INFORMATION TO USERS

This manuscript has been reproduced from the microfilm master. UMI films the text directly from the original or copy submitted. Thus, some thesis and dissertation copies are in typewriter face, while others may be from any type of computer printer.

The quality of this reproduction is dependent upon the quality of the copy submitted. Broken or indistinct print, colored or poor quality illustrations and photographs, print bleedthrough, substandard margins, and improper alignment can adversely affect reproduction.

In the unlikely event that the author did not send UMI a complete manuscript and there are missing pages, these will be noted. Also, if unauthorized copyright material had to be removed, a note will indicate the deletion.

Oversize materials (e.g., maps, drawings, charts) are reproduced by sectioning the original, beginning at the upper left-hand corner and continuing from left to right in equal sections with small overlaps. Each original is also photographed in one exposure and is included in reduced form at the back of the book.

Photographs included in the original manuscript have been reproduced xerographically in this copy. Higher quality 6” x 9” black and white photographic prints are available for any photographs or illustrations appearing in this copy for an additional charge. Contact UMI directly to order.

UMI
A Bell & Howell Information Company
300 North Zeeb Road, Ann Arbor MI 48106-1346 USA
313/761-4700 800/521-0600
DETECTION OF FREQUENCY MODULATION BY LISTENERS WITH COCHLEAR IMPLANTS

DISSERTATION

Presented in Partial Fulfillment of the Requirements for
the Degree Doctor of Philosophy in the Graduate School of the The Ohio State University by
Ina Rea Bicknell, A. B., M. A.
The Ohio State University 1996

Dissertation Committee:
Lawrence L. Feth
Pamela J. Mishler
D. Bradley Welling
Marios S. Fourakis
Jan M. Weisenberger

Approved by

Adviser
Department of Speech and Hearing Science
UMI Number: 9630853
To Kara and Jessica, my angels.

After you grow your wings, fly as high as you can.
ACKNOWLEDGMENTS

I would like to thank my adviser Dr. Lawrence Feth for his support and advice throughout this study. My thanks also go to Dr. D. Bradley Welling who introduced me to the cochlear implant and who was so generous with his time. Special thanks go to Dr. Pamela Mishler for her friendship and enthusiastic support during my fellowship year at the Dayton Veterans Administration Medical Center. I would also like to thank Dr. David Van Winkle, Chief of Audiology and Speech Pathology, at the Dayton VA for his generosity in providing me with equipment and space for this project.

The preparation of the synthetic speech continuum was done under the supervision of Dr. Rob Fox. I would like to thank him for his forbearance while I learned to synthesize the stimuli.

My thanks go to Dr. R. Goldenberg and Dr. T. Schneiderman of Soifer, Goldenberg Ear Associates for providing me access to the patients who volunteered to participate in this study.

Many of the decisions about the kinds of signals, the signal intensity, and procedural modifications used in this study were made on the basis of a pilot study run in 1994. I wish to thank Dr. D. E. Gebhart and Dr. H. W. Lowery of Ear-Nose-Throat and Head & Neck Surgeons, Inc. and Dr. Welling for providing the names of patients who volunteered for this pilot project.
Special thanks go to Lynette Roth. She provided me with invaluable practical advice about the implant device and about the patients who use it. I would also like to thank Linda McGinnis of Goldenberg and Soifer Ear Associates and Beth Fais and Paula Beal of Ear-Nose-Throat and Head & Neck Surgeons, Inc. who were so cooperative in providing me names, addresses, and MAPs of the cochlear-implant subjects.

I am indebted to Jayanth Anantharaman and Karl Wilke for their programming help.

During the time this research was done, I was supported by a Predoctoral Fellowship in Audiology at the Dayton Veterans Administration Medical Center. My thanks to the Department of Veterans Affairs for this support. Funds for subjects were provided by a Graduate Student Alumni Research Award from the Graduate School of The Ohio State University. Further support for me and funds for subjects were provided, through Dr. Feth, by the Air Force Office of Scientific Research. The pilot study done in the summer of 1994 was supported by an Interdisciplinary Seed Grant from The Ohio State University which was awarded to Dr. Feth, Dr. Welling and Dr. A. Krishnamurthy.

This study would not have been possible without the participation of volunteers CI 1, CI 2, and CI 3 and the four volunteers who served as subjects
in the pilot study. Their courage as pioneers in the restoration of hearing via electrical stimulation and their cheerful willingness to serve as participants in research directed at improving the cochlear implant are truly remarkable.

Finally, I would like to thank my husband Bill for his unflagging support during my pursuit of this doctoral degree. The wine and flowers never failed to lift my spirits.
VITA

July 29, 1942 ............................................. Born - Indianapolis, Indiana

1964 ............................................................. A.B., Indiana University
Bloomington, Indiana

1968-1969 ................................................... Assistant Professor, Biology
University of California, Chico
Chico, California

1970-1971 .................................................... Research Assistant,
Washington University
St. Louis, Missouri

1971 ............................................................ M.A., Washington University
St. Louis, Missouri

1974-1976 ...................................................... Instructor, Biology
1978, 1987 ................................................... Wright State University
1985-1986	Department of Biochemistry

1985-1986 ...................................................... Instructor, Biology
University of Houston
Houston, Texas

1978-1980 ...................................................... Research Assistant
1987-1989	Department of Biochemistry
Wright State University
Dayton, Ohio

1989-1992 ...................................................... Teaching Assistant
Speech and Hearing Science
The Ohio State University
Columbus, Ohio
1993-1994 ................................................ Graduate Research Assistant
Speech and Hearing Science
The Ohio State University
Columbus, Ohio

1994-1995 ................................................ Pre-doctoral Fellow
in Audiology; Dayton Veterans
Administration Medical
Center; Dayton, Ohio

1995 ......................................................... Visiting Instructor, Audiology
and Speech Pathology
Miami University
Oxford, Ohio

PUBLICATIONS


FIELD OF STUDY

Major Field: Speech and Hearing Science
TABLE OF CONTENTS

DEDICATION .................................................. ii
ACKNOWLEDGMENTS ................................. iii
VITA ............................................................... vi
LIST OF TABLES ........................................... xiii
LIST OF FIGURES ......................................... xv

CHAPTER PAGE

I. INTRODUCTION ......................................... 1
   1.1 Contribution of psychoacoustic studies to understanding speech perception ............... 4
   1.2 Detection of frequency-modulated cues by normal-hearing listeners as determined from psychoacoustic studies ................................................................. 5
   1.3 Detection of frequency-modulated cues by hearing-impaired listeners as determined from psychoacoustic studies ................................................................. 6
   1.4 Detection of frequency-modulated cues by cochlear-implant users .............................. 7
   1.5 Sound perception via electrical stimulation of the VIIIth nerve: The cochlear implant device .......................................................... 9
      1.5.1 Design of the MSP speech processor .......................... 11
      1.5.2 Design of the Spectra speech processor .................. 14
   1.6 Scope of the present research ........................................ 16

II. LITERATURE REVIEW .................................. 18
   2.1 Psychoacoustic studies of temporal acuity in normal-hearing and hearing-impaired listeners ............................. 20
2.2 Psychoacoustic studies of temporal acuity in listeners with cochlear implants ........................................ 31
2.3 Perception of frequency-modulated signals in normal-hearing and hearing-impaired listeners. ............... 54
2.4 Perception of dynamic signals by cochlear-implant users ................................................................. 80
2.5 Encoding of frequency-modulation in auditory-nerve fibers: acoustic stimulation ................................ 84
2.6 Encoding of frequency-modulation in auditory-nerve fibers: electrical stimulation ............................... 100
2.7 The perception of elements of a synthetic speech continuum that differ on the basis of place of articulation . 102
2.8 Comments on potential limiting factors in speech perception in cochlear-implant users ...................... 104

III. SUBJECTS AND METHODS ................................................................. 108

3.1 Subjects ..................................................................................... 108
3.2 Stimuli ...................................................................................... 112
  3.2.1 Signal frequency ................................................................. 113
  3.2.2 Signal level .......................................................................... 115
3.3 Signal generation ..................................................................... 122
3.4 Signal delivery ......................................................................... 122
3.5 Step-versus-glide discrimination task .................................. 123
3.6 Synthetic speech stimuli ............................................................... 128
3.7 Discrimination and identification task using the synthetic speech stimuli ........................................... 132
3.8 Delivery of the synthetic speech stimuli ......................................... 134
3.9 Data analysis ................................................................................ 135

IV. RESULTS ................................................................................................... 137

4.1 Initial signal durations and step sizes used to collect data for the step-versus-glide discrimination task ........ 139
4.2 Intensity levels corresponding to one-half of the dynamic range ................................................................. 141
4.3 Temporal resolution thresholds as determined from the step-versus-glide discrimination task ................. 142
  4.3.1 Temporal resolution thresholds as determined from the step-versus-glide discrimination task: Listeners with cochlear implants ................................................................. 144
  4.3.2 Temporal resolution thresholds as determined from the step-versus-glide discrimination task: Listeners with normal hearing ................................................................. 177
E. INSTRUCTIONS: DISCRIMINATION SYNTHETIC SPEECH . 259
F. INSTRUCTIONS: IDENTIFICATION SYNTHETIC SPEECH . 261
## LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Biographic data for cochlear-implant and normal-hearing subjects</td>
<td>110</td>
</tr>
<tr>
<td>2. Frequency characteristics of step and glide signals for cochlear-implant subjects</td>
<td>114</td>
</tr>
<tr>
<td>3. Initial signal durations and adaptive step sizes used in the step-versus-glide discrimination task for cochlear-implant subjects</td>
<td>125</td>
</tr>
<tr>
<td>4. Initial signal durations and adaptive step sizes used in the step-versus-glide discrimination task for normal-hearing subjects</td>
<td>126</td>
</tr>
<tr>
<td>5. Initial frequencies for the synthetic consonant formant transitions</td>
<td>130</td>
</tr>
<tr>
<td>6. Subject CI 1. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep and downsweep signals</td>
<td>156</td>
</tr>
<tr>
<td>7. Subject CI 3. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals</td>
<td>164</td>
</tr>
<tr>
<td>8. Subject CI 3. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downsweep signals</td>
<td>166</td>
</tr>
</tbody>
</table>
9. Subject CI 2. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals .............................................. 172

10. Subject CI 2. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for down-sweep signals .............................................. 174

11. Subject NH 1. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals .............................................. 181

12. Subject NH 1. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for down-sweep signals .............................................. 182

13. Subject NH 2. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals .............................................. 183

14. Subject NH 2. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for down-sweep signals .............................................. 184

15. Subject NH 3. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals .............................................. 185

16. Subject NH 3. Mean signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for down-sweep signals .............................................. 186

17. d' as an index of the sensitivity of listeners with cochlear implants and listeners with normal hearing to the difference between two-step pairs of elements in a synthetic speech continuum. .......... 206
# LIST OF FIGURES

<table>
<thead>
<tr>
<th>FIGURES</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The Nucleus 22-Channel Cochlear Implant System.</td>
<td>10</td>
</tr>
<tr>
<td>2. Schematic representation of a glide and a step signal</td>
<td>116</td>
</tr>
<tr>
<td>3. Visual display as seen by subjects during testing session</td>
<td>118</td>
</tr>
<tr>
<td>4. First seven elements of synthetic /ba/, /da/, /ga/ continuum</td>
<td>131</td>
</tr>
<tr>
<td>5. Comparison of mean temporal resolution thresholds for CI 1 and NH 1</td>
<td>145</td>
</tr>
<tr>
<td>6. Comparison of mean temporal resolution thresholds for CI 2 and NH 2 for upsweep signals</td>
<td>147</td>
</tr>
<tr>
<td>7. Comparison of mean temporal resolution thresholds for CI 2 and NH 2 for downsweep signals</td>
<td>149</td>
</tr>
<tr>
<td>8. Comparison of mean temporal resolution thresholds for CI 3 and NH 3 for upsweep signals</td>
<td>151</td>
</tr>
<tr>
<td>9. Comparison of mean temporal resolution thresholds for CI 3 and NH 3 for downsweep signals</td>
<td>153</td>
</tr>
<tr>
<td>10. Mean temporal resolution thresholds for cochlear-implant user CI 1</td>
<td>155</td>
</tr>
<tr>
<td>11. Mean temporal resolution thresholds for subject CI 1</td>
<td>159</td>
</tr>
<tr>
<td>12. Mean temporal resolution thresholds for cochlear-implant user CI 3</td>
<td>161</td>
</tr>
<tr>
<td>13. Mean temporal resolution thresholds for subject CI 3</td>
<td>163</td>
</tr>
</tbody>
</table>
14. Mean temporal resolution thresholds for cochlear-implant user CI 2 ....................................................... 168
15. Mean temporal resolution thresholds for subject CI 2 ................................................................. 170
16. Mean temporal resolution thresholds for subject CI 2 at signal levels that produced a sensation of vibration and at lowered levels that produced an auditory sensation .................................................. 171
17. Mean temporal resolution thresholds for normal-hearing listener NH 1 ......................................................................................... 178
18. Mean temporal resolution thresholds for normal-hearing listener NH 2 ......................................................................................... 179
19. Mean temporal resolution thresholds for normal-hearing listener NH 3 ......................................................................................... 180
20. Mean temporal resolution thresholds for subject NH 1 ........................................................................... 188
21. Mean temporal resolution thresholds for subject NH 2 ........................................................................... 190
22. Mean temporal resolution thresholds for subject NH 3 ........................................................................... 191
23. Five-step signals on which subject NH 1 performed inconsistently ..................................................... 193
24. Identification functions for listeners with normal hearing .............................................................. 197
25. Identification functions for cochlear-implant subjects ........................................................................ 198
26. Average identification function for all listeners with normal hearing .................................................. 199
27. Discrimination functions for listeners with normal hearing ............................................................ 201
28. Average discrimination function for listeners with normal hearing .................................................. 202
29. Discrimination functions for cochlear-implant listeners ................................................................... 203
CHAPTER I
INTRODUCTION

Acoustic events in an individual’s environment are rarely static but rather are dynamic, changing with time. Dynamic sounds can change over time in amplitude, i.e., are amplitude modulated and/or in frequency, i.e., are frequency modulated (reviewed in Kay, 1982). Some of the most significant dynamic sounds, at least for humans, are the amplitude and frequency modulations that occur in speech. During speech production, the flow of air produced by the lungs is modified by the transitory characteristics of the vocal tract. The pharynx, hard and soft palates, tongue, jaw, teeth and the nasal passage are all part of the articulatory mechanism acting upon the air stream, which may or may not have been set into vibration by the vocal folds, producing sound that we identify as "speech." As the mobile components of the articulatory mechanism move rapidly to and from their target positions, changes in the length of the vocal tract and in the location and degree of constrictions result in rapid changes in the natural modes of vibration, or resonances, of the vocal tract. The imposition of these resonances upon the source air stream can be visualized in spectrograms as formants. Steady-state formants are a feature of vowel production in speech. Rapidly changing formants, formant transitions, are a feature of consonant-vowel or vowel-consonant production.
Classic studies in the 1950s at the Haskins Laboratory have shown that formant transitions, which can be viewed as frequency-modulated signals, are important in the identification of speech (Cooper, Delattre, Lieberman, Borst, and Gerstman, 1952; Liberman, Delattre, and Cooper, 1952; Liberman, Delattre, Cooper, and Gerstman, 1954; Delattre, Liberman, and Cooper, 1955). First formant (F1) transitions have been shown to be important as perceptual cues for distinguishing voiced and voiceless stops (Liberman, Delattre, Gerstman, and Cooper, 1956). Rising first formants characterized voiced consonants whereas flat or falling formants characterized voiceless consonants. These experiments showed that the direction and extent of higher frequency transitions, particularly the second (F2) and, to some extent, the third (F3) formants, are nearly all the acoustic information a listener needs to distinguish perceptually among labial, alveolar, and velar stops, and among nasals. In other words, they may be cues to the place of articulation. Studies of synthetic consonants by the Haskins group showed that /b/, /p/, and /m/ are characterized by a rising F2 transition, and /g/, /k/, and /ŋ/ by a falling F2 transition. Second formant transitions in /d/, /t/, and /n/ can be either rising or falling depending on the associated vowel.

The rate of the rise or fall in the formant frequency transitions was also shown by the Haskins group to have an effect on consonant perception (Liberman, Delattre, Gerstman, and Cooper, 1956). As the first and second transitions of the stops, /b/, /g/, and /d/ accompanied by the vowel /a/ were slowed, the consonant perception changed from that of a stop to that of a semivowel to that of a vowel. For example, the perception of /ba/ became
/wa/ and then became a vowel changing color from /u/ to /a/. A similar result was obtained when the vowel was /ɛ.

Using synthetic consonant vowel stimuli (CV) in which C represented a voiced stop consonant, Stevens and Blumstein (1978) compared stimuli having transition-only information, transition-plus-burst information, and burst-only information. In an identification task employing these stimuli, 100% of the normal-hearing listeners could identify the transition-only stimuli in a /ba, da, ga/ continuum. Only 36% could identify the stimuli correctly when the stimuli contained burst-only cues. Stevens and Blumstein concluded that listeners had a difficult time identifying the burst-only stimuli because they contained two onsets, one at the onset of the burst and one at the onset of voicing. The listener, therefore, was receiving conflicting information. The authors hypothesized that the significance of the formant transition as a cue to consonant perception lay in its role of smoothing out the onset spectrum thereby preventing conflicting information from successive onsets of burst and voicing.

Other investigators studying the role of formant transitions in the perception of consonants have arrived at conclusions that are at odds with the findings of the Haskins group and of Stevens and Blumstein. Fant (1973) concluded from his study of natural stimuli containing initial stops in combination with Swedish vowels that F2 and F3 formant transition patterns were not good predictors of place of articulation. Kewley-Port (1982), in an extensive study of formant transitions in naturally produced English syllables composed of /b, d, or g/ followed by one of a series of vowels,
concluded that formant transitions alone did not provide enough information to specify the place of articulation.

1.1 Contribution of psychoacoustic studies to understanding speech perception

As the studies mentioned above illustrate, experiments of basically the same design but employing natural versus synthetic stimuli, can result in diametrically opposed conclusions. One of the advantages to using synthetically generated speech stimuli is that the experimenter can manipulate the parameters in order to control the intelligibility of the speech. This approach enables the researcher to make some statements about which characteristics serve as cues in the perception of speech. The problem with both synthetic and natural speech stimuli is that they are very complex acoustic signals, and there may be interactions among the parameters of which the experimenter is unaware. For this reason, it is often useful to use very simple stimuli that represent the parameters of speech and to employ psychoacoustic tasks as a means to measure the limits of the auditory system with respect to the perception of these representations. In order to address the problem of the role of formant transitions in speech perception, one might consider formant transitions as being essentially frequency-modulated signals. A number of different psychoacoustic tasks can be used to study this problem, ranging from the simple discrimination of a frequency-modulated signal from a steady-state signal to tasks in which the frequency transition of
the signal is modulated in successively smaller steps until it becomes perceptually smoothed by the auditory system.

1.2 Detection of frequency-modulated cues by normal-hearing listeners as determined from psychoacoustic studies

One of the earliest psychoacoustic studies of perception of frequency-modulated signals by normal-hearing listeners was that of Sergeant and Harris (1962). Using unidirectional frequency-modulated signals, that is, signals that always rose in frequency from a fixed point, they showed that listeners were sensitive to signals in which the rise was as small as 5 Hz and that were as brief in duration as 2.5 ms. Since this study was done, numerous investigators (Collins, 1984; Cullen and Collins, 1982; Elliott, Hammer, Scholl, Carrell, and Wasowicz, 1989; Nábelek, 1978; Porter, Cullen, Collins, and Jackson, 1991) have examined the effects of various factors such as masking, rate of frequency change, direction, and duration on the ability of normal-hearing listeners to detect frequency-modulation in an acoustic signal.

Feth, Neil & Krishnamurthy (1989) in a study of temporal acuity asked normal-hearing listeners to distinguish between a signal that moved across a given frequency range in a linear trajectory, a glide, from a step signal that traversed the same range but did so in a series of discrete steps. Each step was a fixed frequency of short duration. At the end of a step, the frequency jumped instantaneously to the next step. As the step number of the signal increased, the duration of each step decreased, and the signal became
temporally smoothed, that is, it became perceptually more like a glide. From this study a temporal resolution threshold of 7-10 ms for normal-hearing listeners was determined for signals with center frequencies (CFs) below 4 kHz.

1.3 Detection of frequency-modulated cues by hearing-impaired listeners as determined from psychoacoustic studies

Once it was established that normal-hearing listeners could detect frequency modulation in a signal, it was natural to ask if hearing-impaired listeners had the same perceptual abilities. In a study designed to measure temporal acuity, Collins (1984) determined the effects of sensorineural hearing loss on the ability of listeners to discriminate frequency-varying sinusoids designed to be analogs of F2 transitions and the contiguous steady-state portions of consonant vowel (CV) syllables and vowel consonant (VC) syllables. She found that a majority of normal-hearing subjects had larger difference limens for the glide portion of the stimuli when the stimuli were glides followed by steady portions than for the reverse situation. Subjects with sensorineural hearing losses were inconsistent in their responses to the two kinds of stimuli but, in general, had larger difference limens than did the normal-hearing listeners.

Madden (1990) and Madden and Feth (1992) using step and glide signals like those employed in Feth et al.'s 1989 study compared the ability of listeners with normal hearing to that of listeners with hearing impairments of cochlear origin to discriminate between these signals. They found that
both normal-hearing subjects and hearing-impaired subjects could discriminate between the glide and step signals. However, the two groups had very different discrimination thresholds. The mean temporal resolution thresholds at the 75% level were found to be frequency dependent and ranged from 7-9 ms for signals centered at 500, 1000, and 2000 Hz for listeners with normal hearing. The temporal resolution thresholds of the hearing-impaired listeners were nearly twice as large as those of the normal-hearing listeners.

1.4 Detection of frequency-modulated cues by cochlear-implant users

The concept of hearing impairment spans a wide range of hearing limitations covering anything from a mild conductive loss that may limit the listener’s ability to hear certain frequencies to the profound sensorineural loss that stems from some dysfunction in the normal processing abilities of the hair cells, VIIIth (auditory) nerve, or a deficiency in the processing of the neural response in the central auditory pathway.

In 1957 a new type of device that electrically stimulates the fibers of the VIIIth nerve, the cochlear implant, was introduced. Restoration of at least some sound-processing capabilities in people who have a profound-to-total bilateral sensorineural hearing loss can be achieved with this implanted device (Bilger and Hopkinson, 1977; Tyler and Tye-Murray, 1991). In addition to improvement in the perception of general environmental sounds, cochlear-implant wearers show improvement in recognition of speech sounds. Many wearers can understand speech, particularly when sound
perception is coupled with visual cues from lipreading (Cooper, 1991). Some very successful wearers can understand speech without the aid of visual cues (Tyler, Moore, and Kuk, 1989; Tyler and Tye-Murray, 1991).

In view of the fact that frequency transitions are important in the perception of the consonants of speech, the question that arises, of course, is whether cochlear-implant users, like hearing-impaired listeners, are impaired in their ability to process the frequency-modulated information that is so important in consonant recognition. Are cochlear-implant users who are successful in their perception of speech tracking formant transitions or are they using some other cues? In a study to determine if cochlear-implant users can track dynamic frequency information, Tong, Clark, Blamey, Busby, and Dowell (1982) examined the ability of cochlear-implant users to discriminate between electrical-pulse trains that were delivered to one electrode only from pulse trains that were swept across differing numbers of electrodes. The results showed that the subjects had perfect discrimination scores even when the transition portion, that is, the sweep across electrodes, was only 25 msec in duration. The mean percent of judgments called “different” was 52% when the reference stimulus and the comparison stimulus differed by one initial electrode position, 80% when stimuli differed by two initial electrode positions, and 92% when the difference was three electrode positions.

In another study designed to investigate the kinds of cues that cochlear implant wearers might be using in their perception of speech, Kirk, Tye-Murray, and Hurtig (1992) used syllables containing different combinations of formant transitions, steady-state and durational cues. They found that word
recognition by cochlear-implant wearers was positively correlated with the use of dynamic vowel cues but not with steady-state cues.

1.5 Sound perception via electrical stimulation of the VIIIth nerve: The cochlear implant device

Today, approximately 10,000 people worldwide have some type of implanted hearing device. Of these, 5000 adults and 2000 children are implanted with the FDA-approved multichannel intracochlear implant device illustrated in Figure 1, the Nucleus 22 Channel Cochlear Implant System (hereafter referred to as the Nucleus 22). The Nucleus 22 is a transcutaneous system with a maximum of 22 active electrodes to which pulsatile electrical signals are delivered in a sequential manner. The external components of the Nucleus 22 consist of a directional microphone, worn behind the ear, which picks up sound and filters it to accept sound energy between 150 Hz and 10 kHz (Spectra 22) or 75 Hz and 6 kHz (Mini Speech Processor). The microphone has a pre-emphasis feature which compensates for the 6 dB per octave dropoff that usually occurs in speech by filtering the input at +6 dB per octave thereby producing a flat spectrum. The filter output is amplified by a fixed 30 dB gain.

The microphone output is sent to the next external component, the speech processor (usually worn by the patient in a pouch at his/her waist or at chest level). The two most current generations of speech processors are the Mini Speech Processor (MSP) and the Spectra 22. In both devices, the conditioned output from the microphone is sent to the Automatic Gain
Cochlear Implant

Audio Input Selector

Receiver/Stimulator

Transmitter

Speech Processor

Microphone

Figure 1: The Nucleus 22-Channel Cochlear Implant System
Control (AGC) of the signal-processing chips which are built into the speech processor. The AGC limits signal peaks to values that produce stimulation at the maximum acceptable loudness level (C level). C levels are determined at the initial activation of the implant device by the implant-team audiologist using psychoacoustic techniques and become a part of the patient's MAP. The MAP is the individualized coding scheme that is established during the activation of the implant device and is modified during the patient's rehabilitation period. T levels, that is, the lowest level of sound that will produce stimulation at threshold, are also established during the MAPping procedure. The dynamic range of acoustic input that will produce stimulation between T and C levels is 30 dB.

1.5.1 Design of the MSP speech processor

At this point, the designs of the two processors diverge considerably. The design of the MSP is based on the concept of feature extraction whereas the design of the Spectra is based on the concept of maximum filter output. In the MSP, the signal is sent from the AGC, conditioned by a low-pass filter, and then sampled by an Analog to Digital Converter. The MSP uses a feature extractor to extract and encode the relevant speech features from the acoustic signal, that is, the fundamental frequency (F0), the frequencies of the first and second formants (F1 and F2), and the amplitudes of the first and second formants (A1 and A2). The filter bandwidth for the F1 component is 280-1000 Hz; for the F2 component, the bandwidth is 800-4000 Hz. Additional energy in the acoustic signal at 2 kHz-2.8 kHz (Band 3), 2.8 kHz-4 kHz (Band
and above 4 kHz (Band 5) is also detected, encoded, and assigned to specific electrodes, usually electrodes 7, 4, and 1. This particular encoding strategy is the MPEAK Strategy (Cochlear Handbook, 1993).

Frequency information is delivered to the electrodes in a tonotopic fashion so that low-frequency information is sent to electrodes at the apex of the cochlea and high-frequency information is sent to electrodes at the base of the cochlea. Electrode 1 is the most basal electrode relative to the longitudinal axis of the cochlea, and electrode 22 is the most apical. The electrodes are stimulated sequentially from high to low frequency. This sequence of delivery is an attempt to mimic the path of a traveling wave moving along the basilar membrane in response to an acoustic stimulus.

The total frequency range of a MSP programmed with the MPEAK Strategy is 280 Hz to +4 kHz and is divided among the active electrodes in a linear-log arrangement in which the eight most apical (low frequency) electrodes each have a range of 100 Hz and the remaining basal (high frequency) electrodes each have a range that is 1.112 times the range of the electrode just apical to it. The actual frequency range of each electrode depends on the number of electrodes that are functional in the particular individual. The fewer the functional electrodes, the wider the ranges.

The acoustic amplitude information that has been encoded by the speech processor must be adapted to "fit" the T and C levels that were determined for each electrode during the MAPping procedure. The resultant amplitude information is reflected in the amount of electrical charge used to stimulate the electrodes. The charge is a function of electrical pulse width (microseconds) and pulse height (microamps). Electrical stimulation of the
electrodes results in a flow of current between an active electrode and an indifferent electrode. The speech processor can be programmed in either the bipolar stimulation mode, in which the active electrode and the indifferent electrode are both within the cochlea or in the common ground mode in which one electrode is the active electrode and all the other electrodes are linked together as a common reference electrode. The method most commonly used is the bipolar mode BP + 1 in which there is one electrode intervening between the active and the reference electrode.

If the acoustic input to the MSP is speech (and the strategy is MPEAK), the rate of electrical stimulation of the electrodes is a function of the presence or absence of F0 information. If the signal is a voiced phoneme, the rate of stimulation is equal to an estimate of the F0. For voiced phonemes, therefore, F1, F2, and two high-frequency band electrodes are stimulated at a rate equal to F0. If the signal is a periodic environmental sound, the rate is equal to any low frequency periodicity detected. If the signal is aperiodic or is an unvoiced phoneme, the rate of stimulation is equal to F1 (that is, the spectral peak falling between 280-1000 Hz). For unvoiced phonemes, F1 information is presented as rate information. F2 information is presented as place information, that is, as frequency assigned to a specific electrode. If F2 is less than 2800 Hz then the electrodes representing Bands 4 and 5 are also stimulated. If F2 is greater than 2800 Hz, the electrodes representing Bands 3 and 5 are stimulated.
1.5.2 Design of the Spectra speech processor

The Spectra 22 has a bank of 20 bandpass filters which are scanned continuously for the highest amplitude output (the maxima). The center frequencies of eight of the filters are linearly spaced from 150 Hz to 1550 Hz. The rest of the filters are logarithmically spaced from 1550 Hz to approximately 10,000 Hz. As in the MSP, these ranges can be adapted to the patient's requirements for the best quality representation of sound. The outputs from the six filters having the highest amplitudes are analyzed and encoded and that information is sent to selected electrodes for every cycle of the scan. Each filter is assigned to one electrode. The number of maxima selected varies with the spectral complexity of the signal. The more complex the frequency content of the signal, the more maxima selected, up to a maximum of ten. The analysis and encoding strategy used in the Spectra 22 is called SPEAK (Cochlear Clinical Bulletin, 1993, 1994).

As in the MSP, the electrodes are arranged tonotopically with low-frequency information delivered to the apical electrodes and high-frequency information delivered to the basal electrodes. Amplitude information in the Spectra is coded in the amount of electrical charge sent to each electrode with the charge being a function of electrical pulse width and pulse height. Also, like the MSP, the Spectra allows a choice of mode of electrical stimulation. The MAP can be programmed for bipolar or common ground mode.

The Spectra 22 uses an adaptive rate of stimulation varying from 180 Hz to 300 Hz. The actual rate depends on the complexity of the signal, that is, the number of maxima detected, the stimulus intensity, and the parameters
established in the patient’s MAP. The more maxima there are, the more electrodes that are stimulated, and, therefore, a longer stimulation cycle results. The stimulus intensity is reflected in the charge delivered to the electrode, and because charge is a function of pulse height and pulse width (duration), a stronger intensity signal may dictate a wider pulse which requires more time. The speech processor offers two alternative methods for controlling the pulse height/pulse width relationship; therefore, the rate of stimulation will depend on which method has been programmed into the MAP.

Both the MSP and the Spectra speech processors amplify, filter, digitize, and encode the acoustic information into signals which are delivered to a transmitting coil worn on the head just behind the ear. The transmitting coil is held in place by a magnet which is positioned over a magnet located beneath the skin. The encoded signals from the speech processor are carried by a radio frequency wave from the transmitter across the skin to a receiver/stimulator implanted in a bed which is created surgically in the mastoid bone. The receiver/stimulator converts the information carried by the radio frequency wave into charge-balanced biphasic electrical impulses which are then delivered to the electrode array in which electrodes are arranged longitudinally. The array is threaded through the round window at the base of the bony cochlea into the scala tympani where it comes to lie along the distal wall of the cochlea. Stimulation of the electrodes produces an electrical field that spreads to the cell bodies of any surviving VIIIth nerve neurons remaining in the modiolus of the cochlea. This electrical current induces an action potential in the neurons.
1.6 Scope of the present research

As noted above, studies have shown that listeners with hearing losses of cochlear origin are impaired in their perception of acoustic signals that change dynamically over time. The hypothesis presented in this dissertation is that cochlear-implant users likewise have impaired temporal acuity; hence, they are unable to track the rapid frequency transitions that are characteristic of consonants in speech. In order to test this hypothesis, the temporal resolution thresholds of cochlear-implant users were obtained from their performance on a psychoacoustic task that was a modification of one used by Madden (1990) involving the discrimination of a step signal from a glide signal. The temporal resolution thresholds of normal-hearing listeners obtained from their performance on the same task was compared to the thresholds of the cochlear-implant users.

Although it was not a goal of this study to compare the efficacy of one speech processor to another, some subjects used the older generation MSP and one subject had the newer generation Spectra.

This study is unique among psychoacoustic studies of cochlear-implant users in that the signals used were acoustic signals, and they were delivered to the cochlear-implant subjects through their speech processors via the Cochlear Audio Input Selector. Most psychoacoustic studies involving cochlear-implant subjects have used electrical signals delivered directly to the receiver/transmitter, bypassing the speech processor.
Because the author realizes that the signals used in this study are simple representations of the formant transitions in speech, measurements of performance by the subjects on a /ba, da, ga/ synthetic speech continuum were included in this study. Synthetic rather than a natural speech was chosen so that burst cues which are present in natural speech could be eliminated. In this way, more pertinent comments about the role of frequency transitions in the perception of speech could be made based on the performance on tasks using step/glide signals and performance on tasks of identification and discrimination of the elements of the /ba, da, ga/ continuum.
The ability of listeners to detect changes in a signal over time is referred to as temporal acuity. There have been several studies that have examined the temporal acuity, sometimes referred to as temporal resolution, of listeners with normal hearing, listeners with hearing impairments, and listeners with cochlear implants. One reason for the interest in temporal resolution of listeners is that a number of studies have shown that a high level of speech recognition is possible when only temporal cues are present. For example, in a study by van Tasell, Soli, Kirby, and Widin (1987) on the role of speech waveform envelope cues in consonant perception, listeners with normal hearing had a high degree of correct identification of stimuli that had been created by multiplying noise with the speech envelopes of /aCa/ natural-speech syllables. Erber (1972) found that the scores of profoundly deaf children on tests of speech reading with only visual cues improved by 7-11% when they received speech-envelope information in addition to the visual cues. Even some users of single channel cochlear implants (e.g., the Vienna device) which allow only temporal cues have performed well above chance on open-set word recognition tests (Tyler and Tye-Murray, 1991).
Some psychophysical measures of temporal processing are gap detection, forward masking, temporal integration, modulation detection, and rate discrimination. In gap detection, the shortest duration of silence between two gap markers that permits discrimination from an uninterrupted signal is taken as a measure of the gap detection threshold (GDT). In forward masking, an intense masker is followed by a brief signal; the threshold of the signal as a function of time delay from the offset of masker to onset of signal is used as a measure of temporal acuity. Temporal integration is measured as a function of the amount of increase in intensity that is required to maintain a signal at a fixed level of loudness as the duration of the signal is decreased. It is a measure of the time over which the auditory system integrates energy. Modulation-detection tasks measure thresholds for the discrimination of a modulated signal from an unmodulated signal. They measure the ability of the auditory system to follow changes in amplitude over time. Rate discrimination involves the detection of small changes in the rate at which information is delivered.

This review of the literature will begin with an examination of some of the basic psychoacoustic studies that have measured temporal acuity in the three listener categories mentioned above. A survey of studies that have addressed the perception of frequency transitions will follow. Next, studies concerning the neural mechanisms involved in coding the temporal aspects of sound will be reviewed. Finally, the literature review will conclude with a discussion of consonant perception in which elements of a synthetic speech continuum are the stimuli.
2.1 Psychoacoustic studies of temporal acuity in normal-hearing and hearing-impaired listeners

A number of psychoacoustic studies have shown that temporal acuity in normal-hearing and hearing-impaired subjects is essentially the same; others have shown it to be different in these two groups. In some instances, the temporal processing abilities of the hearing-impaired were poorer than those of normal-hearing subjects; in some instances it was better. In a comparison of the temporal acuity of normal-hearing listeners and hearing-impaired listeners with moderate sensorineural hearing losses, Jesteadt, Bilger, Green, and Patterson (1976) asked their subjects to distinguish between a Huffman sequence in which the energy at a single frequency region was delayed by a specific amount of time, T, and a sequence with the delay equal to T + ΔT. Because the energy spectra of the two sequences were identical, the listener had to make the discrimination based on when the delay occurred. A four-interval, forced-choice, adaptive procedure was used. The range of center frequencies (CFs) of the Huffman sequences was from 472-4016 Hz for normal-hearing listeners. For the hearing-impaired listeners, the CFs of the Huffman sequences matched a frequency of minimal and of moderate hearing for each individual subject. The intensities of the signals ranged from 50-90 dB SPL.

For both groups studied, temporal acuity increased as the level of the signal increased. There was no center-frequency effect observed. When the stimuli were equated in sensation level (SL), 80% of the hearing-impaired listeners with the greater degree of hearing loss had the higher degree of
temporal acuity. The authors noted, however, that had they evaluated the subjects’ performance on the basis of stimuli in dB SPL those subjects with the greater degree of hearing loss would have shown poorer temporal acuity.

The authors were unable to make an unequivocal conclusion regarding temporal acuity of the hearing-impaired listeners compared to that of normal-hearing listeners. An analysis of the data led Jesteadt et al. to conclude that the amount of training that the subjects had was a confounding factor in the study. The normal-hearing subjects in this study were trained listeners whereas the hearing-impaired subjects were naive listeners. The authors surmised that, with training, many of the hearing-impaired listeners would exhibit better temporal acuity than did the normal-hearing listeners.

Fitzgibbons and Wightman (1982), using a two-interval, forced-choice method (2AFC), measured temporal resolution in normal-hearing listeners and in listeners with sensorineural hearing losses by manipulating the duration of a gap between two successive octave-band noise signals using different CFs. The signals were presented in a continuous broadband noise so that listeners could not use as cues the spectral splatter generated as a result of sharp onset and offset. The signals were presented both at a fixed level of 85 dB SPL and at levels adjusted to correspond to 30 dB SL for each listener. Fitzgibbons and Wightman demonstrated that in both normal-hearing and hearing-impaired listeners the gap detection threshold decreased with increasing octave-band CFs. However, overall, the hearing-impaired listeners had gap detection thresholds that were significantly higher than those of the normal-hearing listeners.
A study of gap detection was also done by Florentine and Buus (1984). In order to separate the role of degree of hearing loss in gap detection from truly impaired temporal resolution, Florentine and Buus measured GDTs as a function of signal level in normal-hearing listeners, hearing-impaired listeners, and in normal-hearing listeners in whom a hearing impairment was simulated by use of a masking noise spectrally shaped to produce a specific audiometric configuration.

A 2AFC procedure was used. The signal was a continuous white noise, low-pass filtered with a cut-off frequency of 7 kHz. A gap was created by simply turning the noise off and on. The results showed that listeners with normal hearing had GDTs that decreased with increasing level. The GDTs for hearing-impaired listeners were higher than those of the normal-hearing listeners. Generally, under conditions that simulated hearing impairment, the GDTs of normal-hearing subjects matched those of the subjects with real hearing impairments. Florentine and Buus attributed the poorer gap detection abilities in the hearing-impaired subjects, therefore, as a function of the elevated hearing thresholds. At high levels, the GDTs of the normal-hearing listeners with simulated losses were lower that those of the subjects with real hearing-impairments. The authors interpreted this observation as indicating that some of the hearing-impaired subjects did have impaired temporal resolution.

Using narrow-band noise both with and without a gap and with CF below 1 kHz, DeFilippo and Snell (1986) compared temporal processing ability in normal-hearing and hearing-impaired listeners. A 2AFC procedure was used. When the signals were presented at levels that were near threshold
(5 dB SL) both groups had GDTs of approximately 100 ms. At intensities above 5 dB SL both groups showed a level effect with GDTs decreasing as the level increased. However, the thresholds for the hearing-impaired subjects were higher than those of the normal-hearing subjects at comparable sensation levels. The authors noted that one hearing-impaired subject had a GDT like that of the normal-hearing listeners and that there was a great deal of variability among the hearing-impaired subjects.

In an effort to address the equivocal results of previous studies with regard to the temporal processing capabilities of hearing-impaired listeners, Tyler, Summerfield, Wood, and Fernandes (1982) did a study in which they compared performance on a number of different temporal-processing tasks using the same listeners for all the tasks. They measured temporal integration, gap detection, temporal difference limens (TDLs), and gap difference limens (GDLs). Temporal integration reflects a listener's ability to monitor changes in a signal over long periods of time. Gap detection, as noted before, involves the ability to detect a silent interval within a signal. TDLs are a measure of the increment in duration required for a listener to detect a change in duration of a signal. GDLs are a measure of the increment in duration required for detection of a change in the duration of a silent gap within a signal. Generally, a three-alternative, forced-choice (3AFC) testing procedure was used on all the tasks.

For the temporal-integration task, 10-ms and 1000-ms pure tones of 500 Hz and 4 kHz were presented to the listeners. Detection thresholds were obtained by having the subjects adjust the level of the tones. The measure of performance was the difference between the thresholds of the 1000-ms and
10-ms tones. In the normal-hearing listeners, the difference between the thresholds of the two signals was approximately 12 dB for both frequencies. Like the subjects in Defilippo and Snell’s study (1986), the hearing-impaired subjects exhibited a great deal of variability. At 500 Hz the difference thresholds of the impaired listeners were as good as or greater than (better than) the difference thresholds of normal-hearing listeners. At 4 kHz, however, the hearing-impaired subjects showed poorer temporal integration with the differences between the thresholds of the 10-ms and the 1000 ms signals being one fourth to one half of the values for the listeners with normal hearing.

For the gap detection task, the gap-difference-limen task, and the temporal difference limen task, narrow-band noise (NBN) centered at either 500 Hz or at 4 kHz was used as the stimulus. The presentation levels for the hearing-impaired listeners were 102 dB SPL for 500-Hz NBN signals and 89 dB SPL for 4 kHz NBN signals. Normal-hearing listeners were tested at a high and a low level for each signal—102 and 82 dB SPL for the 500-Hz NBN signals and 89 and 49 dB SPL for the 4 kHz signals.

Signals of one second duration were used in the gap-detection study. GDTs were obtained by adaptively varying an 80-ms gap in the signal. For normal-hearing listeners, GDTs at the high level were 13 and 7 ms for the 500 Hz and the 4 kHz stimuli respectively. At the low level, the thresholds were 23 and 7 ms respectively. The authors noted that the thresholds for the 500-Hz signal were higher than other values previously reported in the literature. They suggested that the amplitude fluctuations in the NBN stimuli may have made it more difficult to detect the gaps. They also
suggested that the lower thresholds for the higher-frequency stimuli reflected the shorter time constants and wider bandwidths associated with higher-frequency auditory filters.

In the hearing-impaired listeners, GDTs for the 500-Hz NBN tended to be around 15-25 ms. The range of GDTs was from 11-80 ms. The GDTs at 4 kHz tended to be like those of the normal-hearing listeners; however, some hearing-impaired subjects had GDTs that were 2-3 times higher.

Gaps of 30 ms and 100 ms were used in the standard stimuli for the GDL determinations. GDLs were obtained by varying the duration of this gap in a comparison stimulus and asking the listeners to choose the signal with the longest gap. Listeners with normal-hearing had GDLs for the 30-ms 500- and 4-kHz NBN that were poorer at a low level, 66 and 60 ms respectively, than at a high level, 53 and 51 ms respectively. For 500-Hz signals with a 100-ms initial gap, the GDL was 75 ms at both high and low levels. At the higher frequency, the GDLs were 71 and 68 ms for high and low levels respectively. GDLs for the hearing-impaired listeners were extremely variable, ranging from values lower than those of listeners with normal hearing to values that were twice as high.

Five-hundred and 4000-Hz NBN stimuli standards with an initial duration of 30 ms were used in the TDL study. TDLs were obtained by varying the duration of a comparison signal. This was the only task on which every hearing-impaired subject had thresholds that were poorer than those of normal-hearing listeners. The TDLs for the normal-hearing listeners were approximately the same for the 500- and 4000-Hz signal, 18 ms
at high levels and 23 ms at low levels. The TDLs of the hearing-impaired subjects ranged between 20 and 86 msec.

To summarize the results for listeners with normal hearing: the increment required for GDL (100 ms) was significantly shorter at 4 kHz than at 500 Hz when the level of the signal was low. The GDT, TDL, and GDL (30 ms) were all significantly poorer at lower levels than at higher levels.

There was extreme variability in the performance of the hearing-impaired subjects. On all of the tasks except for the temporal-difference-limen task employing the 4-kHz NBN signal, there were hearing-impaired subjects whose thresholds were within or below the limits of normal hearing. Although many of hearing-impaired subjects had normal or near-normal thresholds, overall, the hearing-impaired subjects had reduced temporal resolution when compared to normal-hearing subjects.

As noted above, one explanation for the increase in GDTs with decreasing center frequency of the signal was that narrow auditory filters ring and that ringing obscured the gap. If this were true, then hearing-impaired subjects should have better temporal resolution thresholds since their auditory filters tend to be broader than those of individuals with normal hearing. Moore and Glasberg (1988), however, pointed out that hearing-impaired listeners did not always have smaller GDTs. They suggested that GDTs were influenced by fluctuations in the noise signals. Dips in the noiseband could be confused with a real gap in the signal. If the noiseband were narrow or if the noiseband were centered at a low frequency where the auditory filter would be narrow, fluctuations at the output of the filter would be slow and, therefore, easily confused with a gap. If this situation were compounded by
loudness recruitment in the hearing-impaired listener, then the intensity fluctuations in the noise would be magnified, and the dips would sound even more like gaps.

Moore and Glasberg (1988) measured GDTs in subjects with unilateral sensorineural hearing losses. Stimuli were presented to both the impaired and the unimpaired ear of the same subject; thus, each subject served as his own control. Two types of stimuli used. One was a bandpass noise with a center frequency of 0.5, 1, or 2 kHz and a bandwidth that was one-half of the center frequency. These bandpass noises were presented in a bandreject background noise with a notch of the same width as the bandpass noise. The signals were presented at a level of 84 dB SPL to both the impaired and unimpaired ears. The normal ear was also tested at a level equal to the sensation level of the impaired ear. The second type of stimuli used were sinusoidal markers with an introduced gap. These sinusoidal stimuli were chosen because they had no inherent fluctuations. The markers had frequencies of 0.5, 1, and 2 kHz and were phase preserved, that is, the marker at the end of the gap started at the same phase that would have existed if there had been no gap between it and the marker preceding the gap. These stimuli were presented in a continuous background noise with a sharp notch at the marker frequency. This arrangement masked spectral splatter with minimal masking of on-frequency energy. The signals were presented at 80-90 dB SPL to both ears and at equal SL to the normal ear. An adaptive 2AFC procedure was used to determine the GDTs.

The GDTs for bandpass-noise stimuli for both normal and impaired ears were higher than the GDTs obtained with sinusoidal stimuli. The authors
concluded that fluctuations in noise markers limited performance. They suggested that the narrow dynamic range associated with hearing impairments would also affect performance. As the dynamic range decreased and loudness recruitment increased, the confusion caused by the fluctuations was magnified. With the bandpass noise as the signal, the GDTs were larger for the impaired ears than they were for the normal ears at equal dB SPL. When the comparison was at equal SL, the difference in GDT between the two ears was reduced. For both normal and impaired ears, GDT decreased with increasing frequency. Moore and Glasberg hypothesized that the observed frequency effect in noise could be a reflection of the fact that fluctuations of the low-frequency signals are slow and could be confused with the gap or that the effect could reflect the auditory filter output. If the bandwidth of the noise were greater than the bandwidth of the auditory filter that the listener was using, then fluctuations at the output of the filter would be slower than at the input. For a gap in the stimulus to be detected, therefore, it would have to be longer than the dips in the envelope at the output of the filter.

For normal ears when the stimuli were sinusoidal markers, no frequency effect was observed at equal dB SPL. GDTs for all frequencies were 3.3-4.2 ms. For the impaired ears, at 1 and 2 kHz, the GDTs were 4.7 and 4.3 ms respectively; the GDT at 0.5 kHz was 8.1 ms. Moore and Glasberg noted that this GDT may have been artificially elevated because some of their hearing-impaired subjects complained that the signal was too soft at this frequency and was difficult to hear in the background noise. In the normal ears with sinusoidal stimuli at equal SL, at 0.5 and 2 kHz the GDTs increased by 100%
and 50% respectively as compared to GDTs at dB SPL. With 1 kHz sinusoidal markers at equal SL the GDTs increased about two and one half times. It should be noted that one of the problems with this paper is that no statistics were done. It is very difficult, therefore, to judge if observed differences were significant.

One explanation that Moore and Glasberg offered for the observed increase in GDTs in the normal ear when sinusoidal markers were presented at equal SL was that the ringing of the auditory filters in the normal ear may have obscured the gap. Alternatively, the authors suggested that the difference may have been due to a level effect, that at equal SL the level would have been softer in the normal ear and, therefore, difficult to hear.

In a study designed to continue the investigation of the effects of envelope fluctuations on gap detection, Glasberg and Moore (1992) increased the fluctuations in the envelope of noisebands having various bandwidths between 10 and 500 Hz. The GDTs of both hearing-impaired and normal-hearing listeners increased as a result. When the fluctuations were decreased, the GDTs decreased. Decreasing the fluctuations in signals presented to the hearing-impaired subjects resulted in GDTs that were as small as, or even smaller than, GDTs for subjects with normal hearing when unprocessed signals were the stimuli. Glasberg and Moore interpreted these results as support for the idea that loudness recruitment in hearing-impaired listeners has the effect of magnifying amplitude modulations in the signal and, therefore, leads to greater confusion of signal fluctuations with gaps in the signal.
Bacon and Viemeister (1985) noted that in early temporal acuity studies performance was highly dependent on the level of the signal. These authors proposed that temporal modulation transfer functions (TMTFs) might be a better measure of temporal acuity. In the TMTF, modulation-detection thresholds for sinusoidally amplitude-modulated (SAM) signals are expressed as a function of the frequency modulation of the signal. As the frequency modulation of a signal is increased, the modulations in amplitude are smoothed out; therefore, the listener requires a concomitant increase in depth in modulation in order to still be able to discriminate an amplitude-modulated signal from an unmodulated signal. Bacon and Viemeister pointed out that the TMTF is independent of level in listeners with normal hearing. If TMTFs could be shown to be independent of level of the signal in hearing-impaired subjects, they might serve as a more reliable measure of temporal acuity in hearing-impaired listeners.

Using broadband SAM noise to obtain TMTFs of hearing-impaired subjects having sensorineural hearing losses primarily in higher frequencies. Bacon and Viemeister found that the TMTFs had the same general shape as those of normal-hearing listeners that is, they resembled a low-pass filter function, but the thresholds tended to be higher and were level dependent. The lower the level of the signal, the higher were the thresholds. Also, at very low levels of the signal, the TMTF dropped off very sharply at a low frequency of modulation.

Bacon and Viemeister suggested that in listeners with normal hearing sensitivity to modulation decreases as the bandwidth of the stimulus decreases. The decrease in sensitivity observed in the hearing-impaired
subjects with high-frequency hearing loss may be a reflection of the
narrowness of the frequency range in which they can listen and, hence,
reflect a limitation imposed by the hearing loss upon the effective bandwidth
of the stimulus. Bacon and Viemeister prepared stimuli that simulated
bandwidth limitations such as those found in subjects with a high-frequency
sensorineural hearing loss and presented them to listeners with normal
hearing. They found that these subjects had now high thresholds for these
signals as compared to SAM noise and that performance was level
dependent.

2.2 Psychoacoustic studies of temporal acuity
in listeners with cochlear implants

With the introduction of the cochlear implant device, opportunities are
now available for winnowing out aspects of sound perception that are a
function of mechanical and biochemical peripheral processing from those
that are due to neural processing at the level of the VIIIth nerve and levels
higher in the auditory pathway. In evaluating studies of temporal processing
in listeners using a cochlear implant device, the reader must keep in mind
that extreme variability in performance is a hallmark of cochlear-implant
subjects. Every user has his/her own MAP in his/her processor; therefore,
for the same input, each user receives information at the level of the VIIIth
nerve that is unique. Many of the studies which will be discussed in this
review have shown a high degree of within- and between-subject variability.
One must also remember that while the device can provide the perception of
a signal as complex as speech to individuals classified as profoundly deaf, the user of the device is still hearing impaired. The cochlear prosthesis does not restore normal sound perception. It is also essential to keep in mind when comparing performance of cochlear-implant listeners to performance of listeners with normal hearing that users of the cochlear implant device are receiving electrical stimulation whereas normal-hearing listeners are receiving acoustic stimulation. Finally, the reader needs to be aware that in many studies of cochlear-implant users the electrical stimulus is delivered directly to the receiver/stimulator via a percutaneous connector and, therefore, has not been conditioned by the speech processor. In a few studies, an acoustic signal is input to the speech processor. Here, the acoustic information is converted into a binary code which is then delivered to the receiver/stimulator where the code is converted into electrical parameters. Finally, the reader should keep in mind that statistical analyses of the data are rare in psychophysical studies involving cochlear-implant subjects. While performance on a task may be different from that of normal-hearing or hearing-impaired listeners, it may not be significantly different.

One of the most productive investigators of psychophysical phenomenon in cochlear-implant users has been Shannon (1983, 1989, 1990, 1992, 1993a, 1993b). In his extensive 1983 study, three subjects, each implanted with a 16-channel electrode, were tested on various psychophysical tasks. One channel of the electrode was stimulated at a time; the stimuli were delivered directly to the receiver/stimulator via a percutaneous plug. No speech processors, therefore, were used in this study. Among the various psychophysical measurements made were two that were measures of temporal acuity:
temporal integration and adaptation. To measure temporal integration, a modified method of limits/Bekesy tracking technique was used. Thresholds for a 1-kHz sinusoidal electrical waveform were determined as a function of stimulus duration. Also, as a measure of temporal integration, subjects were asked to estimate the loudness of the signal at suprathreshold levels as function of stimulus duration.

As the stimulus duration was shortened the absolute threshold of the signal increased. In normal listeners, plots of threshold as a function of stimulus duration show a -3 dB slope for each doubling of duration. In other words, the threshold is proportional to the power in the stimulus. In order to maintain the signal at threshold, if the intensity of the signal is halved, the duration must be doubled (Zwislocki, 1969). For the cochlear-implant users in Shannon's study, the plots of threshold as a function of stimulus duration showed a -2 dB slope for each doubling of duration. For threshold level stimuli, one subject had a temporal integration time of 50 ms; the other two subjects had integration times of 100-300 ms. Shannon noted that these times were similar to the 100-200 ms integration times that Zwislocki found in normal-hearing listeners. At suprathreshold levels, with the amplitude of the signal kept constant, shortening the duration had no effect on the loudness of the signal until the duration was shortened to 70 ms for one subject and to 40 ms for the other two. Again, Shannon concluded that these integration times for suprathreshold stimuli were of the same order as the 50-100 ms integration times that Zwislocki found in normal-hearing listeners.
Zwislocki proposed that the difference between integration times for threshold- and suprathreshold-level acoustic stimuli was due to temporal adaptation in the neurons and that adaptation at the level of the VIIIth nerve influenced a higher central integrating mechanism which summed incoming neural energy. Shannon pointed out that Moxon (cited in Shannon, 1983) had shown that neurons of the VIIIth nerve in cats showed adaptation with electrical stimulation. Noting that differences in threshold- and suprathreshold-level integration times in his cochlear-implant subjects were comparable to the differences that Zwislocki observed in normal-hearing listeners, Shannon concluded that the central mechanism in the cochlear-implant subjects was still intact.

Adaptation in the cochlear-implant subjects was measured with stimuli of different frequencies. For all three subjects, loudness decayed as a function of time. When the stimulation frequency was above 300 Hz, loudness decreased from "moderate to loud" to near threshold or inaudibility in 20-40 seconds. For stimulation at frequencies of 200-300 Hz, loudness decay occurred at a slower rate; when the stimulation frequencies were 100-200 Hz no decrease in loudness was observed even when the stimulus was on for as long as 2-3 minutes. In normal-hearing listeners, a 5 dB SL tone of any frequency can be heard for 60 seconds without decay (Martin, 1991). Shannon proposes that these observations can be explained by events that occur in the neurons in the VIIIth nerve. In a normal ear receiving acoustic stimulation, the response of any one neuron to the stimulus is a stochastic event and the amount of adaptation may, therefore, be less than in the implanted ear where the response of the neuron to the stimulus is deterministic. With
electrical stimulation, neurons are driven harder. Harris and Dallos (1979) saw a direct relationship between length of time for recovery from adaptation and firing rate. The perception of loudness of a tone would decay more rapidly, therefore, in the electrically stimulated ear.

In measures of temporal acuity employing forward masking, Shannon determined thresholds for a signal as a function of the delay between masker offset and signal onset for different masker levels. For up to 20 ms following the masker offset, there was no change seen in the signal threshold. Then the threshold decreased in a curvilinear fashion. In Shannon's study, recovery time for the cochlear-implant subjects was longer than the 200-300 ms found in normal-hearing listeners. Even when the delay was as long as 400 ms, a significant amount of masking was still observed. Shannon again related this observation to the long recovery from adaptation that is observed in electrically-stimulated neurons.

In 1985 Hochmair-Desoyer, Hochmair, and Stiglbrunner did a study to determine if a correlation existed between temporal processing and speech understanding in subjects using the Vienna cochlear implant device. Some subjects in the study used the Vienna 4-channel intracochlear electrode; some used the extracochlear device in which the electrode was positioned at the round window niche. The psychoacoustic tasks that the authors used to measure temporal acuity were gap detection and temporal difference limens (TDLs). GDTs for a 250-1 kHz and a 2 kHz-4 kHz bandpass noise were determined. These same signals were used to determine the TDLs. The presentation level of the signals was adjusted to "most comfortable loudness
level" for both implanted subjects and normal controls. An adaptive, three-
interval, forced-choice procedure was used for data collection.

The results of the study showed extreme variability in the performance of
the cochlear-implant subjects. Generally, both GDTs and TDLs were higher
for the low frequency bandpass noise than for the high-frequency noise. As
noted previously, this was a common finding in normal-hearing and in
hearing-impaired listeners. GDTs ranged from 9 ms to 157 ms. The average
GDT for the low-frequency signal was 52 ms and for the high-frequency
signal 29 ms. The GDTs for normal-hearing controls in this study were 6
msec for the low-frequency bandpass noise and approximately 4 msec for the
high-frequency bandpass noise. None of the cochlear-implant subjects had
thresholds that were near or below those of normal-hearing listeners. The
average GDTs for hearing-impaired listeners in the study by Tyler et al. (1982)
were 23 ms for the 500-Hz NBN signal and 12 ms for the 4-kHz NBN signal.
The GDTs for the normal-hearing controls in the Tyler et al. study were
approximately twice those of the controls in the Hochmair-Desoyer et al.
study. This difference in GDTs of the normal-hearing subjects in these two
studies probably reflects the difference in the stimuli used. Fluctuations in
the NBN stimuli in the Tyler et al. study may have interfered with the
detection of gaps in the signal.

TDLs for signals with a duration of 300 ms ranged from 37 to 455 ms in the
Hochmair-Desoyer et al. study. The average TDL for cochlear-implant
subjects was 139 ms for the low-frequency signal and 102 ms for the high-
frequency signal. The average TDL for the controls was 40 ms for the low-
frequency bandpass noise and 38 ms for the high-frequency noise. It is
difficult to compare the performance of the subjects in the Hochmair-
Desoyer et al. study to the performance of the subjects in Tyler et al.'s study as
there was a ten-fold difference in the duration of the signals used in these
studies. In both the Tyler and Hochmair-Desoyer studies, the TDLs were
nearly the same for the low- and high-frequency signals. In the Hochmair
study, however, the TDLs for the control subjects were about two times as
great as the TDLs of the controls in the Tyler study. The TDLs of the cochlear-
implant users in the Hochmair study were about three times as large as the
TDLs of the hearing-impaired subjects in the Tyler study. One might suspect,
therefore, that had comparable signals been used, the TDLs of the cochlear-
implant subjects would have been comparable to the TDLs of the hearing-
impaired subjects.

Shannon (1986) reported on psychophysical measures of temporal
resolution in two additional cochlear-implant users. Temporal integration
was measured by determining the threshold of a 1-kHz pulsed electrical
signal as a function of the duration of the signal. GDTs which, as noted
before, reflect the ability of the listener to discriminate signals containing a
brief period of silence from signals which are continuous in time were
determined. Forward masking was used to measure recovery of sensitivity
to a short tone after the offset of the masker. Finally, the ability of the
subjects to discriminate between steady-state stimuli and amplitude-
modulated stimuli was measured. As in the 1983 study, the electrical stimuli
were delivered directly to the stimulator/receiver of the implant via a
percutaneous connector. The modified Bekesy tracking procedure previously
described was used for the threshold tasks; an adaptive 3AFC procedure was used for the masking and discrimination tasks.

The high degree of between-subject variability that Shannon observed in his 1983 study was also seen in the tasks used in this study. One subject had nearly normal temporal integration. For durations of less than 50 ms, halving the duration required a doubling of intensity to maintain equal loudness. One subject showed no evidence of temporal integration ability. For this subject, the threshold for signals longer than 5 ms was unaffected by the duration of the signal. Shannon noted that this latter subject was a “star” user of the implant device and that he had previously tested both “star” and “nonstar” users and had found impaired temporal resolution in both categories. He pointed out that temporal-integration abilities, therefore, were not predictors of good speech perception.

The stimulus in the gap-detection study contained two 2-kHz sinusoidal markers of 200-ms duration, one preceded and one followed the gap. The results of the gap-detection procedure showed that the GDTs of the implant subjects was dependent on the level of the signal. Near threshold levels, a gap of approximately 30 ms was required for detection. The best GDTs achieved were at a level of 10 dB SL; here, thresholds fell to approximately 1 ms. Shannon noted that the best GDT for listeners with normal hearing occurred at 40-50 dB SL. He concluded that gap detection in cochlear-implant users was similar to that in normal-hearing listeners at similar loudness levels. Shannon noted that the GDTs at threshold levels that Hochmair-Desoyer et al. (1985) observed in their cochlear-implant subjects were about twice as large as those that he observed. He attributed this
difference in GDTs to differences in the stimuli, claiming that fluctuations in broadband noise stimuli interfere with detection of gaps in the signal.

The results of the forward-masking procedure were much like those Shannon observed in his 1983 study. There was no change in masking for delays of up to 20 ms; then, the amount of masking decreased curvilinearly. A significant amount of masking remained even 500 ms after the offset of the masker. A level effect was observed; for a given duration after masker offset, higher signal thresholds occurred at higher levels of the masker. Shannon claimed that if the data from cochlear-implant users and normal-hearing subjects were plotted in terms of loudness, then forward masking results would be similar.

In the amplitude-modulation detection task, modulation thresholds rose rapidly above 150 Hz. Below 150 Hz, the cochlear-implant subjects were able to detect amplitude modulations of only 2-4%. Shannon notes that this was significantly greater sensitivity than he observed in the normal-hearing controls in his 1983 study.

Shannon concluded that the plotting of the results of psychophysical measurements of temporal processing in the same units makes the performance of cochlear-implant users and normal-hearing listeners appear different. He believed that the observed differences merely reflect differences in the dynamic ranges and in loudness recruitment. When performance is compared on the basis of equal loudness levels, then subjects with implants perform about the same as subjects with normal hearing.

The issue of differences in GDTs for low- and high-frequency signals was studied extensively in cochlear-implant users because these subjects
provided an opportunity for investigation of the phenomenon in the absence of the peripheral mechanisms of the ear. The location of temporal processing in the auditory pathway could then be more clearly defined. As noted earlier (Florentine and Buus, 1984), one hypothesis was that GDTs were lower for high-frequency signals and broadband noise, because the response time of the auditory filters for low-frequencies was longer. The ringing of these filters smeared the gaps in the signal; therefore, periods of silence were difficult to detect. Another hypothesis was that GDTs were due to differences in traveling time of the wave of displacement along the basilar membrane (Duifhuis, 1973). If it could be demonstrated that there were no differences in GDTs at high- and low-frequency signals in cochlear-implant subjects, then the differences that were observed in normal-hearing listeners could be attributed to the mechanics of the basilar membrane.

One of the early studies of gap detection in cochlear-implant users was that of Moore and Glasberg (1988) in which they compared gap detection in normal and impaired ears of the same hearing-impaired subject to gap detection in cochlear-implant subjects. The three implanted subjects in this study wore a single-channel extracochlear device in which stimulation was at the promontory. Subjects using this type of device could receive frequency information via the rate of electrical stimulation (Cooper, 1991) but would not have access to the range of frequency selectivity provided by the basilar membrane. The stimuli used were of two types. In one type, noise markers preceded and followed the gap. The markers had CFs of 100 Hz with bandwidths of 100 or 200 Hz, or CFs of 200 Hz with bandwidths of 200 or 400 Hz, or CFs of 400 Hz with bandwidths of 200, 400, or 800 Hz. In the second
type of stimulus, sinusoids were used as markers. The markers were in "standard-phase condition" that is, the marker preceding the gap ended with a positive-going zero crossing; the marker following the gap started with a positive-going zero crossing.

The results showed that with noise stimuli, the GDTs for the implanted subjects were larger than those for normal and impaired ears. The thresholds also increased as the CF of the stimulus increased. This was not consistent with the observation in normal and impaired ears of the other listeners tested; in these subjects, GDTs decreased with increasing center frequency. Moore and Glasberg suggested that the differences between cochlear-implant subjects and subjects with normal/impaired ears were due to a reduced dynamic range at higher frequencies; therefore, the most comfortable loudness level may have been at a lower sensation level at the higher frequencies. The intensity/loudness function for electrical stimulation is very steep (Shannon, 1993b), and loudness recruitment may have been greater at higher center frequencies. Consequently, dips in the noise were magnified and confused with gaps. Another possible explanation for the observation of higher thresholds in cochlear-implant subjects compared to those of normal controls is that the signals used in this study may have been at the limits of frequency perception of the cochlear-implant subjects. Studies have shown that cochlear-implant users can distinguish changes in rate of electrical stimulation for frequency rates of up to only 300-500 Hz (Cooper, 1991).

It was observed that GDTs for noise decreased with increasing marker bandwidth. One might anticipate this result. As the bandwidth increased,
fluctuations in the noise would become more rapid and, therefore, would be less confusable with the gap.

GDTs with sinusoidal markers were measured at 100, 250, 400, and 1kHz for only one implanted subject. The thresholds were observed to be 4 ms for the 100 Hz markers, 3.5 ms for the 400 Hz markers and about 6 ms for the 1kHz sinusoidal markers. The implanted subject experienced a tingling sensation with the 1 kHz signal which may have been a confounding factor in the threshold determination at this frequency. These thresholds were comparable to GDTs of the normal ears of the other subjects. These GDTs for sinusoidal markers were smaller than for the same subject for noise markers. Again, the authors suggested that the GDTs for noise markers were higher because the gaps were confused with dips in the noise. GDTs for sinusoidal signals varied very little with frequency in both the implanted subjects and in the normal ears.

The GDTs were also measured for sinusoidal markers at different levels. Although there was considerable between-subject variability, GDTs decreased with increasing levels. At high levels (3-4 dB re most comfortable loudness level) thresholds of 1-3 ms were achieved by the implanted subject with the best performance. These values are comparable to those found by Shannon (1986) in multichannel implant users.

Because the performance of the cochlear-implant subjects was often similar to that of normal-hearing listeners in this study and because differences in performance could be related to differences in dynamic range, Moore and Glasberg concluded that temporal resolution per se was not severely compromised in listeners with a cochlear implant.
Detection of gaps in electrical sinusoids and in biphasic pulse trains were compared by Shannon (1989) in subjects implanted with either the Nucleus 22-channel device or the Symbion 4-channel device (also known as the Ineraid). Stimuli were delivered via a percutaneous connector in the Symbion; both sinusoidal and biphasic pulsatile stimuli were used. Only biphasic pulsatile stimuli were used with the Nucleus device. An adaptive 2AFC procedure was employed for data collection.

Generally, GDTs were lower for users of the Symbion than for users of the Nucleus device. According to Shannon, there is greater spread of current with the monopolar electrode configuration used in the Symbion than there is with the bipolar electrode configuration used in the Nucleus. Monopolar thresholds are usually lower than bipolar thresholds (Shannon, 1983). As part of the 1989 gap detection study, Shannon compared thresholds obtained from stimulation of adjacent electrodes and from electrodes that were widely separated (thresholds were for only subjects using the Nucleus device). Gap detection thresholds as low as 1.5-3.1 ms were obtained with both arrangements; however, a higher level of stimulation was required when electrodes were closely spaced.

One of the neural factors that might be involved in gap detection is phase locking to the stimulus. Hartman (1984) showed that, in the cat, synchronization of neuron firing was stronger to electrical than to acoustic stimuli and stronger to pulsatile than to sinusoidal stimuli. Shannon pointed out that, based on Hartman’s evidence, one would expect better GDTs in cochlear implant listeners than in listeners with normal-hearing and that the GDTs for pulsatile stimulation would be lower than those for
sinusoidal stimulation. The results of this 1989 study, however, showed lower thresholds for the Symbion users who received mostly sinusoidal stimulation than for the Nucleus users who received only pulsatile stimulation. One Symbion user was tested with both sinusoidal and pulsatile stimulation; the sinusoidal GDTs at the same stimulus level were actually lower than the pulsatile GDTs. For low level signals, the implant users had GDTs of 20-50 ms. When the stimuli were presented at high levels, the GDTs fell to 2-5 ms. Shannon noted that these values were comparable to GDTs for comparable acoustic stimuli in normal-hearing listeners. Values of 30-50 ms near quiet threshold and 3-5 ms for high level signals for listeners with normal hearing had been previously reported in the literature. Because GDTs were similar with electrical and acoustic stimuli and because the performance obtained with pulsatile and sinusoidal stimulation did not fit the predicted results, Shannon concluded that enhanced neural synchrony was not a significant factor in gap detection.

No differences in GDTs were observed between stimulation of apical and basal electrodes. Shannon noted that his results were like those of Moore and Glasberg (1988) and concluded that gap detection processes which take place in the central auditory pathway are independent of tonotopicity in the cochlea.

Normal-hearing controls were not used in this study; however, Shannon analyzed data obtained by other investigators from normal-hearing subjects receiving acoustic stimulation and concluded that it was not proper to compare the two groups on the basis of levels in dB SPL or even dB SL. He thought GDTs might be comparable if the levels of the stimuli were equated
on the basis of loudness levels. In fact, he did ask the implanted listeners to estimate the loudness of the stimuli and used those estimates to construct a plot of the GDTs. Although there was a high degree of variability, this plot was monotonic up to a loudness of 6 on a scale of 1-10. At this point further increases did not produce decreases in threshold. This plot looked very similar to the ones obtained by Florentine and Buus (1984) for normal-hearing listeners receiving acoustic stimulation. The independence of threshold and level at high levels of signal was also observed by Moore and Glasberg (1988) in the implanted subjects in their gap detection study.

Shannon concluded that the similarity of thresholds in cochlear-implant and normal-hearing listeners suggested that peripheral neural activities involved in gap detection are not the significant factors in good temporal resolution and that there is no retrocochlear loss of auditory temporal resolution with a sensorineural hearing loss. He suggested that comparisons of performance in normal-hearing and in hearing-impaired listeners should be made on the basis of perceptual rather than physical units.

A comprehensive study comparing the performance of cochlear-implant subjects using one of five different devices was done by Tyler, Moore, and Kuk (1989). Measures were made of word recognition in word lists and sentences (in the language of the subject), environmental-sound perception, and gap detection. The devices used were the Chorimac, the Nucleus, the Duren/Cologne, the 3M/Vienna, and the Symbion. The Chorimac is a 12-channel device which is characterized by bandpass filters and nonsimultaneous pulsatile electrical stimulation of the electrodes. The Nucleus, as was noted in a previous section, is a 22-channel device
characterized by a feature-extraction strategy and nonsimultaneous pulsatile stimulation of the electrodes. The Duren/Cologne has a 14-electrode plate that is fixed to the bony labyrinth on the medial wall of the middle ear cavity. It features bandpass filters and nonsimultaneous pulsatile stimulation of the electrodes. The 3M/Vienna is a single-channel device that uses analog stimulation of the electrode. The Symbion, as noted previously, is a 4-channel device and employs simultaneous analog stimulation of the electrodes.

The stimuli used in the gap-detection study were octave bands of noise centered at 500 Hz. The total duration of the signal was one second; the first noise marker was 500 ms and the second was 500 ms minus the duration of the silent gap. Gap duration varied between 2 and 300 ms. The stimuli were presented in a sound field at a level of 58 dB SPL. The subjects were instructed to adjust their devices so that the signal was "loud and comfortable". It is pertinent to note here that both Moore and Glasberg (1988) and Shannon (1986) showed that gap detection is level independent at high signal levels.

The mean GDTs for normal controls in this study were 13 ms with a range of 8 to 27 ms. For the Symbion device, mean GDTs were 17 ms with a range of 8-30 ms; for the Nucleus the average GDTs were 19 ms with range of 16-26 ms. With the single-channel 3M/Vienna device thresholds were 22.3 ms with a range of 14-36 ms. GDTs for users of both the Chorimac and Duren/Cologne devices were very high. Chorimac users had mean GDTs of 67 ms with a range of 17-200 ms; Duren/Cologne users had average GDTs of 83 ms with a range of 11-195 ms. As one can see, there was a very high degree
of variability both across and within devices. A significant correlation between gap detection and environmental sound perception and vowel recognition was observed in this study.

The authors noted that in analog devices the stimulus waveform is largely represented in the electrical stimulation waveform and that in the Symbion simultaneous multichannel stimulation would allow correlation of envelope changes across channels; therefore, one might predict better performance with this device. In devices which deliver a pulsatile electrical signal, noise produces a quasi-random pulse train. The authors suggested that this makes detection of gaps more difficult. The problem with these conclusions is that there were no statistical analysis of the GDT data to determine if the differences observed between the Symbion, Nucleus and 3M/Vienna were significant. The average GDT from each of the three devices fell well within the range of performance for normal-hearing listeners. The surprising result is how well the users of the 3M/Vienna device did considering that it is a single-channel device.

Using the same approach as Bacon and Viemeister (1985), Shannon measured temporal resolution in cochlear-implant users by determining TMTFs (1992). As noted in the previous section, a TMTF is the plot of thresholds for detection of amplitude modulation of a signal as a function of the frequency of modulation. Sinusoidal electrical stimuli were used because inherent fluctuations in noise might interfere with the detection of amplitude modulation. Normal-hearing listeners could use the sidebands generated by interaction of the sinusoidal carriers with the amplitude modulator (Kay, 1982), but Shannon points out that this was not a problem
for cochlear-implant users because of their limited spectral resolution. TMTFs were obtained by making measurements of: absolute thresholds of low-frequency sinusoidal electrical stimuli, the ability to detect envelope fluctuations caused by beats between two high-frequency sinusoidal stimuli, and the ability to detect sinusoidal amplitude modulation of a high-frequency carrier waveform. For the beat detection study a pulsed signal was presented in a continuous masker. A modified Bekesy tracking technique (Shannon, 1983) was used to determine the threshold for detection of the pulsed signal. A 2AFC choice procedure was used for determination of modulation thresholds.

The three implant devices used by the subjects in the study were the Ineraid (also known as the Symbion), a four-channel device that stimulates the electrodes with a continuous analog electrical signal, the 22-channel Nucleus which delivers a biphasic pulsatile electrical stimulus to the neurons of the VIIIth nerve, and the UCSF, an experimental 16-electrode device. For the Nucleus device, the pulse duration was modulated sinusoidally while the pulse height was held constant. Shannon noted that, strictly speaking, this was not perceptually equivalent to amplitude modulation but that this method did produce a sinusoidal modulation of the charge/pulse delivered to the electrode.

The results of the beat-detection study indicated that the listeners were actually hearing beats and, therefore, perceiving the envelope fluctuations caused by the interaction of the two sinusoids. They reported that when the frequency of the pulsed signal was below that of the continuous masker, the perceived pitch of the pulsation was lower than it was when the pulsed
signal was above the frequency of the masker. Also, Shannon cited the marked decrease in threshold for the pulsed signal in the presence of the masker as the frequency of the pulsed signal approached that of the masker as evidence that the subjects were using beats as cues. TMTFs generated by plotting the data obtained from the three methods of measuring sensitivity to amplitude modulation as described above had the same shape, and all showed that maximum sensitivity to modulation of the electrical sinusoids was at 50-100 Hz. Greater modulation was required for frequencies above 100 Hz. Shannon concluded that the similarity in TMTFs suggested a common temporal mechanism underlying all three sets of data. Sensitivity to modulation increased as a function of level for both the beat-detection task and the modulation detection task. In fact, some implanted subjects could detect as little as a 1-3% modulation in amplitude at signal levels of 15-20 dB SL.

Shannon made a comparison of the data from his study to that obtained by Bacon and Viemeister (1985) for normal-hearing listeners receiving acoustic stimulation. TMTFs for normal-hearing listeners were low-pass functions whereas TMTFs for implanted listeners were either low-pass or bandpass functions. The 3-dB-down frequency for the TMTFs of the implanted subjects was estimated at 100-200 Hz. For acoustic stimulation, the 3-dB-down frequency, as estimated from Bacon and Viemeister’s data, was 69 Hz. Shannon attributed this difference to differences in dynamic ranges for acoustic and electrical stimulation. A small change in amplitude modulation might be a small perceptual change for the normal-hearing listener but be a large one for the implanted listener. Finally, in normal-
hearing listeners, TMTFs are nearly independent of level above 20 dB SL (Bacon and Viemeister, 1985); in Shannon's study, the TMTFs of the implanted listeners were highly dependent on level. The best acoustic listeners could detect a modulation of 5% (Viemeister, 1979); in Shannon's study, as noted above, the best electrically-stimulated listeners could detect a modulation of 1-3%. Shannon attributed this difference to the type of signal used, pointing out the inherent fluctuations in the noise might result in poorer thresholds.

In another study designed to determine the sensitivity of cochlear-implant users to amplitude modulation, Busby, Tong, and Clark (1993) used a four-interval, forced-choice, adaptive procedure to measure detection thresholds for modulations of pulsatile electrical stimuli. Three different modulation thresholds were determined. First, detection thresholds were determined for modulated-pulse durations of a series of modulation frequencies at each of four pulse rates: 100, 200, 500, and 1000 pps. (The authors noted that higher rates of stimulation produce a better representation of the stimulus waveform.) Second, detection thresholds for modulated-pulse durations using three reference pulse durations were obtained. This was done to determine if the level of the reference might influence the detection thresholds. Variations in acoustic amplitude are coded as variations in pulse duration. Different durations, therefore, represent different levels. Finally, difference limens for discrimination of modulation depth at a series of different modulation frequencies were determined. The subjects were implanted with the 22-electrode Cochlear Pty. Limited device. All the stimuli were 500 ms in duration and all were presented at a comfortable
listening level. Stimuli were presented directly to the receiver/stimulator and to only a single electrode of the array at any one time.

In the first task, subjects were to discriminate between an unmodulated reference signal and a modulated comparison signal. The duration of the comparison signal was modulated at different frequencies at the four pulse rates. Because the minimum modulation period was equal to four pulses of pulsatile electrical stimulation, the maximum modulation frequencies were 25 Hz for 100 pps, 50 Hz for 200 pps, 125 Hz to 500 pps, and 250 Hz for 1000 pps. For a given modulation frequency, shorter pulses resulted in longer silent intervals between pulses.

The second task also required discrimination between an unmodulated reference signal and a modulated comparison signal. In this task, however, the comparison signal was modulated about the reference signal pulse duration which was either 50, 100, or 300 ms. In the third task, difference limens for discrimination of modulation depth were determined for the same frequencies as those used in the first task. The reference stimulus was modulated at a fixed 50% sinusoidal modulation.

The results from all the tasks were highly variable and were influenced by whether the subjects were pre- or post-lingually deafened. Modulation detection thresholds at all pulse rates were lower for post-lingually deafened subjects. For three of the four post-lingually deafened subjects, the TMTFs resembled low-pass filter functions with cut off frequencies of 50-100 Hz. These results were similar to those observed by Shannon (1992). The TMTFs for pre-lingually deafened subjects also resembled low-pass filter functions but were not as clearly defined. The TMTF of one subject had a cutoff
frequency of 4-5 Hz; another had a cutoff frequency of 100 Hz. The third pre-
lingually deafened subject (P7) had a function with two passbands, one at 4-5
Hz and one at 100-125 Hz.

Although the results were quite variable between subjects, the TMTFs for
detection thresholds for modulation of the comparison signal duration about
the pulse duration of the reference signal tended to resemble a low-pass filter
function at 1000 pps and were essentially flat at 500 pps. The thresholds were
somewhat lower when the stimulation rate was 1000 pps. There were
essentially no differences between reference pulse durations at any of the
stimulation rates.

The results for the determination of difference limens for modulation
depth were highly variable. Generally, the TMTF for the difference limens
was similar to the transfer function for detection thresholds i.e., vaguely
resembling a lowpass filter function with a cut off frequency at 100 Hz.
Interestingly, the TMTF for P7 closely resembled his TMTF for modulation
detection thresholds.

In a study of the relationship between auditory temporal resolution and
open speech recognition, Muchnik, Taitelbaum, Tene, and Hildesheimer
(1994) used acoustic stimulation to test the cochlear-implant subjects and
controls with normal hearing. The implanted subjects all had the Nucleus
device and used the Cochlear Corp. Wearable Speech Processor (WSP) which
uses a feature extraction strategy of F0, F1, and F2. Approximately one-half of
the subjects had no significant open-set speech recognition (NOSSR) and half
had open-set speech recognition (OSSR) of different levels. The stimuli in
this study were noise bursts with durations of 10-85 ms. The frequency
spectrum of the noise was 100 Hz-4 kHz with the maximum energy at 100-500 Hz. Stimuli were presented to the subjects through a loudspeaker at a comfortable listening level.

GDTs were significantly lower in cochlear-implant subjects with OSSR than in cochlear-implant subjects with NOSSR. The values ranged from 12-46 ms for durations of 85 to 10 ms for the OSSR group and from 28 to 71 ms for the NOSSR group. GDTs consistently increased with decreasing burst duration only in the cochlear-implant subjects with OSSR. Post-hoc comparisons showed that the two groups differed significantly only for the 85-ms duration stimulus. There was a large amount of variability in performance, and the authors attributed the lack of significant differences at the other durations to this variability.

Although there was a high degree of variability in performance among the subjects of these studies of temporal resolution in cochlear-implant users, it appears, in general, that listeners with cochlear implant devices have temporal-processing abilities that are fairly normal or are even at times better than normal.

Shannon (1993a, 1993b) pointed out that when the differences in the dynamic ranges of listeners with cochlear implants and listeners with normal hearing are taken into account, temporal processing as measured by forward masking, gap detection, and modulation detection appeared to be similar in these two groups of listeners. Two factors might account for those instances in which measurements of temporal resolution were better in implanted listeners than in normal-hearing listeners. The elimination of the time required for the signal to travel along the basilar membrane may be
one factor, and the elimination of the time required for biochemical events involved in the synaptic transmission between hair cells and neurons of the VIIIth nerve may be another.

2.3 Perception of frequency-modulated signals in normal-hearing and hearing-impaired listeners

In most of the studies discussed above, measures of temporal resolution were made using signals that did not change over time. Most auditory events in the human listening environment are dynamic, they do change in time. Conclusions regarding temporal resolution made from data obtained using static stimuli may not be applicable to temporal resolution of dynamic acoustic events. Speech is a highly dynamic event and, as noted previously, some of the dynamic elements of speech, the formant transitions, have been shown to be important in the perception of speech. Many hearing scientists have, therefore, done studies of temporal resolution using frequency-modulated (FM) signals, particularly unidirectional FM signals that mimic frequency transitions.

One of the earliest studies of the sensitivity of listeners to unidirectional frequency modulation was that of Sargeant and Harris (1962). They found that a rise of only 5 to 7.5 Hz was required for listeners with normal hearing to experience the sensation of a "glissando" in signals with durations between one and ten seconds.

In a psychophysical experiment designed to further understanding of the perception of FM tones, Pollack (1968) varied three parameters of the signal:
initial starting frequency, rate of change of frequency with duration, and duration of presentation. In a one interval task, listeners with normal hearing were presented with a pure tone signal that was linearly modulated in frequency in either a positive or a negative direction. The listeners’ task was to indicate the direction of frequency change. The results showed an inverse relationship between the detection threshold for direction of frequency change and duration, the shorter the duration of the signal, the greater the rate of frequency change required to correctly distinguish a positive from a negative frequency change of the same magnitude. In other words, the total frequency sweep had to be longer at short durations for a just noticeable difference in direction. This finding was true for initial frequencies between 125 and 1 kHz. An effect of initial frequency on the detection threshold was observed. For a given duration, thresholds were larger at higher than at lower initial frequencies.

The author thought that because the initial frequency was constant during successive presentations, subjects were using spectral cues, that is, differences in the final frequencies of the signals, rather than directional cues in their discrimination of rising and falling transitions. Random presentations of the starting frequencies were used, therefore, so that the listeners couldn’t easily use the starting frequency as a reference against which the final frequency could be compared. The detection thresholds did not increase with random presentations of the signal initial frequency. Pollack concluded that the listeners were responding directly to the direction of signal movement.

One of the problems that arises in a study of dynamic acoustic signals is a confounding of the effects of duration, extent of the frequency change, and
rate of change. For a fixed duration, as the extent of frequency changes, so does the rate. The question arises as to whether the subject is responding to the change in extent of the sweep or to the difference in rate of change. The same problem arises when the frequency sweep is fixed, but the duration of the transition is varied. Pollack was aware of this problem and in an attempt to circumvent it, he used signals that were swept tones consisting of an initial and a terminal steady state with a frequency transition segment that linked the two steady-state segments. By keeping the total duration of the steady-state portions long relative to the total duration of the transition, the primary variable would be the slope of the transition.

In a task employing these signals, listeners were asked to indicate whether a transition was perceived as "fast" or as "slow". The total frequency difference between initial and final frequencies of the transition was held constant while the duration of the transition was varied. The initial steady-state portion was constant at 1 ms; the listeners were required to respond before the end of the signal so that the duration of the final steady-state portion could not be used as a cue. When the difference between the final and initial frequencies was very small (4 Hz), the threshold for discrimination of the rate of frequency change was independent of duration of the transition. Regardless of whether the transition duration was less than one ms or as large as 100 ms, an increase in transition duration of approximately 1000 ms was required to just detect a difference in rate of frequency change. When the frequency difference was large (512 Hz) the difference in transition duration required to just detect a difference in the
rate of frequency change was approximately a constant proportion of the transition duration.

Nábelek and Hirsh (1969) also were aware of the problems that are involved in psychophysical measurements of sensitivity to frequency modulation. They noted that given a fixed initial and final frequency of a tone burst, manipulating the rate of frequency change by varying the duration of the transition would produce differences in loudness. If duration is held constant and the initial or final frequency varied, then the listener might use pitch cues in discrimination tasks. In a study of discrimination of frequency transitions, Nábelek and Hirsh, used tone bursts which consisted of an initial linear change in frequency followed by a steady-state portion. With this type of signal the transition portion could be varied, yet overall frequency range and duration could be fixed; hence, pitch and loudness could be controlled.

The following variables were examined: effect of transition duration, effect of frequency region, effect of frequency interval, effect of frequency direction. Transitions rose or fell from frequency regions of either 250, 1000, or 4 kHz. Initial frequencies of rising transitions and final frequencies of falling transitions were 1/2, 2/3, or 8/9 of each of these three frequencies. Durations of the transitions were 10, 30, 100, and 300 ms. Total duration of the signal was four times as long as the duration of the transition.

In a same/different (AX) procedure, listeners with normal hearing were presented with a reference tone and a comparison tone whose transition duration was the same or larger than that of the reference tone for a given frequency change.
Discriminability was measured in DLs. There was little difference in DLs for rising and falling glides. The data showed that DLs were smallest when the transition duration of the reference burst was 30 ms. The authors pointed out that this duration was within the range of formant transition durations found in speech. There was an effect due to frequency interval. DLs were largest for small frequency intervals and were approximately equal for the medium and large intervals. The effect due to frequency interval was more pronounced for 10 and 30 ms than it was for 100 and 300 ms durations. An effect due to frequency region was also observed. The smallest DLs were always for signals that rose or fell from 1000 Hz, and the largest were for the low frequency region (250 Hz). It must be noted that no statistical analyses of the data was done.

Tsumura, Sone, and Nimura (1973) examined the ability of listeners with normal hearing to follow frequency transitions by using signals that were linearly modulated in frequency during either the entire duration of the burst or during only a portion of the burst. In an initial experiment in which the transition occupied only a portion of the signal, the frequency of the burst was 1 kHz, and the duration varied from 32-1000 ms. The frequency transition was varied in duration from 5 to 300 ms. Subjects were presented with a reference tone of constant frequency and a comparison tone containing either a rising or a falling transition. They were required to adjust the frequency of the comparison tone until it was just noticeably different from the reference tone. The parameters that were varied were delay of transition onset, duration of the final steady-state portion, or transition duration. The total duration of the tone burst was 1 second.
Unlike the subjects in Nábelek and Hirsh's study (1969) who showed no
difference in sensitivity to rising and falling signals, subjects in this
experiment generally had higher thresholds for detection of rising than for
falling transitions. Frequency DLs were greatest when the transition
occurred near the onset of the signal. When durations of the initial and final
steady-state portions were 10-30 ms, DLs ranged from 30-4 Hz. This spread
was observed because under conditions of short duration steady-state
portions, there was an interactive effect of duration of the steady-state
portion with the duration of the frequency transition. DLs decreased as the
duration of the transition increased. Frequency DLs of 3 Hz were obtained
when the transition occurred near offset i.e., when long (>300 ms) initial
steady-states were used, and when long (>100 ms) final steady states were
used. The DL of 3 Hz is a little smaller than the 5-7.5 Hz rise that Sergeant
and Harris (1962) found was necessary for listeners to perceive a glissando in
signals with durations of 1-10 seconds.

In a second experiment, the DL required for detection of frequency
transitions was obtained using short (32, 102, or 302 ms) tone bursts with the
transition occupying either only a portion of the signal as in the previous
experiment or with the transition occupying nearly all of the tone burst
duration. As in the previous experiment, initial and final steady-state
portions were varied. Seven subjects with normal hearing listened to pairs
of bursts, one member of which was the unmodulated reference tone burst;
the other member, the comparison tone, contained the frequency transition.
Subjects were asked to indicate which interval contained the transition. The
results were similar to those of the previous experiment in that when the
initial steady-state portion was small, i.e., the transition was near the signal onset, there was sharp decrease in the frequency DL as the duration of the transition increased. It should be noted that no statistical analyses were done.

The authors concluded that the pitch difference between onset and offset frequencies was the dominant cue for detection of the frequency transition as long as the final and initial steady-state portions were "sufficiently" long. They further suggested that the detection of frequency transitions is based on two mechanisms: one involves a detection of pitch differences between initial and final frequencies and the other is based on detection of gliding pitch.

Arlinger, Jerlvall, Ahrén, and Holmgren (1977) investigated the threshold for frequency change required for the perception of frequency transitions in subjects with sensorineural hearing losses. The stimulus was a continuous 1 kHz pure tone which was frequency modulated with randomly variable interstimulus intervals (ISIs). The authors did not specify the frequency sweep other than to state that six different ones were used. A rising frequency modulation was followed by a steady-state portion which, in turn was followed by a falling transition. The sequence of rising transition, steady-state portion, falling transition was followed by a interstimulus interval of randomly varied durations of at least 2 seconds. The duration of the transition was varied and was either 10, 20, 100 or 500 ms. The duration of the rising transition plus steady-state portion was fixed at 500 ms. The duration of the falling transition was fixed at 600 ms. The stimulus was presented at each subject's most comfortable listening level. Only data from
rising transitions was collected. The authors noted that they did not collect data from falling transitions because data from listeners with normal hearing indicated that there was no asymmetric difference in sensitivity to the two directions (Arlinger et al., 1976).

The authors compared the data from the hearing-impaired listeners with that obtained from listeners with normal hearing who participated in their 1976 study and found that thresholds for the subjects with sensorineural hearing losses were significantly higher at all durations tested. In both the normal-hearing and hearing-impaired listeners, thresholds were independent of the duration of the transition for durations of 10, 20, and 100 ms. Thresholds for the normal hearing listeners were 3 Hz (Arlinger et al., 1976) and for the listeners with sensorineural hearing losses 8 Hz. When the duration of the transition was 500 ms, however, the thresholds for both groups nearly doubled. If, as the authors indicated, the transition and steady state portions of the signal were fixed at 500 ms, then under these conditions, the rising and falling transitions would have been abutted. It is difficult to compare this study to that of Tsumura et al. (1973) because of the variable ISIs used by Arlinger et al. and the fact that the signal included both rising and falling transitions.

Using listeners with normal hearing, Nábelek (1978) made a comparison of masked thresholds for constant and for upward and downward gliding tones. The frequency regions of the stimuli were 250 Hz-1 kHz, 250 Hz-2 kHz, and 825 Hz-3.3 kHz. Within these frequency regions, comparisons were made of long, 250-2 kHz, and short, 1.8-2 kHz, transitions. A variety of durations from 0.5 to 5000 ms were used. The glides were linearly
modulated in frequency. The wideband noise masker was presented at a level of 70 dB SPL. Masked thresholds were measured by the method of adjustment.

There was little difference in absolute thresholds across frequency regions for a given direction and duration of the signal. For all frequency regions used, and for both short and long frequency transitions, thresholds fell as a function of duration. Generally, for the different frequency regions used, the results showed that for signals of less than 10 ms the thresholds for falling frequency transitions were higher than those for rising transitions; the thresholds for constant tones fell between those of the rising and falling transitions. Between durations of 10 and 200 ms, thresholds for both rising and falling transitions, were higher than the thresholds for constant tones; thresholds for falling transitions were the highest. Beyond 200 ms the thresholds for rising and falling transitions were similar and approached or fell below those of constant tones. Differences between thresholds of transitions and of constant tones were smaller for narrow frequency sweeps than for wide ones.

In the discussion, Nábelek addressed possible hypotheses for the observed results. He noted that a tone whose frequency is changing has the same energy as a constant tone of equal duration and intensity. Perfect integration should produce the same thresholds for constant tones and frequency transitions; however, as Nábelek’s results showed, that does not occur. He notes that the spectrum of a glide is broader than that of a constant tone, therefore, the average power/Hz (average power density = $I_o$) is less and that for transitions of a given duration, the average power density decreases as
the frequency sweep increases. In addition the ratio of $I_o$ to $N_o$ (the average noise power/Hz=$N_o$) also decreases as the width of the frequency sweep increases. If only $I_o/N_o$ were the relevant cue one would expect that thresholds would be the same for rising and falling transitions of the same frequency sweep and duration and that thresholds for constant tones would be lower than for swept tones. As Nábelek's results showed, the thresholds for rising and falling transitions were not the same for a given sweep and duration, and thresholds for rising transitions were lower than those for constant tones at durations <10 ms.

Another hypothesis that the author considered was that summation across critical bands occurred as the signal swept across them. This hypothesis was ruled out because it did not account for the differences between rising and falling glides. Also, the observed transition thresholds were smaller than those predicted by a model of summation across critical bands. The final hypothesis that Nábelek offered was that the results reflected adaptation effects. At long durations, the data showed lower thresholds for glides compared to those of constant tones. He suggested that because the frequency of the stimulus changed constantly in the glide, adaptation effects were eliminated.

Masked thresholds for tone glides were also determined by Cullen and Collins (1982) using glides with durations of 5, 10, 20, and 40 ms. The signals were modulated linearly at rates of 24, 48, 96, and 192 Hz/ms. In this study, rates of frequency change were deliberately varied. For a given signal of a given duration, increasing the rate of frequency change had the effect of lengthening the extent of the frequency sweep. In some of the studies
previously discussed (Ndbelek, 1978; Arlinger et al., 1977; Tsumura et al., 1973) rate of frequency change was incidental to variation of duration of signals with fixed endpoint frequencies. The thresholds for a steady-state tone for each of the durations studied was also obtained. All glides were centered at 2 kHz. A broadband noise presented at 60 dB SPL was used as a masker. Using the method of adjustment, subjects with normal hearing adjusted the level of the signal until it was "just masked."

A statistical analysis of the data revealed significant main effects of signal duration and rate of frequency change. For both rising and falling glides, the thresholds for detection decreased as the signal duration increased. Thresholds for falling glides were higher than those for rising glides except at the lowest rate of frequency change, 24 Hz/ms. Except for the rising 5 ms tone glides, both rising and falling thresholds increased as the rate of frequency (and, hence, as extent of the sweep) increased. As the rate of frequency change increased, the difference between thresholds for rising and falling glides increased. The threshold for the 5-ms rising tone glide decreased slightly as the rate of change increased. Ndbelek (1978) also observed that detection thresholds for falling glides were higher and that detection thresholds for transitions in a noise masker increased as the extent of the sweep increased.

Cullen and Collins suggested that a plot of detection threshold as a function of duration should show parallel curves for the different frequency sweeps if the difference in thresholds for rising and falling glides were dependent on the extent of the frequency sweep. Plots of falling glides were parallel, but plots for rising glides were not. The authors concluded that it
was the rate of frequency change and not extent of the frequency sweep per se that accounted for the differences in thresholds for rising and falling glides.

Another hypothesis proposed by the authors to explain the differences in threshold addressed the relationship between the direction of the glide and the tonotopic arrangement of hair cells along the basilar membrane. When the direction of the glide was opposite to the frequency tonotopicity of the basilar membrane i.e., low to high for rising glides, displacement along the membrane would be more evenly distributed. This would produce synchronized nerve firing, resulting in lower thresholds.

Tyler, Wood, and Fernandes (1983) used two different tasks employing dynamic signals to measure frequency discrimination in normal-hearing and hearing-impaired listeners. In the first task, subjects were required to discriminate between a reference signal with a fixed frequency of either 500 or 4 kHz and a target signal whose frequency changed throughout its presentation. The dynamic signals were always swept from an initial frequency that was <500 or <4 kHz to the final frequency of 500 or 4 kHz. With this arrangement, listeners were not able to use the final frequency as a cue to discrimination. In the second task both the reference and the target signals were dynamic. The reference had either an initial frequency of 500 Hz and a final frequency of 750 Hz or an initial frequency of 4 kHz and a final one of 5 kHz. The target signal in this condition always started below the initial frequency of the reference but ended at the same frequency. DLs were also determined for conditions in which both the reference and the target were fixed-frequency signals. The signals were presented at 94 dB SPL to the
hearing-impaired listeners and at 94 dB SPL and 54 dB SPL to listeners with normal hearing. An adaptive 2AFC procedure was used. Frequency DLs were used as a measure of sensitivity.

For the fixed 500-Hz reference, the average frequency DL for normal-hearing listeners was 45.2 Hz. The average frequency DL was 59.6 Hz when the reference was dynamic. These DLs are three to four times higher than the mean DL of 14.8 Hz when both the reference and target were fixed-frequency signals. Average frequency DLs for the hearing-impaired listeners were 84.3 Hz with the fixed reference, 135.9 Hz with the dynamic reference. Like the DLs of normal-hearing listeners, these values are three to four times the average DLs of 29.6 Hz obtained when both the reference and target signals were fixed in frequency. For normal-hearing listeners, with the 4 kHz signal, the average frequency DL for conditions with fixed frequency reference and target was 175.5 Hz, the DL with fixed reference was 331.4 Hz, and with dynamic reference 734.4 Hz. Comparable average frequency DLs for the hearing-impaired listeners were 374.8 Hz, 912.3 Hz, and 964.2 Hz respectively. It should be noted that both listeners with normal hearing and hearing-impaired listeners had a high degree of variability. The authors noted that these DLs were unusually high when compared to data reported in the literature. They claimed that the poor performance of the subjects in this study reflected the fact that they were inexperienced listeners. They considered the high degree of variability in the normal-hearing controls as being not unusual.

A consideration of performance on all the tasks showed that 25% of the hearing-impaired subjects had DLs that were comparable to the DLs of the
normal-hearing listeners. The authors noted that the degree of hearing loss was not a predictor of performance on these tasks. This was the same conclusion that Arlinger et al. (1977) had arrived at in their study of frequency discrimination. The authors concluded from the data that discrimination of dynamic stimuli was poorer than discrimination of fixed stimuli.

Collins (1984) compared the ability of listeners with normal hearing and listeners with sensorineural hearing loss to discriminate frequency-modulated signals. One of the objectives of this study was to determine if the poorer performance of hearing-impaired subjects on tasks employing signals containing a dynamic portion preceded and followed by a steady-state portion (such as those used by Arlinger et al., 1977) was due to an inability to follow the rapid frequency changes or to masking effects of the steady-state segments.

The signals in this study were designed to be analogs of second formant transitions and adjacent steady-state portions such as are found in consonant-vowel or vowel-consonant syllables in speech. The stimuli consisted of a frequency-varying segment, 30 ms in duration, preceded or followed by a 200-ms fixed-frequency segment. Both rising and falling glides were used. One set of stimuli (Set I) was designated as fixed-segment-following signals. Reference fixed-segment-following signals either rose from a frequency less than 2200 Hz and ended at 2200 Hz or fell from a frequency higher than 2200 Hz and ended at 2200 Hz. A second set of reference stimuli (Set II) was called fixed-segment-preceding signals. These stimuli either rose from a frequency of 2200 Hz and ended at a higher
frequency or fell from a frequency of 2200 Hz and ended at a lower frequency. One reference signal in each set had two contiguous fixed-frequency 2200-Hz segments; in other words, it was the equivalent of a steady-state tone. Two sets of comparison stimuli were designed so that the rate of frequency change of the glide portion was either greater than or less than that of the reference stimuli.

Two tasks employing these stimuli were used in this study. One task required the discrimination between a signal entirely fixed in frequency and a signal containing a glide and a fixed portion. The other task involved discrimination of signals containing glides with different rates of frequency change (or, from another point of view, with different extents of frequency sweeps). An adaptive ABX procedure was used. The signals were presented at a level that was half way between the threshold and loudness discomfort levels for the reference signals. For normal-hearing listeners this level was 75 dB SPL.

Like the subjects in the Tyler et al. (1983) study, both normal-hearing and hearing-impaired listeners found it easier to distinguish between a reference signal with a fixed frequency and a comparison signal containing a glide than between a reference and comparison both containing glides. For listeners with normal hearing, frequency DLs for fixed-reference signals and comparison signals with glides were generally <50 Hz. DLs for discrimination between a reference and comparison signal both containing glides ranged from 100 Hz to >250 Hz. DLs were larger for glide-preceding than glide-following conditions.
Generally, DLs for hearing-impaired listeners were two to four times higher than those of the normal-hearing controls. However, in some glide-following conditions their performance was as good as that of normal-hearing listeners. The results for temporal position were less clear cut than they were for the normal-hearing subjects. Only one hearing-impaired subject showed consistently higher DLs for glide-preceding than for glide-following conditions. Generally, the other subjects had DLs that deviated most from those of listeners with normal hearing when the glide-following signal contained a falling transition whose endpoint frequency fell into the region of greatest hearing impairment.

Collins noted that temporal masking may have accounted for higher DLs in the glide-preceding condition; therefore, a second experiment was designed to determine the effect of temporal masking on DLs for glides. In this experiment, the same stimuli as described above were used, but the duration of the fixed-frequency portion of the stimulus was varied from 100 to 0 ms. The data showed that discrimination thresholds in both normal-hearing and hearing-impaired listeners were lowest when the duration of the steady-state portion was 0 ms. As a group, the listeners with normal hearing showed a great deal of between-subject and task-related variability making it difficult to reach conclusions about the effect of steady-state segment on the threshold of discrimination. Collins concluded that the amount of training that the subjects received may have had a bearing on their performance in this task. In general, regardless of duration of the steady-state portion, DLs tended to be higher for glide-preceding conditions; hence, backward masking had a greater effect than forward masking.
Data for discrimination thresholds in this second experiment were obtained from the impaired left and right ears of a single subject. Generally, as was observed in the listeners with normal hearing, DLs were lowest when the duration of the fixed segments was 0 ms. Both forward and backward masking effects were observed in each ear depending on the endpoint frequency of the glide. Poorer performance tended to occur when the endpoint frequencies coincided with the frequency region in which there was the greatest hearing loss.

One explanation that Collins proposed for the higher DLs with glide-preceding signals was a "recency effect." If pitch were used as a cue, the signals in the glide-preceding condition may have been confused more easily because the glide portions all rose or fell to a final 2200 Hz. As the duration of the fixed portion increased, this effect would be magnified. Indeed, some of the data for the masking experiment did show that as the duration of the glide-preceding fixed-frequency signals increased, the difference in DLs between glide-preceding and glide-following conditions increased. With glide-following signals, the most recent cue pitch would be the endpoint of the glide. Subjects may have used differences in pitch related to the endpoint as cues.

Grant (1987) pointed out that the higher-than-normal frequency DLs for FM signals found in hearing-impaired listeners may have reflected the ability of listeners with sharply sloping losses to use loudness cues for frequency discrimination. He proposed, therefore, adding amplitude modulation to FM signals to minimize the use of differences in loudness as cues. He measured frequency DLs in normal-hearing and profoundly
hearing-impaired subjects using either FM signals or fixed-frequency pulsed sinusoids under conditions of constant amplitude (CA), sinusoidal amplitude modulation (SAM), and random amplitude modulation (RAM). An adaptive three-interval, two-alternative, forced-choice procedure was used to determine DLs for frequency modulation. The stimuli were one second in duration and were presented at levels of 30 dB SL for listeners with normal hearing and at a comfortable listening level for the hearing-impaired listeners.

The FM target tone was triangularly modulated at a rate of 3 Hz so that each FM cycle began and ended at the fixed frequency of a reference tone. Amplitude modulations of both the reference and target signals were fixed and equal in the CA condition. A 3-Hz sinusoid was used to modulate the amplitudes of the tones in the SAM condition. Lowpass-filtered white noise with a cutoff frequency of 3 Hz was used to modulate the signals in the RAM condition. The subjects were intensively trained to recognize the difference between a change in pitch and a change in amplitude. DLs for frequency modulation (DLFMs) were expressed in percent modulation corresponding to 79% correct responses.

For listeners with normal hearing under conditions of CA, DLFMs decreased as the frequency of the signal increased. The percent of modulation required for detection of FM ranged from 2.2% at 100 Hz to 0.15% at 1000 Hz with a mean at 0.4%. DLFMs under conditions of SAM were approximately twice as high as those under CA. Under conditions of
RAM, the DLFMs were three times as high as those under CA. For frequencies above 300 Hz, less than 1% modulation was required for FM detection under all conditions.

In hearing-impaired listeners, DLFMs were higher than those of normals under all amplitude conditions and at all frequencies. The DLFMs under CA conditions were 3.8%, 7.9% under conditions of SAM, and 17.8% under conditions of RAM. The author noted that DLFMs for normal-hearing subjects decreased by approximately one order of magnitude as the frequency of the signal increased from 100 to 1000 Hz; the DLFMs of the hearing-impaired subjects decreased by two to three orders of magnitude across the same range of frequencies under all conditions. The author concluded that the very high DLFMs under conditions of RAM may indicate that hearing-impaired listeners do depend on amplitude cues to make frequency discriminations. The RAM conditions would have made loudness cues unusable.

Four specific questions were addressed by Elliott, Hammer, Scholl, Carrell, and Wasowicz (1989) in their study of discrimination of formant transitions by normal-hearing listeners: what were just noticeable differences (JNDs) for synthetic frequency transitions that resembled second formants in speech, what was the effect of practice on the ability to discriminate frequency transitions, what was the effect of an appended steady-state segment on a preceding glide, how did discrimination of frequency transitions compare to discrimination of steady-state sounds which were similar in complexity.
The frequency transitions used in this study were very different from those used in the studies discussed previously in that they were derived from synthetic speech, and they had an 80-Hz bandwidth. The durations of the transitions were 30, 60, and 120 ms. In some signals, only the transition portion was present; in others a steady-state portion was appended at the end of the transition. The total duration of appended signals was 300 ms. Both rising and falling transitions were employed. The onset frequencies of the rising transitions varied linearly in 17-Hz steps from 942 to 1146 Hz. The onset frequencies of the falling transitions varied linearly in 36-Hz steps from 1772 to 1340 Hz. The endpoint frequency for all transitions was 1240 Hz. The frequency of the steady-state segment appended to some of the transitions was also 1240 Hz. The center frequency of steady-state single formants were equal to the midpoint frequency of the rising and falling formant transitions. These steady-state formants were 300 ms in duration.

An adaptive two-interval same/different procedure was employed. Presentation of stimuli was at approximately 40 dB SL. The reference stimulus was always the one with the shortest frequency sweep (1146 Hz onset for rising transitions; 1340 Hz onset for falling transitions).

For both transition only and transition plus steady-state segment, JNDs were smallest for 120-ms transitions, next largest for 60-ms transitions, and largest for 30-ms transitions. As the duration of the transition decreased, the JNDs generally increased by a factor of three to four. There was no significant difference in JNDs as a function of transition direction and no consistent differences between JNDs for transition only and transition plus steady-state. The results of the study of practice effects showed a great deal of within-
subject variability. There was an improvement with practice for some conditions (i.e., shortest duration stimuli) but not for others. Collins (1984) also observed a high degree of within-subject variability which she attributed to a lack of training on a task requiring the discrimination of transitions with appended steady-state segments of varying duration. The lack of asymmetry in response to rising and falling transitions that Elliott et al. found was in agreement with the findings of Arlinger et al. (1976), Collins (1984), and Nábelek and Hirsh (1969).

Synthetic speechlike transitions consisting of second-formant transition resonances rising or falling to steady-state vowel-like resonances were also used by Porter, Cullen, Collins, and Jackson (1991) in a study of the perception of dynamic stimuli by listeners with normal hearing. The variables used in this study were transition duration, extent of frequency sweep, rate of change, and direction. The steady-state portion of the stimulus was centered at 1800 Hz. The total duration of the signal was 320 ms, and the formant bandwidth was 125 Hz. An adaptive two-cue, two-alternative, forced-choice (2Q, 2AFC) procedure was used. The JNDs in onset frequency of the transition were used as a measure of sensitivity. Comparison stimuli in successive runs in a block alternated between onset frequency less than the reference stimulus and onset frequency higher than the reference. The level of presentation was 70 dB SPL.

In the first experiment that was done, the reference stimuli either rose from 1500 Hz to 1800 Hz or fell from 2100 Hz to 1800 Hz. The frequency sweep, therefore, was always 300 Hz. Rate of frequency change was varied by using transitions with durations of either 30, 60, or 120 ms. Onset
frequencies of the comparison stimuli were varied above or below the onset frequency of the reference. This had the effect of creating rates of change that were either slower or faster than the rates of change of the reference.

JNDs for onset frequency varied as a function of the duration of the transition. The longer the duration the easier the task. Some of the subjects in the study could not consistently discriminate endpoint frequencies of the 30-ms duration transitions. Nábelek and Hirsh (1969), Nábelek (1978), and Elliott et al. (1989) also reported poor discrimination performance at shorter duration frequency transitions. The JNDs for rising transitions were about 75% higher than those for falling transitions. The authors concluded that lower JNDs at longer durations could be interpreted as meaning that listeners were more sensitive at lower rates of frequency change or that listeners were more sensitive at longer durations where they may have been better able to resolve spectral differences. If rate of change were the factor, one would expect that, for a fixed duration, the JNDs would be smaller for changes in onset frequency of comparison stimuli which were below the onset frequency of the reference (i.e., smaller frequency sweeps, consequently, slower rates of change) than the JNDs for changes in onset frequency of comparison stimuli above the onset frequency of the reference (i.e., longer frequency sweeps, therefore, faster rate of change). In fact, just the opposite results were obtained. JNDs for increasingly longer sweeps were smaller than JNDs for increasingly shorter sweeps for both rising and falling transitions and at all durations.

In order to be certain that discrimination performance was not, at least partially, a reflection of rate of change, the authors did a second experiment
in which duration was varied, but rate of frequency change was held constant at 5 Hz/ms. Durations of the formant transitions in this second experiment were 30, 45, and 60 ms; hence, the extent of the frequency transitions were, \( \pm 150 \) Hz for 30 ms, \( \pm 225 \) Hz for 45 ms, and \( \pm 300 \) Hz for 60 ms. Again, the transitions rose or fell from a frequency of 1800 Hz. If rate of frequency change in the reference was the important factor in discrimination, the JNDs in this experiment should be comparable across different durations. The results of the experiment showed that the JNDs were lowest for 30 ms, and highest for 45 ms. However, the JNDs for the 30-ms duration signal for this experiment were one fifth to one half as large as those of the 30-ms transitions in the first experiment. The difference between the signals was the rate of frequency change. In the second experiment the rate of change was 5 Hz/ms; in the first experiment the rate was 10 Hz/ms. The authors noted that they made the rate slower in the second experiment because many subjects had difficulty making consistent responses for durations of 30 ms. The authors concluded that for signals longer than 30 ms, transition duration affects discrimination rather than rate of frequency change, but at short durations (e.g., 30 ms) rate of change is a factor.

With these more slowly changing transitions, asymmetry in the perception of rising and falling transitions was also observed. Falling transitions had lower JNDs. Increments in rates of frequency change (increases in rates of change) were also more discriminable than decrements (decreases in rates of change). These asymmetries were also observed in the first experiment. Based on models of temporal envelope processing which propose that excitation envelopes produced by neurons sensitive to energy
within specific high frequency bands have greater temporal detail than the envelopes produced by neurons sensitive to energy in lower frequency bands, Porter et al. suggested that the differences in JNDs for rising and falling transitions may reflect differences in resolution of temporal envelope cues associated with high and low frequency regions. Discrimination should be easier among transitions with high frequency region onsets than among transitions with low frequency region onsets.

In explaining the differences in thresholds that were observed with 30-ms transitions that had different rates of frequency change, the authors suggested that with these signals the lower limits of temporal resolution are reached and that such short duration transitions are processed as a single nonvarying wideband signal. Discrimination of these signals is dependent upon the pattern of excitation on the basilar membrane. Signals with wide frequency sweeps would produce overlapping excitation patterns that would be more difficult to discriminate than signals with narrow sweeps which would produce more discrete patterns. One would predict, therefore, that the 10 Hz/ms 30-ms transitions would have higher discrimination thresholds than the 5 Hz/ms 30-ms transitions. That was exactly the observation that Porter et al. made in this study.

A very different type of FM signal was used by Feth, Neill, and Krishnamurthy (1989) to measure temporal acuity in listeners with normal hearing. These subjects were asked to discriminate between two FM signals with the same initial and final frequencies but with different trajectories between the initial and final frequencies. One signal, the glide, was linearly modulated in frequency over a given frequency range; the other signal, a step
signal, traversed the same frequency span in a series of steps. As the number of steps increased, both the duration of each step and the frequency difference between successive steps decreased. In effect, as the number of steps increased, the signal became temporally smoothed and perceptually more like the glide. This signal avoids the confounding effects of duration, extent of sweep, and rate of change that were a problem in other studies of perception of frequency transitions. On any given trial, both the reference and comparison signal change in the same way with respect to these parameters.

A nonadaptive 2Q, 2AFC procedure was used. Duration of the frequency transitions were 25, 50, and 100 ms. The frequency extent of the sweeps were either 100, 200, or 400 Hz and were centered on octave frequencies from 250-4 kHz. The level of presentation was 50 dB SL.

Minimum step size (i.e., the temporal resolution threshold) that would just permit discrimination of a glide from a step signal at the 75% correct discrimination point was found to be 7-10 ms for all signals except for those centered at 4 kHz. The temporal resolution threshold for these high-frequency sweeps was 15-20 ms.

In a extension of this study, Madden (1990) and Madden and Feth (1992) used the step and glide signals to compare temporal resolution in listeners with sensorineural hearing losses to that in listeners with normal hearing. The signals in these studies had a fixed duration of 50 ms, a fixed frequency extent of 200 Hz, and were centered at octave frequencies from 500 Hz to 4 kHz. Step number in the step signal varied from two to nine. The level of presentation was 35 dB SL for hearing-impaired listeners. Normal-hearing
listeners were tested under conditions of quiet at sensation levels equal to those of the hearing-impaired listeners and under conditions of masking at equal SL and SPL. A broadband noise was used as the masker. A nonadaptive 2Q, 2AFC procedure was used for data collection.

The mean temporal resolution thresholds for the hearing-impaired subjects were significantly higher than those of the normal-hearing subjects at all frequencies. Thresholds for all subjects increased as the CF of the frequency transitions increased. This frequency effect was only significant in the hearing-impaired subjects; a correlational analysis showed an r of .68 between the temporal resolution threshold and the hearing threshold. Under quiet testing conditions normal-hearing listeners had a temporal resolution threshold of 7-9 ms for CFs of 500-2 kHz. At a CF of 4 kHz, neither the normal-hearing nor the hearing-impaired subjects could achieve the 75% correct performance criterion established. Under masking conditions the thresholds for normal-hearing listeners were 10-12 ms. These values are comparable to the thresholds of the hearing-impaired listeners which ranged from 13-18 ms.

Madden and Feth suggested that temporal resolution at the lower frequencies was a function of the ability of neurons to phase lock at frequencies <4-5 kHz. At higher frequencies, they hypothesized the existence of a second mechanism with a larger time constant than that of the phase-locking mechanism, hence, the higher thresholds observed at 4 kHz.

It is difficult to summarize the results of these psychoacoustic measures of temporal resolution because different stimuli, different levels of presentation, and different testing procedures were used, and different tasks
were required of the subjects. In some studies the subjects were asked to
detect a change in frequency, in others a change in direction, and in still
others, a change in rate. Also adding to the difficulty, is the confounding
effect of duration, frequency sweep, and rate of change. Generally, one can
conclude from these studies that thresholds for detection of frequency
modulation decrease as the duration of the signal increases. Many of the
studies showed that there is an asymmetry in the detection of rising and
falling signals. However, this observation is muddied by the fact that, in a
few studies, there was no difference in thresholds for rising and falling
transitions, and in studies that did show an asymmetry, there was no
agreement on which direction resulted in the lower thresholds.

2.4 Perception of dynamic signals by cochlear-implant users

There have been very few studies of perception of dynamic signals by
cochlear-implant users. As noted previously, Tong et al. (1982) tested
cochlear-implant users on their ability to discriminate a train of biphasic
electrical pulses delivered to only one electrode from a train of pulses swept
across a number of electrodes. The pulse trains were 200 ms in duration; the
first 100 ms of the signal was swept across electrodes; the final 100 ms was a
steady-state train of pulses at the terminal electrode. The rate of delivery was
100 pps. The current levels used produced a sensation described as
"medium" loud. Three sets of stimuli were used. In set A the sweeps of
both reference and comparison stimuli always ended at electrode #1. (In
Tong et al.'s numbering system, low numbers designate electrodes to which
low frequency information is delivered.) The initial electrodes for set A reference signals were either #1, 4, or 7. The initial electrodes of the comparison stimuli in set A were #0 through #8. Set B was like set A except that the final electrode position was at electrode #4; the final electrode position in set C was at electrode #7 (an electrode located in the basal end of the cochlea).

Two subjects implanted with a ten-channel multi-electrode device participated in this study. Delivery of the pulse train was directly to the receiver/stimulator of the implant, probably via a percutaneous connector; therefore, no speech processor was used in this study. The subjects were asked to discriminate between reference and comparison stimuli on the basis of whether they sounded the same or different. As noted in the introduction to this dissertation, the mean percent judged different was 52% when stimuli differed by one initial electrode, 80% when the difference was two initial electrodes, and 92% when the difference was 3 initial electrodes. Another way of expressing this relationship is in terms of total number of electrodes swept by the electrical pulse train. As the difference between total number of electrodes swept in the reference stimulus and total number of electrodes swept in the comparison stimulus increased, the percent judged different increased.

The effect of duration of the sweep portion of the signal on discrimination performance was also examined. Durations of 25, 50, and 100 ms were used; the pulse rate was fixed at 150 pps. For this study only four initial electrode positions were used, #1, 2, 3, and 4. The sweep portion was followed by a 100-ms steady-state train of electrical pulses delivered only to electrode #1. The
stimulus with the same initial and final electrode numbers (#1,1) was the reference. The percent judged different was nearly 100% on every trial regardless of the duration of the sweep.

It is apparent from this study, therefore, that users of the cochlear implant are very sensitive to the difference between single-electrode stimulation and stimulation that sweeps across electrodes and that their sensitivity increases as the difference between number of electrodes swept in the reference and in the comparison stimuli increases.

Kirk, Tye-Murray, and Hurtig (1992) used acoustic rather than electrical stimuli to examine the perception of dynamic cues in cochlear-implant users of either the Nucleus or the Ineraid device. Naturally produced /bVb/ or /wVb/ speech syllables were modified so that they contained dynamic only or both dynamic and static cues. The consonants /b/ and /w/ were chosen so that long and short duration transitions could be compared. Initial transitions in /bVb/ syllables ranged from 33 to 62 ms. In the /wVb/ syllables, initial transitions ranged from 83 to 137 ms. No prevoicing or burst release cues were present in the initial transitions of /bVb/ syllables. Performance was assessed by the ability to consistently identify vowels in the modified syllables. The performance of the implant users was compared to that of listeners with normal hearing. The stimuli were presented through a loudspeaker at a level of 70 dB SPL. Implant users were allowed to adjust their processors to a setting at which the level of the stimuli were comfortable.

The results showed that 80% of the implant users scored above chance in the correct identification of vowels in syllables containing both static and
transitional cues. When the syllables contained only transitional cues, 55% of the implant users performed above chance when the rapid stop /b/ was the initial consonant, but only 20% performed above chance when the long duration glide /w/ was in the initial position. Users of the Ineraid had about the same level of performance for both abutted transition syllables /b-b/ and /w-b/. However, the users of the Nucleus device performed much more poorly in the vowel identification tasks when the abutted transitions were /w-b/. With steady-state only cues, 65% of the cochlear-implant users performed above chance. Listeners with normal hearing performed at levels above 92% when transition and steady-state cues were both present. Like implant users their performance dropped somewhat for certain conditions of steady-state only cues.

The authors offered three possible explanations for the rather dramatic differences in performance with abutted /b-b/ and /w-b/signals (transitional cues only). They noted that Lehiste and Peterson (1961) reported that increasing the duration of the transition in a glide could result in the perception of a diphthong. However, according to the authors none of the listeners reported that perception. A second explanation related to coarticulation. Vowels in the context of /w/ are produced with lip rounding; in the context of /b/ they are not. Rounded vowels have greater formant bandwidths, and the F2 and F3 formants are closer than in unrounded vowels. The effect, therefore, is the production of a single broad spectral peak (Stevens, 1989). Because the Nucleus processor uses a feature extraction strategy to encode speech, it may have difficulty accurately encoding spectral information from this smeared frequency band. It should be noted that
vowel identification did not differ significantly between the two implanted groups. The authors concluded the most probable explanation of the observed performance was related to rate of formant change. The differences in onset of transition frequency in the /w/ and /b/ syllables are large enough, that despite the long duration of the /w/ transition, the rate of change is greater in the /w/ transition. Kirk et al. (1992) suggested that users of the Nucleus processor may have difficulty in resolving dynamic information because in the MSP and WSP (an early version of the MSP speech processor) voicing frequency determines the rate of delivery of electrical pulses. In the presence of a rapidly changing signal, too few pulses may be distributed across channels.

2.5 Encoding of frequency-modulation

in auditory-nerve fibers: acoustic stimulation

Because perception of sound by cochlear-implant users is completely independent of the mechanics and biochemical events of the auditory system from the outer ear to the level of the spiral ganglia, perception begins at the level of the neurons of the VIIIth nerve. To be able to explain the observed perceptual abilities of cochlear-implant subjects as measured by various psychoacoustic tasks, it is necessary to have some knowledge of the responses of the neurons of the VIIIth-nerve to both acoustic and electrical stimuli. With this knowledge, investigators can determine which aspects of speech and sound perception by cochlear-implant users reflect differences in nerve fiber response to acoustic and electrical stimuli, and speech processor strategies can be developed that account for these differences.
Studies of the responses of auditory-nerve fibers to dynamic acoustic stimuli have shown that the rapid changes in frequencies that are characteristic of formant transitions (as noted earlier, the term "frequency-modulated signal" is applicable to formant transitions) can be encoded both in the spatial distribution of the fibers responding and in the temporal pattern of discharge in response to the stimulus (Sinex and Geisler, 1981; Sinex and Geisler, 1983; Miller and Sachs, 1983; Delgutte and Kiang, 1984; Shamma, 1985; Carney and Geisler, 1986; Deng and Geisler, 1987). It must be kept in mind that these studies have all been done with anesthetized cats and, therefore, they can serve only as models of the human auditory system.

The earliest study of the response of auditory-nerve fibers to frequency-modulated signals was done by Sinex and Geisler (1981). Until this study, knowledge of the response of the auditory system to frequency-modulated signals was limited to the responses of fibers in the nuclei of the central auditory system. Sinex and Geisler used as stimuli sweep tones which were shaped by a trapezoidal-modulating voltage. These sweep tones were characterized by both an ascending portion, changing from low to high frequency, and a descending portion, changing from high to low frequency. Sweep rate, sound pressure level (SPL), and the center frequency of the sweep were varied.

From plots of firing rate as a function of instantaneous frequency of the sweep tone, the authors obtained from single fibers two estimates (termed the frequency-modulated-response area, the FMRA) of the nerve-fiber response to the frequency modulation, one for each direction of the sweep. In general, the shape of FMRAs was similar to the response area (RA) of the
fiber to a pure tone. Fibers responded during a sweep only when the instantaneous frequency was within the frequency range over which a pure tone would elicit a response from the fiber. As occurs with pure-tone response, increasing the SPL broadened the response range. Sinex and Geisler concluded that the fibers seemed to be responsive to the instantaneous frequency of the stimulus and not to the long-term spectra of the sweep tone.

Generally, the peak response was greater for the high-to-low frequency sweep than for low to high. The authors suggested this difference was a result of adaptation. The center frequency of the sweep was chosen to be near the characteristic frequency (CF) of the fiber; therefore, when the duration of the sweep was from low to high frequency, the instantaneous frequencies that were nearest the fiber’s CF occurred after the fiber had been responding to the less effective lower frequencies. By the time the more effective frequencies had been reached, the fiber had adapted. The authors noted that “the response to a given narrow frequency range within a sweep tone is greater when that range occurs at the beginning of driven activity” (Sinex and Geisler, 1981, p. 139).

Sweep rate had little effect on the FMRA. Higher sweep rates produced higher firing rates, but the shape of the FMRA did not change. There was, however, an observed interaction between SPL and the direction of the sweep. As the SPL was increased, the peak response area of the fiber tended to shift toward the earliest occurring frequency for each duration of the sweep. That is, the peak response of the FMRA for the ascending sweep
shifted to slightly lower frequencies and that for the descending sweep to slightly higher frequencies.

Sinex and Geisler observed phase locking in the auditory-nerve fibers in response to the stimuli. The authors looked at interspike intervals (ISI) as a function of time of occurrence during a complete cycle of the sweep tone. The ISIs followed the sweep of the waveform with ISI times decreasing as the sweep ascended (frequency increasing) and then increasing as the sweep descended (frequency decreasing). A plot of the reciprocals of the ISI times as a function of time during the cycle showed, in general, a good correspondence between the trajectory of the data points and the trajectory of the sweep.

Sinex and Geisler's study showed that a fiber's response to frequency-modulated tones was like that characteristic of the fiber's response to pure tones in terms of frequency selectivity and adaptation. The authors concluded that the timing of discharges might encode the frequency trajectory of a frequency-modulated tone more efficiently than would simple firing rate.

In a second study of the responses of auditory-nerve fibers to dynamic stimuli, Sinex and Geisler (1983) used synthetically-generated stop consonant-vowel syllables as the stimuli. The formant transitions in the syllables corresponded to frequency-modulated signals. In this study, a comparison was made between formant transitions of 25-, 50-, and 75-ms durations. The data was organized into PST histograms (peristimulus time histograms in which the time of occurrence of spike discharges was recorded relative to the onset of the stimulus). Fast Fourier Transforms were done on
another set of PST histograms in order to provide estimates of the response to particular frequency components of the stimulus.

Profiles of average discharge rate (spikes/sec) as a function of time relative to stimulus onset were generated either for single fibers or for several fibers whose CFs were in a limited range. These profiles showed that, in general, increasing SPL had little effect on discharge rate during the vowel portion of the stimulus but that during the onset of the formant transition portion of the stimulus increasing SPL resulted in an increased discharge rate especially in high CF fibers. The observed relationship between SPL and discharge rate during the transition for high CF fibers probably reflects the fact that components of the higher formants tended to have lower amplitudes. Hence, when the SPL was increased, the amplitude of the formants increased and, consequently, the discharge rate increased.

The authors noted differences in the responses of high-spontaneous rate fibers (>18 spikes/sec) and low-spontaneous rate fibers (<18 spikes/sec). High-spontaneous rate fibers were more active at signal onset and continued to be more active throughout the syllable. Low-spontaneous rate fibers had highest discharge rates during the vowel or at the midpoint of the transition. The paucity of the response of low-spontaneous rate fibers at the onset of the syllable may be a reflection of the higher thresholds in these fibers compared to thresholds in the high-spontaneous rate fibers.

Sinex and Geisler compared average discharge rates at the onset of /ba/, /da/ and /ga/ and found that average discharge rates could be used to differentiate among these syllables. For all of the syllables tested, a selected low-CF fiber responded in about the same fashion. Because the first formant
was the same in all of the syllables tested, this result was not unexpected. A selected high-CF fiber, however, responded differently to different syllables. For example, in a 1kHz-CF fiber responding to /ba/, the discharge-rate response was high at the onset of the transition as the second formant of /ba/ was near the CF of the fiber. In response to /da/, however, the same fiber had a much lower rate of discharge, as the second formant onset frequency was well above the fiber’s CF.

During the formant transition, fibers had different discharge rate responses to different stop-consonant vowel stimuli at a given time period relative to the onset of the CV syllable. For example, in response to /ba/ the discharge rate of a 1kHz fiber was highest at 12 msec after onset when the second formant frequency components were near the fiber’s CF and then fell as the second formant continued to rise in frequency with time. At the same time, the onset response of a fiber with a CF of 1900 Hz was low at the onset of /ba/ relative to its response at 32 msec at which time the second formant frequency components were nearer that fiber’s CF.

A different pattern emerged when the stimulus was /da/ which has falling second and third formants with high onset frequencies. The discharge rate of a fiber with 1kHz CF increased late in the transition relative to its response at onset as the frequency of the transition fell towards the fiber’s response range. The response of the 1900 Hz CF fiber continued to fall relative to its onset response rate as the second formant fell.

Synchronization of the discharge rate was also found to have a role in encoding formant transitions. At low SPL the synchronized response was highest to frequency components of the stimulus which were near the CF of
the fiber. As the SPL increased, however, the highest degree of synchronized response tended to shift to frequency components of the formant transitions which were away from the CF of the fiber, and the response range narrowed. The fiber, therefore, more effectively signaled the presence of a single component of the stimulus at higher SPLs. At moderate SPLs most of the fibers tended to synchronize to one of two or three dominant components in the syllable.

Synchronization to second formant frequencies tended to be in fibers whose CFs were at or above the formant frequencies. Synchronization to first formant frequencies tended to be in fibers whose CFs were both above and below F1 frequencies because low-CF fibers have wider tuning curves. Synchronization occurred both during formant transitions and in the steady-state vowel portions of the syllables. The results also showed that fibers with CFs greater than the third formant frequencies tended to synchronize their discharge responses to the first formant even when there were F3 components whose frequencies were near the CF of the fiber. At the onset of a syllable, fibers whose CFs were in ranges for which there were no strong components in the acoustic signal tended to synchronize to frequency components near their CF. As time progressed, however, the responses of the fibers became synchronized to a component of the F1 or F2. In plots of the responses of several fibers to dominant components versus duration of the syllable for /ba/, /da/, and /ga/, one can see that the response pattern closely followed the second and third formant trajectories of the syllables. There was some scatter in the distribution of data points around the first formant trajectory at the onset of the transition because, as noted above,
fibers whose CFs are below the second and third formant frequencies tended to synchronize to components near their CFs at onset.

Miller and Sachs (1983) examined the representation of /da/ and /ba/ in the discharge responses of auditory-nerve fibers. In addition to generating PST histograms for the fibers, the authors did Fourier Transforms of the histograms in order to show the fiber responses to particular frequency components of the stimulus. In addition, they computed the average localized synchronized rate (ALSR) which was a measure of the average temporal response in a population of fibers whose CFs were within ± 0.25 octave of the stimulus frequency component of interest.

In response to /da/, a single fiber with a CF of 570 showed strong synchrony to 500 Hz, the onset frequency of the first formant, during the first 20 msec of the stimulus. During the last 20 msec of the syllable, the first formant frequency rose to 700 Hz. This same fiber with a CF of 570 Hz showed a strong synchronization to the 700 Hz frequency component of the Fourier Transform of the PST histogram at that time. A fiber with a CF of 1400 Hz showed a dominant response component that shifted from 1450 Hz in the first 20 msec of the stimulus to 1150 Hz during the last 20 msec. This shift in response reflected the falling frequency of the second formant transition in /da/. A fiber whose CF was near the third formant showed a similar response to frequency shifts in the third formant. The authors concluded from their data that temporal short-term responses of the fibers reflected the short-time spectral compositions of the stimuli. The response of fibers with CFs near the formant frequencies was dominated by those frequencies and followed the formant trajectory.
Miller and Sachs also examined the temporal representation of /ba/ and /da/ in populations of fibers. Plots which showed the distribution of synchronization within a population indicated that at the onset of the first formant there is a strong synchronization among fibers whose CFs were near the frequency of the first formant. Generally, synchronization to instantaneous formant frequency tended to spread to fibers whose CFs were above the frequency of the formant. In other words, fibers with high CFs tended to phase lock to formants that had low frequencies. As the frequency of the first formant rose with increasing duration of the stimulus, the fiber CFs at which peaks of synchronization occurred shifted upward in accord with the upward movement of the first formant trajectory.

Plots of data obtained by Miller and Sachs from populations of fibers showed fairly close correspondence between the stimuli power spectra and the ALSR plots. Peaks in the ALSR shifted in correspondence to the frequency trajectories of the first, second, and third formants. For example, in /da/ the movement of the ALSR peaks reflected the rising first formant and the falling second and third formant transitions. In addition, plots of the normalized average firing rate of unit fibers versus the CF of the units showed peaks in frequency regions that corresponded to the first, second, and third formants. When these plots were analyzed as a function of time, there was a shift in the peak of the average rates that corresponded to the trajectory of the three formants. This representation by the average discharge rate of the temporal features of the formants was not found to hold for the steady-state vowel portion of the syllables.
In a study by Delgutte and Kiang (1984) the response of auditory-nerve fibers to a complex stimulus composed of three parts was measured. The initial portion of the stimulus, called the preceding context, varied in formant frequency composition, duration, silent interval, and amplitude. The central portion of the stimulus, the test interval, always had /da/-like formants. The final portion was always the vowel /a/. The focus of the study was on the effect of the preceding context on the average discharge rate of the fibers in response to the /da/-like formant transitions.

The profiles of discharge rate versus CF during the test interval varied according to the spectrum of the preceding context. All preceding contexts produced a decrease in the characteristic discharge rate of the /da/-like test interval. Those fibers whose CFs were in frequency regions where the context had significant energy showed the largest decrease in discharge rate. The authors noted that such short-term adaptation would serve to increase contrast between successive speech segments.

Shamma (1985) re-analyzed auditory-nerve response data previously gathered by other investigators. Spatio-temporal response patterns were constructed from single unit recordings of large populations of auditory-nerve fibers responding to /da/ and /ba/ stimuli. Three-dimensional plots were created by ordering the PST histograms of each unit according to its CF. These 3-D spatio-temporal response patterns indicated that the overall firing rates of fibers decreased significantly 1-2 ms after the onset of the response to both /da/ and /ba/ stimuli. Most of the fibers showed phase locking (i.e. synchrony) to one or more harmonic components of the stimuli. Fibers with low CFs (CFs greater than 200 Hz but less than 300 Hz) phase locked to the
half-wave rectified waveform of the second harmonic. The response pattern also showed strong phase locking to the fourth and fifth harmonics (which are close in frequency to the first formant of the stimuli) and to the twelfth and thirteenth harmonics when the stimulus was /da/ and to the seventh and eighth harmonics when the stimulus was /ba/ (harmonics close in frequency to the second formants of the respective stimuli). With time, the phase locking shifted. The first and second formants in the /da/ stimulus shifted to frequencies of the sixth and tenth harmonics respectively. The phase-locked response to these formants, likewise, shifted to these frequencies. Shamma noted that units also phase locked to the combined waveform of several adjacent harmonics. Units with CFs larger than 3kHz phase locked to lower frequencies. Shamma interpreted this as a response to the temporal envelope that resulted from the interaction of frequency components that fell within the bandwidth of the fiber. In general, however, synchrony to frequencies above 2 kHz was sharply reduced.

Shamma also analyzed the profiles of the average firing rates during the first 30 msec of response to the stimulus. The envelopes of the profiles showed, to some degree, the locations of the formants. However, as the stimulus continued in time, the effects of short-term adaptation began to appear and the formants became less distinct. Shamma concluded that the average firing rate profiles were not especially good representations of the frequency components of formant transitions nor of the following vowel. His conclusion was in contrast to that of Miller and Sachs (1983) and of Sinex and Geisler (1983) who claimed the average-firing-rate profiles did provide good representation of the transition spectra.
In all of the studies discussed above synthetic stimuli were used. Carney and Geisler (1986) used natural, spoken stimuli in their study of the responses of auditory-nerve fibers to speech. Unique to their study was the inclusion of both voiced (b,d,g) and unvoiced (p,t,k) consonants in the consonant-vowel stimuli. The authors also used an unusual technique for the analysis of their data. They constructed digital spectrograms of the stimuli waveforms and compared them to digital spectrograms constructed from neural responses. These spectrograms showed the relationships between time, frequency, and intensity of the response.

The spectrograms of the stimuli waveforms showed areas of localized energy notably along the trajectory of the second formant. These areas occurred as the formant trajectory crossed and interacted with harmonics associated with voicing. These areas were seen throughout the duration of the voiced consonant syllables as voicing begins before or near the release of the stop in voiced consonants; however, in the voiceless consonant syllables, the energy in the formants before voicing was spectrally diffuse. Once voicing associated with the vowel began, however, the areas of localized energy also appeared wherever the formant trajectory crossed harmonics of the fundamental frequency (F0). Other acoustic features of the digital spectrogram that were considered to be possible correlates of perceptual cues were voice-onset time, formant transitions, and initial spectral shape.

The responses of the nerve fibers could be categorized on the basis of whether the fibers had high spontaneous rates of discharge (>18 spikes/sec) or low spontaneous rates (<18 spikes/sec). Recall that Sinex and Geisler (1983) also made this distinction in behavior of the fibers in the samples they
studied. Generally, high-spontaneous-rate (SR) fibers showed a strong synchronized response to the onset transient (that is, the burst associated with the release of the stop) regardless of the fiber CF. During aspiration and frication in unvoiced stops, the high-SR fibers tended to synchronize to spectral peaks that were near their CF. However, some of the stimuli had energy that was strong enough to entrain fibers whose CFs were not near the frequencies of the spectral peaks of the stimuli. Spectrograms of single fiber responses showed that the fibers' synchronized discharges followed the dynamic frequency trajectory of first, second, and third formants and the response was enhanced when the frequency of the trajectory interacted with harmonics of the F0 as the path of the trajectory crossed the harmonics. Synchrony of high-SR fibers was evident throughout the duration of the stimulus. For a population of high-SR fibers the pattern of response was nearly the same for voiced and unvoiced stimuli. Within a population of high-SR fibers, however, low-CF and high-CF fibers responded differently. Low-CF, high-SR fibers did not show strongly synchronized responses during the voiceless part of a voiceless-consonant syllable. They did, however, synchronize to the early onset of voicing in the voiced stimuli and to the late onset of voicing in voiceless stimuli. The authors suggested that the strong response to the late onset of voicing may have been due to the fact that the fibers did not respond during the voiceless part of the syllable, therefore, they were not adapted. High-CF, high-SR fibers synchronized to frequency components in aspiration and frication noise and to formants; in other words, they synchronized to components throughout the syllable duration. They did not synchronize to the onset of voicing. Again, the authors
suggested adaptation as a possible explanation for this observed effect. Fibers with CFs above 3300 Hz only synchronized to the F0 frequency.

Low/medium-SR fibers often did not exhibit synchronization throughout the duration of the stimulus. There appeared to be three general patterns of response for low/medium-SR fibers: low-CF fibers that synchronized only to low-frequency energy in voiced segments of the stimuli; fibers with higher CFs which synchronized to frequencies present before but not during voicing; and fibers that synchronized their discharge rate to frequency components of the stimulus throughout the duration of the stimulus.

One of the problems with this study was that it was difficult to compare composite spectrograms because the fibers used to construct the spectrograms had different CFs from spectrogram to spectrogram. The authors were aware of this problem. Basically, this study showed that, in most cases, a fiber synchronized its discharge to a band of frequencies concentrated about the fiber's CF. The fiber's threshold, dynamic range, and response to naturally occurring changes in SPL during the duration of the stimulus influenced synchrony. Carney and Geisler noted that in a stop consonant + vowel there can be a 20 dB difference in SPL before and during voicing. They noted that other investigators have shown that low/medium-SR fibers respond to weak sounds with low discharge rates.

The authors concluded that the dynamic formant trajectories were well represented in the synchronized discharge rate responses of the auditory-nerve fibers, particularly in the responses of high SR fibers. They also noted that their findings were in good agreement with those of previous investigations in which synthetic stimuli were used. The functional
significance of the distinction between the responses of the high-SR and low/medium-SR fibers is unknown but the authors point out that these fibers do have anatomical differences.

In another study of the temporal features of the discharge pattern in auditory-nerve fiber response to speech stimuli, Deng and Geisler (1987) used natural nasal-consonant vowel syllables as stimuli. The formant transitions of these syllables were similar to those for voiced stop-consonant vowel syllables. As previous investigators had done, Deng and Geisler examined the data in terms of the responses of high-SR fibers and of low/medium-SR fibers. They also examined the response patterns of single fibers and of arrays of fibers that included both high-SR and low/medium-SR fibers.

Individual high-SR fibers were found to respond throughout the duration of the syllable. These fibers synchronized their discharges throughout the murmur, the formant transition, and the vocalic portions of the stimuli. An interesting observation was that synchronization tended to be as strong through the murmur as through the vowel even though the murmur had much less energy. The formant trajectories were “faithfully encoded”. Similar to the observation by Miller and Sachs (1983) regarding fiber entrainment to first formants, fibers whose CFs were higher than second formant frequencies were entrained by these lower F2 frequencies.

Synchronization in low/medium-SR fibers was strong in response to the F0 of the stimuli and to its lower harmonics. This strong synchronized response to low harmonics of the F0 was rarely observed in high-SR fibers. High-SR fibers whose CFs and threshold tuning curves were similar to those of low/medium SR fibers showed synchrony to the second formant of the
same syllable. The authors made a systematic comparison of pairs of high-SR and low-SR fibers having similar CFs and threshold tuning curves. Their general conclusions were that both low-CF high-SR and low-CF low/medium-SR fibers synchronized to low-frequency formants whose frequencies were near these fibers' CFs. Low/medium-SR fibers responded less strongly during the low acoustic energy murmur portion of the syllable than did high-SR fibers. As mentioned previously, the differences in threshold for these two groups of fibers may account for this response.

Deng and Geisler used Shamma's technique (1985) to construct 3-D displays that showed the spatio-frequency characteristics of the responses of populations of auditory-nerve fibers. These displays showed that fibers synchronized their responses to both the murmur and the vowel of the syllable. In /mu/ the bandwidth of fibers entrained to the second formant was much wider than that in /nu/; fibers with CFs up to 3200 Hz were entrained to the second formant in /mu/ while only fibers whose CFs were at or slightly above the F2 in /nu/ were entrained to the second formant frequencies. In the syllables /mi/ and /ni/ there was strong synchrony to harmonics of the F0 that lay near the CFs of the fibers whose CFs were <1600 Hz. This effect was especially notable in the vowel. Deng and Geisler attributed this observed synchrony in /mi/ and /ni/ to the fact that formant frequencies in /i/ are very high and far away from the CFs and, therefore, had less influence on those lower CF fibers. This effect had previously been observed by Sinex and Geisler (1983), Delgutte and Kiang (1984), and Carney and Geisler (1986).
Deng and Geisler concluded that the spectral characteristics of nasal-consonant vowel syllables were encoded by discharge patterns of the auditory-nerve fibers in the murmur, vocal, and transition regions. There was one significant difference between this study and the Carney and Geisler (1986) study. A comparison of the vocalic portion of /mu/ to that of /tu/ by Deng and Geisler showed that fibers with CFs between the second formant frequencies and 4kHz entrained to the F2, this included fibers whose CFs were near the frequencies of the third formant. In the Carney and Geisler study, fibers with CF near the frequencies of the third formant entrained to the F3 not to the F2. Deng and Geisler attributed this difference to the differences in amplitudes of the F2 and F3 in the vocalic portions of the stimuli used in the two studies. In their study, there was a 18-dB difference between the second and third formants in /mu/; there was a difference of 8 dB between the formants in /tu/. The authors stressed the significance of a consideration of the relative heights of the different formants in the stimuli when analyzing the responses of auditory-nerve fibers.

2.6 Encoding of frequency modulation
in auditory fibers: electrical stimulation

In the cochlear implant device, acoustic information extracted from the speech signal is encoded by the circuitry of the speech processor and that information is used to establish the parameters for the train of electrical pulses that stimulate the neurons of the VIIIth nerve. While a great deal is known about the responses of the auditory nerve fibers to the acoustic
parameters of speech, not much data is available about the responses of auditory-nerve fibers to speech-like electrical stimuli (Abbas, 1993). Studies have been done using sinusoidal or pulsatile electrical stimuli. These studies have shown that the tuning curves of auditory fibers in response to electrical stimuli are broad. The range of threshold of the response to electrical stimuli is only about 12 dB for all fibers and varies little with stimulus frequency. Nearly all the fibers have their lowest threshold in response to electrical stimulation at approximately 100 Hz (van den Honert and Stypulkowski, 1987). It should be recalled that in response to acoustic stimuli, auditory-nerve fibers have sharp tuning curves. Whereas under conditions of electrical stimulation all VIIIth nerve fibers have their lowest threshold when the frequency of the stimulus is 100 Hz, under conditions of acoustic stimulation the frequency to which each fiber is most responsive is one of a broad range of frequencies. Fibers respond differently to changes in intensity level of acoustic and of electrical stimuli. Electrically-stimulated fibers have extremely steep growth functions with very narrow dynamic ranges, 2-6 dB, versus 30 dB for acoustically-stimulated single fibers. Also, the maximum discharge rate of the fiber is greater for electrical stimulation (Abbas, 1993).

Electrically-stimulated nerve fibers show a stronger degree of phase locking than do acoustically-stimulated fibers (van den Honert and Stypulkowski, 1987; Javel, 1990). PST histograms also show differences in the responses of auditory-nerve fibers to electrical and to acoustic stimulation. The latency of response time is shorter with electrical stimulation, reflecting the fact that there is no time delay introduced by the traveling wave as it
moves along the basilar membrane. Also, the response peaks are narrow because the basilar membrane does not ring, and there is no delay due to the time involved for release and binding of the neurotransmitter at the synapse between the hair cells and the neurons (Abbas, 1993). PST histograms also show two peaks for each phase of electrical stimuli. Because the polarization/depolarization response that reflects the bending of the hair cells by the basilar membrane is not a factor with direct stimulation of the auditory nerve, the nerve fiber responds to each phase of the stimulus (van den Honert and Stypulkowski, 1987).

Adaptation does occur in electrically-stimulated neurons, with responses to the stimulus decreasing over time. However, this occurs only within very narrow intensity ranges. At high stimulus levels, adaptation has not always been observed (Abbas, 1993).

2.7 The perception of elements of a synthetic speech continuum that differ on the basis of place of articulation

Numerous analyses have been done of consonant perception by cochlear-implant users (Blamey, Dowell, Brown, and Clark, 1987; Blamey et al., 1992; Dorman, Hannley, McCandless, and Smith, 1988; Dorman et al., 1990; McKay and McDermott, 1993; Rosen, Walliker, Brimacombe, Edgerton, 1989; Skinner et al., 1991; Tye-Murray and Tyler, 1989; Tyler and Moore, 1992;) Natural speech stimuli were used in all of these studies except the one by Dorman et al. (1988) in which a synthetic /ba/-/da/ continuum of eight elements was created using the Klatt algorithm (1980). The general conclusion from these
studies was that the implant users could consistently identify and
discriminate consonants on the basis of manner of production, that is, how
they are produced (stops, fricatives, nasals, affricates, glides), but had
difficulty distinguishing consonants on the basis of place, that is, the location
of the consonantal constriction in the vocal tract.

As an adjunct to the psychacoustic study under discussion in this thesis,
we tested the subjects on their ability to identify and discriminate between
synthetic speech stimuli that differ on the basis of place. We wished to see if
cochlear-implant subjects, using the Nucleus device, would have the same
difficulty differentiating synthetic stop consonants as did the subjects in the
study by Dorman et al. (1988) in which the subjects all used the Symbion
device. We also wished to determine if our subjects’ performance with
speech stimuli containing transition-only cues would reflect their
performance on tasks employing nonspeech frequency-modulated stimuli
that resemble frequency transitions.

The synthetic continuum used in this study was based on parameters of a
/ba/, /da/, /ga/ continuum specified by Stevens and Blumstein (1978). This
continuum was chosen because the elements of the continuum had
transition only cues; release burst cues were absent. It was felt that these
stimuli resembled the frequency-modulated signals used in the step-versus-
glide psychoacoustic task.

Stevens and Blumstein used stimuli that consisted of transitions only,
release burst plus transitions, and burst only. The modified consonants were
accompanied by one of the following vowels, /a/, /i/, or /u/. When the
stimuli consisted of release burst plus transition, 90% of the subjects with
unimpaired hearing could consistently label the initial consonant as /b/, /d/, or /g/. For transition-only stimuli, the percent of subjects who consistently labeled the initial consonant as /b/, /d/, or /g/ dropped to 81%. Only 18% could consistently label stimuli with burst-only cues.

The authors hypothesized that the poor performance of the subjects with stimuli having burst-only cues was a function of confusion caused by the close proximity of two onsets, the onset of the burst and the onset of voicing. They suggested that the role of formant transitions in the perception of consonant place was not a direct one; rather, transitions functioned to provide a smooth continuum between the onset spectrum and the vowel thereby preventing the conflict between the onset cues of voicing and burst.

Van de Graff Turek, Dorman, and Franks (1980) used a synthetic speech /ba/, /da/, /ga/ continuum to investigate the identification of consonants differing on the basis of place of articulation by listeners with sensorineural hearing losses. Under conditions of both monotic and dichotic presentations of F1 and F2/F3 information, the authors found a high degree of variability in labeling performance. A few of the subjects had nearly normal identity functions, but most had identity functions with varying degrees of abnormality.

2.8 Comments on potential limiting factors in speech perception in cochlear-implant users

As noted earlier, the Haskins Laboratory studies showed that rapid shifts in the second and third formants can be cues for the distinction of place in
stop consonants. As cochlear-implant users have been shown to have temporal resolution capabilities as good as or even superior to those of normal-hearing listeners as measured by psychoacoustic tasks which require detecting gaps in a signal, changes in duration of the signal, or amplitude modulation of a signal, one must question whether performance on tests that measure the ability to distinguish consonants is a reflection of an inability to follow rapid frequency changes in the signal or whether it is a reflection that current measures of speech perception in cochlear-implant users are inadequate.

Factors, other than impaired temporal resolution, that could limit cochlear-implant wearers in their use of dynamic information as perceptual cues in speech: 1) inherent differences in the responses of auditory-nerve fibers to acoustic and to electrical stimulation and 2) limitations imposed by the design of the speech processors and the strategies used in them.

As noted above, several investigators have shown that auditory-nerve fibers synchronize, or phase lock, to components of an acoustic stimulus; fibers also have been shown to phase lock even more strongly to electrical stimuli. However, phase locking to acoustic stimuli is half-wave rectified; that is, the fiber discharges in response to only the upward deflection of the basilar membrane. Phase locking to electrical stimuli is full-wave rectified; that is, the fiber responds to both phases of the electrical stimulus waveform (this is true for both sinusoidal and pulsatile biphasic stimuli). If synchronization is important for the encoding of frequency information, it would appear that the encoded frequency of a given electrical signal would be twice that of the encoded frequency of an acoustic signal of the same actual
frequency. [van den Honert and Stypulkowski (1987) showed this would be true only if the first phase of the electric pulse were upgoing.] It should be kept in mind that the rate of stimulation of the nerve cell bodies in the spiral ganglia is only in part a function of the incoming acoustic signal. In the MSP (MPEAK strategy) the rate of stimulation for aperiodic/voiceless sounds is determined by F1 information in an incoming speech signal and is limited by the filtering characteristics of the speech processor to a maximum of 1000 Hz. If the sound is voiced, however, the rate of stimulation is determined by the F0. Rate of stimulation for periodic environmental sounds is determined by any low frequency periodicity that is detected. In the Spectra, rate depends on the spectral complexity of the signal and varies between 180-300 Hz.

Another limitation to encoding dynamic change that is inherent in the different responses of auditory-nerve fibers to acoustic and to electrical stimuli might be in the differences in response to SPL. Several of the authors whose studies have been discussed noted that changes in the discharge rates of fibers and changes in which particular fibers were entrained to components of the stimuli occurred as the SPL changed during the stimulus. Because of the steep loudness/intensity growth function with electrical stimulation, the changes in firing rate and fiber entrainment with changes in SPL may occur too quickly and, therefore, a given intensity contrast between the beginning and end of a transition may be encoded too quickly relative to the onset of the transition and not reflect the true time course of intensity contrasts in the signal.

The ability of cochlear-implant users to follow rapid changes in the acoustic signal could be limited by the design and processing strategies used
in the speech processors. Kent and Read (1992) note that the average formant transition duration for stops is 50 ms. Madden and Feth (1992) showed that hearing-impaired listeners had temporal resolution thresholds of 13-18 ms. In a 50-ms transition, therefore, the hearing-impaired listener could have access to at least 3 points of information. This may explain why Madden and Feth's hearing-impaired subjects could discriminate between the step and glide stimuli which were 50 ms in duration. The reason that cochlear-implant listeners have difficulty distinguishing consonants on the basis of place may be that they have even higher temporal resolution thresholds than do the hearing-impaired subjects in Madden and Feth's study. They may not be able to distinguish between the step and glide signals simply because they are not getting enough points of information along the frequency transition trajectory.
CHAPTER III
SUBJECTS AND METHODS

In this chapter, the biographic data of the subjects are listed. A detailed description of the frequency characteristics of the signals and the procedure for determining presentation levels of the signals is given. The two-cue, two-alternative, forced-choice psychoacoustic task which was used to determine the temporal resolution threshold is discussed in detail. The methods used for generation of the synthetic speech continuum and for the collection of data in the identification and discrimination of the speech stimuli are also described.

3.1 Subjects

The subjects used in this experiment were volunteer human adults. The cochlear-implant users are patients of R. Goldenberg, M.D. and T. Schneiderman, M.D. of Soifer, Goldenberg Ear Associates and have been diagnosed as having a profound sensorineural hearing loss and as having no open-set speech perception with appropriate conventional hearing aids. The
patients have been implanted in one ear only with a Cochlear Corporation multichannel intracochlear implant device known as the Nucleus 22 (FDA approved).

The implanted subjects used in this study had been implanted for at least five months before testing was initiated and had undergone post-implant rehabilitative procedures and were judged to be fully acclimated to the device before participating in the psychoacoustic experiments. None had previously participated in psychoacoustic studies.

Controls were normal-hearing subjects, matched for age and sex to the cochlear-implant subjects, with pure-tone air-conduction thresholds for frequencies of between 0.5 kHz and 4 kHz of less than or equal to 15 dB HL (ANSI, 1969). Controls NH 1 and NH 2 had no previous experience as subjects in psychoacoustic tasks. NH 3 was an experienced listener. Biographic data for both the implanted subjects and the control subjects are listed in Table 1.

Three implanted subjects and three normal-hearing subjects (controls) were paid for their participation in the study. Implanted subjects were paid at a rate of $10.00 per hour, and controls were paid at a rate of $5.00 per hour. Subjects were asked to sign consent forms approved by the Wright State University Human Subjects Review Committee (SC# 1477) and by The Ohio State University Human Subjects Review Committee (Protocol No. 92H0190). At the time the consent forms were signed, the procedure was fully explained to the participants.

Generally, both implanted and control subjects were tested in two sessions per week, two hours per session. Sessions for subject CI 2 were shortened to
Table 1. Biographic data for cochlear-implant and normal-hearing subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (Yrs.)</th>
<th>Sex</th>
<th>Etiology</th>
<th>Length of Deafness</th>
<th>Length of implant use*</th>
<th>Processor</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI 1</td>
<td>70</td>
<td>M</td>
<td>Fever/high altitude flying</td>
<td>50 years</td>
<td>5 months</td>
<td>Spectra</td>
</tr>
<tr>
<td>CI 2</td>
<td>37</td>
<td>F</td>
<td>trauma or fever</td>
<td>35 years</td>
<td>18 months</td>
<td>MSP</td>
</tr>
<tr>
<td>CI 3</td>
<td>57</td>
<td>F</td>
<td>trauma</td>
<td>30 years</td>
<td>3.5 years**</td>
<td>MSP</td>
</tr>
</tbody>
</table>

Normal-hearing subjects

<table>
<thead>
<tr>
<th>Subject</th>
<th>Age (Yrs.)</th>
<th>Sex</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH 1</td>
<td>63</td>
<td>M</td>
</tr>
<tr>
<td>NH 2</td>
<td>34</td>
<td>F</td>
</tr>
<tr>
<td>NH 3</td>
<td>52</td>
<td>F</td>
</tr>
</tbody>
</table>

*At the time of initial testing
**Single channel implanted in 1985; device failed
one and one-half hours per session. Because she was tested at the end of her regular working day, a two-hour session proved to be too fatiguing for her. Control subject NH 2 was tested as often as three times per week because of a time constraint. Subjects received no extended training on the psychoacoustic task; rather, short 10-trial blocks were used for training. Once it was clear that the subjects understood the mechanics of the task, testing was begun.

The total length of time over which data was collected from implanted subjects depended on the point at which a subject became unable to distinguish between a glide and a step signal. If the subject was unable to distinguish between the two signals at the two-step level then discrimination at the three-step level was not tested. There was no cochlear-implant subject who could distinguish between the step and glide signals for every one of the seven signals on which he/she was tested. One of the problems with testing cochlear-implant subjects is that they are extremely variable in their responses. The original goal for the psychoacoustic task was to collect, for each signal, 12 blocks of data with percent correct averages on each block of 75% to 85% and with at least six to eight reversals on each block. On some signals, approximately 25 to 30 blocks of data were collected in order to get 12 blocks that met these criteria. The collection of data from cochlear-implant subjects on the psychoacoustic task required approximately 15-17 weeks. For normal-hearing subjects, data collection required approximately 13 weeks to meet the goals. Data was collected from these control subjects on all seven signals using all parameters i.e., upsweeps, downsweeps, two-, three-, and five-steps.
All testing was carried out in the Audiology and Speech Pathology Service at the Dayton Veterans Administration Medical Center in Dayton, Ohio.

3.2 Stimuli

The exact center frequencies and excursions of the stimuli selected for the cochlear-implant subjects were based on the frequency ranges of the active electrodes. These ranges were specified in the MAP (the individualized coding scheme that is established during the activation of the implant device and during the patient's rehabilitation period) that the individual cochlear-implant patient was using at the time of the initial testing. Each cochlear-implant subject, therefore, had his/her own unique set of seven signals with which he/she was tested. An attempt was made to match as closely as possible the center frequencies and frequency ranges of the signals across the cochlear-implant subjects.

Some of the frequency transitions were designed to fall within the frequency range of a single electrode; some were designed to sweep across at least three adjacent electrodes. The following criteria were used for selecting the center frequencies and frequency ranges of the signals used in this study:

1) The center frequency of a signal had to be near 500, 1000, 2000, or 3000 Hz.
2) The range of frequencies of a set of signals was from 350-4000 Hz.
3) Multi-electrode sweeps had to cross at least three electrodes.
4) The range of a within-electrode sweep had to fall within the range of the middle electrode of a corresponding three-or-more electrode sweep.

5) All signal sweeps had to be at least 20 Hz inside the boundaries of the electrode frequency ranges established during the MAPping process.

The frequency characteristics of each set of signals for each cochlear-implant subject are summarized in Table 2. Once the signal parameters were defined for a particular cochlear-implant patient these same parameters were used to define the signals that were presented to his/her age- and sex-matched normal-hearing control.

3.2.1 Signal frequency

Glide signals  The glide signals were sinusoidal tones which traversed a given frequency range in 20 discrete steps. Because an adaptive procedure was used in the step-versus-glide discrimination task, the signal duration varied. The rise/fall time was 10 ms. Onsets and offsets were shaped by a cosine-squared function. Subjects were tested with signals that were both upsweeps, that is, signals changing from lower to higher frequencies, and downsweeps, that is, signals changing from higher to lower frequencies.

According to Madden (1994), the detection of the steps in a 20-step signal exceeds the temporal resolution capabilities of normal-hearing listeners; therefore, they are incapable of distinguishing a 20-step signal from a linearly
Table 2: Frequency characteristics of step and glide signals for cochlear-implant subjects

<table>
<thead>
<tr>
<th>CF (Hz)</th>
<th>Sweep</th>
<th>Range</th>
<th>Electrode number</th>
<th>CF (Hz)</th>
<th>Sweep</th>
<th>Range</th>
<th>Electrode number</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>575</td>
<td>230</td>
<td>460-690</td>
<td>19,18,16*</td>
<td>450</td>
<td>200</td>
<td>350-550</td>
<td>20,19,18</td>
</tr>
<tr>
<td>1070</td>
<td>300</td>
<td>920-1220</td>
<td>15,14,13</td>
<td>1065</td>
<td>290</td>
<td>920-1210</td>
<td>14,13,12</td>
</tr>
<tr>
<td>2000</td>
<td>600</td>
<td>1700-2300</td>
<td>11,10,9</td>
<td>2025</td>
<td>600</td>
<td>1725-2325</td>
<td>8,7,6</td>
</tr>
<tr>
<td>1950</td>
<td>180</td>
<td>1860-2040</td>
<td>10</td>
<td>2000</td>
<td>160</td>
<td>1920-2080</td>
<td>7</td>
</tr>
<tr>
<td>3000</td>
<td>600</td>
<td>2700-3300</td>
<td>8,7,6</td>
<td>3100</td>
<td>600</td>
<td>2800-3400</td>
<td>4,3,2,(7?)</td>
</tr>
<tr>
<td>2950</td>
<td>300</td>
<td>2800-3100</td>
<td>7</td>
<td>3400</td>
<td>300</td>
<td>3250-3550</td>
<td>2</td>
</tr>
<tr>
<td>3200</td>
<td>1500</td>
<td>2450-3950</td>
<td>5,6,7,8</td>
<td>3305</td>
<td>1350</td>
<td>2630-3980</td>
<td>4,3,2,1,(7?)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>CF (Hz)</th>
<th>Sweep</th>
<th>Range</th>
<th>Electrode</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>450</td>
<td>200</td>
<td>350-550</td>
<td>19,18,17**</td>
</tr>
<tr>
<td>1070</td>
<td>300</td>
<td>920-1220</td>
<td>13,12,11</td>
</tr>
<tr>
<td>1900</td>
<td>600</td>
<td>1600-2200</td>
<td>8,7,6</td>
</tr>
<tr>
<td>1890</td>
<td>180</td>
<td>1800-1980</td>
<td>7</td>
</tr>
<tr>
<td>3100</td>
<td>600</td>
<td>2800-3400</td>
<td>4,3,2,(6?)</td>
</tr>
<tr>
<td>3000</td>
<td>300</td>
<td>2850-3150</td>
<td>3</td>
</tr>
<tr>
<td>3300</td>
<td>1500</td>
<td>2550-4050</td>
<td>4,3,2,1,(6?)</td>
</tr>
</tbody>
</table>

* Electrode 17 turned off
** Electrode 20 turned off
modulated signal which makes the same frequency transition in a smooth linear path generated in 1-Hz increments. In order to be certain that the 20-step glide was temporally smoothed for all subjects participating in the study, each was tested on his/her ability to discriminate a 20-step signal from a linearly modulated signal using a two-cue, two-alternative, forced-choice procedure (2Q,2AFC) such as that described later in this proposal. The particular signal used to test for this discrimination was the one that, for a given subject, had a center frequency near 1000 Hz. The signal had a fixed duration of 100 ms for the cochlear-implant subjects and of 30 ms for the control subjects.

**Step signals** The step signals were sinusoids with the same characteristics as the glides except they traversed the frequency range in two, three, or five discrete steps (Figure 2).

### 3.2.2 Signal level

When a cochlear implant device is activated, as a part of the MAPping procedure, the threshold and comfort levels for each electrode in the electrode array are established. Whereas the signals used in this experiment swept across different regions of the array and across different numbers of electrodes, it seemed imperative to devise a method whereby a level of presentation could be established that would be comparable across the entire set of signals. To accomplish this, the threshold and upper loudness level (ULL) for every signal in the particular set used for each cochlear-implant subject and each normal-hearing subject were determined in order to
Representation of Glide Signal with 10-ms onset/offset

Three-step signal with 10-ms onset/offset

Figure 2: Schematic representation of a glide (top) and a step signal (bottom)
establish a dynamic range for that signal. The signals were then presented at a level that corresponded to the midpoint in the dynamic range.

**Threshold Determination** A three-down, one-up, two-alternative, forced-choice procedure (2AFC) (Levitt, 1971) that adapted on intensity was used for threshold determination. Subjects were given written instructions (Appendix A) for this procedure. The experimenter also reviewed the instructions verbally. Subjects were seated in a double-walled sound-attenuating chamber (Industrial Acoustics Company, Inc.) that met ANSI standards. The subject faced a computer screen on which was a display as shown in Figure 3. Each listening interval box was illuminated for 200 ms. There was a 200-ms interstimulus interval between the listening intervals. The two middle intervals illuminated in succession.

A fixed-duration signal of 100 ms for cochlear-implant subjects and of 30 ms for normal-hearing controls was presented randomly so that it occurred approximately 50% of the time in the left middle interval and approximately 50% of the time in the right middle interval. Each signal was presented so that its duration was centered on the duration of illumination of the interval i.e., for a 100-ms signal, illumination corresponded to 50 ms of silence, 100 ms of signal, followed by 50 ms of silence. If the subject heard a signal associated with the left middle interval, he/she pressed the left button on a computer mouse. If the signal was heard in the right interval, the subject pressed the right mouse button.

When the subject correctly identified the location of a signal three times in succession, the signal level was reduced. When the subject made a single error, the signal level was increased. For cochlear-implant subjects, the
Figure 3: Visual display as seen by subjects during testing session
reduction or increase in the increment was generally 0.5 dB. After the first run of this task, these subjects were asked if the 0.5 dB increment was sufficient to produce a noticeable increase or decrease in loudness. If it was not, the increment was increased to 1.0 dB. For normal-hearing listeners, the increment was 4 dB. The threshold obtained with this three-down one-up rule corresponds to the 79.4% correct point on the subject’s psychometric function (Levitt, 1971). The subject was given visual feedback as to which illuminated interval was associated with the signal.

In order to familiarize the subjects with the signals, a short block of 20-25 trials was presented at an above-threshold level prior to the actual testing. This level varied with the signal and from subject to subject. Threshold determinations were begun after this short familiarization period. If the percent correct \( p(c) \) on the first block was above 85%, the level of the signal was reduced on successive blocks until a \( p(c) \) of 75%-85% was attained. If the \( p(c) \) on the first block was below 75%, the level of the signal was increased on successive blocks until a \( p(c) \) of 75%-85% was reached.

Each presentation of the signal was considered to be a trial. Stimuli were presented in blocks of three with 50 trials per block for a total of 150 trials. Only blocks on which the \( p(c) \) was between 75% and 85% were included in the calculations. The first two reversals (a reversal is a change in direction) in the run in each block were discarded. If there was a total of eight reversals in a run, the final six were averaged to obtain the average signal intensity for the block. If the total number of reversals was nine, the last seven were averaged. If the number of reversals was ten or more, the average signal intensity for each block was calculated by averaging the last eight reversals of
the 50-trial run. Blocks of data were discarded when there were fewer than six usable reversals. An overall mean signal intensity for each signal was determined from the mean of the average signal intensities from six blocks of trials (a total of 300 trials).

**ULL Determination** Prior to the determination of ULL, subjects were again presented with written instructions (Appendix B) about the task. The investigator reviewed the instructions verbally with the subject. In this task, the goal was to have the subject increase the loudness of the signal just until it became uncomfortably loud. Subjects were cautioned not to let the signal become painful but to allow the signal to just reach the point where they would not want to listen to it for a very long time. Signals of 100 ms fixed duration were used for the ULL determination for cochlear-implant subjects; signals were 30 ms for normal-hearing listeners.

The subjects faced a computer screen with a bank of boxes as shown in Figure 3. For the ULL determination, the sound was presented only in the left middle interval. The initial level of the signal was set at the previously determined threshold level. Pilot studies had shown that starting the task at a level above threshold often resulted in the cochlear-implant subjects incrementing the level only a few times. Because we wanted the subjects to have more time during which the level was actually increasing, the starting point was chosen to be at threshold. From this point on, the level of the signal was under the control of the subject. If the subject felt that a louder signal could be tolerated, he/she pressed the left button on a computer mouse. When the signal reached a level that was just uncomfortably loud,
the subject pressed the right mouse button, and the presentation of the signal ceased.

Because incrementation of the level was under the control of the subject, the number of times a signal was incremented varied from signal to signal and from subject to subject. At the end of the procedure, the subject was asked to describe the signal in terms of whether it sounded soft, comfortable, loud, or uncomfortably loud. The endpoint values from three runs of the procedure were averaged, and this averaged value was used as the ULL level.

With some signals, for both cochlear-implant and normal-hearing subjects, the goal of an uncomfortably loud signal was never attained before the limits of the equipment was reached. Most subjects reported that, if the equipment had allowed them to increment the signal one or two more times, they would have terminated the incrementing process themselves. Occasionally, some cochlear-implant subjects reported that they would reach a level at which the signal did not continue to increase in loudness. Both the Spectra and the Mini Speech Processors have as part of their circuitry an Automatic Gain Control (AGC) feature. Apparently, as a result of compression by the AGC, some signals reached a loudness asymptote before the limits of the equipment were exceeded. When either of the above situations occurred, the maximum output of the equipment was considered as the ULL and the dynamic range was calculated as the difference between this ULL and the threshold for a given signal.

**Signal Level** To measure signal levels in dBSPL, the output from a Wilsonics attenuator (signal delivery described in Section 3.4) was delivered to Senneheiser HD430 headphones coupled to a Quest Electronics Impulse
Integrating Sound Level Meter (Model 2800) with a Quest Electronics earphone coupler Model EC-9A (Quest Electronics, Oconomowoc, Wis.). A Quest Model OB-300 1/3-1/1 Octave Filter plug-in module with a selectable set of filters was used to filter extraneous background noise during the dBSPL measurement of each signal. The input voltage to the Cochlear Corporation Audio Input Selector which was used to deliver the signal to the speech processors of the cochlear-implant listeners was measured with a Tektronix 7613/R7613 Storage Oscilloscope. Signals of 300-ms duration were used for the measurements with the sound level meter and the oscilloscope. A check was made to confirm that the levels of the 30-ms signals were 10 dB lower than those of the 300-ms signals.

3.3 Signal Generation

Both step and glide signals were generated using an Ariel DSP-16 signal acquisition board with a 16-bit D-A converter mounted in a Dell 486D/32 computer. A sampling rate of 100 kHz was used. The resulting signals were low-pass filtered at 20 kHz.

3.4 Signal delivery

The signals were sent from the Ariel board to a programmable Wilsonics attenuator. For the cochlear-implant subjects, the attenuated signal was delivered, without noise reduction, from the attenuator to the processor of the subject via a Cochlear Corporation Audio Input Selector. The Audio
Input Selector was set to "audio only." This setting cut off the external microphone of the processor so that no sounds other than the signal input were delivered to the processor. The volume control of the input selector was set at its highest setting. The subject's processor settings were those that were used for normal everyday sound perception by that particular subject. For normal-hearing subjects, the stimuli were delivered from the attenuator directly to Senneheiser HD430 headphones. Signals were presented monaurally.

3.5 Step-versus-glide discrimination task

Prior to testing, the subject was given a written description of the step-versus-glide discrimination task (Appendix C); the experimenter verbally reviewed this handout with the subject. The subject was seated, in a double-walled sound-attenuating chamber, facing a computer screen on which seven boxes were displayed (Figure 3). The four lower boxes were designated as listening intervals. As each signal was presented, a listening interval illuminated on the screen. When the answer box illuminated, the subject pressed the left or the right button on a computer mouse to record his response. If his response occurred within the time limits of the answer box, the accept box was illuminated.

A 2Q, 2AFC procedure that adapted on duration was used to determine the subject's sensitivity to differences in the glide and step signals. The signal duration decreased following three correct responses and increased following one incorrect response. Adaptation continued until the step signal was just
discriminable from the glide signal. The threshold which is obtained with this three-down one-up rule corresponds to the 79.4% correct point on the subject's psychometric function (Levitt, 1971).

There was a 5-ms floor written into the computer program for this task; hence, the total duration of a signal could be no shorter than 5 ms. In a pilot study, this floor was not included in the computer program. Unfortunately, very good listeners could make the signal disappear entirely.

Establishing signal duration and step size for the cochlear-implant subjects required a period of trial and error for each signal. A duration of 100 ms and adaptive step sizes of 35, 20, and 10 ms were always used the first time a signal was presented to a cochlear-implant listener. Adjustments in duration were made by estimating the temporal resolution threshold from the midpoints of every other reversal beginning with the second reversal. The maximum starting duration of a signal (except as noted later in Chapter IV) was never greater than 300 ms. As will be discussed in a later section, excessively long signals sometimes produced auditory carryover effects that interfered with signals presented in successive intervals. Signal durations and step sizes for the cochlear-implant subjects are listed in Table 3.

Subjects with normal hearing were initially tested with signals that were 30 ms in duration with adaptive step sizes of 20 ms, 10 ms, and 5 ms. Signal duration and step size were adjusted until the subjects could meet the criteria of 75-85% correct responses and six to eight reversals in a 50-trial run. It was necessary to lengthen the initial duration and adaptive sizes for some control subjects for some of the three- and five-step signals. The durations and step sizes for the signals used by the normal-hearing controls are listed in Table 4.
Table 3: Initial signal durations and adaptive step sizes used in the step-versus-glide discrimination task for cochlear-implant subjects. Signal durations are listed above and adaptive step sizes are listed below in parentheses. Both values are in ms.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step up</th>
<th>2-step down</th>
<th>3-step up</th>
<th>3-step down</th>
<th>5-step up</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>460-690</td>
<td>100</td>
<td>30</td>
<td>50</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(20, 10, 5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>920-1220</td>
<td>100</td>
<td>30</td>
<td>150</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1700-2300</td>
<td>100</td>
<td>30</td>
<td>200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1860-2040</td>
<td>100</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2700-3300</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2800-3100</td>
<td>150</td>
<td>150</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2450-3950</td>
<td>30</td>
<td>30</td>
<td>100</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(20, 10, 5)</td>
<td>(20, 10, 5)</td>
<td>(20, 10, 5)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>CI2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>350-550</td>
<td>100</td>
<td>100</td>
<td>300</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>920-1210</td>
<td>100</td>
<td>350</td>
<td>200</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 15, 10)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>1725-2325</td>
<td>100</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1920-2080</td>
<td>100</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2800-3400</td>
<td>100</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3250-3550</td>
<td>200</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2630-3980</td>
<td>100</td>
<td>200</td>
<td>200</td>
<td>500</td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td>(50, 25, 15)</td>
<td></td>
</tr>
<tr>
<td>CI3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>350-550</td>
<td>100</td>
<td>200</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>920-1220</td>
<td>100</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1600-2200</td>
<td>300</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2550-4050</td>
<td>30</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>(20, 10, 5)</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Table 4: Initial signal durations and adaptive step sizes used in the step-versus-glide discrimination task for normal-hearing subjects. Signal durations are listed above and adaptive step sizes are listed below in parentheses. Both values are in ms.

For all normal-hearing subjects, the initial signal durations were 30 ms and the adaptive step sizes were 20, 10, and 5 ms for the 2-step upsweep and 2-step downsweep signals at all frequencies.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>3-step up</th>
<th>3-step down</th>
<th>5-step up</th>
<th>5-step down</th>
</tr>
</thead>
<tbody>
<tr>
<td>NH1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>460-690</td>
<td>100</td>
<td>30</td>
<td>100</td>
<td>150</td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>920-1220</td>
<td>100</td>
<td>30</td>
<td>100</td>
<td>120</td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>1700-2300</td>
<td>30</td>
<td>30</td>
<td>70</td>
<td>70</td>
</tr>
<tr>
<td></td>
<td>(20, 10, 5)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>1860-2040</td>
<td>100</td>
<td>100</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(50, 35, 20)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>2700-3300</td>
<td>100</td>
<td>30</td>
<td>100</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>2800-3100</td>
<td>100</td>
<td>30</td>
<td>200</td>
<td>200</td>
</tr>
<tr>
<td></td>
<td>(35, 20, 10)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
<tr>
<td>2450-3950</td>
<td>30</td>
<td>30</td>
<td>70</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td>(20, 10, 5)</td>
<td>(20, 10, 5)</td>
<td>(35, 20, 10)</td>
<td>(35, 20, 10)</td>
</tr>
</tbody>
</table>

NH 2: Initial signal durations were 30 ms, and the adaptive step sizes were 20, 10, and 5 ms for signals at all frequencies except for the signals listed below.

1920-2080  200  (35, 20, 10)
2800-3400  100  (35, 20, 10)
3250-3550  200  (35, 20, 10)

NH 3: Initial signal durations were 30 ms, and the adaptive step sizes were 20, 10, and 5 ms for signals at all frequencies except for the signals listed below.

1800-1980  150  (35, 20, 10)
2850-3150  70  (35, 20, 10)
A single trial consisted of the presentation of a stimulus in each of four listening intervals. Three of the four intervals had identical signals, the glide. One interval, at random either the second or third interval, contained the step signal. The interstimulus interval time (ISI) was 400 ms. Pilot studies had shown that a long ISI was necessary to avoid auditory “carryover” effects from one interval to another. The subject gave his responses by pressing one of two buttons on a computer mouse. He/she was given feedback as to whether the correct listening interval was chosen.

Stimuli were presented in blocks of 50 trials. Each signal was presented in three consecutive blocks of 50 trials each for a total of 150 trials. The same method described in Section 3.2.2 was used to determine the average signal duration for each block. The mean temporal resolution threshold (MTRT) was determined by dividing the average signal duration by the number of steps in the signal. In cases where the subject could consistently keep the signal duration at or near the 5-ms floor, blocks with averages higher than 85% were included in the calculations. Generally, twelve blocks (600 trials) obtained from two-step signals and nine blocks (450 trials) obtained from three- and five-step signals were averaged to determine the mean temporal resolution thresholds (i.e., the minimum ms/step) necessary to distinguish a step signal from a glide signal.

For some signals, the cochlear-implant subjects could not discriminate between the step and the glide signal. Either the duration of the signal became so long that it exceeded the limits of the equipment, or the p(c)s were always below 75%. In order to ensure that we did not miss learning effects, six blocks of data were collected on these types of signals. If there was no
improvement in performance at the end of the six blocks, further data collection was terminated.

The order of presentation of signals was as follows: two-step upsweeps compared to glide upsweeps, two-step downsweeps compared to glide downsweeps, three-step upsweeps compared to glide upsweeps, three-step downsweeps compared to glide downsweeps, five-step upsweeps compared to glide upsweeps, and five-step downsweeps compared to glide downsweeps. In a given session, all seven signals of a particular sweep direction and step number were presented at least once per session. Signals within a session were presented in a random order.

Generally, a total of 21 blocks was run per two-hour session with the subject taking short breaks every six to nine blocks. As noted previously, one subject could tolerate sessions of only one and one-half hours. At the beginning of the second testing session, cochlear-implant subjects were given a short questionnaire (Appendix D) which was aimed at determining if there were any residual effects from the previous testing session. Testing conditions were always adjusted so that the subjects never experienced any discomfort with the tasks.

### 3.6 Synthetic speech stimuli

A continuum of synthetic voiced-stop consonant-vowel (CV) speech stimuli was produced using the Klatt formant-synthesizer software (Klatt, 1980). The continuum consisted of stimuli composed of an initial stop consonant ranging from [b] to [d] to [g] accompanied by the vowel [a]. The
continuum was generated by systematically manipulating the second (F2), and third (F3) formant transitions according to the parameters of Stevens and Blumstein (1978). Fourteen CV stimuli comprised the continuum; the steady-state frequencies of the first five formants in the vowel [a] were always 720 (F1), 1240 (F2), 2500 (F3), 3600 (F4), and 4500 (F5) Hz. The F1 starting frequency in each of the CV stimuli was 200 Hz and rose in a linear trajectory to the steady-state frequency of 720 Hz. The duration of the F1 transitions was varied systematically from 20-44 ms with the shorter durations at the [ba] end of the continuum and the longer durations at the [ga] end. The F2 and F3 starting frequencies were varied systematically as shown in Table 5. The trajectories in these transitions varied linearly and had durations of 40 ms. The F4 and F5 transitions in all stimuli were flat with both initial and final frequencies at 3400 Hz for the F4 transition and at 4500 Hz for the F5 transitions. A portion of the CV stimulus set is represented diagrammatically in Figure 4.

The burst amplitude parameters in the Klatt synthesizer program were set to 0 in all the stimuli to eliminate the possibility that bursts might be used as cues in identification and discrimination tasks. The first glottal pulse of these synthetic voiced consonants was set to coincide with the beginning of the formant transition. The fundamental frequency (F0) was set initially at 120 Hz and declined systematically across the 300-ms duration of the tokens to 100 Hz. The first formant bandwidth was constant throughout the transition at 130 Hz; the second formant bandwidth was constant at 70 Hz; the third formant bandwidth was constant at 160 Hz; and the fourth formant bandwidth was constant at 200 Hz.
Table 5: Initial frequencies for the synthetic consonant formant transitions

<table>
<thead>
<tr>
<th>Stimulus number</th>
<th>Initial F2 frequency</th>
<th>Initial F3 frequency</th>
<th>F1 Duration (ms)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>900</td>
<td>2000</td>
<td>20</td>
</tr>
<tr>
<td>2</td>
<td>1010</td>
<td>2110</td>
<td>22</td>
</tr>
<tr>
<td>3</td>
<td>1120</td>
<td>2220</td>
<td>24</td>
</tr>
<tr>
<td>4</td>
<td>1240</td>
<td>2340</td>
<td>26</td>
</tr>
<tr>
<td>5</td>
<td>1350</td>
<td>2450</td>
<td>28</td>
</tr>
<tr>
<td>6</td>
<td>1470</td>
<td>2570</td>
<td>30</td>
</tr>
<tr>
<td>7</td>
<td>1580</td>
<td>2680</td>
<td>32</td>
</tr>
<tr>
<td>8</td>
<td>1700</td>
<td>2800</td>
<td>34</td>
</tr>
<tr>
<td>9</td>
<td>1680</td>
<td>2630</td>
<td>36</td>
</tr>
<tr>
<td>10</td>
<td>1670</td>
<td>2460</td>
<td>38</td>
</tr>
<tr>
<td>11</td>
<td>1650</td>
<td>2280</td>
<td>40</td>
</tr>
<tr>
<td>12</td>
<td>1640</td>
<td>2100</td>
<td>42</td>
</tr>
<tr>
<td>13</td>
<td>1620</td>
<td>1920</td>
<td>44</td>
</tr>
<tr>
<td>14</td>
<td>1610</td>
<td>1750</td>
<td>44</td>
</tr>
</tbody>
</table>
Figure 4: First seven elements of synthetic /ba/, /da/, /ga/ continuum
A 12-bit quantization was used in the creation of the stimuli. The sampling rate was 10,000 Hz. The amplitude of voicing was kept constant at 60 dB. The digitized files were normalized so that all the vowels had the same amplitude (equal maximum excursions). Ten-msec onramps (sine-squared) and 10-ms offramps (cosine-squared) were added. A Turbo editor was used to create blocks of stimuli from the digitized files, and the software program, Hypersignal, was used to randomize the stimuli for the discrimination and identification tasks. A 4.5 kHz filter was used at the computer output to filter out high-frequency components that might be generated during the digital to analog conversion.

3.7 Discrimination and identification task

using the synthetic speech stimuli

Two tasks were used as measures of consonant perception, a same/different discrimination task and an identification task. For the discrimination task, the 14 stimuli were arranged into all possible combinations of 2-step pairs producing a total of 12 pairs (e.g., CV1 CV3; CV2 CV4; CV3 CV5, etc.) With the Turbo editor, four combinations of the members of each pair were generated (e.g., CV1 CV3, CV3 CV1, CV1 CV1, CV3 CV3). Ten randomizations of each combination resulted in the production of 480 stimuli. These 480 stimuli were arranged into two blocks of 240 stimuli per block. The interstimulus interval (ISI) between the members of each pair was 500 ms, and the intergroup interval (IGI) between each pair of stimuli was three seconds. Twenty-second pauses inserted after
every ten pairs served as markers which prevented subjects from getting lost as they performed the task.

Subjects were given written instructions about the discrimination task (Appendix E). They were instructed to guess when they were unsure about whether the members of the pair sounded “the same” or “different” and to try to guess “the same” equally as often as they guessed “different”. The subjects were given a response sheet which had 20 consecutively numbered same/different response items per page. Ten numbered items were listed on the left side of the page, ten on the right. The 20-second pauses were identified in bold print at the end of each column of 10 items. Subjects were instructed to circle their responses during the three second intervals between each presentation of pairs. A practice session consisting of 10 pairs was given before the real discrimination task started. The subject was permitted to listen to the practice block more than once if he/she desired. A short break was given between the administration of blocks one and two. The discrimination task required approximately 45 minutes to complete. After the subject completed the task, he/she was given a short break before starting the next procedure.

The second task was an identification task. Each of the 14 CV stimuli was randomly presented 10 times for a total of 140 stimuli. The stimuli were presented in groups of ten with an ISI of three seconds and a pause of 20 seconds between every group. Written instructions similar to those for the discrimination task (Appendix F) were provided. The subjects were given a response sheet which had 20 consecutively numbered [ba], [da], [ga] response items per page. Ten numbered items were listed on the left side of the page,
ten on the right side. Twenty-second pauses were identified in bold print at the end of each column of ten items. As noted previously, subjects were instructed to circle their response during the 3-second ISI. A practice session with ten stimuli to identify was given prior to the administration of the actual identification task. The identification practice session and the identification task required approximately 15 minutes.

3.8 Delivery of the synthetic speech stimuli

The stimuli were generated by the same Ariel DSP-16 board mounted in the Dell 486D/32 computer that was used for generating the step/glide signals. The computer output was sent to the programmable Wilsonics attenuator set at an attenuation that produced stimuli at a comfortable listening level of 75 dBSPL as measured with the Quest sound level meter as previously described. Normal-hearing listeners received the stimuli monaurally through Senneheiser HD430 headphones. Stimuli were presented to cochlear-implant subjects CI 1 and CI 2 at the same 75 dBSPL level via the Audio Input Selector as described in a previous section. The voltage level corresponding to this dB level was 350 mV measured with the Tektronix oscilloscope at the input to the Audio Input Selector. Subject CI 3 found this level to be too soft; therefore, the listening level was adjusted with the Wilsonics attenuator to a level she described as comfortable. This level was measured at 2000 mV at the input to the Audio Input Selector.
3.9 Data analysis

The only statistical procedure employed in the analysis of the psychoacoustic data obtained in this study was the two-tailed $t$ test. The use of more sophisticated statistical methods was precluded because assumptions underlying an analysis of variance (Keppel, 1991) were violated by the design of the experiment and by the performance of the subjects.

Analysis of variance models are based on the assumption that subjects are randomly assigned to different treatments. In the design of this experiment, the subjects were not randomly assigned to different treatments. Each cochlear-implant subject had a set of signals whose center frequencies and frequency sweeps were based on the MAPs in his/her own processor. Each cochlear-implant subject, in other words, had a unique "treatment". Each normal-hearing control, age- and sex-matched to a cochlear-implant subject, listened to the same signals as did his/her match. Hence, each control subject had his/her unique "treatment."

A second assumption of the analysis of variance models is that the variances in the scores of the treatment populations are equal. As will be shown in Chapter IV, the degree of variability in the performance of the cochlear-implant subjects was extremely high. Even among the control subjects, there was a high degree of variability under the five-step conditions. The second assumption of analysis of variance models was, thus, violated by the performance of the subjects in the study.

A third assumption of analysis of variance models is that the samples in an experiment are drawn randomly from the treatment population. There
were just three cochlear-implant subjects participating in this study. One has to consider that the total population of hearing-impaired listeners with cochlear implants is small; therefore, considered as a proportion of the total population, the population of three probably did not violate this third assumption. Because the other two assumptions were not met, however, it was deemed more appropriate to apply a case studies approach to the analysis of the data.
CHAPTER IV
RESULTS

The results for the step-versus-glide discrimination task are presented in this chapter in both graphical and tabular form. The original goal in this research was to collect 12 blocks of data per condition for each subject. As will be noted later in this section, establishing a signal duration and adaptive step size that would permit the cochlear-implant subjects to meet the criteria of 75-85% correct responses and at least six to eight reversals per block of 50 trials was a time-consuming process. Contributing to the extended-time problem was the high degree of variability in the performance of the cochlear-implant subjects on the step versus glide discrimination task. The combination of these two factors made the collection of 12 blocks of data that met the criteria, a formidable task.

Subjects participating in this study were initially told that the experiment would last for 10-12 weeks. Due to the unanticipated problems with the data collection, testing time extended well beyond the original estimate. In order to accommodate the schedules of the listeners, nine blocks of data were collected for the three-step-up and three-step-down conditions and six to
nine blocks for the five-step-up and five-step-down conditions. In some instances fewer than six blocks of data were used to calculate the mean temporal resolution threshold (MTRTs). Instances in which fewer then six blocks were used are indicated by a "#" in Tables 6-16.

One of the difficulties of using cochlear-implant subjects in long-term studies is that occasionally conditions beyond the experimenter's control arise that necessitate adjustments in, or the replacement of, a subject's MAP. This situation occurred with subject CI 1. During the course of the data collection, CI 1 became ill with a cold. At this time, certain sounds in his everyday environment became intolerably loud. CI 1's audiologist attempted to alleviate this problem by adjusting the T and C levels in the MAP. When these adjustments failed, the subject's MAP was replaced with a new one. Because the new MAP had electrode frequency ranges that were different from those in the previous MAP, testing of CI 1 on the step-versus-glide discrimination task could not be continued beyond the three-step-upswep condition.

Data collection in the normal-hearing subjects also became difficult when the step number of the signal was increased to five. All of the subjects reported that it was difficult to "hold onto" the discrimination cue at this step level. If a normal-hearing listener could easily discriminate between the step and glide signals early in a run of 50 trials, the adaptive step size decreased to 5 ms early in the run. If the subject then lost the discrimination cue late in the run, the duration of the signal would increase in 5-ms increments until the signal once again became long enough for the subject to pick up the cue he/she used for discrimination. By the time this point was
reached, the run was often near its end, and there would be too few trials remaining for the subject to again reach his/her threshold. Because the subject might lose the cue at any point in the run, the final thresholds were extremely variable.

Due to the high degree of variability in performance, the data for cochlear-implant listeners on the psychoacoustic tasks were analyzed on the basis of individual rather than group performance. As noted in Chapter III, each normal-hearing control listened to a slightly different set of signals; hence, their performance on the step-versus-glide discrimination could not be evaluated statistically as a group.

Performance of both groups of listeners on the identification and discrimination of elements of the synthetic speech continuum is presented graphically and in terms of $d'$, a measure of the sensitivity of the observer to differences between stimuli.

4.1 Initial signal durations and step sizes used to collect data for the step-versus-glide discrimination task

In the early stages of testing, cochlear-implant subjects required initial durations longer than 300 ms for certain signals in order to distinguish the step from the glide signal 75%-85% of the time. The use of initial durations longer than 300 ms was abandoned for three reasons. First, signals of such long durations made the task inordinately long. Second, under conditions of long duration, all of the cochlear-implant subjects complained of hearing auditory afterimages. For example, they described hearing, in interval three,
a step signal presented in interval two or hearing, in interval four, a step signal presented in interval three. Third, equipment-related limits precluded using initial durations longer than 300 ms. Because the psychoacoustic procedure adapted on duration, each time the subject missed a trial the signal increased in length. With a large step size, e.g. 35-50 ms, the signal could become as long as 650 ms. At this point, the limits of the testing equipment were reached. The computer program would "roll over" and begin presenting signals of very short duration.

Table 3 lists the initial signal durations and the step sizes for signals on which the cochlear-implant subjects could meet the criteria of 75%-85% correct responses and six to eight reversals/run of 50 trials. As can be seen, initial durations ranged from 30-300 ms with the adaptive step size series ranging from 20, 10, and 5 ms to 50, 35, and 20 ms. In two instances, both occurring with cochlear-implant subject CI 2, the initial signal duration exceeded 300 ms, being 350 ms for the two-step 920-1210 Hz downsweep and 500 ms for the single five-step upsweep on which she could distinguish the step and the glide signals.

For subjects with normal hearing, the initial signal duration was usually 30 ms, and the size of the adaptive step size series in a run was usually 20, 10, and 5 ms. Table 4 shows that the main exceptions to these durations were the five-step upsweep signals with center frequencies near 2000 Hz or higher. As was noted in the introduction to this section, the normal-hearing controls experienced difficulty in holding onto the cue for the discrimination of the five-step signal from the glide signal. Subject NH 1 generally required somewhat longer initial signal durations and adaptive step sizes than NH 2.
and NH 3 on the three-step upsweeps and downsweeps. Like NH 2 and NH 3, he required initial durations for the five-step signals that were three to five times longer than those of signals with smaller step numbers.

4.2 Intensity levels corresponding to one-half of the dynamic range

For subject CI 1, the peak-to-peak voltages corresponding to one half of the dynamic range of each of the seven signals used in the step-versus-glide discrimination task, measured at the input to the Audio Input Selector, ranged from 450-920 mV. The peak-to-peak voltages for the seven signals used by subject CI 2 ranged from 200-600 mV, and for subject CI 3 from 290 to 1200 mV. For all three subjects, the highest voltages were associated with the signal with the lowest center frequency and with the signal with the shortest frequency sweep. For example, for subject CI 3, 1200 mV corresponded to one half the dynamic range for the signal with a frequency sweep of 1800-1980 Hz. The voltage at one half of the dynamic range for the signal with a frequency sweep of 350-550 Hz was 660 mV. Voltage measurements were not made at the output of the Audio Input Selector because the available measuring equipment was not sufficiently sensitive. Rough estimates of the output were made, however, and were on the order of 10 mV or less, peak-to-peak amplitude.

For the normal-hearing controls, levels corresponding to one half the dynamic range for the series of seven signals ranged from 72-82 dBSPL for NH 1, from 70-80 dBSPL for NH 2, and from 61-75 dBSPL for NH 3. As was observed in the cochlear-implant subjects, the highest levels tended to be
associated with the signal with the lowest center frequency and with the signal with the shortest frequency sweep. There was no difference in level or voltage for the same signals having different step numbers, nor was there a difference between signals traversing the same frequency sweep in an upward or a downward direction.

4.3 Temporal resolution thresholds as determined from the step-versus-glide discrimination task

MTRTs of each cochlear-implant subject are compared directly to those of his/her matched normal-hearing control on a condition-by-condition basis in Figures 5-9. The performances of each subject are presented individually in Figures 11, 13, and 15 where MTRTs, expressed as ms/step, are compared across all frequencies, step numbers, and directions. The data from which these figures were produced are shown in tabular form in Tables 6-16. Mean signal durations, standard deviations, and ranges; MTRTs, standard deviations, and ranges; and average percent correct across all blocks of data in the step-versus-glide signal discrimination task are shown for the cochlear-implant subjects in Tables 6-10 and for the normal-hearing subjects in Tables 11-16.

It should be noted that a durational floor was written into the computer program for the discrimination task. In a pilot study this experimental paradigm was used without a durational floor. When subjects were successful at discriminating the step from the glide, the signal, because it adapted on duration, became progressively smaller until it disappeared.
Because of the floor, the true MTRTs under some conditions used in this current study are, therefore, actually lower than the MTRTs reported. These conditions are noted with a double asterisk in the tables. Because the total duration of the signal could get no lower than 5 ms, the minimum MTRT for the two-step signal is 2.5 ms/step; for the three-step signal, 1.7 ms; and for the five-step signal, 1.0 ms/step. If the signal could decrease below 5 ms in duration, then, under some conditions, the MTRTs for the two-and three-step signals could become as low as those of the five-step signals. The floor effect on the MTRT can be recognized by a signal that has the same total duration across increasing step numbers but that has a decreasing step duration.

The reader should bear in mind that one might anticipate that as the number of steps in the signal increases, the signal becomes smoothed, and it becomes more difficult for the subject to distinguish between a step and a glide. Because the procedure adapts on duration, the subject may be able to compensate for the temporal smoothing by increasing the total duration of the signal just to the point where the MTRT (ms/step) is the same for signals with high step numbers as it is for signals with low step numbers; in this case, the duration of each step will remain the same across the step number range. On the other hand, the temporal smoothing may become so great for certain signals, that the subject must increase both the duration of the step and the total duration of the signal across the step number range.

It should be noted that in some instances the data in Tables 6-16 show percent correct responses equal to or greater than 85%. Because of the durational floor in the program, under certain conditions, the subjects were
operating above threshold and, hence, the percent correct responses were artificially high.

4.3.1 Temporal resolution thresholds as determined from the step-versus-glide discrimination task:

Listeners with cochlear implants

Figures 5-9 show that, with few exceptions, the cochlear-implant listeners had higher MTRTs than did listeners with normal hearing. Bars with asterisks indicate, that all three cochlear-implant subjects were unable to distinguish between the step and glide signals at certain frequencies and sweep extents at initial signal durations shorter than 300 ms. There were some conditions, however, under which the cochlear implant listeners had MTRTs that were near the 10-ms upper limit of the MTRTs of the normal-hearing listeners.

MTRTs of 10.5 ms or less were achieved by CI 1 on five signals, four of which were downsweeps (Figures 5 and 10). The lowest MTRT (4.8±1.0 ms) that he achieved was with the 2-step downsweep with a CF of 3200 Hz and a sweep extent of 1500 Hz. The only upsweep signal for which his MTRT was less than 10.5 ms was the three-step signal centered at 475 Hz.

The results from the performance of CI 1 showed that increasing the step number had the effect of temporally smoothing the signal except for the signals with a CF of 475 Hz and a CF of 1070 Hz. CI 1 was able to maintain the same MTRT on both the two-step and the three-step upsweep having a CF of 1070 Hz and sweep extent of 300 Hz. The respective MTRTs of 33.3±19.7 ms
Figure 5: Comparison of mean temporal resolution thresholds for CI 1 and NH 1. I=460-690 Hz; II=920-1220 Hz; III=1700-2300 Hz; IV=1860-2040 Hz; V=2700-3300 Hz; VI=2800-3100 Hz; VII=2450-3950 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Figure 5
Figure 6: Comparison of mean temporal resolution thresholds of CI 2 and NH 2 for upsweep signals. I=350-550 Hz; II=920-1210 Hz; III=1725-2325 Hz; IV=1920-2080 Hz; V=2800-3400 Hz; VI=3250-3550 Hz; VII=2630-3980 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Figure 6
Figure 7: Comparison of mean temporal resolution thresholds of CI 2 and NH 2 for downsweep signals. I=350-550 Hz; II=920-1210 Hz; III=1725-2325 Hz; IV=1920-2080 Hz; V=2800-3400 Hz; VI=3250-3550 Hz; VII=2630-3980 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Figure 7
Figure 8: Comparison of mean temporal resolution thresholds of CI 3 and NH 3 for upsweep signals. I=350-550 Hz; II=920-1220 Hz; III=1600-2200 Hz; IV=1800-1980 Hz; V=2800-3400 Hz; VI=2850-3150 Hz; VII=2550-4050 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Figure 8
Figure 9: Comparison of mean temporal resolution thresholds of CI 3 and NH 3 for downsweep signals. I=350-550 Hz; II=920-1220 Hz; III=1600-2200 Hz; IV=1800-1980 Hz; V=2800-3400 Hz; VI=2850-3150 Hz; VII=2550-4050 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Figure 9
Temporal resolution thresholds for 2-, 3-step upsweeps and downsweeps for subject CI 1

Figure 10: Mean temporal resolution thresholds for cochlear-implant user CI 1. I=460-690 Hz; II=920-1220 Hz; III=1700-2300 Hz; IV=1860-2040 Hz; V=2700-3300 Hz; VI=2800-3100 Hz; VII=2450-3950 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Table 6: Subject CI 1. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep and downsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step up</th>
<th>3-step up</th>
<th>2-step down</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I</td>
<td>49.1±41.7/10.0-113.6</td>
<td>29.1±8.2/14.4-36.3</td>
<td>18.7±6.2/12.1-31.9</td>
</tr>
<tr>
<td>460-690</td>
<td>24.6±20.9/5.0-56.8</td>
<td>9.7±2.8/4.8-12.1</td>
<td>9.4±3.1/6.1-16.0</td>
</tr>
<tr>
<td>2-step down</td>
<td>82.5±1.9</td>
<td>77.8±3.4</td>
<td>81.2±1.6</td>
</tr>
<tr>
<td>II</td>
<td>66.5±39.4/21.3-161.3</td>
<td>117.2±50.4/55.0-202.5</td>
<td>13.9±3.5/10.0-20.0</td>
</tr>
<tr>
<td>920-1220</td>
<td>33.3±19.7/10.7-80.7</td>
<td>39.1±16.8/18.3-67.5</td>
<td>7.0±1.7/5.0-10.0</td>
</tr>
<tr>
<td>1700-2300</td>
<td>86.2±22.5/38.8-128.8</td>
<td>222.7±68.0/112.3-280.0</td>
<td>20.7±12.1/10.6-50.0</td>
</tr>
<tr>
<td>III</td>
<td>43.1±11.3/19.4-64.4</td>
<td>74.2±22.7/37.4-93.3</td>
<td>81.0±3.2</td>
</tr>
<tr>
<td>1700-2300</td>
<td>81.3±1.8</td>
<td>77.2±1.1#</td>
<td>82.8±1.3</td>
</tr>
<tr>
<td>IV</td>
<td>116.0±18.8/87.5-150.0</td>
<td>*</td>
<td>129.6±34.9/78.8-186.3</td>
</tr>
<tr>
<td>1860-2040</td>
<td>58.0±9.4/43.8-75.0</td>
<td>78.5±2.7</td>
<td>64.8±17.5/39.4-93.2</td>
</tr>
<tr>
<td>V</td>
<td>123.0±24.8/95.0-171.3</td>
<td>*</td>
<td>94.3±32.8/42.9-142.5</td>
</tr>
<tr>
<td>2700-3300</td>
<td>61.5±12.4/47.5-85.7</td>
<td>47.2±16.3/21.5-71.5</td>
<td>80.5±2.4</td>
</tr>
<tr>
<td>VI</td>
<td>150.5±40.1/93.8-225.0</td>
<td>*</td>
<td>162.8±44.1/105.0-228.3</td>
</tr>
<tr>
<td>2800-3100</td>
<td>75.3±20.0/46.9-112.5</td>
<td>81.4±22.0/52.5-114.2</td>
<td>78.8±3.1</td>
</tr>
<tr>
<td>VII</td>
<td>31.9±18.2/13.1-72.5</td>
<td>99.5±22.5/81.3-142.9</td>
<td>9.6±2.1/5.0-11.9</td>
</tr>
<tr>
<td>2450-3950</td>
<td>16.0±9.1/6.6-36.3</td>
<td>33.2±7.5/27.1-47.6</td>
<td>4.8±1.0/2.5-6.0</td>
</tr>
<tr>
<td>2-step down</td>
<td>81.7±2.2</td>
<td>77.3±3.9#</td>
<td>86.5±5.5**</td>
</tr>
</tbody>
</table>

*Subject unable to distinguish step from glide at durations less than 300 ms  
#Fewer than nine blocks of data  
**Floor effect
and 39.1±16.8 ms were not significantly different (predicted $t_{19,0.5}=2.09$; observed $t_{19,0.5}=0.73$). For the remaining signals in the set, both mean total signal duration and MTRT increased as the step number increased from two to three. The smoothing effect was evident from an observed increase in mean total duration of the signal ranging from nearly 40% to 70% (Table 6) to a total inability to distinguish some step and glide signals at the three-step level. Surprisingly, the mean total duration for the 460-690 Hz three-step upsweep was about 40% lower than the mean total duration for the two-step upsweep with the same center frequency. One would expect, based on the subject's performance on the other signals in the set, the three-step signal to be more temporally smoothed and, therefore, more difficult to distinguish from a glide.

As Table 6 shows, CI 1's MTRTs for the two-step upsweeps were between 1.3 and 4.8 times greater than those for the two-step downsweeps except for the very narrow 180-Hz sweep centered at 1950 Hz and the 300-Hz sweep centered at 2950 Hz. Both of these signals were within-electrode sweeps. The MTRT for the 180 Hz two-step upsweep was 58.0±9.4 ms and for the downsweep was 64.8±17.5 ms. A two-tailed t test showed that these MTRTs are not significantly different (predicted $t_{22,0.5}=2.07$; observed $t_{22,0.5}=1.19$). The MTRTs for the 300-Hz sweep centered at 2950 Hz were 75.3±20.0 ms for the upsweep and 81.4±22.0 ms for the downsweep. This difference was also not significant ($t_{22,0.5}=0.71$).

Under three-step conditions, there were three signals on which CI 1 was unable to distinguish the step from the glide for signal durations of less than 300 ms. These included the within-electrode sweep centered at 1950 Hz and
the two signals with CFs near 3000 Hz and with frequency sweeps of 600 Hz or less. If the extent of the sweep was extended to 1500 Hz, however, the MTRT for the three-step upswing (33.2±7.5 ms) was within the range of MTRTs for the two-step signals.

In Figure 11, CI 1’s data is presented so that upsweeps and downsweeps can be compared for each step number. The two-step upsweeps and downsweeps are shown in the top half of the figure and the three-step upsweeps in the bottom half. The only clear evidence for a frequency effect in any of the data collected from the cochlear-implant subjects was for CI 1. For both two-step and three-step upsweeps, his MTRT increased with increasing center frequency when the frequency sweep was 600 Hz or smaller.

Except for the 180-Hz signal, there did not appear to be a frequency effect in the two-step downsweeps for signals with CFs less than 2000 Hz (all MTRTs near or less than 10 ms). The 180-Hz signal, as noted above, was designed to fall within the frequency range of a single electrode. Restricting the sweep to a single electrode had the effect of producing an MTRT that was nearly six times greater (64.8±17.5 ms versus 10.4±6.0 ms) than the MTRT of a signal with a similar CF but with a frequency sweep of 600 Hz extending across three electrodes i.e., the 1700-2300 Hz signal (compare Signal IV to Signal III, Figure 11). There was a frequency effect observed when the CF of the two-step downsweep was increased from 2000 Hz to 3000 Hz (compare Signal III to Signal V, Figure 11). The MTRT of the 2700-3300 Hz signal (centered at 3000 Hz) was 47.2±16.3 ms, nearly five times greater than the MTRT (10.5±6.0 ms) of the 1700-2300 Hz two-step downsweep centered at 2000
Figure 11: Mean temporal resolution thresholds for subject CI 1. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Hz. Both of these signals had a frequency sweep of 600 Hz extending across three electrodes.

As occurred in the two-step downsweeps with CFs of less than 2000 Hz, the MTRT of signals with CFs near 3000 Hz increased when the extent of the sweep fell within a single electrode. The MTRT of the 2800-3100 Hz signal (CF of 2950 Hz and a within-electrode sweep extent of 300 Hz) was 1.7 times the MTRT of the 2700-3300 Hz signal with an across-electrode sweep extent of 600 Hz (81.4±22.0 ms versus 47.2±16.3 ms). In summary, in two-step downsweeps with CFs above 2000 Hz, both an increase in the CF and restriction of the sweep to a single electrode produced an increase in the MTRT. In two-step downsweeps with CFs below 2000 Hz, restriction of the sweep to a single electrode, but not an increase in the signal CF, produced an increase in the MTRT.

CI 1 could discriminate between both upswing and downswing two-step and glide signals of less than 300 ms duration at all center frequencies and sweep extents. A comparison of Figures 12 and 14 to Figure 10 shows that the other cochlear-implant users were unable to distinguish between the step and glide at all CFs and sweeps at the two-step level. The striking differences in temporal resolution abilities may reflect the differences in speech processors. CI 1 used the Spectra; both CI 2 and CI 3 used the MSP.

CI 3 had extremely poor temporal resolution abilities on most of the signals. However, as Figure 12 shows, she did have MTRTs of near 10 ms on the two-step and three-step downsweeps with a CF of 3400 Hz and a sweep extent of 1500 Hz. The MTRTs for these signals were 9.6±4.8 ms and 11.0±4.8 ms respectively and were not significantly different (observed $t_{19,0.5}=0.66$).
Figure 12: Mean temporal resolution thresholds for cochlear-implant user CI 3. I=350-550 Hz; II=920-1220 Hz; III=1600-2200 Hz; IV=1800-1980 Hz; V=2800-3400 Hz; VI=2850-3150 Hz; VII=2550-4050 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
It appears from Figure 13 that CI 3 was unable to distinguish between the three-step upsweep and glide with a frequency sweep of 350-550 Hz. Inspection of Table 7, however, shows a MTRT of 43.7±12.7 ms for this same signal. This apparent discrepancy reflects an rather unusual event that occurred during the testing of this subject. When CI 3 was first tested on the three-step 350-550 Hz upsweep, she could not consistently discriminate between the step and glide signals. On two successive sessions that occurred later in the testing period, however, six blocks of data were collected that consistently met the criteria of 75% to 85% correct responses with six to eight reversals in a run of 50 trials. The MTRT of these six blocks was 43.7±12.7 ms. In sessions after these two, the subject’s performance again became inconsistent, and she was generally unable to meet the criteria. Her MTRT for all blocks of data collected after the two sessions mentioned was 99.2±18.3 ms with a p(c) of 66.0±5.4%. In discussing this decline in performance with the subject, she mentioned that during the week in which data meeting the criteria was collected she was able to understand several conversations without depending on lip reading. Because there was a period during which CI 3 could discriminate between the two signals, this data was included in Table 7.

The only upsweep signals on which CI 3 could consistently distinguish a step from a glide signal were the two-step centered at 450 Hz and the two- and three-step signals centered at 1070 Hz (Figure 13). An analysis of the performance of CI 3 on the basis of the data discussed above shows that the mean total signal duration for the three-step signal centered at 450 Hz was 1.7 times that of the two-step signal at the same CF. However, the MTRTs
Figure 13: Mean temporal resolution thresholds for subject CI 3. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Table 7: Subject CI 3. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 350-550</td>
<td>75.4±28.3/41.3-131.3</td>
<td>130.9±38.1/76.3-173.8</td>
<td>*</td>
</tr>
<tr>
<td>II 920-1220</td>
<td>83.6±31.7/45.0-150.0</td>
<td>294.9±58.3/211.3-403.3</td>
<td>*</td>
</tr>
<tr>
<td>III 1600-2200</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV 1800-1980</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI 2850-3150</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII 2550-4050</td>
<td>*</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Subject unable to distinguish step from glide at durations less than 300 ms.

#Fewer than nine blocks
(43.7±12.7 ms and 37.7±14.2 ms, Table 7) for these two signals were not significantly different (predicted $t_{16,0.5}=2.12$; observed $t_{16,0.5}=0.9$). For this signal, therefore, CI 3 could compensate for the smoothing of the signal by increasing the total signal duration. When the CF was raised to 1070 Hz, however, both the mean total signal duration and the MTRT increased with increased step number. The MTRT for the three-step upsweep at CF of 1070 Hz, was more than double that of the two-step upsweep at the same CF (98.3±19.4 ms versus 41.8±15.9 ms) thus, increasing the step number at this CF made discrimination more difficult for the subject.

There was no effect due to increasing the CF of the upsweep signal at the two-step level until the CF was increased to 1900 Hz (Figure 13). The MTRT of 37.7±14.2 ms for the two-step upsweep centered at 450 Hz was not significantly different from the MTRT of 41.8±15.9 ms for the same signal centered at 1070 Hz ($t_{22,0.5}=0.66$). At CFs higher than 1900 Hz, the subject was unable to discriminate between step and glide when the signal was an upsweep. It was striking to see, however, that the signal having a CF of 1900 Hz and a frequency extent of 600 Hz was the only downsweep with an extent of less than 1500 Hz on which CI 3 could distinguish the step from the glide. The MTRT for this signal was 146.1±33.8 ms (Table 8). It appears from Table 7 that there was a frequency effect for the three-step upsweep conditions, but because the subject could consistently discriminate between the step and glide signal with CF of 450 Hz for only a short time during the testing period, one cannot state with confidence that such an effect existed.

The frequency extent of all of the signals on which CI 3 could distinguish the step and glide spanned at least three electrodes of the implant device.
Table 8: Subject CL3. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 350-550</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II 920-1220</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>III 1600-2200</td>
<td>292.1±67.5/190.0-397.9</td>
<td>*</td>
<td></td>
</tr>
<tr>
<td></td>
<td>146.1±33.8/95.0-199.0</td>
<td>81.7±2.8</td>
<td></td>
</tr>
<tr>
<td>IV 1800-1980</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI 2850-3150</td>
<td>*</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII 2550-4050</td>
<td>19.1±9.6/8.1-38.1</td>
<td>33.1±14.4/20.6-46.4</td>
<td>*</td>
</tr>
<tr>
<td></td>
<td>9.6±4.8/4.1-19.1</td>
<td>11.0±4.8/6.9-15.5</td>
<td>80.7±2.9</td>
</tr>
</tbody>
</table>

*Subject unable to distinguish a step from a glide signal at durations less than 300 ms.
The frequency extent of the signals on which she had the lowest MTRTs spanned the frequency ranges of at least four adjacent electrodes. It is interesting to note that CI 3’s MTRTs for the two-step upsweep signals with CFs of 450 Hz and 1070 Hz are not significantly different from the MTRTs for comparable signals of subject CI 1 ($t_{22,0.5}=1.79$ and 1.16 respectively) even though these subjects used different processors.

CI 3 often commented that she experienced a delay between the actual physical presentation of the auditory signal and her perception of the signal. For example, with signals in which she was sure she could hear the difference between the step and the glide, it would appear to her that the step signal occurred in interval three when the interval three box was illuminated, and she would chose interval three as the correct one. However, the feedback interval on the computer screen would tell her that the step had occurred in interval two. If the step really occurred in interval three, she would hear it when interval four was illuminated, an interval which was programmed to always contain a glide signal. As a result of this delay, the subject’s ability to perform the required psychoacoustic task was impaired. This phenomenon became a problem particularly when the signals were long in duration. All of the cochlear-implant subjects experienced this problem to some degree but none to the extent that CI 3 did.

As can be seen from Figure 14, cochlear-implant subject CI 2, unlike subjects CI 1 and CI 3, had no MTRTs near 10 ms. Her temporal resolution capabilities seemed to fall between those of CI 1 and CI 3. Overall, her MTRTs were higher than those of CI 1; on the other hand, she could discriminate between many more step and glide signals than could CI 3. It
Temporal resolution thresholds for 2-, 3-, and 5-step upsweeps and downsweeps for subject CI 2

Figure 14: Mean temporal resolution thresholds for cochlear-implant user CI 2. I=350-550 Hz; II=920-1210 Hz; III=1725-2325 Hz; IV=1920-2080 Hz; V=2800-3400 Hz; VI=3250-3550 Hz; VII=2630-3980 Hz. Asterisks indicate that subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
should be noted here that, as was the case with CI 3, there appears to be a discrepancy between data presented graphically and in tabular form. As shown in Figures 14 and 15, CI 2 was unable to distinguish between the step and the glide at the two-step level when the signal was an upsweep having a frequency range of 1725-2325 Hz. This signal was presented at an intensity level equal to one-half of the subject's dynamic range for that signal. When the subject was initially tested on this signal, she commented that she was perceiving a physical vibration rather than a sensation of sound. Further testing with this signal was immediately stopped. When CI 2 was tested at a later time using the two-step upsweep with a frequency range of 1920-2080 Hz at a level equal to one-half of the dynamic range, she did not complain of any vibratory sensation. However, the MTRT was near 150 ms. Considering that the level equal to one-half of the dynamic range was essentially the same for both the 1725-2325 Hz and the 1920-2080 Hz signals and that the frequency range of the 1920-2080 Hz signal fell within the frequency range of the 1725-2325 signal, it seemed probable that the subject might also be experiencing a vibratory sensation with this signal that interfered with her discrimination abilities. As shown at the bottom of Figure 16, when the level of the 1920-2080 Hz upsweep was lowered by 10 dB, the subject's MTRT fell by two-thirds, decreasing from 150.5±19.3 ms to 52.2±6.6 ms. The MTRT obtained at the lower presentation level is the value that is shown in Table 9. The bar graph at the bottom of Figure 16 also shows that when the direction of the sweep was down, the subject was unable to discriminate the step from the glide at either the level equal to one-half the dynamic range or at the reduced level.
Figure 15: Mean temporal resolution thresholds for subject CI 2. Asterisks indicate subject was unable to distinguish a step from a glide at signal durations shorter than 300 ms.
A comparison of the performance of subject CI 2 at signal levels perceived as a vibration and at reduced levels

- 2-step up @ 1/2 dynamic range
- 2-step up level 10 dB lower
- 2-step down level 10 dB lower

Signal with frequency sweep of 1725-2325 Hz (III)

Signal with frequency sweep of 1920-2080 Hz (IV)

Figure 16: Mean temporal resolution thresholds for subject CI 2 at signal levels that produced a sensation of vibration and at lowered levels that produced an auditory sensation. Asterisks indicate subject unable to distinguish a step from a glide at signal durations shorter than 300 ms.
Table 9: Subject CI 2. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 134.4±32.4/81.3-201.4</td>
<td>67.2±16.2/40.7-100.7</td>
<td>78.0±2.7</td>
<td></td>
</tr>
<tr>
<td>350-550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>II 92.3±14.0/62.5-114.3</td>
<td>46.2±7.0/31.5-57.2</td>
<td>79.2±2.0</td>
<td>159.7±46.8/125.0-267.1</td>
</tr>
<tr>
<td>920-1210</td>
<td></td>
<td>53.2±15.6/41.7-89.0</td>
<td>82.2±2.1</td>
</tr>
<tr>
<td>III 62.5±24.6/39.0-102.5</td>
<td>31.0±12.1/19.5-51.3</td>
<td>85.7±2.0#</td>
<td></td>
</tr>
<tr>
<td>1725-2325</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV 104.5±13.2/80.0-122.5</td>
<td>52.2±6.6/40.0-61.3</td>
<td>81.3±2.1</td>
<td></td>
</tr>
<tr>
<td>1920-2080</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>V 155.2±30.2/115.0-222.5</td>
<td>77.6±15.1/57.5-111.3</td>
<td>78.7±3.0</td>
<td>304.9±62.2/218.3-407.1</td>
</tr>
<tr>
<td>2800-3400</td>
<td></td>
<td>101.6±20.7/72.8-135.7</td>
<td>77.3±1.7</td>
</tr>
<tr>
<td>VI 219.6±65.5/162.5-352.5</td>
<td>109.8±32.7/81.3-176.3</td>
<td>79.8±2.3</td>
<td></td>
</tr>
<tr>
<td>3250-3550</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>VII 79.5±23.6/51.3-133.1</td>
<td>195.3±33.2/153.8-227.5</td>
<td>386.1±74.6/299.2-468.0</td>
<td></td>
</tr>
<tr>
<td>2630-3980</td>
<td>39.8±11.8/25.7-66.6</td>
<td>65.1±11.1/51.3-75.8</td>
<td>77.2±14.9/59.8-93.6</td>
</tr>
</tbody>
</table>

*Subject unable to distinguish step from glide at durations less than 300 ms.
#Fewer than nine blocks
It seemed reasonable to assume that the subject could distinguish the 1725-2325 Hz two-step signal from the glide if the level for that signal were also lowered. The bar graph in the top of Figure 16 proves this was a valid assumption. Lowering the level by approximately 10 dB resulted in a MTRT of 31.0±12.1 ms for this signal. According to the subject, lowering the level eliminated the sensation of vibration. It should be noted in Table 9 that the mean percent correct for the 1725-2325 Hz 2-step upsweep is higher than 85%. This signal was not tested at the lowered level until the end of the experiment. Because of a time constraint, adjustments in the initial signal duration that would drive the subject to 75%-85% correct performance were not made.

In order to avoid more problems with vibration, testing with the remaining signals having a frequency range of 1725-2325 Hz, the two-step downsweep and the three-upsweeps and downsweeps, was done at the reduced level.

If one takes into consideration the MTRT obtained using the reduced signal level for the 1725-2325 Hz signal, the data in Table 9 demonstrate that CI 2 could discriminate the step from the glide on every two-step upsweep. A comparison of the two-step data in Table 10 to that in Table 9 shows that the only two-step upsweep MTRT that was larger than the MTRT for the comparable two-step downsweep was for the signal centered at 450 Hz. The MTRT of the two-step downsweep centered at 1065 Hz was three times larger than that of the comparable upsweep (150.3±24.6 ms versus 46.2±7.0 ms). The MTRT of 49.6±11.7 ms for the downsweep centered at 2025 Hz was significantly higher than the MTRT of 31.0±12.1 ms for the same upsweep.
Table 10: Subject CI 2. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
</tr>
<tr>
<td>I 54.5±16.0/38.1-75.8</td>
<td>350-550</td>
<td>27.3±8.0/19.1-37.9</td>
</tr>
<tr>
<td>350-550</td>
<td>81.8±1.8</td>
<td>27.3±8.0/19.1-37.9</td>
</tr>
<tr>
<td>II 300.5±49.1/225.0-381.7</td>
<td>920-1210</td>
<td>150.3±24.6/112.5-190.9</td>
</tr>
<tr>
<td>920-1210</td>
<td>79.8±3.7</td>
<td>150.3±24.6/112.5-190.9</td>
</tr>
<tr>
<td>III 99.1±23.3/70.0-156.3</td>
<td>1725-2325</td>
<td>49.6±11.7/35.0-78.2</td>
</tr>
<tr>
<td>1725-2325</td>
<td>79.7±2.1</td>
<td>49.6±11.7/35.0-78.2</td>
</tr>
<tr>
<td>IV 1920-2080</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V 2800-3400</td>
<td></td>
<td></td>
</tr>
<tr>
<td>VI 287.5±30.9/248.6-358.8</td>
<td>3250-3550</td>
<td>143.8±15.5/124.3-179.4</td>
</tr>
<tr>
<td>3250-3550</td>
<td>81.0±2.5</td>
<td>143.8±15.5/124.3-179.4</td>
</tr>
<tr>
<td>VII 204.0±52.8/125.0-311.7</td>
<td>2630-3980</td>
<td>102.0±26.4/62.5-155.9</td>
</tr>
<tr>
<td>2630-3980</td>
<td>80.2±4.1</td>
<td>102.0±26.4/62.5-155.9</td>
</tr>
</tbody>
</table>

*Subject unable to distinguish step from glide at durations less than 300 ms.
signal \((t_{16,0.5}=3.11)\). The MTRT of the 3250-3550 Hz downsweep was also significantly higher than that of the upsweep \((143.8\pm15.5\text{ ms} \text{ versus} 109.8\pm32.7\text{ ms}; t_{22,0.5}=3.26)\). For the 1350-Hz sweep centered at 3305 Hz, the MTRT for the downsweep was 2.6 times larger than the MTRT for the upsweep. This pattern of higher MTRTs on the two-step downsweeps than on the upsweeps is in contrast to the performance of CI 1 who had lower MTRTs on the two-step downsweeps.

The only two-step signals on which CI 2 was unable to distinguish the step from the glide were downsweeps, the within-electrode sweep centered at 2000 Hz and the across-electrodes sweep centered at 3100 Hz.

There were only three three-step upsweep signals on which CI 2 could discriminate the step and the glide signals. There did not appear to be any pattern that related to which signals she was able to discriminate. The mean total duration for the three-step signal centered at 1065 Hz was 1.7 times greater than that of the two-step signal, but the MTRT \((\text{ms/step})\) of 53.2±15.6 ms for the three-step signal was not significantly different from the MTRT of 46.2±7.0 ms for the two-step signal \((t_{19,0.5}=1.25)\) indicating that, although the signal was temporally smoothed by the increase in step number, CI 2 could compensate for the smoothing simply by increasing the total duration of the signal. When the step number was increased to five, however, the subject was unable to do the task. She complained that the five-step signal centered at 1065 Hz had a static quality. For the 600-Hz sweep centered at 3100 Hz she had to increase the step duration by 25% in order to discriminate the step and the glide. An even greater increase of 40% in step duration was required when the step number was raised from two to three in the 1350-Hz signal.
centered at 3305 Hz. When the step number was increased to five for this same signal, however, the MTRT of 77.2±14.9 ms did not differ significantly (predicted $t_{11,0.5}=2.20$; observed $t_{11,0.5}=1.46$) from that of the three-step signal (65.1±11.1 ms). It must be pointed out, however, that the MTRT from the five-step signal was based on 200 trials whereas, that of the three-step signal was based on 300 trials. Note in Table 9 that the percent correct for the five-step upsweep was 85.5%. This indicates that the initial signal duration should have been lowered in order to make the task more difficult for the subject. CI 2 reported that she heard the steps in the step signal as a "ruffle." Unfortunately, there was not enough time to test the performance of the subject at shorter initial signal durations.

When the step number was increased to three in the downsweep signals, CI 2 was unable to distinguish the step from the glide regardless of signal CF or sweep extent. It should be noted, though, that the subject continued to show improvement in performance during the collection of the 12 blocks of data for the signal centered at 450 Hz. With continued practice, she may have been able to discriminate this three-step downsweep from the glide.

The limitation on temporal resolution abilities imposed by restricting the frequency sweep of the signal to one electrode that was observed in the performance of CI 1 was also seen in the performance of CI 2. Two of the upsweep signals and one downsweep signal on which CI 2 could discriminate the step and glide at the two-step level but not at the three-step level were within-electrode sweeps. These were the 160-Hz upsweep centered at 2000 Hz and the 300 Hz upsweep and downsweep centered at 3400 Hz. Confining the sweep of the 1920-2080 Hz downsweep to one electrode
prevented CI 2 from discriminating the step and the glide at even the two-step level.

As shown in Figure 14 and 15, there does not appear to be any consistency in the characteristics of the signals for which CI 2 had the lowest MTRTs. Her lowest MTRTs were 27.3±8.0 ms for the two-step 350-550 Hz downswep with a 300-Hz sweep, the two-step 1725-2325 Hz upswep with a 600-Hz sweep (31.0±12.1 ms), and the two-step 2630-3980 Hz upswep with a 1350 Hz sweep (39.8±11.8 ms). Neither do there appear to be any identifiable frequency effects.

The rather erratic performance of CI 2 may be a reflection of the time of day at which she was tested. Her testing sessions were at the end of her working day, and she often commented that she had difficulty concentrating. It was for this reason that her testing sessions were shortened to 60-90 minutes. Two signals, three-step upswep with CF of 2025, and five-step upswep with CF of 3100 were not tested because of time limitations.

4.3.2 Temporal resolution thresholds as determined from the step-versus-glide discrimination task: Listeners with normal hearing

As the data in Figures 17, 18 and 19 and Tables 11-16 show, there was little difference in the performance of the three subjects with normal hearing on the two-step-versus-glide discrimination task. In all but one instance, the MTRTs, averaged from 600 trials were less than 10 ms for all center frequencies, sweep extents and directions of sweep. These MTRTs are within or below the 7-10 ms MTRTs obtained by Madden and Feth (1992) using a
Temporal resolution thresholds for 2-, 3-, and 5-step upsweeps and downsweeps for subject NH1

Figure 17: Mean temporal resolution thresholds for normal-hearing listener NH1. I=460-690 Hz; II=920-1220 Hz; III=1700-2300 Hz; IV=1860-2040 Hz; V=2700-3300 Hz; VI=2800-3100 Hz; VII=2450-3950 Hz.
Temporal resolution thresholds for 2-, 3-, and 5-step upsweeps and downsweeps for subject NH2

Figure 18: Mean temporal resolution thresholds for normal-hearing listener NH2. I=350-550 Hz; II=920-1210 Hz; III=1725-2325 Hz; IV=1920-2080 Hz; V=2800-3400 Hz; VI=3250-3550 Hz; VII=2630-3980 Hz.
Temporal resolution thresholds for 2-, 3-, and 5-step upsweeps and downsweeps for subject NH3

Figure 19: Mean temporal resolution thresholds for normal-hearing listener NH3. I=350-550 Hz; II=920-1220 Hz; III=1600-2200 Hz; IV=1800-1980 Hz; V=2800-3400 Hz; VI=2850-3150 Hz; VII=2550-4050 Hz.
Table 11: Subject NH 1. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 460-690</td>
<td>17.2±5.1/8.1-27.5</td>
<td>36.3±15.1/18.1-63.8</td>
<td>115.5±33.0/73.8-155.0</td>
</tr>
<tr>
<td>II 920-1220</td>
<td>8.6±2.6±4.1-13.8</td>
<td>12.1±5.6/6.0-21.3</td>
<td>22.3±6.6/14.8-28.0</td>
</tr>
<tr>
<td>III 1700-2300</td>
<td>81.0±2.0</td>
<td>82.7±2.0</td>
<td>78.7±1.6#</td>
</tr>
<tr>
<td>IV 1860-2040</td>
<td>16.6±4.8/8.1-23.3</td>
<td>32.2±6.5/23.2-45.0</td>
<td>69.4/60.0-79.3</td>
</tr>
<tr>
<td>V 2700-3300</td>
<td>8.3±2.4/2.4-11.7</td>
<td>10.7±2.2/7.8-15.0</td>
<td>13.9/12.0-15.9</td>
</tr>
<tr>
<td>VI 2800-3100</td>
<td>80.8±2.2</td>
<td>81.8±2.1</td>
<td>78.7#</td>
</tr>
<tr>
<td>VII 2450-3950</td>
<td>16.0±3.1/10.6-21.3</td>
<td>25.2±2.3/21.9-28.1</td>
<td>56.2±15.9/33.8-70.0</td>
</tr>
<tr>
<td></td>
<td>8.0±1.6/5.3-10.7</td>
<td>8.4±0.8/7.3-9.4</td>
<td>11.3±3.2/7.9-14.0</td>
</tr>
<tr>
<td></td>
<td>81.8±2.5</td>
<td>80.9±1.8</td>
<td>79.7±3.2#</td>
</tr>
<tr>
<td></td>
<td>26.9±6.0/18.1-36.3</td>
<td>72.1±31.0/45.0-133.8</td>
<td>inconsistent</td>
</tr>
<tr>
<td></td>
<td>13.6±2.9/9.1-18.2</td>
<td>24.0±10.3/15.0-44.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>81.3±1.8</td>
<td>80.4±3.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>14.1±3.7/8.1-21.9</td>
<td>51.7±23.1/28.8-87.5</td>
<td>76.9±6.7/68.6-77.5</td>
</tr>
<tr>
<td></td>
<td>7.1±1.8/4.1-11.0</td>
<td>17.2±7.7/9.6-29.2</td>
<td>15.4±1.4/13.7-17.8</td>
</tr>
<tr>
<td></td>
<td>82.0±2.1</td>
<td>81.3±2.0</td>
<td>82.0±2.8#</td>
</tr>
<tr>
<td></td>
<td>17.6±2.2/13.8-22.5</td>
<td>36.5±11.4/27.9-65.0</td>
<td>inconsistent</td>
</tr>
<tr>
<td></td>
<td>8.8±1.1/6.9-11.3</td>
<td>12.1±3.8/9.3-21.7</td>
<td></td>
</tr>
<tr>
<td></td>
<td>81.2±2.8</td>
<td>82.0±2.2</td>
<td></td>
</tr>
<tr>
<td></td>
<td>11.9±1.9/8.8-15.0</td>
<td>18.4±5.0/13.1-27.5</td>
<td>53.1±10.6/40.6-67.5</td>
</tr>
<tr>
<td></td>
<td>6.0±1.0/4.4-7.5</td>
<td>6.1±1.7/4.4-9.2</td>
<td>10.6±2.1/8.1-13.5</td>
</tr>
<tr>
<td></td>
<td>84.2±2.3**</td>
<td>82.0±2.0</td>
<td>78.3±2.3#</td>
</tr>
</tbody>
</table>

**Floor effect #Fewer than nine blocks
Table 12: Subject NH 1. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downswep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean signal duration (ms±st. dev./range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean step duration (ms±st. dev./range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I 460-690</td>
<td>15.1±2.0/11.9-18.8</td>
<td>26.0±3.2/20.6-30.0</td>
<td>65.0/53.3-76.7</td>
</tr>
<tr>
<td></td>
<td>7.6±1.0/6.0-9.4</td>
<td>8.7±1.1/6.9-10.0</td>
<td>13.2/10.7-15.9</td>
</tr>
<tr>
<td></td>
<td>82.0±1.5</td>
<td>78.0±2.4</td>
<td>79.3#</td>
</tr>
<tr>
<td>II 920-1220</td>
<td>13.8±1.9/11.4-16.9</td>
<td>24.2±7.0/15.0-37.5</td>
<td>60.7/50.0-71.3</td>
</tr>
<tr>
<td></td>
<td>6.9±1.0/6.0-8.5</td>
<td>7.7±2.5/5.0-12.5</td>
<td>12.1/10.0-14.3</td>
</tr>
<tr>
<td></td>
<td>81.7±2.1</td>
<td>80.9±2.8</td>
<td>81.0#</td>
</tr>
<tr>
<td>III 1700-2300</td>
<td>9.9±1.9/8.1-12.9</td>
<td>18.3±4.8/10.6-27.1</td>
<td>52.9/35.0-68.8</td>
</tr>
<tr>
<td></td>
<td>5.0±1.0/4.1-6.5</td>
<td>6.1±1.6/3.5-9.0</td>
<td>10.6/7.0-13.8</td>
</tr>
<tr>
<td></td>
<td>87.3±2.3**</td>
<td>80.9±2.7</td>
<td>79.3#</td>
</tr>
<tr>
<td>IV 1860-2040</td>
<td>14.9±3.7/9.4-22.5</td>
<td>36.2±16.2/21.3-69.4</td>
<td>inconsistent</td>
</tr>
<tr>
<td></td>
<td>7.5±1.9/4.7-11.3</td>
<td>12.1±5.4/7.1-23.1</td>
<td></td>
</tr>
<tr>
<td></td>
<td>82.3±1.7</td>
<td>82.4±2.2</td>
<td></td>
</tr>
<tr>
<td>V 2700-3300</td>
<td>13.9±1.7/10.6-17.5</td>
<td>20.5±3.8/16.3-31.3</td>
<td>79.6±8.7/68.6-88.8</td>
</tr>
<tr>
<td></td>
<td>7.0±0.9/5.3-8.8</td>
<td>7.2±1.7/5.4-10.4</td>
<td>15.9±1.8/13.7-17.8</td>
</tr>
<tr>
<td></td>
<td>82.7±1.3</td>
<td>80.0±3.0</td>
<td>80.5±4.4#</td>
</tr>
<tr>
<td>VI 2800-3100</td>
<td>16.6±2.9/12.5-23.1</td>
<td>41.1±8.7/31.9-58.8</td>
<td>inconsistent</td>
</tr>
<tr>
<td></td>
<td>8.3±1.5/6.3-11.6</td>
<td>13.7±2.9/10.6-19.6</td>
<td></td>
</tr>
<tr>
<td></td>
<td>82.0±1.9</td>
<td>77.3±2.0</td>
<td></td>
</tr>
<tr>
<td>VII 2450-3950</td>
<td>8.7±4.4/5.0-18.8</td>
<td>13.1±4.3/8.8-20.6</td>
<td>38.5±12.0/26.3-55.0</td>
</tr>
<tr>
<td></td>
<td>4.4±2.2/2.5-9.4</td>
<td>4.4±1.5/2.9-6.9</td>
<td>7.7±2.4/5.3-11.0</td>
</tr>
<tr>
<td></td>
<td>91.7±6.0**</td>
<td>84.0±2.8**</td>
<td>83.0±2.0#</td>
</tr>
</tbody>
</table>

**Floor effect  #Fewer than nine blocks
Table 13: Subject NH 2. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 350-550</td>
<td>15.6±5.0/7.5-23.1</td>
<td>15.7±4.3/7.5-23.1</td>
<td>41.4±7.5/33.8-55.0</td>
</tr>
<tr>
<td>II 920-1210</td>
<td>13.4±1.5/11.3-17.5</td>
<td>22.2±5.4/14.4-30.6</td>
<td>39.9±10.1/16.4-51.9</td>
</tr>
<tr>
<td>III 1725-2325</td>
<td>8.7±0.2/8.5-9.1</td>
<td>5.2±1.0/3.8-6.5</td>
<td>81.2±1.3</td>
</tr>
<tr>
<td>IV 1920-2080</td>
<td>18.2±1.0/17.5-20.0</td>
<td>28.6±5.5/20.6-40.0</td>
<td>213.0±37.6/147.0-277.5</td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>8.7±1.2/6.0-10.7</td>
<td>10.9±3.2/5.6-14.0</td>
<td>10.7±3.4/6.7-16.7</td>
</tr>
<tr>
<td>VI 3250-3550</td>
<td>18.8±3.1/12.5-23.8</td>
<td>50.3±13.3/32.5-75.0</td>
<td>180.5±53.2/107.5-255.0</td>
</tr>
<tr>
<td>VII 2630-3980</td>
<td>13.4±2.3/9.4-17.5</td>
<td>19.6±2.8/15.6-24.4</td>
<td>35.3±4.4/28.8-40.6</td>
</tr>
<tr>
<td></td>
<td>81.5±1.2</td>
<td>81.8±1.2</td>
<td>80.2±2.9</td>
</tr>
</tbody>
</table>

#Fewer than nine blocks
Table 14: Subject NH 2. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean signal duration (ms±st. dev./range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mean step duration (ms±st. dev./range)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Percent correct</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I 350-550</td>
<td>13.3±1.5/10.0-15.0</td>
<td>19.5±4.4/13.3-26.3</td>
<td>47.0±15.7/22.9-66.3</td>
</tr>
<tr>
<td>6.7±0.7/5.0-7.5</td>
<td>6.5±1.5/4.4-8.8</td>
<td>9.4±3.1/4.6-13.3</td>
<td></td>
</tr>
<tr>
<td>81.3±1.3</td>
<td>80.9±1.5</td>
<td>78.0±3.1#</td>
<td></td>
</tr>
<tr>
<td>II 920-1210</td>
<td>9.7±1.7/7.5-12.5</td>
<td>13.7±2.1/11.3-18.1</td>
<td>32.5±5.8/24.4-41.3</td>
</tr>
<tr>
<td>4.9±0.8/3.8-6.3</td>
<td>4.6±0.7/3.8-6.0</td>
<td>6.5±1.2/4.9-8.3</td>
<td></td>
</tr>
<tr>
<td>83.0±1.4</td>
<td>81.6±0.9</td>
<td>79.3±2.6#</td>
<td></td>
</tr>
<tr>
<td>III 1725-2325</td>
<td>8.2±1.3/7.5-11.9</td>
<td>9.9±2.8/7.5-13.8</td>
<td>15.5±5.9/7.5-27.5</td>
</tr>
<tr>
<td>4.2±0.7/3.8-6.0</td>
<td>3.3±1.0/2.5-4.6</td>
<td>3.1±1.2/1.5-5.5</td>
<td></td>
</tr>
<tr>
<td>85.5±1.2**</td>
<td>84.7±2.5**</td>
<td>83.3±2.2**</td>
<td></td>
</tr>
<tr>
<td>IV 1920-2080</td>
<td>13.3±1.6/11.3-16.9</td>
<td>18.8±1.7/15.6-21.3</td>
<td>54.1±13.5/33.6-74.4</td>
</tr>
<tr>
<td>6.7±0.8/5.7-8.5</td>
<td>6.3±0.6/5.2-7.1</td>
<td>10.8±2.7/6.7-14.9</td>
<td></td>
</tr>
<tr>
<td>82.0±1.5</td>
<td>80.7±2.0</td>
<td>76.8±2.6#</td>
<td></td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>10.3±1.7/7.9-13.3</td>
<td>15.5±3.4/10.0-20.0</td>
<td>28.9±10.0/8.8-37.5</td>
</tr>
<tr>
<td>5.2±0.9/4.0-6.7</td>
<td>5.1±1.1/3.3-6.7</td>
<td>5.8±2.0/1.8-7.5</td>
<td></td>
</tr>
<tr>
<td>85.9±3.8**</td>
<td>82.4±1.3</td>
<td>79.1±3.4#</td>
<td></td>
</tr>
<tr>
<td>VI 3250-3550</td>
<td>17.1±0.9/15.0-18.1</td>
<td>34.1±6.7/27.5-50.0</td>
<td>53.8±7.3/42.5-62.5</td>
</tr>
<tr>
<td>8.6±0.4/7.5-9.1</td>
<td>11.4±2.2/9.2-16.7</td>
<td>10.8±1.5/8.5-12.5</td>
<td></td>
</tr>
<tr>
<td>81.8±1.8</td>
<td>80.0±2.5</td>
<td>79.2±2.3#</td>
<td></td>
</tr>
<tr>
<td>VII 2630-3980</td>
<td>8.3±0.3/8.0-8.8</td>
<td>8.8±1.7/7.5-12.9</td>
<td>7.8±0.5/7.5-8.8</td>
</tr>
<tr>
<td>4.2±0.1/4.0-4.4</td>
<td>2.9±0.6/2.5-4.3</td>
<td>1.6±0.1/1.5-1.7</td>
<td></td>
</tr>
<tr>
<td>91.0±2.2**</td>
<td>85.8±1.2**</td>
<td>85.1±2.3**</td>
<td></td>
</tr>
</tbody>
</table>

**Floor effect #Fewer than nine blocks
Table 15: Subject NH 3. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for upsweep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>I 350-550</td>
<td>13.3±2.6/7.5-17.5</td>
<td>20.2±2.1/16.9-23.8</td>
<td>28.5±9.7/14.4-46.3</td>
</tr>
<tr>
<td>II 920-1220</td>
<td>12.2±3.9/7.5-18.1</td>
<td>17.0±4.1/8.8-22.5</td>
<td>29.4±2.9/25.6-35.0</td>
</tr>
<tr>
<td>II 1600-2200</td>
<td>11.7±1.3/8.8-13.8</td>
<td>16.2±4.0/12.5-22.9</td>
<td>22.5±6.6/13.1-33.1</td>
</tr>
<tr>
<td>IV 1800-1980</td>
<td>18.6±3.2/13.1-25.6</td>
<td>30.8±4.3/25.0-36.9</td>
<td>128.1±25.6/92.1-170.0</td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>14.7±3.1/10.0-21.3</td>
<td>19.3±3.2/14.4-24.4</td>
<td>29.8±5.6/22.5-39.3</td>
</tr>
<tr>
<td>VI 2850-3150</td>
<td>18.1±4.0/11.3-23.1</td>
<td>28.4±4.6/20.7-37.5</td>
<td>69.4±35.6/31.3-127.1</td>
</tr>
<tr>
<td>VII 2550-4050</td>
<td>9.5±3.4/7.5-20.0</td>
<td>13.8±2.7/10.6-18.6</td>
<td>23.6±1.7/19.4-25.0</td>
</tr>
</tbody>
</table>

**Floor effect
Table 16: Subject NH 3. Mean total signal duration and range, mean step duration (=MTRT) and range, standard deviations, and average percent correct across all blocks of data collected for downswep signals.

<table>
<thead>
<tr>
<th>Signal (Hz)</th>
<th>2-step</th>
<th>3-step</th>
<th>5-step</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mean signal duration (ms±st. dev./range)</td>
<td>Mean step duration (ms±st. dev./range)</td>
<td>Percent correct</td>
</tr>
<tr>
<td>I 350-550</td>
<td>12.7±3.3/8.1-20.6</td>
<td>27.1±4.3/21.9-34.4</td>
<td>38.8±14.6/20.6-63.8</td>
</tr>
<tr>
<td>II 920-1220</td>
<td>6.3±1.7/4.1-10.3</td>
<td>9.0±1.4/7.3-11.5</td>
<td>7.8±2.9/4.1-12.8</td>
</tr>
<tr>
<td>IV 1800-1980</td>
<td>5.0±0.8/4.1-6.6</td>
<td>6.0±1.0/4.4-7.3</td>
<td>3.3±0.8/2.5-4.9</td>
</tr>
<tr>
<td>V 2800-3400</td>
<td>8.4±0.9/7.5-10.0</td>
<td>10.7±2.0/8.1-13.8</td>
<td>84.7±1.0**</td>
</tr>
<tr>
<td>VI 2850-3150</td>
<td>4.2±0.4/3.6-5.0</td>
<td>3.6±0.7/2.7-4.6</td>
<td>1.9±0.5/1.5-2.9</td>
</tr>
<tr>
<td>VII 2550-4050</td>
<td>8.0±1.5/7.5-11.9</td>
<td>9.1±1.4/7.3-11.5</td>
<td>91.5±2.1</td>
</tr>
</tbody>
</table>

**Floor effect
nonadaptive 2Q, 2 AFC step-versus-glide discrimination task. The slightly higher MTRTs obtained in the current study reflect the fact that the MTRTs in Madden and Feth's study were defined as 75% correct discrimination. In the our study, the performance of the subjects was at the level of 79.4% correct discrimination. The only instance of a two-step-versus-glide MTRT greater than 10 ms was subject NH 1's MTRT of 13.6±2.9 ms for the narrow 180-Hz sweep.

With two exceptions, subjects NH 2 and NH 3 also had MTRTs near, or less than, 10 ms for the three-step conditions. NH 2's MTRT for the downsweep signal with a CF of 3400 Hz and a sweep extent of 300 Hz was 11.4±2.2 ms, an increase of approximately 33% above the MTRT under two-step conditions. Her MTRT for the 3400-Hz-CF upsweep was 16.8±4.4 under three-step conditions, a 77% increase over the MTRT for the two-step upsweep (Figure 18).

Several of NH 1's MTRTs were higher than 10 ms when the step number was increased to three (Tables 11 and 12). There was a 76% increase in MTRT, from 13.6±2.9 ms to 24.0±10.3 ms, for the narrow 180-Hz upsweep centered at 1950 Hz (t_{19,0.5}=2.97). For the same downsweep signal, the increase in MTRT was 60%, from 7.5±1.9 ms to 12.1±5.4 ms (t_{19,0.5}=2.45). A 38% increase in MTRTs was observed when the step number was increased from two to three in the 300-Hz sweep centered at 2950 Hz under upsweep conditions (8.8±1.1 ms versus 12.1±3.8 ms; t_{19,0.5}=2.54). The increase in MTRT for the same signal under downsweep conditions was 65% (t_{19,0.5}=5.09).

A particularly large increase in MTRT was observed when the step number was increased from two to three for the upsweep signal centered at
Figure 20: Mean temporal resolution thresholds for subject NH1
3000 Hz with a sweep extent of 600 Hz (Figure 20). Under these conditions, the three-step MTRT was 2.4 times as great as that of the two-step MTRT, 17.7±7.7 ms and 7.1±1.8 ms respectively. The 12.1±5.6 ms MTRT for the three-step upsweep centered at 575 Hz was not significantly greater than the 8.6±2.6 ms MTRT for the same two-step signal ($t_{19,0.5}=1.75$).

When the step number was increased to five, the MTRTs of NH2 and NH3 generally remained near or below 10 ms. Both subjects, however, showed large increases in the MTRTs of the upsweep with the narrowest sweep extent (Figure 21 and 22). For Subject NH2, the five-step MTRT for the 160-Hz upsweep centered at 2000 Hz increased by more than 4 times, rising from 9.6±1.8 ms to 42.6±7.5 ms. When the signal was a downsweep, the increase in MTRT was only 70%. NH3's MTRT for the comparable five-step signal, 180-Hz upsweep centered at 1890 Hz, was approximately 2.5 times as great as the MTRT of the three-step signal, rising from 10.3±1.4 ms to 25.6±5.1 ms.

For the 300-Hz upsweep centered at 3400 Hz, NH2's MTRT more than doubled, rising from 16.8±4.4 ms to 36.1±10.7 ms, when the step number increased from three to five. While it appears that NH3 had a moderately large increase in the MTRT for the comparable signal, the difference between the MTRT of 13.9±7.1 ms for the five-step signal and 9.5±1.5 ms for the three-step signal was not statistically significant (predicted $t_{16,0.5}=2.12$; observed $t_{16,0.5}=1.8$). Both NH2 and NH3 commented that it was difficult to hold onto the cue for discriminating the step from the glide under five-step conditions.

NH1 also found it very difficult to hold onto the cue that he used for discriminating the step from the glide signal when the number of steps in
Figure 21: Mean temporal resolution thresholds for subject NH2
Figure 22: Mean temporal resolution thresholds for subject NH3.
the signal was five. Because data collection had already extended well beyond the time frame to which this subject had committed, six blocks of data for each of the five-step-upswEEP and downswEEP frequencies were collected. From these six, only those blocks that met the established criteria were used to calculate NH 1's MTRTs. In Tables 11 and 12, those five-step MTRTs for which standard deviations are not listed, were calculated from only two or three blocks of data (100-150 trials). MTRTs for five-step downswEEP signals V and VII were based on four blocks of data (400 trials); the remaining five-step upswEEP signals (I, III, V, VII) were based on six blocks of data (300 trials). The MTRTs which were based on fewer than six blocks are not included in Figure 17 and are plotted in a separate figure (Figure 23).

Unlike NH 2 and NH 3 who had MTRTs of near 10 ms on most of the five-step signals, NH 1 had MTRTs near 10 ms on only three five-step signals: the 1700-2300 Hz upswEEP and the 2450-3950 Hz upswEEP and downswEEP. MTRTs for these three signals were 11.3±3.2 ms, 10.6±2.1 ms, and 7.7±2.4 ms respectively. Because the MTRT of the downswEEP of 1700-2300 Hz was based on only 150 trials and the degree of variability was high, it is difficult to assign much significance to the value of 10.6 ms obtained for this MTRT.

By increasing the duration of the signal, there were some five-step signals on which NH 1 was able to compensate for the temporal smoothing which results from an increase in step number. The MTRT for the three-step 1700-2300 Hz signal (11.3±3.2 ms) was not really significantly different from the five-step MTRT of 8.4±0.8 ms (predicted \( t_{13,0.5}=2.16 \); observed \( t_{13,0.5}=2.18 \)).
Inconsistent responses of subject NH 1 to 5-step signal

Figure 23: Five-step signals on which subject NH 1 performed inconsistently. White bars represent thresholds of signals on which the subject could meet the criteria of 75-85% correct response and six-eight reversals in a 50-trial run. The addition of black bars indicates that criteria were met on fewer than six blocks of data. I=460-690 Hz; II=920-1220 Hz; III=1700-2300 Hz; IV=1860-2040 Hz; V=2700-3300 Hz; VI=2800-3100 Hz; VII=2450-3950 Hz.
The 2700-3300 Hz five-step upsweep was not significantly different from that for the comparable signal at the three-step level ($t_{13,0.5} = .69$). For all of the remainder of the five-step signals, both upsweeps and downsweeps, NH 1's five-step MTRTs were 1.5 to 2.2 times greater than the three-step MTRTs.

As was observed with subjects NH 2 and NH 3, the 180-Hz signal centered at 2050 Hz and the 300-Hz signal centered at 2950 Hz proved to be the most difficult for NH 1. In fact, when the signal frequency extent was 180 Hz, this subject could not meet the criteria on any blocks. Likewise, the subject met the criteria on only one block for the 2800-3100 Hz five-step upsweep. The same pattern of performance was observed with the five-step downsweep signals (Figure 23).

On every five-step downsweep signal, NH 1's MTRTs were higher than those of the other two control subjects. It is difficult to make comparisons between the normal-hearing subjects for the five-step downsweep signals. Some of the data from NH 1 was based on as few as 100-150 trials whereas data of NH 2 and NH 3 was based on 450 trials. As occurred under five-step upsweep conditions, NH 1's performance was so inconsistent on the 180-Hz downsweep signal and the 300-Hz downsweep signal centered at 2950 Hz, no data could be collected for these signals. For many of these five-step downsweep signals that proved to be difficult for NH 1, the average signal duration on some of the 50-trial blocks was 200-300 ms.

It is interesting to note that increasing the frequency sweep to 1350-1500 Hz produced low MTRTs for the normal-hearing subjects as it did for CI 1 and CI 3. All three control subjects showed floor effects when the sweep extent was unusually long. NH 2 and NH 3 were even able to consistently decrease the
duration of this signal to the 5-ms floor when the signal was a five-step
downsweep. NH 1 could approach the 5-ms floor only when the 1500-Hz
signal had two or three steps.

4.3.3 Perceptual asymmetry in normal-hearing listeners

There were several instances of asymmetry in the perception of upsweeps
and downsweeps among the control subjects. The most consistent
occurrence of asymmetry was for the 160-180 Hz signal centered near 2000 Hz.
For all three normal-hearing listeners, the MTRT for the upsweep was 50%
to 500% higher than the MTRT for the same signal presented as a
downsweep. For subjects NH 1 and NH 2, the asymmetry appeared as early
as the two- and the three-step conditions (Figures 17 and 18). In both of these
subjects, the difference in perception of upsweeps and downsweeps became
more pronounced as the number of steps in the signal increased. For subject
NH 3, asymmetry in the perception of upsweeps and downsweeps did not
appear until the the step number of the signals was increased to five.

4.4 Identification and discrimination of the elements
of the synthetic /ba/, /da/, /ga/ continuum

The identification function for listeners with normal hearing is shown in
Figure 24 and for listeners with cochlear implants in Figure 25. The
identification functions show that, generally, listeners with normal hearing
perceived the stimuli categorically, that is, they consistently labeled certain
stimuli. The first four stimuli of the continuum were generally heard as /ba/; stimuli 6, 7, 8, and 9 were heard as /da/; stimuli 11, 12, 13, and 14 were heard as /ga/. There were some discrepancies in labeling. As can be seen from the individual identification functions in Figure 24 subject NH 2 occasionally labeled stimulus 6 as /ga/ (10%) and subject NH 3 occasionally labeled both stimuli 5 (20%) and 6 (10%) as /ga/. Subject NH 1 labeled stimulus 9 as /ba/ 20% of the time and stimulus 11 as /ba/ 30% of the time. These small discrepancies in the labeling functions are probably a reflection of the small number of repetitions of each stimulus. When performance is averaged across all subjects with normal hearing, the identification function, as shown in Figure 26, has a much smoother appearance. This averaged identification function is based on 30 presentations of each stimulus (3 subjects x 10 presentations per stimulus).

The individual identification functions for each cochlear-implant subject, shown in Figure 25, illustrate that none of these subjects perceived the stimuli categorically. They also show that there was an extreme amount of between-subject variability in the labeling of the stimuli. (It should be noted that the MAP in CI 1's processor at the time he was tested with the synthetic speech stimuli was different from the MAP that he had for the psychoacoustic task.) Two trends among the cochlear-implant subjects should be noted. The first is that all of the subjects tended to identify stimulus 1 as /ba/. CI 1 identified stimulus 1 as /ba/ 90% of the time, and both CI 2 and CI 3 labeled 1 as /ba/ 60% of the time. CI 1 and CI 2 also identified stimulus 2 as /ba/ fairly consistently, 60% and 80% respectively.
Figure 24: Identification functions for listeners with normal hearing.
Figure 25: Identification functions for cochlear-implant subjects.
Figure 26: Average identification function for all listeners with normal hearing
The second trend was the tendency among the cochlear-implant subjects to classify stimuli 12, 13, and 14 more often as /da/ than as /ga/.

An assumption of categorical perception is that stimuli classified similarly are discriminated at chance levels, but stimuli not classified into distinct categories should be easily discriminable. The discrimination functions for each of the normal-hearing listeners are shown in Figure 27. Two distinct peaks appear in the discrimination functions of each of the listeners. Averaging discrimination performance across all of the normal-hearing listeners, results in a discrimination function with two peaks, one that falls at the 4/6, 6/4 stimuli pair and one at the 9/11, 11/9 stimuli pair. Examination of the average identification function (Figure 26) shows that the crossover points for the /ba/ and /da/ categories occurred between stimuli 5 and 6; the broadness of the first peak in the average discrimination function reflects this slight shift of the crossover point to between stimuli 5 and 6. The crossover point in the average identification function for /da/ and /ga/ categories fell at stimulus 10. The highest discrimination of stimuli, therefore, occurred at the points in the synthetic speech continuum where stimuli were least often labeled as similar. The points on the average discrimination function where discrimination performance was lowest, 2/4, 4/2; 7/9, 9/7; and 12/14, 14/12 corresponded to stimuli that were centered in the /ba/, /da/, and /ga/ categories in the average identification function shown in Figure 26. This result is as expected, because stimuli which are similarly labeled should be discriminated poorly.

Figure 29 shows that the discrimination performance of each cochlear-implant subject was different. Whereas CI 1 discriminated the stimuli at
Figure 27: Discrimination functions for listeners with normal hearing.
Figure 28: Average discrimination function for listeners with normal hearing (NH1, NH 2, and NH 3).
Figure 29: Discrimination functions for cochlear-implant listeners. Dotted line indicates level of chance performance.
levels slightly above chance, subject CI 3 discriminated at levels below chance. Subject CI 2 discriminated the stimuli at, or nearly at, chance levels.

4.5 Calculation of $d'$ as a measure of discrimination ability

The $d'$ in an AX task is an unbiased measure of the ability of a subject to discriminate between pairs that are the same and pairs that are different. Stimuli which are classified in the same category should be indiscriminable, and $d'$ for discrimination of such stimuli should equal zero. On the other hand, pairs of stimuli along a continuum of stimuli which fall at the crossover points in the identification function should be highly discriminable, and the $d'$ for discrimination of these stimuli should be high.

The $d'$ for the AX discrimination of the elements of the synthetic speech continuum by both normal-hearing subjects and cochlear-implant subjects is shown in Table 17. The calculation of $d'$ was based on the presentation of 240 pairs of "same" and 240 pairs of "different" stimuli and was done according to the formula:

$$z(H)-z(F) = d' \quad \text{(Macmillan and Creelman, 1991)}$$

in which:  
$z$=z score
$H$=hit rate: calculated as the percentage of pairs correctly identified as "different"

$$(S_1S_2/S_2S_1)$$
F = false alarm rate: calculated as 1 - the percentage of pairs correctly identified as "same"
\[(S_1S_1/S_2S_2)\]

Adjusted d’s were included in Table 17 because a same/different task is more difficult than a yes/no task (Macmillan and Creelman, 1991). Adjusted d’s take the increased demands on the subject into account. Table 17 shows that, with d’s near zero, all the cochlear-implant listeners had a marked lack of sensitivity to the difference between elements of two-step pairs taken from the synthetic speech continuum. CI 3, in fact, had a negative d’. This is characteristic of what is known as a "perverse observer," that is, an observer who says "yes" when she knows the answer is "no." As was seen in Figure 29, all points on the discrimination function for this subject were below the level of chance. Macmillan and Creelman (1991) point out that statistical fluctuations can easily lead to below-chance performance. As the results of the rest of this study have indicated, a high degree of variability in performance was characteristic of cochlear-implant subjects.

The average d’ for all three listeners with normal hearing was 1.75. A d’ that is this high indicates a sensitivity to the difference between pairs composed of elements that are the same and pairs composed of elements that are different that is well above chance. As can be seen from the data in the Table 17, normal-hearing controls NH 1 and NH 2 both exhibited a strong response bias toward saying "same" Bias is zero when the false alarm rate (F) and the miss rate (1-H) are equal.
Table 17: $d'$ as an index of the sensitivity of listeners with cochlear implants and listeners with normal hearing to the difference between two-step pairs of elements in a synthetic speech continuum. Adjusted $d'$ for same-different design from Table A5.3 in Macmillan and Creelman (1991).

<table>
<thead>
<tr>
<th>Subject</th>
<th>H</th>
<th>F</th>
<th>calculated $d'$</th>
<th>$d'$ for same-different design</th>
</tr>
</thead>
<tbody>
<tr>
<td>CI 1</td>
<td>.60</td>
<td>.60</td>
<td>0</td>
<td>0.10</td>
</tr>
<tr>
<td>CI 2</td>
<td>.49</td>
<td>.47</td>
<td>0.05</td>
<td>0.36</td>
</tr>
<tr>
<td>CI 3</td>
<td>.40</td>
<td>.43</td>
<td>-0.077</td>
<td>-0.44</td>
</tr>
<tr>
<td>NH 1</td>
<td>.26</td>
<td>.05</td>
<td>1.002</td>
<td>1.75</td>
</tr>
<tr>
<td>NH 2</td>
<td>.40</td>
<td>.05</td>
<td>1.392</td>
<td>2.14</td>
</tr>
<tr>
<td>NH 3</td>
<td>.63</td>
<td>.25</td>
<td>1.006</td>
<td>1.76</td>
</tr>
</tbody>
</table>

Average $d'$ for all normal subjects $\text{NH 1 + NH 2 + NH 3}$

$\begin{array}{cccc}
.43 & .12 & 0.999 & 1.75 \\
\end{array}$
The results of this study show that under some conditions listeners with cochlear implants have temporal resolution abilities that are quite comparable to those of listeners with normal hearing, yet under other conditions, their temporal resolution is impaired. It has also been shown that, under certain conditions, temporal resolution is limited even in listeners with normal hearing. In this section, an attempt will be made to relate these differences in temporal resolution abilities to mechanical and physiological factors in the auditory system and to limitations imposed by the speech processors used by the cochlear-implant users.

5.1 Differences in temporal resolution abilities between listeners with normal hearing and those with cochlear implants: narrow-sweep signals

The most unanticipated observations from this study were for those signals with 160- to 180-Hz sweeps centered near 2000 Hz and signals with 300-Hz sweeps centered near 3000 Hz. Under some conditions, the MTRTs of
the normal-hearing listeners for these signals were greater than 10 ms. These elevated MTRTs first appeared at the two-step level for NH 1, the three-step level for NH 2, and at the five-step level for NH 3 (Tables 11, 13, 15). These signals were designed to be within-electrode sweeps for the matched cochlear-implant subjects, so the most obvious explanation for the elevated thresholds in the normal-hearing subjects was that they stemmed from some relationship between the extent of the sweep and equivalent rectangular bandwidth (ERB) of the auditory filter at the CF of the sweep.

According to the equation described by Glasberg and Moore (1990):

\[ \text{ERB} = 24.7(4.37F + 1) \]

where \( F \) is frequency in kHz

the ERB for a signal frequency of 2000 Hz, is approximately 240 Hz; for a signal frequency of 3000 Hz, the ERB is approximately 350 Hz. It appears, therefore, that signals designed to be within-electrode sweeps for the cochlear-implant subjects were, serendipitously, signals whose sweeps were smaller than the ERBs of the auditory filters at the CF of the sweep for listeners with normal hearing (hereafter, these signals will be referred to as smaller-than-ERB signals). The results from the study suggest that the perception of signals that change in frequency over time by listeners with normal hearing is dependent to some extent upon the relationship between the width of the sweep of the signal and the ERB of the auditory filter associated with that signal's CF.
The truly surprising observation in this study was that the cochlear-implant subjects were able, in some cases, to distinguish the step from the glide when the frequency sweep was a within-electrode sweep. Intuitively, it does not seem that this should be possible. The explanation of their ability to do so may lie in their ability to use off-frequency energy generated by the instantaneous jumps in frequency that occurred in the step signals. The listeners may have been able to use this spectral splatter as a cue for discriminating the two signals. No masking was used in this study, and the step signals were not modified in any way to eliminate the generation of off-frequency energy. The question of whether cochlear-implant listeners are using off-frequency energy cues might be answered by repeating the step/glide discrimination task with signals presented in a notched-noise masker.

It is not likely that the normal-hearing listeners were using off-frequency energy cues in distinguishing the step and the glide signals. If they had been using the splatter as an additional cue, one would expect that the MTRTs observed in this study would be lower than those observed in the studies of Madden (1990), and Madden and Feth (1992) in which the step signals were generated with rounded corners in order to eliminate off-frequency energy. The thresholds in the current study were not lower than those observed in the Madden and Madden and Feth studies.

The use of dynamic signals having smaller-than-ERB sweeps may explain why the MTRTs for the signals centered at 4 kHz in Feth et al.’s 1989 study of the discrimination of step and glide signals were twice as high as those for signals centered on 250 Hz, 500 Hz, 1 kHz, and 2 kHz. The maximum
frequency transition for these signals was 400 Hz. This sweep would have been smaller than the 456-Hz ERB of the auditory filter at a CF of 4 kHz (Glasberg and Moore, 1990). It may also explain why the subjects in Madden and Feth's study (1992) were unable to achieve a p(c) of 75% when the CF of the signal was 4 kHz. The sweep extent of the signals used in this study was 200 Hz, again smaller than the ERB of the auditory filter associated with a CF of 4 kHz.

5.2 Differences in temporal resolution abilities between listeners with normal hearing and those with cochlear implants: broad-sweep signals

It is not possible to make any sweeping generalizations about the temporal resolution abilities of listeners with cochlear implants compared to those of listeners with normal hearing. As noted above, although the cochlear-implant subjects generally had higher MTRTs than those of their matched controls for a given set of signals, there were certain conditions under which the temporal resolution abilities of the implanted subjects were comparable to those of the control listeners. For example, CI 3, whose abilities to distinguish the step and the glide signals were extremely limited, had MTRTs for the 1500-Hz-sweep signal that were within the range of the MTRTs of the normal-hearing listeners. It is not even possible to make sweeping statements about temporal resolution abilities within the population of cochlear-implant users. A comparison of the MTRTs for the two-step downsweep centered at 450 Hz or 575 Hz shows that CI 3 was completely
unable to distinguish the step from the glide signal even when the total
duration of the signal was 300 ms (Figure 9); the MTRT for CI 2 for the same
signal was approximately three times that of the normal listeners (Figure 7),
and the MTRT of CI 1 for this signal was comparable to that of listeners with
normal hearing (Figure 5). It is clear, therefore, that an identification of the
specific limitations on the temporal resolution abilities of users of the
implant device must be made.

These limitations could arise from two sources, one from the existing
integrity of the inner ear and the VIIIth nerve of the individual listener and
the other from the device itself. At the present time, we have no way of
establishing, with any degree of certainty, the relationship between the
amount of damage in the cochlea and the amount of impairment in
perceptual processes. Limitations imposed by the implant device itself are
more easily assessed. In fact, an analysis of the discrimination of step-versus-
glide data from this study indicates that this appears to be a source of reduced
temporal resolution abilities of some of the implant subjects. In most
instances, the cochlear-implant subjects had significantly higher MTRTs on
the three-step signals as compared to the two-step; in some instances, they
were completely unable to discriminate the step from the glide under three-
step conditions. In two cases (Signal II, Table 6; Signal VII, Table 9), the three-
step MTRTs were higher than those under two-step conditions but not
significantly so because of the high degree of variability. In contrast, subjects
with normal hearing were unaffected by temporal smoothing when the step
number of the signal was increased from two to three. Those instances in
which the MTRTs of the normal-hearing listeners were greater than 10 ms
for the three-step signal were nearly always associated with sweeps smaller than the ERBs of the auditory filters associated with the CF of the signal.

An analysis of the electrode stimulation pattern under the two-step and three-step conditions shows that the source of apparent impaired temporal resolution abilities in the cochlear-implant subjects lies in the manner in which information about the signal is presented by the speech processor to the electrode array lying within the cochlea. Each electrode of the array has assigned to it a specific band of frequencies. Because the spectral resolution capabilities of the listener with an intact basilar membrane are lacking in the implanted subject, all frequencies within the assigned band are perceptually "the same." The glide signal, therefore, is represented in the implant device in "frequency chunks" across the electrode sweep rather than as a nearly continuous frequency-to-frequency glide.

The broad sweep signals used in this study were designed so that the frequency extent of the glides crossed at least three electrodes in the electrode array of the implanted device. The step signal nominally swept across the same electrodes as did the glide. In actuality, however, under some conditions, these step signals did not sweep the same number of electrodes. For example, the 920-1220 Hz upswing glide (Signal II) presented to CI 1 presumptively stimulated electrodes 15, 14, and 13. The two-step signal with the same frequency sweep, however, stimulated only electrode 15 (920 Hz) for the first step, then made an instantaneous jump midway through the signal to 1220 Hz to generate the second step. This frequency fell within the range of electrode 13. This electrode was stimulated for the remaining duration of the signal. The subject, therefore, was actually discriminating between a
signal that stimulated three electrodes and a signal that stimulated two electrodes. All of the two-step signals, with the exception of the within-electrode-sweeps and the 1350-1500 Hz-sweep signals, followed this pattern.

A very different situation arose with the three-step signals. Using the same example as above, the 920-1220 Hz upswing (Signal II), the glide stimulated electrodes 15, 14, 13. The first step of the three-step signal (920 Hz) stimulated electrode 15; the frequency changed instantaneously one-third of the way through the signal to 1070 Hz, resulting in the stimulation of electrode 14. The final instantaneous frequency change from 1070 Hz to 1220 Hz, that is the third step, stimulated electrode 13. For CI 1, then, under the three-step condition the 920- to 1220-Hz signal stimulated the same number of electrodes for the glide and for the step signal. Unlike the situation described above for the two-step signal, the subject was now attempting to discriminate between two signals that both stimulated the same three electrodes.

In the normal-hearing listener, the rate of temporal smoothing is slow. Because the auditory filters of the basilar membrane overlap and because spectral resolution is sharp, greater smoothing of the signal is required before the glide and step signal begin to "sound" alike to the listener. Because frequencies within the frequency ranges of single electrodes in the implant device are smeared, the rate of temporal smoothing is very fast for the cochlear-implant listener. The results of this study showed that temporal smoothing did occur more quickly in the cochlear-implant users as their MTRTs for the three-step signals were higher than those of the two-step signals. In the normal-hearing listeners, except for the signals with smaller-
than-ERB sweeps, the MTRTs for the two-step and the three-step signals were less than 10 ms. The effect of temporal smoothing wasn't seen until the step number increased to five.

The situation for the 1500-Hz-sweep signal (Signal VII) for CI 1 was a bit different. The upswing glide swept four electrodes, presumably electrodes 8, 7, 6, and 5. Under the two-step condition, the frequency maintained throughout the first step stimulated electrode 8, and the frequency throughout the final step stimulated electrode 5. Only two electrodes were stimulated in the step signal as compared to four electrodes in the glide. Under three-step conditions, electrodes 8, 6, and 5 were stimulated in response to the steps of the signal. Here, three electrodes were stimulated in the step signal compared to four in the glide. The elevated MTRT for the 1500-Hz three-step signal compared to that of the two-step signal may simply reflect the fact that the subject could more easily discriminate between a signal that stimulated four electrodes and one that stimulated two than he could discriminate between a signal that stimulated four electrodes and one that stimulated three. In fact, this was exactly the observation that Tong et al. (1982) made in their study of discrimination of time-varying electrode position.

Increasing the step number from two to three also resulted in temporal smoothing of the signal for subjects CI 2 and CI 3 in the same manner as described for CI 1 above. For the extended frequency sweep (Signal VII) of 1500 Hz for CI 3 and 1350 Hz for CI 2, the ratio of the number of electrodes stimulated by the glide to the number of electrodes stimulated by the step signal, under two-step conditions was 4:2. Under three-step conditions, it was
4:3, and under five-step conditions, it was 4:4. Thus, increasing step number for Signal VII for CI 3 (Table 8) and CI 2 (Table 9) also produced a temporal smoothing that actually was an artifact of a lack of contrast in the electrodes stimulated by the glide and the step signal.

Temporal smoothing was observed for CI 2 and CI 3 when the step number was increased from three to five in the 920-1210/1220-Hz upswEEP (Table 7 and Table 9). For this sweep, electrodes 13, 12, and 11 were stimulated in CI 3 and electrodes 14, 13 and 12 for CI 2 under both three-step and five-step conditions. One would anticipate that if identical patterns of electrode stimulation for the glide and the step signal were the only explanation for the temporal smoothing observed in the three-step conditions, then the subjects should have had MTRTs for the three-step and the five-step conditions that were very similar as the same electrodes were stimulated under each condition. Instead, under five-step conditions, neither subject could distinguish the step from the glide.

Listeners with normal hearing generally could compensate for temporal smoothing in the step signal by increasing the total duration of the signal so that the ms per step, the MTRT, remained at the same level. While this mechanism was effective for signals having sweeps larger than the ERBs for the auditory filters at the signal's CF, it did not always result in compensation for temporal smoothing caused by increasing the step number in signals with smaller-than-ERB sweeps. In some instances, the subject had to increase the total duration of the signal to such an extent that the ms per step also increased (Table 11, Signals IV, VI; Table 12, Signals IV, VI; Table 13, Signal IV, VI; Table 15, Signal IV).
As noted above, listeners with cochlear implants frequently were unable to discriminate between the step and glide signals when the step number was increased from two to three or from three to five. However, in some instances they were able to employ the same mechanisms used by the normal-hearing subjects to compensate for temporal smoothing. CI 1 used the mechanism of increasing the total signal duration so that the MTRT remained unchanged for the signal centered at 1070 Hz. CI 2 also used this mechanism for that signal. CI 3 applied this mechanism to compensate for temporal smoothing in the 1500-Hz downsweep. The compensation mechanism of increasing both the total signal duration and the ms per step was also used by all three cochlear-implant subjects (Table 6, Signal III, VII; Table 7, Signals I and II; Table 9, Signal V, VII). The ability to use the same mechanisms to compensate for increased temporal smoothing in a signal would appear to support the notion that the innate temporal resolution capabilities of the cochlear-implant subjects are intact.

It can be concluded from these results that temporal resolution in the cochlear-implant user is, in part, a function of contrast in the patterns of electrode stimulation produced by the representation of the dynamic acoustic signal. Thus, limitations of temporal resolution abilities are a function of that representation. At first glance, it might appear that the results demonstrated that the increase in the actual number of electrodes stimulated in response to an acoustic signal with an extended frequency sweep resulted in better temporal resolution. A closer examination of the data shows, however, that it was not the long sweep alone that was important but that the long sweep resulted in stimulation of more electrodes and coincidentally
resulted in a greater contrast in the electrode stimulation pattern of the two-step and the glide signals.

Temporal resolution, particularly in regard to the tracking of frequency changes which occur in a dynamic signal, needs to be viewed from a different perspective in listeners with a cochlear implant than it does in listeners with normal hearing. In a study of formant transitions in CNC words, Lehiste and Peterson (1961) showed that for the bilabial stop consonant /b/, the average F2 onset in the presence of the vowel /i/ was 1780 Hz. For the dental-alveolar consonant /d/, the average onset in the presence of /i/ was 1785 Hz. The average F2 formant position for /i/ was 2205 Hz. In the cochlear implant devices used by the subjects in our study, the pattern of electrode stimulation representing these 1780-2205-Hz and 1785-2205-Hz transitions would be same. Based on the results of our study, because there would be no contrast in the pattern of electrode stimulation, we would predict that these two frequency-modulated signals would be indistinguishable to the listener. This lack of contrast in the representation of the important second formant may explain why cochlear-implant listeners have difficulty distinguishing consonants on the basis of place of production. It is possible that the F2 formants of natural speech produce the same or very similar patterns of electrode stimulation in the implant device. It was observed in the current study that as the number of steps in the signals increased, the pattern of contrast in the electrodes stimulated decreased, and, as a result, the MTRTs rose.

In thinking of how rapid changes in frequency in acoustic signals can be better represented, we need to think not in terms of how much information
about each frequency transition we can impart to the cochlear-implant listener but should think in terms of how much we can contrast the information present in a /b/ F2 or a /d/F2. For the cochlear-implant user, representation of the maximum information in speech may not be the answer to speech perception unaided by lipreading. Better representation may be achieved by finding ways to maximally contrast the cues used in perception of speech.

As the data show the importance of representing information in a signal by means of the contrast in the pattern of electrode stimulation, it seems rather remarkable that the cochlear-implant subjects were able to distinguish between step and glide signals crossing the same electrodes. When there was no contrast, it is possible that the implanted subjects were using cues other than temporal ones as help in distinguishing between the step and glide. Perhaps they were using off-frequency energy as a cue to the difference between the two signals. As will be discussed in more detail later, this may be the reason that CI 1 was more successful in discriminating the step from the glide signals than were CI 2 and CI 3. He may have obtained better resolution of high frequency information with the Spectra.

5.3 The effect of vibratory sensation on perception

As noted in the results, CI 2 stated that she perceived a sensation of vibration rather than of sound for one of the signals and, consequently, was unable to distinguish the step from the glide signal. When the intensity level of that signal was lowered, she was then successful in discriminating
the two signals. Based on the performance of CI 1 and CI 3 with the two-step signal with an extended frequency sweep, one might have anticipated that CI 2 would have a very low MTRT for this same signal. However, this was not the case. Recall that she experienced the vibration problem with Signal III whose frequency sweep of 1725-2325 Hz resulted in stimulation of electrodes 8, 7, and 6. In the MSP with the MPEAK strategy, if the peak frequency detected by the F2 filter (filters frequencies between 1000 and 4000 Hz) is greater than 2.8 kHz, then the electrode designated as Band 3 is also stimulated in every cycle. The electrode corresponding to Band 3 is, by default, electrode 7. The frequency range of the long-sweep signal was from 2630-3980 Hz; thus, electrode 7 also would have been stimulated during a substantial proportion of the total number of stimulation cycles which occurred during that sweep. If electrode 7 were a source of the vibration in Signal III, then it is possible the vibration from that electrode, even though it was unnoticed by the subject, also interfered with the perception of the 1350-Hz-sweep signal. This interference might account for the fact that the lowest MTRT obtained from CI 2 was not for the extended-sweep signal as it was for the other two subjects.

5.4 Perceptual asymmetry in listeners using the cochlear implant device

Asymmetry in the perception of upsweep and downsweep signals was strongly apparent in only one of the three cochlear-implant subjects, CI 1. It is possible that one could attribute the asymmetry in the perception of
upsweeps and downsweeps observed in CI 1 to practice effects. Because we wished to maximize the performance of the cochlear-implant subjects, we chose to present only one type of signal at a time; that is, the subjects were tested with the entire set of two-step upsweeps prior to testing with the set of two-step downsweeps. In retrospect, a better design would have been to randomly present upsweep and downsweep signals of a given step number. An argument in support of the observation that the differences in upsweep and downsweep thresholds were real is based on the performance of the other two cochlear-implant subjects. CI 3 had lower thresholds for the two-step low-frequency upsweeps (Figure 13, Signals I & II), and CI 2 had lower thresholds for four of the seven two-step upsweeps. Because all three listeners were naive, and all three had the same method of presentation of the complete set of two-step upsweeps followed by the complete set of two-step downsweeps, we conclude that the higher thresholds that CI 1 had for the two-step upsweeps were not due to practice effects.

It is possible that the lower thresholds observed for across-electrode two-step downsweeps in Subject CI 1 and the striking difference between the two-step 1500-Hz upsweep and downsweep for subject CI 3 related to the flow of current in the cochlea relative to the direction of the signal sweep. Kasper, Pelizzone, & Montandon (1991) and Lim, Tong, & Clark (1989) have shown that the major current flow with electrical stimulation of the cochlea is toward the base of the cochlea. Although the BP + 1 mode for the active/ground arrangement of electrodes which was used by all the cochlear-implant subjects in this study minimizes current flow to other electrodes, there is no claim that it eliminates current flow. The pattern of electrode
stimulation for upsweep signals was from apex to base, that is from low to high frequency ranges. This pattern of electrode stimulation will be referred to as isocurrent; that is, it is in the same direction as the flow of current in the cochlea. For the purpose of illustration, assume an upsweep signal stimulates electrodes 18, 17, and 16. If electrode #18 is stimulated first (#18 is a low-frequency electrode), some current may flow toward electrode #17. When electrode #17 is stimulated next in the sequence, the electrical stimulus at that electrode may be masked by current flowing basally from the more apical electrode (Lim et al., 1989). The electrical stimulus at electrode #16 may be masked by current flow from electrode #17. For isocurrent signals, current flow is toward the next electrode to be stimulated.

Downsweep signals, on the other hand, are countercurrent signals; that is, the direction of electrode stimulation is opposite to the flow of current in the cochlea. Again, for purposes of illustration, suppose a downsweep signal stimulates high-frequency electrode #3 followed by the successively more apical, lower-frequency electrodes, #4 and #5. The direction of stimulation is opposite to the flow of current in the cochlea; masking of each higher number electrode by current flow from the preceding electrode is, therefore, not a factor. For countercurrent signals, current flow is away from the next electrode to be stimulated.

Further support for the hypothesis that masking is a factor in asymmetrical perception of upsweeps and downsweeps is provided by MTRTs for the within-electrode signals obtained from CI 1. The statistical analysis of these signals showed that the upsweep and downsweep thresholds for within-electrode sweeps were not significantly different.
Because the frequency sweep of the signal stimulates only one electrode, there would be no masking related to the direction of electrode stimulation.

As there was no consistent pattern to the performance of CI 2 and performance in CI 3 was limited, it is somewhat difficult to interpret instances of perceptual asymmetry in these two subjects. Like CI 1, CI 2 had lower thresholds for the downsweeps on signal I with CF of 450 Hz and on signal III with CF of 2025 Hz. On the remainder of the signals, however, the downsweep thresholds were higher than the upsweep thresholds (Figure 15, two-step signals II, V, VI, VII; three-step signals II and VII). Also in contrast to subject CI 1, CI 3 had higher downsweep thresholds for the two signals with the lowest CFs (Figure 13, Signals I and II).

5.5 Perceptual asymmetry in listeners with normal hearing

In all instances of perceptual asymmetry in normal-hearing listeners, the MTRT for the upsweep signal was greater than that for the downsweep. The increases observed in MTRTs for the upsweeps ranged from 50% to 500%. Whereas masking in CI 1 was induced by the direction of current flow relative to the direction of electrode stimulation, in the listeners with normal hearing masking may have been induced by the action of the traveling wave as it moved along the basilar membrane of the cochlea. In response to an upsweep, a signal that moves from low to high frequency, the traveling wave induced by the low frequency end of the stimulus moves towards the apex of the basilar membrane and in doing so passes through
regions of the membrane associated with high frequency. It is well known that low frequencies are effective maskers of high frequencies. The low-frequency-response sides of tuning curves are broad; hence, high-CF hair cells and neurons located at the base of the basilar membrane respond to a broad range of frequencies below their CF.

The results also showed that perceptual asymmetry rarely occurred in signals with sweeps wider than the ERB of the auditory filter associated with the signal CF and that the masking effect was most consistently seen when the sweep of the signal was smaller than the ERB of the auditory filter (Figures 17, 18, and 19). It appears that when the temporal resolution abilities of the listener are already taxed by a signal with a smaller-than-ERB sweep, as evidenced by higher MTRTs compared to those obtained with wider-than-ERB sweeps, masking then becomes a significant factor in perception.

The inconsistency in reports of perceptual asymmetry in earlier studies of the perception of frequency-modulated signals probably reflects the relationship of the extent of the sweeps used in these studies to the ERBs of the auditory filters of the basilar membrane. If the signals were of sweep extents that were smaller than the ERBs of the auditory filters associated with the CF of the signal, then the subjects would demonstrate asymmetry in their perception of upsweeps and downsweeps. As shown above, masking becomes a factor in perception when the limits of temporal resolution are pushed by a narrow-sweep signal. If the signals happened to have sweeps wider than the ERBs of the auditory filters associated with the CF of the signal, then perceptual asymmetry would not be observed. Many of these previous studies used signals that adapted on frequency; it is quite possible,
therefore, that the subjects were listening to sweeps that varied between being smaller than, and wider than, the ERBs of the auditory filters associated with the signal CF.

5.6 Phase locking and perception of frequency-modulated signals.

As was discussed in the review of the literature, the rapid changes in frequency that are characteristic of formant transitions in speech can be encoded by the synchronization of the discharge rate in VIIIth-nerve neurons to the instantaneous frequencies of the dynamic transitions. In addition to showing that fibers phase lock their discharge rate to frequencies near their CFs, it was shown that fibers with CFs above the instantaneous frequencies of the formants phase locked their response to the frequencies of the transitions (Carney, & Geisler, 1986; Deng & Geisler, 1987; Miller and Sachs, 1983; Sinex & Geisler, 1981). Could an impairment in the phase-locking properties of the surviving VIIIth-nerve neurons in the hearing-impaired subjects account for the observation that MTRTs of the implanted listeners were generally higher than those of normal-hearing subjects in this current study? Also, does phase locking account for the observation that among the cochlear-implant subjects lower MTRTs were often associated with signals having lower CFs? The last question is particularly difficult to answer as the pattern of MTRTs with respect to the frequency of the signals was different in all three cochlear-implant subjects. The MTRTs of CI 1 and CI 3 increased as the CFs increased for the signals with CFs below 1900-2000 Hz; the MTRTs of CI 2 generally decreased as the CF increased. The
differences in response pattern may relate to the number of surviving neurons in a particular population of neurons. It is possible that both CI 1 and CI 3 have more surviving neurons at the apex of their cochleas than does CI 2.

A consideration of the way in which information about the frequency content of the signal is delivered to the electrodes of the implant device leads the author of this study to conclude that, for the cochlear-implant users, phase locking played a limited role in the detection of rapid frequency changes in the signals used in this study. In the implant device the actual rate at which the electrodes are stimulated, and, consequently, the rate at which the cell bodies of the VIIIth-nerve neurons associated with each specific electrode are stimulated carries little to no information about the specific frequency content of the incoming acoustic signal. In the MSP, for unvoiced speech sounds, the rate of electrical stimulation is derived from F1 information present in the signal (according to the MPEAK strategy which was used by both CI 2 and CI 3 in this study). The signals used in the step-versus-glide discrimination task were not speech signals, so it is difficult to say with certainty anything about the rate of stimulation at the electrodes. It is possible that the rate of stimulation was even aperiodic. In the MPEAK strategy, the F1 encompasses frequencies from 280 to 1000 Hz. The only signals having frequencies equivalent to these F1 frequencies were Signal I (350-550 Hz) and the first 80 Hz of Signal II (920 to 1210 or 1220 Hz). The instantaneous frequencies of all the remaining signals in the set of seven signals used by CI 2 and CI 3 had no relationship, according to the MPEAK strategy, to the rate of electrical stimulation of the electrodes. The only
information about the frequencies in most of the signals analyzed by the MSP, therefore, was place information which was encoded by stimulating electrodes with the appropriate frequency ranges along the tonotopically arranged array of electrodes.

In the Spectra, the rate of stimulation of the electrodes is between 180 and 300 Hz regardless of the frequency of the incoming acoustic stimulus. The specific rate of stimulation depends upon the spectral complexity of the signal. The more spectrally complex the signal, the slower the rate of electrical stimulation; the less spectrally complex the signal, the faster the rate of stimulation. The signals used in this study were not spectrally complex, and they were all at the same level of spectral complexity. In addition, the level of spectral complexity was constant throughout a given signal. The rate of stimulation of the electrodes in the Spectra should have been the same for all of the signals regardless of their frequency content. With the Spectra, therefore, all information about specific frequencies in the signal is encoded by the stimulation of the appropriate electrodes according to their position in the electrode array.

If the surviving neurons in the cochlear-implant subjects were phase locking to the frequency of the electrical stimulation, then the only signals for which rate of electrode stimulation actually carried information about the frequency content of the signal was, as noted above, the 350-550 Hz signal and the first 80 Hz of the 920-1210 or 1220 Hz signal for subjects using the MSP with the MPEAK strategy. For all other signals used in this study encoded with MPEAK or SPEAK strategies, the rate of stimulation of the population of neurons under the influence of a particular electrode had no relationship
to the actual frequencies of the signals. In both the MSP and the Spectra, information about the specific spectral content of most of the signals was encoded as place information. Phase locking to specific frequencies of most of the simple acoustic signals used in this study had no significance for the VIIIth-nerve neurons stimulated according to a strategy based on spectral complexity or extraction of F1 information.

One might argue that the listeners were using differences in the entrainment of surviving fibers with different CFs to the electrical stimulation at each successively stimulated electrode during the encoding of the dynamic acoustic signal. As noted in the literature review, several of the investigators mentioned above showed that, in the cat, fibers tended to synchronize their rate of discharge to frequencies at or near to their CFs or even entrained to frequencies well below their CFs when the stimuli were acoustic. One might argue that synchronization to the rate of stimulation of the electrode would be greatest in those fibers under the influence of the electrode with a frequency range closest to the rate of stimulation. As electrodes with successively higher (using upsweps as an example) frequency ranges were stimulated, those fibers under their influence would have higher CFs and, hence, would be less likely to entrain to the frequency of the electrical stimulation of the electrode. Listeners might use differences in the degree of the phase locking at a succession of electrodes to detect movement within a signal.

Knauth, Hartmann, and Klinke (1994) demonstrated in the cat, however, that, with electrical stimulation, the neural responses of the VIIIth-nerve neurons were independent of the CF of the fiber. Action potentials were
phase locked, but they were phase locked to the frequency of the electrical stimulus. When acoustic stimuli were presented to the same animals, Knauth et al. found that fibers with CFs near the F0 of the stimulus had a discharge rate close to the F0; fibers with CFs near a formant frequency had a discharge rate whose frequency was near the frequency of the formant. Because electrically stimulated fibers can phase lock only to the frequency at which the pulses are presented to the electrode, any phase locking of the fibers along the sequence of electrodes of a given signal used in this study, except for the 350-550 Hz signals and first 80 Hz of the 920-1210/1220 Hz signals, would have been the same and would have had no relationship to the actual frequency content of that signal. The cochlear-implant subjects could not have used phase locking to frequencies of the higher CF sweeps as a means of tracking the changes in frequency in these signals. It is possible that the low thresholds observed for CI 2 and CI 3 for the 350-550-Hz signals and the first 80 Hz of the 920-1210/1220 Hz signals resulted from phase locking of the discharge of the neurons to the rate of stimulation because that rate would have been determined by F1 (frequencies below 1000 Hz) information in the signal and, therefore, would have changed in response to changes in acoustic sweeps which contained instantaneous frequencies lower than 1000 Hz.

As was noted in the literature review, Shamma (1985) observed that single fibers phase locked to the combined waveform of several adjacent harmonics in the speech stimuli. This may explain why the listeners with normal hearing in the current study had higher MTRTs for smaller-than-ERB signals. If the frequencies of the smaller-than-ERB sweeps interacted, the
neurons located within the auditory filter on the basilar membrane, may have phase locked their responses to the temporal envelope produced by the interaction of the frequencies rather than to the actual frequencies of the signal. It is possible that the cue(s) signaling rapid changes in the frequency of a signal, are reduced in the smaller-than-ERB signals. Shamma’s observations may also explain why the normal-hearing subjects had such low MTRTs for the 1350- and 1500-Hz sweeps. If the sweep of these signals were wider than the ERB of the auditory filter associated with the CF of the signal, the effect of any interaction of the frequencies of the sweep may have been weakened. The neurons may have phase locked to the frequencies of the sweep rather than to the frequency of fluctuation in the temporal envelope produced by an interaction of the frequencies.

5.7 The persistence of auditory images in cochlear-implant listeners

It was reported in Section 4.1 that, when the duration of the signal was greater than 300 ms, all of the cochlear-implant subjects reported hearing auditory afterimages during the step-versus-glide discrimination task. This phenomenon has been reported previously by Smoorenburg (1990). Implanted subjects in a task measuring frequency discrimination reported that discrimination was difficult when the second signal was presented less than 500 ms after the first signal. The afterimage of the first signal affected the perception of the second signal. In his 1983 study of masking, Shannon reported that the cochlear-implant subjects had recovery times that were approximately twice as long as those of the normal-hearing controls. In
another masking study, Shannon (1986) reported that an implanted subject showed a significant amount of masking 500 ms after the offset of the masker. Once again, we are reminded that great care must be exercised in the design of psychoacoustic experiments which test the perceptual abilities of subjects with cochlear implants.

5.8 Identification and discrimination of the elements of the continuum of synthetic stop consonants

The results of the identification and discrimination of elements of the synthetic /b,d,g/ continuum showed that the cochlear-implant users did not perceive the elements of the continuum categorically; that is, they did not consistently attach the appropriate phonemic label to the elements of the continuum. The results also showed that the cochlear-implant listeners were insensitive to the difference between pairs of stimuli that were the "same" and pairs that were "different."

The acoustic cues that listeners with normal hearing used for identification and discrimination of the elements of the synthetic speech continuum were apparently not available to the users of the implant device. Either the speech processor was not effective at capturing acoustic cues from the stimuli and mapping them onto the electrical stimulus, or the cochlear implant-listeners were limited in the use of the available cues by their hearing impairment, or the contrast between the electrode patterns representing the formant transitions was not great enough to allow the subjects to distinguish one element from another. An analysis of the pattern
of electrode stimulation representing the center frequencies of the F2 formants of each of the last seven elements in the continuum showed that the pattern would have been the same for every element. For example, for CI 2, the center frequencies representing F2 information in the formants of elements #7 to #14, would have been assigned to electrodes 9, 10, and 11. This lack of contrast may have contributed to the inability of the cochlear-implant subject to discriminate among and identify the stimuli in the continuum.

The performance of the cochlear-implant subjects in this current study was compared to that of subjects using the Symbion implant device in a study of phonetic categorization of synthetic elements of a /ba/ to /da/ continuum by Dorman et al. (1988). Both groups of subjects identified the elements in the /da/ portion of the continuum at near chance levels, and both groups identified the elements at the /ba/ end of the continuum at higher-than-chance levels. The identification functions of the listeners with cochlear implants were very similar to those obtained by Van deGrift Turek, Dorman, and Franks (1980) in a study of the identification of the elements of a synthetic /b, d, g/ speech continuum by listeners with various degrees of hearing impairment. As was the case among the cochlear-implant subjects in this study, Van deGrift Turek et al. observed a high degree of variability in performance among their hearing-impaired subjects. Performance of the cochlear-implant subjects on the identification and discrimination of the synthetic phonemes did not reflect the performance of the subjects in discriminating nonspeech signals having formant-like frequency transitions.
The identification functions of the normal-hearing listeners in the current study were comparable to those of listeners with normal hearing in the Van deGrift Turek et al. study and in a study of the perception of synthetic stop consonants by Stevens and Blumstein (1978). Although the discrimination function for normal-hearing listener NH 1 was similar in shape to those of the other two normal-hearing listeners, the percent correct discrimination of those elements located at the crossover points on the identification function was surprisingly low.

The conclusion that one can draw from this small study of synthetic speech perception by cochlear-implant listeners is that even though the implant device restores the sensation of sound, the users of the device must still be regarded as hearing-impaired listeners.

5.9 Performance with the Spectra compared to that with the MSP

Although it was not the intent of this study to compare the efficacy of one speech processor to another, the results did show that CI 1 was able to distinguish the step from the glide at more CFs and/or had lower MTRTs than those of the other two subjects, particularly on signals with CFs at 2000 Hz or higher. Based on the results of a study by Skinner, Fourakis, Holden, Holden, & Demorest (in press) in which the ability to identify pure and r-colored vowels was compared in listeners using either the Spectra or the MPEAK (=MPEAK 640) strategy, one could conclude that CI 1 was receiving more complete information about the acoustic signal from his Spectra speech processor than were the other subjects from their Mini speech processors.
These authors showed that the SPEAK strategy was better for transmitting durational and second formant information. Included in their paper were electrodograms, graphic displays of the pattern of electrode stimulation as a function of time, which showed that with the SPEAK strategy the representation of the token "hard" was characterized by more frequent stimulation of individual electrodes during the movement of the F2. There was also a much better representation of the narrowing of the distance between the F3 and the F2 formants that occurs with time in r-colored vowels.

With the MPEAK strategy, there was little discernible change in the F2 and F3 with time, and the representation of the spectral composition of the F2 was very thin. The better representation of the F3 frequencies of the signal by the Spectra may also explain CI 1’s better performance on the psychoacoustic task. If the subjects were using high frequency spectral splatter as a cue to distinguishing the step from the glide signal, CI 1 may have simply gotten more of that information with the Spectra than did the other subjects using MSPs.

The authors also concluded from their study that the MPEAK (640) strategy was superior for transmitting F1 information. At first glance it might appear that this observation could explain why CI 2 and CI 3, both users of the MPEAK strategy, had their lowest MTRTs for the three-electrode-sweep signals with CFs at or below 1070 Hz. Skinner et al. hypothesized that the better representation of the F1 was because the MPEAK strategy maps the lowest 1000 Hz of the signal onto more electrodes than does the SPEAK strategy. This hypothesis is pertinent for speech signals which have the
potential to result in the stimulation of any or all of the six-eight electrodes assigned to frequencies below 1000 Hz. In our study, however, the across-electrode nonspeech signals were designed to stimulate three electrodes only, irrespective of the signal CF.

The CIDIAG software developed by Bogli, Dillier, Lai, & Rohner (cited in Skinner et al., in press) used to generate the electrodograms discussed above would be extremely useful for psychoacoustic studies of listening behavior in cochlear-implant users. Speech processor strategies are based on the analysis of an incoming speech signal, but in psychoacoustic tasks, the stimuli used for the purposes of testing basic perceptual abilities are generally not speech signals. It is difficult to say with complete certainty, therefore, which electrodes are stimulated. An analysis of the electrodograms derived from the nonspeech stimuli could eliminate this ambiguity.

5.10 Performance of control subject NH 1

The performance of control subject NH 1 was somewhat different from that of control subjects NH 2 and NH 3. He had much more difficulty holding onto the cue for the discrimination of the five-step signal from the glide than did the other two subjects. As noted in the previous section, his performance in the discrimination of the elements of the synthetic speech continuum was quite different from that of the other two subjects. Age may have been one factor contributing to his "poorer" performance. This subject was 63 years old, ten years older than NH 3, the next oldest subject. A more likely explanation was he was taking medication for high blood pressure.
This subject was a patient at the Audiology Clinic at the VA hospital where he was being monitored for side effects from the medication. When he first was given the medication for the high blood pressure, he experienced tinnitus as a side effect. His dosage was reduced, and he was sent to the Audiology Clinic for periodic monitoring. At the time he participated in this study, he was having no recurrence of the tinnitus. It is possible that, while he had no physical manifestation of the tinnitus, there was some effect on his hearing abilities which was expressed in his performance on the step-versus-glide task and the discrimination and identification tasks. One of the difficulties this subject had during the two-hour testing periods was that he often became quite sleepy. For a period of time, he was taking his high-blood-pressure medication before he came to the testing session. When he changed his medication schedule so that he took the medicine after the session, the sleepiness was alleviated.

5.11 Variability in performance of cochlear-implant subjects

The extreme degree of variability in performance observed in this study is a hallmark of studies investigating the perceptual abilities of cochlear-implant users. Such variability should not be surprising when one considers that little is known about the kind and extent of damage to the hair cells and auditory nerve of the individual implant user. In addition, each implant user has an individualized MAP; hence, the frequency ranges of the electrodes, the amount of current delivered to each electrode, and the extent of the spread of current will vary from subject to subject. In this study, we
had hoped to avoid some of this variability by tailoring the frequencies of the stimuli to the frequency ranges of the electrodes and by presenting each signal at a level corresponding to the midpoint of the dynamic range for that signal. As the results showed, even with these precautions, we were unable to overcome the problem of variability in performance.

This variability is particularly a hindrance in psychoacoustic studies in which the goal is to drive the subject to his best performance. Ordinarily, in these studies it requires a fairly substantial commitment of time on the part of the subject to achieve this goal. With cochlear-implant users, because of their variability in performance, the time needed to collect reliable data can become a burden both in terms of physical time for the subject and in cost to the project. There are several possible approaches to this problem. One is simply to collect data from greater numbers of subjects. This approach has drawbacks for two reasons. One, of course, is that, compared to the hearing population in general, there simply are not many listeners with cochlear implants. Also, because implant centers tend to be localized, their patients are drawn from a wide area. It is not uncommon for an implant subject to drive one to two hours from his/her home in order to participate in a research study. A second drawback is that in psychoacoustic studies the usual procedure is to test subjects every day for a period of several weeks. Many of the implant patients, however, are working adults with full-time jobs and families. It becomes nearly impossible, therefore, to schedule subjects for more than a couple of sessions per week. A Catch-22 situation develops: in order to compensate for the limited weekly testing time, the total period over
which testing takes place has to be expanded. If there are extended periods between testing sessions, the subjects lose their skills.

One way to address the problem of using cochlear-implant subjects in psychacoustic studies is to develop methods that drive the subject to his/her best performance in a shorter time. This method would require testing and evaluating alternative methods. Another, perhaps easier method, is to take the testing to the subject. Use of the Audio Input Selector for delivery of the signals makes this a viable alternative. The Audio Input Selector can be set so that the user receives only the desired test stimulus with no extraneous auditory input from the testing environment. The software programs for the testing paradigms can be loaded onto laptop computers, and the subjects can be trained in a couple of laboratory sessions to run the programs themselves. The subjects are allowed to take the laptops home and are instructed to test themselves everyday at the same time in an environment free from distractions. The investigator can visit the subject once or twice a week to download the data and to check for any problems that may have arisen. This method allows the subject to chose a testing time that fits his/her schedule and frees the subject from having to travel long distances to the test site.
5.12 Conclusion

The results of this study demonstrated that under conditions which maximize the contrast in the frequency modulation of two signals temporal resolution in the cochlear-implant user, as measured by the ability to detect changes in frequency over time, is comparable to that of listeners with normal hearing. This study also showed that in the design of experiments for the purpose of examining the perceptual abilities of cochlear-implant users it is especially important to consider the limitations that the speech processor and the device itself may impose upon the listener.
BIBLIOGRAPHY


APPENDIX A

INSTRUCTIONS: DETERMINATION OF THRESHOLD
DETERMINATION OF THRESHOLD

In this experiment we want to present sounds that are comfortable for you to hear. To do this we need to determine two things:

a) where you first start to hear sound, that is, the softest level at which you can hear a sound and

b) the loudest level that you can comfortably listen to for a long time.

1) Look at the screen. We will first concentrate on determining the softest level at which you can hear a sound. We will start at a level where you can clearly hear the sound.

2) In the upper left corner is a red warning box. You can ignore this box.

3) There are 4 boxes at the bottom of the screen. The 2 middle boxes are going to turn brown, first the left box and then the right box.

4) A sound will always occur in one of these 2 middle boxes. There might be times when you think there is no sound there because you just can't hear it. However, there will always be a sound present in one of the 2 boxes. It just might be too soft for you to hear it.

5) You will make your response by pressing one of the two buttons on the device called a mouse. If you think you hear the sound in the left box press the left mouse button. If you think you hear the sound in the right box press the right mouse button.
6) Remember there will always be a signal present in one of the boxes even if you think you can't hear it. If you think you can't hear the sound, you must guess which box it is in. You must respond to every trial.

7) There is a time frame within which you must respond. Push the mouse button only when the answer box is green. If you respond too quickly or too slowly, your response will not be recorded. The box will remain green long enough for you to easily respond.

8) We are going to have some practice runs so you can get used to the procedure. If you have any questions feel free to ask us.

9) If you respond when the answer box is green, the accept box will light up.

10) After you respond one of the two middle boxes at the bottom of the screen will turn blue. This is the way the computer lets you know where the sound really was.

11) We will practice for 5 trials. Then we can answer any questions you might have.

12) After these practice runs we'll do 50 trials at a time.
APPENDIX B

INSTRUCTIONS: DETERMINATION OF ULL
DETERMINATION OF COMFORT LEVEL

Now we will concentrate on the second part of the procedure. This time we want to determine the level at which the sound is uncomfortably, but not painfully, loud.

We are going to use the same computer screen; however, this time the sounds will get louder. Also, we are going to use a slightly different method to determine the loudest level that you can tolerate for a short time.

1) You will still use the left and right mouse buttons. This time the sound will only be in the left middle box. You will control how loud the sound gets by pushing the left mouse button. Each time you press the left mouse button, the sound will get louder.

2) The sound you hear might be in one of the categories listed here.

    LEVELS OF LOUDNESS
    Painfully loud
    Extremely uncomfortable
    *Uncomfortably loud
    Loud but OK
    Comfortable, but slightly loud
    Comfortable
    Comfortable, but slightly soft
    Soft
    Very soft
3) We are trying to zone in on the item marked with the *. When the sound becomes uncomfortably loud, press the right mouse button. This will stop the sound so that it does not get any louder.

4) Think of uncomfortably loud as a sound that you would NOT like to listen to for a long time. Do not wait to press the right mouse button until the sound becomes extremely uncomfortable or painfully loud.

5) Remember, we want the sound to stop when it becomes uncomfortably loud. You should press the stop button on the mouse when you feel that the sound is uncomfortably loud.

6) Any questions?
APPENDIX C

INSTRUCTIONS: DISCRIMINATION TASK
DISCRIMINATION TASK

In this experiment we want to determine whether you can distinguish between two different sounds. One sound we call a glide and the other we call a step signal. To give you an idea of what each sound is, there is a picture of each of them below.

1) Look at the screen. You will see a series of boxes.

2) In the upper left corner is a red warning box. You can ignore this box.

3) There are 4 boxes at the bottom of the screen. These four boxes will turn brown, first the box on the left, then the second box, then the third box, and finally the fourth box which is on the right.

4) A sound will occur in each box as it turns brown. The sound that you will hear in three of the four boxes will be the same. A sound that is different from the other three will be heard in either box 2 or in box 3.

5) You will make your response by pressing one of the two buttons on the device called a mouse. If you think you hear the different sound in box 2 press the left mouse button. If you think you hear the different sound in box 3 press the right mouse button.
6) Even if the sound in box 2 sounds nearly the same as the sound in box 3, you must guess which one you think is different.

7) There is a time frame within which you must respond. Push the mouse button only when the answer box is green. If you respond too quickly or too slowly, your response will not be recorded. The box will remain green long enough for you to easily respond.

8) We are going to have some practice runs so you can get used to the procedure. If you have any questions feel free to ask us.

9) If you respond when the answer box is green, the accept box will light up.

10) After you respond one of the two middle boxes at the bottom of the screen will turn blue. This is the way the computer lets you know where the different sound really was.

11) We will practice for 10 trials. Then we can answer any questions you might have.

12) After these practice runs we'll do 50 trials at a time.
APPENDIX D

PATIENT QUESTIONNAIRE
PATIENT QUESTIONNAIRE

This questionnaire is to determine if you have experienced any uncomfortable or unusual effects that you feel could have resulted from your participation in the previous testing session.

1) Have you noticed the appearance of tinnitus (that is, a ringing in your ears) or an increase in tinnitus?

2) Have you had any spells of dizziness?

3) Do you have strong afterimages of the sounds after you leave the session?

4) If so, have they bothered you in any way? For example, are you unable to sleep because you hear the test sounds in your head?

5) Please remember that you are a volunteer in this study and that you are free to withdraw at any time.

If you feel that the sessions are too long, too tiring, or are causing you to have unpleasant auditory afterimages please notify the experimenter. The test sessions can be modified to make you more comfortable.
APPENDIX E

INSTRUCTIONS: DISCRIMINATION SYNTHETIC SPEECH
The task you will be doing is a discrimination task.

You will be asked to listen to computer-generated speech samples of ba, da, and ga. The speech samples will be presented in pairs, and you will be asked to decide whether the members of a pair sound "the same" or "different" to you. For example, the pair might be ba and ba or ga and ga in which case you would respond "the same", or the pair might be da and ba in which case you would respond "different." Because these speech samples are computer generated, they may sound a little odd to you. Don't worry about this, just concentrate on whether the members of a pair sound "the same" or "different." You will record your responses on a numbered sheet. You will have 3 seconds between each presentation of a pair in which to record your response and to prepare to listen to the next pair. There will be a 20-second pause after every ten pairs. If you lose your place, stop circling items, wait for the 20-second pause and then begin circling items in the next set of ten pairs. The speech samples will be presented in a block of 240 pairs. There will be two blocks, and you will be given a short break between the two blocks.

**If you are unsure** about whether the members of the pair sound "the same" or "different", you should guess. If you must guess, try to guess "the same" equally as often as you guess "different."

There will be a practice session before the real discrimination task begins.
APPENDIX F

INSTRUCTIONS: IDENTIFICATION SYNTHETIC SPEECH
The task you will be doing is an identification task.

You will be asked to identify whether randomly presented computer-generated speech samples sound like ba, da, or ga. Record your responses on the numbered sheet. You will have three seconds in which to record your response and to prepare to listen to the next speech sample. If you are unsure of your response, you should guess. When you must guess, try to guess ba, da, or ga equally often. Because these are computer-generated speech samples, they probably will sound a little odd to you. Don't worry about this. Base your response on knowing that the speech samples are either ba, da, or ga.

There will be 140 random presentations of the speech samples. These samples will be presented in groups of ten. There will be a 20-second pause between each group of ten samples. If you lose your place on the response sheet, stop circling items. Wait for the 20-second pause and then begin circling items in the next group of ten samples.