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

PARAMETRIC STUDY OF THE ACOUSTIC CHANGE COMPLEX ELICITED BY SECOND FORMANT CHANGE IN SYNTHETIC VOWEL STIMULI

by

JODI M. OSTROFF

A dissertation submitted to the Graduate Faculty in Speech and Hearing Sciences in partial fulfillment of the requirements for the degree of Doctor of Philosophy, The City University of New York

1999

This manuscript has been read and accepted for the Graduate Faculty in Speech and Hearing Sciences in satisfaction of the dissertation requirements for the degree of Doctor of Philosophy.

4/23/99
Date

Arthur Boothroyd, Ph.D.
Chair of Examining Committee

4/23/99
Date

Irving Hochberg, Ph.D.
Executive Officer

Supervisory Committee:
Walter Ritter, Ph.D.

Lawrence Raphael, Ph.D.

Outside Reader:
Terence W. Picton, Ph.D.

THE CITY UNIVERSITY OF NEW YORK

ABSTRACT

Parametric Study of The Acoustic Change Complex Elicited by Second Formant Change in Synthetic Vowel Stimuli

by

Jodi M. Ostroff

Advisor: Professor Arthur Boothroyd

The Acoustic Change Complex (ACC) is an N1-P2 event-related potential elicited by acoustic change that occurs in an ongoing stimulus. As part of an effort to develop an electrophysiologic test of speech discrimination capacity, this dissertation examined the smallest second formant (F2) change within a synthetic vowel stimulus that elicited the ACC and compared those threshold values to behavioral responses. Stimuli were created by generating a synthetic 800 ms 3-formant vowel with 7 separate degrees of F2 change from 1050 Hz. The F2 change, occurring at 393 ms, was either 9, 19, 38, 75, 150, 300, or 600 Hz. A control condition with no spectral change was also included. Electrophysiologic responses were recorded from 8 adult subjects at 5 scalp locations. The smallest F2 change that elicited an ACC in the group waveform was: a) 75 Hz (range 75 to 300 Hz) based on waveform morphology and amplitude criteria, and b) 28 Hz (range 24 to 131 Hz) based on a curve fit and interpolation technique. Behavioral responses were recorded from 7 of the 8 adult subjects who demonstrated an ACC in the electrophysiologic experiment. Using the same curve fit and interpolation technique used in the electrophysiologic experiment, the binary choice, change/no change detection task revealed a mean just noticeable difference threshold of around 5 Hz (range 4.5 to 24.5

Hz). The mean detection threshold was around 20 Hz when taken from percent correct scores at 75% (range 6.6 to 39.3 Hz). The range of threshold values for the ACC from the group waveform correspond with mean detection probabilities between 86 and 97% correct in the behavioral task. These findings support the conclusion that the ACC can be elicited by F2 changes that are perceived with confidence. An ACC elicited by F2 changes between 30 and 75 Hz is likely sufficient for the detection of F2 changes between English vowels which are on the order of 100 Hz. These data are encouraging in terms of the clinical utility of the ACC as a measure of the potential for the development of speech discrimination capacity.

ACKNOWLEDGEMENTS

Acknowledging those who have helped me to achieve this doctorate seems to me to complete some imaginary circle. The top of the circle begins first and foremost, with an expression of gratitude to my parents, Michael and Celia Ostroff, who sacrificed much in an effort to provide the best possible education for their children.

Sincere appreciation to the Graduate School, Dr. Irving Hochberg, and Dr. Arthur Boothroyd for providing the financial support, without which I could not have completed my doctoral studies.

The knowledge, insight, and expertise provided to me through interactions with my committee members Dr. Larry Raphael and Dr. Walter Ritter was invaluable and much appreciated throughout the progress and completion of this dissertation. Specials thanks also to my outside reader Dr. Terry Picton.

Colleagues and friends at the Graduate Center (many of whom were also subjects for this dissertation) deserve many thank yous, especially Brett Martin, Eddy Yeung, and Gary Chant. A special debt of gratitude to Bethany Mulhearn for her wit and to Rupa Balchandran for her memory. Respect and admiration to Constantine Chernenko and to Debbie Ostroff.

The circle is only completed by acknowledging the chairman of my committee Dr. Arthur Boothroyd, who is not only a professor of speech and hearing sciences – he is a professor of life. I will remember most fondly and am most grateful for the times he spent with me at the computer teaching me "stuff". There are no exams to test that kind of knowledge.

I am indebted to everyone who has helped me along the way.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION AND PURPOSE

General Purpose

To measure speech perception capacity on the basis of cortical auditory event-related potentials (ERPs) elicited by speech sounds.

Background

Increasing numbers of young hearing-impaired children are candidates for cochlear implants. This trend has heightened the need for valid, reliable tests of speech perception capacity in children for whom behavioral assessment may be inappropriate. Cortical auditory ERPs may fill this need. The N1-P2 component of the auditory ERP is a negative-positive wave occurring around 100–200 ms after the onset of acoustic stimulation. There is a body of literature demonstrating that the N1-P2 evoked potential can be elicited by a change in amplitude or frequency during an ongoing tonal stimulus. This potential is time-locked to the onset of the change. The author and colleagues have demonstrated this N1-P2 potential to acoustic changes within a naturally-produced speech stimulus ("say") (Ostroff, Martin, & Boothroyd, 1998). The author and colleagues have also demonstrated it to formant changes within quasi-synthetic vowel stimuli (/u-i/ and /u-U/) (Ostroff, Martin & Boothroyd, in preparation) and Martin & Boothroyd (1999) have demonstrated it to changes of periodicity within a stimulus of constant spectral envelope.

This N1-P2 potential, which we have termed the Acoustic Change Complex (ACC), may provide an index of peripheral auditory discrimination capacity. Furthermore, we

believe the ACC has potential clinical utility as a means of assessing the possibility of the development of phonetic contrast perception in children who have yet to develop a phonology (and in other difficult-to-test populations). The data collected thus far suggest that the N1-P2 potential is a more efficient and robust measure of auditory discrimination than a more established test of discrimination activity: the mismatch negativity evoked potential. The magnitude of a spectral change required to produce a detectable ACC, however, has not been determined. Hence the purpose of the present two-part study:

Part I: Electrophysiologic Study

i. To measure the ACC elicited by changes in the frequency of the second formant within a synthetic vowel stimulus as a function of the magnitude of the F2 shift (spectral change).

ii. To determine the least amount of spectral change (i.e., threshold) within a vowel needed to elicit a reliable ACC.

Part II: Behavioral Study

iii. To measure behavioral discrimination of changes in the frequency of the second formant of a synthetic vowel stimulus as a function of the magnitude of the F2 shift (spectral change).

iv. To determine the least amount of spectral change (i.e., threshold) within a vowel needed to elicit a reliable behavioral response.

v. To determine the relationship between the threshold of the ACC potential and the threshold on the behavioral task.

Definition of Terms

For the purposes of the present document, the following definitions will apply:

Event-related potential (ERP) - Variations in voltage recorded from the surface of the scalp (electroencephalogram) that are time-locked to the presentation of a stimulus. ERPs represent the net electrical fields associated with the synchronous activity of a group of neurons within the brain. The neurons may be configured in such a way that their individual electrical fields summate to yield a dipolar field. ERPs are believed to reflect post-synaptic action potentials (Rugg & Coles, 1995).

Acoustic Change Complex (ACC) - N1-P2 event-related potential complex elicited by acoustic change that occurs in an ongoing stimulus.

Mismatch Negativity (MMN) - Cortical event-related potential elicited by a deviant stimulus presented among a train of standard stimuli.

Discrimination Capacity – The ability of the system to demonstrate two different responses to two different stimuli. It is a prerequisite for the development of discrimination.

Discrimination – An internal "decision" (conscious or unconscious) that two stimuli belong to two different categories.

CHAPTER 2

LITERATURE REVIEW

This literature review is divided into three sections. The first deals with the acoustic and perceptual characteristics of vowels. Although the stimuli in the proposed study were synthetic vowels, a thorough understanding of naturally produced vowels is necessary. Given that the general purpose of the proposed study was to predict speech perception on the basis of cortical auditory event-related potentials (ERPs), the second section reviews the literature that deals with event-related potentials and speech perception. Finally, the third section reviews the literature demonstrating that the N1-P2 complex can be elicited by a change in amplitude or frequency during an ongoing stimulus. This last body of literature relates directly to the proposed study.

I. Acoustic and Perceptual Characteristics of Vowels

The literature dealing with the acoustic characteristics of vowels generally relates to theories of vowel perception. Therefore, any discussion of the acoustics of vowels must include theoretical constructs regarding the ways in which a listener identifies a vowel, as well as data supporting the various theories.

i. Static Cues to Vowel Perception

Perhaps the first and most widely accepted theory asserts that the most important cues to the perception of vowels lie in their steady state or static properties, which include the formant frequencies. For example, Delattre, Liberman, Cooper, & Gertsman (1952) attempted to find the two formant frequencies that provided sufficient acoustic

information for the perceptual identification of a vowel. By varying the frequencies of the first two formants of their synthesized steady state vowels, they determined that listeners only required the first and second formants to identify the vowels of English. Additionally, they determined that although both F1 and F2 were required for the correct identification of front vowels, a single formant, between F1 and F2, was sufficient for the correct identification of back vowels.

The two-formant theory was also advanced by the work of Peterson & Barney (1952) who recorded ten vowels in /hVd/ context spoken by men, women, and children. Acoustic measurements from wide-band spectrograms and amplitude spectra (narrowband) consisted of formant frequencies (F1-F3) and formant amplitudes. These measurements were believed to be close to the steady state values. When the /hVd/ syllables were presented to listeners for identification, the results showed a strong relationship between the intended vowel and the formant frequency pattern. There was, however, a good deal of variability in the formant frequencies among speakers and overlap in the F1/F2 plane among adjacent vowels.

As the Peterson & Barney data illustrate, there is a considerable amount of variability in the production of vowels. How is it that listeners arrive at the correct percept in spite of wide variations in the acoustic patterns? This is known as the problem of lack of invariance. There are essentially two sources of variability involved in vowel production. The first is the variability of formant frequencies across groups of speakers (men, women, and children) which results in an overlap of vowel categories in the F1/F2 dimensions. This variability arises primarily from differences in vocal tract size across groups of speakers. The second source of variability is that which is found within

speakers. The formant frequencies for a given vowel within the same speaker vary with dialect, phonetic context, stress, and rate of speech. One source of this variability has been called the "target undershoot" problem whereby steady-state formant frequencies are rarely achieved when vowels are produced in context. (Lindblom & Studdert-Kennedy, 1967).

ii. Dynamic Cues to Vowel Perception

Given these sources of variability, it seems unlikely that listeners would be able to rely solely upon the static portions of a vowel in order to identify it correctly. Listeners must be relying on other perceptual cues contained in an utterance in order to compensate for the fact that vowels are not acoustically invariant. There have thus been other approaches aimed at understanding the normalizing mechanisms that might be involved in adjusting for the talker and context-dependent variations in spectral patterns. One such approach asserts that dynamic properties, including a vowel's inherent spectral change and consonantal context guide a listener's perception of a vowel. For example, various studies demonstrated that vowels presented in consonant-vowel-consonant (CVC) context are identified more accurately than vowels presented in isolation despite the acoustic variability associated with the acoustic context.

Strange, Verbrugge, Shankweiler, & Edman (1976) compared the identifiability of isolated (natural) vowels with vowels in /pVp/ context when 1) a single talker produced all tokens and 2) multiple talkers produced the tokens. Results demonstrated that vowels in CVC context are identified with far greater accuracy than comparable isolated vowels in both single and multiple-talker conditions. This result held even when the initial and

final consonants in the CVC context were varied in a way that could not be easily predicted by the subjects. The authors concluded that listeners were using the dynamic acoustic information distributed over the syllable as a cue to vowel identity.

Generally, studies examining whether vowel perception is affected by the consonantal context in which a vowel occurs use similar methodological procedures. This methodology, as pointed out by Rakerd, Verbrugge, & Shankweiler (1984) may account for the finding of a significant advantage for vowel identification in CVCs. For example, subjects are required to circle a written alternative or equivalent to the stimulus they heard (for example "dad" as an alternative for the correct response for /ae/). This places a bias towards the context effect because CVC combinations are more "real" than isolated vowels. This procedure also places strong demands on the short-term memory of subjects because the stimulus must be compared with each alternative. Vowels in context might be better remembered than vowels in isolation.

In light of these methodological considerations, Rakerd et el., (1984) attempted to compare the identifiability of vowels in CVC context and in isolation using a new kind of perceptual "monitoring" task. The task does not rely on the selection of orthographic alternatives and does not place strong demands on short-term memory. Subjects were asked to monitor a list of vowels in /pVp/ and /V/ contexts for occurrences of nine vowels by checking "yes" on an answer sheet if an item was an instance of the vowel being monitored. Error rates for open vowels in CVC context were lower than for open vowels in isolation. Identification of close vowels was slightly (though not significantly) hindered by context. This finding is accounted for by the fact that the identification of open vowels, which have a greater degree of spectral change than close vowels (because

they are marked by more extensive formant transitions), might be aided by context to a greater degree than identification of closed vowels.

The spectral change contained within a vowel has been implicated in several studies as the major perceptual characteristic leading to vowel recognition. Lindblom & Studdert-Kennedy (1967) presented vowels in CVC context in which the formants varied along a continuum from /I/ to /U/. They demonstrated that the identification of a vowel is determined by the rate and direction of the adjacent F2 transitions. That is, the category boundaries for the vowels shifted as a function of the formant transition context in which the vowel was placed.

Hillenbrand & Gayvert (1993) also noted that the pattern of spectral change throughout an utterance is essential for the identification of a vowel. They attempted to determine how well the Peterson & Barney vowels could be identified on the basis of steady-state information alone. They synthesized the Peterson & Barney steady-state vowels based on fundamental and formant measures from the original dataset. A second set of the vowels was created which was identical to the first except that a falling pitch contour was included in order to introduce spectral change information but not a spectral envelope change. Error rates for the identification of synthesized steady-state vowels with a falling pitch contour were significantly lower than for those with a constant F0 indicating that the dynamic spectral change information aided in the identification of the vowels. The authors pointed out the limitations of the Peterson & Barney vowel database as a means of evaluating theories of vowel perception because the omission of duration and spectral change makes it impossible to determine how listeners use this information.

Hillenbrand, Getty, Clark, & Wheeler (1995) attempted to replicate the findings of Peterson & Barney (1952). Recordings were made of men, women, and children producing the Peterson & Barney vowel stimuli. The vowel identification scores for this set of stimuli were compared with those from the Peterson & Barney data. Although there were some differences in average formant frequencies of F1 and F2 and in the degree of overlap among the adjacent vowels, as in the original study, the signals were nearly always identified as the vowel intended by the talker. Discriminant analysis showed that the vowels could be separated into categories with a higher degree of accuracy if duration and spectral change information was included in addition to steady-state information about the vowels.

Dynamic properties are also implicated in the identification of diphthongs. Gay (1970) attempted to determine whether the phonetic identity of the formant targets or the second formant transition served as the primary cue for identifying three phonemically distinct American English phonemes: /ɔI/, /aI/, /au/. Each diphthong was reduced in duration from 250 to 100 ms in steps of 10 ms. The first and third formant frequencies were held constant while the second formant contained a transition. Shifts in the perception from simple vowel to diphthong occurred as a function of transition duration; 50% of the identifications of a diphthong occurred when the stimulus duration reached approximately 150 ms in all three cases. This finding implies that the transition duration (rather than the specific changes in frequency) separates the perception of a vowel from a diphthong.

iii. Intrinsic Relational Properties as Cues to Vowel Perception

The relationships among the fundamental and formant frequencies within a vowel provide another source of information believed to provide cues to vowel perception. The invariance problem is deemed not to exist when the correct representation of spectral properties of vowels is considered. For example, when certain linear or non-linear transformations (i.e., log, bark, mel) of the F0 and formant frequencies are applied, the overlap in the F1/F2 plane is believed to be reduced. The main goal of these normalization procedures is 1) to maximize differences between phonologically distinct vowel categories and 2) to minimize differences in the same vowel spoken by different talkers.

Syrdal and Gopal (1986) developed such a normalization scheme that forms the basis for their two-stage quantitative perceptual model of vowel perception. At an intermediate stage of auditory processing the pattern of peripheral auditory excitation is represented as the distance in critical bands (barks) by transforming the fundamental and formant frequencies into the bark scale. Three bark-difference measures are calculated from the bark-transformed formant frequencies: F1 minus F0 (F1-F0), F2 minus F3 (F2 – F3), and F3 minus F2 (F3-F2). The second or higher phonetic stage of processing involves classifying the vowels according to whether or not they exceed a three bark critical distance. Using the Peterson and Barney (1952) data, Syrdal and Gopal demonstrated that the vowel transformations correspond well to the major phonetic dimensions by which vowels are traditionally classified. The F1-F0 dimension relates to vowel height and high vowels have an F1-F0 difference less than three Bark. The F3-F2 dimension relates to tongue advancement and front vowels have an F3-F2 difference less than three Bark. The

model also reduced the variability between speakers while providing good vowel clustering.

Hillenbrand & Gayvert (1993) attempted to evaluate several normalization schemes (including that of Syrdal & Gopal, (1986)) using the Peterson & Barney (1952) vowel database. Vowels were classified into categories using a statistical technique called quadratic discriminant analysis. In terms of category separability, it was found that there was no advantage for the nonlinear auditory transforms over linear frequency, and no advantage in category separability for spectral difference measures compared to absolute frequency measures of formants. However, the differences between phonetically equivalent vowel tokens within a talker group were reduced using nonlinear transforms and spectral difference measures.

Nearey (1989) pointed out the limitation of the application of models of normalization to explain the perceptual mechanisms underlying vowel recognition. These types of approaches relate indirectly to perception since they do not involve comparisons with perceptual data. Notably, these perceptual models performed significantly better (around 85%) than did human listeners in classifying vowels based on static spectral cues (75% for human listeners in the Hillenbrand & Gayvert data, (1993)). This discrepancy between performance by perceptual models and human performance indicates that the perceptual models may be classifying vowels based on some cue or cues to which human listeners are not sensitive and which they therefore can not use. Additionally, although these models propose a strong relationship between the fundamental frequency and the frequency of F1, speech that has been altered to conceal the speaker's voice (for television broadcast) is still intelligible even as the F0 and normal formant relations are

substantially altered. The same is true of whispered speech that has no fundamental frequency.

iv. Extrinsic Relational Properties as Cues to Vowel Perception

Several investigators have proposed that listeners establish a frame of reference from the information that is distributed across the vowels of a given speaker in order to narrow down the speaker's vocal tract size and/or formant ranges and thus identify the vowels. This type of information is obtained through relational properties such as relative vowel duration and relative formant frequencies of a vowel compared to those of other vowels of the same speaker. For instance, Ladefoged & Broadbent (1957) demonstrated that the categorization of vowels placed in CVC context could be altered systematically by manipulating the formant structure of the introductory sentence "please say what this word is." These results suggest that the linguistic information conveyed by a given vowel does not depend solely on the absolute values of its formant frequencies alone, but also on the relationship between the formant frequencies of that vowel and the formant frequencies of other vowels produced by that speaker.

Fry, Abramson, Eimas, & Liberman (1962) demonstrated that the perception of a vowel could be influenced by other vowels that precede it. For example, they demonstrated that the phoneme boundaries in the case of 13 synthetic vowels created along a continuum from /I/ to /ɛ/ to /ae/ were less sharply defined than in the case of stop consonants. Moreover, the results showed that the sequence or acoustic context in which the vowels were placed considerably influenced subjects' perceptions of the vowels. This influence was always in the direction of contrast. That is, when a stimulus was paired

with any other that was more closed in articulatory character than itself, it was judged to sound more open, or /ae/-like. When a stimulus was paired with any other that was more open than itself, it was judged to sound more closed, or /I/-like. These findings suggest that listeners tend to judge a vowel by comparing it with one they have just heard. This might reflect the process by which listeners build a reference for a particular speaker against which they identify all other vowels uttered by that speaker.

Finally, Verbrugge and Rakerd (1986) demonstrated that the dynamic portions of a vowel might be used to determine the articulatory characteristics that are common to productions of the same vowel by speakers of the same dialect which could in turn point to vowel identity. Results showed that listeners were able to correctly identify a CVC syllable (about 75% of the time) whose center (vowel) was silenced leaving only the initial and final portions of the syllable. They found the same result when the initial portion of the silent-center syllable was produced by a speaker of a different gender than was the final portion of the syllable. The authors maintain that these results indicate that the listeners integrate the initial and final portions of the syllable because of the common articulatory style, dialect, prosodic features of the male and female speakers. They conclude that vowel identification is achieved by integrating dynamic properties of the stimulus in addition to any steady-state information available.

II. Event-Related Potential Measures of Speech Perception

Much of the work on speech perception using ERP techniques has focused on a well-known phenomenon in speech perception called "categorical perception". This is possibly because categorical perception sparked a debate that has lasted for about fifty

years regarding the mechanisms that underlie man's ability to perceive speech. Categorical perception refers to the non-linearity in the perception of speech sounds. For example, subjects are asked to identify stimuli varying in equal steps along a frequency continuum from /da/ to /ga/. They are then required to discriminate pairs of stimuli drawn from that continuum. The exercise reveals that listeners can discriminate two speech sounds that they have identified as different phonemes with far greater accuracy than two sounds that they have identified as the same phoneme, although the acoustic difference between the pairs of sounds is equivalent (Liberman, Harris, Hoffman, & Griffith, 1957).

Two competing traditions emerged from the debate surrounding categorical perception. The phonetic/cognitive tradition contends that "speech is special" and that a speech continuum is perceived categorically in terms of language-specific or linguistic features. The auditory/psychoacoustic tradition on the other hand, contends that phonetic boundaries arise from the same mechanisms that underlie the processing of all acoustic input. Neither tradition has been able to resolve the issue completely using the experimental methodology available to it. In recent years a novel approach to the study of speech perception has been taken by some. This approach involves using electrophysiologic techniques that are capable of examining the processing of phonemes in real time by the perceptual and cognitive mechanisms.

Aaltonen, Niemi, Nyrke, & Tuhkanen (1987) were among the first to use the event-related brain potential (ERP) called the mismatch negativity response to determine whether speech is processed in terms of its acoustic or its phonetic/linguistic features. The mismatch negativity is a response to stimulus change, elicited by a rare (deviant) sound presented in a train of homogeneous (standard) stimuli. It is believed to originate

from the supratemporal auditory cortex and it is identified in the electrical waveforms as a negative potential, largest at the frontal midline site, occurring around 200 ms after stimulus onset.

Three synthetic stimuli that varied in the transition of the second formant were used for the Aaltonen et al. study. The stimuli were endpoints of the Finnish /i/-/y/ continuum. The first was a category exemplar of /i/ (pure stimulus), the second was a category exemplar of /y/ (pure stimulus), and the third was a category boundary stimulus /i-y/ which was equally identified by Finnish listeners as /i/ or /y/ (border stimulus). The ERPs (MMN, and P300) were recorded in three conditions: 1) pure-pure condition, when each pure stimulus served as the standard and the other as the deviant (i.e., /i/ standard, /y/ deviant and vice versa), 2) border-pure condition, when the border stimulus served as the standard and the pure stimulus served as the deviant (/i-y/ as the standard, /i/ or /y/ as the deviant), and 3) pure-border condition, when the pure stimulus served as the standard and the border stimulus served as the deviant (/i/ or /y/ as the standard and /i-y/ as the deviant). All three stimulus conditions elicited the MMN, although the MMN to the pure-pure condition was the largest in amplitude and earliest in latency. The MMN thus varied with the extent of the acoustic difference between the standard and the deviant because the pure-pure condition contained a larger acoustic difference than did the border-pure/pure-border condition. These results suggest that the MMN is more sensitive to acoustic than to the phonetic features of the stimuli.

In a similar study, Aaltonen, Paavilainen, Sams, & Näätänen, (1992) confirmed the conclusion originally reached by Näätänen, Gaillard, & Mäntysalo (1978); that the MMN reflects the automatic discrimination of auditory information. Synthetic steady-state

vowels from the Finnish continuum /i/ to /y/ were used. The MMN was recorded under two conditions: 1) Phonetic center condition - a category exemplar of /i/ with a second formant frequency of 2400 Hz was the standard and the deviants varied randomly having second formant frequencies of 2450, 2550, and 2600 Hz, and 2) Phonetic boundary condition - when the category boundary stimulus /i-y/ (equiprobable identifications of /i/ and /y/ by Finnish listeners) with a second formant frequency of 2100 Hz was the standard and deviants varied randomly with second formant frequencies of 2150, 2250, and 2300 Hz. The MMN was elicited by both conditions, but only when the acoustic (F2) difference between the standard and the deviant was the largest (i.e., standard = 2100 or 2400 Hz, deviant = 2300 or 2600 Hz). The only significant factor affecting the MMN amplitude was the magnitude of the acoustical difference between the standards and deviants.

These two studies by Aaltonen and colleagues raise a methodological problem regarding the use of vowels in studies aimed at examining ERPs and categorical perception. Fry et al., (1962) demonstrated that vowels are perceived less categorically, that is more continuously, than are consonant stimuli. Moreover, vowels presented in isolation have a tendency to be perceived more continuously than vowels presented in CVC context, which are likely to be perceived more categorically (Rakerd et al., 1984 and Strange et al., 1976). Therefore, vowels are not likely to reveal the mechanisms that underlie categorical perception as well as stop-consonants.

Several other investigators have also demonstrated evidence of speech sound processing based on acoustic information using other stimuli. Sams, Aulanko, Aaltonen, & Näätänen (1990) argued that their data revealed continuous rather than categorical

processing of speech sounds. A phonetic continuum comprising synthetic stimuli from /bae/ to /dae/ to /gae/ was created whereby the frequency of the second formant was increased in steps of 200 Hz from 400 to 2800 Hz. The syllable /bae/ always served as the standard stimulus. ERPs were recorded in an ignore condition and a discrimination condition. The physical difference between the standard and the deviant was found to determine the magnitude of the MMN response. The amplitude variations in this potential showed no consistent features that could easily be related to categorical boundaries between /bae/, /dae/ and /gae/. When the subjects attended to the stimuli, there were some reflections of categorical perception in the later ERP components N200 and P300.

Sharma, Kraus, McGee, Carell, & Nicol (1993) predicted that if the MMN was reflective of categorically perceived stimuli as a result of phonetic analysis, crossing a phonetic boundary would yield a MMN with a larger amplitude than a MMN to stimuli within the same phonetic category. Nine synthetic stimuli from the phonetic continuum from /da/ to /ga/ were created. The stimuli differed only in the frequencies of the 2nd and 3rd formants. Two sets of stimuli were selected for the electrophysiologic measures; a within-category contrast and a between-category contrast. The acoustic difference between pairs of stimuli used in the two stimulus conditions was equivalent. A MMN response was detectable in all subjects for both between-and within-stimulus contrasts with no significant differences in latency or amplitudes between the contrast conditions. These results failed to demonstrate categorical perception on the basis of phonetic aspects of the speech signal.

A series of experiments by Maiste, Wiens, Hunt, Scherg, & Picton (1995) used the same phonetic continuum as the one used by Sharma et al., (1993). They reported that

identification scores revealed that the /ba/ to /da/ continuum was perceived categorically, although they did not perform discrimination tests. The MMN, obtained in the across-category condition when the standard stimulus came from the /ba/ end of the continuum, was not obtained in this condition when the standard came from the /da/ end of the continuum. A categorically perceived continuum from /ba/ to /da/ should not be influenced by the end of the continuum from which the standard is chosen. The /da/ stimulus, however, had a broader frequency content than did the /ba/ stimulus. Therefore, the authors speculated that when /da/ was presented as a deviant in the /ba/ standard condition, its presentation elicited an MMN because of the addition of spectral information. Moreover, the MMN evoked by a small change in intensity was much larger in amplitude than that evoked by a change in a phonetic category (i.e., from /ba/ to /da/). The authors concluded therefore that the MMN evoked by speech sounds indicates that acoustic rather than phonetic changes are being detected.

Several studies have introduced some interesting results contrary to those finding no evidence of phonetic/linguistic speech sound processing at the cortical level. For example, Aulanko, Hari, Lounasmaa, Näätänen, & Sams (1993) demonstrated that the magnetic counterpart of the MMN (MMNF) can be elicited by phonetically relevant features embedded in extensive irrelevant acoustic variation. Synthesized two-formant syllables with either steeply rising or steeply falling F2 transitions were generated. Stimuli with rising F2 transitions were identified as /bae/ and those with falling F2 transitions were identified as /gae/. Although the first formant was the same for both /bae/ and /gae/, they were produced with 16 different fundamental frequencies which were presented randomly. Standard tones were presented 10 dB lower than deviant tones. The

MMNF was elicited by an across-category /bae-gae/ contrast despite the extensive F0 variation of both standard and deviant stimuli. The authors concluded that the MMNF is sensitive to phonetically relevant F2 transitions despite extensive irrelevant pitch variation.

Näätänen, Lehtokoski, Lennes, Cheour, Huotilainen, Iivonen, Vainio, Alku, Ilmoniemi, Luuk, Allik, Sinkkonen, & Alho (1997) demonstrated that the MMN and MMNF reflect language-dependant memory traces that are activated during the processing of speech. They first determined the phonemes best representing the vowel prototypes /ɛ/, /ö/, /ɑ̈/, and /o/, by asking Finns and Estonians to phonemically categorize stimuli varying only in the frequency of the second formant; /ɛ/ = 1940, /ö/ = 1533, /ɑ̈/ = 1311, and /o/ = 851. Finns and Estonians judged the vowels common to both languages in a similar manner although Finns showed a gap in the /ɑ̈/ position of the profile because Finns do not have this vowel in their language. In the ERP study, the prototype /ɛ/ became the standard stimulus and all of the other vowel prototypes (/ö/, /ɑ̈/, /o/) became deviants. The MMNF amplitude was enhanced as a function of the F2 frequency deviance from the /ɛ/ standard. For the Finnish subjects, however, the deviant /ö/ (prototype) elicited a larger MMN than did the deviant /ɑ̈/ (non-prototype) although /ö/ deviated acoustically less from /ɛ/ than did /ɑ̈/. The MMN amplitude elicited by /ö/ was not larger than the MMN amplitude elicited by /ɑ̈/ in the responses from the Estonian subjects which was interpreted as being due to the fact that /ɑ̈/ is a prototype in Estonian but not in Finnish. Therefore, a vowel prototype of the subject's native language

presented as the deviant stimulus elicited a considerably larger MMN than could have been expected on the basis of the acoustic (F2) distance from the standard stimulus alone. Moreover, the MMNF was larger over the left hemisphere than over the right hemisphere and the dipole moment was considerably greater for all prototype deviants than for the non-prototype deviant /ɑ̃/. Although there is the suggestion that the prototype deviants elicited some language processing in the left hemisphere, there is a need to rule out the possibility that the complex nature of the vowel stimuli, not necessarily their linguisitic salience, produced the larger response over the left hemisphere.

A recent study by Winkler, Lehtokoski, Alku, Vainio, Czigler, Csepe, Aaltonen, Raimo, Alho, Lang, Iivonen, & Näätänen (1999) suggests that the MMN reveals the access of both sensory and categorical representations of speech stimuli in parallel. The MMN was recorded to synthesized standard and deviant vowels representing an across-category contrast in Hungarian, but a within-category contrast in Finnish and vice versa. The MMN was elicited in both Finnish and Hungarian subjects to either contrast, although speakers of each language demonstrated a larger MMN amplitude to the contrast that crossed a vowel boundary in their native language than to the within category contrast. These results suggest that the MMN amplitude was enhanced in the across-category contrast because of access to information that is phonetically relevant in the language of the listener/subject.

Steinschneider, Schroeder, Arezzo, & Vaughan (1995) presented evidence that the auditory system places constraints upon neural encoding of speech that allows for categorical perception. Multi-unit activity (the net action potential activity of neuronal ensembles within a sphere of approximately 100 μm in diameter surrounding a recording

electrode) was measured from the auditory cortex of an awake monkey. The stimuli consisted of clicks and syllables with four different voice onset times (VOT): 0 ms (/da/), 20 ms (/da/), 40 ms (/ta/), 60 ms (/ta/). Two main response types were produced in relatively equal proportions: 1) "phase-locked" response patterns to the onset of the syllable followed by responses time-locked to the F0 of the syllable, and 2) "double-on" patterns which contained a peak at stimulus onset and a 2nd peak at the onset of voicing. Responses to 40 and 60 ms VOT showed this "double-on" response. Responses to 0 and 20 ms VOT only showed a peak at the onset of the syllable. The authors suggested that the "double-on" response demonstrates evidence for categorical perception at a neuronal level because the category boundary is marked when the response gains a second peak. Within the /da/ category, responses only contained a single peak to the onset of voicing, and within the /ta/ category two peaks were displayed. The "phase-locked" responses however did not show categorical properties because phase locking to the onset of voicing was progressively delayed as the VOT was increased and there was no indication of a boundary between the /da/ and /ta/ categories. The authors noted that although both response types were equiprobable in the neurons sampled, the psychoacoustic boundary probably represents an average of the temporal activity patterns in the auditory cortex.

Kukoranta, Hari, & Lounasmaa (1987) demonstrated a double-peaked response in the ERP waveform elicited by speech stimuli. Neuromagnetic and electric responses were recorded to a Finnish CV word /hei/ and a noise burst. Both stimuli were 320 ms in duration. The fricative portion of the CV syllable was 100 ms in duration and the frequency of the fundamental was 110 Hz. The spectral content of the noise burst peaked at 0.5 to 3KHz. The noise burst evoked N100 and P200 responses. The CV syllable

evoked the N100 as well as another deflection of the same polarity (N200) at 200 ms (100 ms after the onset of the vowel). When a 200 ms noise burst imitating the fricative was abruptly changed into a square wave tone, simulating the vowel onset, the N200 response was obtained 100 ms after the onset of the tone. Subjects noted that this stimulus did not sound like speech. The amplitude relationship between the noise and the tone was similar to that between /h/ and /ei/, therefore it could not have been the amplitude increase triggering the N200. The authors concluded that the evoked potentials at these latencies are sensitive to the acoustic rather than to the phonetic features of speech stimuli because the stimulus that simulated the CV syllable elicited the same evoked potential pattern, as did the CV syllable.

The use of magnetic recordings by this team of investigators allowed them to determine that the N200 dipole was significantly anterior to the N100 dipole in both hemispheres which suggested separate generator sites for N100 and N200. The localization of generator sites responsible for the event-related potentials elicited by speech sounds represents a new direction in ERP and speech perception research. For example, Kuriki & Murase (1989), investigated the characteristics of the magnetic responses in the human auditory cortex to pure tones and short speech sounds /a/ and /ka/. The field patterns over the scalp at the peak latencies (N100 and P200) indicated a single current dipole as an equivalent field generator. No statistically significant differences were found between the dipole locations of the pure tones. The dipoles for /a/ were anterior to the dipole for a 1000 Hz pure tone in the left hemisphere. The dipoles for /ka/ were lateral-posterior to the dipoles for /a/, also in the left hemisphere. This study was

one of the first to provide evidence that speech sounds and other acoustic input might be processed in distinct regions of the auditory cortex.

Poeppel et al., (1997), were not able to demonstrate that speech sounds, namely vowels, and pure tones are differentiated spatially (by their dipoles) in the auditory cortex. They evaluated the ERPs to vowels as a function of pitch (F0), phonetic category (formants), and hemisphere. Three-formant synthetic vowels /a/, /i/, and /u/ of two different fundamental frequencies, 100 and 200 Hz, were used to elicit auditory neuromagnetic fields (N100m or M100). The results demonstrated that the localization for vowels (based on equivalent current dipole modeling) spatially overlapped with the localization for independently presented tones. The location was a well-circumscribed area in the supratemporal auditory cortex. The localization didn't distinguish between speech and non-speech stimuli and there was no hint of a vowel space or any underlying vowel phonotopy. The data also pointed to the lateralization of vowels in the left hemisphere. Additionally, no significant effect of F0 on M100 amplitude or latency was demonstrated suggesting that vowels are processed independently of their pitch.

Eulitz, Diesch, Pantev, Hampson, & Elbert (1995) suggested that the sensitivity to the linguistic features of speech could be demonstrated in the ERP in the latency interval subsequent to the N100m. If an auditory stimulus continues for more than 150 – 200 ms, a stimulus-locked DC shift lasting for the duration of the stimulus can be recorded. This is commonly referred to as the sustained field (SF). The electric counterpart is the sustained potential (SP). The N100/N100m and the SP/SF potentials were recorded in humans to the synthetic German vowels (/a/, /i/, /u/, /ae/, /oe/) and a tone of 1000 Hz. The fundamental frequency of all vowels was the same (129 Hz). The sustained field rms and

sustained potential amplitudes as well as the dipole strength of the sustained field were larger for vowel-evoked than for tone-evoked stimuli. This effect was larger in the left hemisphere. No significant interaction of hemisphere and stimulus type for N100/N100m was found suggesting that speech-related processing doesn't occur in the N100 latency range.

Most of the literature in the area of ERPs and speech perception to date suggests that the auditory cortex does not respond as a detector of speech-specific components. More likely, it responds in a general way to acoustic transients that are used in the analysis of temporal speech features. This conclusion might stem from the lack of innovative methodological techniques that may not be sensitive enough to reveal phonetic level processing. Additionally, the common use of the N100 and MMN evoked potentials in ERP research may reveal a more peripheral level of processing that precedes speech processing at a phonetic level. Thus the use of different methodological approaches and the investigation of less common ERP components such as the sustained field are beginning to provide some evidence for the level at which the linguistic feature of speech may be processed.

III. N1-P2 Potential to Change in an Ongoing Stimulus

i. N1-P2 Potential to Change in an Ongoing Tonal Stimulus

Early investigations of the N1-P2 event-related potential, commonly referred to then as the "vertex or V potential", were aimed at determining the relations between the potential and certain parameters of acoustic stimuli. A goal of some of the attempts was to quantify and describe the response so that it could be used as a tool for measuring

hearing thresholds in a more objective manner than traditional audiometric techniques. The N1-P2 potential is a negative-positive-going pair of responses that peaks between 100 and 200 ms post-stimulus onset. It was at first termed the vertex potential because it is recorded best from the vertex of the skull. The work on this potential as a measure of hearing thresholds was largely abandoned in the 1970's for two main reasons: The response was determined to be too variable for this purpose and the auditory brainstem response was discovered. Nonetheless, the papers on the subject provide some interesting conclusions about the characteristics of the potential in response to stimuli containing various types of change. The types of changes typically investigated included frequency and amplitude modulation, and changes in the rise time and plateau of the signal. These dynamic changes are interesting because they are likely to be of high functional significance to the neural analysis of acoustic signals, including speech sounds. Speech sounds contain significant ongoing variations in frequency and amplitude. The N1-P2 potential to these types of variation might therefore provide some useful descriptive information about the manner in which the brain processes the signals necessary for speech perception.

Davis and Zerlin (1966) were among the first to investigate the N1-P2 potential in response to various parameters of an acoustic signal. They presented 1200 Hz tone bursts of various durations and rise times in the soundfield at 90 dB SPL to young listeners. They examined the duration of the plateau by varying it from 2-320 ms while rise and fall times were held constant at 5 ms. They also examined rise and fall times of 2.5, 5, 10, 25, 50, and 100 ms while the plateau duration was held constant at 5 ms. They concluded from the data that there was no systematic increase in N1-P2 amplitude with changes in

the duration of the plateau. There was also no relation between N1-P2 amplitudes or N1 latencies and rise/fall time.

Onishi & Davis (1968) also examined the effects of stimulus duration and rise time on the N1-P2 potential. 1000 Hz tone bursts were used in two experiments. The first experiment investigated various plateau durations of 0, 3, 10, 30, 100, or 300 ms with a rise time of either 3 or 30 ms at 25, 45, 65, and 85 dB SPL. The second experiment investigated rise times of 3, 10, 30, 50, 100, 300 ms, with a plateau duration of 2.5 ms at intensities of 45, 65, and 85 dB SPL. The amplitude of the N1-P2 potential increased and the latency decreased as the intensity of the signal increased. The amplitude of the response increased as the plateau duration increased up to 30 ms, after which response amplitude remained constant with further increases in plateau duration. The latency of the response was independent of plateau duration except at the lowest signal intensity. The amplitude of the response decreased as the rise time increased above 30 ms and the latency of the response decreased as the rise decreased (especially below 30 ms). Therefore, the authors concluded that the duration of the plateau becomes important only when it is less than 30 ms and the rise time becomes important only when it is greater than 30 ms.

Skinner & Jones, (1968), recorded the P1 and N2 components of the vertex potential from electrodes at the vertex and mastoids of 40 subjects to tone bursts of 1000 Hz in two different experiments. The first used signal durations of 10, 25, 45, 75, 100, and 150 ms with a 10 μsec rise time at 10 and 15 dB SL. The second used rise times of 10 μsec, 5, 10, 25, 50, and 75 ms with a plateau duration of 75 ms at 30, 50, 70, and 90 dB SL. No clear trends were apparent in the amplitudes of the P1-N2 with increases in signal

duration, although latencies appeared to be shorter for increases in duration up to 50-75 ms. Latencies were not affected by rise time but they did decrease with increases in sensation level. The P1-N2 amplitude increased as the rise time decreased and the sensation level increased. The authors therefore concluded there is no relationship between signal duration and the amplitude of P1-N2. Larger responses are elicited by signals with fast rise times.

Ruhm & Jansen (1969) concluded that the amplitude of the N1-P2 potential is sensitive to rate of change, although N1 latency is not. They employed 3000 Hz trapezoidal shaped tones at 40 dB SL with rise times of 10, 50, 100, 250, 300, 500, and 1000 ms. Electrodes were located at the vertex and right earlobe. Behavioral thresholds were estimated at each rise time by having subjects press a switch as soon as they heard the tone. This reaction time measure (i.e., behavioral threshold) was then subtracted from the N1 latency of the AER and became the latency measure. There was no trend of latency change with increase in rise time. The amplitude of the response steadily decreased as rise time increased from 10 to 1000 ms. Inter-subject variability was noted to be high in terms of the amplitude and latency measures.

Clynes (1969) was among the first to describe the N1-P2 potential to frequency (FM) and amplitude modulated (AM) signals. The modulated signals were of triangular, trapezoidal, and square wave shape presented to subjects at 40 dB SL. The N1-P2 potential was elicited by stimuli whose frequency and amplitude was modulated in a trapezoidal and square wave manner but not by stimuli modulated in a triangular manner. Moreover, the N1-P2 potential was elicited when the frequency or amplitude changed from a steady-state to a state of modulation and not when it went from a state of

modulation to a steady-state. The steady-state portion of the stimulus had to be maintained for at least 300 ms before the modulation occurred in order to initiate the N1-P2. The N1-P2 potential decreased in amplitude as the constant portion of the stimulus deviated from its steady-state. Although both increases and decreases in frequency elicited the N1-P2, the response was greatest when the amplitude of the stimulus was increased rather than decreased.

Similar findings were demonstrated by Kohn, Lifshitz, & Litchfield (1978) whereby FM stimuli were presented at 40 dB SL to four subjects. Two types of FM stimuli were employed. The first was a square wave stimulus, either "positive", which started at 1000 Hz and ramped up to 1500 Hz and back to 1000 Hz, or "negative", which started at 1500 Hz, ramped down to 1000 Hz and back up to 1500 Hz. The second type was "ramp modulated" whereby a tone swept between 1000 and 1500 Hz. Sometimes the "ramp-modulated" tones were preceded by increasing or decreasing frequency tones. The direction of the square wave stimulus (positive or negative) did not influence the amplitude or latency of the N1-P2 potential. Responses elicited by "ramp modulated" tones were similar to those elicited by square wave stimuli, although responses to descending frequency ramps were smaller than to ascending frequency ramps. When "ramp modulated" tones were preceded by tones of increasing or decreasing frequency, no N1-P2 response was observed.

AM and FM stimuli were used in an experiment by Jerger & Jerger (1970) to demonstrate that individuals with sensorineural hearing loss display heightened sensitivity to intensity change and impaired sensitivity to frequency change. A 53 year-old man with sensorineural hearing loss and a 19 year-old girl with normal hearing were

subjects. Ongoing pure tone signals of 500, 1000, and 4000 Hz were frequency-and amplitude-modulated in increments at 2, 3, and 5 second intervals randomly. The auditory evoked response (AER) was recorded from the vertex and mastoids in an "attend" condition where subjects were asked to push a button when they heard the change. The subject with the sensorineural hearing loss showed steeper functions than did the normal hearing subject at all frequencies. That is, he displayed better detection of the intensity modulation that was reflected in larger N1-P2 amplitudes to the change. The subject with normal hearing displayed better detection of the frequency-modulated signals than did the subject with sensorineural hearing loss. This was also reflected in larger N1-P2 amplitudes to the frequency-modulated signals.

As the literature developed in the field of auditory evoked potentials, several investigators began to examine the effect of the rate and magnitude of the absolute frequency change on the N1-P2 potential. Spoor, Timmer, & Odenthal (1969) recorded the potential from an electrode at the vertex to three types of 1000 Hz trapezoidal-shaped tones each with a rise and fall time of 20 ms and a plateau duration of 40 ms: The first type consisted of short tone bursts presented at 10, 20, 30, 40, 50, 60, 70, 80, and 90 dB SL. The second type consisted of intensity-modulated tones presented at 30 and 70 dB SL, whereby the increment of intensity was varied from 1 to 15 dB during the rise time. The third type consisted of frequency-modulated tones, presented at 30 and 70 dB SL, whereby the maximal frequency shift reached in the plateau portion varied from 1 to 10% of the base frequency (1000 Hz). N1-P2 amplitudes increased as sensation level increased for the short tone bursts, as the change of intensity increased for the intensity modulated tones, and as the change of frequency increased for the frequency modulated tones. The

N1 latencies decreased for these three stimulus types. Equivalent increments of intensity in dB produced much larger responses at 70 dB SL than at 30 dB SL while latencies were shorter at the 70 dB level compared to the 30 dB level for intensity modulated pulses. Frequency modulated tones, on the other hand, presented at 70 dB SL, elicited larger responses with longer latencies than did frequency modulated pulses presented at 30 dB SL. Overall, frequency modulated tones showed larger N1-P2 amplitudes and longer latencies than did the intensity modulated tones.

Ruhm (1970) also reported that FM signals elicited longer N1 latencies than did AM signals as revealed by a single subject who listened to both FM and AM stimuli. Ramp-modulated tones presented at 50 dB SL were swept upward and downward every 7.5 seconds and the duration of the sweep was always 100 ms. The base frequency was 600 Hz and the sweep rates were 0.1, 0.25, 0.5, 1, and 2.5 cycles per second. Therefore, the frequency was increased by 100, 250, 500, 1000, and 2500 Hz. The latency of the N1 decreased as a function of the sweep rate, although the fastest sweep rate of 2.5 cycles/sec evoked latencies that were longer than those evoked by clicks or fast rising tones (Onishi & Davis, 1968). The amplitude of the N1-P2 response increased as the rate of frequency change increased up to 1 cycle/sec after which the amplitude fell slightly. Responses were evoked by upsweeps but not by downsweeps.

Lenhardt (1971) could not demonstrate increases in response amplitude with increases in the magnitude of the frequency change as did Spoor et al. (1969), although he did demonstrate that N1-P2 amplitude increased and N1 latency decreased as the rate of frequency change increased (i.e., ramp duration decreased). The evoked potential was recorded in two subjects with normal hearing from electrodes at the vertex and the left

mastoid. Two kinds of stimuli were used: low frequency stimuli that started at 500 Hz and ramped up to 550 Hz, and high frequency stimuli that started at 2000 Hz and ramped up to 2200 Hz. The initial and final frequency was maintained for five seconds, the duration of the upramp was either 25, 100, 500, or 2000 ms, and the duration of the downramp was 10 ms. The stimuli were presented at 40 and 60 dB SL. The N1-P2 amplitude increased and N1 latency decreased as the rate of frequency change increased. The authors suggested that there was no effect of the magnitude of the frequency change on the N1-P2 amplitude since the lower frequency region with a change of 50 Hz elicited a larger N1-P2 response than the 200 Hz change in the 2000 Hz region. Both low and high frequency stimuli presented at the higher sensation level (60 dB) elicited a larger N1-P2. The N1 latency was shorter for high frequency stimuli than for low frequency stimuli and for stimuli with the higher sensation level.

In an attempt to sort out the confusing relationship between the N1-P2 potential and various changes in stimulus parameters, a series of experiments was conducted by Arlinger and his associates. Arlinger, Jerlvall, Ahren, & Holmgren (1976) used trapezoidally-modulated pure tone stimuli characterized by a base frequency (Fo), rise time (Tr), plateau time (Tp), and fall time (Tf). The total duration of Tr and Tp was 500 ms, and the duration of Tf was 600 ms. An electrode was located at the vertex. Four stimulus parameters were examined: frequency change (ΔF), rise time, base frequency, and sound level. The direction of ΔF (increasing vs. decreasing) did not affect the latency or amplitude of the N1-P2. N1 Latencies decreased with increases in ΔF and increased with increases in the rise time. The N1 latency was determined mainly by the slope of ΔF ($\Delta F/Tr$) except when the rise time was less than 50 ms and ΔF was less than 50 Hz.

Under these circumstances, ΔF determined the N1 latency. The amplitude of the response increased with increasing ΔF and decreased with increases in the rise time. The duration of the plateau had no effect on response amplitudes or latencies. N1 latencies were found to be shortest when the base frequency was 2000 Hz. The larger the change in frequency from the base (in percent), the shorter was the N1 latency. No systematic relationship was found between the base frequency and the amplitude of N1-P2. The N1 latency increased and the N1-P2 amplitude decreased as sound level was reduced from 80 to 20 dB HL. The authors concluded that slope of the frequency change is the most important factor affecting the N1-P2 potential except for ramp durations less than 50 ms and frequency changes less than 50 Hz.

In a subsequent study, Arlinger & Jerlvall (1979) examined amplitude- as well frequency-modulated pure tone signals. The pure tone signal was characterized by its base frequency or amplitude (Fo or Po). The linear frequency or amplitude change had the magnitude ΔF (up to 500 Hz) or ΔP (up to 30 dB) and the duration Tr. After remaining at a plateau, the signal then returned to its base during Tf. The duration of Tr was varied from 10 to 1000 ms and Tf was always 600 ms. For FM stimuli, Fo was 1000 Hz, and various sensation levels (20, 40, 60, 80) were investigated. For AM stimuli, Po was 60 dB HL and the frequency remained at 1000 Hz. Electrodes were placed at the vertex, referenced to the earlobe. Results for amplitude-and frequency-modulated signals were similar. The amplitude of the N1-P2 potential was independent of rise time for small ΔF (less than 100 Hz) until rise time reached 100 ms after which the amplitude of the response decreased with increasing rise time. The influence of rise time appeared at shorter values (20 ms) when ΔF was large (500 Hz). For AM signals, the amplitude of the

N1-P2 potential was independent of rise time until rise time was high (500 ms) and then the amplitude decreased with increases in rise time. N1 latency decreased with increases in ΔF and ΔP and decreased with increases in rise time. The amplitude of the response increased as ΔP increased. No differences were found in the N1-P2 potential between increasing and decreasing frequency glides although lower N1-P2 amplitudes were elicited by decreasing amplitude glides than by increasing amplitude glides. N1 latency increased when the sensation level went below 60 dB and the N1-P2 amplitude decreased monotonically as sensation level decreased from 80 dB to 20 dB. N1 latencies were lowest when the base frequency was 1000 – 2000 Hz and increased when the base frequency dropped below 1000 Hz. N1 latencies increased as the base amplitude decreased. The authors revised their hypothesis regarding what determines the detectability of FM and AM signals. They proposed that at short glide durations, less than 100 - 200 ms, the detection of a frequency or amplitude modulated signal is determined by the total frequency or amplitude change (ΔF or ΔP). As the glide duration increases above 200 ms, the rate of frequency or amplitude change (ΔF/Tr or ΔP/Tr) determines detectability.

Arlinger, Elberling, Bak, Kofoed, & Saermark (1982) extended their findings by recording cortical magnetic fields to the same type of frequency-modulated signals they had employed previously. They confirmed the findings that N1 latencies decreased and amplitudes increased with increases in ΔF. N1 latencies were larger and amplitudes lower for longer rise times and for stimuli presented at low vs. high sensation levels. The rate of frequency change was the most important determiner of N1-P2 potential. Additionally, different dipole locations were found for a 1000 Hz base frequency in the right and left

hemispheres along the antero-superior dimension. Average dipole localization in the right hemisphere was 15 mm anterior to T4 and 7 mm posterior to T3 in the left hemisphere. Larger current dipole moments for this base frequency were found in the left hemisphere although latencies were shorter on average in the right hemisphere.

Finally, Yingling & Nethercut (1983) demonstrated that the N1-P2 potential grew with increases in the size of the frequency shift. Event-related potentials were recorded from 13 adults with normal hearing in response to a wide range of frequency shifts of a constant amplitude tone in four separate experiments. The N1-P2 amplitude increased and N1 latencies decreased as the frequency shift increased. Additionally, they demonstrated that the ratio between the frequencies rather than the absolute frequency shift determines response amplitude. A one-octave shift, from 1000 Hz to 2000 Hz, elicited about the same response amplitude as did a one-octave shift from 200 to 400 Hz.

Several conclusions can be drawn from these studies. First, the N1-P2 potential can be elicited by a change in frequency or intensity that occurs during an ongoing stimulus. The N1-P2 amplitude increases and the N1 latency decreases with increases in the magnitude of the change - ΔF and ΔI (Arlinger et al., 1982; Arlinger & Jerlvall 1979; Arlinger et al., 1976; Spoor et al., 1969). Moreover, FM signals produce response amplitudes that are larger, although with longer latencies than AM signals (Ruhm 1970; Spoor et al., 1969).

The duration of the signal rise (rise time) is related to rate of frequency or amplitude change in that as the rise time increases, the rate of change decreases for a fixed change. The effect of the rise time and rate of frequency or amplitude change on the N1-P2 potential generally points to increases in response amplitude and decreases in response latency as the rise time is decreased (and rate of frequency or amplitude change is

increased) (Arlinger et al., 1982; Arlinger & Jerlvall, 1979; Arlinger et al., 1976; Ruhm, 1970, Lenhardt, 1971). Davis & Zerlin (1966) concluded, however, that there was no relationship between signal rise time and response latency or amplitude although, as noted by Onishi & Davis (1968), their instrumentation was probably not sophisticated enough to demonstrate the effect. Additionally, when N1 latencies were measured from the behavioral threshold measured for each rise time, latency remained constant with increases in rise time (Ruhm & Jansen, 1969). Arlinger and his colleagues (1979) propose that the rate of frequency or amplitude change determines the detectability of the signal modulation for rise times greater than 100-200 ms. The detectability of the signal modulation is determined by the total frequency or amplitude change for rise times below this critical duration of 100-200 ms.

There is rather good agreement among the findings that response latency and amplitude are not affected by the duration of the signal plateau except possibly when it very short, below about 30-50 ms (Arlinger et al., 1976; Skinner & Jones, 1968; Onishi & Davis, 1968).

Most investigations have found that the direction of the frequency change has no effect on the response amplitude or latency of the N1-P2 potential (Arlinger & Jerlvall, 1979; Arlinger et al, 1976; Clynes, 1969), although, Kohn et al., (1978) demonstrated that responses to ramp-modulated tones that increased in frequency were larger than those that decreased in frequency. Amplitude-modulated signals elicit larger responses when the amplitude is increased rather than decreased (Arlinger & Jerlvall, 1979; Clynes, 1969).

The amplitude of the N1-P2 potential decreases and the N1 latency increases as the intensity of the signal is decreased (Arlinger et al., 1982; Arlinger & Jerlvall, 1979, Arlinger et al., 1976; Onishi & Davis, 1968, Skinner & Jones, 1968). The latency of the N1 response is generally lowest when the base frequency is around 1000 to 2000 Hz and N1 latency increases when the base amplitude decreases. No systematic relationship has been demonstrated between base frequency and the amplitude of the response (Arlinger et al., 1982; Arlinger & Jerlvall, 1979; Arlinger et al., 1976).

ii. N1-P2 Potential to Change in an Ongoing Speech Stimulus

There is a very small body of literature demonstrating the N1-P2 potential to a change in an ongoing speech stimulus (Kaukoranta et al., 1987). Recent investigations by this author and colleagues (Ostroff, Martin, & Boothroyd, 1998) have been able to replicate the findings of Kaukoranta et al., 1987. Cortical potentials N1 and P2 were obtained in eight adults with normal hearing. Three naturally-produced speech stimuli were used: i) the syllable [sei], ii) the sibilant [s], extracted from the syllable, iii) the vowel [ei] extracted from the syllable. The isolated sibilant and vowel preserved the same time relationships to the sampling window as they did in the complete syllable. Clear responses were observed to both the sibilant and the isolated vowel. Although the response to the [s] was weaker than that to [ei], both had N1 and P2 components with latencies, in relation to sound onset, appropriate to cortical onset potentials. The vowel onset response was preserved in the response to the complete syllable, though with reduced amplitude.

In a subsequent unpublished study, Ostroff, Martin & Boothroyd demonstrated that N1-P2 complex could be produced by a formant change in an ongoing quasi-synthetic vowel stimulus. A single cycle from each of the vowels /u/, /i/, and /U/ was extracted from digitally recorded words, and iterated over 150 ms. The quasi-synthetic vowels were equalized for fundamental frequency and rms amplitude. They were then concatenated to produce two stimuli; [ui] and [uU], each with a duration of 300 ms. These stimuli were selected because they permit the evaluation of electrophysiologic responses to a spectral change in the middle of an ongoing stimulus based on an F1 ([uU]) and an F2 ([ui]) contrast. The N1-P2 complex was elicited by the spectral change contained within both the F1 and F2 contrasts, although the response to the F2 contrast was larger. This could be accounted for by the fact that the F2 difference between /u/ and /i/ is much larger than the F1 difference between /u/ and /U/. The latency of the response was approximately 150 ms in relation to the onset of the change for both types of formant change.

In a follow-up study, Martin & Boothroyd (1999) demonstrated that the N1-P2 complex could be elicited by stimuli that changed in periodicity while the rms amplitude and spectral envelope were held constant. The stimuli consisted of a noise band and a tonal complex (Fo = 100 Hz) band pass filtered, and matched for rms level at 80 dB SPL. The sounds were then concatenated to produce two test stimuli that changed in the middle (noise-tone, tone-noise), and two control stimuli that did not (noise-noise, tone-tone). The N1-P2 complex was elicited by the change contained in the two test stimuli. The response was not elicited by the two control stimuli, indicating that the N1-P2 change component is not an artifact of the concatenation process.

In summary, it is clear that the N1-P2 complex, which we have termed the Acoustic Change Complex (ACC), can be elicited by a change in frequency or intensity that occurs in an ongoing simple stimulus. Both the amplitude and the latency of the response are related to the magnitude and rate of the change. The ACC can also be elicited by changes occurring in ongoing complex stimuli including speech. The changes in speech stimuli can be of spectrum only (constant rms) or periodicity only (constant spectrum and rms). What is not clear however, is the sensitivity of this response. It is not known whether the response can be demonstrated to changes that are close to the limits of discriminability as assessed behaviorally. Moreover, it would prove useful to determine whether this response could be elicited by minimal changes, of the kind that occur in speech sounds (i.e., changes in formant frequencies, smaller than those used in previous research by the author and colleagues), which are likely to elicit a change of vowel percept. Given that the ACC can be elicited by the types of changes that occur in ongoing speech, it is potentially a clinically viable measure of speech perception capacity, in a gross sense, at the very least.

As indicated earlier, the goals of the two-part study to be described were:

i) To measure the ACC elicited by changes in the frequency of the second formant within a vowel stimulus as a function of the magnitude of the F2 shift (spectral change).

ii) To determine the least amount of spectral change (i.e., threshold) within a vowel needed to elicit a reliable ACC.

iii) To measure behavioral discrimination of changes in the frequency of the second formant of a synthetic vowel stimulus as a function of the magnitude of the F2 shift (spectral change).

iv) To determine the least amount of spectral change (i.e., threshold) within a vowel needed to elicit a reliable behavioral response.

v) To determine the relationship between the threshold of the ACC potential and the threshold on the behavioral task.

CHAPTER 3

STIMULI

Stimuli containing a spectral change in the middle of a synthesized vowel (/u/) elicited both behavioral and electrophysiologic responses in the present experiment. As shown in Figure 3.1, the synthetic vowel was created by generating a harmonic series composed of 35 harmonics with a fundamental frequency of 150 Hz. The duration of the vowel was 400 ms with a rise and fall time of 13 ms. The formants were produced using a formant filter whose shape (in log frequency) consisted of a half sine wave defined by five points. The 6 dB down point of this filter had a bandwidth of 300 Hz. The formant filter is displayed in Figure 3.2. As shown in Figure 3.1 the frequency of F1 was centered at 300 Hz, (i.e., the second harmonic) and the frequency of F3 was centered at 3000 Hz, (i.e., the 20^{th} harmonic). The frequencies of F1 and F3 remained fixed. The frequency of F2 was initially centered at 1050 Hz, creating the "standard" stimulus. The amplitude of F1 was 10 dB and 20 dB above the amplitudes of F2 and F3 respectively. The amplitude of F3 was 30 dB above the amplitude of all harmonics between F1 and F3. The decrease of amplitude with increase of frequency was intended to simulate the natural amplitude relationship between the formants that occurs in natural speech. The amplitudes of all three formants remained fixed.

In order to produce several stimuli each having a different F2 center frequency, small shifts in the frequency of the second formant were created. For example, the first change in F2 (from the standard of 1050 Hz) was accomplished by shifting the center frequency of F2 by $1/16^{th}$ of the harmonic spacing. Therefore the center frequency of F2 became 1050 Hz plus 150 Hz (the fundamental) divided by 16 which is equal to 1059 Hz. The

other changes in F2 are shown in Table 3.1 and are displayed in Figure 3.3. The largest shift in F2 frequency that could be accomplished before F2 overlapped F3 was 1200 Hz, which corresponded to an F2 center frequency of 2250 Hz.

The standard stimulus was then concatenated (with overlap) with each of the stimuli containing higher F2 frequency values to create thirteen stimuli that contained a spectral change at temporal midpoint (393 ms). The overlap used in the concatenation process was of two cycles of the waveform (approximately 14 ms). This overlap of the offset of the first stimulus with the onset of the second stimulus was intended to produce a smooth transition and to avoid spectral splatter at the point of concatenation. The two concatenated portions of the stimuli were equalized for rms intensity. The results of the concatenation process are illustrated in Figure 3.4 which shows wide band and narrow band spectra of the stimulus containing a 75 Hz spectral change. Figure 3.5 shows details of the stimulus waveform in the region of transition for an F2 change of 300 Hz. The rms amplitude of the waveform (measured over 100 ms) fluctuates by less than 0.1 dB during the course of the transition.

The standard stimulus was also concatenated with itself to produce a control stimulus with no spectral change. The total duration of each concatenated stimulus was 786 ms. A total of 14 stimuli (control and 13 degrees of spectral change) were originally created in order to ensure a sufficient number of stimuli for detection of threshold. The stimuli selected to elicit electrophysiologic responses for the present study, indicated by asterisks in Table 3.1, were based on results from a pilot study indicating thresholds around 75 Hz. In addition to the control stimulus, these stimuli involved spectral changes of 9, 19, 38, 75, 150, 300, and 600 Hz.

The synthetic vowel stimuli were 16 bit, binary, signed integer files with a sampling rate of 22050 Hz. The stimuli were created using the Dadisp digital signal processing software package.

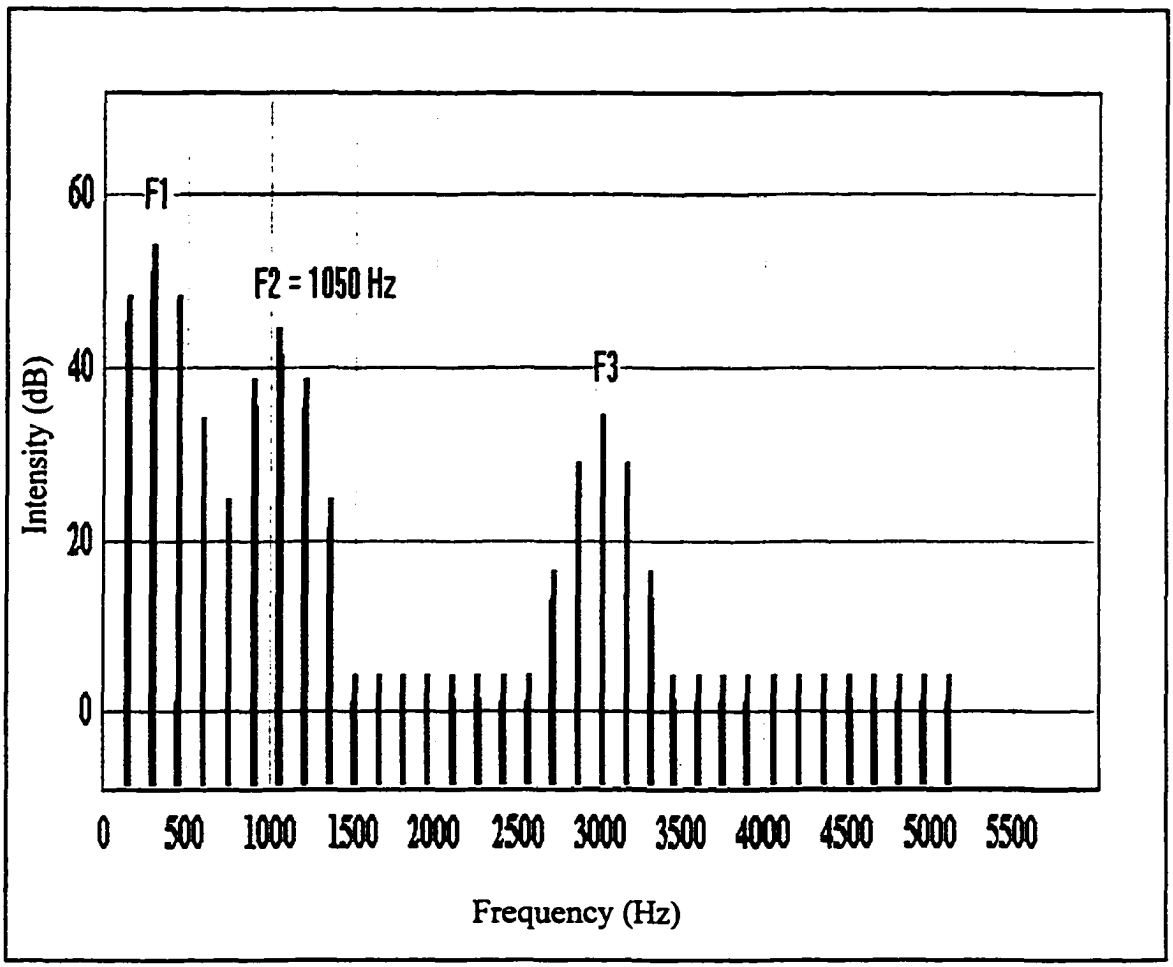


Figure 3.1. Harmonic series with a fundamental frequency of 150 Hz. F1, F2, and F3 centered respectively at 300, 1050, and 3000 Hz.

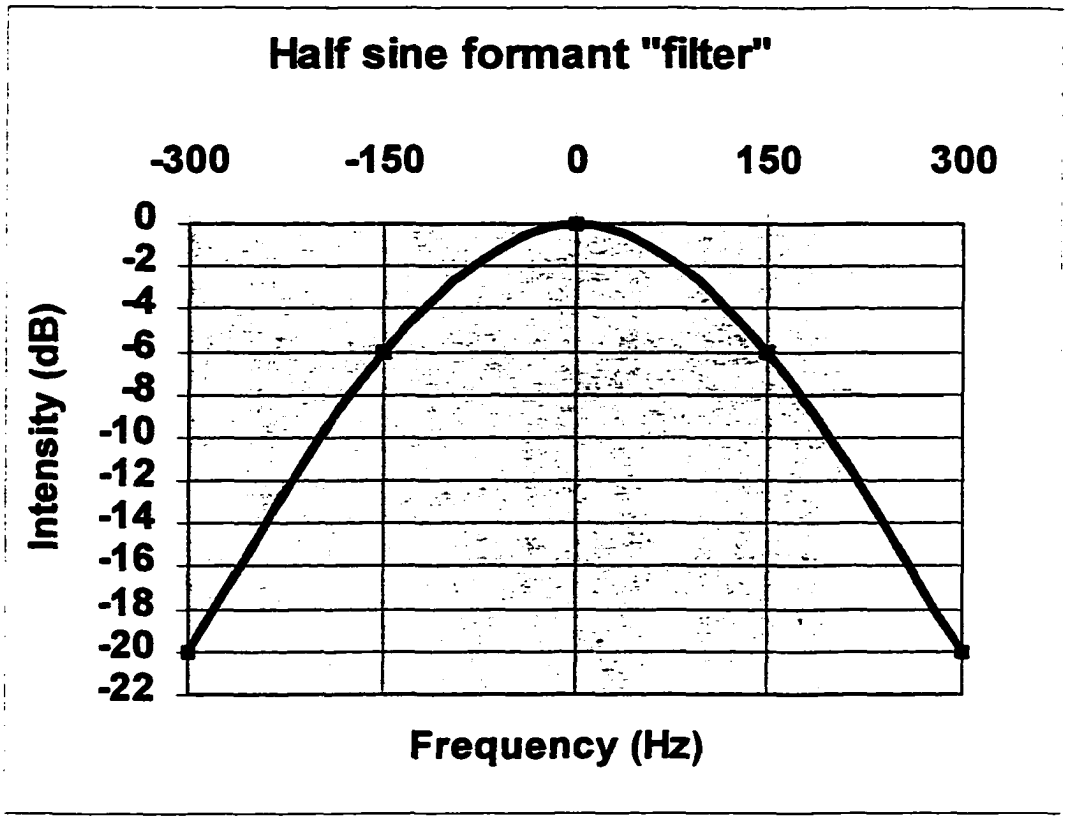


Figure 3.2. Formant filter composed of a half sine wave defined by 5 points. The 6 dB bandwidth is 300 Hz.

Table 3.1. Stimuli created by shifting the frequency of F2 in successive steps from the standard stimulus (1050 Hz). Asterisks denote stimuli selected for the current study.

Stimulus Number	Shift Size (re: 1 harmonic)	Shift Size (Hz)	Center Frequency (Hz)	
*1	0.0625	9	1059	(150 * 0.0625) + 1050
*2	0.125	19	1069	(150 * 0.125) + 1050
*3	0.25	38	1088	(150 * 0.25) + 1050
*4	0.5	75	1125	(150 * 0.5) + 1050
5	0.75	113	1163	(150 * 0.75) + 1050
*6	1	150	1200	(150 * 1) + 1050
*7	2	300	1350	(150 * 2) + 1050
8	3	450	1500	(150 * 3) + 1050
*9	4	600	1650	(150 * 4) + 1050
10	5	750	1800	(150 * 5) + 1050
11	6	900	1950	(150 * 6) + 1050
12	7	1050	2100	(150 * 7) + 1050
13	8	1200	2250	(150 * 8) + 1050

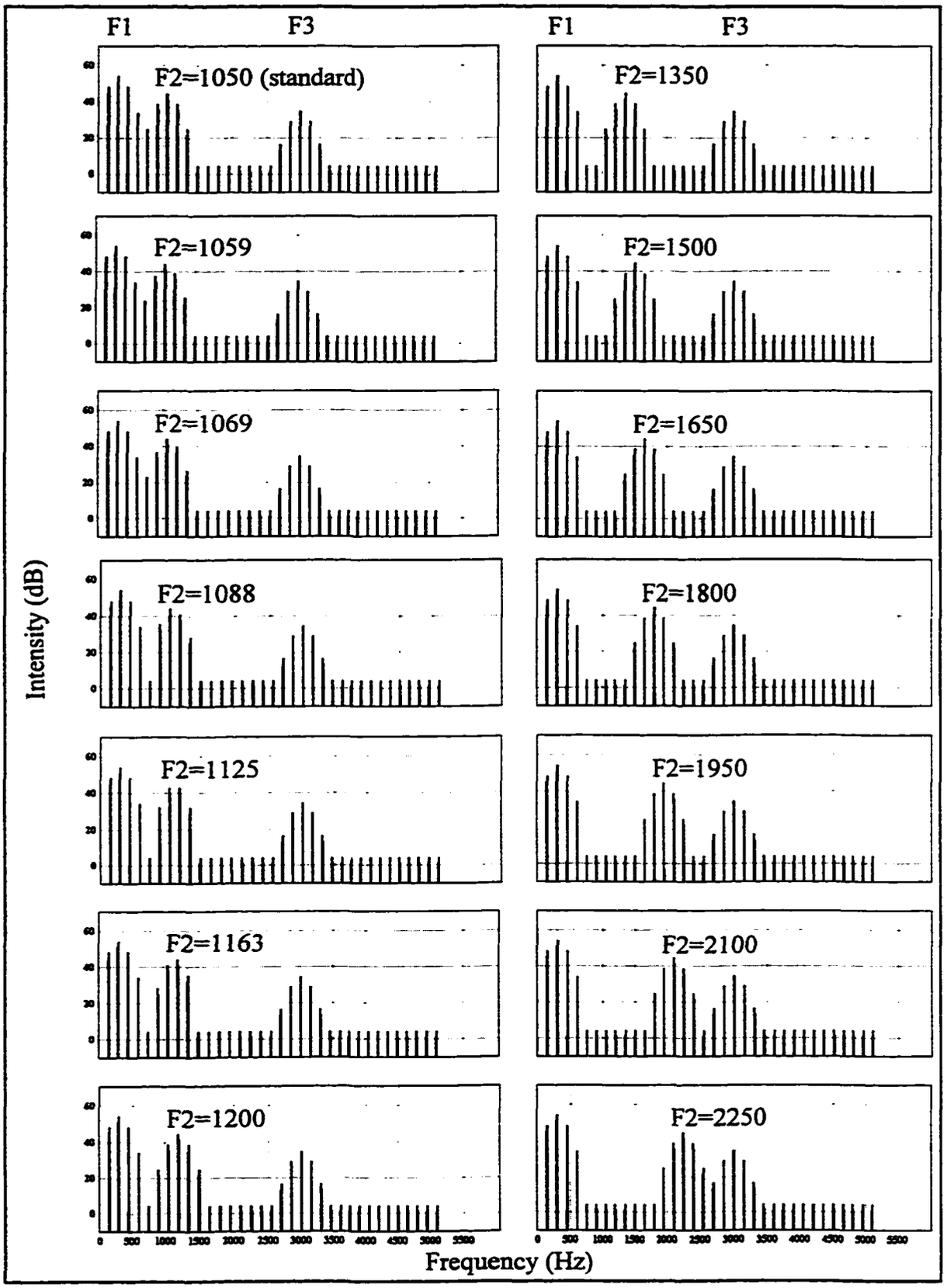


Figure 3.3.
The standard stimulus and 13 degrees of shift in the frequency of the second formant.

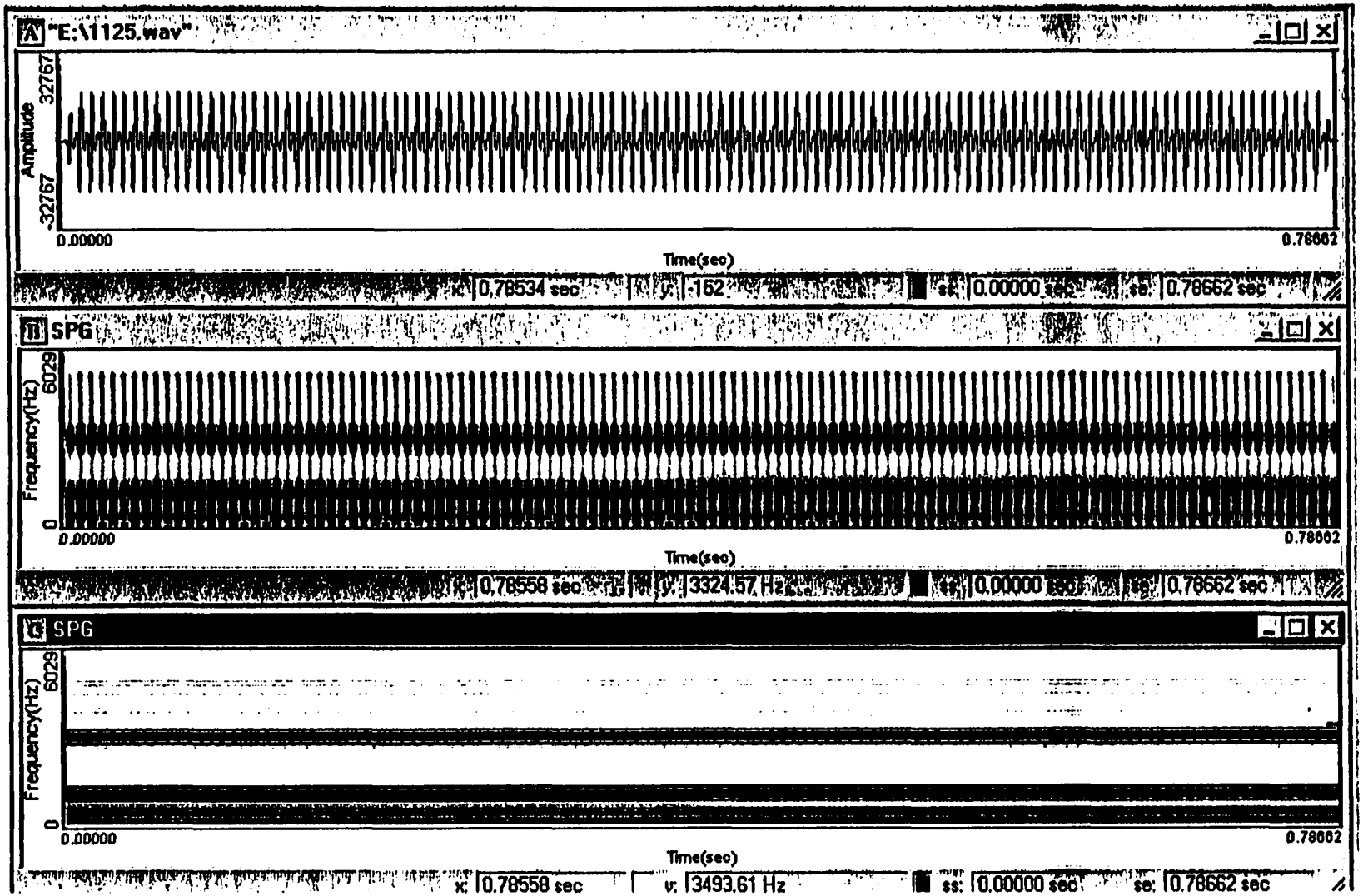


Figure 3.4. Upper trace: acoustic waveform of 75 Hz spectral change. Middle trace: wide band spectrogram (323 Hz) of 75 Hz spectral change. Lower trace: narrow band spectrogram (63 Hz) of 75 Hz spectral change.

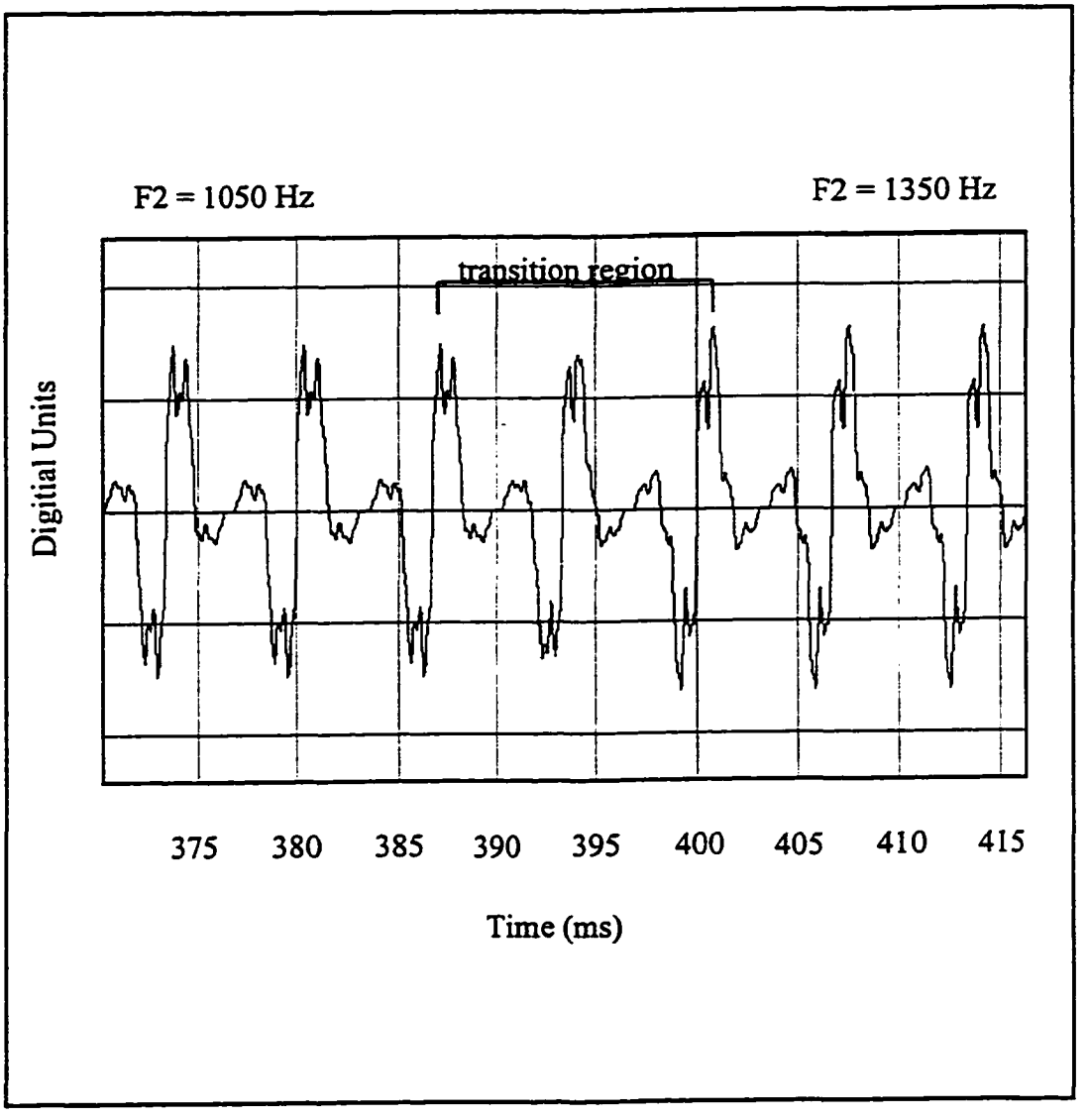


Figure 3.5. The standard stimulus concatenated with a stimulus having an F2 frequency of 1350 Hz. The spectral change is 300 Hz.

CHAPTER 4

ELECTROPHYSIOLOGIC STUDY

I. Methods

Subjects

Event-related potentials (ERPs) were recorded from eight adult volunteers (four female, four male) between the ages of 18 and 45 years. Four of the eight subjects were non-native speakers of English. All subjects passed an audiologic screening at 25 dB HL for frequencies between 250 through 8000 Hz, and had no history of neurological impairment.

Electrophysiologic Measures

Event-related potentials were recorded using the Neuroscan evoked potentials system by Neurosoft, Inc. This system is made up of two components: the Stim computer and the Scan computer. The Stim computer enables the user to design the experimental protocol (i.e., number of stimulus presentations, inter-stimulus-interval, presentation level, randomization of stimuli) and presents the stimuli to the subject. The Scan computer records the electrical activity that is picked up from the electrodes placed in various locations on the scalp. The Scan computer is connected to a series (8 channels in this experiment) of AC amplifiers that amplify the electrical brain activity. The scan computer is also capable of handling the recorded activity through sophisticated processing algorithms. For example, the Scan computer manages the data in such ways as filtering, averaging, sorting, correcting baseline drift, and removing excessive artifact from eye movements.

Stimuli were delivered through the Stim computer using a digital-to-analog rate (D/A) of 22050 Hz (10 000 Hz anti-imaging filter, 60 dB/octave), and presented binaurally through EARlink 3A insert earphones at 80 dB SPL. The onset-to-onset (SOA) interval was 3000 ms. Each of the eight stimuli was presented in two trial blocks, yielding a total of 16 trial blocks. For each trial block, the same stimulus was presented 400 times. Therefore the total number of stimulus presentations for each stimulus condition was 800. Presentation order of the trial blocks was randomized across stimulus conditions and subjects. Each trial block lasted 20 minutes and the total actual testing time was approximately five and one half hours divided into two test sessions.

Calibration

Calibration of the speech sounds and electroencephalography (EEG) amplifiers was accomplished before each recording. The speech sounds were calibrated in dB peak rms sound pressure level (SPL) using a Bruel and Kjaer sound level meter (model 2235), Bruel and Kjaer "ear simulator" (model 4157) for measurements using insert earphones, and a Bruel and Kjaer ½" microphone (model 4134). The calibration signal was the control stimulus that did not contain a spectral change. Amplifier channels were calibrated by sending a 1volt sine wave generated by the Neuroscan Scan software through the Grass amplifiers and back into Scan. All channels were calibrated within +/- 0.05 volts of 1 volt.

Electrophysiologic Recordings

ERPs were recorded from surface electrodes at five scalp locations (Fz, Cz, Pz, A1, and A2) with the nose as a reference (Vaughan & Ritter, 1970). An electrode at Fpz served as a ground. Vertical eye movements were recorded from two electrodes placed above and below the right eye. Electrode impedances were maintained below 5 Kohms. EEG signals were amplified 20 000 times, with the exception of the eye channel which was amplified 5 000 times, filtered (0.1-100 Hz, 6 dB/octave), and digitized (301 Hz A/D rate, 512 points per epoch). The recording window of 1601 ms included a 100 ms pre-stimulus period. Data were stored for off-line processing.

Procedure

Subjects were tested in an acoustically treated and electrically shielded room while seated in a recliner. Subjects watched a captioned videotape of their choice and were instructed to ignore the stimuli and to remain as still and as quiet as possible. Breaks between trial blocks were provided to subjects upon request. It should be noted that this experiment, unlike a P300 experiment, was designed to ascertain ACC thresholds in test conditions without attention to the stimuli. This test procedure was developed in order to achieve one of the ultimate goals of this research, which is to develop an electrophysiologic test of speech discrimination capacity to be used clinically where attention to task may be inappropriate or impossible to demand from patients.

Data Analysis

Single-trial ERP waveforms from the individual subjects were baseline corrected (-100 to 1601 ms), digitally filtered (0.1-30 Hz, 12 dB/octave), and artifact rejected (+/- 100 μV). Averaged waveforms were then created for each trial block for every subject. Additionally, averaged waveforms were created for the mean of the two trial blocks for each of the eight stimuli. Averaged waveforms elicited by the control stimulus (containing no spectral change) were compared to averaged waveforms elicited by stimuli containing a spectral change in order to identify the ACC.

II. Results

Response Waveforms

For every subject, the average voltage, collapsed across two trial blocks, was calculated as a function of time for each stimulus (i.e., each degree of spectral change). The individual waveforms were then averaged to create group waveforms. The averaged group waveforms, taken from electrode site Cz at each spectral change condition, are shown in Figure 4.1. Note that positive is plotted up in this and in all subsequent figures. Responses from the group waveforms display a large N1-P2 onset response elicited by all stimuli on the order of 10 μvolts peak to peak. The onset response is followed by a sustained negativity that appears to stabilize by 300 ms. The ACC is visible in the group waveforms down to a spectral change of 75 Hz. The ACC elicited by the largest (600 Hz) spectral change condition is approximately 3.4 μvolts peak to peak. There is an offset response in the group waveforms beginning at around 850 ms.

The scalp distributions of the onset N1-P2 response and the ACC elicited by the 600 and 75 Hz spectral change conditions are shown in Figure 4.2 for the group waveform.

The data from Fz, Cz, Pz, A1 and A2 are displayed. Both the onset response and the ACC are largest in amplitude at Cz, with diminishing amplitudes at Fz and Pz. There is however, only a slight difference in amplitude between Cz and Fz for the ACC elicited by the 75 Hz spectral change. The polarity of both the onset response and the ACC invert at the earlobe electrode sites in the waveforms elicited by both change conditions.

The averaged waveforms from every subject at each spectral change condition are shown in Appendix B. Response waveforms from all subjects display a clear N1-P2 response at onset on the order of about 10 μvolts peak to peak. In many waveforms the N1-P2 onset response is followed by a negative dip at around 300 ms. The ACC, which is clearly evident in all response waveforms to the largest spectral change condition of 600 Hz, is characterized by the appearance of an N1 component at around 480 ms. The P2 component, which is more variable than the N1 in terms of latency across subjects, peaks at around 580 – 600 ms. The ACC to the largest spectral change conditions (i.e., 150, 300, and 600 Hz) is on the order of about 4 μvolts peak to peak. Under these conditions, the magnitude of the ACC is roughly half the amplitude of the onset response. The amplitude of the ACC decreases with decreasing magnitude of the F2 shift. The ACC becomes difficult to detect (by visual inspection) for small degrees of F2 shift. Response identification and threshold estimation will be discussed in greater detail later. Response waveforms from subject 2, in Appendix B.2, elicited by the three largest spectral change conditions contain very good examples of the ACC and are reproduced in Figure 4.3. In contrast, the presence of an ACC in the response waveforms from subject 8 (reproduced in Figure 4.4) is questionable. Note that although it appears that the ACC is present in the waveform elicited by the largest spectral change condition because of a negative-positive

swing beginning around 420 ms, this response pattern is also present in the waveform elicited by the control condition. Responses from this subject were discarded due to their uncertain nature and the group waveforms of Figure 4.1 were based on the remaining seven subjects.

Response Identification and Threshold Estimation

Two approaches were used to determine whether the acoustic change complex (ACC) was present in the individual and group waveform to a particular degree of spectral change. These approaches include

i) response identification based on overall waveform morphology and rms amplitude, and

ii) threshold estimation based on a curve-fitting and interpolation technique.

These approaches were devised in an attempt to formalize decisions based on visual inspection of the response waveforms. These objective approaches were used to estimate the least amount of spectral change that can elicit an observable ACC (i.e., a threshold for elicitation of ACC).

i. Response Identification Based on Waveform Morphology and Amplitude

The basic assumption underlying the first approach to identifying the ACC is that the ACC has the same waveform morphology as does the onset N1-P2 potential. The onset N1-P2 potential was therefore used as the "template" against which all responses believed to be an ACC were compared.

The latencies, re stimulus onset, of the maximum values of N1 and P2 respectively were determined. A temporal window, based on the group waveform elicited by the 600

Hz spectral change condition, was chosen from 70 to 233 ms re stimulus onset, to include the maxima for all subjects as determined by the values chosen for N1 and P2. A second temporal window was chosen to have the same temporal relationship to the onset of the change – i.e., 70 to 233 ms post onset of change (or 463 to 626 ms re onset of stimulus). For each degree of spectral change a Pearson product-moment correlation coefficient was calculated for the section of waveform in the two temporal windows. This internal correlation gives a measure of the similarity of morphology between the onset response and the possible change response.

A second correlation was calculated for the section of waveform in the "change" window and the corresponding section of the waveform for the next higher amount of change (i.e., the change waveform for the 9 Hz change was correlated with that for the 19 Hz change, which was correlated with that for the 38 Hz change and so on). This adjacent correlation gives a measure of the similarity of morphology between two waveforms differing in terms of one step of change magnitude.

Finally, a standard deviation (i.e., rms amplitude) was calculated for the section of waveform in the "change" window. This standard deviation gives a measure of the magnitude of voltage fluctuation about the mean and is assumed to reflect the combined effect of the true change response and uncorrelated noise. This procedure is illustrated in Figure 4.5.

The three measures so obtained (calculations achieved in Microsoft Excel and Statistica by Statsoft, Inc.) were used to decide whether an apparent response was most likely to be due to noise alone or whether it could be assumed to represent noise plus a

genuine ACC. For the waveform at a particular level of spectral change to be accepted as a genuine response the data had to meet the following three criteria:

(a) The adjacent correlation had to be above 0.5 ($p<0.0001$ for a one-sided test with 50[1] observations, $p<0.0002$ for a two-sided test with 50 observations).

(b) The internal correlation had to be above 0.5.

(c) The standard deviation had to be higher than the average standard deviation for all lower amounts of spectral change.

Note that these criteria were chosen to reflect the kinds of decisions that were being made by visual inspection of the response waveforms.

Starting at the largest spectral change condition, the above criteria were applied to the waveforms for successively smaller changes. When one criterion was failed, the waveform at that level and for all smaller amounts of spectral change was assumed most likely to represent noise alone. The results of this procedure are shown in Table 4.1. When threshold values from all subjects are averaged together, the smallest spectral change that yields a reliable ACC based on these response criteria is 129 Hz with a range of 75 to 300 Hz. This is not to be confused with the threshold estimate based on the group waveform (i.e., an average of all individual subject waveforms), which is 75 Hz. Individually, subjects 2, 3, 4, and 7 produce a reliable ACC down to 75 Hz. An ACC is likely elicited in subjects 5 and 6 when presented with a spectral change on the order of 150 Hz and subject 1 produces an ACC when the spectral change reaches 300 Hz.

[1] Temporal window of 163 ms (from 70 to 233 ms post stimulus/change onset) divided by 3.3 ms per point

ii. Threshold Estimation Based on Curve-fitting and Interpolation

The second approach to threshold estimation was based on waveform amplitude alone. The standard deviations (rms amplitude) from all subjects and the group mean for all "change" windows were calculated. The results were plotted as a function of F2 frequency shift and the data were fit to a sigmoid transition function. The function is defined by the following equation:

$$y = a + (b-a)/(1+e^{-(x-c)/d}) \quad \text{.....(1)}$$

where:

y = voltage
a = lower asymptote
b = upper asymptote
e = base of natural logarithms
x = log(F2 change)
c = x value for midpoint of transition
d = an inverse measure of slope

The resulting curves are displayed in Figure 4.6. The transition function assumes that rms amplitude rises from a lower asymptote – representing the rms amplitude of noise alone, to an upper asymptote – representing a maximum response amplitude, following a symmetrical cumulative Gaussian function. Threshold is defined as the value of frequency change for which the ACC rms amplitude (without contamination by noise) is equal to the rms noise amplitude. Assuming power summation, the point at which the rms amplitude – interpolated from this function - equals the estimated noise amplitude multiplied by the square root of two (1.414) is defined as the threshold for the practical identification of the ACC. To explain further:

First we assume that the variance in the window of interest is equal to the sum of the variances due to noise and to the underlying ACC. Thus:

$$\sigma^2_{s+n} = \sigma^2_s + \sigma^2_n \quad (2)$$

where σ_{s+n} = rms amplitude measured

σ_s = rms amplitude of ACC

σ_n = rms amplitude of noise

$$\text{or } \sigma_{s+n} = \sqrt{\sigma^2_s + \sigma^2_n} \quad (3)$$

Given that $\sigma_s = \sigma_n$ in the definition of threshold

$$\sigma_{s+n} = \sqrt{2 * \sigma^2_n} \quad (4)$$

$$\sigma_{s+n} = \sqrt{2} * \sigma_n \quad (5)$$

The estimated thresholds for all subjects and for the group waveform obtained by this technique are displayed in Figure 4.6. The values of these thresholds are also included in Table 4.2. The mean threshold value obtained by averaging individual threshold values from all subjects is 67 Hz with a range of 24 to 131 Hz. The estimated threshold based on the group waveform is 28 Hz. We can see that the threshold values based on the curve-fitting technique correspond fairly well with the smallest spectral change that was determined to yield an ACC as defined by the waveform morphology and amplitude criteria. The only exception is seen in subject 6 where there is a large discrepancy between the threshold estimate of 39 Hz based on the curve-fit, and 150 Hz based on waveform morphology and amplitude. In general, the estimate of threshold based on the curve-fitting technique yields a lower frequency value than can be obtained by applying the waveform morphology and amplitude criteria. For example, according to the waveform morphology and amplitude criteria, the smallest spectral change that elicited an ACC in subject 2 is 75 Hz. According to the threshold estimate based on the curve-

fitting technique, an ACC can be elicited by spectral changes down to 69 Hz. It is probable that the smallest frequency change that yields an observable ACC in individual subjects lies somewhere between the frequency value determined based on the curve-fit values (average of 67 Hz) and that determined by the more conservative waveform morphology and amplitude criteria (average of 129 Hz). From the group waveform, however, it is clear that an F2 change in the region of 28 to 75 Hz elicits an observable ACC.

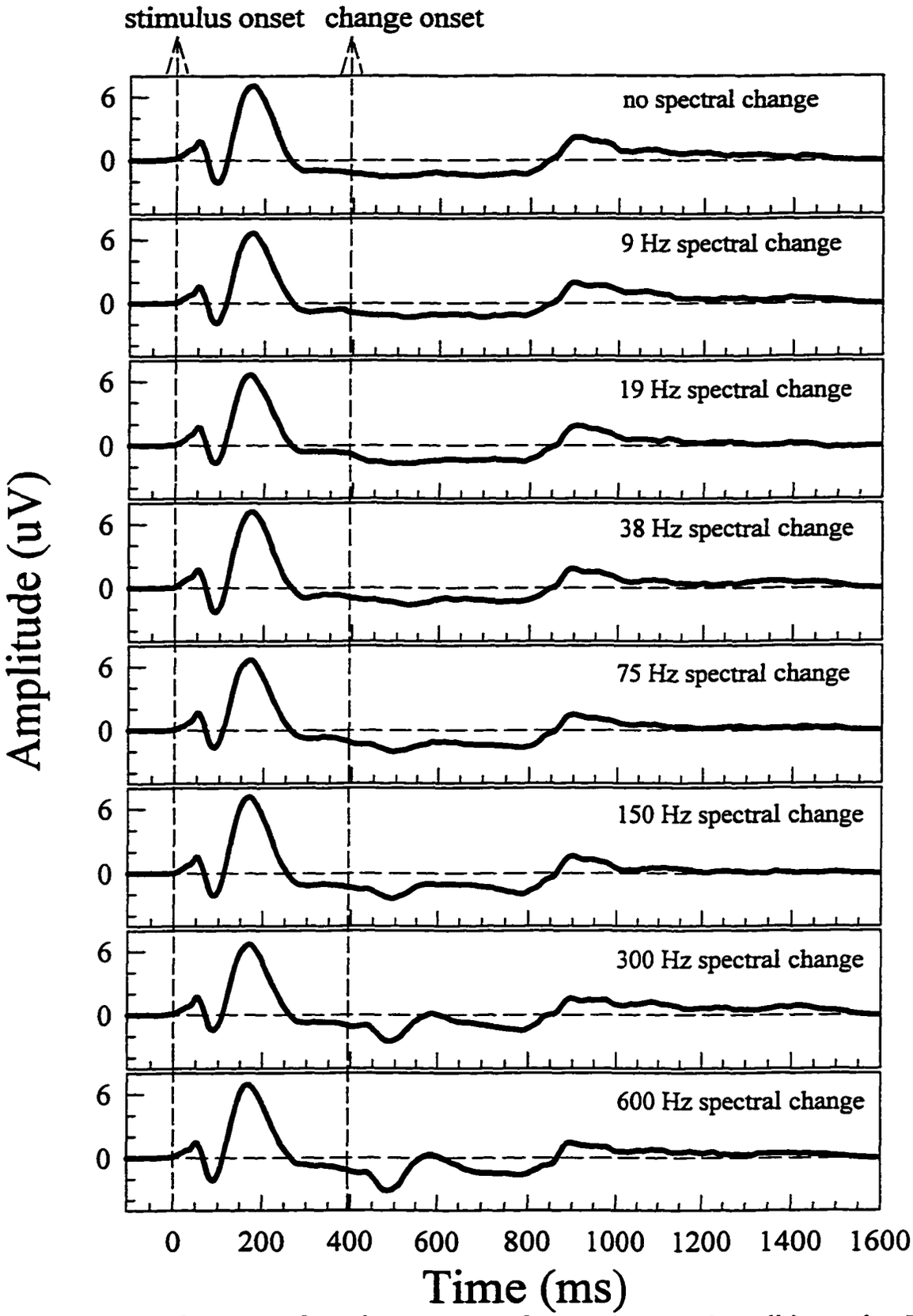


Figure 4.1. Responses from the group waveform to the control condition and to 7 degrees of spectral change taken from electrode site Cz. Positive is plotted up.

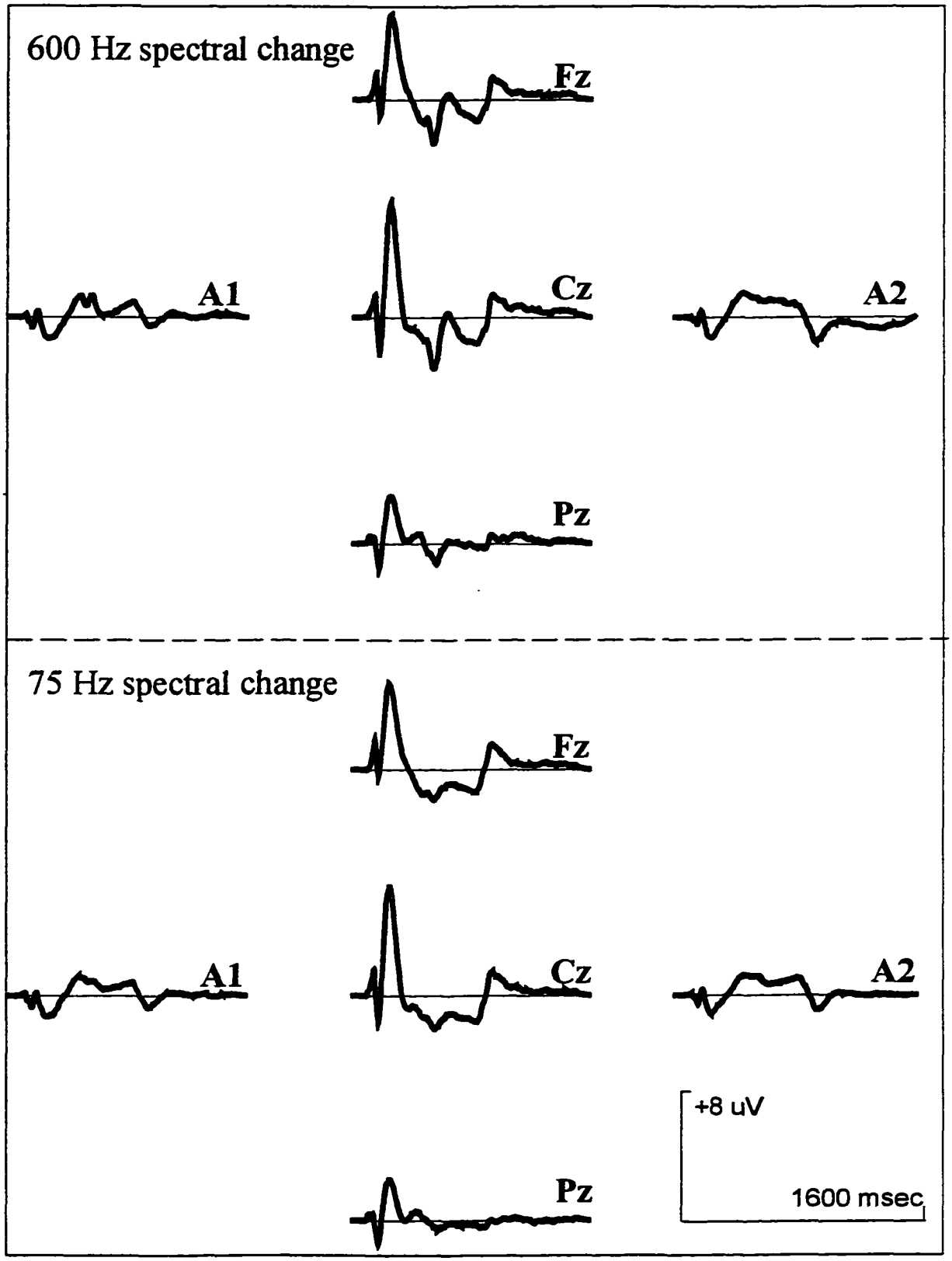


Figure 4.2. Scalp distribution for the group waveforms elicited by the 600 Hz (top) and 75 Hz (bottom) spectral change conditions.

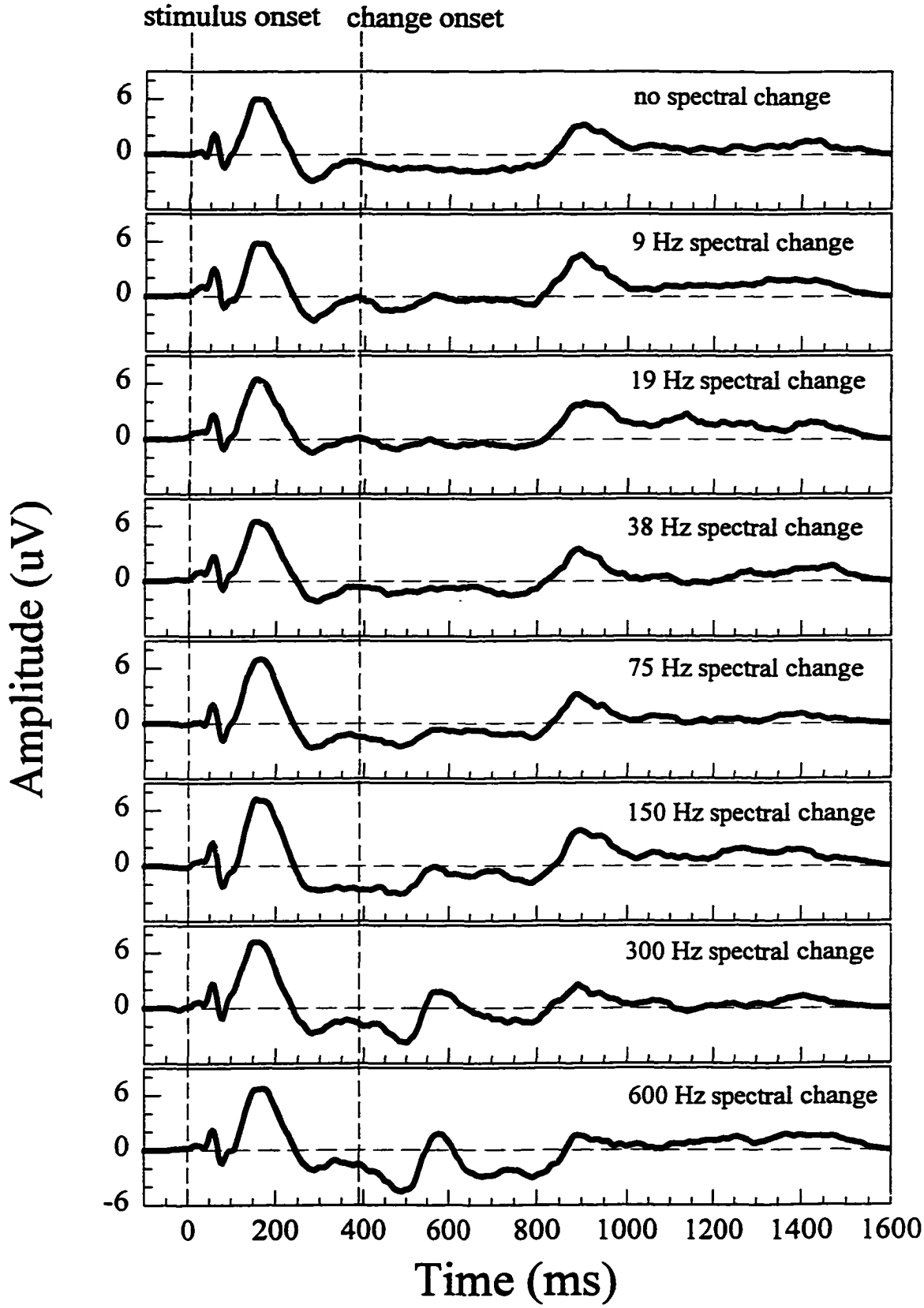


Figure 4.3. Responses from subject 2 to the control condition and to 7 degrees of spectral change taken from electrode site Cz.

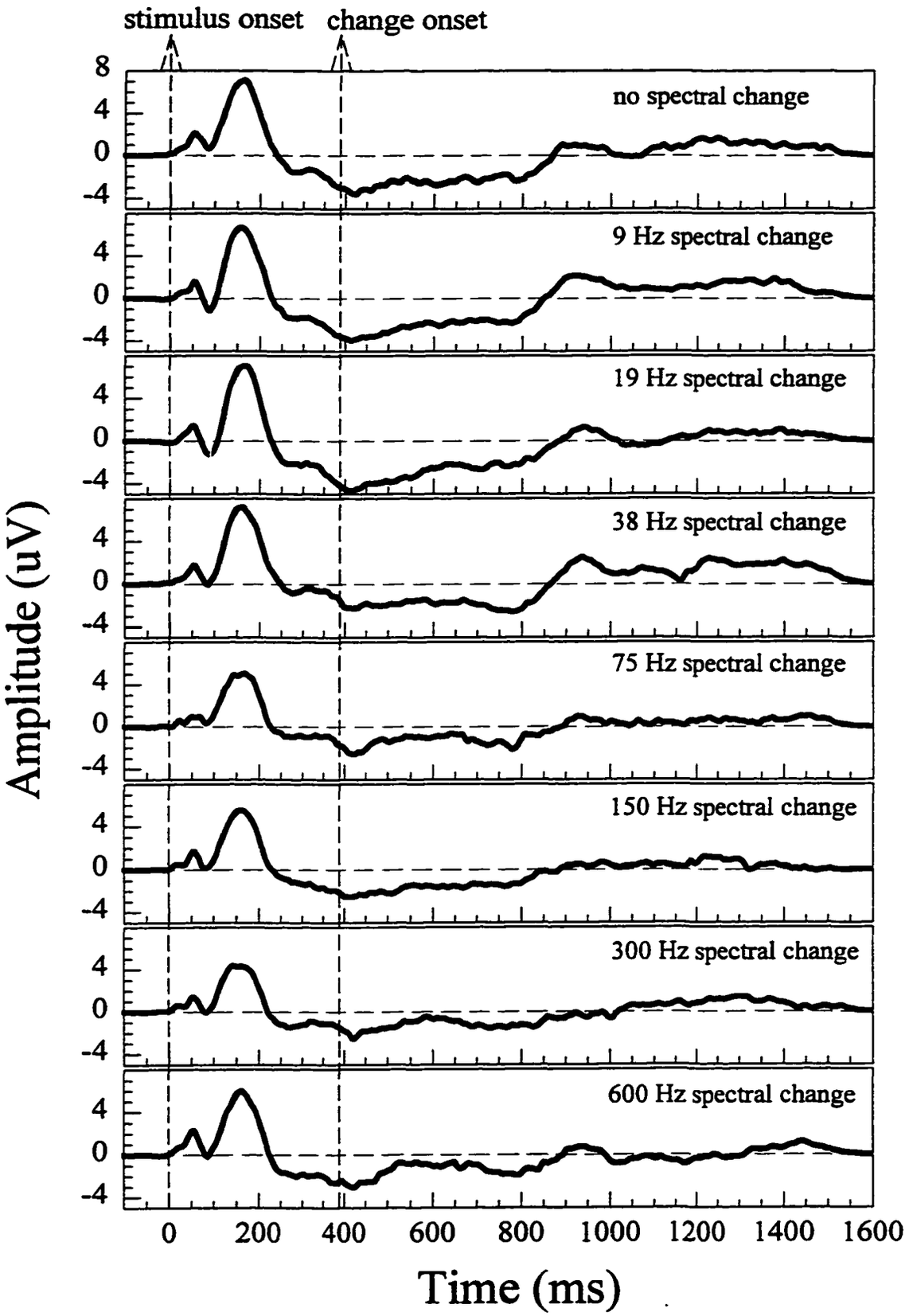


Figure 4.4. Responses from subject 8 to the control condition and to 7 degrees of spectral change taken from electrode site Cz.

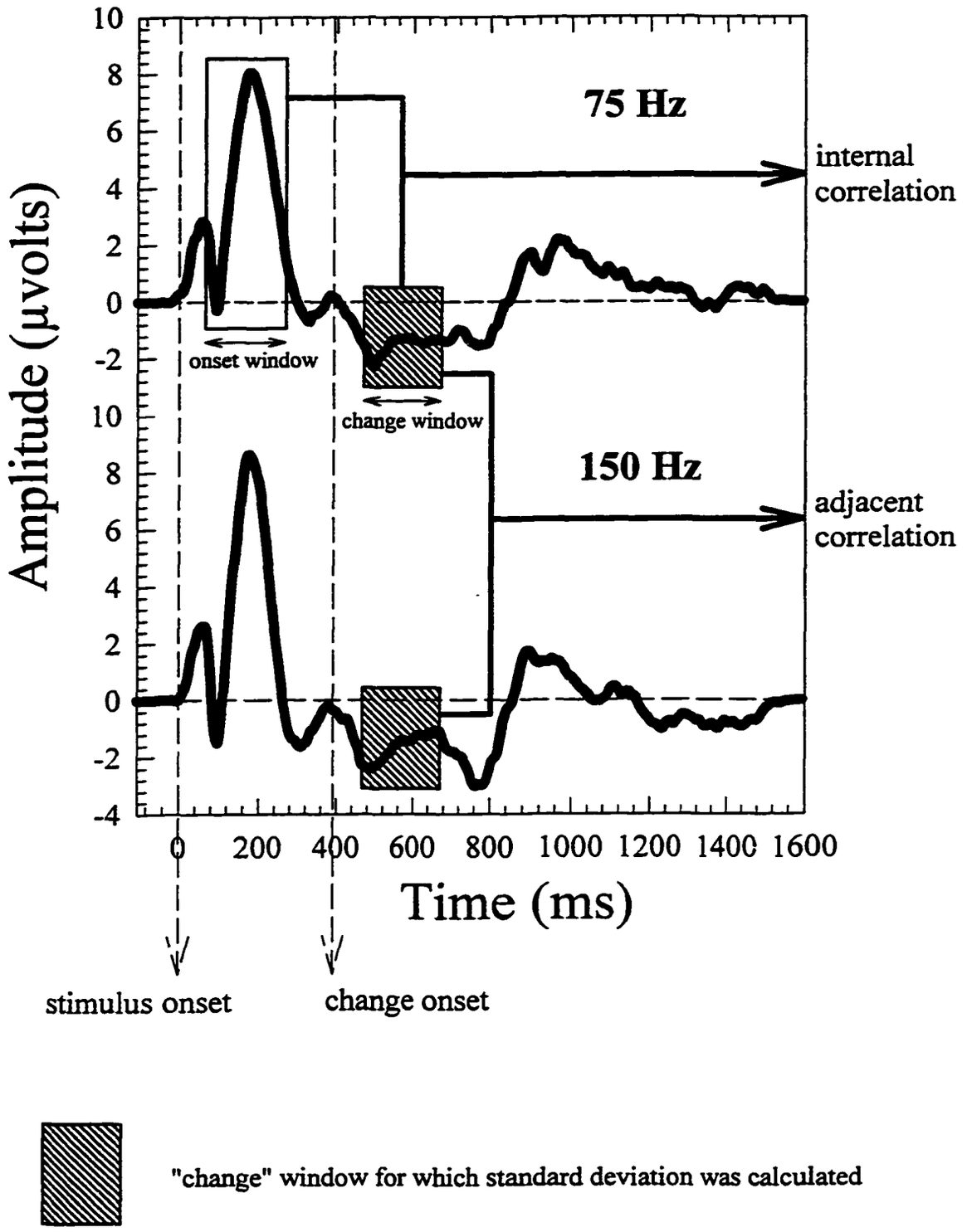


Figure 4.5. Procedure for response identification. Upper waveform shows the response to the 75 Hz spectral change condition. Bottom waveform shows the response to the 150 Hz spectral change condition.

Table 4.1. Application of criteria for response identification based on waveform morphology and rms amplitude. Areas shaded in gray are true ACC responses as defined by the criteria. Standard deviation values are in microvolts.

		Spectral Change (Hz)							
Subject	Criteria	0	9	19	38	75	150	300	600
S1	standard deviation	0.27	0.27	0.39	0.50	0.51	0.56	1.21	1.70
	internal correlation	-0.61	-0.91	-0.64	-0.94	-0.59	0.06	0.78	0.93
	adjacent correlation	0.54	0.80	0.71	0.63	0.87	0.61	0.96	
S2	standard deviation	0.59	0.13	0.32	0.22	0.71	1.10	2.25	2.43
	internal correlation	0.87	0.22	0.94	0.61	0.76	0.91	0.53	0.81
	adjacent correlation	-0.31	0.21	0.56	0.91	0.98	0.96	0.99	
S3	standard deviation	0.16	0.22	0.32	0.39	0.51	0.53	0.82	1.40
	internal correlation	0.00	-0.36	0.04	0.25	0.52	0.80	0.54	0.83
	adjacent correlation	0.28	0.55	0.90	0.85	0.81	0.81	0.84	
S4	standard deviation	0.19	0.13	0.16	0.29	0.33	0.43	0.83	1.25
	internal correlation	-0.26	0.17	0.04	-0.79	0.84	0.85	0.91	0.90
	adjacent correlation	0.35	-0.31	0.07	-0.60	0.82	0.96	0.98	
S5	standard deviation	0.38	0.32	0.21	0.38	0.24	0.53	0.79	0.87
	internal correlation	0.93	0.81	0.86	0.93	0.77	0.89	0.95	0.93
	adjacent correlation	0.85	0.72	0.89	0.43	0.74	0.98	0.97	
S6	standard deviation	0.25	0.24	0.42	0.43	0.30	0.44	0.78	1.14
	internal correlation	-0.03	0.24	-0.20	0.40	0.46	0.67	0.60	0.67
	adjacent correlation	0.80	0.78	0.81	0.88	0.83	0.95	0.99	
S7	standard deviation	0.32	0.26	0.17	0.36	0.44	0.67	0.81	0.92
	internal correlation	0.46	-0.40	0.76	-0.56	0.67	0.57	0.90	0.87
	adjacent correlation	0.02	-0.11	-0.19	0.14	0.97	0.88	0.97	
Group	standard deviation	0.14	0.09	0.14	0.21	0.34	0.54	1.00	1.31
	internal correlation	0.74	-0.16	0.29	-0.12	0.57	0.80	0.83	0.88
	adjacent correlation	0.44	0.71	0.84	0.63	0.96	0.99	0.99	

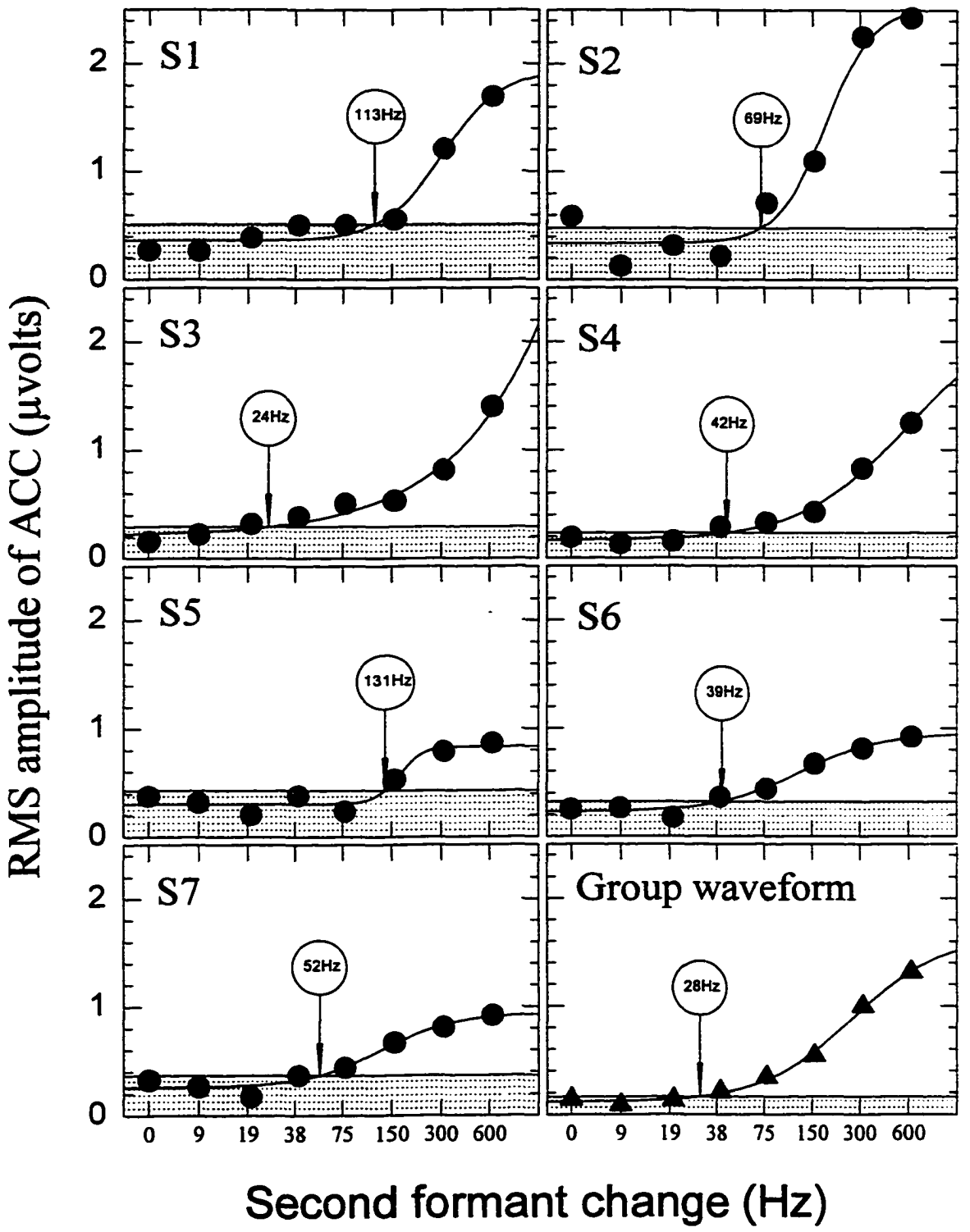


Figure 4.6. ACC thresholds for detection of F2 frequency change as a function of magnitude of change. Data are shown for 7 subjects and the group waveform. Shaded areas show the rms noise level. Lines show least squares fit to a sigmoid function.

Table 4.2. Threshold values in Hz for detection of the ACC based on waveform morphology and rms amplitude and on a curve fit and interpolation technique.

	Response Criterion	
Subject	waveform morphology & amplitude	curve fit and interpolation
S1	300	113
S2	75	69
S3	75	24
S4	75	42
S5	150	131
S6	150	39
S7	75	52
Avg. of all subjects	129	67
Threshold based on group waveform	75	28

III. Discussion

One goal of the present study was to measure the Acoustic Change Complex elicited by changes in the frequency of the second formant within a synthetic vowel stimulus as a function of the magnitude of the F2 shift. The ACC was reliably elicited in seven of the eight subjects tested in this experiment.

A second goal of the present study was to determine the smallest change of second formant frequency that could elicit an observable ACC. It is clear from the findings, however, that threshold estimation depends, in part, on the criteria used to define threshold. For the moment, we can conclude from the group data that the ACC is certainly present for a second formant change of 75 Hz and may be present for a change as small as 28 Hz.

The difficulty in threshold estimation lies not in the actual magnitude of the ACC or the fact that the ACC may be difficult to elicit. The limitation is due to the noise inherent in the averaged waveform. The noise is a product of many sources including physiological artifact, equipment/computer-generated noise, and protocol parameters (i.e., total number of accepted trials). The threshold estimate for the ACC would, presumably, be lowered if the noise could be reduced.

Several cautions must be applied when generalizing from the findings of this study. For example, the findings might only apply to the specific stimuli and protocols used in this study. In addition, it will be recalled that the waveforms from an 8th subject failed to meet the response criterion and were excluded from further analysis. The issue of individual differences will be dealt with later. The procedural issues are addressed in the next section.

Procedural Issues

Stimuli

Each of the stimuli used in this study has a duration of 786 ms with the spectral change beginning at 393 ms. It should be noted that the changes that occur in ongoing speech are generally of shorter duration. However, the change at 393 ms in the present study was selected in an attempt to keep the N1-P2 onset response separate from the ACC. In fact, there is some concern as to whether this delay between the onset of the stimulus and the onset of the change is sufficient for the ACC to remain clear of the transition from onset response to sustained negativity. The sustained negativity is a stimulus-locked DC shift that is elicited if an auditory stimulus continues for more than 150-200 ms. This DC shift lasts for the duration of the stimulus (Eulitz et al., 1995). There is some indication from the data obtained in this experiment that a sustained negativity is present in some of the waveforms. The sustained negativity can be seen clearly in the response waveform from the group mean presented in Figure 4.1. The transition to the sustained negativity is especially important when basing decisions about response presence/absence on rms amplitude alone. Any baseline slope within the presumed response window will increase rms amplitude and lead to an overestimate of the noise floor (i.e., the rms amplitude within the response window in the absence of an acoustic change). The sustained negativity could be especially relevant to the small spectral change conditions where the magnitude of the elicited ACC was small. Note that the decision to use a 393 ms delay in the present study was based on results of earlier studies with shorter offsets in which the ACC became confused with the onset response. The sustained negativity present in some of the response waveforms from this study does

not appear to compromise the ability to elicit the ACC because in the group waveform, at least, the negativity appears to have stabilized by 300 ms.

The stimuli used in the present study contain an F2 change created by concatenating two stimuli, each having a different F2 center frequency. The F2 change is therefore a step function with an overlap of two cycles to avoid an abrupt transition from the first portion of the stimulus to the second. Every attempt was made to create stimuli as close to natural speech as possible while still maintaining control over all stimulus parameters. Nonetheless, natural speech contains formant changes that are more gradual than the ones contained in the stimuli used in this study. Additionally, the present stimuli are composed of a three-formant synthetic vowel with fixed formant amplitudes. In natural speech, the frequency and amplitude of formants are neither fixed nor independent. There is no reason to conclude that the peripheral discrimination capacity revealed by the ACC in the present study is irrelevant to the perception of natural speech. It would, however, be helpful to know whether these findings can be replicated with stimuli that are closer to natural speech.

Individual Differences

Somewhat like fingerprints, each subject seems to provide his or her own distinct set of electrophysiologic response waveforms. These differences can make response identification difficult. For example, the waveforms from Subject 1, seen in Appendix B.1, display a kind of "ringing" characterized by a negative dip following the onset P2 and before a return to baseline. This peculiar response pattern typical of this one subject, made response identification difficult with the response identification procedures used in

this study. This is evident in the threshold estimates obtained from this subject, which are on the order of 100 Hz higher than those of the other six subjects tested. There is also a large discrepancy between the threshold provided by the waveform morphology and amplitude criteria and that provided by the curve-fit procedure. The duration of the stimulus could have contributed to this response pattern. Had the duration of the portion of the stimulus prior to the spectral change been longer, this rather severe negative dip following the onset response might have been completed before the onset of the ACC. Considering that this subject has the largest onset responses of any other subject (on the order of 15 μvolts peak to peak), it is possible that this "ringing" pattern is reducing the amplitude of the ACC in this subject. The "ringing" probably also contributes to the lower (and even negative) internal correlation coefficients, which limit the threshold estimate based on waveform morphology to 300 Hz.

There is also a rather large discrepancy (on the order of 110 Hz) between the ACC threshold provided by the waveform morphology and amplitude criteria and that provided by the curve-fitting procedure in the data from Subject 6. Upon visual inspection of the waveforms, displayed in Appendix B.6, one could argue that the ACC appears to be present down to spectral changes as small as 19 Hz because of the presence of an N1 component at about 550 ms in this condition. However, the ACC amplitude is weakened in the 75 Hz spectral change condition causing the rms amplitude value to fall below that of the rms amplitude for the 38 Hz spectral change condition. This violates one of the criteria specified above for the acceptance of an ACC which requires that the standard deviation for a given response waveform be higher than the average standard deviation for all lower amounts of spectral change. Therefore the threshold of the ACC in this

subject using the rms amplitude criteria is limited by the decrease in amplitude of what appears to be an ACC in the 75 Hz spectral change condition.

Response waveforms from Subject 5 displayed in Appendix B.5 display a peculiar morphology when compared with response waveforms from the other six subjects. The slope of the waveform immediately following the P2 component of the onset response is rather shallow, indicating a slow attenuation of the onset response. Despite this unique response pattern typical of this one subject, the ACC was elicited, although both techniques for identifying the ACC provided rather high thresholds (150 Hz by waveform morphology and amplitude, 131 Hz by curve-fit). It is possible that the slow attenuation of the onset response diminished the strength of the ACC, especially for small spectral change conditions where the ACC would be small to begin with, thereby increasing the ACC threshold.

The ACC could not be reliably identified in the response waveforms from Subject 8 (Appendix B.8). The response waveforms elicited by the two largest spectral change conditions do contain a negativity (at around 420 ms). However, this negativity is probably too early to be the N1 of an ACC – it occurs only 27 ms after the onset of the change (at 393 ms). Moreover, this negativity is also present at around 420 ms in the response waveform elicited by the control condition that does not contain a spectral change. Many of the responses from this subject were contaminated by physiological noise (mostly muscle artifact). Of 800 total sweeps per condition, an average of 659 were accepted for this subject compared with an average of 765 for all other subjects. The noise is unlikely to have been equipment/environment related, because the recordings were done on two separate days. The noise however, cannot entirely account for the

response pattern displayed in the waveforms from this subject. It is possible that the negativity present around 420 ms is an artifact of the onset response peculiar to this subject, which eliminates the ACC or confounds its presence. Therefore, uncertainties about the presence of an ACC even to the 600 Hz spectral change required the exclusion of the responses from this subject from further analysis.

Responses from Subject 2, displayed in Appendix B.2 are by far the best behaved of the group. The ACC is easily detected upon visual inspection down to spectral changes of 75 Hz. The amplitude of the ACC is also large, especially when elicited by the two largest spectral change conditions (approximately 6 μvolts peak to peak). Of particular interest is the fact that the two response identification procedures provide very similar ACC threshold estimates. The ACC threshold estimated provided by the waveform morphology and amplitude criteria is 75 Hz and it is 69 Hz as estimated by the curve-fit procedure.

It should be noted that although every attempt was made to devise threshold estimation procedures which were as objective as possible, there was still an element of subjectivity involved which could account for some of the findings. For example, upon thorough investigation of individual and group mean responses, it was determined that the most logical temporal window for response identification based on waveform morphology and amplitude was from 463 to 626 post-stimulus onset. This window was selected based on the latency of the ACC in the group waveform elicited by the largest spectral change condition. However, given that individual differences have just been pointed out, it is likely that this temporal window does not "fit" the response waveforms elicited by every subject. The occasional discrepancy between ACC thresholds expected

based on visual inspection and those provided by the threshold estimation procedures might be accounted for by the selection of response window. For example, had the response window been different, it is possible that all or a part of the "ringing" demonstrated in the responses from Subject 1 would not have been included, thereby changing such things as the internal correlation coefficients.

It is possible to make the argument that a customized response window should have been selected for each subject (perhaps based on the response waveforms elicited by the 600 Hz spectral change condition) in order to account for individual waveform idiosyncrasies. Although this method was considered, it was rejected because it would have increased the number of decisions to be made by the experimenter.

ACC Identification and Threshold Estimation

The choice of response identification and threshold estimation procedures can also be discussed in terms of other important factors. The two approaches selected to identify the ACC and to estimate threshold from the set of data presented in this study involved a combination of criteria based on correlation measures and a curve-fitting procedure using a sigmoid transition function. Recall that for a waveform at a particular level of spectral change to be accepted as an ACC, the data had to meet the following criteria:

(a) The (internal) correlation with the N1-P2 onset response had to be above 0.5.

(b) The (adjacent) correlation with the waveform for the next biggest spectral change had to be above 0.5.

(c) The standard deviation had to be higher than the average standard deviation for all lower amounts of spectral change.

The first criterion relies on the assumption that the onset response and the ACC are similar – i.e., that they are essentially produced by the same generators and therefore display the same morphological, amplitude, and temporal characteristics. Although the ACC appears to be in fact a simple onset response (to change) based on the waveforms from most subjects; there are latency and morphology differences between the two responses. An example of one difference can be seen in the responses from Subject 1, which contain a peculiar extra component following the onset response. The generator sites of the ACC and N1-P2 onset response can only be confirmed using source analysis and dipole mapping techniques.

The second criterion relies on the assumption that there is morphological similarity between responses to different amounts of change. Unfortunately, the shape of a component can change or become gradually degraded as the signal-to-noise ratio gets smaller. That is the reason the correlation is made with the next largest spectral change, which should theoretically have a greater signal-to-noise ratio. There is however, no guarantee of morphological similarity between responses, and a resulting correlation coefficient of less than 0.5 would not necessarily indicate that the ACC is not present. For example, this possibility appears most likely to be the case in the response waveforms from subject 6 where there is a reduction in the amplitude of what appears to be the ACC at the 75 Hz spectral change condition. The use of topographical analysis could demonstrate that two waveforms had similar topography (albeit somewhat dissimilar morphology) thereby providing additional evidence that the ACC was present. However, because of equipment limitations (limiting the number of available recording channels) and the lack of the appropriate software, topographical analysis was not possible when

the data from the present experiment were analyzed.

Note that although the assumption of morphological similarity between responses may not be absolute, the required correlation of 0.5 for the first two criteria is not very demanding. Therefore it appears to have been a reasonable coefficient to select.

The third criterion, which takes into account standard deviation over a specified response window, has its advantages over the ordinary base-to-peak or peak-to-peak measures. The problem in using peak measures is that 1) only two samples are being used thereby wasting data, 2) the samples are chosen on the basis of their values (in μvolts) – similar to choosing extreme members of a sample and claiming they are significantly different from the mean, and 3) the measured amplitude is a combination of signal and noise although the noise has been measured as the difference between two points that have not been randomly selected.

The relative advantages and disadvantages of the procedures used in this study could provide endless debate. Their primary justification is that they are based on reasonable theoretical constructs and they gave plausible results for the waveforms generated in this study.

Conclusions of the Electrophysiologic Study:

The threshold for detection of an observable ACC to F2 changes in normally hearing adults is dependent upon the criteria used to define threshold. However, using the materials and protocols of the present study, individual thresholds lie between:

1) 75 and 300 Hz when using the waveform morphology and rms amplitude criteria.
2) 24 and 131 Hz when using the curve-fitting procedure based on rms amplitude.

When using the group waveforms, thresholds are in the region of:

1) 75 Hz when using the waveform morphology and rms amplitude criteria.

2) 28 Hz when using the curve-fitting procedure based on rms amplitude.

These are upper estimates because the sensitivity of response detection is limited by residual noise in the averaged waveforms.

CHAPTER 5

BEHAVIORAL STUDY

I. Methods

Subjects

Behavioral responses were measured from seven adult volunteers who demonstrated an Acoustic Change Complex in the electrophysiologic study. Subjects underwent behavioral testing upon completion of the electrophysiologic study, in a single session on a separate day.

Behavioral Measures

Behavioral testing was accomplished in a psychoacoustic paradigm using a customized computer program written in Multimedia ToolBook by Asymetrix Corporation (Appendix A). On each trial, a pair of stimuli was presented in which one member of the stimulus pair contained a spectral change and the other was the control stimulus. The interval between members of the stimulus pair was 1 second. There was a pause of one second after the subject responded and before the next stimulus pair was presented. Subject's had an unlimited amount of time to respond. The subject's task was to determine whether the spectral change was present in the first stimulus or the second stimulus of the pair ("binary choice" paradigm). Subjects received practice trials (approximately 3) using the 75 Hz spectral change condition to ensure that they understood the task.

Spectral change stimuli of 9, 19, 38, and 75 Hz were presented to subjects in the behavioral test.[2] The stimuli were presented through TDH 50P earphones at 80 dB SPL. The same degree of spectral change was presented within a trial block. A trial block consisted of 20 pairs of stimuli. A total of 5 trial blocks was presented for each stimulus condition yielding a total of 100 stimulus presentations for each degree of spectral change. Presentation order of the trial blocks began from the largest spectral change condition (and therefore easiest to perceive) and proceeded to the smallest (and most difficult to perceive). Behavioral testing was accomplished in approximately one hour.

Calibration

The speech sounds for the behavioral study were calibrated before each test. The speech sounds were calibrated in dB peak rms SPL using a Larson Davis sound level meter with a 1" microphone (model 2575). The calibration signal was the control stimulus.

Procedure

Subjects were tested in an acoustically treated and electrically shielded room. Subjects were seated in front of a computer monitor, keyboard and mouse, and were instructed to select the stimulus (i.e., "first" or "second" of the pair) that contained the spectral change. Breaks between trial blocks were provided to subjects upon request.

---

[2] Presenting spectral changes larger than 75 Hz proved to be superfluous in a pilot study because subjects can perceive this degree of spectral change 100% of the time.

Data Analysis

Percent correct responses were calculated for each stimulus condition and each subject. A mean score was calculated across subjects and stimulus conditions. Individual and group values of behavioral threshold were calculated (described below) and compared with the thresholds derived from the electrophysiologic component of the study.

II. Results

Percent correct scores, on the binary choice - "change detection" task, are presented in Table 5.1 and displayed graphically in Figure 5.1 for the seven subjects giving observable ACCs. Also shown are the group means for the four spectral change conditions.

The group mean score for the largest (75 Hz) spectral change condition tested was close to 100%. As a group, subjects performed very well (93%) discriminating the 38 Hz spectral change condition, but scores fell to 74% and 62% respectively for the 19 and 9 Hz spectral change conditions.

Thresholds for detection of the F2 spectral change, determined for each subject and for the group mean, were obtained from three sets of estimates: the just noticeable difference (JND), percent correct scores from the 75% performance level, and percent correct scores from the 95% performance level. These estimates were derived using the same curve-fit and interpolation technique that was used to estimate ACC thresholds from the electrophysiologic data. The percent correct scores from all subjects and the group mean for each spectral change condition were fit to a sigmoid transition function. The function is defined by the same equation given in Chapter 4:

$$y = a + (b-a)/(1+e^{-(x-c)/d}) \quad \text{.....(1)}$$

where:

y = recognition probability in percent
a = 0
b = 100
e = base of natural logarithms
x = log(F2 change)
c = x value for midpoint of transition
d = an inverse measure of slope

The resulting curves are displayed in Figure 5.2. This function assumes 50 % and 100 % asymptotes. The noise floor was estimated from the 95% confidence limits based on application of the binomial theory to chance performance. The first estimates of threshold are taken from the point at which the curve-fit to the data points intersected with the upper confidence limit of the noise floor. These JND threshold values based on the first estimate for all subjects and the group mean are shown in the first column of Table 5.2.

Using the first estimates, averaging JND threshold values from all seven subjects, the threshold is 15.3 Hz with a range of 4.5 Hz to 24.5 Hz. The threshold based on the group mean data (i.e., averaging percent correct scores across all subjects for each condition) is 5.4 Hz.

The second set of threshold values is based on a more traditional psychoacoustic estimate of change detection: Percent correct scores from the 75% performance level are included in the second column of Table 5.2. The average threshold at 75% is 21 Hz with a range of 6.6 Hz to 39.3 Hz. The group mean threshold at 75% is 17.9 Hz.

The third set of threshold values is based on conservative estimates from the 95% correct performance level. This "confidence" threshold is assumed to be the point at which subjects are confident that they have detected the F2 formant change. Thresholds

based on 95% performance are included in the last column of Table 5.2. The average threshold at 95% is 36.3 Hz with a range of 11.9 Hz to 86.7 Hz. The group mean threshold at 95% is 50.9 Hz.

Table 5.1. Percent correct scores for seven subjects and the group mean on four F2 spectral change conditions.

	Subject							
Condition	S1	S2	S3	S4	S5	S6	S7	group mean
9 Hz	88	61	65	55	46	65	54	62
19 Hz	99	78	98	57	70	61	53	74
38 Hz	100	99	99	72	100	100	84	93
75 Hz	100	100	100	95	100	100	95	99

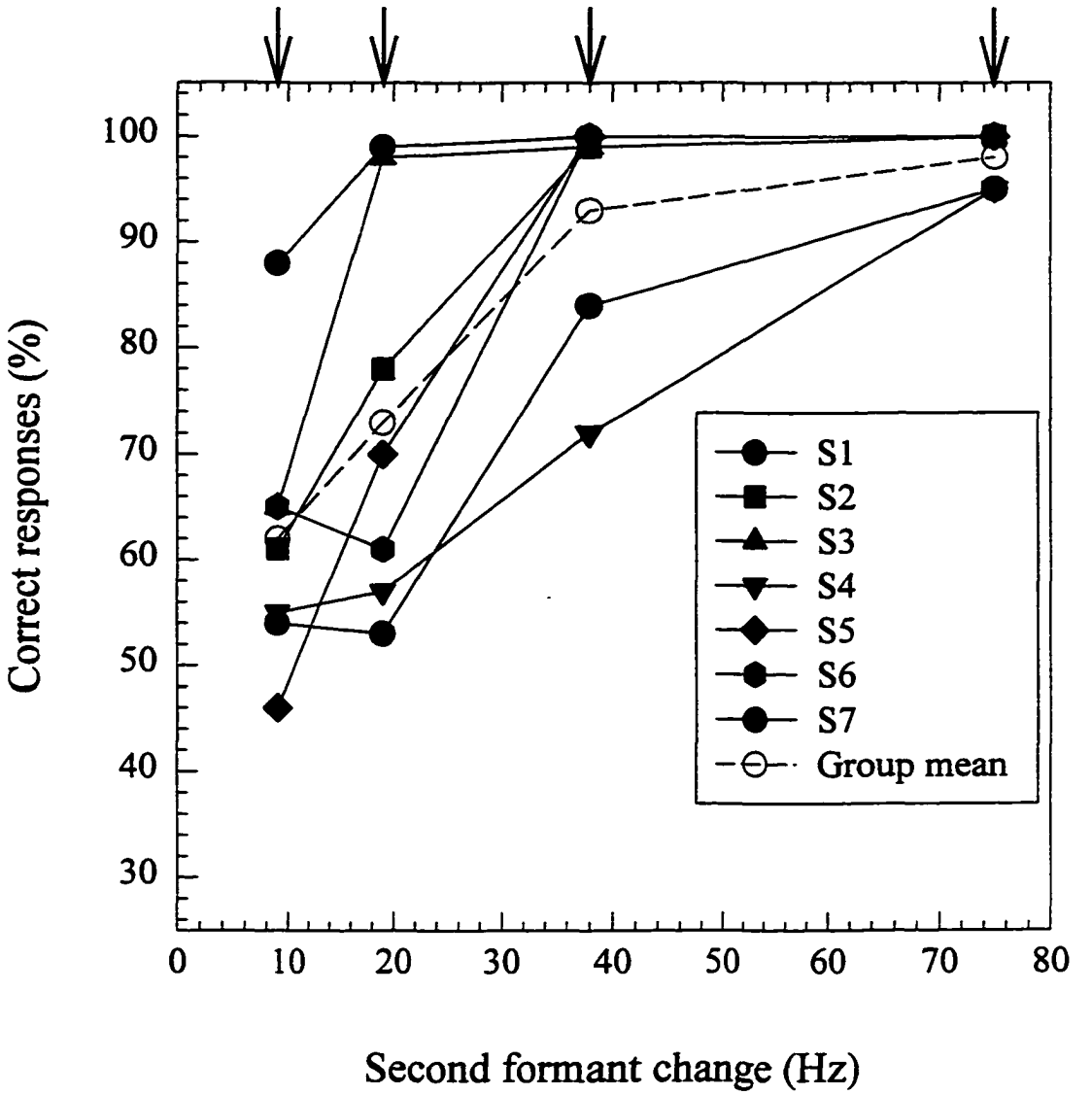


Figure 5.1. Percent correct scores for seven subjects and the group mean on four F2 spectral change conditions. Arrows indicate change conditions.

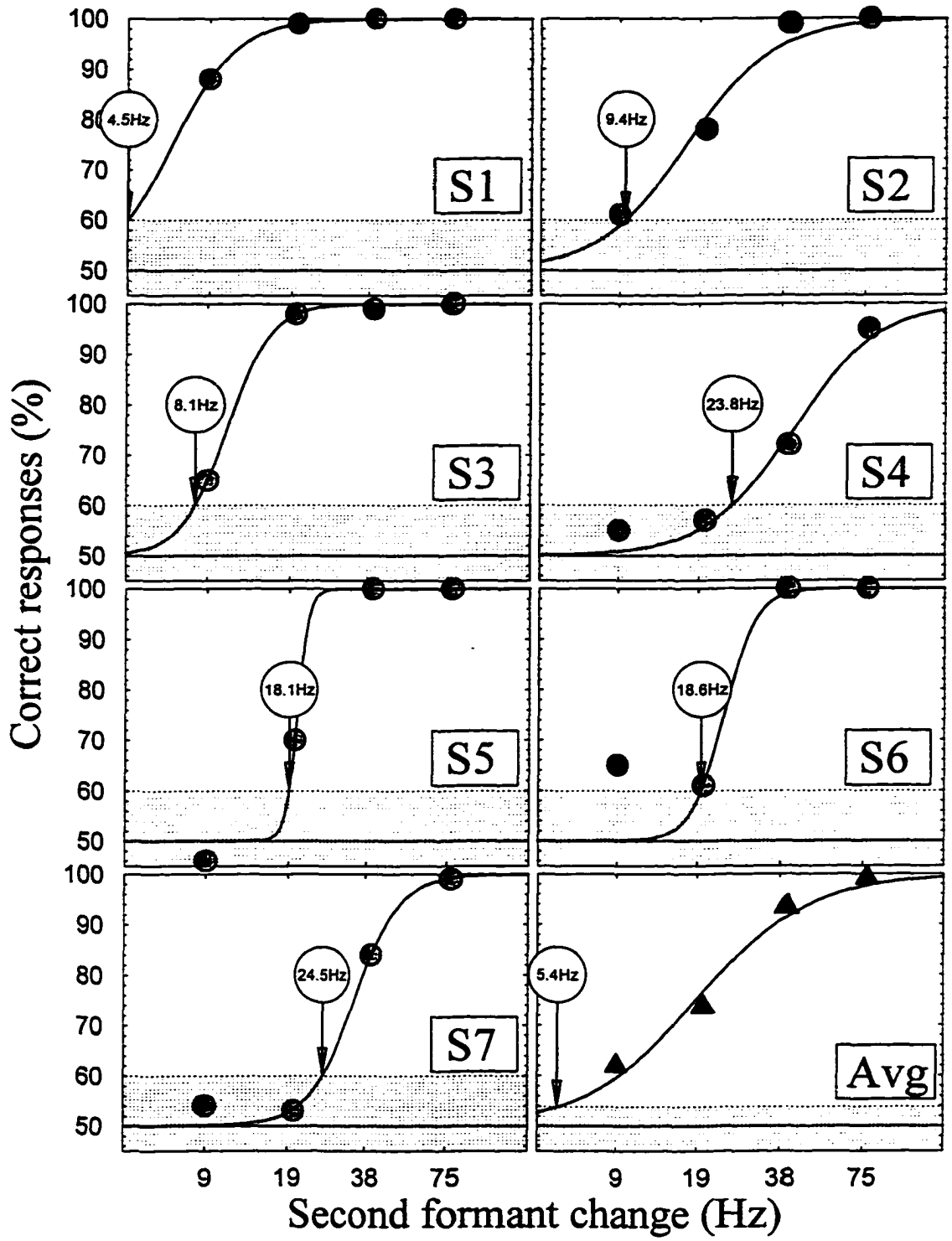


Figure 5.2. Probability of detection of change in second formant frequency as a function of magnitude of the change. Data are shown for 7 subjects and for the group mean. Chance performance is 50%. The shaded areas show the 95% confidence limits for performance based on chance alone. Lines show least squares fit to sigmoid transition functions. Intersection of fitting curve with upper confidence limit defines threshold for just noticeable difference.

Table 5.2. Threshold values in Hz based on the curve fit and interpolation technique for just noticeable differences (JND), 75% and 95% performance on the binary choice - F2 change detection task.

	Threshold measure		
Subject	JND	75%	95%
S1	4.5	6.6	11.9
S2	9.4	16	36.6
S3	8.1	10.5	15.9
S4	23.8	39.3	86.7
S5	18.1	19.4	21.6
S6	18.6	22.4	30.1
S7	24.5	32.6	51
Avg. of all subject thresholds	15.3	21.0	36.3
Group mean	5.4	17.9	50.9

III. Discussion

The first goal of the behavioral study was to measure discrimination of changes in the frequency of the second formant of a synthetic vowel stimulus as a function of the magnitude of the F2 shift. All of the seven subjects tested could complete the binary-choice task. The issue of individual variability will be discussed in greater detail later.

The second goal of the behavioral study was to determine the least amount of spectral change within a vowel needed to elicit a reliable behavioral response. However, this behavioral threshold must be defined by some logical principle much the same way that the ACC threshold is dependent upon its definition. Assuming that the behavioral threshold is defined as the point at which the F2 change becomes *just* detectable, the point on the curve (fit to a sigmoid function) at which percent correct scores are significantly greater than chance performance would represent a logical value of behavioral - just noticeable difference - threshold. In the case of the data from the present experiment, the behavioral threshold for F2 change detection from the group mean data would be 5.4 Hz.

The behavioral threshold could just as logically be defined as the point on the curve at which performance on the F2 change detection task has reached 75%. This second, more stringent definition of F2 change detection threshold places the threshold value for the group mean data at 17.9 Hz. It is of interest that this threshold value of F2 frequency discrimination corresponds closely with that found by Kewly-Port & Watson (1994). The results of their research demonstrate that in an adaptive two-alternative forced choice task with threshold determined at 71% correct performance, subjects were able to discriminate F2 changes of 18.4 Hz in the region of 1175 Hz.

Individual Variability

The differences between subjects can be explained by two sources of variability that contribute to behavioral testing of human subjects. The first source of variability results from true differences in detection ability between subjects. Some subjects may be very good at perceiving small frequency changes and some subjects do not have the "talent" for it. The second source of variability results from "noise" within individual subjects that takes place from moment to moment during the detection task. For example, subjects' responses can be influenced by their state at the time of testing. Subject-state encompasses such factors as attention and motivation to the task, and alertness at the time of testing. The levels of physiological noise inherent to the subject also influence the results of experiments involving audition, such as the present one. Physiological noise includes such things as the sound of blood flow and heartbeat, which can be heard in a sound-treated environment. These noises can influence a subject's ability to detect sounds or to discriminate small differences in sounds, especially if the noise coincides with the presentation of a stimulus (or stimuli).

A subject's comprehension of the task, or perception of what is expected on the task, can also influence the results of the experiment. The task was explained to the subjects in the present study and they received several practice trials before the actual experiment began. From all indications the subjects understood the task. It is possible, however, that the perception or interpretation of the speech stimuli differed among subjects, some listening for acoustic differences and others listening for phonetic differences. Subjects were not asked to identify the stimuli and/or to categorize them phonetically. Moreover, the measured JND's for F2 change were miniscule compared with the kinds of change

involved in phonemic distinctions. Nonetheless, there is no way to determine how subjects perceived the stimuli. The variability in behavioral thresholds among subjects might be accounted for by differences in the way they approached the perceptual task.

The behavioral threshold for Subject 1 is also of interest because it is by far the lowest of the group at 4.5 Hz. This threshold value is actually suspect because of the amount of extrapolation involved, and a more accurate threshold would require data points at spectral changes between 9 and 19 Hz. However, the threshold for Subject 1 is not far from the thresholds of Subjects 2 and 3 at 9.4 and 8.1 Hz respectively.

The discrimination thresholds from this group of subjects, despite some variability, are actually quite interesting because they reflect a tremendous capability of the auditory system to detect very small changes in formant frequencies on the order of 5 – 100 Hz, depending on the criterion of F2 change detection threshold that is used.

Conclusions:

Thresholds for behavioral detection of F2 changes in normally hearing adults are dependent upon the estimate used to define threshold. However, using these materials and protocols, individual thresholds lie between:

1) 4.5 and 24.5 Hz when thresholds are estimated from percent correct scores that are just significantly greater than chance performance.
2) 6.6 and 39.3 Hz when thresholds are estimated based on 75% correct performance.
3) 11.9 and 86.7 Hz when "confidence" thresholds are estimated based on 95% correct performance.

When using the group mean data, thresholds are in the region of:

1) 5.4 Hz when thresholds are estimated from percent correct scores that are just greater than chance performance.
2) 17.9 Hz when thresholds are estimated based on 75% correct performance.
3) 50.9 Hz when "confidence" thresholds are estimated based on 95% correct performance.

CHAPTER 6

COMPARISON OF ELECTROPHYSIOLOGIC WITH BEHAVIORAL RESULTS

The last goal of the present study was to determine the relationship between the thresholds of spectral change detection as measured electrophysiologically and the thresholds as measured behaviorally. A comparison will be made between group mean data presented in Chapters 4 and 5.

I. Results

For the sake of clarity, the group mean responses, as a function of F2 change, obtained in the two studies are displayed in Figure 6.1. This Figure consists of the panels at bottom right of Figures 4.5 and 5.2.

II. Discussion

It will be seen that the F2 change at which the group mean emerges from the noise floor was 28 Hz for the electrophysiologic study and 5.4 Hz for the behavioral study. At first sight, these findings suggest that the ACC is a poor index of discriminability. To compare these two data points is, however, of little value. Both threshold estimates are partly dependent on the noise floor, and the nature of this noise floor is different in the two studies.

In the behavioral study the estimate of noise floor was the standard error of group mean scores based on random guessing. Two factors determine the value of the standard error: the number of observations contributing to an individual score and the number of subjects. Lowering either of these values would have elevated the threshold estimate.

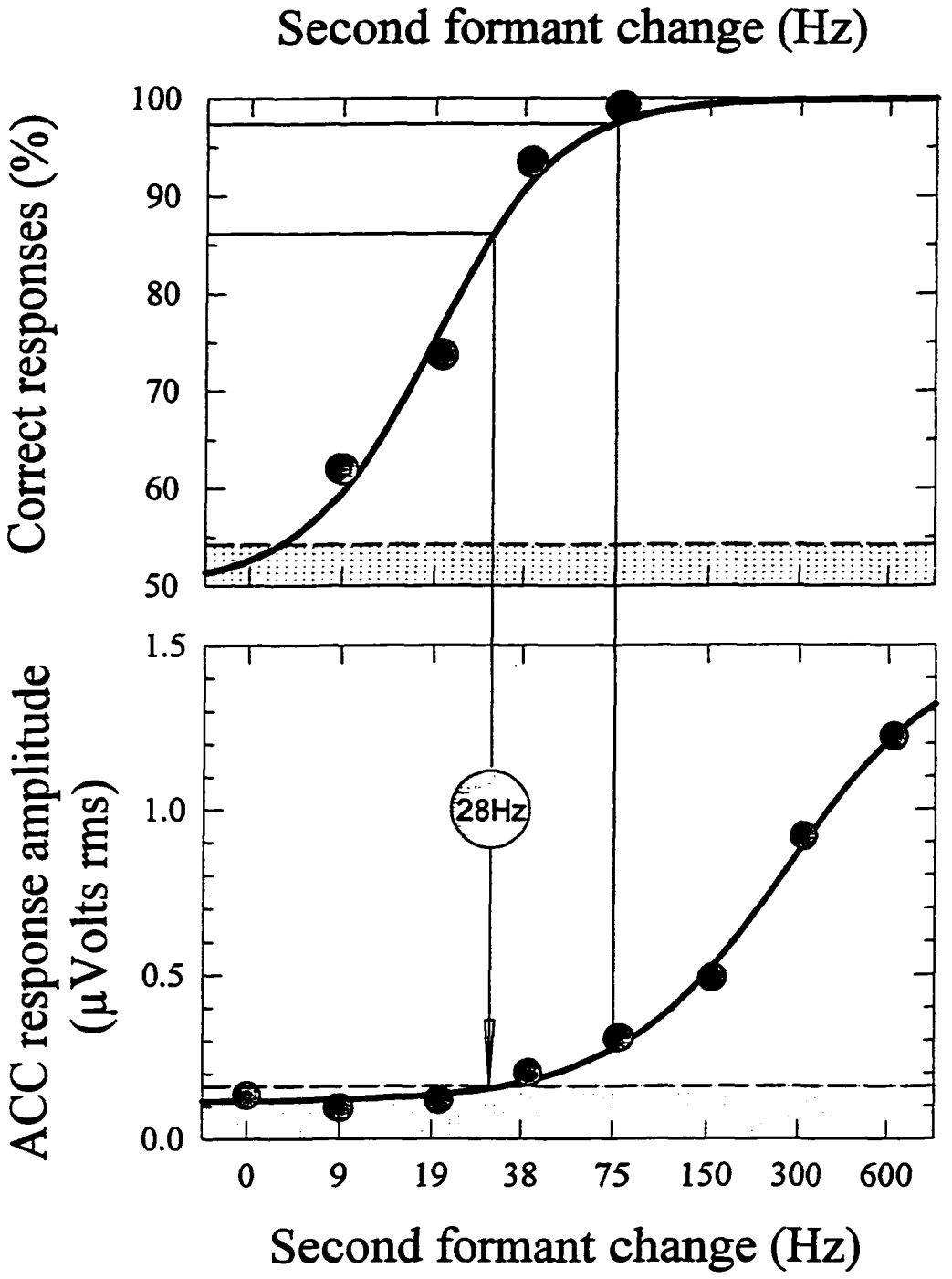


Figure 6.1. Top panel: percent correct scores on the F2 change detection task fit to a sigmoid function. Bottom panel: rms amplitude of ACC fit to a sigmoid function. Data are from the group mean and the group waveform. ACC threshold range between 30 and 75 Hz corresponds with 86 to 97 % correct performance on the behavioral task.

In the electrophysiologic study the noise floor was the rms response amplitude of the group mean EEG waveform in the assumed response window. This quantity is determined by at least six factors: background EEG activity, electromyographic potentials, background electrical fields, baseline drift, the number of sweeps used to compute an individual average, and the number of subjects averaged to produce the group mean. Increases in background noise, increases in baseline drift, and decreases in the numbers of sweeps and subjects would increase the estimate of threshold.

The only factor in common to the two noise floors is the number of subjects. In other respects, the two estimates have nothing in common. Because the noise floor affects the threshold estimate there is no basis for comparison of the two threshold estimates. It cannot be concluded, therefore, that the ACC is inherently incapable of revealing detection of an F2 change with the same sensitivity as is a binary-choice behavioral task – although with the protocols used in these two studies, that was certainly the case.

One could argue that an F2 change that is detectable at slightly better than chance levels in a binary-choice task should not be expected to produce an ACC response. A more reasonable comparison might be the F2 change for which an ACC is reliably generated with an F2 change that listeners can detect with confidence. Depending on the threshold criterion, the group waveform from the electrophysiologic study gave thresholds between 28 and 75 Hz. Examination of the group behavioral data shows that these two values correspond with detection probabilities of around 86% and 97%, respectively (see Figure 6.1). These last values certainly meet the criterion of "detection with confidence". It can be concluded, therefore, that, with the stimuli and protocols used

in these studies, a reliable ACC could be elicited in the group mean for amounts of F2 change that are detected with confidence in a binary-choice change/no-change task.

CHAPTER 7

GENERAL DISCUSSION

I. The Behaviorai Task and The Limits of Psychoacoustics

Every attempt was made in the present study to design a behavioral task that would best ascertain subjects' psychophysical threshold. Presumably, this threshold is limited by internal noise. This assumption is taken from signal detection theory where performance other than perfect is due to a certain level of noise internal to the system that affects the system's decisions about signal presence vs. absence.

The binary choice task used in the present study required subjects to determine which of two stimuli contained an F2 formant change. The task was designed to be as insensitive as possible to such variables as subject attention, memory, categorical perception, and higher order cognitive processing. Although the task was well controlled, the results from individual subjects are not entirely free of the influence of these variables. The threshold value obtained from the group mean data at the traditional 75% performance level is 17.9 Hz. This frequency value translates to 0.027 of an octave, which is 1/3 of a semitone. In other words, the group mean threshold is very small. A frequency change on the order of 17.9 Hz can also be put in perspective by considering that the critical bandwidth of the human ear at 1000 Hz is approximately 100 Hz, which is 0.i 3 of an octave or 1½ semitones. The discrimination ability of subjects on this task is therefore very good. It is reasonable to conclude that the task used here was effective in terms of revealing psychoacoustic capacities without serious contamination by phonetic-level processing.

II. Spectral Change vs. Amplitude Change

The two concatenated portions of the stimuli (producing stimuli containing various degrees of F2 spectral change) were equalized for rms intensity. It is possible, however, that rms intensity is not the critical property that determines the amount of synchronous neural excitation. Other possibilities might include instantaneous peak level. Martin & Boothroyd (in preparation) have demonstrated, however, that with a protocol similar to that used in the present study, the ACC is not confounded by changes of rms amplitude until the rms change reaches at least +/- 2 dB. The two portions of the stimuli used in the present study contain a difference in peak amplitude of only 0.79 of a dB for the 300 Hz spectral change and 0.1 of a dB for the 75 Hz condition. Therefore, it is unlikely, for the spectral change conditions around threshold values reported here, that peak amplitude change contributed to the ACC. The ACC responses elicited in the present study can reasonably be attributed to a change of spectrum alone.

III. Theoretical Implications

The ACC hypothetically reflects an onset response from cortical regions not previously demonstrating synchronous neural firing, possibly supplemented by an offset response from regions no longer excited. With small changes in the frequency of F2 considerable overlap between cortical regions of excitation for the two values of F2 is expected. Hence there will be small cortical regions producing onset and offset responses. As the change in F2 frequency increases, the overlap should diminish and the regions providing onset and offset responses should increase. If the excited regions do not overlap at all, a complete offset response will be elicited by the first portion of the

stimulus, and a complete onset response by the second portion. This simple model would predict an increase of ACC amplitude with increasing F2 change, but no further increase in the magnitude (in μvolts) of the ACC once the change in F2 frequency becomes sufficiently large for a separation of excited cortical regions from the two values of F2. From an acoustic point of view, this condition is not satisfied until an F2 shift of 1200 Hz is reached. That is, the two formant regions do not overlap until the separation of center frequencies reaches 1200 Hz (see Figure 3.3). Unfortunately, this condition was not included in the present study so it is not possible to test this prediction. This model does explain, however, the absence of evidence that ACC amplitude reaches an asymptote for an F2 change of 600 Hz. In a pilot study, a wider range of F2 change was used and there was no evidence that the ACC amplitude increased for F2 changes beyond 1200 Hz. The data were too few, however, to conclude that an asymptote had been reached. This issue provides a topic for further research.

IV. Attentional State of Subjects

There is evidence in the literature of an enhancement of the N1 component of the onset response when subjects attend to the stimulus presentations (Näätänen & Picton, 1987). The influence of attention on the ACC is not yet known. Although subjects in the present electrophysiologic experiment were instructed to ignore the stimuli there is no way to determine the attentional state of subjects during testing. If attention does influence the magnitude of the ACC, it is possible that the differences in threshold values for the ACC between subjects represent differences in their attentional states. In other words, a subject providing a very low value of threshold for the ACC may have been

attending to some or all of the stimuli, thereby increasing the amplitude of the ACC.

These considerations raise the possibility that the attentional state of subjects could be influenced by the phonemic quality of the stimuli. For instance, the ACC may have emerged in the response waveforms once the second formant change was enough such that the two halves of the stimulus were perceived as two different phonemes and thus captured the subjects' attention. This increased attention to phonemically salient stimuli, if it exists, would essentially determine the measured threshold. To the extent that the primary goal of the present research is to develop electrophysiological alternatives to behavioral psychoacoustics, such an interpretation is discouraging. Clearly this is a topic for further research.

V. Vowel Distinctions

Both behavioral and electrophysiologic data suggest that the seven subjects in this experiment can discriminate, with confidence, frequency changes on the order of about 30 to 75 Hz depending on the threshold criterion that is used. Assuming that one can demonstrate sensitivity to F2 changes in the region of 30 to 75 Hz, would that degree of discriminability provide for the ability to accurately perceive phonemic distinctions among vowels? The data from Peterson and Barney (1952) suggest that second formant changes between adjacent vowels are on the order of 100 to 150 Hz. The smallest second formant change of 30 Hz, is between the vowels /u/ and /ɔ/ but this is accompanied by a large F1 change on the order of about 120 Hz. Therefore the sensitivity of the ACC is such that it could reveal the potential for the ability to discriminate vowel changes that are important for speech perception.

VI. Future Research

i. Identify the source of the ACC:

Future research should be aimed at determining the location of the generators of the ACC within the cortex. The similarity of the ACC waveform to the N1-P2 onset response suggests that they share the same cortical generators. Scalp topographies from the onset and change responses to both pure tones and tonal complexes could be compared at various scalp locations. The results could be used to derive the intracerebral sources of the two responses. Additional studies could determine whether the ACC elicited by changes of spectrum, amplitude, and periodicity have different sources - implying that different cortical regions are involved in the processing of different acoustic features.

ii. Test the asymptote hypothesis:

Based on a simple tonotopic theory of cortical excitation, it is hypothesized that the ACC would demonstrate no further increases in response amplitude once the change in F2 frequency is large enough that separate cortical regions are excited by each portion of the stimulus containing the change. This plateau occurs because no additional neural units are being turned on or off. Conversely, it is expected that as the magnitude (in decibels) of the acoustic change increases, the total new volume of cortical activation increases monotonically. The amplitude of the ACC might not, therefore, display a plateau to changes in stimulus amplitude, at least within the limits of available stimulable cortex. Note that previous research has shown that some subjects demonstrate a plateau and some do not in the amplitude of the N1-P2 onset response with increases in stimulus intensity (Näätänen & Picton, 1987). These predictions could be tested by increasing the

amount of spectral change beyond that used in the present study (600 Hz) and by examining the ACC to increasing changes of amplitude.

iii. Separate onset from offset response:

The relative contributions of offset and onset effects to the ACC elicited by spectral change should be determined. Based on the known tonotopic organization of the auditory cortex, it is hypothesized that a shift in center frequency of a tonal complex of fixed bandwidth involves both the excitation of new neural regions, and the inhibition of other areas already active. The observed ACC must, therefore, be a combination of onset and offset responses. To determine the relative contributions of these components, the ACC could be measured to changes involving addition and removal of single tones at the edges of a harmonic tonal complex with a rectangular spectral envelope. It is hypothesized that the addition of information to a signal is of far greater significance to the auditory system than the removal of information and therefore the onset response mostly accounts for the ACC.

iv. Modify protocol for increased sensitivity:

Decreasing the noise floor could increase the sensitivity of the ACC. The noise floor is dependent in part, on the total number of stimulus presentations. The total number of stimulus presentations is directly related to the amount of testing time: the greater the number of stimulus presentations, the longer the test takes. At present, eight stimulus conditions of 800 stimuli each were presented to subjects in the electrophysiologic experiment. The total time for testing was approximately five and one half hours. The

number of stimulus presentations could be increased without increasing testing time (or perhaps decreasing it) by eliminating the inter-stimulus-interval between stimulus presentations. This could be achieved by creating a long sustained vowel with a square wave modulated F2 frequency.

A second factor influencing response sensitivity relates to the potential problem of habituation. This phenomenon has been noted in the literature relative to the N1-P2 onset response, (Näätänen & Picton, 1987) which decreases in amplitude over several minutes of stimulation. Upon closer examination of the data from the present experiment, it was found (not reported here) that on average, the peak-to-peak amplitude of the ACC elicited by the 600 Hz spectral change condition was 10-20% larger when elicited by the first half of the stimulus presentations compared with the second half. This raises the possibility that decreasing the total number of stimulus presentations contained in a single trial block (thus reducing the degree to which the ACC habituates), and increasing the total number of trial blocks might increase the sensitivity of the ACC. Habituation might also be avoided by simply randomizing the degree of F2 change contained within a single trial block.

v. Establish the reliability of the ACC:

The repeatability of the ACC within subjects across test sessions is not yet known. In order for the ACC to be useful as a clinical measure of speech discrimination capacity, the reliability of the ACC must be established. It would be important to determine the subject factors that influence presence vs. absence of the ACC as well as factors affecting the magnitude of the response in order to maximize response reliability.

vi. Establish the developmental course of the ACC:

Although the influence of maturation on the onset N1-P2 response has not been entirely elucidated, there is research to suggest that the onset response is sensitive to developmental changes. At the very least, there may be differences with age in the neural generators contributing to the auditory evoked potentials in the latency range of the N1 response. For example, Bruneau, Roux, Guérin, Barthélémy, & Lelord (1997) demonstrated maximal amplitude N1 responses at the midtemporal sites in four to eight year old children, whereas the N1-P2 complex is largest at the vertex in adults (Vaughan & Ritter, 1970). It is possible, however, that the major anatomical and physiological changes affecting the N1 wave take place before eight years of age (Tonnquist-Uhlen, Borg, & Spens, 1995). Nevertheless, if the ACC is to be elicited in young children, the nature of changes in the ACC waveform as they relate to such variables as latency, amplitude, morphology, and scalp topography must be established as a function of age.

vii. Test persons with sensorineural hearing impairment:

The peripheral frequency resolution of persons with sensorineural hearing loss is usually poor. In pursuit of clinical validation, it will be important to measure both ACC and behavioral responses in this population of subjects in order to determine how well the two threshold measures correlate.

VII. Clinical Implications of the ACC

There is a general need for objective tests of speech discrimination capacity in young children with hearing impairment. The results of the present study have demonstrated that the sensitivity of the ACC, on the order of 30 to 75 Hz is probably good enough to confirm the basic perceptual requirements for the development of a vowel system. However, these values are based on the group waveform where the noise floor has been reduced by the square root of seven (total number of subjects). To obtain the equivalent signal-to-noise ratio within recordings from individual subjects, the total number of stimulus presentations would have to be increased seven times. Nonetheless, when threshold estimates are based on the curve-fit procedure, it can be argued that even thresholds on the order of 113 Hz (S1) and 131 Hz (S5) are sufficiently low for confirming the potential to discriminate F2 changes among vowels.

There is an additional difficulty (as with many other cortical potentials) of individual variability. Moreover, an inability to detect the frequency change can not be inferred because a subject/patient does not provide an ACC (such as Subject 8 in the current study).

Although not formally addressed in the present experiment, the time required for test administration should be considered for the purposes of determining the clinical utility of the ACC. As noted above, it would be possible, and probably prudent, to reduce the total number of stimulus presentations within a trial block in order to avoid habituation of the ACC. An additional way to decrease testing time would be to remove the inter-stimulus-interval by square-wave modulating the F2 spectral change in a continuous way. That is, subjects would be presented with a continuous stimulus lasting several minutes in which

there was a change that flipped back and forth from /u/ to /i/. The responses time-locked to the change in one direction (e.g. from /u/ to /i/) would be sorted separately from responses time-locked to changes in the other direction (e.g. from /i/ to /u/). Moreover, as a clinical screening tool for determining the potential for the ability to discriminate among vowel changes important for speech perception, the ACC need only be elicited by two F2 change conditions: perhaps 75 and 150 Hz. These modifications to the test protocol could theoretically produce a clinical test of speech discrimination capacity that requires no more time than other commonly used electrophysiologic tests (of auditory threshold) such as the auditory brainstem response test. It is hoped that improving the test protocols will increase the sensitivity of electrophysiologic tests, particularly the ACC, for clinical purposes.

Conclusions

1. A clear ACC was elicited by F2 changes in seven of eight Subjects.
2. Threshold values between 30 and 75 Hz (depending on the criterion) were obtained based on the group waveform.
3. In a binary-choice change/no-change task, detection threshold was around 5 Hz, when defined as the change giving mean scores that were significantly ($p<0.025$) above chance.
4. In the same task, detection threshold was around 20 Hz when defined as the change giving a mean score of 75%.
5. The range of thresholds for reliable detection of ACC in the group waveform corresponded with mean recognition probabilities in the range 86% to 97% in the behavioral task.

6. These findings support the conclusion that reliable ACC responses can be elicited by F2 changes that are perceived with confidence.

7. For both the ACC and behavioral data, there was considerable inter-subject variability.

8. The ACC thresholds, both group and individual, are largely determined by noise amplitude.

9. To the extent that vowel distinctions in English require F2 differences of the order of 100 Hz, these data are encouraging in terms of the potential use of the ACC as an index of speech perception capacity.

10. For such application, however, ways must be found of increasing sensitivity when applied to individual subjects.

APPENDIX A

SOFTWARE CODE – BINARY CHOICE CHANGE DETECTION TASK
Written in the Openscript language for Toolbook from Asymetrix

---

Page id 0 of Book "C:\JODIPARA\JODI.EXE"

```
to handle enterpage
    clear text of field "subject"
end
```

Clears the field where the subject's name gets entered when you enter the page.

Field id 16 ("show") of Page id 0 (p. 1)

```
to handle buttonclick
    show group "trials"
    hide self
end
```

Allows you the option of showing or hiding the number of trials to go.

Field id 14 of Page id 0 (p. 1)

```
to handle buttonclick
    show field "show"
    hide group "trials"
end
```

Button id 12 of Page id 0 (p. 1)

```
to handle buttonclick
    system  int h1,h2,m1,m2
    clear text of field "count"
    h1=0;h2=0;m1=0;m2=0;rndcnt=0
```

h1 = hits when the stimulus is in position 1,
h2 = hits when the stimulus is in position 2,
m1 = misses when the stimulus is in position 1, etc.
This counter is set to zero.

```
    n = text of combobox "runsize"
    X = round(1050+150*text of combobox "stepsize")
        xxx = x&".wav"
        pth = ""
        zzz = pth&xxx
```

runsize (n) can be selected: 10, 20, 30, 40 or 50 stimuli.
the stepsize (1050 + 150 / stepsize) is selected (and rounded up) The *.wav file corresponding to the stepsize is selected.

```
    step d from 1 to n
        rnd = random(2)
        if rnd = 2
        increment rndcnt
        if rndcnt >n/2
        rnd=1
        end if
        end if
```

Stimuli are presented from 1 to n (runsize) and are randomly placed into the 1st or the 2nd stimulus position, but ½ of the total number of stimuli must go into the 1st position.

```
    step c from 1 to 2                          There is a 1 sec. pause after the subject responds
        pause 1 seconds                         (and before the next stimulus is presented) and
        if c = rnd                              between presentations of the 2 stimuli.
        get playsound(zzz,true)                 Either a stimulus with a change is played
        else                                    or the stimulus without the change is played (1050.wav).
        get playsound(pth&"1050.wav",true)
        end if
        end step
        request "Which stimulus had the change?" with "First" or "Second" or "Quit"
        conditions
                when it is "First" and rnd = 1      The program asks the subject to decide
                        increment h1                which stimulus contained a change in the center:
                when it is "First" and rnd = 2      "first" or "second".
                        increment m2                The counter starts to record hits & misses.
                when it is "Second" and rnd = 2
                        increment h2
                when it is "Quit"                   The subject can stop the run by choosing "quit".
                        break
                else
                        increment m1
                end conditions
                text of field "count" = n-d
        end step
        request "h1="&h1&&"    h2="&h2&&"
                m1="&m1&&"  m2="&m2             The program will display hits & misses on
        send savedata                           screen.
end
-----------------------------------------------
to handle savedata                              Data gets saved to a *.csv file called:
        system  int h1,h2,m1,m2                 paratest.csv.
                tm=""
                        n=0
                        st=systime
                        step c from 1 to charcount(st)
                                if character c of st =":"
                                        n=n+1
                                end if
                                        if n<2
                                                tm=tm&character c of st
                                                else
                                                break step
                                        end if
                        end step
                                if "P" is in st
                                        tm=tm&"pm"
                                else
                                        tm = tm&"am"
                                end if
        syssuspend = false
        openfile "paratest.csv"
        syssuspend = true
                if syserror <> ""
                        createfile "paratest.csv"
                         output= \
                         "Subject"          &","&\
```

```
                    "Date"          &","&\
                    "Time"          &","&\
                    "Stepsize"      &","&\
                    "H1"            &","&\
                    "H2"            &","&\
                    "M1"            &","&\
                    "M2"            &","&CRLF
                Writefile output to "paratest.csv"
                end if

        output = Text of field "Subject"                        &","&\
                sysdate                                         &","&\
                tm                                              &","&\
                text of combobox "stepsize"                     &","&\
                h1                                              &","&\
                h2                                              &","&\
                m1                                              &","&\
                m2                                              &","&CRLF
Writefile output to "paratest.csv"
closefile "paratest.csv"
end
-----------------------------------------------
```

APPENDIX B

INDIVIDUAL RESPONSE WAVEFORMS

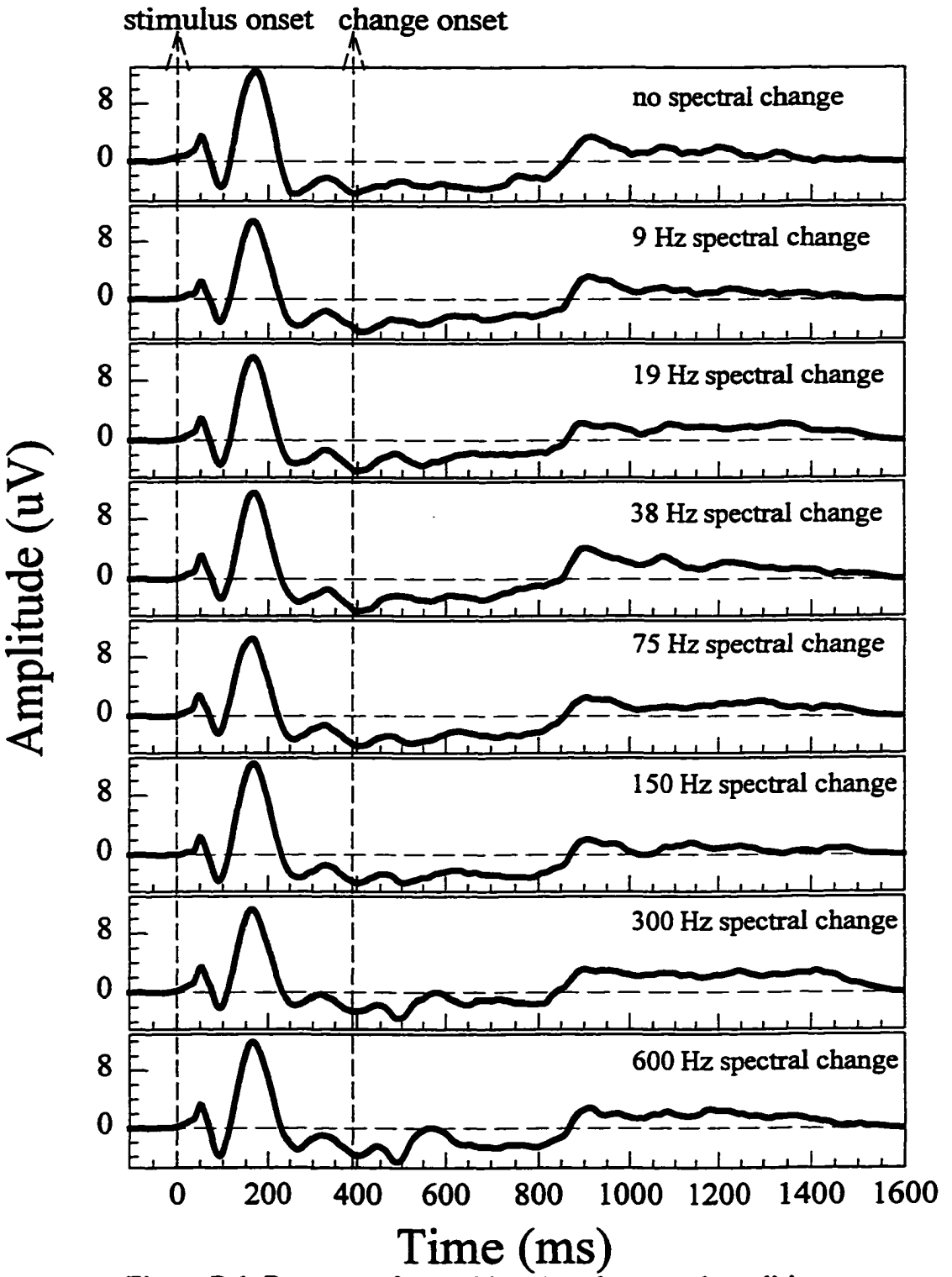


Figure B.1. Responses from subject 1 to the control condition and to 7 degrees of spectral change.

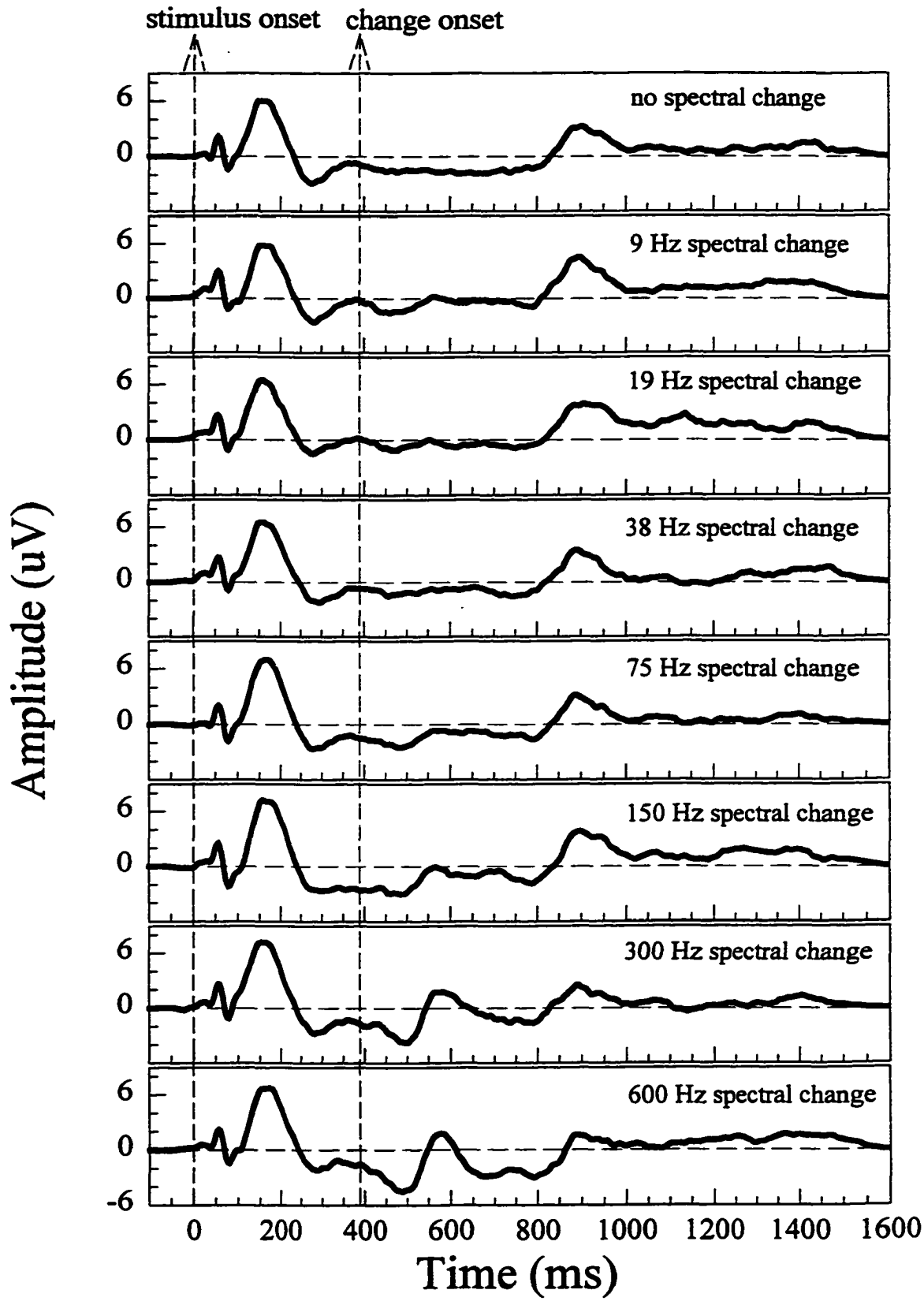


Figure B.2. Responses from subject 2 to the control condition and to 7 degrees of spectral change.

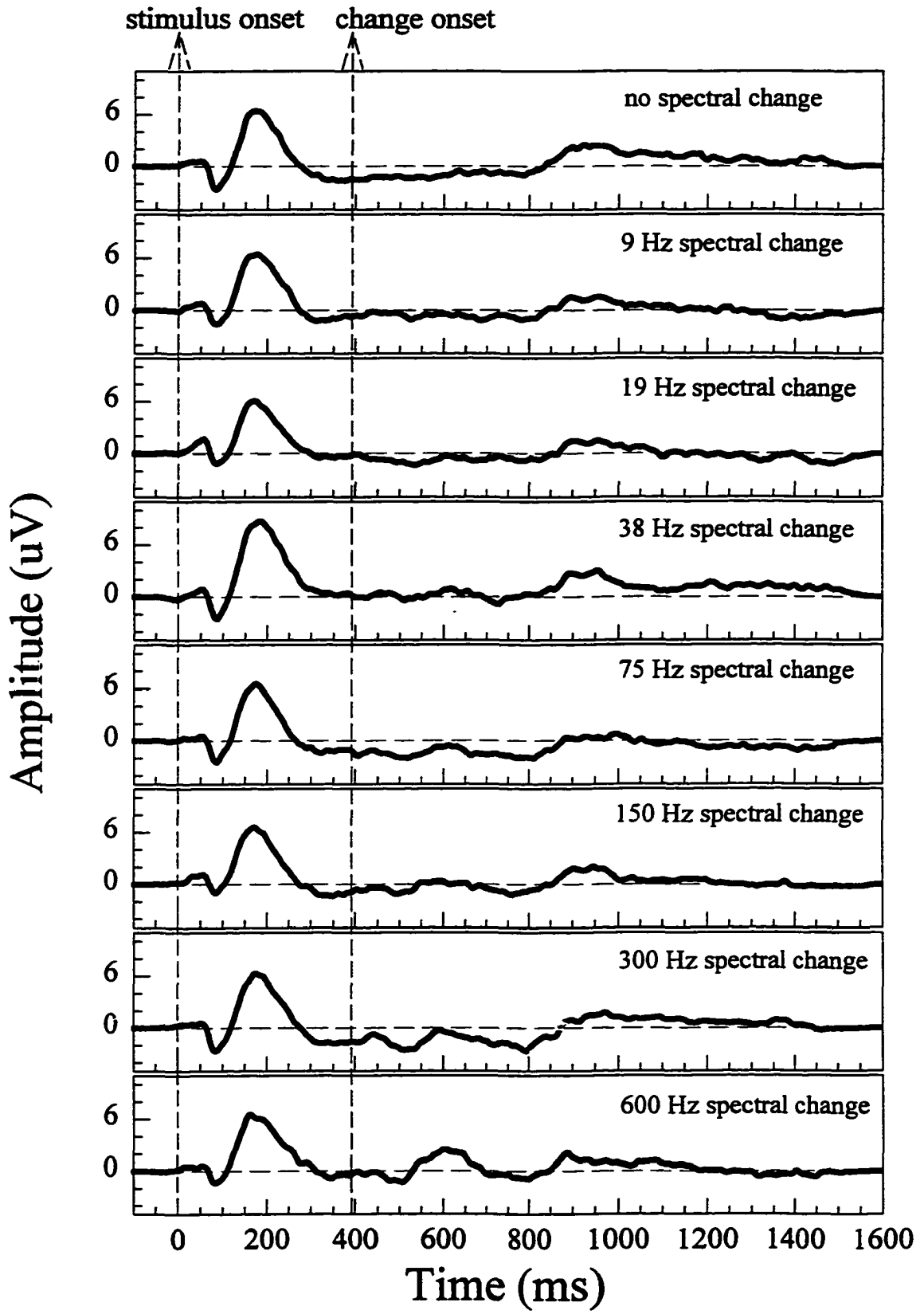


Figure B.3. Responses from subject 3 to the control condition and to 7 degrees of spectral change.

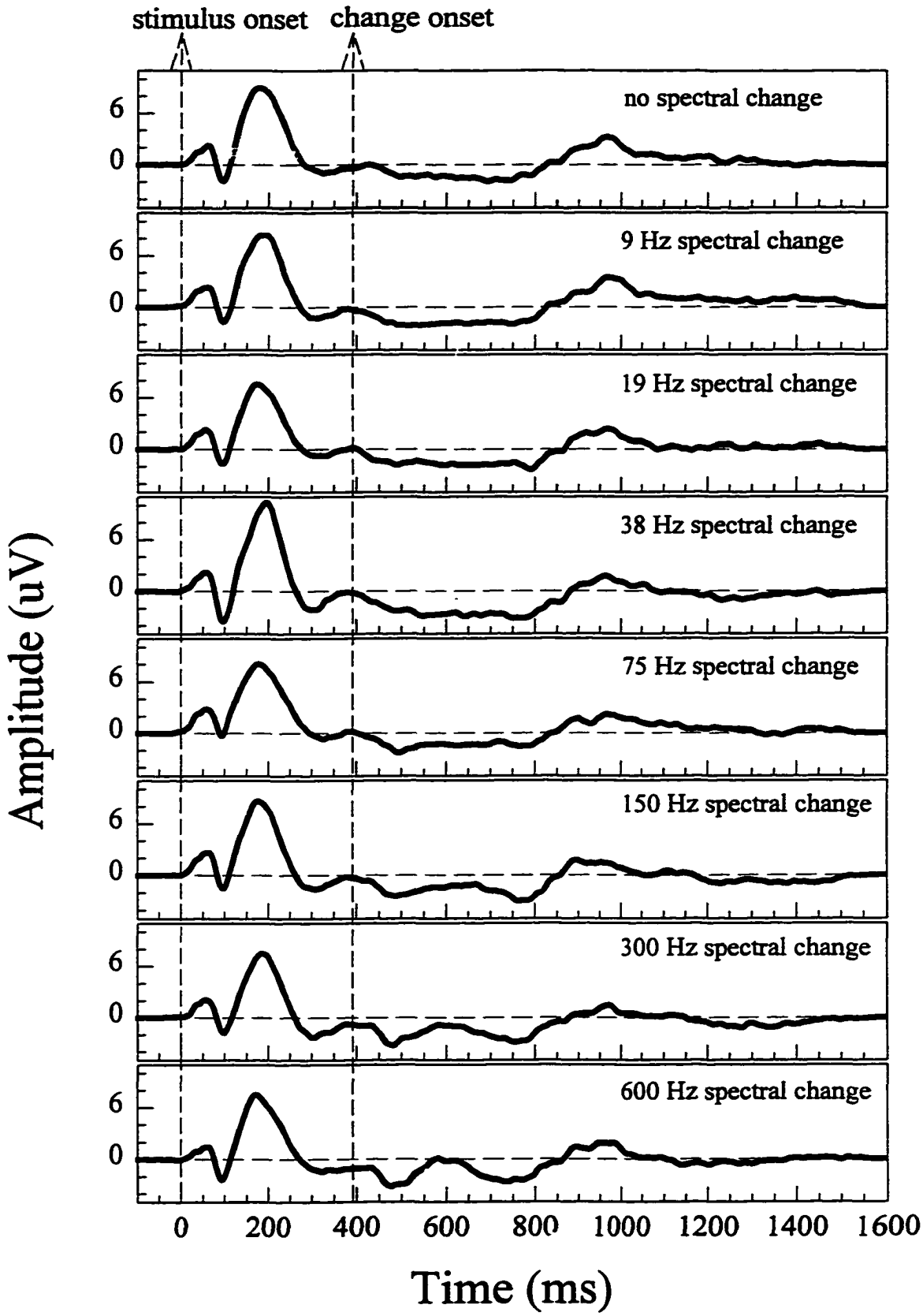


Figure B.4. Responses from subject 4 to the control condition and to 7 degrees of spectral change.

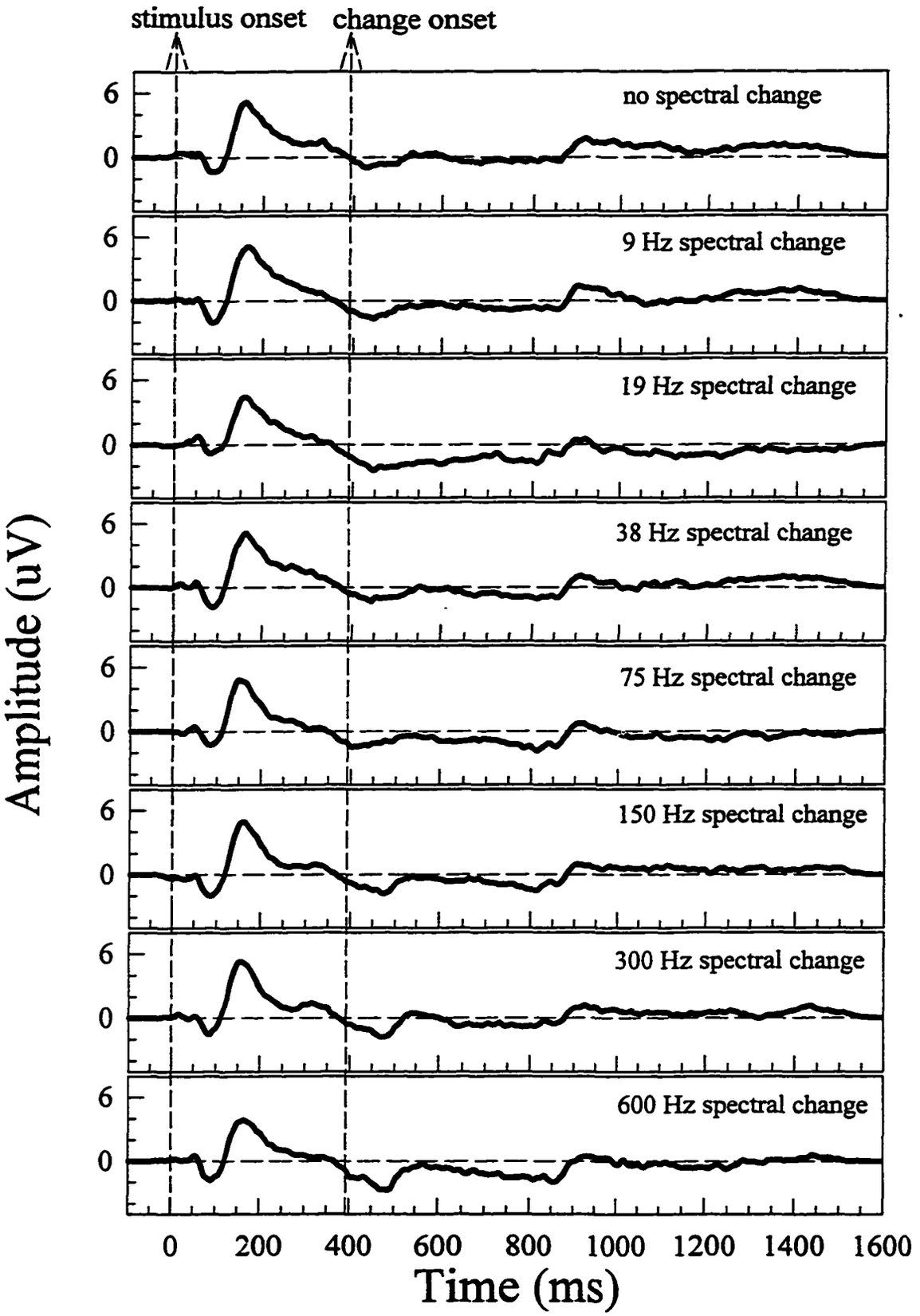


Figure B.5. Responses from subject 5 to the control condition and to 7 degrees of spectral change.

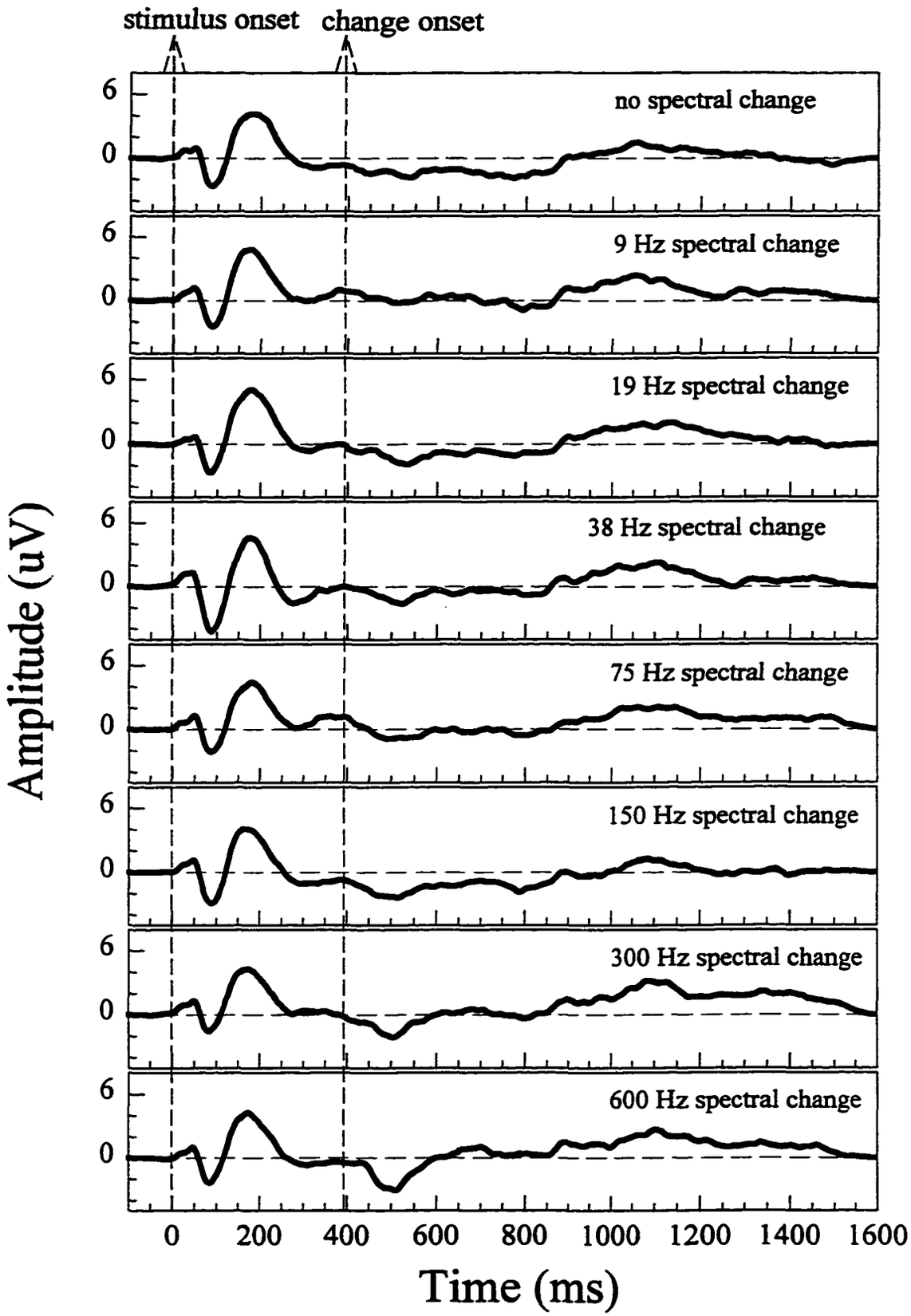


Figure B.6. Responses from subject 6 to the control condition and to 7 degrees of spectral change.

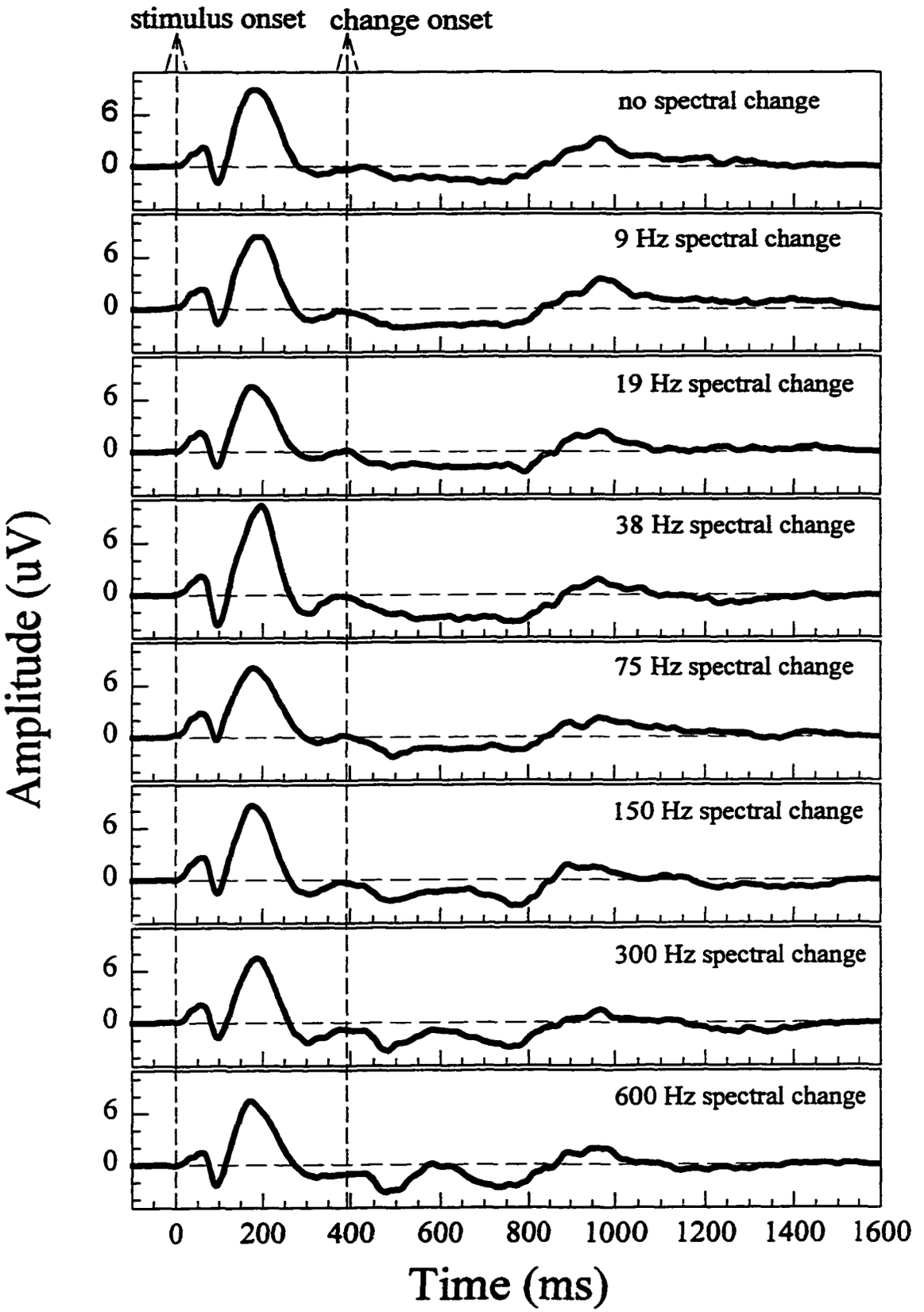


Figure B.7. Responses from subject 7 to the control condition and to 7 degrees of spectral change.

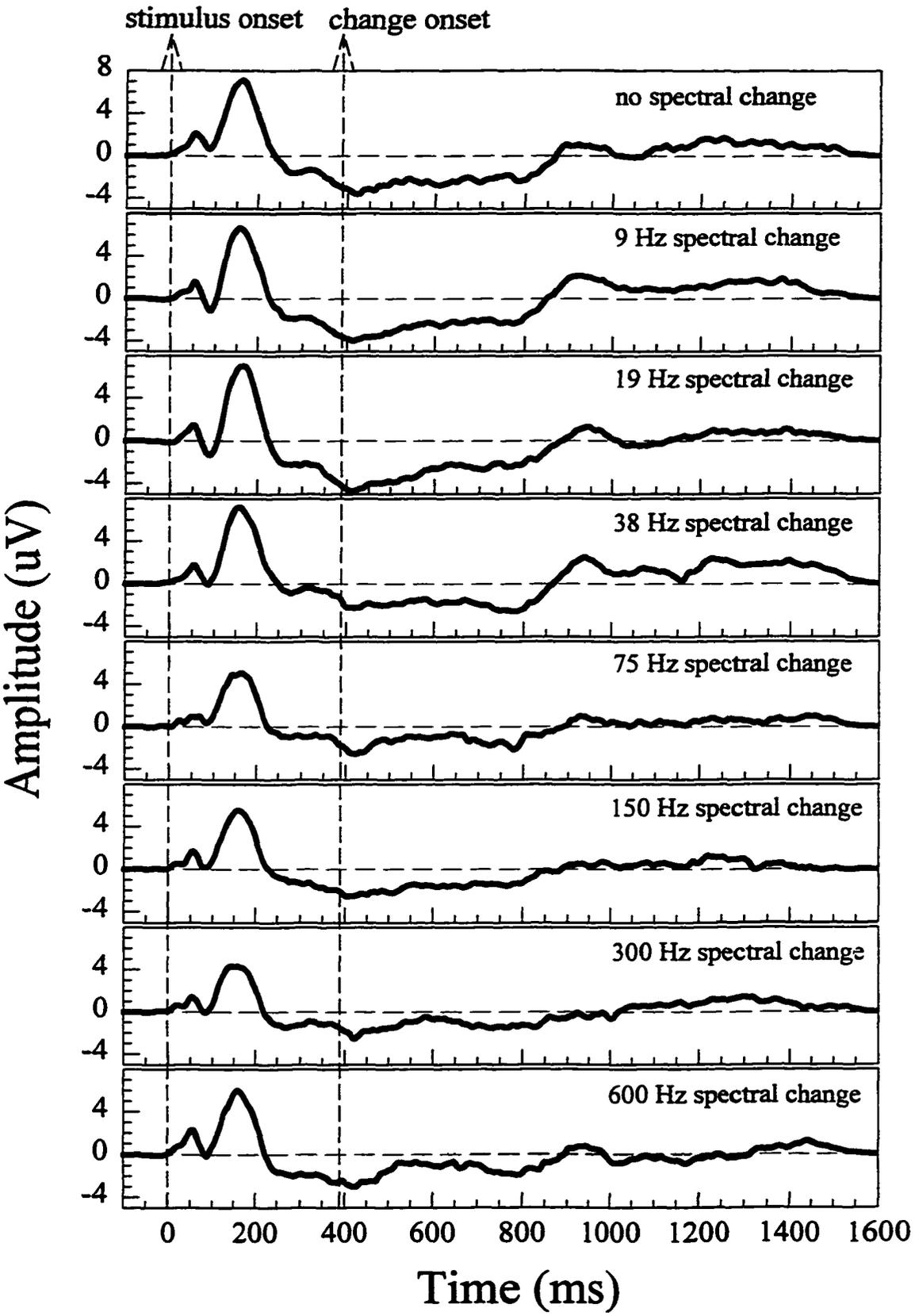


Figure B.8. Responses from subject 8 to the control condition and to 7 degrees of spectral change.

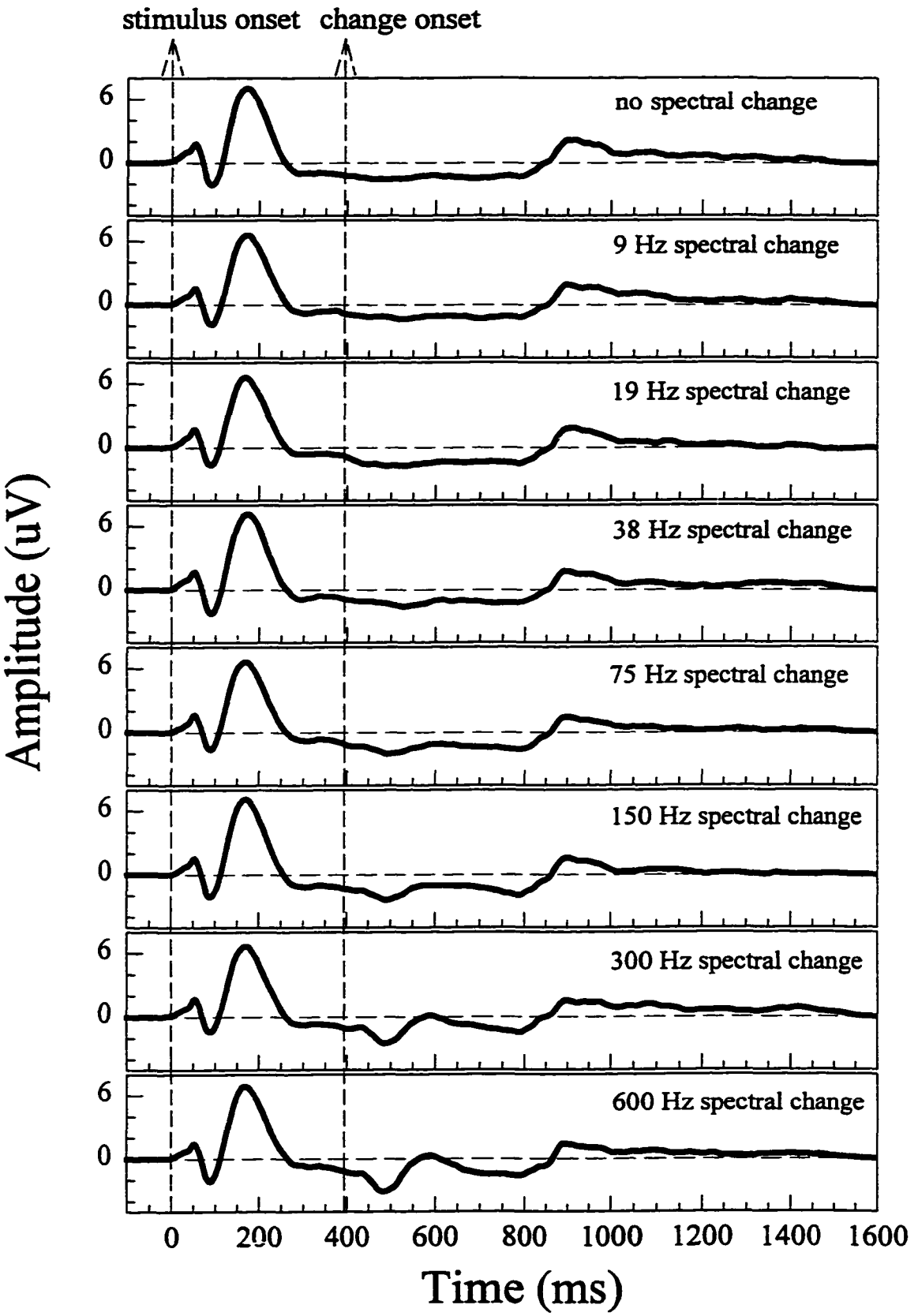


Figure B.9. Responses from the group waveform to the control condition and to 7 degrees of spectral change.

REFERENCES

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. Biological Psychology, 24, 197-297.

Aaltonen, O, Paavilainen, P., Sams, M., & Näätänen , R. (1992). Event-related brain potentials and discrimination of steady-state vowels within and between phoneme categories: A preliminary study. Scandinavian Journal of Logopedics and Phonetics, 17, 107-112.

Arlinger, S.D., Elberling, C., Bak, C., Kofoed, B., & Saermark, K. (1982). Cortical magnetic fields evoked by frequency glides of a continuous tone. Electroencephalography and Clinical Neurophysiology, 54(6), 642-653.

Arlinger, S. D., Jerlvall, L. B. (1979). Results of psychoacoustic and cortical evoked potential experiments using frequency and amplitude modulated stimuli. Scandinavian Audiology, Suppl 9, 229-239.

Arlinger, S.D., Jerlvall, L.B., Ahren, T., & Holmgren, E. C. (1976) Slow evoked cortical responses to linear frequency ramps of a continuous pure tone. Acta Physiologica Scandinavica, 98, 412-424.

Aulanko, R., Hari, R., Lounasmaa, O.V., Näätänen, R., & Sams, M. (1993). Phonetic invariance in the human auditory cortex. Neuroreport, 4, 1356-1358.

Bruneau, N., Roux, P., Guérin, C., Barthélémy, & Lelord, G. Temporal prominence of auditory evoked potentials (N1 wave) in 4-8 year-old children. Psychophysiology, 34, 32-38.

Clynes, M. (1969). Dynamics of vertex evoked potentials: The R-M brain functions In E. Donchin & D.B. Lindsley (Eds.), Average evoked potentials: Methods, results, and evaluations (pp.363-374). NASA SP-191, Washington: US government printing office.

Davis, H., & Zerlin, S. (1966). Acoustic relations of the human vertex potential. Journal of the Acoustical Society of America, 29, 109-116

Delattre, P., Liberman, A.M., Cooper, F.S., & Gertsman, L. (1952). An experimental study of the acoustic determinants of vowel color: Observations on one and two formant vowels synthesized from spectrographic patterns. Word, 8, 195-210.

Eulitz, C., Diesch, E., Pantev, C., Hampson, S., & Elbert, T. (1995). Magnetic and electric brain activity evoked by the processing of tone and vowel stimuli. The Journal of Neuroscience, 15, 2748-2755.

Fry, D.B., Abramson, A.S., Eimas, P.D., & Liberman, A.M. (1962). The identification of synthetic vowels. Language and Speech, 5, 171-189.

Gay, T. (1970). A perceptual study of American English diphthongs. Language and Speech, 13, 65-88.

Hari, R. (1991). Activation of the human auditory cortex by speech sounds. Acta Otolaryngologica, Suppl. 491, 132-138.

Hillenbrand, J. & Gayvert, R.T. (1993). Identification of steady-state vowels synthesized from the Peterson & Barney measurements. Journal of the Acoustical Society of America, 94, 668-674.

Hillenbrand, J. & Gayvert, R.T. (1993). Vowel classification based on fundamental frequencies. Journal of Speech and Hearing Research, 36, 694-700.

Hillenbrand, J., Getty, J., Clark, M.J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. Journal of the Acoustical Society of America, 97, 3099-3111.

Jerger, J., & Jerger, S. (1970). Evoked responses to intensity and frequency change. Archives of Otolaryngology, 91, 433-436.

Jewett, D.L., Romano, M.N., & Williston, J.S. (1970). Human auditory evoked potentials: Possible brainstem components detected from the scalp. Science, 167, 1517-1518.

Kaukoranta, E., Hari, R., & Lounasmaa, O.V. (1987). Responses of the human auditory cortex to vowel onset after fricative stimulation. Experimental Brain Research, 69, 19-23.

Kewley-Port, D. & Watson, C.S. (1994). Formant-frequency discrimination for isolated English vowels. Journal of the Acoustical Society of America, 95, 485-496.

Kohn, M., Lifshitz, K., & Litchfield, D. (1978). Average evoked potentials and frequency modulation. Electroencephalography and Clinical Neurophysiology, 45, 236-243.

Kuriki, S., & Murase, M. (1989). Neuromagnetic study of the auditory responses in right and left hemispheres of the human brain evoked by pure tones and speech sounds. Experimental Brain Research, 77, 127-134.

Ladefoged, P., & Broadbent, D.E. (1957). Information conveyed by vowels. Journal of the Acoustical Society of America, 29, 98-104.

Lenhardt, M.L. (1971). Effects of frequency modulation on auditory averaged evoked responses. Audiology, 10, 18-22.

Liberman, A.M., Harris, K.S., Hoffman, H.S., & Griffith, B.C. (1957). The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology, 54, 358-368.

Lindblom, B., & Studdert-Kennedy, M. (1967). On the role of formant transitions in vowel recognition. Journal of the Acoustical Society of America, 42, 830-843.
Maiste, A., Wiens, A., Hunt, M.J., Scherg, M., & Picton, T.W. (1995) Event-related potentials and the categorical perception of speech sounds. Ear & Hearing, 16, 68-90.

Martin, B.A., & Boothroyd, A. (1999). Cortical, auditory, event-related potentials to changes in periodicity. Ear and Hearing, 20, 33-44.

Näätänen, R., Gaillard, A.W.K., & Mäntysalo, S. (1978). Early selective attention effect on evoked potential reinterpreted. Acta Psychologica, 42, 313-329.

Näätänen, R., Lehtokoski, A., Lennes, M., Cheour, M., Huotilainen, M., Iivonen, A., Vainio, M., Alku, P., Ilmoniemi, R., Luuk, A., Allik, J., Sinkkonen, J., & Alho, K. (1997), Language-specific phoneme representations revealed by electric and magnetic brain responses, Nature, 385, 432-434.

Näätänen, R., & Picton, T. (1987). The N1 wave of the human electric and magnetic response to sound: A review and an analysis of the component structure. Psychophysiology, 24, 375-425.

Neary, T.M. (1989). Static, dynamic and relational properties in vowel perception. Journal of the Acoustical Society of America, 85, 2088-2113.

Onishi, S., & Davis, H. (1968). Effects of duration and rise time of tone bursts on evoked V potentials. Journal of the Acoustical Society of America, 44, 582-591.

Ostroff, J.M., Martin, B.A., & Boothroyd, A. 1998. Cortical evoked response to acoustic change within a syllable. Ear & Hearing. 19, 290-297.

Peterson, G.E., & Barney, H.L. (1952). Control methods in a study of vowels. Journal of the Acoustical Society of America, 24, 585-594.

Poeppel, D., Phillips, C., Yellin, E., Rowley, H.A., Roberts, T.P.L., & Marrantz, A. (1997). Processing of vowels in supratemporal auditory cortex. Neuroscience Letters, 221, 145-148.

Rakerd, B., Verbrugge, R.R., & Shankweiler, D.P. (1984). Monitoring for vowels in isolation and in consonantal context. Journal of the Acoustical Society of America, 76, 27-31.

Rugg, M.D. & Coles, M.G.H. (1995). Event-related brain potentials: an introduction. In M.D. Rugg & M.G.H. Coles (eds.), *Electrophysiology of Mind: Event-related brain potentials and cognition*: Oxford University Press, 1-26.

Ruhm, H. B. (1970). Rate of frequency change and the acoustically evoked response. Journal of Auditory Research, 10, 29-34.

Ruhm, H. B., & Jansen, J. W. (1969). Rate of stimulus change and the evoked response: 1: signal and rise-time. Journal of Auditory Research, 3, 211-216.

Sams, M., Aulanko, R., Aaltonen, O., & Näätänen, R. (1990). Event-related potentials to infrequent changes in synthesized phonetic stimuli. Journal of Cognitive Neuroscience, 2, 344-357.

Semlitch, H.V., Anderer, P., Schuster, P., & Presslich, O. (1986). A solution for reliable and valid reduction of ocular artifacts applied to the P300 ERP. Psychophysiology, 23, 695-703.

Sharma, A., Kraus, N., McGee, T., Carell., & Nicol, T. (1993). Acoustic vs. phonetic representation of speech as reflected by the mismatch negativity event-related potential. Electroencephalography & Clinical Neurophysiology, 88, 64-71.

Skinner, P. H. & Jones, H. C. (1968). Effects of signal duration and rise time on the auditory evoked potential. Journal of Speech and Hearing Research, 11, 301-306.

Spoor, A., Timmer, F., & Odenthal D.W. 1969). The auditory evoked responses to intensity modulated and frequency modulated tones and tone bursts. International Audiology, 8, 410-415.

Steinschneider, M., Schroeder, C.E., Arezzo, J.C., & Vaughan, H.G. (1995). Physiologic correlates of the voice onset time boundary in primary auditory cortex (A1) of the awake monkey: Temporal response patterns. Brain & Language, 48, 326-340.

Strange, W., Verbrugge, R.R., Shankweiler, D.P., & Edman, T.R. (1976). Consonant environment specifies vowel identity. Journal of the Acoustical Society of America, 60, 213-224.

Syrdal, A.K., & Gopal, H.S. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. Journal of the Acoustical Society of America, 79, 1086-1100.

Tonnquist-Uhlen, I., Borg, E., & Spens, K.E. (1995). Topography of auditory long-latency potentials in normal children, with particular reference to the N1 component. Electroencephalography & Clinical Neurophysiology, 95, 34-41.

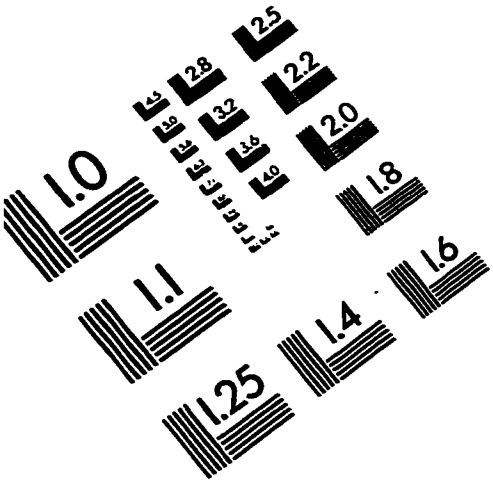
Vaughan, H.G. & Ritter, W. (1970). The sources of auditory evoked responses recorded from the human scalp. Electroencephalography & Clinical Neurophysiology, 28, 360-367.

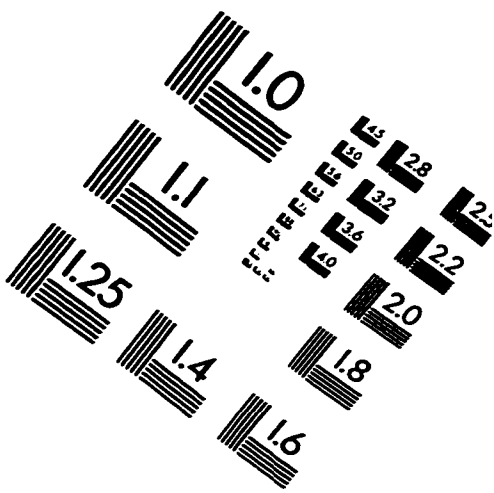
Verbrugge, R.R., & Rakerd, B. (1986). Evidence of talker-independent information for vowels. Language and Speech, 29, 39-57.

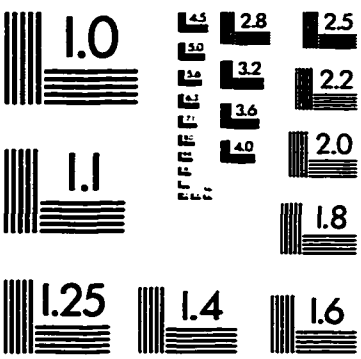
Winkler, I., Lehtokoski, A., Alku, P., Vainio, M., Czigler, I., Csepe, V., Aaltonen, O., Raimo, I., Alho, K., Lang, H., Iivonen, A., & Näätänen, R. (1999). Pre-attentive detection of vowel contrasts utilizes both phonetic and auditory memory traces. Cognitive Brain Research, 7, 357-369.

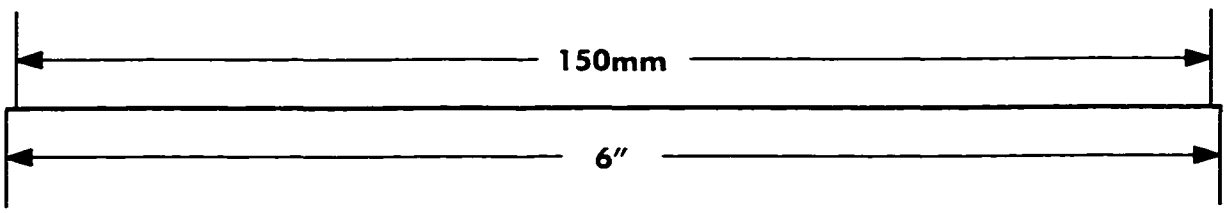
Yingling, C.D., & Nethercut, G.E. (1983). Evoked responses to frequency shifted tones: Tonotopic and contextual determinants. International Journal of Neuroscience, 22(1-2), 107-118.

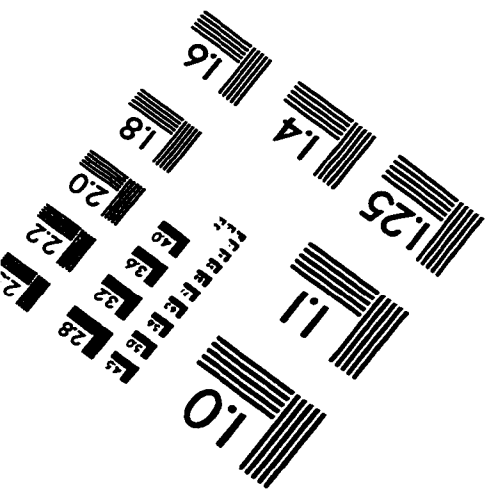
IMAGE EVALUATION
TEST TARGET (QA-3)

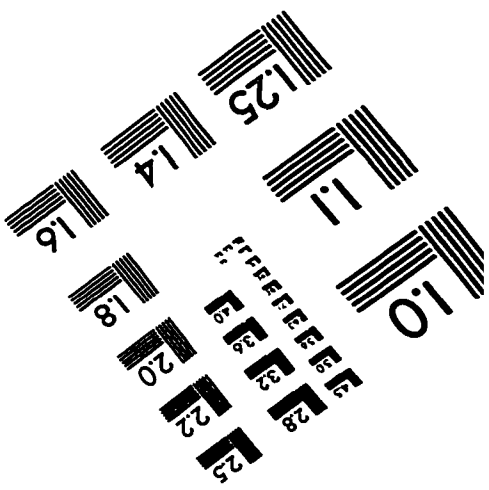












APPLIED IMAGE, Inc
1653 East Main Street
Rochester, NY 14609 USA
Phone: 716/482-0300
Fax: 716/288-5989