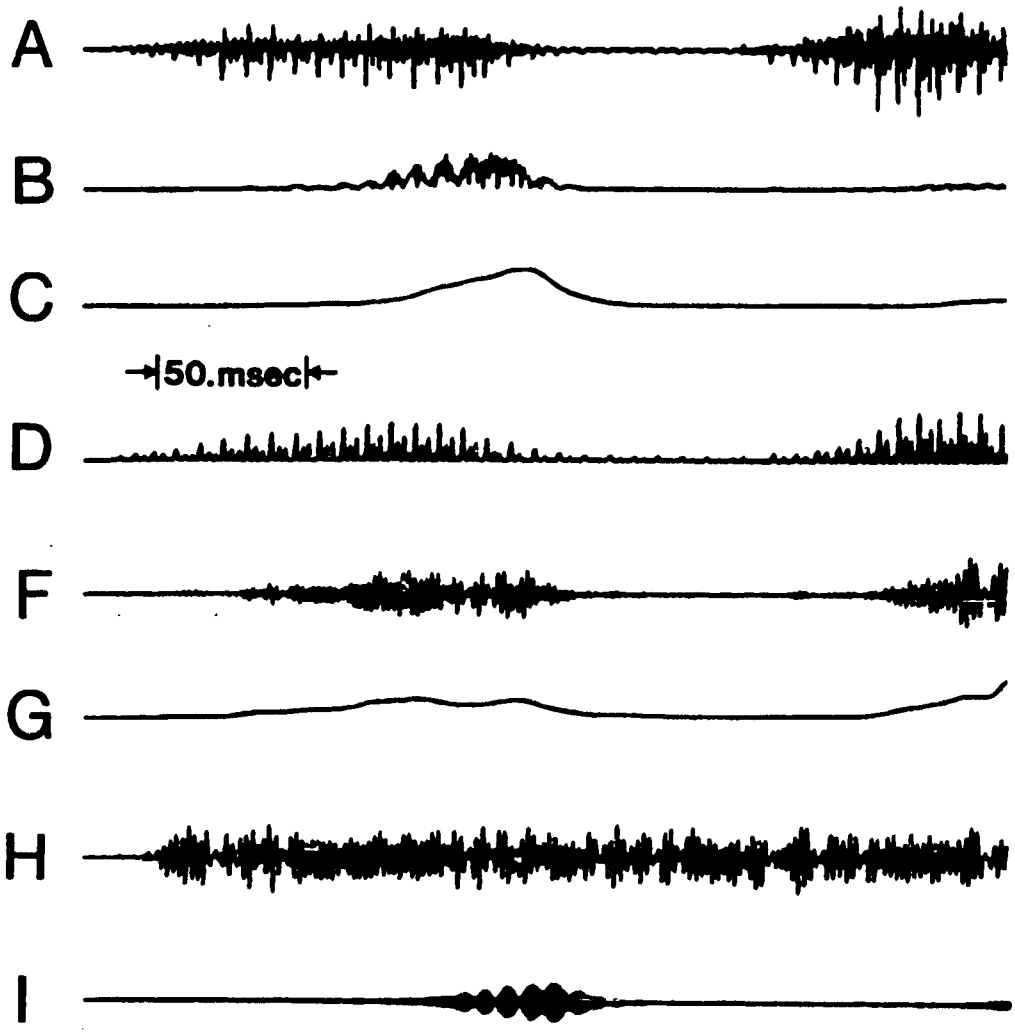

A
B
C
50.msec
D
F
G
H
I

Figure 4.18. Impulse response plots of the three bandpass filters used to generate the previous illustration. Note that filters "J", "K", and "L" have a longer impulse response than those illustrated in Figure 4.16.

"J" is the analysis filter's impulse response,

"K" is the synthesis filter's impulse response.

"L" is the output interpolating filter's impulse response.

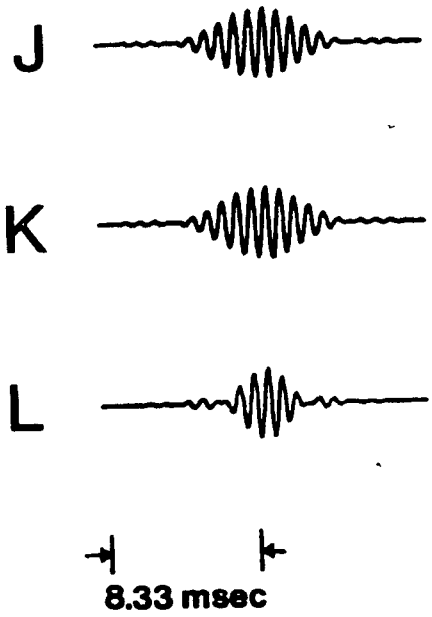

J
K
L
8.33 msec

Figure 4.19. Plots of time series data obtained at intermediate locationsin the channel vocoder processing chain.

This illustration is similar to Figure 4.15. The processing parameters are identical except for those relating to the lowpass filter and the decimation-interpolation ratio. The lowpass filter was changed to a Kaiser window convolved with itself in the time domain. After convolution, the impulse duration was 61 msec. The transition width was 80 Hz. (with a 12 dB cutoff at 40 Hz.) and minimum stopband attenuation was 67.4 dB. In this example, the bandpass decimation ratio was changed from 6 to 5. The ratio change had little consequence. Due to the new lowpass filter, curves "C" and "G" do appear more "smoothed" and delayed than those of Figure 4.15. However, little more difference is readily apparent from these plots.

"A" is the delayed input signal.

"B" is the "A" signal after it has been bandpass filtered and decimated by the analysis filter and then full wave rectified.

"C" represents the lowpassed version of "B". "C" is the output of the analyzer's envelope detector.

"D" represents the spectrally expanded baseband input to the vocoder channel.

"E" is the bandpass filtered and decimated version of "D".

"G" is an envelope estimate of "E".

"F" is a delayed version of "E". This delay compensates for the envelope detector's lowpass filter delay.

"H" is similar to "E", except that all slowly changing amplitude variations have been minimized by the automatic gain control.

"I" represents the vocoder channel output after interpolation.

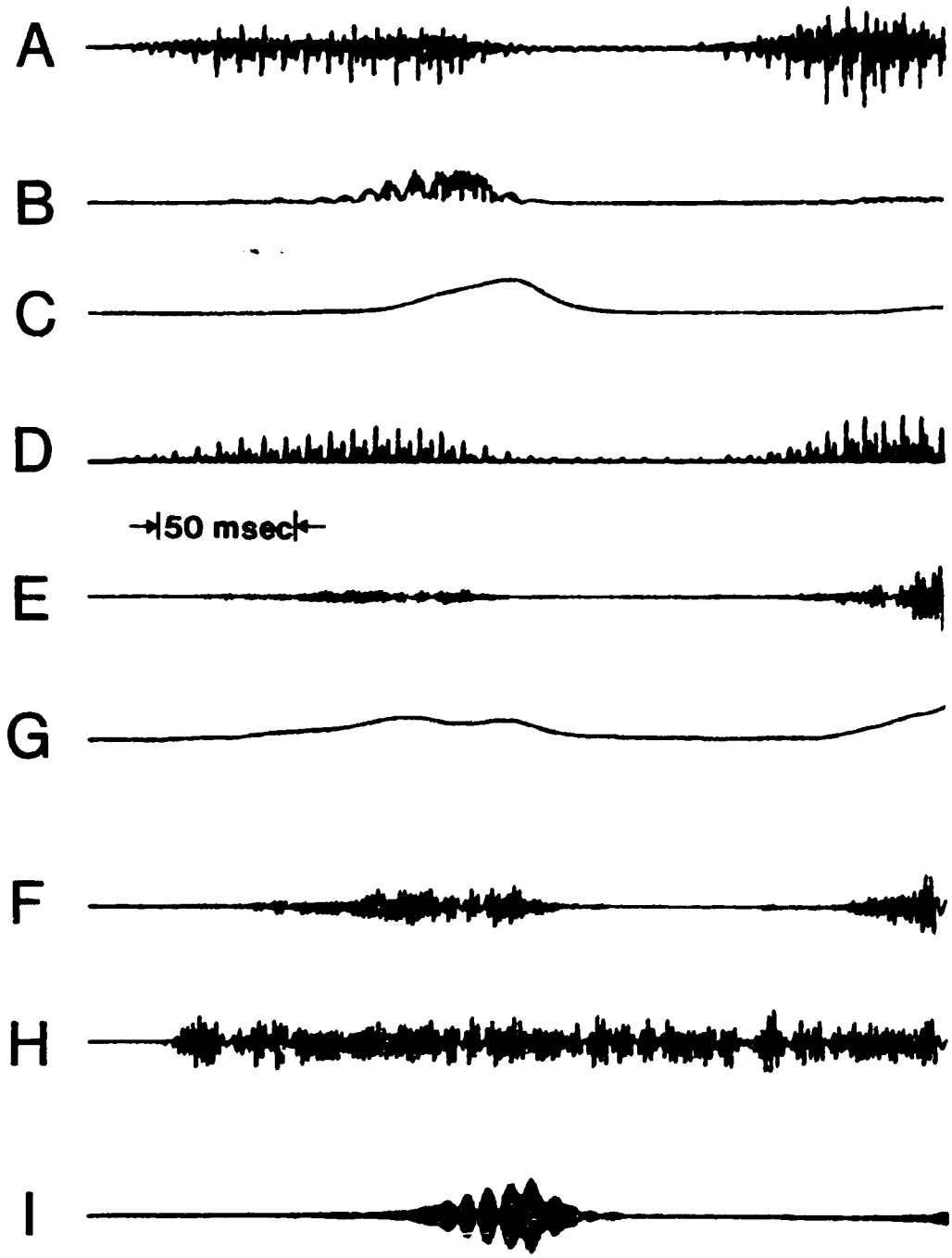

A
B
C
D
50 msec
E
G
F
H
I

Figure 4.20. Plots of time series data obtained at intermediate locations in a sixteen channel, vocoder example.

The next sixteen pages comprise this illustration. Processing parameters are summarized in Tables 4.4 through 4.6.

"B" represents a bandpass filtered and decimated and rectified version of the delayed speech input signal.

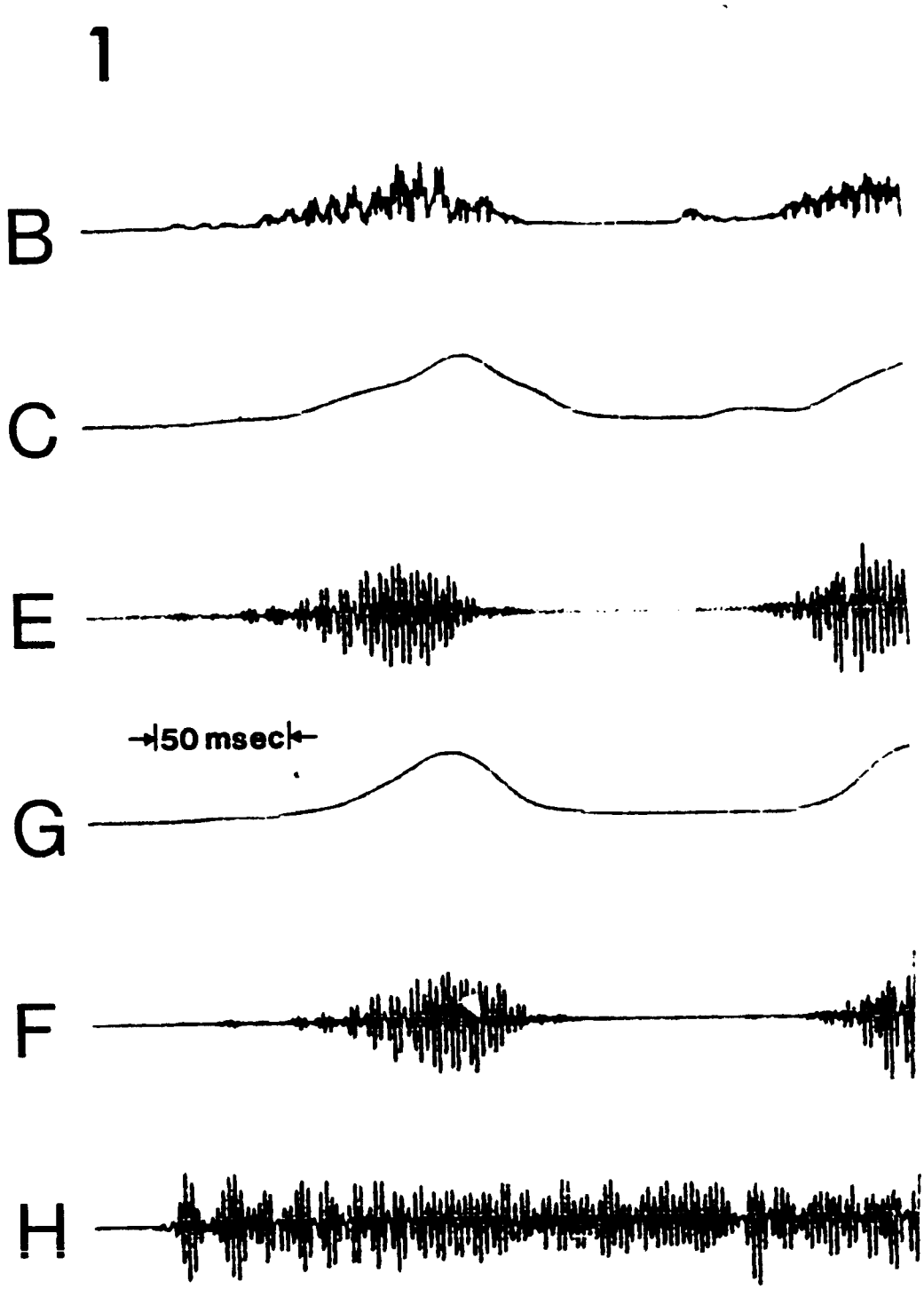
"C" is the output of the analyzer's envelope detector. It is a lowpass filtered version of "B". "C" is the analyzer's output.

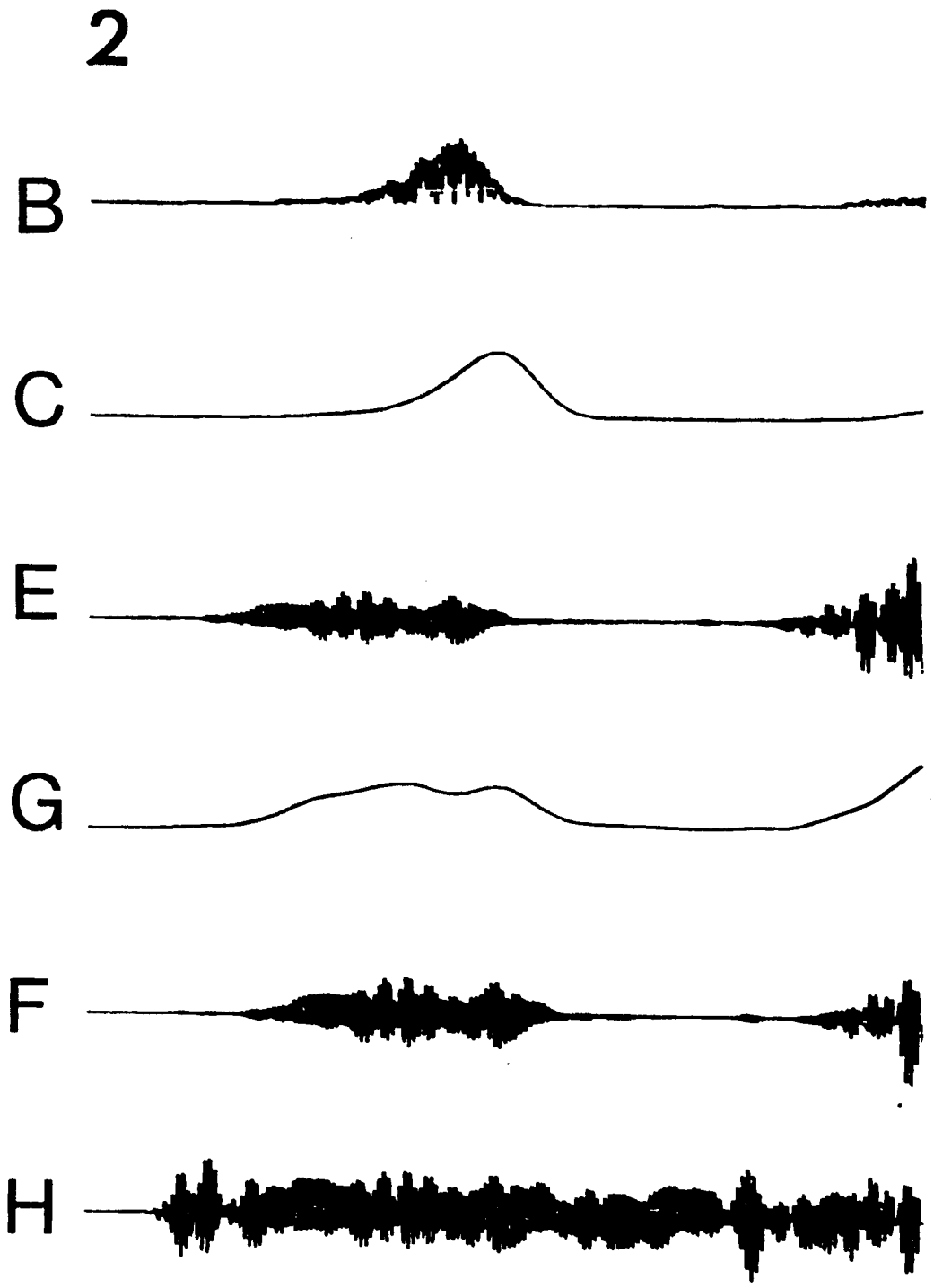
"E" is the bandpass filtered and decimated version of the expanded baseband signal.

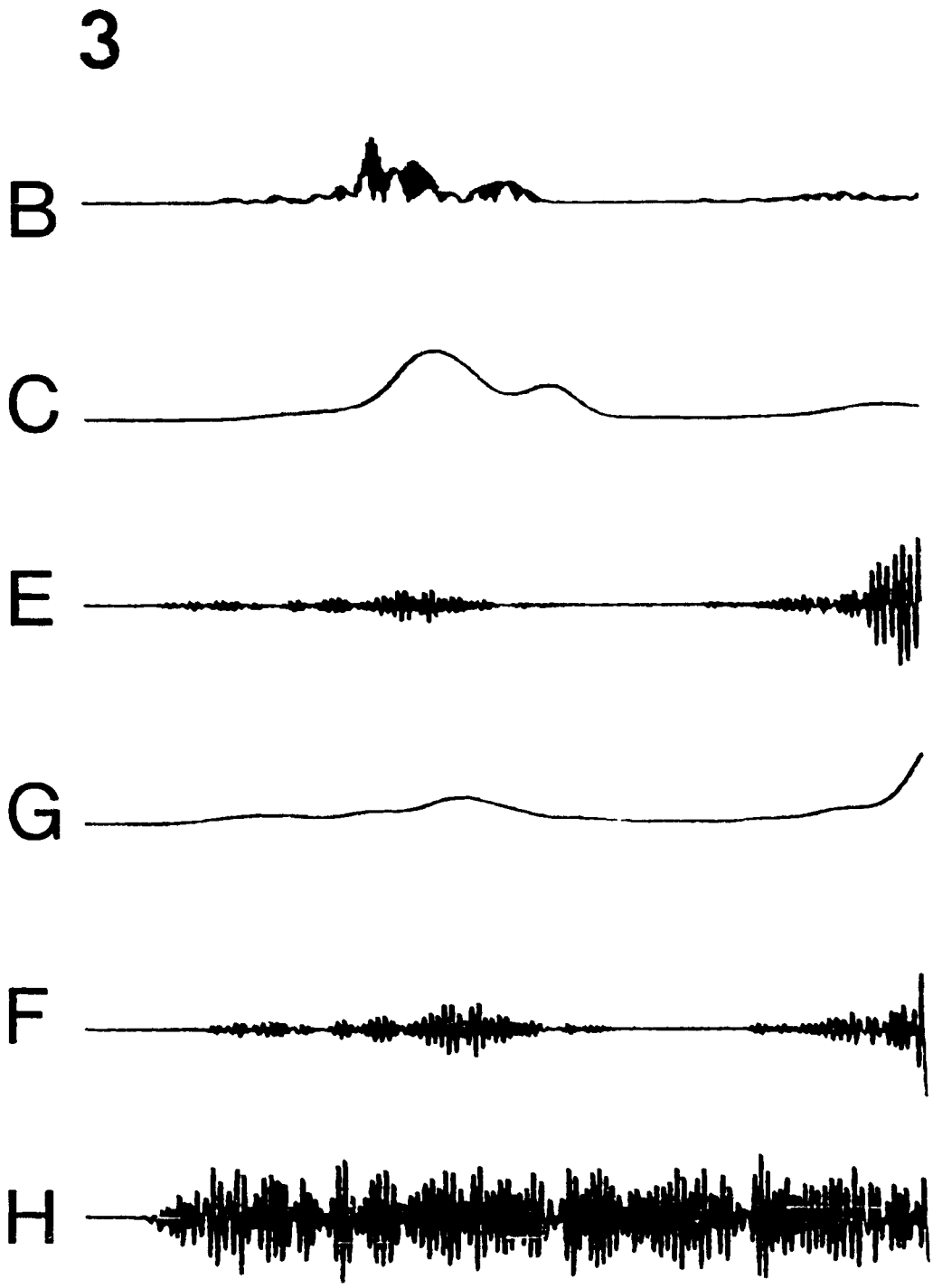
"G" is the output of the AGC's envelope detector.

"F" is a delayed version of "E".

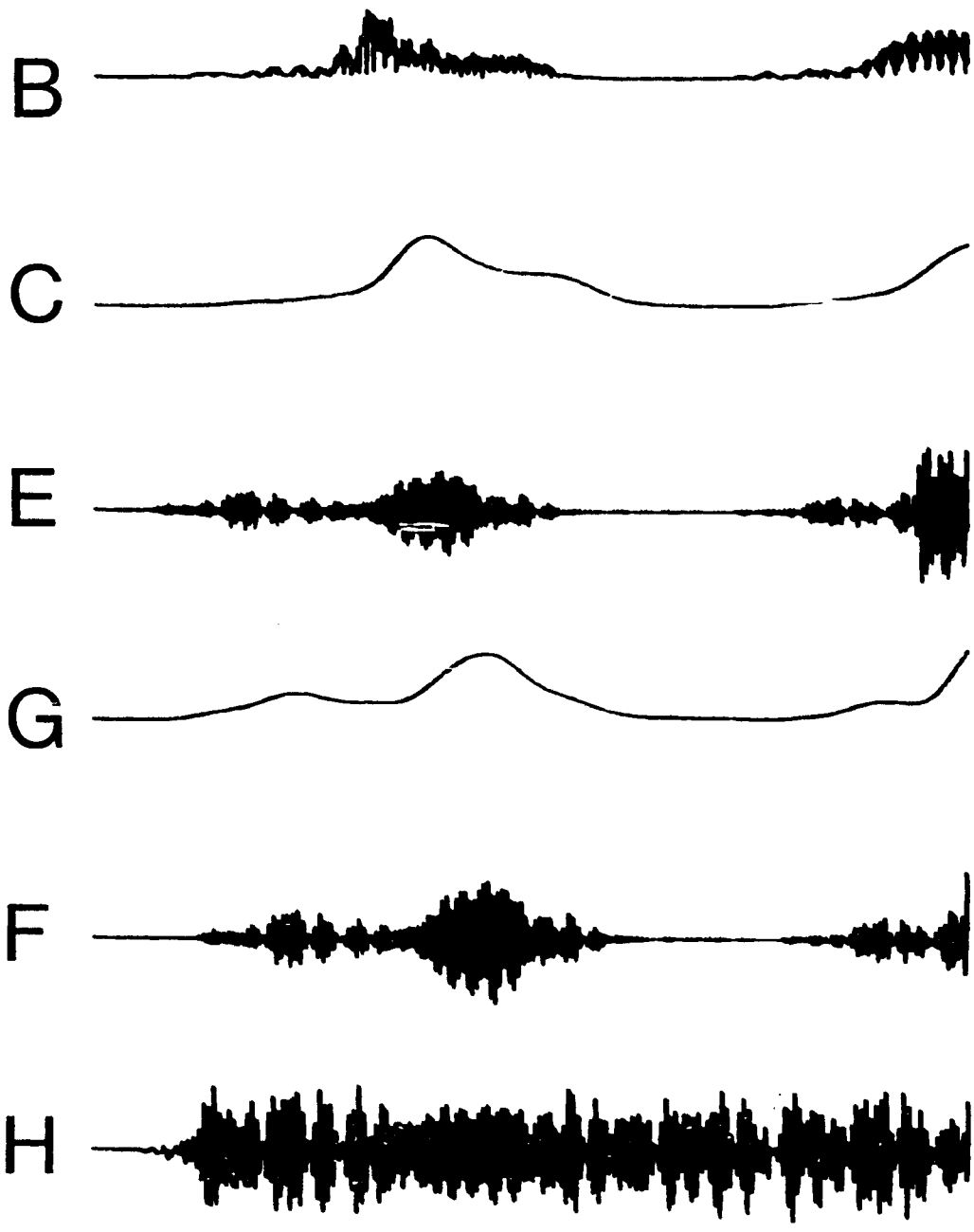
"H" is similar to "E" except that all slowly changing amplitude variations in "E" have been minimized by the AGC.


1
B
C
E
150 msec
G
F
H

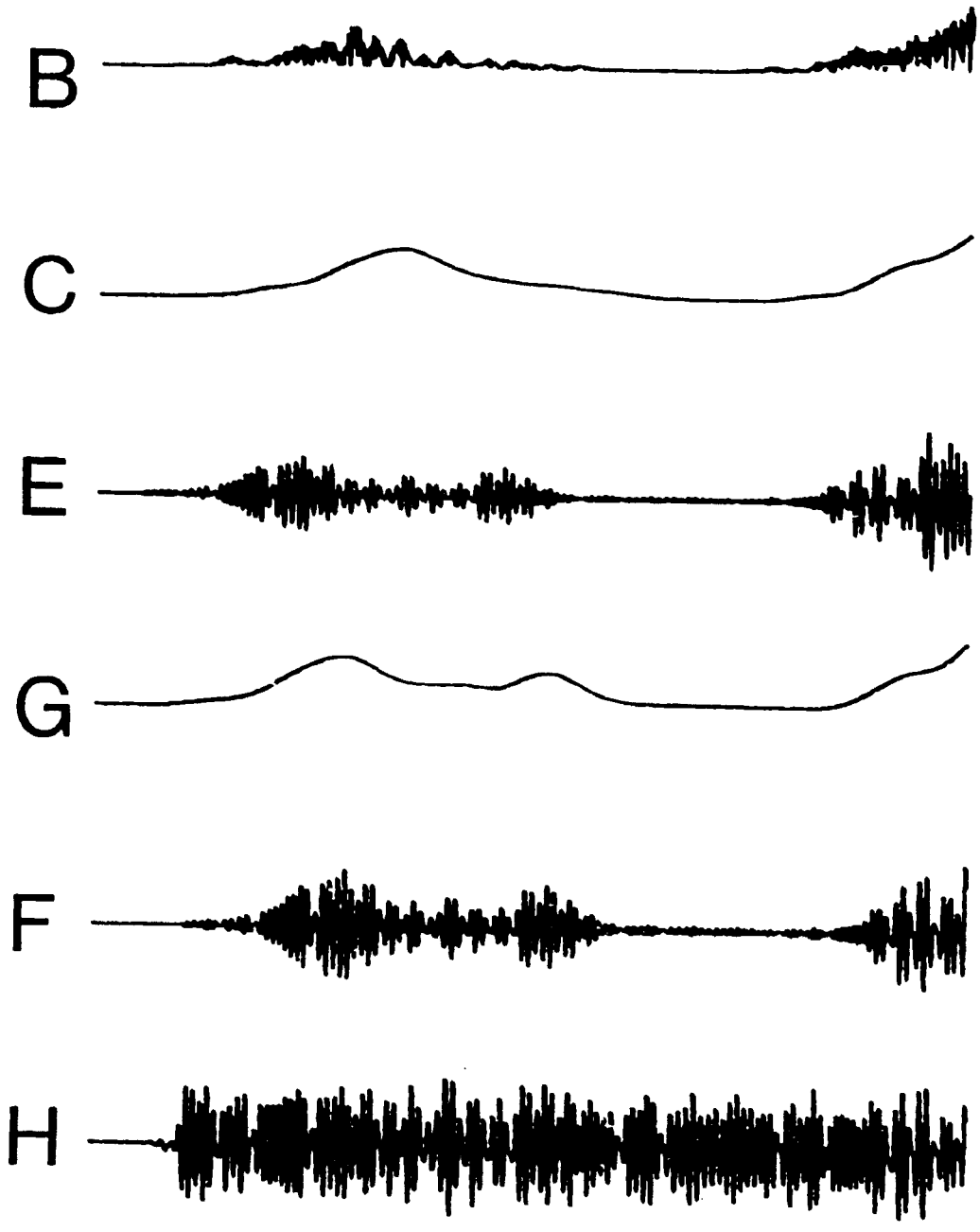

2
B
C
E
G
F
H


3
B
C
E
G
F
H

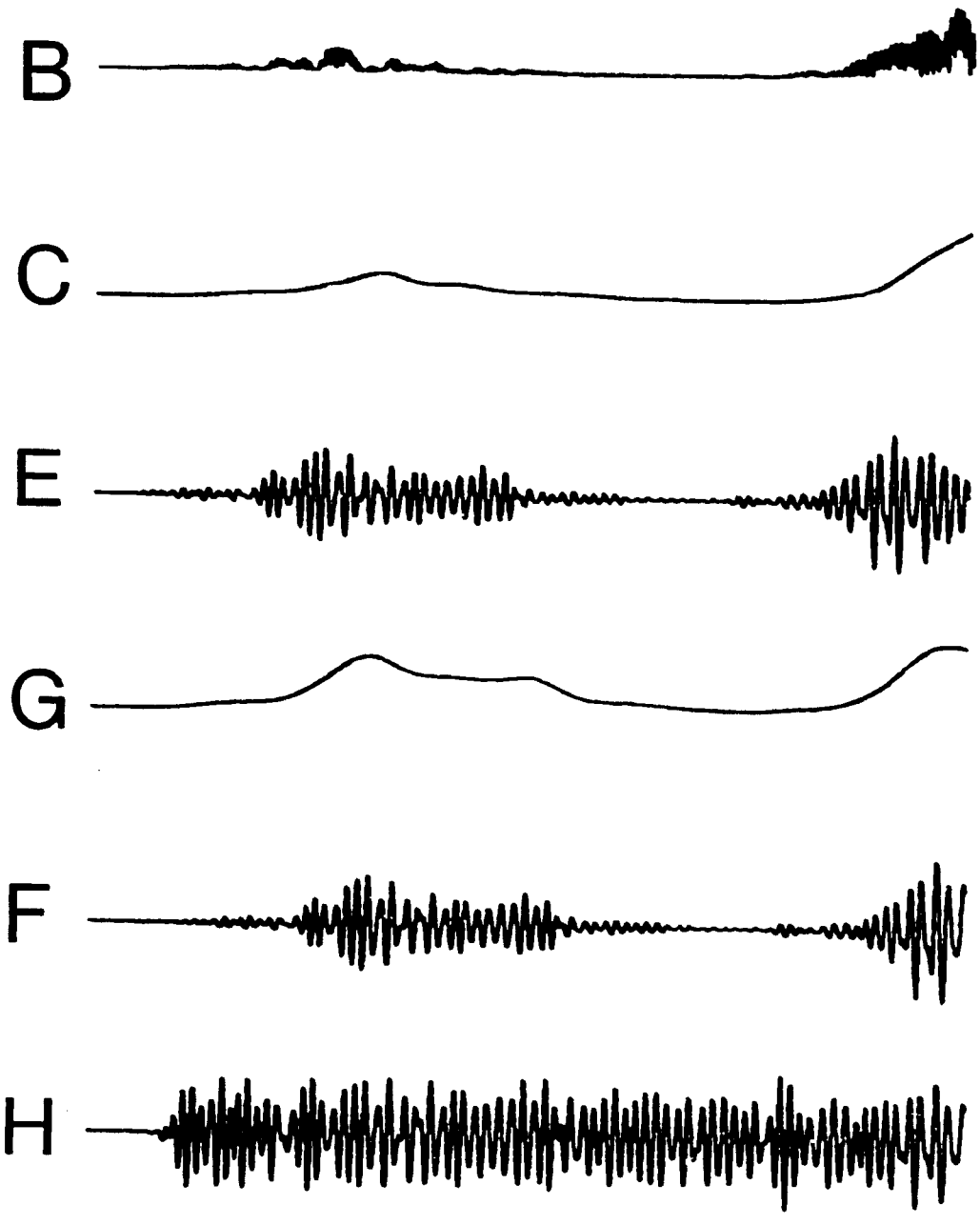
4



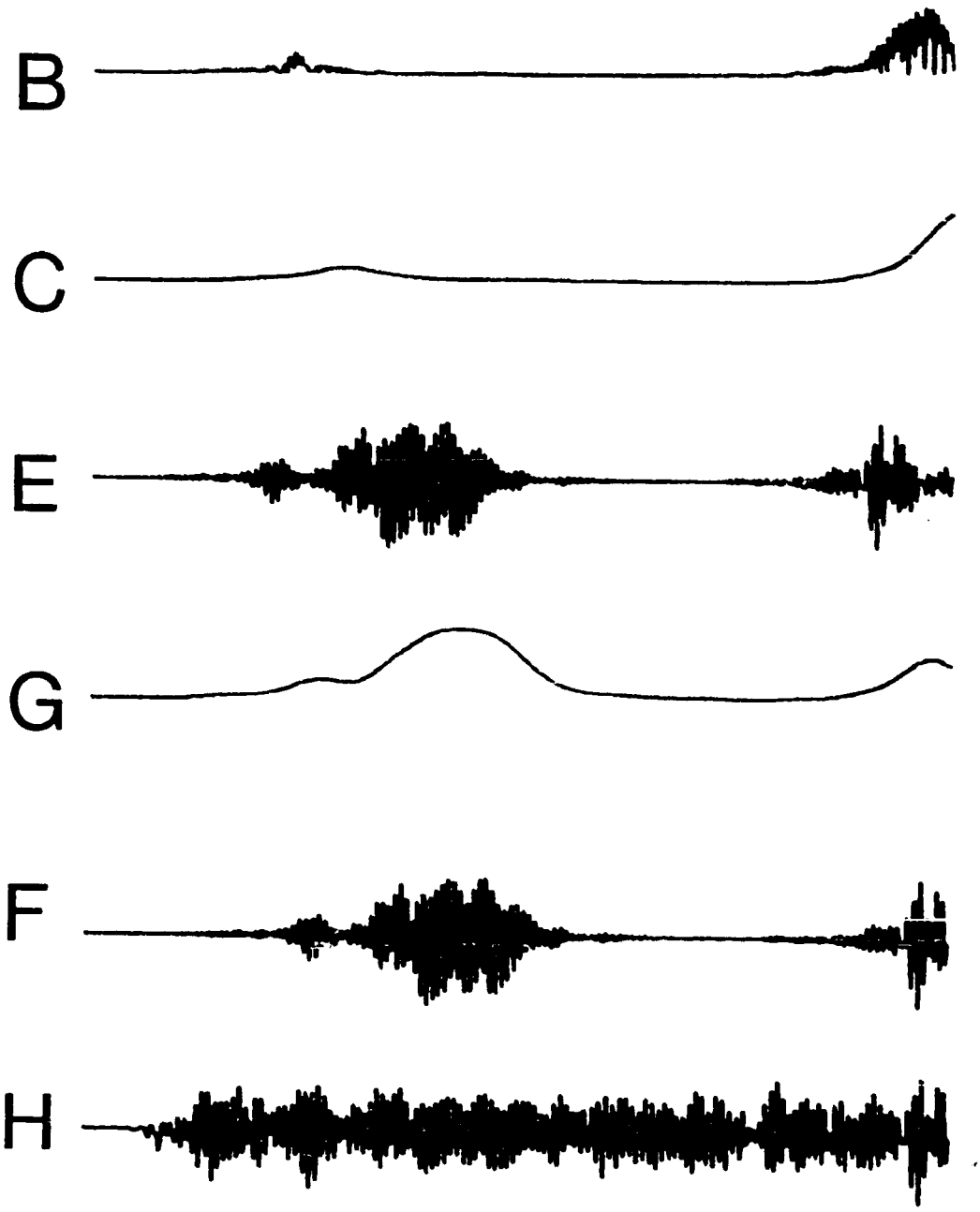
5



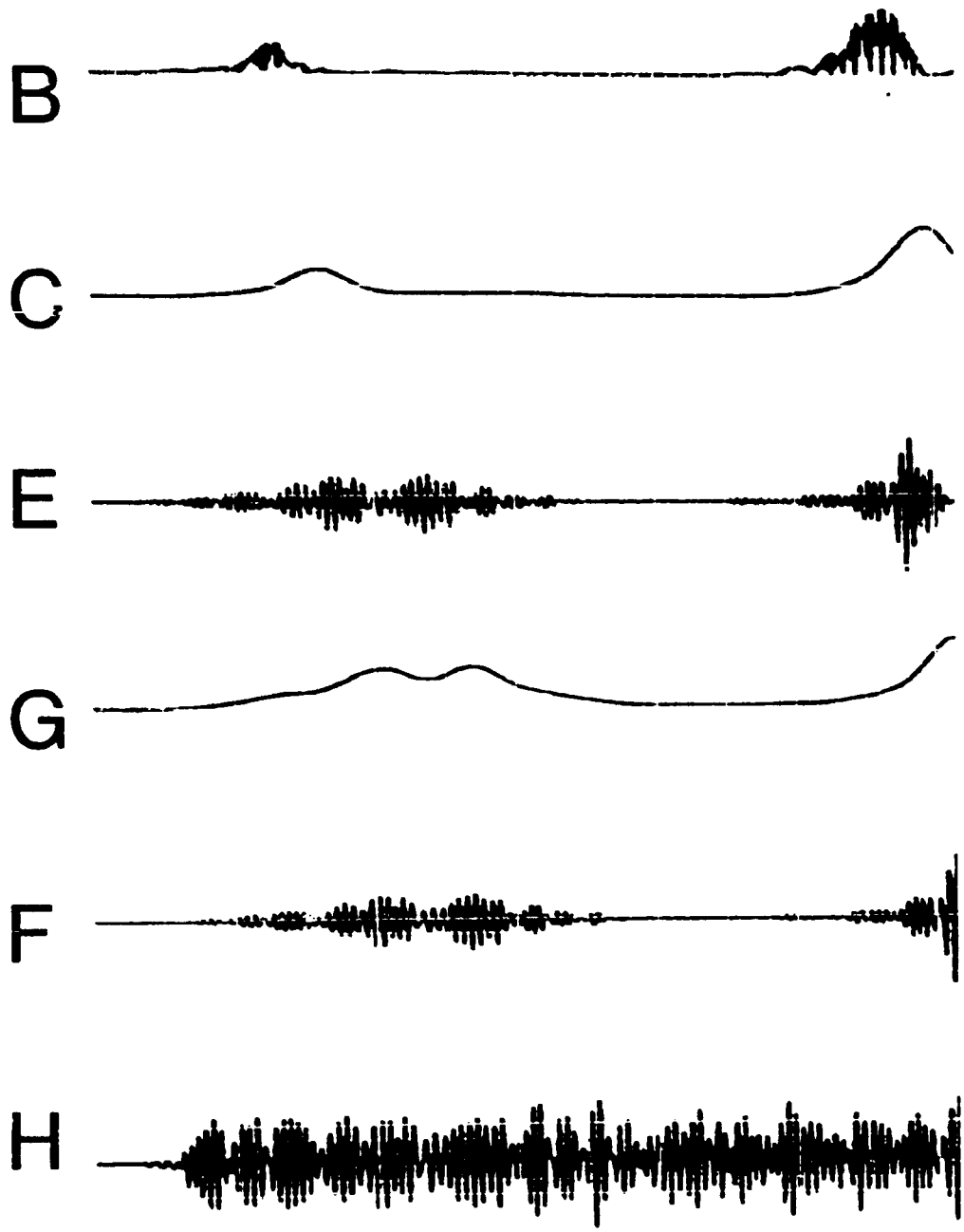
6



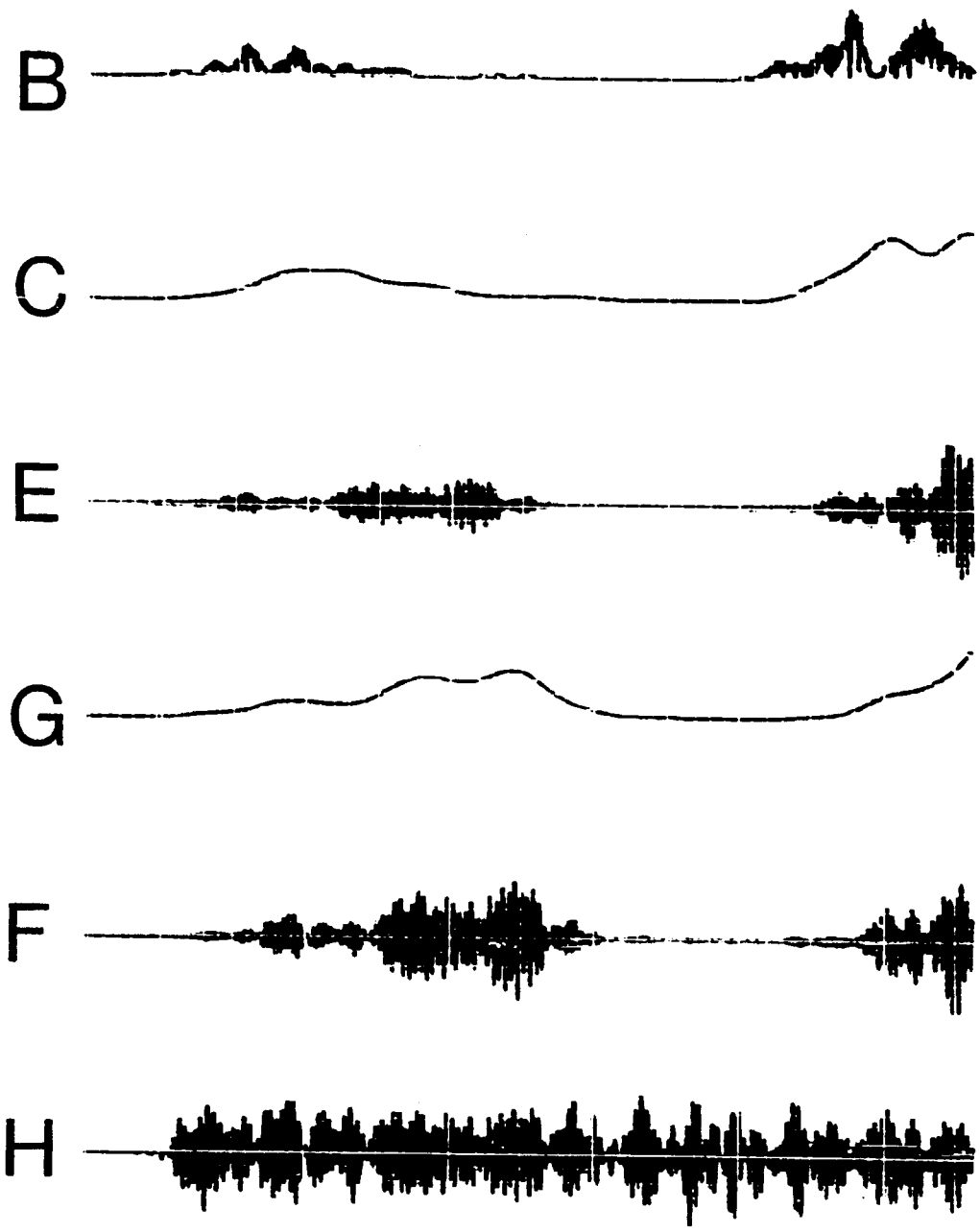
7



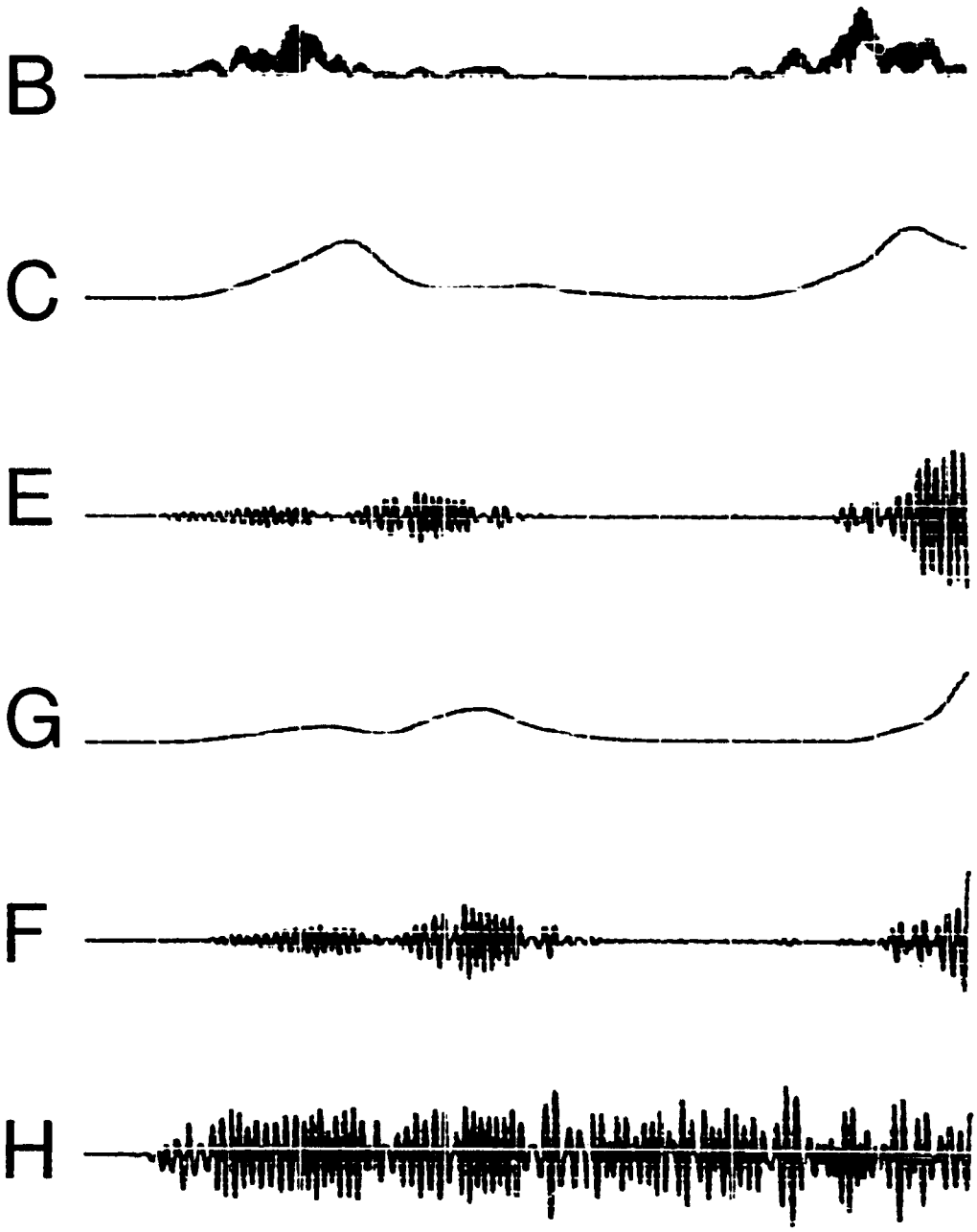
8



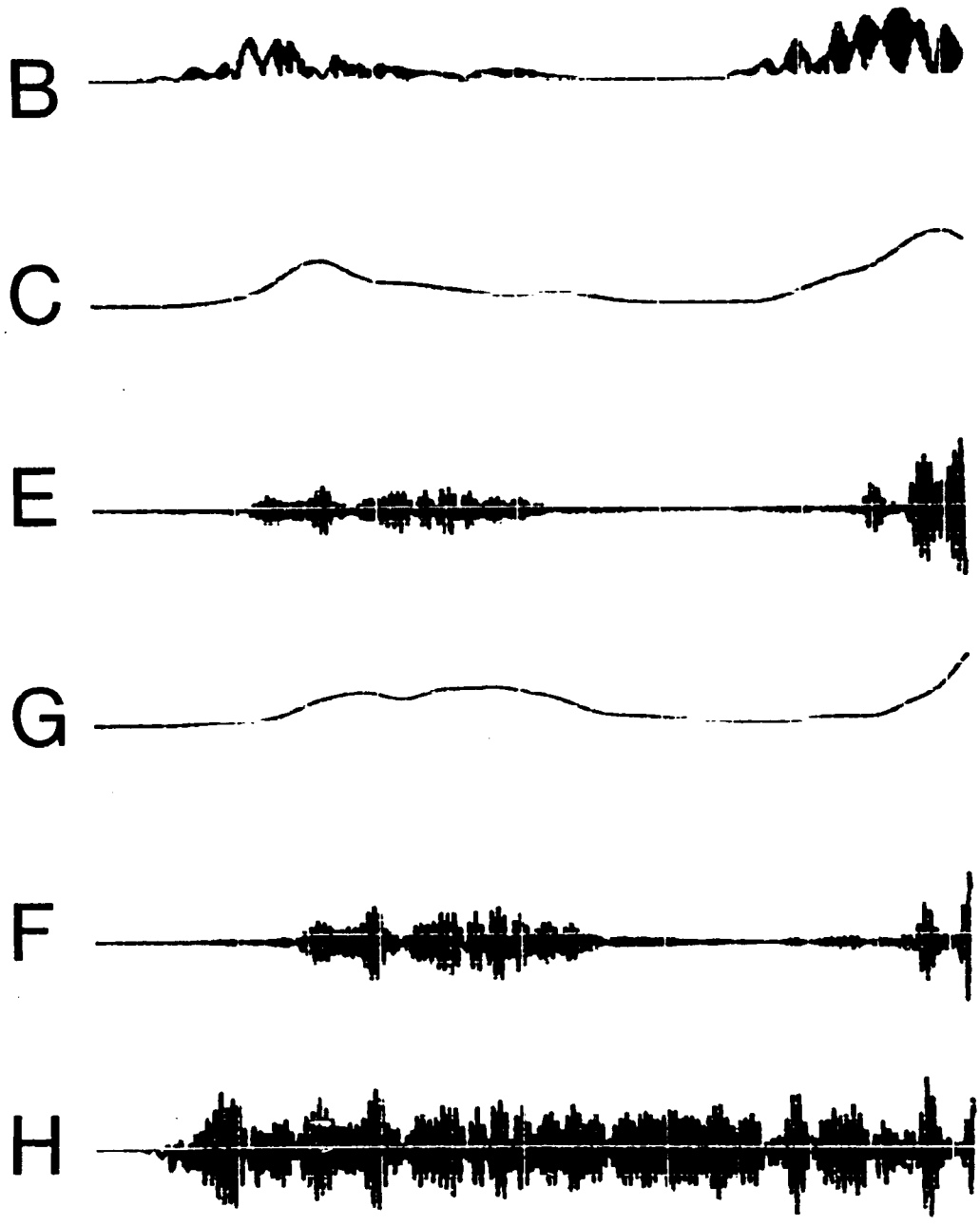
9



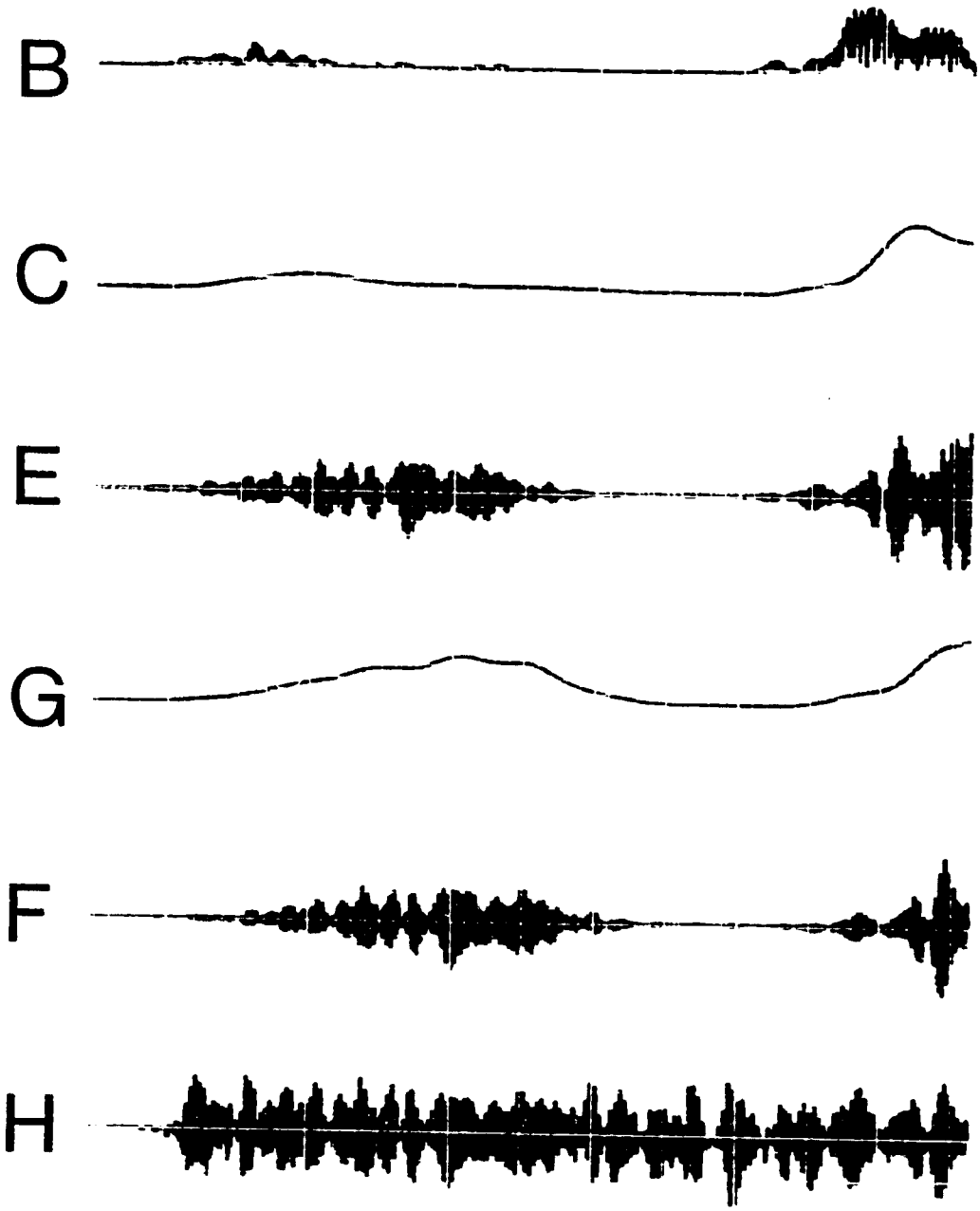
10



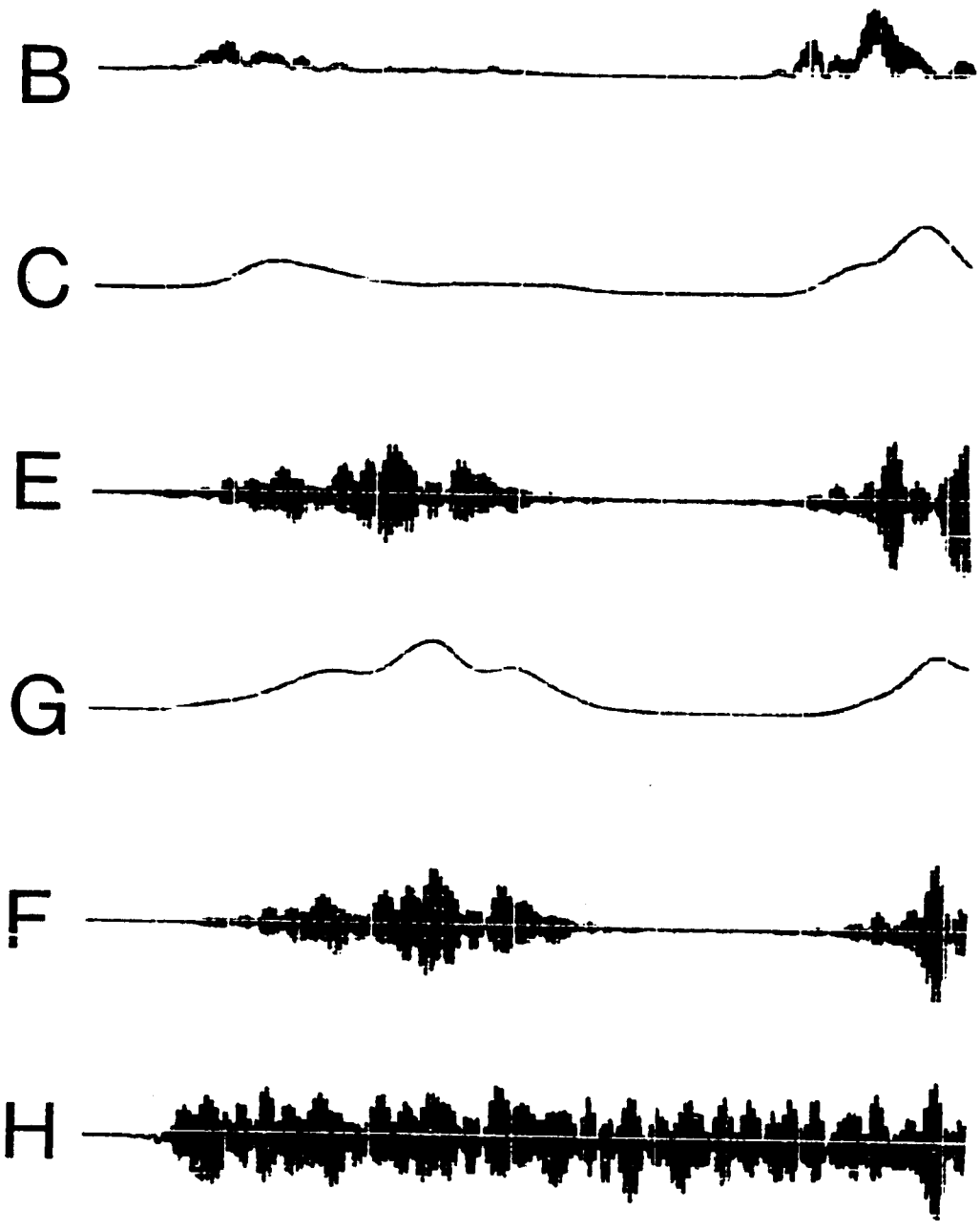
11



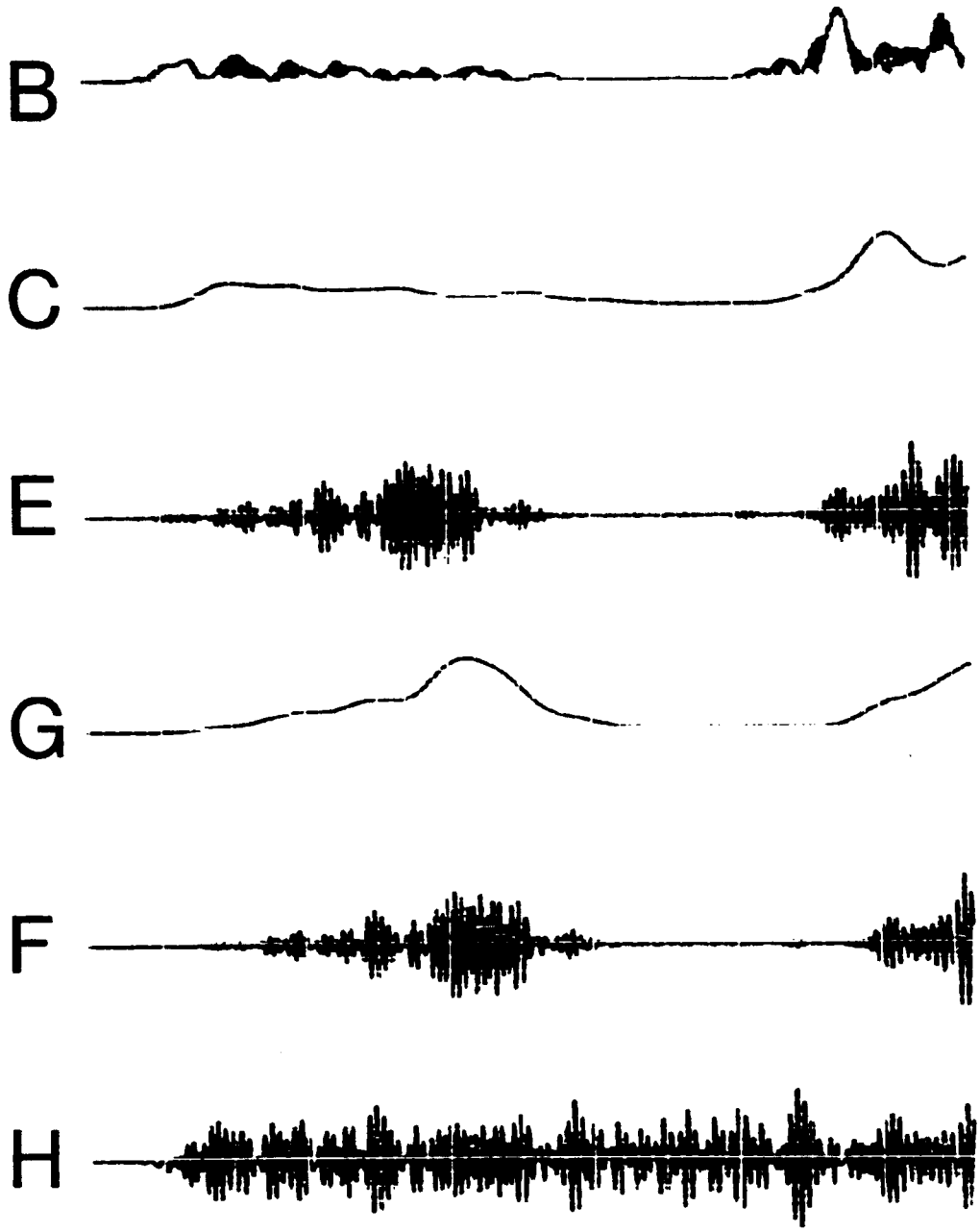
12



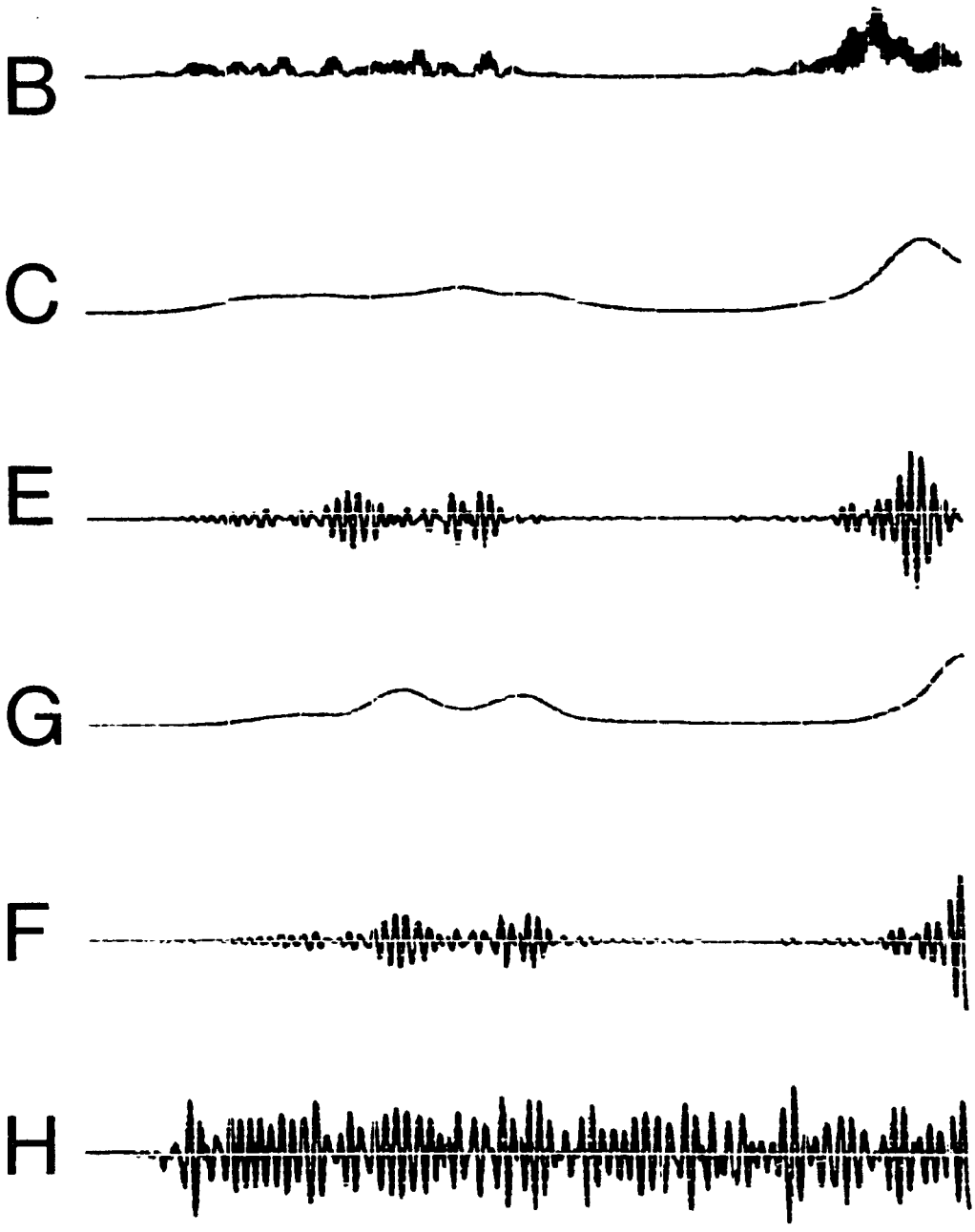
13



14



15



16

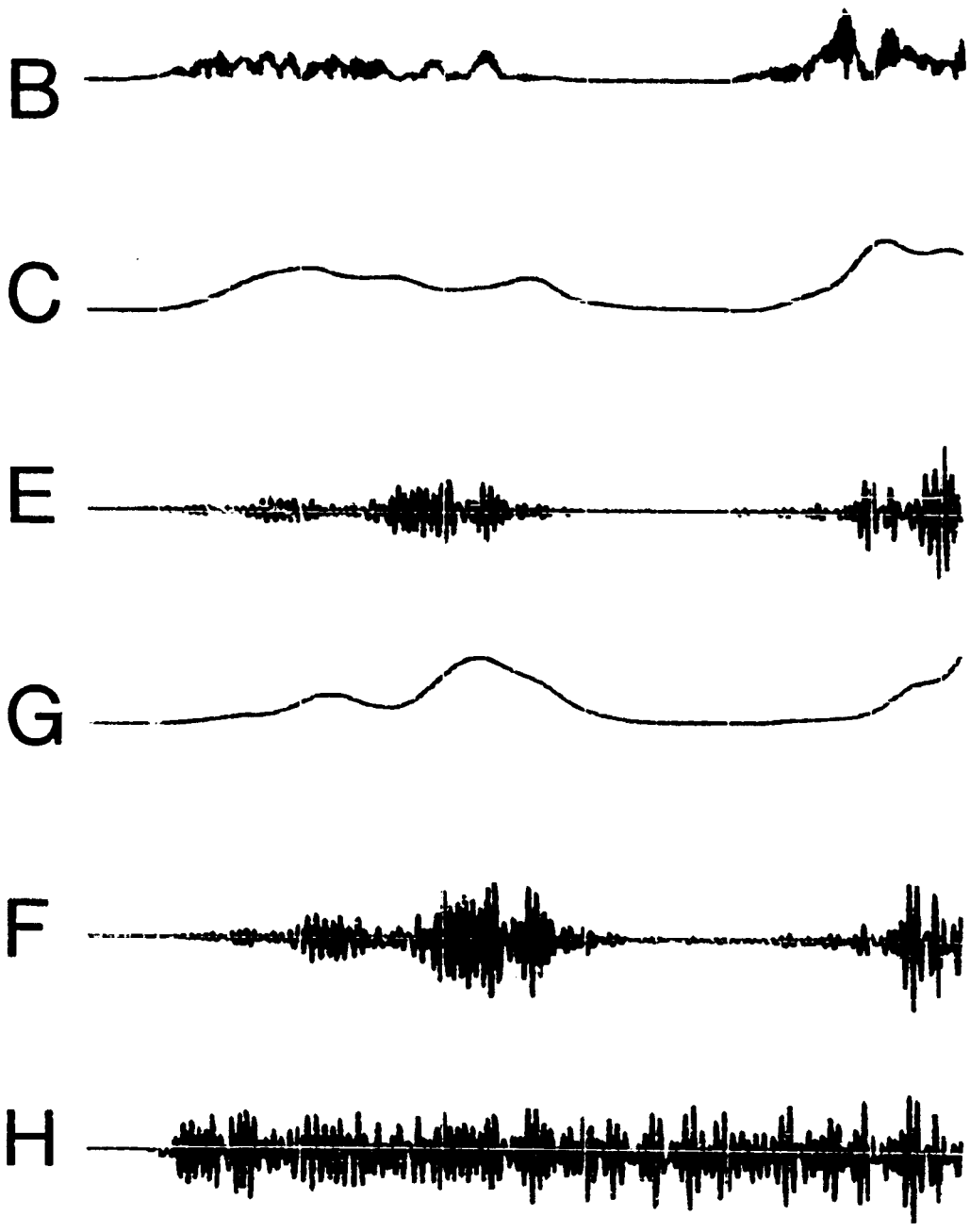


Table 4.4. Baseband, analysis, and synthesis filter characteristics for the constant channel bandwidth vocoder example.

Baseband filter characteristics:

Type of filter	= Kaiser window design, bandpass filter
Filter length	= 185 pts. (i.e. an 18.5 msec. impulse response)
Transition width (delta F)	= 150 Hz.
-6 dB frequencies	100 Hz. and 900 Hz.
Stopband ripple	= passband ripple = 53.1 dB down

Analysis and synthesis filter characteristics:

In this vocoder example, each channel's synthesis bandpass filter is identical to its analysis bandpass filter.

Input sampling rate	= 10 KHz.
Type of filter	= Kaiser window design, bandpass filter
Filter length	= 185 pts. (i.e. an 18.5 msec. impulse response)
Transition width	= 150 Hz.
Buffer zone	= 100 Hz.
Stopband ripple	= passband ripple = 53.1 dB down

Table 4.4 continued:

The other parameters vary across channels. They are tabulated according to channel number:

Channel Number	Lower-6dB Cutoff (Hz)	Upper-dB Cutoff (Hz)	Unity Gain Frequency	Decimation Ratio	Decimation Ratios Meeting Non-aliasing Criteria
1	900	1050	975	7	8,7,4,3,2
2	1050	1200	1125	6	7,6,3,2
3	1200	1350	1275	6	6,5,3,2
4	1350	1500	1425	5	5,2
5	1500	1650	1575	5	8,5,4,2
6	1650	1800	1725	5	7,5,4,2
7	1800	1950	1875	7	7,4,2
8	1950	2100	2025	6	6,4,3,2
9	2100	2250	2175	6	8,6,4,3,2
10	2250	2400	2325	5	5,3
11	2400	2550	2475	5	9,7,5,3
12	2550	2700	2625	5	5,3
13	2700	2850	2775	6	8,6,4,3,2
14	2850	3000	2925	6	6,4,3,2
15	3000	3150	3075	6	9,6,4,3,2
16	3150	3300	3225	7	7,4,2

Table 4.5. Lowpass filter characteristics for the constant channel bandwidth vocoder example.

The analysis and AGC lowpass filters are identical.

Type of filter = Hamming window

Impulse response duration = 33 msec.

Cutoff frequency = 39 Hz. (-3 dB)

Minimum attenuation in stopband = 42 dB (at 90 Hz.)

Decimation ratio = 1

The other parameters vary across channels. They are tabulated according to channel numbers:

Channel Number	Sampling Rate Into Lowpass (Hz.)	Number of Filter Coefficients
1	1428	47
2	1666	55
3	1666	55
4	2000	67
5	2000	67
6	2000	67
7	1428	47
8	1666	55
9	1666	55
10	2000	67
11	2000	67
12	2000	67
13	1666	55
14	1666	55
15	1666	55
16	1428	47

Table 4.6. Output interpolating bandpass filter characteristics for the constant channel bandwidth vocoder example.

Output sampling rate = 10 KHz.

Type of filter = Kaiser window design, bandpass filter

Number of filter coefficients = 185

Impulse response duration = 18.5 msec.

Transition width = 150 Hz.

Stopband ripple = passband ripple = 53.1 dB

Buffer zone = 0 Hz.

The other parameters vary across channels. They are tabulated according to channel number:

Channel Number	Lower-6 dB Cutoff (Hz.)	Upper-6 dB Cutoff (Hz.)	Unity Gain Frequency	Interpolation Ratio
1	800	1150	975	7
2	950	1300	1125	6
3	1100	1450	1275	6
4	1250	1600	1425	5
5	1400	1750	1575	5
6	1550	1900	1725	5
7	1700	2050	1875	7
8	1850	2200	2025	6
9	2000	2350	2175	6
10	2150	2500	2325	5
11	2300	2650	2475	5
12	2450	2800	2625	5
13	2600	2950	2775	6
14	2750	3100	2925	6
15	2900	3250	3075	6
16	3050	3400	3225	7

Chapter 5

Stimulus Delivery System

Definition of the optimum speech processing format for gaining maximum intelligibility with a given kind of multielectrode array requires a versatile speech processing, stimulus sequencing, patient response analysis, and stimulus delivery system. This system will ultimately be employed in an extensive series of psychophysical experiments conducted in a few carefully selected patients. In these experiments, the consequences of changing design parameters of processor models can be systematically evaluated. A consideration of speech processing methods suitable for this testing was presented in the previous section. Development of stimulus sequencing and response analysis procedures has been initiated in this laboratory. The sequencing and response analysis considerations will not be discussed in this dissertation.

Much of the test material will first be recorded on audio tape. Test material will then be processed and delivered to the patient in a manner appropriate for the particular test protocol. In many cases the processed material, ready for presentation, will be stored on a digital magnetic tape or a disk cartridge for future patient testing. Laboratory equipment capable of converting processed and stored test material into currents flowing through the patient's electrode array has been designed. All of the delivery system molules have been constructed and tested. Final system integration awaits delivery of the computer system dedicated to patient testing.

The initial patients will wear a temporary percutaneous pyrolized carbon electode connector (Mooney et al, 1977). This connector will enable an appropriate set of stimulus processing techniques to be tested. After this initial period, a transcutaneous receiver (Gweelala et al, 1976) will be implanted and the connector removed.

The stimulators are controlled by a programmable computer interface. This stimulator and interface system is extremely safe. The stimulators are presently being applied in experiments using animals. The stimulator and interface system enables the testing of a variety of stimulus waveforms; does not limit the implementation of the numerous models of appropriate speech processing; contains multiple hardware safeguards against inadvertent overstimulation; is capable of delivering uninterrupted high transfer rate multichannel stimuli of very long duration and is capable of accepting a variety of stimulator control coding conventions such that stimulus storage requirements may be minimized for sets of processing options. The stimulators may be used independently of the interface. For example, a prototype hardware speech processor can be substituted for the interface.

Stimulators

The output current of the stimulator closely approximates the waveshape of the control voltage input. Maximum stimulator output current is adjustable from 50 to 3000μa. (see Figure 5.1.) Measured output risetime is 15 μsec.

To limit the maximum charge deliverable and to maintain a good dynamic range, it is noted that the neuron's excitability is a threshold function of stimulus charge over most of the normal parametric range of stimulation (see Fig. 4.2.). In other words, for most practical stimulus waveforms (e.g. 100-2000μsec biphasic pulses), a nerve fiber will be excited after a specific amount of charge is delivered regardless of the waveform. This approximation is in agreement with models of neural excitability vs. electrical excitation (Hill, 1936a). Empirically, the charge concept holds remarkably well over the practical stimulus range (Fig. 5.2). However, with both electrodes placed extracellularly near a myelinated nerve, the stimulus intensity vs. pulse duration curve theoretically deviates slightly from the simple transmembrane stimulation model of Hill (McNeal, 1976). This relatively inconsequential effect is largely due to the current's path through the finite axonal resistance.

If the stimulator is in the controlled current mode, integrating the stimulus control signal before it enters the saturatable, but normally linear element "U4" and then differentiating it after the signal passes through this element, sets an upper limit on charge delivered and still maintains a "flat" frequency response over the linear range of the stimulator. This technique has been utilized in the stimulators to preset the maximum charge deliverable, irreguardless of the stimulator's input.

Figure 5.1. Circiut Schematic of the Stimulator

Functional description for each stage:

Stage 1. This stage incorporates operational amplifier U1. The major function of this stage is to reject common mode components that may have been introduced in the input cabling and to reduce the effect of ground noise.

Stage 2. This stage incorporates U6. The major function of this stage is amplification and filtering. The integration, referred to in the text, is approximated in this stage.

Stage 3. This stage incorporates U2, U3, U4, and Q1. This stage is the optical isolation stage. It incorporates a feedback scheme in order to linearize the response characteristics of the optical isolator (Rutkin, 1975; Winston, 1975). This feedback also improves the frequency response of the stage.

Stage 4. This stage incorporates U5, the high voltage operational amplifier. This stage is the output stage and functions as a voltage to voltage amplifier or a voltage to current amplifier. In most stimulation work it is used as a voltage to current amplifier. A high pass filter is incorporated to approximate the differentiation referred to in the text. The battery supply can be varied from ±25 to ±100 volts. The electrode leads are relay switched. The electrode leads may be disconnected from the stimulator in this manner. The reeds are controlled by a patient activated switch. These reeds are also controlled by power monitoring circuitry, such that the stimulator will be disconnected from the electrodes before any minor, power-down transient current may enter the electrodes. Also, on "power-up", the reed relays are not closed until transients have subsided.

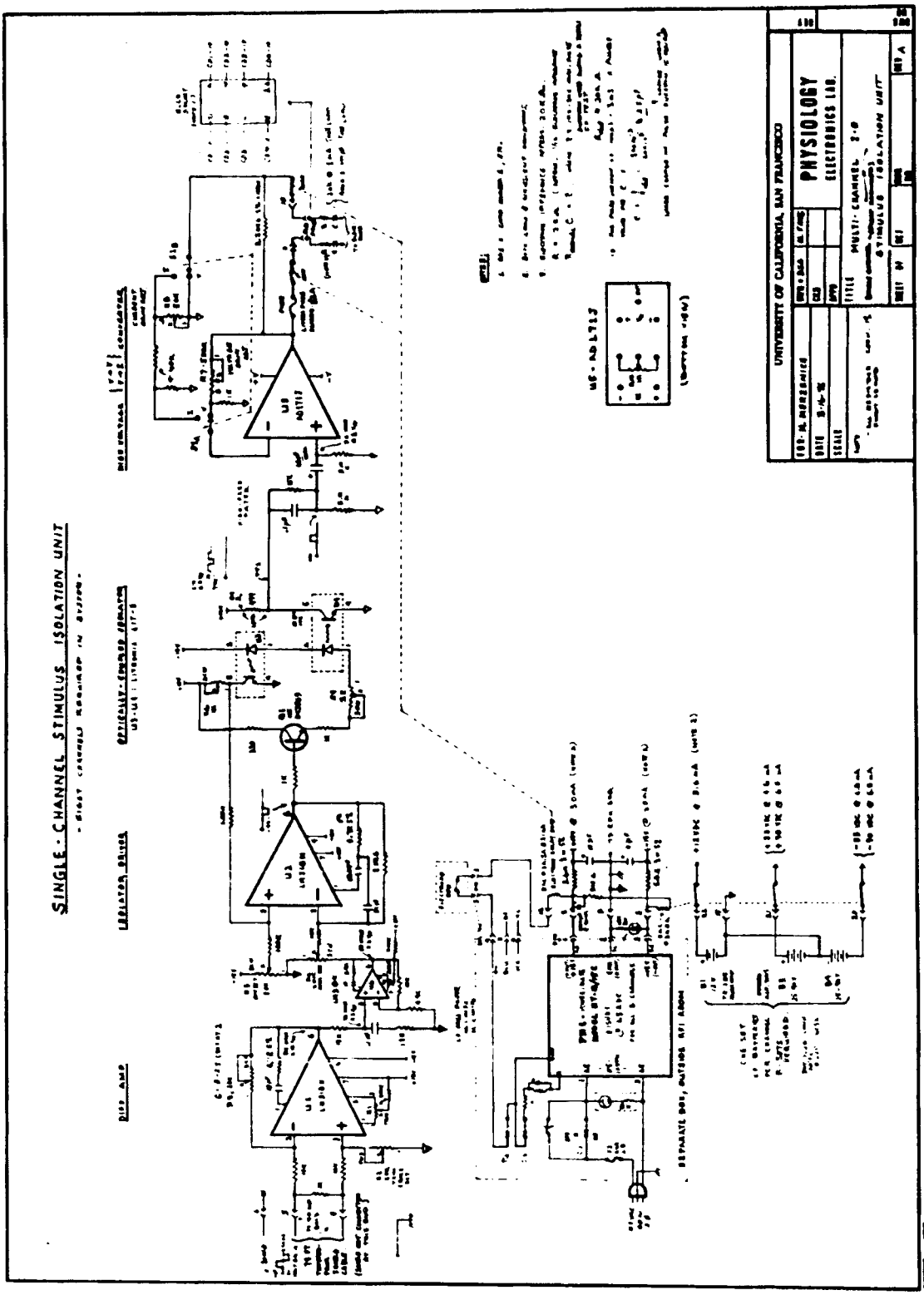
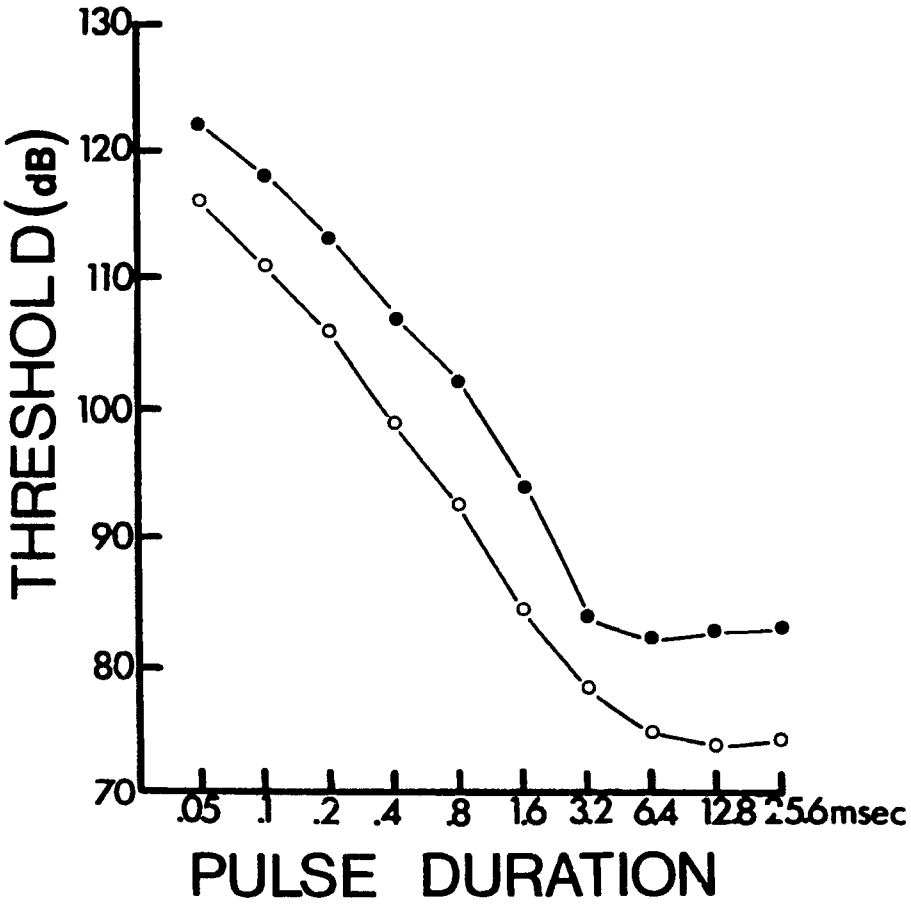

SINGLE-CHANNEL STIMULUS ISOLATION UNIT
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
PHYSIOLOGY
ELECTRONICS LAB.

Figure 5.2. Strength-duration curves for electrically stimulated auditory neurons derived from determination of threshold current as a function of pulse duration, for two neurons within the central nucleus of the inferior colliculus (105 dB = 100 μa). All strength-duration data derived (for monopolar electrodes, in different cats) were similar to those shown. Neurons most sensitive to stimulation were used to obtain the data. For the range of .05 to 3.2 milliseconds there is an approximately inverse relationship between pulse duration and threshold current. For this range, the neuron will be excited once the threshold charge has been delivered.

These strength-duration curves exhibit chronaxies on the order of 2 msec. Cell bodies and initial segments of proximal dendrites commonly display chronaxies greater than 1 msec. (BeMent and Ranck, 1969a).


130
120
110
100
90
80
70
THRESHOLD(dB)
.05
.1
.2
.4
.8
1.6
3.2
6.4
12.8
25.6msec
PULSE DURATION

Stimulator safety considerations

Stimuli are generated using battery powered photo-coupled isolators with current or voltage as the controlled output variable. The isolators are capacitively coupled to the electrodes. There is also an output fusing device. No physiologically significant leakage through the patient to other circuits is possible using low leakage (10^{11} ohms), low capacitance (1pfd.), high breakdown voltage (2,500 volts) photo-isolators and good construction techniques. Very low net charge is transferred (much less than .01% of maximum charge delivered per pulse phase per interpulse interval) during long term stimulation. This level is considered to be well within platinum dissolution limits, if the maximum current and densities per pulse phase do not exceed proper levels (Brummer, 1977; White, 1974). Current limiting resistors, dual output blocking capacitors of low capacitance and low leakage, isolator output fusing, variable voltage battery supply, electrode disconnect relays at the stimulator-electrode junction, careful experimental design and thorough on-going equipment testing virtually eliminate the possibility of stimulating above safe levels. Indeed, in the very unlikely event that three worst-case components fail simultaneously, these precautions eliminate the possibility of stimulating above the limits set for long term stimulation using conservative metal dissolution and oxidation criteria. These limits are considerably more stringent than any other stimulus limits that pertain to patient safety.

Programmable Stimulator Interface

The stimulator interface converts encoded and stored test material into the voltages that control the stimulator's output currents. The interface is easily expanded from eight to sixteen channels. With the present disk device, the upper limit on transfer rate is 800,000 bits/sec. This is more than adequate for all speech processing models proposed. However, if a higher transfer rate and/or a longer stimulus duration is desired, only an improved disk system would be required. No changes in the other system components or associated software would be required.

Interface Design

The interface is divided into two parts: 1) A Data General Corp, programable data control unit (DCU) with direct memory access to CPU memory. The DCU's instruction set is identical to the CPU's. Also, the DCU's I/O buss is identical in design to the CPU's I/O buss. The DCU contains a 1024 word bipolar memory. All effective addresses below 1024 reference this local high speed memory. Effective addresses above 1023 reference the main CPU memory just as though the DCU was accessing its own memory. Consequently, this module is a fast, very versatile, relatively inexpensive, programable, direct memory access interface. 2) A locally constructed digital to analog converter and demultiplexer (DA-DEMUX). The DA-DEMUX board converts the decoded (decoded by the DCU) digital numbers to analog voltages and routes these voltages to the specified stimulators.

These two modules, in conjunction with the stimulators, the

central processing unit (CPU) and the disk memory will generate high data rate, continuous multichannel stimuli of long duration (see Fig. 5.3 and 5.4). In this output system the DCU is the initiator of all operations. When a processed speech segment is to be delivered, the DCU interrupts the CPU, causing the CPU to load CPU memory buffers "A" and "B" with the first and second segments, respectively, of the stored and coded stimulus on the disk. Once loading is completed, the DCU begins real-time output operations on the coded words in buffer "A". The DCU retrieves, decodes, and sends the decoded information to the DA-DEMUX board for digital to analog conversion and routing to the proper stimulator channel. Once buffer "A" has been converted, the DCU commands the CPU to load buffer "A" with the third segment from disk while the DCU is decoding and transferring buffer "B". Buffers "A" and "B", alternately, are loaded and unloaded until the entire stimulus has been generated. Because the DCU is fully programmable, optimized, "memory saving" encoding for each processing option is easily implemented.

At the 800,000 bit/sec maximum rate, one inefficient form of stimulus coding (i.e. pulse code modulation) would allow a transfer rate of 12,500 samples/sec (at 8 bits per sample) for each channel of an eight channel delivery system. The system is easily expandable to 16 channels, with a proportionate decrease in transfer rate per channel.

More efficient coding schemes will be used in most cases. One such coding technique would deliver: 1) A constant waveshape (e.g. a biphasic pulse waveform), 2) an onset resolution of 100 microseconds,

3) an amplitude resolution of up to 12 bits, and 4) the capacity to simultaneously stimulate multiple logical channels. Such a coding scheme requires approximately two 16-bit words per waveform generated. If one assumes an average of 100 p.p.s. for each of 50 channels, a transfer rate of 160 Kbits/sec. is required. This represents a considerable savings in storage requirements over the pulse code modulation technique. It also reduces I/O overhead on a time-sharing operating system. Also, in most processing models, silence intervals within the speech segments would require a much lower transfer rate. This coding method allows one to specify the logical channel number, the stimulus amplitude, time of stimulus onset, and allows multiple (2-3) logical channels to be stimulated simultaneously. All logical channels are defined by software routines before stimulation begins. Each logical channel is defined by specifying the portion of stimulating current that is to pass through each one of the 16 electrode contacts. For most logical channel specifications, only two to four nearby electrode contacts would be given non-zero proportions. For example, logical channel #30 might be defined by giving electrode contact #8 a +1 proportion, electrode contact #7 a -.5 proportion, and contact #9 a -.5 proportion. Thus, if a stimulus amplitude were specified at 100 μamp, electrode contact #8 would receive a +100 μamp biphasic stimulus pulse and contacts #7 and #9 would each receive a -50 μamp biphasic stimulus pulse.

If only one or a few channels need to be stimulated simultaneously, electrode switching could minimize the complexity of the

Figure 5.3. Block Diagram of a Stimulus Delivery System

Outline of Operation:

A. The CPU relinquishes control to the Data Control Unit (DCU).

B. The DCU commands the CPU to load CPU memory buffers "A" and "B".

C. After this loading, the DCU starts real time output operations. The DCU accesses words in buffer "A", decodes them, sends the decoded words, and commands to the Digital to Analog converter and Demultiplexor (DA-DEMUX). Figure 5.4 contains a block diagram of the DA-DEMUX.

D. The analog voltages generated by the D-A are routed to the specified stimulators by the demultiplexor.

E. After buffer "A" has been exhausted, the DCU commands the CPU to reload buffer "A" with the third segment from disk. The CPU then generates the commands necessary for this disk to CPU memory transfer. The transfer then starts.

F. Simultaneously with operations outlined in step "E", the DCU is accessing (by direct memory access) buffer "B", decoding, and outputing this information to the DA-DEMUX.

G. After buffer "B" has been exhausted (it is assumed that the disk transfer to buffer "A" referred to in step "E" has been completed before buffer "B" has been exhausted), the DCU commands the CPU to reload buffer "B" with the fourth segment of the coded stimulus from disk.

H. Simultaneously with step "G", the DCU is accessing buffer "A", decoding, and outputting this information to the DA-DEMUX.

I. Steps "E" through "H" repeat until the entire stimulus has been generated.

J. The DCU relinquishes control to the CPU.

K. The patient response is recorded and analyzed by the CPU.

L. The next stimulus is determined, either by the experimenter or by the CPU program and steps "A" through "L" are repeated.

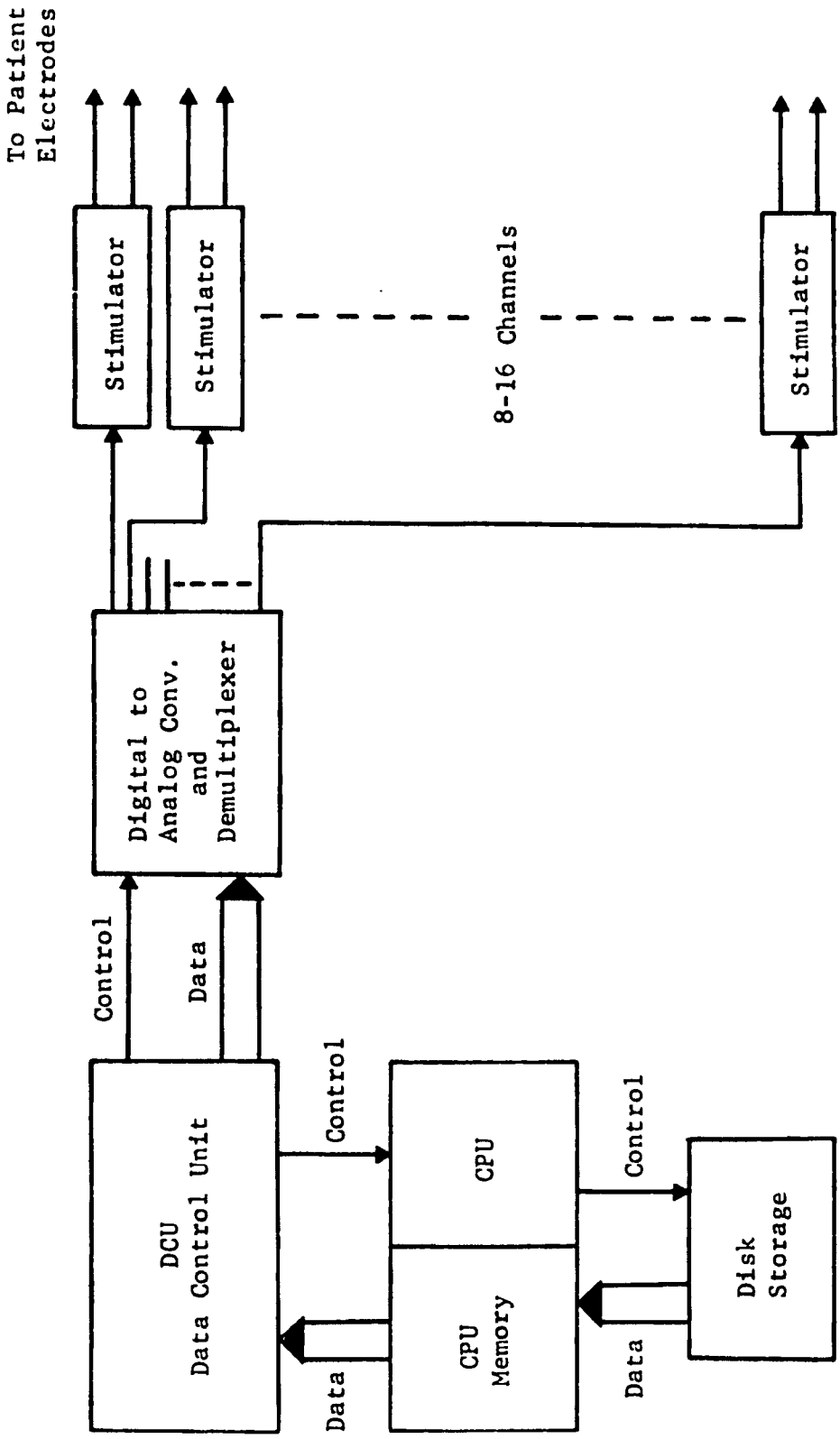

To Patient Electrodes
Stimulator
Stimulator
8-16 Channels
Stimulator
Digital to Analog Conv. and Demultiplexer
Control
Data
DCU Data Control Unit
Control
CPU
Control
Disk Storage
Data
CPU Memory
Data

Figure 5.4. Block Diagram of the D-A and Demultiplexor Interface

Outline of Operation:

A. Decoded data is placed on the DCU buss by the DCU.

B. Digital to analog converters convert this to analog voltages.

C. A control signal from the DCU causes one of the sample-and-holds of each set of 4, to sample the analog voltage generated by its D-A converter. The sample-and-holds are expandable to eight per set; 2 sets per interface.

D. This same signal toggles each set's 2-bit counter (expandable to a 3-bit counter per set) to update the sample-and-hold address.

E. This "round-robin" technique continues until the entire stimulus is delivered.

The circiut diagram of this module is presented in Figures 5.5 and 5.6.

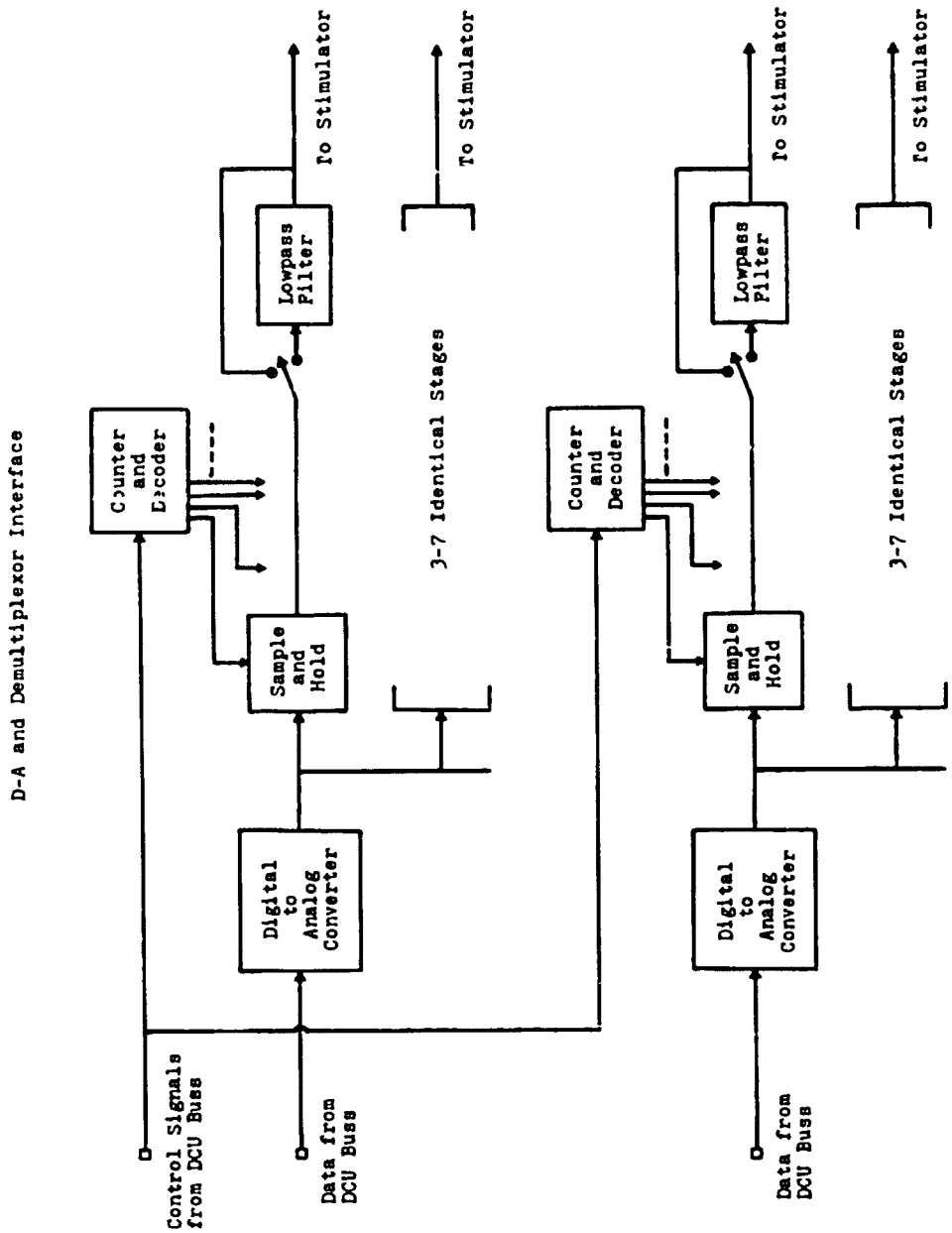

D-A and Demultiplexor Interface
Control Signals from DCU Buss
Data from DCU Buss
Digital to Analog Converter
Sample and Hold
Counter and Decoder
Lowpass Filter
To Stimulator
3-7 Identical Stages
To Stimulator
Digital to Analog Converter
Sample and Hold
Counter and Decoder
Lowpass Filter
To Stimulator
3-7 Identical Stages
To Stimulator
Data from DCU Buss

Figure 5.5. Schematic diagram of the digital to analog and demultiplexor circuitry.

An expanded schematic of the sample-and-hold circuitry (i.e. blocks labelled "S & H") is presented in Figure 5.6.

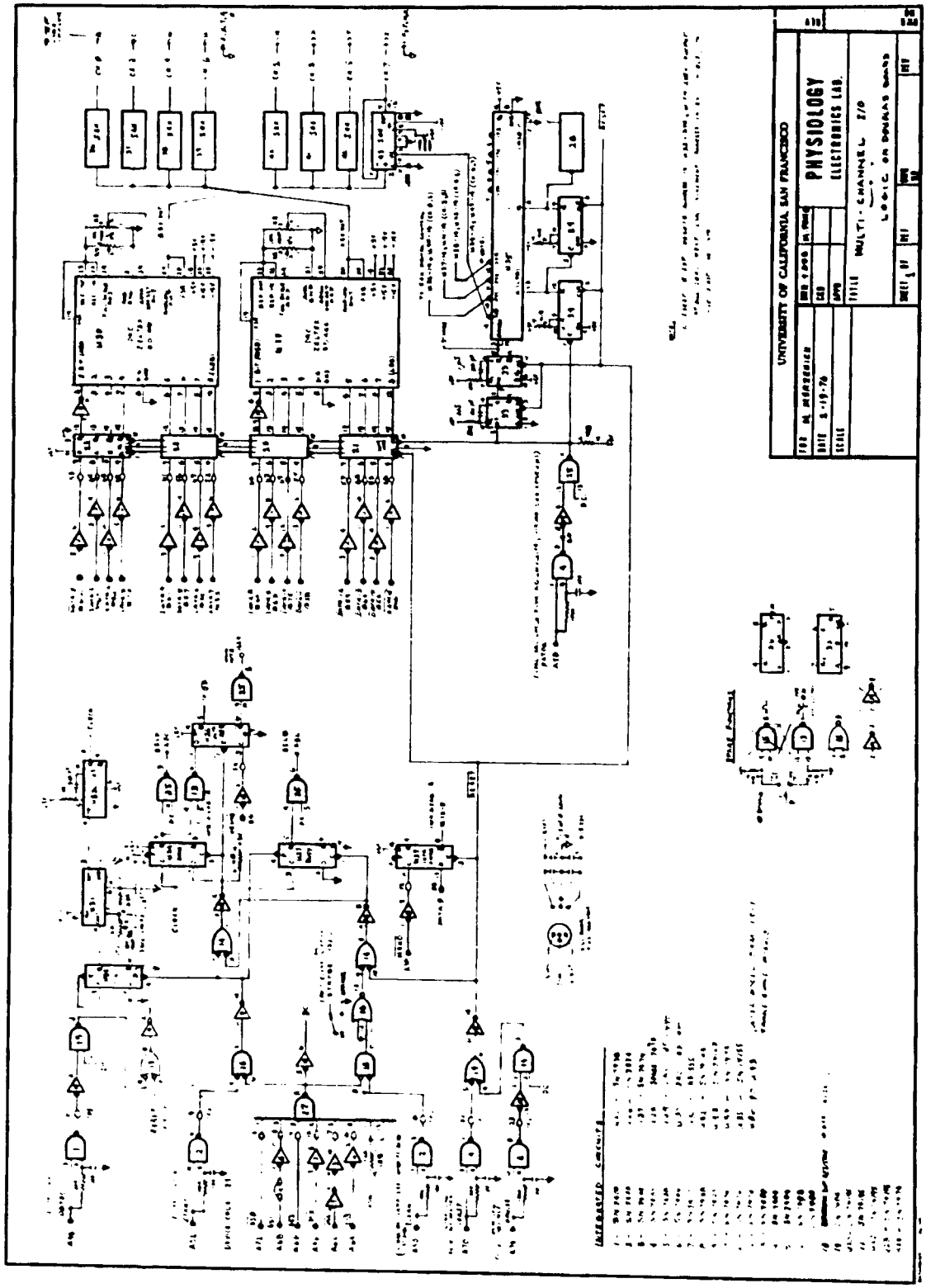
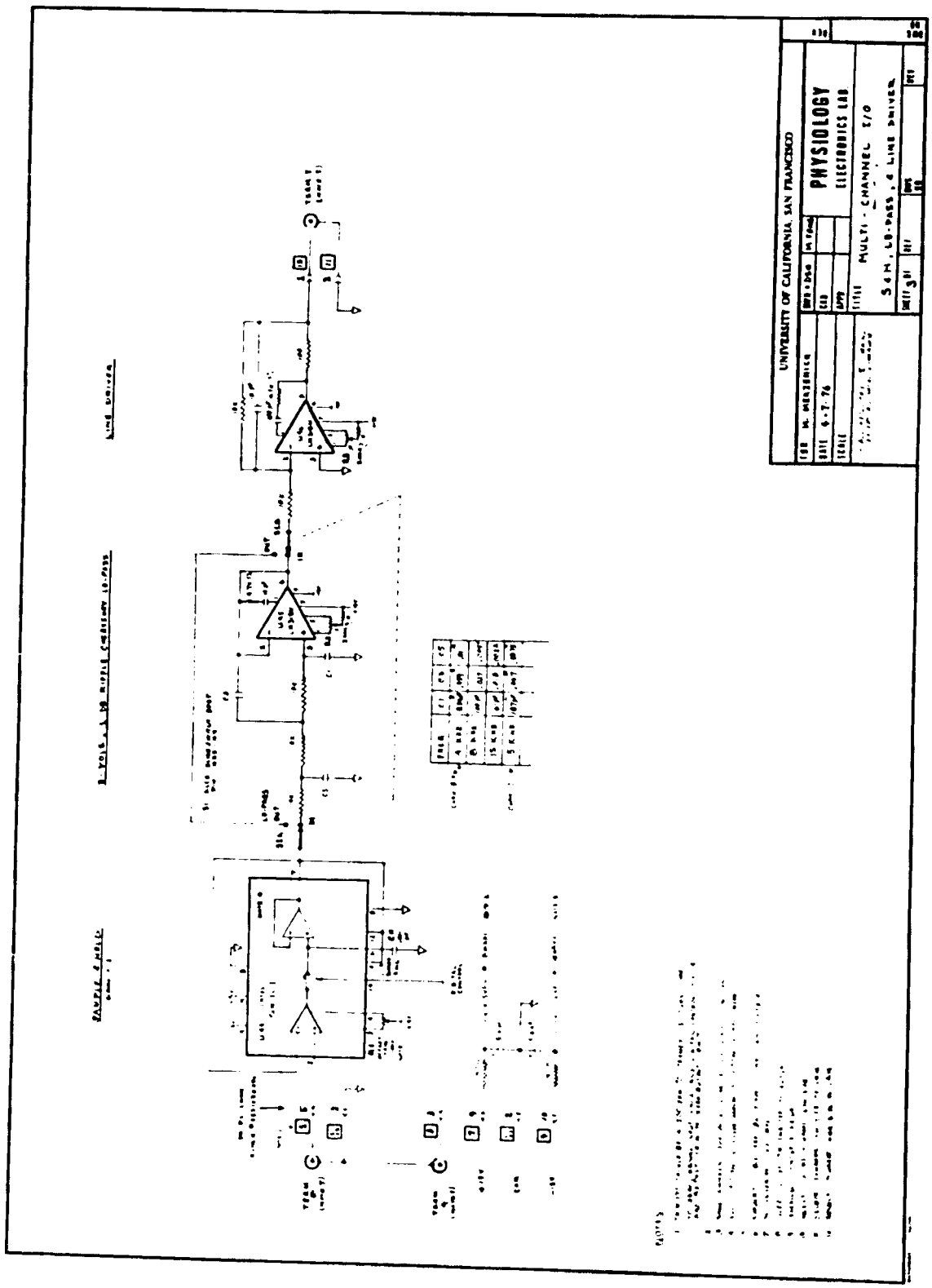

UNIVERSITY OF CALIFORNIA SAN FRANCISCO
PHYSIOLOGY
ELECTRONICS LAB.
MULTI-CHANNEL I/O

Figure 5.6. Schematic diagram of the sample-and-hold section of the DA-DEMUX circuitry.


UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
PHYSIOLOGY
ELECTRONICS LAB
MULTI-CHANNEL I/O
S & H, LO-PASS, & LINE DRIVER

stimulation hardware. Figure 5.7 illustrates such a system. However, to obtain maximum versatility, one stimulator for each electrode contact is required.

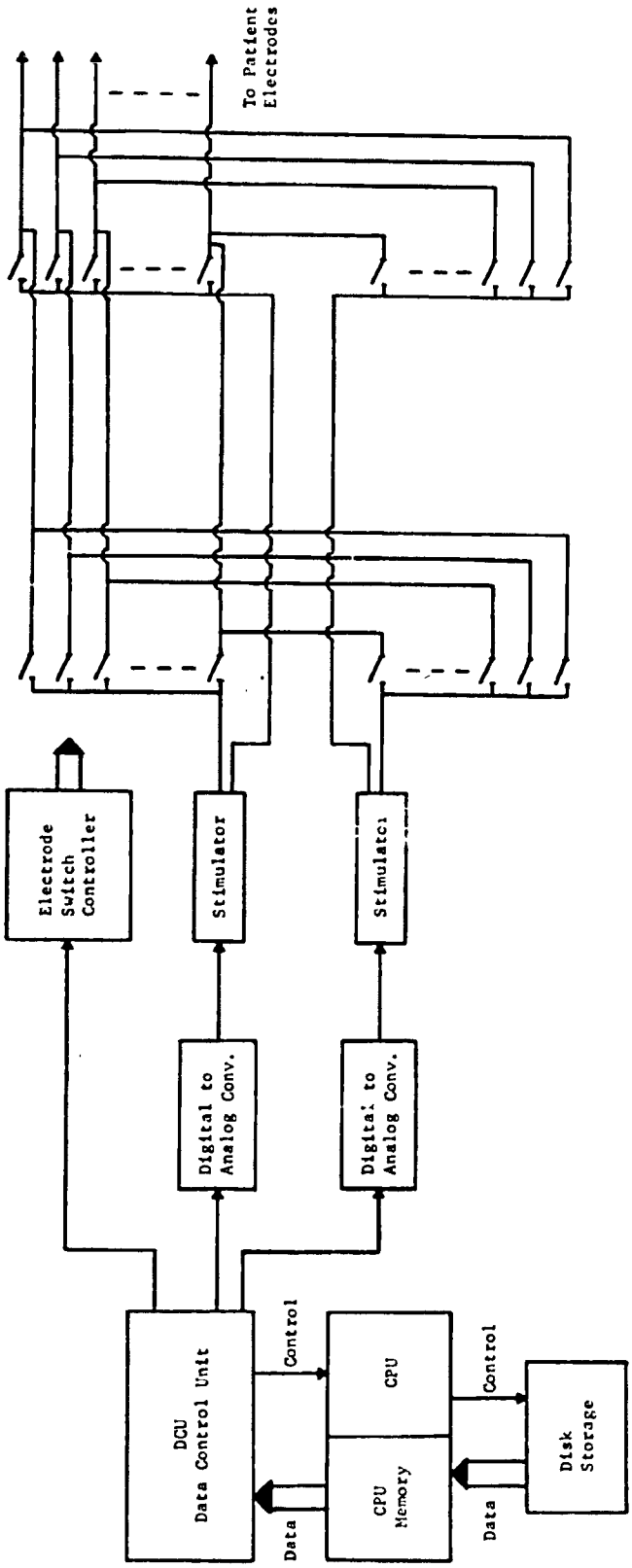
System Software

At present all software for the pulse code modulation, stimulus coding technique has been written. These routines include: 1) A CPU interrupt handler, written in FORTRAN, for loading memory buffers from disk at a 50Kword/sec rate (using FORTRAN callable system disk, I/O subroutines); 2) A general purpose DCU program loading routine (FORTRAN callable); 3) A DCU program (written in ASSEMBLER) for accessing CPU memory, decoding, and DA-DEMUX control. The existing routine generates stimuli assuming a pulse code modulation implementation.

Only program 3 need be changed for different stimulus coding options. Pulse code modulation, with appropriate output filtering, is useful for generating continuous waveform stimulation. As mentioned previously, if pulse or other fixed waveform stimulation is desired, more economical coding can be implemented. At least one other routine will be written for this purpose. It will be substituted for routine (3), mentioned above, whenever pulse stimulation is used.

Figure 5.7. Diagram of an electrode switching stimulation system

In this example system only two channels can be simultaneously stimulated. However, many electrode channels may be sequentially stimulated with this switching arrangement. For example, if a 16 wire electrode is implanted, 120 bipolar electrode contact combinations could be stimulated. If more than two electrode contacts are stimulated simultaneously, even more precise control of the eight nerve excitation pattern should be possible.


To Patient Electrodes
Electrode Switch Controller
Stimulator
Stimulator
Digital to Analog Conv.
Digital to Analog Conv.
DCU Data Control Unit
Control
Data
CPU
CPU Memory
Control
Data
Disk Storage

References

B.A. Blesser, "Perception of Spectraly Transformed Speech", Doctoral Dissertation, MIT, EE, 1969.

A.S. Bregman, J.L. Campbell, "Primary Auditory Stream Segregation and Perception of Order in Rapid Sequences of Tones", J.E.P. 89: 244-249, 1971.

D.J. Broad, J.E. Shoup, "Concepts for Acoustic Phonetic Recognition", in Speech Recognition, ed. D. Raj Reddy, Academic Press, 1975.

S.L. BeMent and J.B. Ranck, Jr., "A Quantitative Study of Electrical Stimulation of Central Myelinated Fibers", Exper. Neurology, 24:147-170, 1969.

S.L. BeMent and J.B. Ranck, Jr., "A Model for Electrical Stimulation of Central Myelinated Fibers with Monopolar Electrodes", Exper. Neurol., 24:171-186, 1969.

R.C. Bilger, "Evaluation of Subjects Presently Fitted with Implanted Auditory Prosthesis", Ann Oto-Rhino-Laryng Suppl 38: Vol. 86, 1977.

R.W. Brodersen, H-S. Fu, R.C. Frye, D.D. Buss, "A 500-Point Fourier Transform Using Charge-Coupled Devices", IEEE International Solid-State Circuits Conf., February, 1975.

S.B. Brummer, M. Turner, "Electrical Stimulation of the Nervous System: The Principle of Safe Charge Injection with Noble Metal Electrodes", Bioelectrochem. and Bioenergetics 2:13-25, 1975.

S.B. Brummer, M.J. Turner, "Electrochemical Considerations for Safe Electrical Stimulation of the Nervous System with Platinum Electrodes", IEEE Trans Biomed Engin BME-24:59-63, 1977.

D. Childers, A. Durling, Digital Filtering and Signal Processing, West Publishing Co., St. Paul, 1975.

A. Chouard, M. Lubin, Correspondence, Paris, France, 1977.

R.A. Cole, B. Scott, "Toward a Theory of Speech Perception", Psychological Review 81:348-374, 1974.

P.E. Crago, P.H. Peckham, J.T. Mortimer, J.P. Van der Meulen, "The Choice of Pulse Duration for Chronic Electrical Stimulation via Surface, Nerve, and Intermuscular Electrodes", Ann Biomed Engin 2:252, 1974.

J.L. Danhauer, S. Singh, Multidimensional Speech Perception by the Hearing Impaired: A Treatise on Distinctive Features, University Park Press, Baltimore, 1975.

H. Davis, "Principles of Electric Response Audiometry", Ann Oto-Rhino-Laryng Suppl 28: Vol. 85, 1976.

E. DeBoer, "On the Principle of Specific Coding", Journal of Dynamic Systems, Measurement, and Control, 265-273, Sept., 1973.

P. Divenyi, I.J. Hirsch, "Identification of Temporal Order in Three-Tone Sequences", J.A.S.A., 56:146-151, 1974.

P. Divenyi, I.J. Hirsch, "Some Figural Properties of Auditory Patterns", submitted to Journal of Acoustical Society of America.

P.E.K. Donaldson, "The Encapsulation of Microelectronic Devices for Long-Term Surgical Implantation", IEEE Trans Biomed Engin BME-23:281-285, 1976.

H. Dudley, "The Vocoder", Bell Lab Record 17:122-126, 1939.

A.M. Dymond, "Characteristics of the Metal-Tissue Interface of Stimulation Electrodes", IEEE Trans. Biomed. Engin., Vol. BME-23: No. 4, 274-280, July, 1976.

D.K. Eddington, W.H. Dobelle, D.E. Brackmann, M.G. Mladejovsky, J.L. Parkin, "Auditory Prosthesis Research with Multiple Channel Intracochlear Stimulation in Man", Preprint Current address: Massachusetts Eye and Ear Infirmary, 243 Charles Street, Boston, Mass 02114, 1977.

E.F. Evans, J.P. Wilson, "Frequency Sharpening of the Cochlea: The Effective Bandwidth of Cochlear Nerve Fibers", Seventh International Congress on Acoustics, Budapest, 1971.

E.F. Evans, "The Frequency Response and Other Properties of Single Fibers in the Guinea-Pig Cochlear Nerve", J. Physiol. 226: 263-287, 1972.

E.F. Evans, "Auditory Frequency Selectivity and the Cochlear Nerve", in Facts and Models in Hearing, ed. by Zwicker and E. Terhardt, Springer-Verlag, New York, 1974.

E.F. Evans, J.P. Wilson, "Cochlear Tuning Properties: Concurrent Basilar Membrane and Single Nerve Fiber Measurements", Science 190:1218-1221, 1975.

E.F. Evans, "Cochlear Nerve and Cochlear Nerve", in Auditory System, Physiology (CNS)- Behavioral Studies - Psychoacoustics, ed. by W.D. Keidel and W.D. Neff, Springer-Verlag, Berlin, 1975b.

E.F. Evans, "Frequency Selectivity at High Signal Levels of Single Units in Cochlear Nerve and Nucleus", in Psychophysics and Physiology of Hearing, ed. E.F. Evans and J.P. Wilson, Academic Press, New York, 1977.

R. Fitzhugh, "Impulses and Physiological States in Theoretical Models of Nerve Membrane", Biophys Journal 1:445-466, 1961.

J.L. Flanagan, "A Difference Limen for Vowel Formant Frequency", J.A.S.A. 27, 613-617, 1955.

J.L. Flanagan, Speech Analysis Synthesis and Perception, (second edition), Springer-Verlag, New York, 1972.

B. Frankenhaeuser, A.F. Huxley, "The Action Potential in the Myelinated Nerve Fiber of Xenopus Laevis as Computed on the Basis of Voltage Clamp Data", J. Physiol. 171:302, 1964.

C.D. Geisler, J.M. Goldberg, "A Stochastic Model of the Repetitive Actimety of Neurons", Biophys Journal 6:53-69, 1966.

C.D. Geisler, "A Model of the Peripheral Auditory System Responding to Low-Frequency Tones", Biophysical Journal 8:1-15, 1968.

C.D. Geisler, W.S. Rhode, D.T. Kennedy, "Responses to Tonal Stimuli of Single Auditory Nerve Fibers and Their Relationship to Basilar Membrane Motion in the Squirrel Monkey", J. Neurophysiology 37:1156-1172, 1974.

T.J. Goblick, R.R. Pfeiffer, "Time-Domain Measurements to Cochlear Nonlinearities Using Combination Click Stimuli", J.A.S.A. 46: 924-938, 1969.

B. Gold, C.M. Rader, Digital Processing of Signals, McGraw-Hill Book Co., New York, 1969.

R.M. Golden, "Digital Computer Simulation of a Sampled-Data Voice-Excited Vocoder", 35: No. 9, 1358-1366, September, 1963.

R.M. Golden, "Vocoder Filter Design: Practical Considerations", J. Acoust Soc Am 43:803-810, 1968.

J.L. Goldstein, N.Y. Kiang, "Neural Correlates of the Aural Combination Tone $2f_1 - f_2$", Proc IEEE 56:981, 1968.

S. Glasser, J. Miller, N.G. Xuong, A. Selverston, "Computer Reconstruction of Invertebrate Nerve Cells", in Computer Analysis of Neuronal Structures, ed. by R. Lindsay, UCLA, Plunum Press, 1977.

P.R. Gray, "Conditional Probability Analysis of the Spike Activity of Single Neurons", Biophys J. 7:759-777, 1967.

D.M. Green, An Introduction to Hearing, John Wiley and Sons, New York, 1976.

D.D. Greenwood, "Critical Bandwidth and the Frequency Coordinates of the Basilar Membrane", J. Acoust Soc Am 33:1344-1356, 1961.

D.L. Guyton, F.T. Hambrecht, "Theory and Design of Capacitor Electrodes for Chronic Stimulation", Med Biol Eng 12:613-620, 1974.

T.R. Gweelala, A CMOS Implantable Auditory Prosthesis Using Integrated Circuits Technology, Tech Rept No. 5301-1, Stanford Electronics Lab, Stanford.

C.C. Hansen, O.M. Lauridsen, "Electrical Stimulation of the Inner Ear", Nordic J Medico-Tech, No. 9, 1975.

A.V. Hill, "Excitation and Accommodation in Nerve", Proc Roy Soc b 119:305-355, 1936a.

A.V. Hill, B. Katz, D.Y. Soandt, "Nerve Excitation by Alternating Current", Proc Roy Soc b 121:74-133, 1936b.

A.L. Hodgkin, A.F. Huxley, "A Quantitative Description of Membrane Current and Its Application to Conduction and Excitation in Nerve", J. Pysiol 117:500-544, 1952.

A.S. House, K.N. Stevens, T.T. Sandel, J.B. Arnold, "On the Learning of Speechlike Vocabularies", J Verbal Learn and Verbal Behavior 1:133-143, 1962.

D.L. Jewett, J.S. Williston, "Auditory Evoked Far Fields Averaged from the Scalp of Humans", Brain 94:681-696, 1971.

J.F. Kaiser, "Design Methods for Sampled Data Filters", Proc First Allerton Conf on Circuit and System Theory, 221-236, November, 1963.

J.F. Kaiser, "Digital Filters", in System Analysis by Digital Computer, F.F. Kuo, J.F. Kaiser, ed., John Wiley and Sons, New Yrok, 1966.

N.Y.S. Kiang, T. Watanabe, E.C. Thomas, et al, Discharge Patterns of Single Fibers in the Cat's Auditory Nerve, Cambridge, MIT Press, 1965.

N.Y.S. Kiang, E.C. Moxon, "Physiological Considerations in Artificial Stimulation of the Inner Ear", Ann Otol 81:714-730, 1972.

D.P. Kolba, T.W. Parks, "A Prime Factor FFT Algorithm Using High-Speed Convolution", IEEE Trans ASSP-25:281-294, 1977.

K.D. Kryter, "Methods for the Calculation and Use of the Articulation Index", J Acoust Soc Am 34:1689-1697, 1962.

J.C.R. Licklider, "Effects of Amplitude Distortion upon the Intelligibility of Speech", J.A.S.A. 18:429-434, 1946.

J.C. Lilly, "Injury and Excitation by Electric Currents. The Balanced Pulse-Pair Waveform", in Electrical Stimulation of the Brain, ed., D.E. Sheer (Austin, Texas: Texas Univ. Press), pp 60, 1961.

J.J. Lussier, W.A.H. Rushton, "The Excitability of a Single Fibre in a Nerve Trunk", J Physiol 117:87-108, 1952.

A.M. Liberman, F.S. Cooper, D.P.Shankweiler, M. Studdert-Kennedy, "Perception of the Speech Code", Psychological Rev, 74:431- 461, November, 1967.

J.D. Markel, A.H. Gray, Linear Prediction of Speech, Springer-Verlag, Berlin, 1977.

P.Z. Marmarelis, G.D. McCann, "Development and Application of White-Noise Modeling Techniques for Studies of Insect Visual Nervous System", Kybernetic 12:74-89, 1973a.

P.Z. Marmarelis, K.I. Naka, "Nonlinear Analysis and Synthesis of Receptive-Field Responses in the Catfish Retina. I. Horizontal Cell to Ganglion Cell Chain", J Neurophys 36:605-618, 1973b.

J.H. McClellan, T.W. Parks, L.R. Rabiner, "A Computer Program for Designing Optimum FIR Linear Phase Digital Filters", IEEE Trans on Audio and Electroacoustics AU-21:506-526, 1973.

D.R. McNeal, "Analysis of a Model for Excitation of Myelinated Nerve", IEEE Trans Biomed Engin BME-23:329-337, 1976.

D.R. McNeal, paper presented in IEEE Trans. Biomed. Eng., January, 1978.

M.M. Merzenich, R.P. Michelson, C.R. Pettit, R.A. Schindler, and M. Reid, "Neural Encoding of Sound Sensation Evoked by Electrical Stimulation of the Acoustic Nerve", Ann Otol, 82, No. 4:486-503, 1973.

M.M. Merzenich, M.D. Reid, "Representation of the Cochlea Within the Inferior Colliculus of the Cat", Brain Res 77:397-415, 1974.

M.M. Merzenich, D.N. Schindler, M.W. White, "Symposium on Cochlear Implants II. Feasibility of Multichannel Scala Tympani Stimulation", Laryngoscope 84:1887-1893, 1974.

M.M. Merzenich, "Studies on Electrical Stimulation of the Auditory Nerve in Animals and Man; Cochlear Implants", in The Nervous System, D.B. Tower, ed., Vol. 3: Human Communication and Its Disorders, Raven Press, New York, 1975.

M.M. Merzenich, M.W. White, "Cochlear Prosthesis: The Interface Problem (10pp. 12 fig.) (to be published in proc of an NIH-RSA sponsored conf on development of neuroprosthetic devices) Marcel Dekker, New York, 1977.

R.P. Michelson, "The Results of Electrical Stimulation of the Cochlea in Human Sensory Deafness", Annals Otolcgy, Rhinology, and Laryngology, 80, No. 6, 914-919, December, 1971.

R.P. Michelson, M.M. Merzenich, C.R. Pettit, R.A. Schindler, "A Cochlear Prosthesis: Further Clinical Observations; Preliminary Results of Physiological Studies", Laryngoscope 83:1116-1122, 1973.

N. Miller, P.D. Jensen, A.K. Meyers, "Injury and Excitation by Electric Currents B, A comparison of the Lilly waveform and the Sixty-Cycle Sinewave", in Electrical Stimulation of the Brain, ed., D.E. Sheer (Austin, Texas: Texas Univ. Press, 1961), pp. 73, 1961.

A.R. Moller, "Frequency Selectivity of the Basilar Membrane Revealed from Discharges in Auditory Nerve Fibers", in Psychophysics and Physiology of Hearing, ed., E.F. Evans and J.P. Wilson, Academic Press, New York, 1977.

V. Mooney, S.A. Swartz, A.M. Roth, M.J. Gorniowsky, "Percutaneous Implant Devices", Ann Biomed Engin 5:34-46, 1977.

R.N. Noyce, "Large-Scale Integration: What Is Yet To Come?", Science 195:1102-1106, 1977.

F.W. Nye, J.H. Gaitenby, "Intelligibility of Synthetic Monosyllable Words in Short Syntactically Normal Sentences", in Haskins Laboratory Status Report on Speech Research SR-37/38, 1974.

E. Peterson, F.S. Cooper, "Peak-picker: A Bandwidth Compression Device", J.A.S.A. 29:777 (A), 1957.

R.R. Pfeiffer, "A Model for Two-Tone Inhibition of Single Cochlear-Nerve Fibers", J.A.S.A. 48:1373-1378, 1970.

R.R. Pfeiffer and D.O. Kim, "Cochlear Nerve Fiber Responses: Distribution Along the Cochlear Partition", J. Acoust Soc Am 58, No. 4, 867-869, October, 1975.

L.R. Rabiner, Schafer, Speech Processing, Prentice-Hall, Englewood Cliffs, N.J., 1977.

E.A. Robinson, Multichannel Time Series Analysis with Digital Computer Programs, Holden-Day, San Francisco, 1968.

J.E. Rose, "Electrical Activity of Single Auditory Nerve Fibers", Adv Oto-Rhino-Laryng 20:357-373, 1973.

J.E. Rose, "Discharges of Single Fibers in the Mammalian Auditory Nerve", in Frequency Analysis and Periodicity Detection in Hearing, R. Plomp and G.F. Smoorenburg, ed., Sijthoff, Leiden, Netherlands, 1969/1970.

J.E. Rose, J.F. Brugge, D.J. Anuerson, J.E. Hind, "Some Possible Neural Correlates of Combination Tones", J Neurophysiol 32:402-423, 1969.

J.E. Rose, J.F. Brugge, D.J. Anderson, J.E. Hind, "Phase-Locked Response to Low-Frequency Tones in Single Auditory Nerve Fibers of the Squirrel Monkey", J Neurophys 30:769-793, 1967.

W.A.H. Rushton, "The Effect Upon the Threshold for Nervous Excitation of the Length of Nerve Exposed, and the Angle Between Current and Nerve", J Physiol (London), 82:357-377, 1927.

B. Rutkin, Private Communication, 1975.

M.B. Sachs, N.Y.S. Kiang, "Two-Tone Inhibition in Auditory-Nerve Fibers", J.A.S.A. 43:1120-1128, 1968.

M.B. Sachs, P.J. Abbas, "Rate Versus Level Functions for Auditory-Nerve Fibers in Cats: Tone-Burst Stimuli", J Acoust Soc Am 56: 1835-1847, 1974.

R.A. Schindler, M.M. Merzenich, "Chronic Intracochlear Electrode Implantation: Cochlear Pathology and Acoustic Nerve Survival", Ann Otol 83:202-215, 1974.

R.A. Schindler, "The Cochlear Histopathology of Chronic Intracochlear Implantation", J Laryng Otol 90:445-457, 1976.

R.A. Schinler, M.M. Merzenich, M.W. White, B. Bjorkroth, "Multi-electrode Intracochlear Implants: Nerve Survival and Stimulation Patterns", to be published in Archives of Otolaryngology.

M.R. Schroeder, J.L. Hall, "A Model for Mechanical to Neural Transduction in the Auditory Receptor", in Facts and Models in Hearing, E. Zwicker and E. Terhardt, ed., Springer-Verlag, New York, 1974.

E. Schubert, S. Walsh, Private Communication, 1976.

M.J. Shantz, G.D. McCann, "Computational Morphology: Three-Dimensional Computer Graphics for Electron Microscopy", IEEE Trans Biomed Eng, Vol. BME-25, January, 1978.

W.M. Siebert, "Hearing and the Ear", in Engineering Principles in Physiology, Vol. 1, J.H.U. Brown and D.S. Gann, eds., Academic Press, New York, 1973.

W.M. Siebert, "Stimulus Transformations in the Peripheral Auditory System", in Recognizing Patterns, ed., P.A. Kolers and M. Eden, MIT Press, Cambridge, Mass, 1968.

F.B. Simmons, "Electrical Stimulation of the Auditory Nerve in Man", Arch Otolaryng 84:24-76, 1966.

R.L. Smith, J.J. Zwislocki, "Shot-Term Adaptation and Incremental Responses of Single Auditory-Nerve Fibers", Biol Cybernetics 17:169-182, 1975.

H. Spoendlin, "Innervation Patterns in the Organ of Corti of the Cat", Acta Oto-Laryng 67:239-254, 1969.

H. Spoendlin, "Innervation Densities of the Cochlea", Acta Oto-Rhino-Laryng 73:235-248, 1972.

L. Stark, "Functional Analysis of Pupil Nonlinearities", in Neurological Control Systems, Studies in Bioengineering by L. Stark, Plenum Press, New York, 1968.

T.G. Stockham, "Highspeed Convolution and Correlation", 1966 Spring Joint Computer Conf, AFIPS Proc, 28:229-233, 1966.

A.A. Verveen, "On the Fluctuation of Threshold of the Nerve Fiber", in Structure and Function of the Cerebral Cortex, D.B. Tower, J.P. Schade eds., Proc Second Internat Meeting of Neurobiologists, Amsterdam, 1959.

W. Voiers, "Reasearch on Diagnostic Evaluation of Speech Intelligibility", Final Report, January 24, 1973, Tracor, 6500 Tracor Lane, Austin, Texas.

S. Walsh, Dissertation, Stanford University, 1978.

T.F. Weiss, "A Model of the Peripheral Auditory System", 3 No. 4: 153-175, November, 1966.

R.L. White, T.J. Gross, "An Evaluation of the Resistance to Electrolysis of Metals for Use in Biostimulation Microprobes", IEEE Trans Biomed Engin 21:487-490, 1974.

G. Winston, Private Communications, 1975.