NOTICE

The quality of this microform is heavily dependent upon the quality of the original thesis submitted for microfilming. Every effort has been made to ensure the highest quality of reproduction possible.

If pages are missing, contact the university which granted the degree.

Some pages may have indistinct print especially if the original pages were typed with a poor typewriter ribbon or if the university sent us an inferior photocopy.

Reproduction in full or in part of this microform is governed by the Canadian Copyright Act, R.S.C. 1970, c. C-30, and subsequent amendments.

AVIS

La qualité de cette microforme dépend grandement de la qualité de la thèse soumise au microfilmage. Nous avons tout fait pour assurer une qualité supérieure de reproduction.

S'il manque des pages, veuillez communiquer avec l'université qui a conféré le grade.

La qualité d'impression de certaines pages peut laisser à désirer, surtout si les pages originales ont été dactylographiées à l'aide d'un ruban usé ou si l'université nous a fait parvenir une photocopie de qualité inférieure.

La reproduction, même partielle, de cette microforme est soumise à la Loi canadienne sur le droit d'auteur, SRC 1970, c. C-30, et ses amendements subséquents.

HUMAN AUDITORY EVENT-RELATED POTENTIALS TO FREQUENCY CHANGES IN SPEECH AND NON-SPEECH SOUNDS

by ANITA MAISTE

A dissertation submitted to the School of Graduate Studies of the University of Ottawa in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Psychology.

School of Psychology

University of Ottawa

Ottawa, Ontario

ISBN 0-315-60010-1

UNIVERSITÉ D'OTTAWA
UNIVERSITY OF OTTAWA

ACKNOWLEDGEMENTS

I am most grateful to my thesis supervisor Dr. Terence Picton for his advice and guidance. His clear and vigorous thinking was greatly appreciated throughout the work on my thesis.

I am very thankful to the Natural Sciences and Engineering Research Council of Canada for years of funding.

Above all I thank Dean Linden and my parents Arved and Salme Maiste for their continuous and unfailing support and encouragement.

ABSTRACT OF THE DISSERTATION

HUMAN AUDITORY EVENT-RELATED POTENTIALS TO FREQUENCY CHANGES IN SPEECH AND NON-SPEECH SOUNDS

Anita Maiste

University of Ottawa

This thesis presents two approaches investigating how the human auditory system processes the brief frequency changes that occur in speech sounds. Section 1 of the thesis consists of a critical review of the literature on human auditory event-related potentials (ERPs) and speech perception.

Section 2 consists of three experiments evaluating ERPs to non-speech frequency changes. The experiments evaluated steady state and transient auditory evoked potentials (EPs) to tones that were sinusoidally modulated in frequency and to tones that alternated between two frequencies with a linear ramp. The tones were presented at modulation rates typical of average syllable production. The steady state responses to sinusoidal FM were small and difficult to record at both the first and second harmonics. Ramp FM evoked larger and more consistent second harmonic steady state responses than the sinusoidal FM. Only the ramp FM stimuli elicited transient EPs and

these only at low modulation rates. These responses were larger to upward ramps than to downward ramps. The response to two simultaneously presented ramp FM tones differed from the sum of responses to the individual tones indicating some interaction in the processing of the two stimuli.

Since the first study found that steady state responses were not as reliable as ERPs to discrete frequency changes, the second study used speech sounds containing discrete frequency changes. In Section 3 of the thesis computer-modified speech sounds from the /ba/ to /da/ continuum were presented to reading subjects as a train of standard speech sounds interspersed with two types of infrequent deviant speech sounds. One deviant stimulus lay within the same category as the standard and the other lay across the categorical boundary from the standard, but both were acoustically equidistant from the standard in terms of the second formant transition. When the standard stimulus was drawn from the /ba/ end of the continuum, the across-category deviants elicited a clear mismatch negativity (MMN) in the auditory event-related potential whereas the within-category deviants did not. This MMN began at about 60-120 ms after stimulus onset and was present for several hundred ms. These results suggest that categorical processing of speech sounds occurs independently of attention at an early echoic memory stage. When the standard stimulus was drawn from the /da/ end of the continuum, the MMN to across-category deviants was not larger than the MMN to within-category deviants. Grand mean waveforms suggested that both

deviant stimuli elicited small MMNs. Although this may indicate processing along an acoustic continuum, the results of the psychophysical tests suggest that the standard stimulus in this condition was too close to the category boundary for the deviants to evoke a consistent categorical mismatch.

TABLE OF CONTENTS

INTRODUCTION

How does the human auditory system process the brief frequency changes that occur in speech sounds? This thesis describes two approaches to this question using Event-related Potentials (ERPs).

Section 1 is an introduction to the topic of ERPs and speech perception. It reviews the literature on human auditory ERPs and speech perception in adults. It concludes that few reliable ERP indices unique to speech perception have emerged from studies comparing ERPs elicited by speech and by non-speech sounds, because the stimuli have differed on acoustic as well as phonological parameters. Studies that have controlled for acoustic differences between speech and non-speech stimuli have yielded mixed findings. Some studies have suggested that ERPs reflect some aspects of phonological processes such as categorical perception. This section discusses the issues and methodological flaws and suggests guidelines are suggested for research in this area.

Continuous human speech consists of a slowly varying flow of vowel sounds that is periodically amplitude and frequency-modulated by the rapid closures of consonants. The vocal tract acts as a resonator, that is, it emphasizes certain frequencies more than others. Resonances of the vocal tract are called formants and their frequencies are the formant frequencies. Each configuration of the vocal tract involved in producing vowel sounds has its own set of characteristic formant frequencies. The transitions in these formants that occur during the onset and offset of consonantal

modulation are important, since it is the modulation that carries the major share of speech information (Daniloff, Schuckers, & Feth, 1980, p. 306). The identification of stop consonants may involve processing channels in the auditory system that are sensitive to changes in frequency. Section 2 consists of three experiments evaluating ERPs to frequency changes in non-speech sounds. These experiments used frequency-modulated (FM) continuous tones presented at modulation rates typical of average syllable production. The major purpose of this study was to see whether reliable steady-state responses could be recorded at the rates of FM characteristic of normal syllable production, ordinarily between .5 and 11 syllables per second (Hoops, 1972). The following questions were evaluated:

1 - Are there reliable auditory steady state evoked responses to these stimuli? Steady state responses are more easily recorded and more objectively measured than transient ERPs (Picton, 1987). A response of this type would be useful as an objective evaluation of frequency discrimination and could prove useful as an objective (although indirect) measure of speech discrimination.

2 - What are the effects of different rates of FM on steady state and transient ERPs? Picton, Skinner, Champagne, Kellett and Maiste (1987) reported small and noisy steady state response to sinusoidal FM at rates of 3 to 8 Hz. This study extended these results by using a larger number of averages and more individual modulation rates in the 3 to 8 Hz region, and also compared the effectiveness of sinusoidal FM with more speech-like ramp changes.

3 - What are the effects of upward and downward frequency changes, or two simultaneous frequency changes? The simultaneous ramp-changes would be similar to the simultaneous frequency transitions that occur during stop consonants.

This study has already been completed and thus Section 2 is in the form of the paper that has been published in Ear and Hearing. The study found that the steady state responses were not as reliable as the ERPs evoked by discrete frequency changes.

In order to better evaluate questions about how the human auditory system processes the frequency transitions in stop consonants we decided to use real speech stimuli containing these transitions. Thus the experiment in Section 3 of this thesis used speech sounds containing brief frequency transitions, namely simple consonant-vowel speech sounds. The major question was: Is there a unique mode of speech perception or is speech perceived like other auditory stimuli? Some of the strongest evidence to support a special mode of speech perception is the phenomenon of categorical perception (CP) for frequency transitions in speech sounds. This study used ERPs to determine whether at an early pre-attentive stage speech sounds are processed categorically or along an acoustic continuum. This study measured the Mismatch Negativity (MMN) to speech sounds within and across CP boundaries. A separate set of identification tasks was studied to evaluate how subjects identified the computer-modified speech sounds used in the ERP experiment.

References

Daniloff, R., Schuckers, G. & Feth, L. (1980). The physiology of speech and hearing. Englewood Cliffs, N.J.: Prentice-Hall.

Hoops, R.A. (1972). Speech science (pp. 67-72). Springfield: Charles C. Thomas.

Picton, T.W. (1987). Human auditory steady state responses. In C. Barber and T. Blum (Eds.), Evoked potentials III (pp. 117-124). Boston: Butterworths.

Picton, T.W., Skinner, C.R., Champagne, S.C., Kellett, A.J. & Maiste, A.C. (1987). Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. Journal of the Acoustical Society of America, 82, 165-178.

SECTION 1: EVENT-RELATED POTENTIAL STUDIES OF SPEECH PERCEPTION: A CRITICAL REVIEW

INTRODUCTION

This paper reviews the research on event-related potentials and speech perception, discusses the issues and methodological problems involved, and suggests principles to be followed in further research.

A wide body of research exists on the particular neural structures which underlie man's language capacity. Brain areas important in language processes are located in broadly defined language zones in the left hemisphere, including Brocas and Wernicke's areas, in addition to other areas of auditory association cortex. The right hemisphere is involved in restricted language functions (Marin, Schwartz, & Saffran, 1979). A role is also played by subcortical structures, including the basal ganglia and posterior thalamus (Lenneberg, 1967; Whitaker, 1976; Kolb & Wishaw, 1980). Evidence for the specialized role of the left hemisphere in language processes has been obtained from brain lesion studies, the Wada tests, commissurotomy effects and dichotic and split-field studies in normal adults (Marin et al., 1979). However, each of these approaches has shortcomings: explanations vary for the phenomena found in dichotic listening studies (Springer, 1979), and the study of abnormal brains may not provide accurate information on normal language processing. This has led many researchers to conclude with Whitaker (1976) that "there is one gap which seems particularly germane to an understanding of the neurobiology of langage: this is

the relationship between electrical activity in the brain and concurrent linguistic behaviour "(p. 122).

Research in this area is based on two assumptions: one is that tasks which require the extraction of linguistic information from auditory signals will activate unique populations of neurons involved in such an extraction and second, that this linguistic process will in turn produce interpretable patterns of electrical potentials at the scalp (Donchin, McCarthy & Kutas, 1977). Event-related potentials (ERPs) would appear to be an ideal tool for studying the moment-by-moment electrical activity occurring during language processes. These scalp-recorded electrical potentials are associated with the central nervous system activity elicited by a physical stimulus or psychological event. In principle, ERPs could provide information on the relative activation and timing of different regions of the brain during sensory and linguistic processing of language stimuli. ERPs could supply an unobtrusive method of recording language processing stages from normal adults in real-time, instead of inferring language processes from indirect performance measures (eg. word recall on a dichotic listening task). Ideally, ERPs can be recorded when subjects are engaged in a language perception task that also yields performance measures so that the information from both sources can be compared.

One major issue surrounding research on speech perception has been the question of whether or not there is something unique to the perception of speech sounds. In their classic paper Liberman,

Cooper, Shankweiler, and Studdert-Kennedy (1967) state that speech perception does not have a one-to-one relationship with the acoustic signal that conveys it. The speech mode is thought to be a special, complex manner of perceiving that uniquely extracts linguistic information from the acoustic signal. For example, although speech is perceived as a series of discrete, non-overlapping, meaningful units, the acoustic signal viewed on a spectrogram shows that speech sounds are fused to one another, silent periods do not always occur at word boundaries, and that the same phoneme varies in its physical structure depending on the context in which it occurs. Therefore, many researchers assume, that as Rosenfield (1985) states "the brain processes auditory and visual information even remotely associated with language in ways significantly different from its capacity to process pure auditory and visual information" (p. 54). However this issue is unresolved and some researchers have argued that there may be nothing peculiar to perceiving in the speech mode (Cutting, 1978).

The Auditory Event-related Potentials

Auditory event-related potentials are characterized by waveforms with negative and positive peaks identified by a unique nomenclature. The following components have been identified: a vertex-positive P1 peak (50-70 ms in latency), a negative peak N1 at 100-150 ms, and a positive peak P2 at 170-200 ms. The N1-P2 complex is usually followed by a negative peak (N2) centered around 200-275 ms. A third positive peak P3 occurs about 300 ms post-stimulus. The

auditory ERPS are sensitive to many acoustic parameters of the stimuli (Spehlman, 1985). Amplitudes usually decrease and latencies increase with faster stimulus rates, increasing stimulus rise time, and shorter stimulus durations. As stimulus intensity is increased, the ERPs generally reach a maximum amplitude after which further increases in intensity have little effect. This depends on the particular deflection in the waveform and the interstimulus interval.

ERP components can be viewed as a "source of experimenter controlled variability" (Donchin, Ritter & McCallum, 1978, p.352). The search for the processes each component signifies is an ongoing pursuit (see review by Picton & Hillyard, 1988). The N1 may manifest automatic encoding of auditory stimuli. When a subject directs attention to a stimulus, a broad negative shift in potential overlaps the N1 (Hansen & Hillyard, 1980; Näätänen, Gaillard & Mäntysalo, 1978). The N2 component may index the first signs of stimulus evaluation processes and may manifest the detection of stimulus change (Näätänen, Simpson & Loveless, 1982). The N2-P3 complex has been associated with a type of involuntary shift in attention to a stimulus (Näätänen et al., 1982), thus it appears to be equivalent to an orienting response. The late P3 response is elicited by improbable task-relevant stimuli that require a cognitive decision by a subject. Duncan-Johnson and Donchin (1982) found that the P3 closely indexed stimulus evaluation processes. In addition, the P3 may manifest updating of memory processes involved in the formation of a subject's future strategies for information processing

(Pritchard, 1981).

The sources of the auditory ERPs have not been unequivocally established. They have a wide scalp distribution. A spatio-temporal dipole model presented by Scherg & Von Cramon (1985) suggests that the predominant source of the late auditory evoked potential is the sequential activation of primary and secondary auditory cortex areas. The posterior-superior temporal area and adjacent parietal areas are also implicated. Thus the scalp recorded auditory ERP is probably produced by multiple generators, including the auditory cortex (Spehlman, 1985)

This paper reviews ERP research in two broadly defined areas. Early research investigated how ERPs elicited by speech stimuli might differ from responses to non-speech stimuli, and this research was characterized by a search for ERP indications of hemispheric asymmetries during language processing. These experiments can be divided into two groups: (1) experiments in which the speech and non-speech stimuli were not acoustically matched, and (2) those experiments in which either speech and non-speech stimuli were acoustically matched on a number of parameters, or the control was exerted through manipulation of the experimental task required of the subject. This distinction is necessary because auditory evoked potentials are sensitive to a wide range of acoustic parameters, so that any ERP differences between unmatched speech and non-speech sounds could stem from the acoustic differences between stimuli and not language-specific effects.

A third area of research has gone beyond speech/non-speech distinctions to focus on ERPs and phonology, the study of the system of sound patterns that occur in language. This research has not been concerned as much with the localization of language processes, as it has been with the timing and nature of the physiological processes occurring in speech perception. It has moved beyond ERP validation of language related asymmetries, to using ERPs to help resolve phonological issues of current theoretical importance. This area of research will be reviewed in Section 3 of the thesis.

Throughout this review, electrode placement will be specified in the widely used International 10-20 System (Jasper, 1958).

ERPS TO SPEECH AND NON-SPEECH STIMULI

Studies that have evaluated ERPs in response to speech and non-speech sounds, and have not controlled for the acoustic differences between these stimuli, have yielded mixed and inconsistent results.

Haaland (1974) compared ERPs in nine right-handed subjects, in response to consonant-vowel-consonant syllables or tones. Stimuli were presented dichotically, diotically, or monaurally. Electrodes were placed between temporal and frontal sites over the two hemispheres. Haaland found no asymmetries in the N1 component (69-153 ms), whereas the P2 component was larger over the right hemisphere in all conditions. Furthermore, the ERPs did not reflect response accuracy; there was a diotic right ear advantage for identifying words, but no concurrent left hemisphere ERP change.

Using a similar paradigm, Neville (1974) used trains of spoken digits and click stimuli presented either dichotically or monaurally. Recordings were from left and right temporal and central electrodes, referenced to the linked mastoids, in ten subjects. The subjects either counted clicks or reported the digits. In the dichotic listening task subjects had a right ear advantage for the verbal stimuli, and the N1 and N2 latencies were shorter over the left than the right hemisphere for verbal stimuli, whereas click stimuli did not elicit asymmetries. P1-N1 amplitude was larger at left than at

right electrode sites in the verbal condition. In the monaural task condition, P1-N1-N2-P3 components at both hemispheres had shorter latencies to verbal than to nonverbal stimuli. The reported latency differences ranged from 1.2 to 19.6 ms. Although Neville concluded that there were ERP signs of functional asymmetry in the processing of dichotically presented verbal stimuli, the ERPs may have reflected differences in the nature and difficulty of the tasks; subjects were able to report the correct number of clicks but were less accurate in recalling the string of digits.

In another study, Galambos, Benson, Smith, Schulman-Galambos, and Osier (1975) binaurally presented syllables (pa,ba) or pure tone stimuli that were matched in duration. Electrodes were placed at Cz, and over the presumed left and right parietal association areas, referenced to the linked mastoids. No differences among conditions occurred in the peak amplitude measures and there were no response asymmetries within or between any condition. At both hemispheres, the speech stimuli elicited longer N1 and P3 latencies. This might indicate that different processes occurred in the identification of speech stimuli than in the identification of pure tones. However, it may also be related to differences in the onset-time for the syllables and the tones.

Friedman, Simson, Ritter and Rapin (1975) presented five different speech words and five human sounds (eg. whistles, coughs) matched in duration. Subjects listened to the sounds passively or detected designated "signal" stimuli. Electrodes were placed at Pz

and left and right temporo-parietal sites midway between Pz and the mastoid, referenced to the tip of the nose. There were no effects on ERP amplitude. However, when subjects passively listened to stimuli, N1 latency at the left hemisphere was longer for words than for sounds, and sounds evoked longer latency P3's than words, at all electrode sites. Therefore, in this study no systematic asymmetries occurred.

Hink, Hillyard, and Benson (1978) recorded ERPs at Cz and from left and right sites lateral to C3 and C4. Eye movement activity was also recorded. Subjects listened to four different tape-recorded consonant-vowel syllables, or tone pips and white noise. Stimuli were matched in duration (100 ms) and in the 25 ms on-and-offset. Stimuli were presented to both ears and a subject attended to one ear or the other and selected certain designated signal stimuli. Results showed that the tones elicited larger N1 components over the left hemisphere when subjects passively listened to the tone and white noise stimuli, than did the phonemic stimuli when subjects actively attended to them.

Neville (1980) recorded responses from twelve subjects in four different tasks that tested the recall of dichotically presented words or melodies. ERPs were recorded from P3, P4, C3, and C4 referenced to the right mastoid. Neville reported that the N1 to word pairs was larger from left than right parietal leads.

Finally, Grabow, Aronson, Rose and Greene (1980b) presented binaural phonemic stimuli and tone pips to ten right-handed subjects.

Electrodes were placed at right and left mid-temporoparietal sites. Although no significant N1-P2 asymmetries occurred for speech sounds or pure tones, individual subjects did show various lateralization effects.

To summarize the findings of these studies, the main observation was the lack of any consistent speech-related ERP effects. Effects centered around the N1 component, but the effects varied. Neville (1974, 1980) demonstrated that a larger left hemisphere N1 was evoked by verbal stimuli than by non-verbal stimuli. Neville (1980) reported a shorter left hemisphere N1 latency to verbal stimuli whereas Friedman et al. (1975) found that N1 latency was longer for words. However, the two studies differed in stimuli and tasks. Haaland (1974), Galambos et al. (1975), Hink et al. (1978) and Grabow et al. (1980b) found no speech-related ERP asymmetries, only variable asymmetries in some subjects, or findings difficult to account for such as larger left hemisphere amplitudes for the non-speech stimuli (Hink et al.,1978).

Speech sounds differ from non-speech sounds in terms of frequency, formant structure and duration, number of formants, frequency transitions and rise and fall times, as well as multiple frequency modulations. Most of the studies either failed to control for acoustic differences between speech and non-speech sounds or controlled only one or two parameters. Therefore, language-related ERP effects may have resulted from the acoustic properties of stimuli instead of reflecting unique language processing events.

These experiments had other methodological flaws. The tasks were relatively undemanding for the subjects and therefore may not have engaged natural language processes. The studies used too few electrode sites. The binaural presentation of stimuli used by some studies may have actually confounded any ERP asymmetries effects (Picton & Stuss, 1984). These problems are discussed in greater detail in the final section of this introduction.

FURTHER ERP STUDIES WITH SPEECH STIMULI

Some researchers have attempted to control the acoustic differences between speech and non-speech stimuli by either matching them on a number of acoustic parameters, or by using only speech stimuli and manipulating the task of the subject.

Hillyard and Woods (1979) recorded ERPs to natural spoken messages having a high information content in order to engage language processing systems in a less artifical task. Stimuli were electronically edited words in a poem and sequences of control tones precisely matching the words in loudness, inter-stimulus interval, duration and rise/fall times. Stimuli were presented monaurally, and subjects answered questions about the poetry after listening to it. Electrodes were placed over the vertex, over Wernicke's area and over its right hemisphere homologue. The reference electrode was the right mastoid, and EOG activity was monitored. Results showed that the monosyllables elicited a larger N1 than did tones, and that this difference was larger at the left hemisphere. The N1 evoked by

speech sounds was 34% larger in amplitude over the left than the right hemisphere. Hillyard and Woods noted that there was considerable intersubject variability and that the overall ERP asymmetries resulted from seven of twelve subjects.

These results suggest that a more natural linguistic situation and high information load for the subjects may be important in discovering ERP indices of speech processing. However, it is possible that the ERP effects reflected attentional processes, since subjects were more engaged in the poetry task, whereas they simply listened passively to the tones. Therefore, this study is in need of replication with a more appropriate control condition.

Wood and his coworkers have reported hemispheric ERP asymmetries in response to speech sounds when subjects attended to the phonemic aspects of the stimuli and not when they attended to acoustic dimensions of the same stimuli. Wood, Goff and Day (1971) compared ERPs to the same speech sound /ba/ in two different tasks. In the phonemic task subjects discriminated between the low-pitched speech sounds /ba/ and /da/. In the pitch discrimination task, subjects detected a low-pitched /ba/ from a high-pitched /ba/. The responses evoked by the low-pitched /ba/ were compared between the tasks in which the subjects were supposedly making a phonemic or an acoustic discrimination. Recordings in ten subjects were from T3, T4, C3 and C4 referenced to linked ear electrodes. The ERP waveforms were analyzed with Wilcoxon tests computed at each of the 256 individual time points in the responses combined across subjects. Only the

initial 250 ms response was used to avoid confounding by a motor response. The ERP responses over the right hemisphere from 100-200 ms were identical for both tasks, whereas left hemisphere responses were larger in amplitude for the phonemic task than for the acoustic task.

In a replication of this study, Wood (1977) found similar results using the speech sounds /bae/ and /gae/. At the right hemisphere there were no differences in ERPs between the phonetic and pitch discrimination tasks. At the left hemisphere ERPs for the phonetic task were larger at 60-80 ms and 120-140 ms than for the pitch discrimination. Since /bae/ and /gae/ only differ in the initial rapid change in the second formant (the F2 transition), in a second experiment Wood (1977) altered the four syllables so that all other acoustic information was removed except the F2 transitions derived from /bae/ and /gae/ at the two pitches (low and high). Therefore, the stimuli were four non-speech pitch-glides that differed only in F2 transition and pitch. The ERPs did not differ between stimuli, suggesting that the ERP differences between the phonetic and pitch discrimination tasks were related to the process of phonetic categorization, not to detection of specific acoustic cues such as formant transitions.

Using the Wood et al. (1971) paradigm, Grabow, Aronson, Offord, Rose and Greene (1980a) failed to replicate the findings. Ten right-handed males listened to computer-generated standardized speech sound stimuli. Electrodes were placed at T3, T4, C3, C4, C5 and C6,

referenced to linked mastoids. Since no baseline ERP was available, responses were expressed in microvolts as a deviation from the mean value over all conditions for a particular area, hemisphere, task and subject. Grabow et al. found no within hemispheric effects for the same stimulus in the phonemic or acoustic task. Regardless of the task, the left hemisphere temporal response was smaller in amplitude than the right hemisphere response, whereas responses at C3 and C4 had the same amplitude for both tasks. Grabow et al. speculated that a greater level of asynchronous firing may actually result in an attenuation of responses in the dominant hemisphere when it is engaged in processsing of language stimuli, similar to the dominant hemisphere attenuation of the EEG power spectrum that occurs during language processing.

The studies by Wood et al. and Grabow et al. can be criticized on several grounds. Wood et al. (1971) and Wood (1977) used Wilcoxon matched pair signed rank tests without the appropriate adjustment of the alpha level for the number of tests they were performing. In other words, the analysis was biased towards finding significant differences. Electrode recordings were referenced to linked ears and this could have cancelled out any asymmetry at temporal regions. Full scalp distribution analyses were not done. Responses to those stimuli that were designated as target stimuli were not separately analyzed even though they may have more accurately shown signs of actual speech discrimination. Binaural presentation of stimuli might confound signs of lateralized processing (Picton & Stuss,

1984). Furthermore, as Wood (1977) cautioned, it is not certain that subjects were able to make a phonemic discrimination without being aware of the pitch changes. Wood (1977) has presented some evidence from behavioural studies that subjects are able to selectively process the pitch dimension while ignoring the place dimension. However, there is no unequivocal evidence that acoustic and phonemic discrimination are mutually exclusive processes, and they may engage many of the same hemispheric processing units.

ERP STUDIES OF PHONOLOGY

Phonology is the study of the system of sounds that occur in language. Speech scientists have identifed a number of acoustic cues in the speech signal which convey differences in meaning to a listener (Borden & Harris, 1984; Daniloff, Schuckers & Fether, 1980; Fromkin & Rodman, 1974). This literature will be discussed in more depth in the third section of this thesis.

CONCLUSIONS

When language-related ERPs effects do occur they are very small (usually just a few fractions of a microvolt), inconsistent from study to study, and do not appear to be different from responses to other auditory stimuli. Linguistic processes may not be well-represented in the scalp-recorded ERPs. Speech processing may occur in neural networks that do not generate stimulus-locked electrical activity recordable from the scalp. In particular, in a "closed field" neuron pool organization the simultaneously activated pools generate dipole layers for which no scalp recordable activity is evident (Lopes da Silva & Van Rotterdam, 1982).

It is also possible that ERP signs of language function are time-locked to intervening processing operations, rather than to the stimuli. If this is so, language-specific ERPs would have highly variable latencies, and produce an unclear and attenuated ERP signal (Hillard & Woods, 1979).

However, ERPs may accurately reflect the processes occuring during speech perception. One explanation for the lack of robust ERP signs of language-related processes, is that there is no special speech-mode of perception. James Cutting (1978) and his coworkers have demonstrated that many of the effects thought to be peculiar to the speech mode, such as categorical perception, also occur for such musical sounds as plucked or bowed notes from a stringed instrument. Thus, it is possible that ERP components do not differ between speech

and non-speech sounds because there is no unique processing mode for linguistic features.

The lack of major ERP asymmetries in language processing is also gaining agreement from studies of cerebral blood flow and PET scans showing that many language functions can and do involve both cerebral hemispheres (Larsen, Skinhoj & Lassen, 1978).

Before the value of ERPs in the study of speech perception can be realized, a number of methodological flaws must be eliminated. The experiments reviewed here can be criticized on a number of grounds, including stimuli and the mode of presentation, task demands, recording methodology and ERP measurement. Each of these issues will be considered in the following sections along with suggested guidelines for ERP studies of language perception.

(1) Stimuli

One of the main shortcomings inherent in ERP research on speech perception is the repeated number of presentations of speech sounds required to obtain an averaged evoked response out of the surrounding EEG noise. The repeated presentations are not natural and may not engage normal linguistic processes in subjects, since repetitions of even complex words and phrases cause major changes in subjective intelligibility. ERP habituation and adaptation effects may also occur to repeated speech stimuli (Woods & Elmasian, 1986).

One solution to this problem is shown by the Hillyard & Woods (1979) study in which more naturalistic stimuli (word of poetry) were

used. Another promising method is to embed the auditory stimulus for the ERP recording in a verbal passage. Shucard, Cummins, Thomas and Shucard (1981) found that the ERPs to irrelevant tone pips reflected language-related processing when the tones were interspersed in a spoken verbal passage.

Different findings between studies could be attributed to differences in stimuli. Some researchers have relied on tape-recorded edited versions of human speech. This speech is highly natural but allows only a minimum of acoustic control (Melvyn Hunt, personal communication). Research that relies on human speech might produce ERP results differing in non-trivial ways from synthesized speech that is low in naturalness, but high in acoustic control.

(2) Mode of Stimulus Presentation

Neville (1974) and Haaland (1974) have shown that hemispheric asymmetries occur to binaurally presented language stimuli but not to monaurally presented stimuli. However, language-related hemispheric asymmetries could in turn be masked by binaural interaction effects. Furthermore, interaction waveforms are more variable from subject to subject, background noise levels may increase during binaural recording, and the choice of reference electrode is even more difficult (Picton & Stuss, 1984).

(3) Task Demands

As Neville (1980) has stated, ideally ERPs should be recorded from an individual when that individual is engaged in a task designed to produce behavioural indices of language-related effects. In many experiments, no concurrent performance measure of language-related behaviour was evaluated. Experiments should be designed to produce measurable behavioural asymmetries or differences in performance between speech and non-speech stimuli. In contrast, Wood et al. (1971) reported no differences in reaction time between the phonetic and acoustic discrimination tasks. While this was helpful in reducing the contamination of ERPs by motor responses, it provides no basis for rejecting the null hypothesis that subjects discriminated all the stimuli in the same way.

Furthermore, many of the tasks in the experiments reviewed could be characterized as too easy, or sometimes subjects were only required to listen passively to stimuli. Subjects could then react in a number of ways; some might listen attentively and other subjects might tune in and out to the stimuli. All of these factors could add considerably to ERP variability within and between subjects in the attention-sensitive ERPs (Maiste & Picton, 1987).

(4) Recording Methodology

Many of the reported language-related effects involved the N1 component occurring at 110 to 152 ms which is the time range at which

attention-related effects are often manifested. This raises the possibility that attentional factors contributed to the ERP asymmetries that were reported.

Many of the early studies recorded from only one electrode located at Cz. Studies of asymmetry often included two temporal sites. However, the use of a multichannel scalp topography (ie. 16 electrodes) is essential in separating language-related ERP effects from more general perceptual or attention-related processes.

The choice of a reference electrode may also affect results (Picton & Stuss, 1984). Almost any cephalic reference may be affected by far-field potentials. Even the use of linked-ear or linked-mastoid references may very well cancel out asymmetries in the recording. Therefore, the use of a non-cephalic electrode (e.g. larynx, neck, chest) is recommended.

It is necessary to reduce, eliminate, or at least account for sources of artifact. Subjects may move their eyes horizontally toward the stimulated ear in a dichotic listening task (Anderson, 1977). Although temporal and parietal areas are not as greatly contaminated by the fieldspread of eye-movement potentials as frontal regions are, the fieldspread can still affect posterior recordings. None of the studies reviewed eliminated eye movement artifact by the use of a computerized eye movement artifact compensation program, and most studies failed to record both horizontal and vertical eye movements. Therefore, any study of ERP language-related effects should record both vertical and horizontal eye movements.

(6) ERP Measurement

Few studies reported a pre-stimulus baseline against which ERP changes are measured. Asymmetries that occur during language processing can only be evaluated properly if baseline levels are measured first. Furthermore, some subjects may show subtle hemispheric asymmetries for any auditory stimulus (Donchin et al., 1977). The data could also be highly variable due to individual differences in cerebral asymmetry and the neuroanatomical configuration of the speech areas (Hillyard and Woods, 1979).

Concluding Remarks

Research on ERPs and speech perception has not yielded a common paradigm, perhaps because no series of studies has produced a robust yet flexible paradigm. However, despite inconsistent results and lack of replication there have been enough promising results to justify further exploration of this area. Results have been encouraging when researchers have used more naturalistic tasks (Hillyard & Woods, 1979) and when they have interspersed probe stimuli in language passages (Shucard et al., 1981). Interesting results have emerged when researchers have moved beyond simple speech/non-speech distinctions to use ERPs as measures of processes occurring during speech perception (as reviewed in Section 3). Although these findings are not easy to interpret, ERPs should not

be expected to present a simplistic picture of what are undoubtedly highly complex processes of language perception.

In the future, if researchers eliminate some of the methodological flaws discussed in the previous section, and use the extensive language behavioural research to help in the design of experiments, then research in ERPs and speech perception may yet make a unique contribution to the understanding of human language processes.

REFERENCES

Anderson, S.W. (1977). Language-related asymmetries of eye-movement and evoked potentials. In S. Harnad, R.W. Doty, L. Goldstein, J. Jaynes and G. Krauthamer (Eds), Lateralization in the nervous system. New York: Academic Press.

Borden, G.J. & Harris, K.S. (1984). Speech science primer (2nd ed.). Baltimore: Williams & Wilkins.

Cutting, J.E. (1978). There may be nothing peculiar to perceiving in a speech mode. In J. Requin (Ed.), Attention and performance VII (pp. 229-243). Toronto: John Wiley & Sons.

Daniloff, R., Schuckers, G. & Feth, L. (1980). The physiology of speech and hearing. Englewood Cliffs, N.J.: Prentice-Hall, Inc.

Donchin, E., McCarthy, G., & Kutas, M. (1977). Electroencephalographic investigations of hemispheric specialization. Progress in Clinical Neurophysiology, 3, 212-242.

Donchin, E., Ritter, W., & McCallum, W.C. (1978). Cognitive psychophysiology: the endogenous components of the ERP. In

E.Callaway, P. Tueting & S.H. Koslow (Eds.), Event-related brain potentials in man (pp. 349-411). New York: Academic Press.

Duncan-Johnson, C.C. & Donchin, E. (1982). The P300 component of the event-related brain potential as an index of information processing. Biological Psychology, 14, 1-52.

Friedman, D., Simson, R., Ritter, W. & Rapin, E. (1975). Cortical evoked potentials elicited by real speech words and human sounds. Electroencephalography and Clinical Neurophysiology, 35, 13-19.

Fromkin, V. & Rodman, R. (1974). An introduction to language. New York: Holt, Rinehart and Winston.

Galambos, R., Benson, P., Smith, T.S., Schulman-Galambos, C. & Osier H. (1975). On hemispheric differences in evoked potentials to speech stimuli. Electroencephalography and Clinical Neurophysiology, 39, 279-283.

Grabow, J.D., Aronson, A.E., Offord, K.P., Rose, D.E., & Greene, K.L. (1980a). Hemispheric potentials evoked by speech sounds during discrimination tasks. Electroencephalography and Clinical Neurophysiology, 49, 48-58.

Grabow, J.D., Aronson, A.E., Rose, D.E. & Greene, K.L. (1980b). Summated potentials evoked by speech sounds for determining cerebral dominance for language. Electroencephalography and Clinical Neurophysiology, 49, 38-47.

Haaland, K.Y. (1974). The effect of dichotic, monaural and diotic verbal stimuli on auditory evoked potentials. Neuropsychologia, 12, 339-345.

Hansen, J.C. & Hillyard, S.A. (1980). Endogenous brain potentials associated with selective auditory attention. Electroencephalography and Clinical Neurophysiology, 49, 277-290.

Hillyard, S.A. & Woods, D.L. (1979). Electrophysiological analysis of human brain function. In M.S. Gazzaniga (Ed.), Handbook of behavioural neurobiology Vol. 2. (pp. 345-378). New York: Plenum Press.

Hink, R.F., Hillyard, S.A., Benson, P.J. (1978). Event-related brain potentials and selective attention to acoustic and phonetic cues. Biological Psychology, 6, 1-16.

Jasper, H.H. (1958). Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology, 10, 370-375.

Kolb, B. & Wishaw, I.Q. (1980). Fundamentals of human neuropsychology. San Francisco: W.H. Freeman.

Larsen, N.A., Skinhoj, E. & Lassen, N.A. (1978). Variation in regional cortical blood flow in the right and left hemispheres during automatic speech. Brain, 101, 193-209.

Lenneberg, E.H. (1967). Biological foundations of language. New York: John Wiley & Sons.

Liberman, A.M., Cooper, F.S., Shankweiler, D.P., & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74, 431-461.

Lopes da Silva, F. & Van Rotterdam, A.G. (1982). Biophysical aspects of EEG and MEG generation. In F. Lopes da Silva & E. Niedermeyer (Eds.), Electroencephalography: Basic principles, clinical applications and related fields (pp 15-26). Baltimore: Urban & Schwarzenberg.

Maiste, A.C. & Picton, T.W. (1987). Selective auditory attention. C. Barber & J. Blum Eds.), Evoked Potentials III. NY: Butterworth. 385-391.

Marin, O.S.M., Schwartz, M.F., Saffran, E.M. (1979). Origins and distribution of language. In M.S. Gazzaniga (Ed.), Handbook of behavioural neurobiology Vol. 2. (pp. 179-213). New York: Plenum Press.

Näätänen, R., Gaillard, A. & Mäntysalo, S. (1978). The N1 effect of selective attention reinterpreted. Acta Psychologica (Amst.), 42, 313-329.

Näätänen, R., Simpson, M. & Loveless, N.E. (1982). Stimulus deviance and evoked potentials. Biological Psychology, 14, 53-98.

Neville, H. (1974). Electrographic correlates of lateral asymmetry in the processing of verbal and nonverbal auditory stimuli. Journal of Psycholinguistic Research, 3(2), 151-163.

Neville, H. (1980). Event-related potentials in neuropsychological studies of language. Brain and Language, 11, 300-318.

Picton, T.W. & Hillyard, S.A. (1988). In T.W. Picton (Ed.), Human event-related potentials, EEG Handbook (Vol. 3) (pp. 361-426). Amsterdam: Elsevier.

Picton, T.W., & Stuss, D.T. (1984). Event-related potentials in the study of speech and language: A critical review. In D. Caplan,

Picton, T.W., & Stuss, D.T. (1984). Event-related potentials in the study of speech and language: A critical review. In D. Caplan, A.R. Lecours, & A. Smith (Eds.), Biological perspectives on language (pp. 303-360). Cambridge, MA: MIT Press.

Pritchard, W.S. (1981). The psychophysiology of P300. Psychological Bulletin, 89, 506-540.

Rosenfield, I. A hero of the brain. (1985, Nov. 21) The New York Review of Books, pp. 49-55.

Scherg, M. & Von Cramon, D. (1985). Two bilateral sources of the late AEP as identifed by a spatio-temporal dipole model. Electroencephalography & Clinical Neurophysiology, 62, 32-44.

Shucard, D.W., Cummins, K.R., Thomas, D.G. & Shucard, J.L. (1981). Evoked potentials to auditory probes as indices of cerebral specialization of function -- replication and extension. Electroencephalography & Clinical Neurophysiology, 52, 389-393.

Spehlman, R. (1985). Evoked potential primer. Boston: Butterworth Publishers.

Springer, S.P. (1979). Speech perception and the biology of language. In M.S. Gazzaniga (Ed.), Handbook of behavioural

neurobiology (Vol. 2), (pp. 153-173). New York: Plenum Press.

Whitaker, H.A. (1976). Neurobiology of language. In E.C. Carterette & M.P. Friedman (eds.), Handbook of perception (Vol. 7) (pp. 121-144). New York: Academic Press.

Wood, C.C. (1977). Average evoked potentials and phonetic processing in speech perception. In J.E. Desmedt (Ed.), Language and hemispheric specialization in man: Cerebral ERPs. Progress in clinical Neurophysiology (pp 73-86). Basel: S. Karger.

Wood, C.C., Goff, W.R. & Day, R.S. (1971). Auditory evoked potentials during speech perception. Science, 173, 1248-1251.

Woods, D.L., & Elmasian, R. (1986). The habituation of event-related potentials to speech sounds and tones. Electroencephalography and Clinical Neurophysiology, 65, 447-459.

in Ear and Hearing

Vol. 10(3), 1989

SECTION 2:

HUMAN AUDITORY EVOKED POTENTIALS TO FREQUENCY-MODULATED TONES

Anita Maiste, M.A. and Terence Picton, M.D., Ph.D.

Human Neurosciences Research Unit

University of Ottawa

Ottawa, Ontario

Telephone: (613) 737-8159

Short Title: Evoked Potentials to FM

Correspondence: Anita Maiste

Human Neurosciences Research Unit

University of Ottawa

451 Smyth Road

Ottawa, Ontario

CANADA K1H 8M5

ABSTRACT

Three experiments evaluated steady state and transient auditory evoked potentials (EPs) to tones that were sinusoidally modulated in frequency and to tones that alternated between two frequencies with a linear ramp. The steady state responses to sinusoidal FM were small and difficult to record at both the first and second harmonics. Ramp FM evoked larger and more consistent second harmonic steady state responses than the sinusoidal FM. Only the ramp FM stimuli elicited transient EPs and these only at low modulation rates. These responses were larger to upward ramps than to downward ramps. The response to two simultaneously presented ramp FM tones differed from the sum of responses to the individual tones indicating some interaction in the processing of the two stimuli.

Key Words: auditory evoked potentials
steady state potentials
sinusoidal frequency-modulation
frequency ramps

Acknowledgements: The research was supported by the Medical Research Council of Canada and the Ontario Deafness Research Foundation. Anita Maiste was supported by the Natural Sciences and Engineering Research Council of Canada. Adrian Kellett provided technical assistance and Edna Leech provided computer assistance.

INTRODUCTION

Psychophysical studies have demonstrated the human auditory system has channels that specifically process frequency-changes. Exposure to frequency-modulated (FM) tones causes adaptation of the FM-channel with higher thresholds for detecting FM stimuli, whereas exposure to amplitude-modulated (AM) tones does not change the thresholds for FM stimuli (Kay and Matthew, 1972; Regan and Tansley, 1979; Tansley and Suffield, 1983). The FM-channel appears to be sensitive to the direction of a frequency change. Collins and Cullen (1978) and Gardner and Wilson (1979) reported that subjects had a higher threshold for detecting brief (75 ms or less) downward "glides" (tones that decreased in frequency) than for detecting upward glides. Both Gardner and Wilson (1979) and Tansley and Regan (1979) found FM adaptation specific to the direction of a frequency change. After subjects adapted to upward glides they had an elevated threshold for upward glides but not for downward glides, and vice versa.

Whitfield and Evans (1965) first reported that units in the primary auditory cortex of the unanesthetized cat were specifically sensitive to frequency-changes. They found that some units exhibited a directional preference, a finding replicated by Phillips, Mendelson, Cynader, and Douglas (1985). Since the early studies on the cortex, units sensitive to FM-stimuli have been found in the auditory nerve (Sinex and Geisler, 1981), cochlear nucleus (Erulkar, Butler and Gerstein, 1968; Møller, 1969, 1971, 1972, 1974), and inferior colliculus (Rees and Møller, 1983).

Several studies have described the human transient EPs to ramp changes in frequency (Spoor et al., 1969; Ruhm, 1970, 1971; Lenhardt, 1971; Arlinger, Jerlvall, Ahren and Holmgren, 1976). Ruhm (1971) and Kohn, Lifshitz and Litchfield (1978) observed larger EPs in response to upward than downward frequency ramps. In contrast, Arlinger et al. (1976) and Arlinger and Jerlvall (1979) did not find any changes in EPs to the direction of the frequency change. In 1969 Clynes suggested that the N1-P2 response to a change in the frequency of a continuous tone is evoked by the onset of a change from a stable state - the "Rest-Motion" response. He found that the N1-P2 was not elicited by tones modulated by a triangular wave so as to be continuously changing in frequency. The Rest-Motion response has a refractory period of 300 ms, the stimulus requiring a stable frequency for greater than this time before any change could evoke a potential. Similarly, Kohn et al. (1978) found that when a slow change in frequency preceded a ramp-change, the N1-P2 response to the ramp was attenuated or disappeared. According to these findings, a sinusoidal FM tone should not elicit an N1-P2 response.

Steady state evoked potentials are evoked by stimuli presented at a rate fast enough to cause an appreciable overlap of successive responses (Regan, 1982; Picton, 1987). Picton, Skinner, Champagne, Kellett, and Maiste (1987) recorded steady state responses to sinusoidal FM stimuli. The responses at modulation rates in the 3 to 7 Hz region were small and much more difficult to recognize than the responses at 40 Hz. The Clynes (1969) and Kohn et al. (1978) findings with transient responses suggest that an FM stimulus consisting of ramp frequency-changes alternating with stable

frequency intervals may be more effective at eliciting a steady state response than sinusoidal FM.

The purpose of the present study was to evaluate whether a reliable steady state response could be recorded to slow rates of FM similar to the rates of syllable production - ordinarily between .5 and 11 syllables per second (Hoops, 1972). A reliable steady state response to FM stimuli could prove useful in evaluating frequency-discrimination. A second purpose was to compare the effectiveness of sinusoidal FM with more speech-like ramp changes. A third purpose was to evaluate responses to simultaneous ramp-changes similar to the frequency-transitions of stop-consonants.

EXPERIMENT 1

The purpose of the first experiment was to see whether a reliable steady state response could be evoked by sinusoidal FM stimuli modulated at rates between 3 to 8 Hz. This experiment extended the results of Picton et al. (1987) who had found small and noisy responses in this frequency range. In the present experiment we used more EP averages at each modulation rate in order to increase the signal-to-noise ratio, and more individual modulation rates in the 3 to 8 Hz region.

Methods

Stimulus

The stimulus was a FM tone with a centre frequency of 1000 Hz. The tone was sinusoidally modulated across the range 750 to 1250 Hz. The modulation rate was set at 2 cycles per recording sweep; the duration of the sweep was varied to obtain modulation rates of 3.1, 3.9, 5.2, 6.5 and 7.8 Hz. The tone was presented to the subject's right ear through a TDH-49 earphone. Every 10 seconds a pause of between 1 and 2 seconds interrupted the tone.

Stimuli were presented at 2 intensities (40 dB HL and 70 dB HL) and at 5 modulation rates (3.1, 3.9, 5.2, 6.5, and 7.8 Hz). The stimuli were randomly presented in two blocks that were counterbalanced for intensity across subjects.

Recording

Eight adults (23-42 years) with normal hearing participated in this experiment. During the experiment the subject sat comfortably in a sound-insulated and electrically shielded room. The subject read during the recordings, and we monitored the EEG to ensure that a subject did not fall asleep.

Electroencephalographic (EEG) activity was recorded from gold-plated cup electrodes fixed to the scalp with collodion-soaked gauze. Recordings were taken between the vertex (Cz) and the right mastoid (M2). The EEG signals were amplified and filtered with a band-pass of 0.16 to 100 Hz. Negativity recorded at Cz relative to M2 produced an upward deflection in the waveform.

Since we changed the length of the sweep to produce the

different modulation rates, the sweep ranged from 256 ms to 896 ms. The number of data points per sweep was 256. Rotating buffers allowed continuous stimulation and recording - while one sweep was recorded the previous sweep was analyzed.

Analysis

The recorded data were subaveraged in 16 blocks of 50 waveforms. These waveforms were then evaluated by Fourier analysis (Regan, 1977; Stapells, Linden, Suffield, Hamel and Picton, 1984). This procedure effectively filters the response with a narrow bandpass at the frequency of stimulation, and provides the amplitude and phase of the response at that frequency. The reliability of the response was assessed using Hotelling's T^2 test (Anderson, 1958; Picton et al., 1987). A Fast Fourier Transform provided the frequency-spectrum of the average response (800 sweeps). This allowed us to determine the amplitude and phase of the response at both the frequency of the stimulus (first harmonic - f_o) and the second harmonic ($2f_o$) of the stimulus.

Average amplitudes were calculated by vector-averaging. This procedure takes into account both the amplitude and phase of the individual measurements (Picton et al. 1987).

The phase-delay in degrees (Rodriguez, Picton, Linden, Hamel and Laframboise, 1986) was calculated as 360° minus the phase of the response at zero-time. This measurement is more homologous to latency than the phase at zero-time. An apparent latency (Regan, 1972) was calculated as the slope of the regression of phase-delay on the rate of the response divided by 360°. The formula for apparent

latency was thus:

$$\frac{(d\theta/df)}{360N}$$

where θ is phase, f is the frequency of the stimulus, and N is the number of the harmonic (Regan, 1988b).

Statistical Analyses

The amplitudes and phase-delays were evaluated in a two-way (harmonic by modulation-rate) repeated measures analysis of variance (ANOVA). Geisser-Greenhouse adjustments were used to calculate the probability levels (Keppel, 1982). Statistically significant interactions were evaluated for simple main effects and significant main effects were tested post-hoc with Newman-Keuls procedures. The significance level was set at $p<0.05$. Linear regressions of phase-delay on response-frequency provided the apparent latency of the responses (Regan, 1972).

Results

Reliability

There were too few reliable responses to the 40 dB HL stimulus to justify further analysis. The individual and group T^2 values for the 70 dB HL responses are summarized in Table 1. Only a few individual responses were reliably different from noise. Responses at f_o were most convincing at the 3.9 and 5.2 Hz modulation rates. Although individual T^2 values were not available for the $2f_o$

response, the group T^2 analysis showed consistent responses at all frequencies except 5.2 Hz.

Insert Table 1 about here

Amplitude

The vector-averaged amplitudes of the responses plotted over response rate are shown in Figure 1. The analysis of the 70 dB HL data yielded a harmonic x rate interaction ($F=6.98$; $df=4,28$; $p<0.01$). Simple main effects showed that modulation rate had no significant effect on the f_o response amplitude whereas the $2f_o$ response decreased as the rate increased.

Insert Figure 1 about here

Apparent Latency

The apparent latency of the f_o response to the 70 dB HL stimuli, over the rates showing a reliable response (3.9, 5.2, and 6.5 Hz) was 125 ms. The apparent latency for the $2f_o$ component, calculated over all five response rates, was 41 ms.

Discussion

At the fundamental frequency, 70 dB HL stimuli evoked consistent responses between 3.9 and 6.5 Hz with 3.9 Hz being the most effective stimulus at evoking a reliable steady state response. However, even this response was reliable in only six of the eight subjects.

Second harmonic responses were consistent across the group for

the 70 dB HL stimulus except at the 5.2 Hz modulation rate. The absence of a clear $2f_o$ response at the 5.2 Hz rate was perhaps related to the higher level of background noise at 10.4 Hz - the EEG alpha rhythm. The average $2f_o$ response for the 5.2 Hz stimulus was not lower in amplitude than other responses.

The apparent latency gives the latency of that part of the response contributing most prominently to the steady state response (Picton, 1987). Our results suggest that an N1 component may contribute to the f_o steady state response and that middle latency responses may contribute to the $2f_o$ steady state response.

We conclude that, although there are f_o and $2f_o$ responses to sinusoidal FM between 3 and 8 Hz, these responses are small and not easy to record even at high intensities.

EXPERIMENT 2

Brief frequency ramps might be more effective than sinusoidal FM in eliciting reliable steady state responses. We therefore compared responses to frequency ramps with the responses to sinsuoidal FM. At slow rates the ramp stimuli should elicit transient as well as steady state responses whereas sinusoidal FM should only elicit steady state responses. In this experiment we calculated T^2 values for individual subjects at both the f_o and $2f_o$. We recorded responses from right and left temporal electrodes as well as from the vertex.

Methods

Stimulus

Two types of FM stimuli were presented to a subject's right ear. One stimulus was the 70 dB HL "sinusoidal" FM tone used in the first experiment. The other stimulus - "ramp" FM - was a continuous 70 dB HL tone that alternated in frequency between 750 and 1250 Hz. The ramp change between the two frequencies was always 42 ms in duration. Two frequency ramps occurred within one cycle of this alternating tone and the modulation rate was defined as the number of cycles per second. We used modulation rates of .49, .98, 1.95, 3.9 and 7.8 Hz. The duration of the constant frequency portion thus varied from 982 ms for a .49 Hz modulation rate to 22 ms for 7.8 Hz.

Recording

Electrodes were located at Cz and over left and right temporal sites (T3 and T4). The reference was the midpoint of a 20 kOhm potentiometer linking electrodes located 2 cm below the right and left mastoids. Eye movements (EOG) were recorded between electrodes located medially above and laterally below the right eye. Recordings were obtained from eight adults (23-42 years old) with normal hearing.

Analysis

The analysis for the Cz channel was the same as for Experiment 1, except that for each experimental condition the sweeps were a constant 2048 ms in duration, 240 rather than 800 sweeps were

recorded, and the on-line T^2 tests were based on 10 subaverages of 24 sweeps. For this experiment the analysis-programs calculated T^2 values for each subject at both f_o and $2f_o$.

The responses from Cz, T3 and T4 and the EOG were also averaged simultaneously using a sweep of 1900 ms.

At modulation rates of .49, .98 and 1.95 Hz the ramp stimulus elicited transient P1-N1-P2 waveforms. The latencies of the peaks were measured relative to the onset of the ramp. The amplitudes were measured relative to a 50 ms baseline preceding the stimulus: P1 as the maximum positive peak between 40 and 100 ms, N1 as the maximum negative peak between 60 and 160 ms, and P2 as the maximum positive peak between 100 and 250 ms.

Statistical Analysis

The amplitudes and phase-delays for the f_o and $2f_o$ responses were separately analyzed with 2-way stimulus type (sinusoidal or ramp FM) x modulation-rate repeated measures ANOVAs. Linear regressions of phase-delay on response rate provided the apparent latencies. The amplitudes and latencies for the P1, N1, P2 components were separately analyzed using 2-way modulation-rate x ramp-direction repeated measures ANOVAs.

Results

Steady State Responses

Reliability. Table 2 shows that most individual subjects did not show reliable steady state responses at f_o or $2f_o$ to

sinusoidal FM stimuli. Although there were only a few reliable f_o responses to the ramp stimuli, reliable responses at $2f_o$ occurred at all modulation rates. Grand mean waveforms are shown in Figure 2.

Insert Figure 2 and Table 2 about here

Amplitude. The responses at f_o and $2f_o$ were separately analyzed. The f_o responses to both sinusoidal and ramp stimuli decreased in amplitude as modulation rate increased ($F=8.52$; $df=4,28$; $p<0.01$). Figure 3 shows the vector-averaged amplitudes of the $2f_o$ responses. For $2f_o$ responses there was a significant stimulus-type x modulation-rate interaction ($F=6.98$; $df=4,28$; $p<0.01$). The amplitude of $2f_o$ responses to both the ramp and sinusoidal stimuli decreased as rate increased but the ramp stimuli evoked larger responses than the sinusoidal stimuli at all rates except for 7.8 Hz.

Apparent Latency. Vector-averaged phase-delays are shown in Figure 3. Apparent latencies were calculated over the .98 to 7.8 Hz range. The $2f_o$ response had an apparent latency of 59 ms and 41 ms for the sinusoidal and ramp FM respectively.

Insert Figure 3 about here

Transient Responses

Figure 2 shows the average P1-N1-P2 waveforms evoked by the stimulus ramps at modulation rates of .49, .98 and 1.95 Hz.

Latency. Table 3 shows the average latencies of the P1, N1, and P2 responses to upward and downward ramps presented at different

stimulus modulation rates. The increase in P1 latency with increases in modulation rate did not quite reach significance ($F=4.36$; $df=2,14$; $p=.06$). N1 latency was not affected by changes in rate. P2 latency decreased as rate increased ($F=6.66$; $df=2,14$; $p<0.05$). Ramp direction did not affect the latency of any of the components.

Amplitude. The average amplitudes for the components appear in Table 3. P1 and P2 amplitude were not affected by the direction of the ramp, whereas as Figure 2 shows, the N1 component was larger for upward than for downward ramps ($F=34.70$; $df=1,7$; $p<0.01$). P1 ($F=4.93$; $df=2,14$; $p<0.05$), N1 ($F=42.18$; $df=2,14$; $p<0.01$), and P2 ($F=9.04$; $df=2,14$; $p<0.01$) all decreased in amplitude as modulation rate increased.

Insert Table 3 about here

Scalp Distribution

Steady state and transient responses evident at Cz could generally also be observed in waveforms recorded from T3 and T4, but the waveforms were smaller and the recordings noisier than at Cz. Figure 4 shows the waveforms recorded at T3 and T4 for the .49 Hz ramp stimulus. The N1 latency of 120 ms at T3 and T4 was similar to the mean N1 latency of 116 ms at Cz. The average N1 amplitude at T3 and T4 was 25% of the N1 average N1 amplitude at Cz. The morphology of the response was different at T3 and T4 but this could not be evaluated because of the small size of the responses.

Insert Figure 4 about here

Discussion

As in the first experiment, the sinusoidal stimulus did not elicit reliable steady state responses. Even the most reliable individual response at the $2f_o$ of the 3.9 Hz rate occurred in only half the subjects. There was some consistency across subjects in the $2f_o$ response elicited by sinusoidal FM in the .98 to 3.9 Hz region, which are evident in the average waveforms (Figure 2), and there was a regular increase in phase-delay over rate, suggesting that small responses to the sinusoidal stimulus are present but buried in EEG noise. Regan (1988a) has recently pointed out that significant improvements in signal-to-noise can derive from increasing the duration of the sweep analyzed by the Fourier transform. The frequency resolution of the transform is 1/(Nt) where N is the number of points in the sweep and t is the time per point. In our recording this was 0.49 Hz. Thus any signal at 3.9 Hz would be combined with noise in adjacent frequencies such as 3.8 and 4.0 Hz. If we had used a longer sweep we could have reduced the range of frequencies contained within each measurement and thus have significantly reduced the noise levels.

The ramp FM stimuli evoked larger and more consistent steady state responses than sinusoidal FM. The ramp stimulus elicited reliable $2f_o$ responses at all rates in the .49 to 7.8 Hz region, but did not elicit f_o responses. The $2f_o$ responses to the ramp stimulus were about twice as large as responses to the sinusoidal stimulus, at all modulation-rates except 7.8 Hz.

Although the sinusoidal stimulus did not elicit transient responses, the ramp stimulus elicited clear transient EPs at modulation rates of .5, .98 and 1.95 Hz. The rate-effects suggest that the P1-N1-P2 waves are perhaps composed of a negative wave riding on a slow positive wave. The N1 is more sensitive to rate and at rapid rates the underlying positive wave becomes more evident. This two-component view of the response would explain why at rapid rates the P1-latency became longer and the P2-latency (unexpectedly) shorter.

Clynes (1969) indicated that frequency must be stable for at least 300 ms before a change would elicit a transient response. This fits with our data showing no clear transient responses when the stable frequency lasted less than 214 ms (at the 1.95 Hz rate).

We found that the N1 amplitude was larger for upward than for downward ramps. This agrees with previous findings by Ruhm (1971) and Kohn et al. (1978), but differs from Arlinger et al. (1976) and Arlinger and Jerlvall (1979). Arlinger and his colleagues measured N1-P2 peak-to-peak whereas we measured P1, N1 and P2 from a pre-stimulus baseline. Since P1 and P2 were not affected by ramp direction, the peak-to-peak measurements may have hidden a significant N1 sensitivity to ramp direction. In addition, the stimuli in the Arlinger studies differed from ours by being followed by a slow (600 ms) return to baseline frequency and by having an interstimulus interval of 2 to 6 s.

The N1 wave might reflect the contribution of FM neurons in the auditory system that are specifically activated by changes in frequency (Whitfield and Evans, 1965; Erulkar et al., 1968; Phillips

et al., 1985). Our experiments suggest that there may be more neurons sensitive to upward than to downward frequency-changes. One possible explanation is that there is a greater synchronization of afferent fibres for a frequency change from low to high frequency than for a change from high to low frequency because of basilar membrane mechanics (Collins and Cullen, 1978). However, since electrocochleographic findings suggest greater sensitivity for downward frequency ramps (Eggermont, 1976), it is possible that our findings are specific to the cortex.

EXPERIMENT 3

Our third experiment evaluated responses to simultaneous ramp frequency changes. The purpose of the experiment was to see how responses to combined ramps differed from responses to single ramps. The rationale for the experiment is that stop-consonants are characterized by simultaneous changes in different frequency-bands.

Methods

Stimulus

The method of stimulus presentation and stimulus intensity were the same as in the previous experiments. A "high" FM tone alternated between 2000 and 1000 Hz. A "low" FM tone alternated between 250 and 500 Hz. The linear ramp between the two alternating frequencies was constant at 41 ms (which differed from the 42 ms of Experiment 2 because of the particular computer program). Two frequency ramps

occurred within one cycle of modulation. Modulation rates were .98, 1.95 and 3.9 Hz. The constant frequency portion varied from 471 ms for the .98 Hz modulation rate to 87 ms for the 3.9 Hz modulation rate.

Three stimulus conditions were presented at each modulation rate. In one condition only the high tone was presented; in another condition only the low tone was presented; in the combined stimulus condition, the high and low tone were presented simultaneously and the linear ramps increased and decreased in phase. Since the tone frequencies were harmonically related, no obtrusive beats were perceptable in the combined stimulus. The nine trials were presented to the subject in randomized order.

Recording

Recordings were obtained from Cz using a reference at M2. Subjects were eight adults (23-42 years old) with normal hearing.

Analysis

The analysis was identical to that of Experiment 2, except that for each experimental condition the sweeps were 1024 ms in duration, 500 sweeps were recorded and on-line T^2 tests were based on 10 subaverages of 50 sweeps. The P1-N1-P2 components were scored by the same criteria as in Experiment 2. Since we found that the f_o steady state response affected baseline-to-peak measurements, we also analyzed N1-P2 peak-to-peak measurements. The evaluation of the transient response was limited to .98 Hz because the responses at faster rates were too small to measure.

Statistical Analysis

The steady state amplitude and phase-delay data were analyzed with a 3 x 3 stimulus presentation x modulation rate ANOVA. N1-P2 amplitude and P1, N1 and P2 amplitudes and latencies were analyzed with ANOVAs comparing the three stimulus conditions and the two ramp directions.

Results

Steady State Responses

Amplitude. There was a main effect of stimulus condition on $2f_o$ amplitude ($F=17.59$; $df=2,14$; $p<0.01$). Newman-Keuls comparisons showed that the amplitude in the combined stimulus presentation condition (1.78 uV) was significantly greater than the amplitude evoked by the high (1.40 uV) or the low (1.32 uV) stimulus. Amplitude decreased as rate increased ($F=6.20$; $df=2,14$; $p<0.01$).

Apparent Latency. The apparent latencies of the $2f_o$ responses were 65, 67, and 69 ms respectively for the low, high, and combined stimulus conditions.

Transient Responses

Latency. Latencies for N1 (121 ms) and P2 (173 ms) were not affected by modulation rate or stimulus presentation. A stimulus presentation effect on P1 latency ($F=4.29$; $df=2,14$, $p<0.05$), followed by Newman-Keuls, showed that P1 latency was longer for the low stimulus than for the high or combined conditions.

Amplitude. P1: A stimulus-presentation x ramp-direction

interaction (F=7.05; df=2,14; p<0.01) analyzed for simple main effects showed that for the low stimulus the downward ramp elicited a larger P1 (1.9 uV) than the upward ramp (1.0 uV).

N1: Mean N1 amplitude was -1.2 (± 1.1) uV for upward ramps and 0.1 (± 1.3) uV for downward ramps. The ANOVA yielded a stimulus-presentation x ramp-direction interaction (F=4.74; df=2,14; p<0.05). As shown in Figure 5, the N1 was larger in response to upward than to downward ramps in all conditions. The N1 to the low tone downward ramp was smaller than the N1 evoked by either the high or combined in-phase stimulus downward ramps. These amplitudes were very variable. This variability may have been partly caused by a steady state response at f_o which caused the baseline before the downward ramp to be more negative than the baseline for the upward ramp, particularly for the low tone. We therefore examined the N1-P2 amplitude as well. There was a stimulus presentation effect on N1-P2 amplitude (F=7.92; df=2,14; p<0.01). Newman-Keuls showed that N1-P2 amplitude for the combined stimulus condition (2.9 uV) was significantly larger than for the low tone (1.6 uV), but not for the high tone (2.3 uV).

P2: There were no significant effects of stimulus presentation or ramp direction on P2 amplitude.

Insert Figure 5 about here

Discussion

The ramp FM stimuli in Experiment 3 evoked reliable steady state

responses at the $2f_o$ of the modulation rate. Response amplitude was larger for the combined stimulus condition than for the single stimuli. This indicates that the response varies with the amount of change and is not just a nonspecific indication of change. The response in the combined condition is smaller than the sum of the steady state responses evoked individually by the high and low stimuli indicating that there was some interaction in the processing of the high and low FM tones when they were presented simultaneously. Such interaction could be caused by an inhibition between different response-regions or by an overlap in the receptive fields of the responding cells.

As in Experiment 2, the N1 amplitude was larger to upward than to downward ramps. The morphology of the response suggests again that the N1 wave rides on (and splits in two) a larger positive wave. The N1 is clearly affected by the direction of the ramp whereas neither P1 nor P2 are affected. Like the steady state responses the N1 showed interactions in the processing of simultaneously presented ramps. The changes in baseline may have been caused by changes in the loudness of the tones - at the intensities we used, a tone of 500 Hz is louder than one of 250 Hz, and would have evoked a larger sustained potential.

CONCLUSIONS

Sinusoidal FM stimuli presented at slow modulation rates did not elicit consistent, reliable steady state responses. It is possible

that steady state responses were present but were obscured by EEG noise, in which case a significant increase in analysis-time would be required to obtain reliable responses in individual subjects.

In agreement with Clynes' (1969) formulation of the Rest-Motion response, a stimulus with a constant frequency portion alternating with frequency-changes proved more effective at eliciting EPs than the sinusoidal FM. This is true for both transient and steady state responses.

The steady state responses were most easily recognized at the second harmonic of the modulation rate. The apparent latency of this $2f_m$ response evoked by either sinusoidal or ramp FM was between 40 and 70 ms. These latencies suggest an origin in the cortex.

Transient P1-N1-P2 responses were evoked by ramp stimuli at modulation rates of less than 2/s (4 ramps/s). Upward ramps evoked a larger N1 than downward ramps. Two simultaneous ramps evoked an N1 that was larger than the N1 evoked by either ramp alone but smaller than the sum of these individual N1s. The effects of stimulus rate suggest that P1 and P2 may be part of a larger positive wave, upon which the N1 is superimposed.

These responses may prove helpful in the objective evaluation of frequency-discrimination. The optimal stimulus is a ramp-change in frequency, which can evoke both transient and steady state responses.

REFERENCES

Anderson TW. Introduction to Multivariate Statistical Analysis. New York: Wiley, 1958:107-108.

Arlinger SD, and Jerlvall, LB. Results of psychoacoustic and cortical evoked potential experiments using frequency and amplitude modulated stimuli. Scan Audiol Suppl 1979;9:229-239.

Arlinger SD, Jerlvall LB, Ahrén, T, and Holmgren, EC. Slow evoked cortical responses to linear frequency ramps of a continuous pure tone. Acta Physiol Scand 1976;98:412-424.

Clynes, M. Dynamics of vertex evoked potentials: the R-M brain function. In Donchin E, and Lindslay DB, Eds. Averaged Evoked Potentials, Washington, D.C.: NASA SP-191, 1969:363-374.

Collins MJ, and Cullen JK. Temporal integration of tone glides. J Acoust Soc Am 1978;63:469-473.

Eggermont, JJ. Electrocochleography. In Keidel WD and Neff WD, Eds. Handbook of Sensory Physiology, Vol. 5/3. Berlin: Springer-Verlag, 1976: 625-705.

Erulkar SD, Butler RA, and Gerstein GL. Excitation and inhibition in cochlear nucleus. II. Frequency-modulated tones. J Neurophysiol

1968;31:537-548.

Gardner RB, and Wilson JP. Evidence for direction-specific channels in the processing of frequency modulation. J Acoust Soc Am 1979:66:704-709.

Hoops RA. Speech Science. Springfield: Charles C Thomas, 1972:67-72.

Kay RH, and Matthews DR. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. J Physiol 1972;225: 657-677.

Keppel G. Design and Analysis. Englewood Cliffs, NJ: Prentice-Hall, 1982:470-471.

Kohn M, Lifshitz K, and Litchfield D. Averaged evoked potentials and frequency modulation. Electroencephal Clin Neurophysiol 1978; 45:236-243.

Lenhardt ML. Effects of frequency modulation on auditory averaged evoked response. Audiology 1971;10:18-22.

Møller AR. Unit responses in the cochlear nucleus of the rat to sweep tones. Acta Physiol Scand 1969;76:503-512.

Møller AR. Unit responses in the rat cochlear nucleus to tones of rapidly varying frequency and amplitude. Acta Physiol Scand

1971;81:540-556.

Møller AR. Coding of amplitude and frequency modulated sounds in the cochlear nucleus of the rat. Acta Physiol Scand 1972;86:223-238.

Møller AR. Coding of sounds with rapidly varying spectrum in the cochlear nucleus. J Acoust Soc Am 1974:55:631-640.

Phillips DP, Mendelson, JR, Cynader, MS, and Douglas, RM. Responses of single neurones in cat auditory cortex to time-varying stimuli: frequency-modulated tones of narrow excursion. Exp Brain Res 1985;58:443-454.

Picton TW. Human auditory steady state responses. In Barber c, and Blum T, Eds. Evoked Potentials III. Boston: Butterworths, 1987:117-124.

Picton TW, Skinner CR, Champagne SC, Kellett AJ, and Maiste AC. Potentials evoked by the sinusoidal modulation of the amplitude or frequency of a tone. J Acoust Soc Amer 1987;82:165-178.

Rees A, and Møller AR. Responses of neurons in the inferior colliculus of the rat to AM and FM tones. Hear Res 1983;10: 301-330.

Regan D. Evoked Potentials in Psychology, Sensory Physiology and Clinical Medicine. London: Chapman & Hall, 1972:p. 77.

Regan D. Fourier analysis of evoked potentials; some methods based on Fourier analysis. In Desmedt JE, Ed. Visual Evoked Potentials in Man: New Developments. Oxford: Clarendon Press, 1977:110-117.

Regan D. Human Brain Electrophysiology. Amsterdam: Elsevier, 1988(a).

Regan D. Human visual evoked potentials. In Picton TW, Ed. Human Event-Related Potentials. Amsterdam: Elsevier, 1988(b):159-164.

Regan D, and Tansley BW. Selective adaptation to frequency-modulated tones: Evidence for an information-processing channel selectively sensitive to frequency changes. J Acoust Soc Am 1979;65:1249-1257.

Rodriguez R, Picton T, Linden D, Hamel G, and Laframboise G. Human auditory steady state responses: Effects of intensity and frequency. Ear Hear 1986;7: 300-313.

Ruhm HB. Rate of frequency change and the acoustically evoked response. J Aud Res 1970;10:29-34.

Ruhm HB. Directional sensitivity and laterality of electroencephalic responses evoked by acoustic sweep frequencies. J Aud Res 1971; 11:9-16.

Sinex DG, and Geisler CD. Auditory-nerve fiber responses to frequency-modulated tones. Hear Res 1981;4:127-148.

Spoor A, Timmer F, and Odenthal DW. The evoked auditory response (ear) to intensity modulated and frequency modulated tones and tone bursts. Int Audiology 1969;8;410-415.

Stapells DR, Linden D, Suffield JB, Hamel G, and Picton TW. Human auditory steady state potentials. Ear Hear 1984;5:105-113.

Tansley BW, and Regan D. Separate auditory channels for unidirectional frequency modulation and unidirectional amplitude modulation. Sens Proc 1979;3:132-140.

Tansley BW, and Suffield JB. Time course of adaptation and recovery of channels selectively sensitive to frequency and amplitude modulation. J Acoust Soc Am 1983;74:765-775.

Whitfield IC, and Evans EF. Responses of auditory cortical neurons to stimuli of changing frequency. J Neurophysiol 1965;28:655-672.

TABLE 1. Reliability of steady state responses to the 70 dB HL sinusoidal FM stimulus.

MODULATION	FUNDAMENTAL FREQUENCY		SECOND HARMONIC	
RATE	N**	GROUP T^2**	GROUP T^2**	
3.1	1	5.1	38.6	*
3.9	6	26.8	23.0	*
5.2	4	12.6	3.2	
6.5	2	12.8	15.9	*
7.8	1	9.1	22.7	*

** N = the number of significant individual T^2s for the eight subjects.

* Group T^2 is significant at $p<.05$ if greater than 12.0.

TABLE 2. T^2 values for ramp and sinusoidal frequency modulation

MODULATION RATE	FUNDAMENTAL FREQUENCY		SECOND HARMONIC	
SINUSOIDAL	N**	GROUP T^2	N**	GROUP T^2
0.49	1	9.4	2	8.1
0.98	1	0.6	2	41.5 *
1.95	1	2.4	3	19.3 *
3.9	3	14.0 *	4	12.2 *
7.8	3	74.1 *	3	9.3
RAMP				
0.49	0	2.2	7	1.6
0.98	1	5.4	8	34.9 *
1.95	1	7.6	6	14.2 *
3.9	1	9.5	8	111.3 *
7.8	3	8.5	4	20.9 *

** N = the number of significant individual T^2s for the eight subjects.

* Group T^2 is significant at $p<.05$ if greater than 12.0.

Table 3. Transient responses to upward and downward frequency-ramps.

LATENCY (ms)

	P1		N1		P2	
RATE(Hz)	UP	DOWN	UP	DOWN	UP	DOWN
.49	67	65	115	116	193	200
.98	70	71	117	113	177	181
1.95	78	73	115	108	161	156

AMPLITUDE (uV)

	P1		N1		P2	
RATE(Hz)	UP	DOWN	UP	DOWN	UP	DOWN
.49	1.7	1.0	- 4.5	- 3.5	4.7	3.4
.98	2.0	2.3	- 0.7	0.5	3.2	2.6
1.95	0.9	1.1	- 0.1	0.4	1.1	1.1

Figure 1. Vector-averaged amplitude and phase-delay of first and second harmonic steady state responses plotted over response rate. The stimulus was a 70 dB HL 1000 Hz tone sinusoidally modulated between 750 and 1250 Hz. The responses were recorded from Cz referenced to the right mastoid. The vector-averaged amplitude was larger between 3.9 and 6.5 Hz than at the other rates. Phase-delay increased as response rate increased.

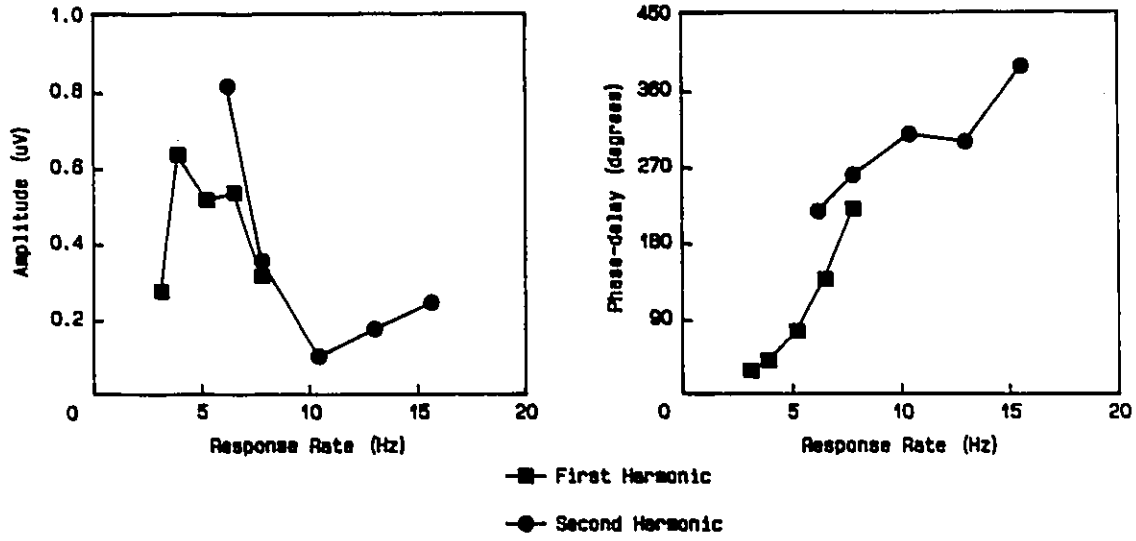

1.0
0.8
0.6
0.4
0.2
0
Amplitude (uV)
5
10
15
20
Response Rate (Hz)
450
360
270
180
90
0
Phase-delay (degrees)
5
10
15
20
Response Rate (Hz)
First Harmonic
Second Harmonic

FIGURE 1

Figure 2. Responses to sinusoidal and ramp FM stimuli. Recordings were between Cz and linked mastoids with negativity at Cz plotted upward. The left side of the figure shows steady state responses to sinusoidal FM. Below each response is shown the modulating frequency at each modulation rate. The stimuli were 70 db HL tones sinusoidally modulated between 750 and 1250 Hz at the rates shown. Responses are most apparent at the second harmonic. The right side of the figure shows steady state responses to FM ramp stimuli. The stimulus was a 70 dB HL tone alternating in frequency between 750 and 1250 Hz with a 42 ms ramp. The top three waveforms show the P1-N1-P2 superimposed on the steady state response. Upward ramps evoked a larger N1 (open triangles) than downward ramps (solid triangles). Ramp FM stimuli evoked a steady state response at the second harmonic. Two replicate averages of 960 trials (from 8 subjects) are superimposed.

FIGURE 2

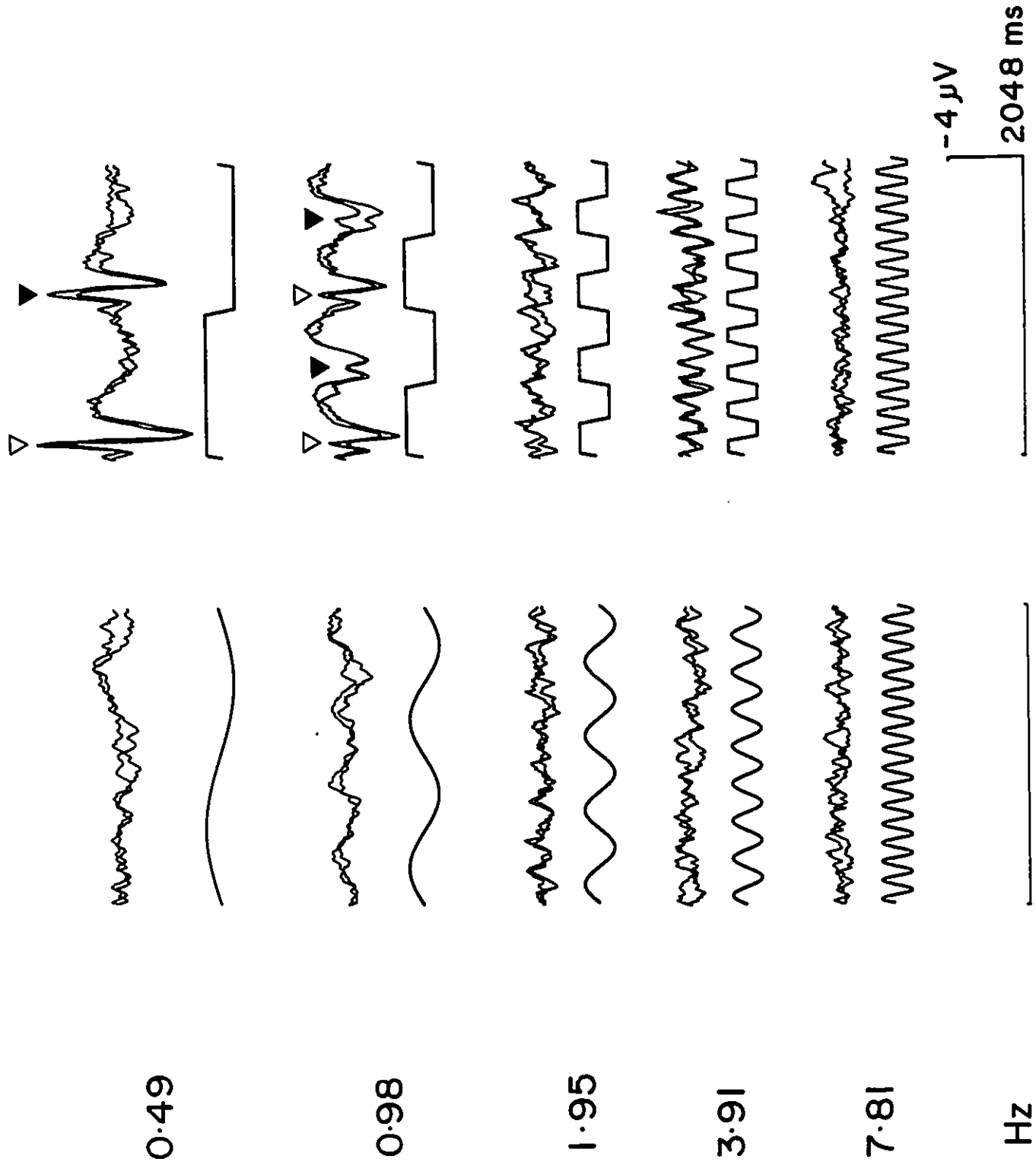


Figure 3. Vector-averaged phase-delay and amplitude of second harmonic responses. The stimuli are described in the legend for figure 2. Ramp FM stimuli generally evoked larger responses than the sinusoidal FM stimuli. Phase-delay increased over response rate. Recordings were from Cz referenced to linked mastoids. Based on data from 8 subjects.

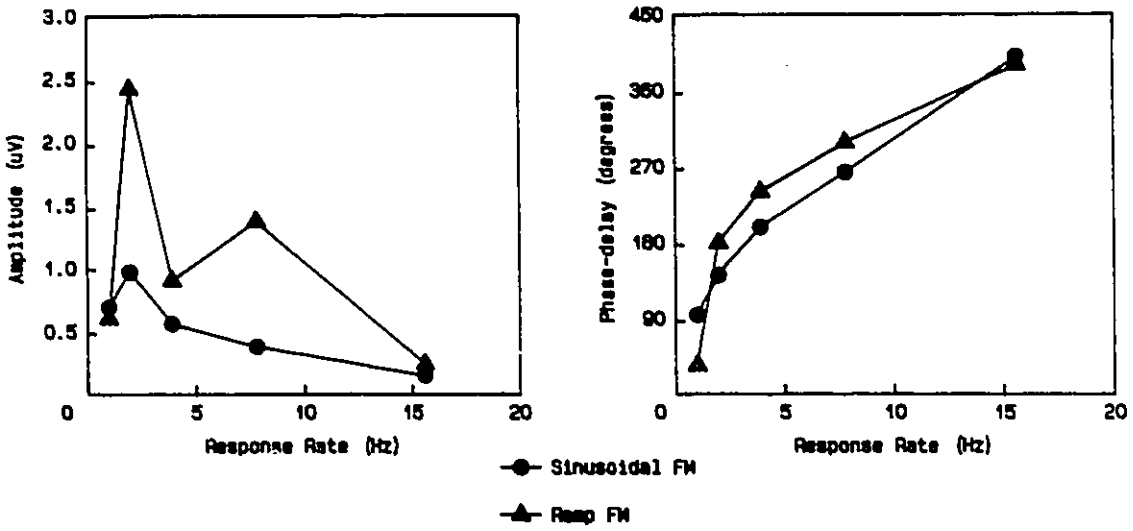

3.0
2.5
2.0
1.5
1.0
0.5
0
Amplitude (uV)
5
10
15
20
Response Rate (Hz)
450
360
270
180
90
Phase-delay (degrees)
0
5
10
15
20
Response Rate (Hz)
Sinusoidal FM
Ramp FM

FIGURE 3

Figure 4. Transient responses to ramp changes in frequency. The stimulus was a 70 dB HL tone alternating between 750 and 1250 Hz at a rate of .49 Hz. A linked mastoid reference was used and negativity at the scalp was plotted upward. Responses at T3 and T4 were smaller than at Cz. The open triangles indicate the larger N1 evoked by upward ramps than by downward ramps (solid triangles). Waveforms are averages of 1920 trials from 8 subjects.

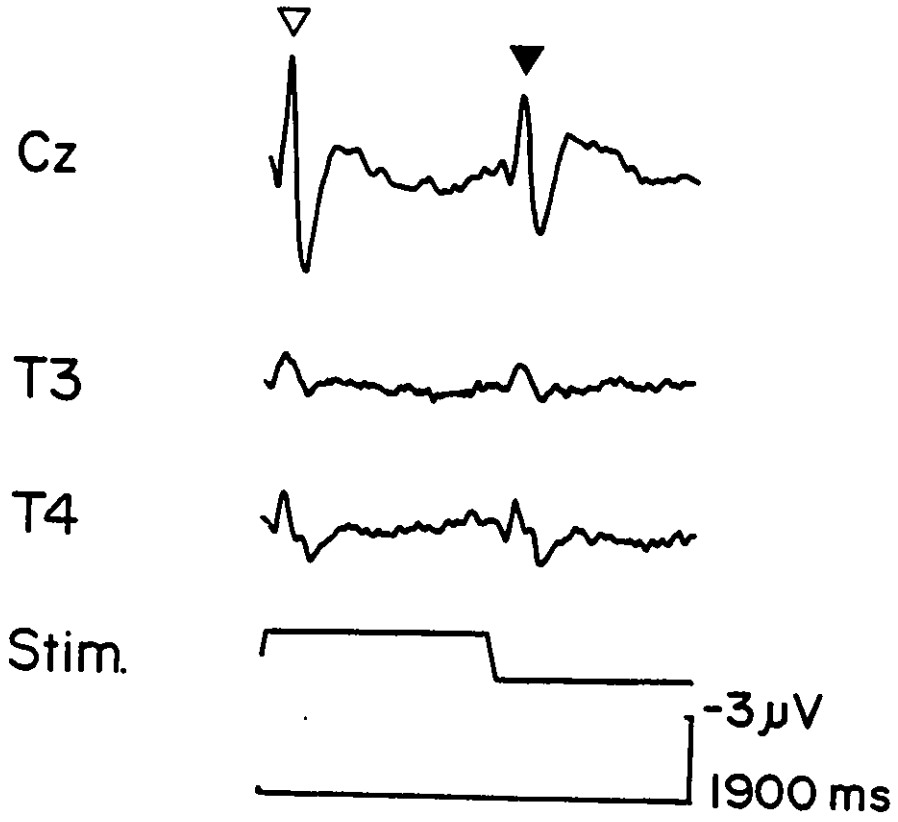


FIGURE 4

Figure 5. Transient responses to combined frequency-changes. This figure shows the average transient responses evoked by the high, low, and combined FM ramp stimuli at a modulation rate of .98 Hz. The high stimulus alternated between 1000 and 2000 Hz, and the low stimulus alternated between 250 and 500 Hz. In the combined condition, the two stimuli occurred simultaneously. Recordings were from Cz referenced to M2 with negativity at Cz shown as an upward deflection. The N1 was larger to upward ramps (open triangles) than to downward ramps (solid triangles). N1 appears superimposed on a larger positive wave. This is most evident following the downward ramp for the low tone. Waveforms are an average of 4000 trials from 8 subjects.

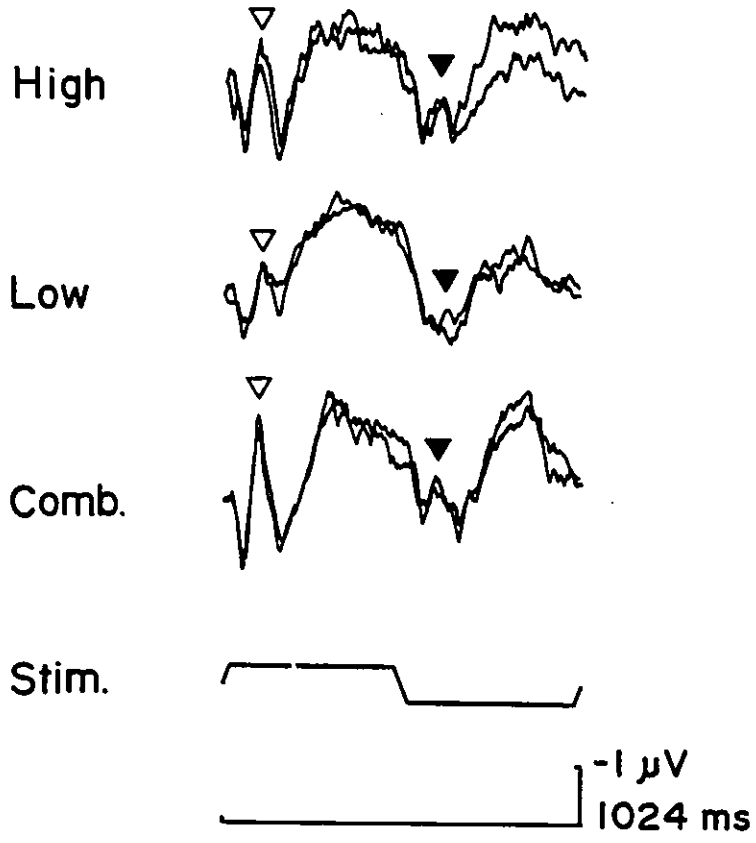


FIGURE 5

SECTION 3: EVENT-RELATED POTENTIALS AND CATEGORICAL PERCEPTION OF SPEECH SOUNDS

ABSTRACT

Computer-modified speech sounds from the /ba/ to /da/ continuum were presented to reading subjects as a train of standard speech sounds interspersed with two types of infrequent deviant speech sounds. One deviant stimulus lay within the same category as the standard and the other lay across the categorical boundary from the standard, but both were acoustically equidistant from the standard in terms of the second formant transition. When the standard stimulus was drawn from the /ba/ end of the continuum, the across-category deviants elicited a clear mismatch negativity (MMN) in the auditory event-related potential whereas the within-category deviants did not. This MMN began at about 60-120 ms after stimulus onset and was present for several hundred ms. These results suggest that categorical processing of speech sounds occurs independently of attention at an early echoic memory stage. When the standard stimulus was drawn from the /da/ end of the continuum, the MMN to across-category deviants was not larger than the MMN to within-category deviants. Grand mean waveforms suggested that both deviant stimuli elicited small MMNs. Although this may indicate processing along an acoustic continuum, the results of the psychophysical tests suggest that the standard stimulus in this condition was too close to the category boundary to evoke a consistent categorical mismatch.

INTRODUCTION

Speech Production

Speech sounds result from the manipulation of air passing through the vocal tract from the lungs. The vocal tract includes all of the air passages above the larnyx from the glottis to the lips, including the mouth, pharynx, and nasal cavity. The vocal tract is essentially a tube with a complicated shape that acts as a resonator, that is, it emphasizes certain frequencies more than others (Borden & Harris, 1984). The vocal tract emphasizes the harmonics of the vocal cord wave at several different frequencies, and the spectrum of the speech wave will have a peak for each of the vocal tract's natural frequencies. Resonances of the vocal tract are called formants and their frequencies are the formant frequencies. The shape of the vocal tract is varied by the movements of the vocal organs, the tongue, pharynx, jaw, lips and soft palate. Each configuration of the vocal tract involved in producing vowel sounds has its own set of characteristic formant frequencies. The lowest formant frequency is called the first formant, and the one with the next highest frequency is called the second formant. A spectrogram (Figure 1) shows the formants during speech production as broad bands of acoustic energy over time.

Insert Figure 1 about here

During the production of stop consonants such as /b/, /d/, or /g/ the oral cavity is first closed at some point in the vocal tract. As the closure is opened, the sound wave passing through the opening changes in its frequency content from the characteristic formant-frequencies of the closure to those of the following vowel. This results in a rapid rising or falling frequency transition at the beginning of the syllable. These changes are clearly visible in the first 50 ms of the spectrogram (Figure 1).

Categorical Perception

Listeners normally perceive the continuum of frequency transitions in speech sounds in a categorical fashion. One acoustic continuum involving frequency transitions is the place-of-articulation continuum. A change in the place of articulation from bilabial (/b/) to alveolar (/d/) to velar (/g/) corresponds to a change in the starting frequency of the second formant, and thus the angle of the second formant transition. As the sound spectrogram in Figure 1 shows, the production of the stop consonant /d/ begins with a rapid rise of the first formant and fall of the second formant to the relatively constant formant-frequencies of the following vowel /a/. The second formant transition is important for how speech sounds are perceived (Liberman, Harris, Hoffman & Griffith, 1957; Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967). The sounds /b/, /d/ and /g/ are known as

phonemes. A phoneme is a family of sounds which functions in a language to signal a difference in meaning (Borden & Harris, 1984). For example, the fact that "bet" and "get" differ in meaning demonstrates that /b/ and /g/ are phonemes in English. Therefore the differences in the second formant transition between these speech sounds is critical for determining differences in meaning.

The acoustic continuum of the second formant transition in stop consonants (/ba/, /da/, /ga/) is categorically perceived. With the graded changes in the second formant transition shown in Figure 2 subjects perceive only two different speech sounds, /ba/ or /da/. This is an example of the phenomenon called Categorical Perception (CP). CP is "a qualitative difference in how similar things look or sound depending on whether or not they are in the same category" (Harnad, 1987, p.2). Another example of CP is the division of the color spectrum into color categories. In both cases, equal physical differences between stimuli are perceived as smaller or larger depending on whether the stimuli are in the same or different categories.

Insert Figure 2 about here

The primary approach to investigating CP has been psychophysics, the study of the relationship between physical stimulation and sensation. Subjects are presented with stimuli that vary along a physical continuum, such as the acoustic continuum of the second

formant transition, and CP is demonstrated by discrimination (telling stimuli apart), and identification (labeling stimuli). The following two criteria are most commonly used to define CP: "(1) a set of stimuli varying along a physical continuum is given one label on one side of a category 'boundary' and another label on the other side and (2) the subject can discriminate smaller physical differences between pairs of stimuli that straddle that boundary than between pairs that are within one category or the other" (Harnad, 1987, p.3). Thus, listeners may be able to discriminate members of the same category, but with more difficulty and less accurately than members of different categories (Massaro, 1987). The identification function in Figure 3 shows how subjects labeled the stimuli along a /ba/ to /da/ continuum. There is a rapid cross-over in this function at the point at which subjects switched from perceiving stimuli as "ba" to "da". This crossover between the two stimulus categories is called the phoneme boundary.

Insert Figure 3 about here

Categorical perception of speech is automatic in that it is performed unconsciously during normal speech perception. However, exposure, practice and selective attention may influence the location of the CP boundary effect as well as the acuity of discrimination and the accuracy of identification (Pastore, 1987). An individual's native language also influences the location of phonemic boundaries.

Voice-onset time is the relationship between the time of vocal cord vibration and consonant release (e.g. the separation of the lips to release a burst of air). The voice-onset time continuum is also categorically perceived (Liberman et al., 1967). There are differences in the boundaries along the voice-onset time continuum between English, French and Thai speakers (Repp & Liberman, 1987). Japanese adults, for whom the /ra-la/ distinction is not phonemic do not show categorical perception for /ra-la/ stimuli (Miyawaki, Strange, Verbrugge, Liberman & Jenkins, 1975). There are however still a limited number of preferred boundary locations. It is not totally arbitrary. Mastery of a new language also implies establishment of new phonemic categories. However lengthy experiments are necessary to show this.

Non-verbal infants have shown the same phonemic boundaries as adults (Eimas, Miller & Jusczyk, 1987). The evidence suggests that infants possess universal potential phonemic boundaries. For example, Werker and Tees (1984) found that prelinguistic American infants were capable of distinguishing phonemic categories foreign to English, but lost that ability around 10 months of age.

Theories of Categorical Perception

One useful distinction to make when discussing categorical perception is that of an auditory level of processing and that of a phonetic level (Kuhl, 1987). The auditory level can be defined as the set of processes common to speech and non-speech, whereas the

phonetic level is assumed to include processing that is unique to speech.

The first influential theory accounting for speech CP was the motor theory of speech perception, which holds that speech perception is different from other auditory perceptions, and that we perceive speech sounds by unconsciously inferring how they would be pronounced (Liberman et al., 1957, Studdert-Kennedy, Liberman, Harris & Cooper, 1970). Within this model, the CP boundaries between speech sounds /ba/ and /da/ arise from the motor discontinuities required to pronounce them. Although there is a continuum of the second formant transition between /ba/ and /da/ there is a discontinuity between the tongue and lip movements required to produce them. Motor theory can account well for several instances of CP (see review by Repp & Liberman, 1987). However, evidence that pre-verbal infants and non-human animals exhibit the same CP boundaries as adult language speakers suggests that some speech perception functions must be innate rather than learned. An animal species with an auditory system similar to man's can provide a good model of man's auditory-processing system in the absence of any higher-order (phonetic) processing (Kuhl, 1987). A number of animal experiments (reviewed by Kuhl, 1987) have shown that the "auditory level" was sufficient to exhibit CP;higher-order processing was not necessary.

The theory of innate sensitivity suggests that due to natural auditory system sensitivities to certain portions of acoustic continua, some regions of the continua are more easily discriminable

than others (Eimas & Corbit, 1973; Stevens, 1981). Sounds drawn from regions with low discriminability will sound more alike than sounds drawn from regions of high discriminability. Rosen and Howell (1987) have reviewed three lines of evidence that innate auditory sensitivities underly CP. First, CP has been reported for non-speech auditory continuua. Second, non-human animals also exhibit CP of some speech continuua. Third, CP can occur in preverbal infants.

Support for natural sensitivities also comes from studies showing that when listeners hear a high repetition of a stimulus from one end of the continuum, /ba/ for example, and are then presented with the random presentation of stimuli along the /ba/ to /da/ continuum, the phoneme boundary shifts towards the /ba/ end of the continuum (Eimas & Corbit, 1973). Listeners identify more of the stimuli as /da/. Results such as these have been interpreted as evidence of selective adaptation of linguistic feature detectors (Eimas & Corbit, 1973; Darwin, 1976). Phoneme detectors are thus viewed as "hard-wired" into the auditory system. A stimulus that can activate two detectors will be perceived as one or the other depending on how much the stimulus activates each detector and how adapted the detector is.

However, Repp and Liberman (1987) have cautioned that selective adaptation effects do not necessarily mean that adaptation takes place at an acoustic level. There is evidence that adaptation effects involve phonetic-level processing. Thai and English speakers differ in voice-onset time adaptaton effects (Foreit, 1977).

There are also individual differences in adaptation effects between speakers of the same language. In a study by Samuel (1982) subjects first chose a prototypical /ga/ on the /ga/-/ka/ contiuum. Subjects showed most adaptation to those stimuli they selected as the prototypical /ga/ and not necessarily the most adaptation to the stimulus at the endpoint of the continuum. These recent results have led investigators such as Repp and Liberman (1987) to suggest that selective adaptation effects do not take place exclusively at a general auditory level but at an auditory level that is phonetically relevant. That is, as long as the adapting stimuli are speech, the adaptation effects reflect the extent to which the stimuli engage the speech-processing apparatus.

The label-learning approach contends that categories are not pre-determined by motor production or natural sensitivity, but arise arbitrarily from learning which sounds belong together in a particular class. Since labels are learned, this theory can account for CP of non-speech stimuli, for example, the perception of musical pitch. Trained musicians have shown more CP of musical pitch than non-musicians (Locke & Kellar, 1973). Label-learning theory also accounts well for context-related shifts in CP boundaries. Some theories present a dual-process memory model of CP in which two processing modes operate in parallel (Fujisaki & Kawashima, 1969, 1971 cited in Rosen & Howell, 1987; Ades, 1977). One process is a quickly decaying sensory-trace store, similar to echoic memory (Neisser, 1966), that is involved in immediate comparisons between

stimuli. The other process is a longer lasting and more abstract memory for the labeling process. This processing mode compares a stimulus with its overall context and is influenced by the range of stimuli. In Fujisaki and Kawashima's (1969, 1971) model, speech perception results from the interaction of the auditory process, and the phonemic (or labeling) process which is based on experience. In this model, categorical processing does not occur at the auditory processing stage but arises from learned phonemic labels for speech. A similar dual-process model described by Ades (1977) includes the possible involvement of natural auditory sensitivities at the sensory-trace stage. In this model, CP can arise from either an innate sensitivity of the auditory system, or from the use of learned labels at the context-dependent stage. The context-dependent stage is affected by the range of stimuli and the relative frequency of stimulus occurance.

In their review of CP theories, Rosen and Howell (1987) conclude that none of the three theories, motor theory, innate sensitivities, or label-learning, can explain all CP findings, nor are they mutually exclusive. They suggest that a complete model of CP must explain both innate auditory sensitivity and the role of experience in CP.

Evidence that categorical processing of auditory speech stimuli occurs at an early pre-attentive stage would support the idea that the quickly decaying sensory-store of the dual-process model plays a role in CP of speech. What evidence is there that CP occurs automatically, at an early pre-attentive stage of processing, without

the involvement of conscious decision-making or even a meaningful linguistic task? Findings from psychophysical tasks do not shed light on this issue because subjects are actively attending and responding to stimuli. However, human auditory event-related potentials (ERPs) provide an ideal tool to study this issue unobtrusively and to obtain a real-time measure of cognitive processing of speech stimuli. The next section provides some general information on ERPs.

Auditory ERPs

Auditory stimuli commonly elicit an ERP waveform consisting of P1-N1-P2 waves evoked by the onset of the stimulus, followed by a negative Sustained Potential (SP) which often persists for several hundred milliseconds throughout the duration of the stimulus

The N1 wave is triggered by the abrupt onset of a change in some physical characteristic from an immediately preceding stable level. The N1 component is not a unitary event. Näätänen and Picton (1987) have suggested that the N1 wave contains several components, that are distinguished by different electrical and/or magnetic fields and their relationship to experimental manipulations. Component 1 is a fronto-central negativity generated in the auditory cortex of the supratemporal plane. It has a peak latency of about 100 ms and is about 10% larger contralateral to the stimulated ear. Processes underlying Component 1 may play a role in informing us that a stimulus is occurring, without indicating precise content of the

stimulus.

Component 2 of the N1 is part of the bi-phasic T-complex (Wolpaw & Penry, 1975, 1977), termed the N1c by McCallum and Curry (1979, 1980), with a peak latency of about 150 ms, that is maximal at temporal electrode sites and larger over the hemisphere contralateral to the stimulated ear. This component appears to be generated in the auditory association cortex on the lateral aspect of the temporal lobe. It is not clear what processes Component 2 reflects.

Component 3, a non-specific N1 component, is largest at the vertex and has a peak latency of 100 ms. Näätänen and Picton (1987) suggest that this component signifies a widespread transient arousal of the cortex for facilitating sensory and motor responses to stimuli. The arousal response probably contributes to conscious perception of the stimulus.

The SP occurs in response to sustained auditory stimuli and can be recorded in the absence of attention to the stimuli (Picton, Woods & Proulx, 1978a & b). Picton et al. calculated an average latency for the SP of 150 ms for toneburst stimuli, and they indicated that the SP is physiologically distinct from the N1 response. The SP scalp distribution was more anterior than that of the N1, and the SP and N1 bore different relationships to some stimuli. Scherg, Vajsar and Picton (in press) reported that a source on the supratemporal plane contributes to both the N1 and the SP. Although the SP has a similar scalp distribution to the Contingent Negative Variation, the two components have different onset latencies

and the SP can occur without subjects' attention and in sleeping subjects.

Scalp Distribution

The scalp distribution of ERPs to speech stimuli is generally the same as the ERP topography for non-speech stimuli (Hillyard & Woods, 1979; Picton & Stuss, 1984). A number of topographic analyses of the late auditory ERP components (Scherg & Von Cramon, 1985, 1986; Wood & Wolpaw, 1982) suggest that at least two active cortical regions from within each hemisphere contribute to the late auditory ERPs. One is situated within the superior temporal plane and another is on the lateral surface of the superior temporal gyrus. These regions are in contact with three areas of the left temporal lobe (in right-handers) involved in speech perception. These are the primary auditory cortex, the secondary auditory area which covers the posterior half of the superior temporal gyrus and includes part or all of the Wernicke speech area, and an area covering most of the planum temporale (Seldon, 1985). Lovrich, Novick and Vaughan (1988) showed that the left hemisphere ERP topography associated with auditory speech stimuli was consistent with 2 cortical generators, one within the supratemporal plane and the other on the lateral surface of the superior temporal gyrus.

ERP Studies of Categorical Perception

Although there have been many studies of ERPs during speech

perception (as reviewed in Section 1), only a few studies have investigated CP. Dennis Molfese and his coworkers have conducted a series of experiments on the processsing of voice-onset time (VOT) and stop consonants (Molfese, 1987). Voice-onset time is the relationship between the time of vocal cord vibration and consonant release (eg. the separation of the lips to release a burst of air). Liberman et al. (1967) showed that adult listeners discriminate changes in VOT only to the extent that they assign unique labels to the sounds. Listeners identify bilabial stop consonants with VOT values of 0 and 20 ms as /ba/ whereas stimuli with VOT of 40 and 60 ms are identified as /pa/. Therefore, although subjects cannot discriminate the 20 ms difference between VOTs of 0 and 20 ms and between 40 and 60 ms, they do discriminate and give different labels to stimuli with VOTs of 20 and 40 ms, i.e. stimuli from different phoneme categories. Therefore, VOT discrimination is categorical.

Molfese (1978) recorded ERPs from the left and right temporal regions of sixteen right-handed adults during a phoneme discrimination task. Subjects identifed synthesized consonant-vowel (CV) syllables with VOT values of 0 or 20 ms as /ba/ and those with VOT of 40 and 60 ms as /pa/. Molfese collected ERPs from 16 repetitions per stimulus and analyzed this data with a Principal Component Analysis with Varimax rotation. Four factors were identified. Right (R) hemisphere effects corresponded with behavioural performance: stimuli with VOT of 0 and 20 ms elicited a different ERP waveform at the R hemisphere site than the 40 and 60 ms stimuli. In contrast, left(L) hemisphere

ERP results did not show categorical processing: components of the L hemisphere ERP differed between 0 and 60 ms and differentiated the 0 and 60 ms VOT from the 20 and 40 ms VOTs. Although these results are difficult to interpret, they suggest that both hemispheres were involved in the processing of the speech sounds, although in different and complex ways.

In another study Molfese (1980a) analyzed ERPs to tone-onset-time (TOT) stimuli in which the two-tone stimuli differed from each other in the onset of the lower tone after the higher tone. The second tone began 0, 20, 40 or 60 ms after the first tone, but both tones ended simultaneously. Data were recorded from sixteen subjects at electrode placements T3, T4, C3, C4, T5, T6, P3, P4, referenced to the linked ears. The data were analyzed with the same PCA method as in the 1978 study. Results showed that nine factors accounted for 81% of the variance in the ERP responses. Once more, the results are not easy to interpret. A portion of the response in the R hemisphere at 330 ms discriminated the 0 and 20 ms TOT from 40 and 60 ms TOT stimuli. L hemisphere ERP responses at this latency distinguished between TOT stimuli within a category. A series of components reflecting categorical processing indicated changes in the first part of the ERP at both hemispheres and later components occuring only at the R hemisphere.

From the experiments on VOT and TOT, Molfese concluded that categorical perception of VOT occurs primarily in the R hemisphere and that this hemisphere is equally sensitive to such cues in non-speech

stimuli.

Molfese (1980b) also conducted a stop consonant study, in which he evaluated what cue enables us to identify a particular consonant independent of context. For example, the initial consonsants of /di/ and /du/ are both perceived as /d/ despite the fact that the acoustic cue, the second formant transition, is different for the two syllables. Molfese demonstrated that ERPs distinguished between phoneme class rather than acoustic differences. Twelve right-handed subjects listened passively to a series of computer-synthesized consonant-vowel syllables varying in the initial consonants /b/ or /g/ and the final vowel /i ,ae, ɔ/. Electrodes were placed at temporal and parietal sites of the L and R hemispheres, referenced to linked ear lobes. Data based on averages of 16 repetitions per stimulus were analyzed with PCA and Anovas. Two ERP factors related to the phoneme difference between /b/ and /g/ regardless of the vowel that followed them. Factor 1 contributed to the first 50 ms and the last 300 ms of the ERPs, and Factor 3 to the ERPs between 100 and 200 ms. An Anova on the Factor 3 responses indicated that only the L hemisphere response differentiated between consonants on the basis of phoneme class. In a more recent study Molfese, Linnville, Wetzel and Leicht (1985) found similar results in both right and left-handed subjects. The L hemisphere ERPs differed for stimuli from different phonetic categories but did not differ between stimuli that were acoustically different but within one phonemic category.

In summary, the studies by Molfese and his coworkers suggest that

ERPs reflect the phonological process of CP, and do not simply reflect the acoustic differences in stimuli. Molfese (1983) concluded that ERPs at both hemispheres index initial discrimination processes, that voicing contrast (VOT) discriminations are distinguished by additional R hemisphere activity, whereas place of articulation contrasts rely on further L hemisphere activity. The R hemisphere ERP activity evoked by the TOT stimuli shows that these effects are not unique to speech stimuli.

These interpretations must be tempered by consideration of the shortcomings of these studies (Picton & Stuss, 1984). The main problem is the evaluation of components resulting from the PCA since it is a data-driven form of analysis. In the Molfese studies no consistent pattern of ERP morphology or peak latencies emerged. Since the ERP indices of speech were of such small amplitude, and small numbers of ERP trials were averaged, the lack of replicability of the components means that the results could have stemmed from random noise and variability in the data. Another problem is that Molfese did not monitor possible noncerebral sources of artifact such as eye movements.

Another study with some relevance to CP is by Lawson and Gaillard (1981). They investigated whether ERPs reflect the behavioural findings that speech sounds differing in more than one feature are easier to discriminate than stimuli differing in only a single feature. Trains of consonant-vowel syllables were presented in a forced two-choice discrimination task, in which a designated target

stimulus differed from non-target stimuli on either, one, two, or three phonetic dimensions such as place of consonant, place of vowel and place of voicing consonant. Recordings were from Cz referenced to linked ear lobes. Results showed that the more a phoneme differed from the non-target stimulus, the shorter the N2 latency and the larger the N2 amplitude. Subjects made faster discriminations the greater the number of dimensions the stimuli varied on. Therefore, although these findings do not constitute any unique speech effects, they show that as with non-speech stimuli, the N2 component is sensitive to the degree of difference between speech sounds.

The Mismatch Negativity ERP

One ERP appears to be closely associated with pre-attentive echoic memory and would be an ideal measure to investigate early pre-attentive processing of speech stimuli. The Mismatch Negativity (MMN) is an automatic, pre-attentive cerebral response to a stimulus physically deviating from stimuli of the immediate past (Näätänen, 1988; Näätänen & Picton, 1986). This negative potential is largest fronto-centrally, begins at about 100 ms, and lasts about 200 ms. The MMN is usually recorded within the oddball paradigm of stimulus presentation, in which a train of "standard" stimuli contains infrequent "deviant" stimuli that differ slightly from the others. With increases in the magnitude of deviance, MMN amplitude increases until it reaches a plateau whereas latency and duration continue to shorten (Näätänen & Picton, 1986). The smaller the probability of

the infrequent stimuli, the larger the MMN; the important factor appears to be the local probability (in a 5-10 s period) (Näätänen & Picton, 1986). MMN is largest at ISIs of less than 4 s, becomes smaller over longer ISIs, and disappears at an ISI of 10 s (Näätänen, Paavilainen, Alho, Reinikainen & Sams, 1987). It is best observed when subjects' attention is directed away from the auditory stimuli because attention-related negative components such as the N2b, within the same latency range, can obscure the MMN (Näätänen, 1988). The MMN is evoked by the relation between the present stimulus and the previous stimulus (Näätänen, 1988); there is no MMN to the first stimulus in a run, to stimulus omission, to stimuli from a different modality than the preceding train of stimuli, to a change in ISI, or at very long ISIs (Näätänen & Picton, 1986; Näätänen et al., 1987). The MMN appears to be generated in the primary auditory cortex and perhaps the secondary auditory cortex (Näätänen, 1988). Support for this comes from magnetoencephalographic recordings (Hari et al., 1984; Sams et al., 1985) and ERP findings that there is a polarity reversal of MMN recorded from above and below the Sylvian fissure. Scherg et al. (in press) reported two distinct intracerebral sources for the MMN. One corresponded to N1 sources on the supratemporal plane, and the second source was located slightly more anteriorly on the supratemporal plane.

These findings suggest that the MMN reflects a pre-perceptual, automatic neuronal mismatch process, closely associated with echoic memory. Näätänan (1984, 1988) has interpreted the findings by

suggesting that each auditory stimulus leaves a 3-4 second duration trace in the auditory cortex that precisely represents physical features of the stimulus. This trace is formed automatically, without conscious perception, or selective attention, and may represent the neurophysiological basis of echoic memory (Neisser, 1966). The more often a stimulus is repeated the stronger the trace becomes. A change in stimuli within the same sensory modality, may cause a neuronal mismatch process, reflected by the MMN, capable of causing an attention switch. To summarize, the MMN appears to reflect a pre-perceptual detection of stimulus change based on a neuronal mismatch with the trace in echoic memory.

Aaltonen, Niemi, Nyrke and Tuhkanen (1987) examined the processing of speech sounds with the MMN. Using the oddball paradigm they presented stimuli drawn from the Finnish /i/ to /y/ continuum. The MMN was larger and occurred sooner in response to deviant stimuli when the two types of stimuli presented in a run consisted of the pure vowels /i/ and /y/, then when the stimuli consisted of the boundary stimulus and one of the pure vowels. This effect occurred equally in subjects that either ignored or attended to the stimuli. The MMN thus varied with the extent of the acoustic difference between the stimuli, just as it does to other auditory stimuli such as tonebursts. If the MMN had not differed between stimulus conditions, Aaltonen et al. suggested that this would have supported the notion that the MMN reflects categorical processing of speech stimuli at an early stage. Instead, they concluded that the

MMN reflects the processing of acoustic features of speech stimuli, and that there was no evidence that categorical processing took place by the time the MMN occured. A problem with this conclusion however, is that vowels are less categorically perceived than consonants (Borden & Harris, 1984).

Purpose and Hypthoses

The purpose of the study was to evaluate whether categorical processing of speech stimuli occurs at a pre-attentive stage. Subjects heard a train of standard speech sounds interspersed with two infrequent deviant speech sounds. To prevent attention-related effects from obscuring the MMN subjects ignored the auditory stimuli and read a book. All stimuli were drawn from the /ba/ to /da/ continuum. Both deviant stimuli were acoustically equidistant from the standard as measured by the slope of the second formant transition. One deviant stimulus lay across the CP boundary from the standard and the other lay within the same category as the standard. The deviant stimuli evoked a MMN in addition to the normal N1-P2-SP waves.

Hypotheses Concerning MMN. The experimental hypothesis was that the MMN to deviant stimuli across the CP boundary would be larger than the MMN to deviant stimuli within the same category as standard stimuli. Support for this hypothesis would suggest that categorical processing of speech stimuli occurs automatically at a pre-attentive stage. It would suggest that categorical perception

of speech does not rely on attention, conscious decision-making, or even that the listener be involved in a meaningful linguistic task. It would support the idea that there exists a special mode of speech perception at the echoic memory stage. If, in contrast, the MMN amplitude does not differ to targets from within or across categories, it would support the notion that at an early, pre-attentive stage, speech sounds are not processed categorically but continuously like other acoustic stimuli. This would indicate that CP of speech occurs at a later stage of processing.

Hypotheses Concerning Responses to Standard Stimuli. The hypothesis was that the P1-N1-P2-SP responses would not be affected by differences between the standard stimuli used in the different conditions of the experiment.

Hypotheses Concerning Scalp Distribution. It was hypothesized that the N1, SP, and the MMN would have different scalp distributions because the literature suggests that these components have different intra-cerebral source-configurations. Furthermore, since the MMN was more related to speech-processing then the N1 or SP, the MMN might be larger over the left hemisphere.

METHODS

Part I: ERP experiment

Subjects

Subjects were 10 English-speaking right-handed adults (5 males) between the ages of 21 and 45 (x = 31 years). Right-handedness was determined with the self-report 13-item Chapman and Chapman Measurement of Handedness (1987) (in Appendix 1). On a sample of 266 Ss this scale had a high internal consistency for both males and females (coefficient = .96 for both sexes), high test-retest reliability (r = .90 for males and .96 for females). On a sample of 53 Ss, the test had a correlation of .83 with a behavioural test in which Ss were observed for right or left-handed performance on 10 different tasks. Chapman and Chapman suggested that those with scores of 13 to 17 are right-handed; a subject reporting right-handed performance on all items scores 13. Subjects had an average score of 13.4 (range from 13 to 15). None of the subjects had reported any hearing difficulties. All subjects signed an informed consent form (Appendix 2).

Stimuli

Stimuli were drawn from 9 speech sounds that have been computer-modified to vary in the slope of the second formant transition along the /ba/ to /da/ continuum. Figure 2 is a sound

spectrogram of the /ba/ to /da/ continuum used in the study. A male speaker produced the sounds /ba/ and /da/ shown in Figure 1 and then the computer constructed seven intervening sounds. Thus, the most "ba"-like sound is stimulus 1 and the most "da"-like sound at the other end of the continuum is stimulus 9. Each stimulus was 350 ms in duration. The stimuli were digitized at 10 kHz and low-pass filtered at 8 kHz. Stimuli were delivered to the subject's right ear via TDH-49 earphones at an intensity of 85 dB peak SPL.

Recording

The electrical signals were recorded from gold-plated cup electrodes filled with electrode gel and fixed to the scalp with collodion-soaked gauze. Non-cephalic electrodes were taped to the body.

Electrodes were placed according to the International 10/20 system at Fz, Cz, Pz, C3, C4, T3, T4, M1, and M2 (Jasper, 1958). The reference was linked across a 20 kOhm potentiometer between the right sternoclavicular joint and the C7 vertebra, and the potentiometer was balanced to reduce EKG artifact (Wolpaw & Wood, 1982). The ground electrode was placed on the lower right side of the neck. Vertical eye movements were recorded with electrodes centered above and below the right eye. Horizontal eye movements were recorded between electrodes placed lateral to the outer canthi of the eyes. Inter-electrode impedances at 10 Hz were below 2000 kOhms.

Raw signals were recorded by a Nicolet EEG 1A97 machine set at a

gain of 20,000 for EEG and 5,000 for EOG. All channels were filtered with a bandpass of 0.15 - 70 Hz. These signals were routed to the computer for digitization and analysis.

Averaging

The recording sweep was 750 ms beginning 50 ms before the onset of the stimulus. The number of data points per sweep was 256.

Those sweeps in which the voltage exceeded a pre-determined level between 100 or 75 uV were rejected prior to averaging and data collection continued until the full number of trials was obtained. The level was adjusted for each subject on the basis of the amplitude of that subject's blink-artifacts. This protocol eliminated trials contaminated by blinking, large eye movements and muscle activity.

A calibration factor was applied to each channel to account for any differences in the calibration of the preamplifiers. These factors were calculated by measuring a 4 uV square wave pulse of 100 ms duration after was has passed through the amplifier and A-D converter.

Waveforms were digitally filtered with a rectangular filter, with a high-frequency cutoff of 20 Hz. Artifact associated with the small vertical or horizontal eye movements that occur during reading were removed from the ERP waveforms by subtracting from the waveforms that fraction of the potential that was equivalent to the field-spread of eye movement artifacts in resting EEG recordings for that particular subject and electrode location (Picton, 1987).

Testing and Experimental Design

During the experiment the subject sat and read in a sound-attenuated, electrically shielded room. The subject ignored the auditory stimuli, remained relaxed but alert, and minimized movement. The EEG was observed between trials to ensure that subjects did not fall asleep.

Stimulus presentation followed the oddball paradigm in which a train of identical standard stimuli are presented with occasional deviant stimuli differing from the standards. There were two conditions in the experiment. ERP Condition 1 (/ba/-standard) consisted of stimuli 1, 4, and 7. As Figure 2 shows, stimulus 4 lies midway between stimuli 1 and 7 according to the slope of the angle of the second formant transition. The identification function in Figure 3 shows that stimului 1 and 4 lie within the /ba/ category, and stimulus 7 lies across the CP boundary in the /da/ category. Stimulus 4, the standard stimulus, was presented at a probability of 80%, and the deviant stimuli 1 and 7, were each presented at probabilities of 10%. Thus stimulus 1 was the within-category deviant, and stimulus 7 was the across-category deviant. Similarly, for ERP Condition 2 (/da/-standard), the standard stimulus 6, usually perceived as a /da/, lies mid-way between the across-category deviant stimulus 3 and the within-category deviant stimulus 9 along the /ba/ to /da/ continuum. The probabilities were the same as in the ba-standard condition.

Blocks of 400 stimuli were presented with an onset-to-onset ISI of 900 ms. Within a block, stimuli were presented randomly except that targets always followed standard stimuli since Sams, Alho and Näätänen (1984) have shown that presenting 2 successive targets reduces the amplitude of the MMN to the second target. Each block was followed by a pause of about 30 s. There were 8 blocks for each condition, for a total of 6400 stimuli. The ordering of the blocks was 1,1,2,2,1,1,2,2 or 2,2,1,1,2,2,1,1. Total testing time was about 1 hour and 45 minutes.

Data Measurement

Baseline-to-peak amplitude of P1, N1, P2, and the SP were measured on the ERPs to the standard stimuli. The baseline was the average amplitude of the waveform for the 50 ms pre-stimulus period. Peaks were identified at Cz, and amplitudes were measured at those latencies at all electrodes. The P1 was the maximum positive peak between 60 to 100 ms; the N1 was the maximum negative peak between 80 and 140 ms; the P2 was the maximum positive peak between 120 and 180 ms. The SP was measured as the average amplitude from 250 to 350 ms to avoid including the stimulus offset response beginning at 350 ms.

The MMN waveforms were obtained from the following subtractions of average waveforms for each subject and condition:

(i) ERP Condition 1 (/ba/-standard):

Stimulus 1 (within-category deviant) - Stimulus 4 (standard)

= MMN within-category

Stimulus 7 (across-category deviant) - Stimulus 4 (standard)

= MMN across-category

(ii) ERP Condition 2 (/da/-standard:

Stimulus 9 (within-category deviant) - Stimulus 6 (standard)

= MMN within-category

Stimulus 3 (across-category deviant) - Stimulus 6 (standard)

= MMN across-category

Since pilot data indicated that the MMN started as early as the N1 did, and that the MMN was often present over the entire recording sweep, average amplitudes of the MMN were measured for the following time periods in ms: 0-60, 60-120, 120-210, 210-350, 350-450, and 450-700.

Statistical Analysis

The amplitudes and latencies were subjected to Analyses of Variance (ANOVA) with repeated measures using BMDP2V (Dixon, 1981). The P1, N1, P2, and the SP responses to standard stimuli were analyzed for the effect of ERP condition (/ba/-standard or /da/-standard) and electrode site. The MMN amplitudes were analyzed with confidence intervals to establish when a MMN was present (i.e. significantly different from zero) at a probability of 0.05 and 0.01. The MMN amplitudes at each time epoch were also analyzed for the effects of ERP condition (/ba/-standard or /da/-standard) and whether a deviant was within or across-category. Geisser-Greenhouse

adjustments were used to calculate the probability levels (Keppel, 1973) with $p<0.05$ considered significant. Newman-Keuls post-hoc tests (Keppel, 1973) were done when necessary.

Scalp Distribution. We compared the scalp distributions of the components using ANOVAs. Normally, a significant interaction of a given condition with the electrode factor is used to indicate that scalp distributions differ between conditions. However, there is an incompatibility between the additive model underlying ANOVA and the multiplicative effects on ERPs voltages produced by differences in source strength (McCarthy and Wood, 1985). McCarthy and Wood note, for example, that a 2-fold increase in source strength produces a 2-fold increase in the voltage at each electrode instead of adding a constant voltage to each location as assumed by the ANOVA model. McCarthy and Wood modelled potential distributions and showed that significant interactions involving scalp distributions can be produced between scalp distributions with identical shapes generated by the same source. Therefore, such interactions do not unambiguously show differences in shape between distributions. McCarthy and Wood suggest scaling the data to eliminate overall amplitude differences between experimental conditions. This preserves genuine differences in distributional shape, but does not confuse amplitude and shape differences.

Therefore, in order to prevent differences in overall amplitude from masquerading as changes in scalp distribution, the data were first normalized to eliminate overall amplitude differences between

waves. The amplitudes of N1, SP, and the MMN at a particular electrode (Xi) for each subject and wave were converted to a percentage of the range between the minimum amplitude (Xmin) and maximum amplitude (Xmax):

$$Xi = (Xi - Xmin)/(Xmax-Xmin)$$

The transformed data were then analyzed with a repeated measures ANOVA. Independent variables were wave and electrode. A difference in scalp distribution between components would show up as a significant Wave x Electrode interaction. If the interaction was not significant, this would indicate that components had similar scalp distributions. A main effect of electrode would mean there were amplitude differences between electrodes and that there thus existed a distinct scalp distribution.

Part II: Identification Performance

The second part of the study consisted of identification tasks with speech sounds drawn from the /ba/ to /da/ continuum. The main purposes of these experiments were to demonstrate (i) that the stimuli used in the ERP study were identified categorically, and (ii) to show that despite the high repetition of the standard stimulus, adaptation effects had little impact on the CP boundaries for the stimuli used in the ERP study.

Subjects

Five of the subjects from the ERP experiment participated in the study.

Stimuli

The stimuli were drawn from the /ba/ to /da/ continuum shown in Figure 2, with the same earphone presentation and intensity as in Part I. Stimuli were presented at a rate of one every 1.5 s.

Testing

A subject sat in the same quiet room used in Part I, and placed the index and middle fingers of the right hand on the two keys of a response panel that rested on the lap. The subject pressed the index finger to any stimulus perceived as "ba" and the middle finger to a stimulus perceived as "da". A subject was told that although some

stimuli were difficult to identify he or she should choose whether the stimulus sounded more like a "ba" or "da", and respond accordingly.

Experimental Design

Stimuli were presented randomly. Tests 1 to 3 consisted of 200 stimuli in order to achieve adequate numbers of stimulus presentations of the 9 different stimuli used in these tests. Tests 4 and 5 consisted of 100 stimuli since only 3 different stimuli were used. The 9 tests described below were presented to a subject in random order except that Tests 5a and 5b were the final tests:

Test 1 - All 9 stimuli with equal probabilities. The purpose of this task was to demonstrate how Ss labelled the sounds within the /ba/ to /da/ continuum.

Series 1 consisted of the following four tests (2a-5a) which used a standard stimulus drawn from the /ba/ end of the continuum.

Test 2a - Adaptation with stimulus 1, which had a 60% probability of occurance. Stimuli 2 to 9 of the continuum were each presented at probabilities of 5%. The purpose was to show that repetition of stimulus 1 causes the CP boundary to shift towards the /ba/ end of the continuum.

Test 3a - Adaptation with stimulus 4 (60% probability) to show that repetition of stimulus 4 does not cause the same shifts in the CP boundary as does repetition of stimulus 1.

Test 4a - Stimuli 1, 4 and 7 at probabilities of 10, 80 and 10%

respectively. One purpose was to demonstrate that with the stimulus presentation used in the ERP experiment, Ss identify stimuli 1 and 4 as "ba" and stimulus 7 as "da". Another purpose was to compare performance on this test with performance using the stimuli from Condition 2 (/da/-standard), to see if the stimuli used in one ERP condition are easier to identify than stimuli in the other.

Test 5a - Stimuli 1, 4 and 7 at the probabilities used in the ERP experiment, but with special instructions. In this task, the Ss were informed that there is a frequently occuring "ba"-like sound (Stimulus 4), an infrequent more distinct "ba" (Stimulus 1), and a "da"-like sound (Stimulus 7). We presented samples of each stimulus to Ss. They were instructed to identify the infrequent, more distinct "ba" and the "da"-like sound. The purpose was to establish that Ss can discriminate between stimulus 1 (the deviant) and stimulus 4 (the standard) even though they lie within the same category, but that this within-category discrimination is made with more difficulty (eg. more slowly and perhaps less accurately) than the identification of the across CP boundary "da".

Series 2 (2b-5b) consisted of homologous tests using standard stimuli drawn from the /da/ end of the continuum.

Test 2b - Adaptation with stimulus 9 to demonstrate CP boundary shifts.

Test 3b - Adaptation with stimulus 6 to show that adaptation boundary shifts are minimal.

Test 4b - Stimuli 3,6,9 at probabilities of 10,80 and 10%.

Test 5b - Stimuli 3,6 and 9 with subjects instructed to identify stimuli 3 and 9.

Recording

The response keys were positive-pressure microswitches. Each key generated a voltage that lasts 50 ms. The recording sweep was 1280 ms with a 50 ms pre-stimulus baseline. If the subject pressed both buttons in response to a stimulus, only the first response was recorded. The raw data were averaged off-line to give reaction time histograms.

Data Analysis

Response measures consisted of the percentage of times a stimulus was identified as "ba" and reaction times. Identification functions were plotted for each test.

The Chi square test (Keppel & Saufley, 1980) was used to compare the identification functions for particular stimuli under different conditions or for different stimuli under the same condition.

The median RTs from each subject to deviant stimuli for Tests 4 and 5 were analyzed with repeated measures ANOVA. The median RT for each subject was used because this value was least affected by very slow or fast button presses that were probably errors. Independent variables were series (1 or 2), and boundary (within or across-category).

RESULTS

Part I: ERP experiment

Response to Standard Stimuli

The speech stimuli evoked a P1, N1, P2 and SP waveform as shown in Figure 4.

Insert Figure 4 about here

Table 1 shows the average latencies and amplitudes for these waves at Cz where they were maximal in response to the standard stimuli. The ANOVAs showed that the latencies and amplitudes of these waves at Cz did not differ between the /ba/-standard and /da/-standard conditions.

Insert Table 1 about here

Mismatch Negativity

Table 2 summarizes the MMN measurements at Cz. Figure 4 shows that in ERP Condition 1 (/ba/-standard), the across-category deviant elicited a MMN that began at about 60 ms after stimulus onset and was present over the full recording sweep, whereas the within-category deviant did not elicit a MMN. As Table 2 shows, the method of confidence intervals indicated that in the /ba/-standard condition the

MMN to across-category deviants was present from 60 to 700 ms. The grand-mean waveforms for the /da/-standard condition in Figure 4 suggest a small MMN to both within and across-category deviants. However, as Table 2 shows, few of these measurements were reliably different from noise and these only for the within-category deviants.

Insert Table 2 about here

The MMN amplitude at each time epoch was analyzed by an ERP Condition (/ba/ or /da/-standard) x Boundary (within or across-category) ANOVA at Fz and Cz, where responses to standards and deviants were most clearly present on the grand-mean waveforms. There were no significant effects on the MMN from 0 to 60 ms. The ANOVAs on the MMN from 60 to 120 ms yielded a significant ERP Condition x Boundary interaction at Fz ($F=14.44$; $df=1,9$; $p<0.01$) and at Cz ($F=11.47$; $df=1,9$; $p<0.01$). The ANOVA on the MMN from 120 to 210 ms also yielded a significant ERP Condition x Boundary interaction at Fz ($F=6.12$; $df=1,9$; $p<0.01$) and at Cz ($F=4.84$, $df=1,9$; $p<0.055$). There were no significant effects on the MMN from 210 to 350 or 350 to 500 ms. For the MMN from 500 to 700 ms there was a significant main effect of boundary at Fz ($F=13.15$; $df=1,9$; $p<0.01$) and at Cz ($F=13.53$; $df=1,9$; $p<0.01$).

Significant interactions were followed by one-way ANOVAs to investigate the effect of boundary on each of ERP Conditions 1 and 2. The main effect of boundary for the MMN from 500 to 700 ms was also

analyzed this way. In ERP condition 1 (/ba/-standard), the MMN amplitude at Cz was larger for across-category deviants than for within-category deviants over time ranges 60 to 120 ms ($F= 9.68$; $df=1,9$; $p<0.01$), 120 to 210 ms ($F=8.12$; $df=1,9$; $p<0.05$) and from 500 to 700 ms ($F=8.90$; $df=1,9$; $p<0.05$). The MMN at Fz was also larger for across-category than within-category deviants in the /ba/-standard condition over the time ranges of 60 to 120 ms ($F=11.62$; $df=1,9$; $p<0.01$), 120 to 210 ($F=9.22$; $df=1,9$; $p<0.05$), and 500 to 700 ($F=6.81$; $df=1,9$; $p<0.05$).

There were no significant effects on the MMN in ERP condition 2 (/da/-standard) except at Fz over 500 to 700 ms, at which period the MMN was larger for across-category than for within-category deviants ($F=5.66$; $df=1,9$; $p<0.05$).

Scalp Distribution

This analysis involved the N1 and SP responses to the standard stimulus in the /ba/-standard condition and the MMN to the across-category deviants in the /ba/-standard condition at the three time epochs (60-120, 120-210, 500-700 ms) where it differed significantly from the response to within-category deviants. Table 3 shows the means and standard deviations of the standardized amplitudes for components N1, SP, and MMN at each electrode; these values are each shown subtracted from 1 so that more negative amplitudes are reflected as greater values.

Insert Table 3 about here

The MMN Scalp Distribution. Figure 5 shows the MMN to the across-category deviant from the /ba/-standard condition mapped on an outline of the head. The initial analysis compared the scalp distribution of the MMN at the three time epochs using a 3 x 9 (Time Epoch x Electrode) ANOVA. The Time Epoch x Electrode interaction was not significant ($F=1.52$; $df=16,144$; $p=.20$) showing that the MMN did not change its scalp distribution over time ranges 60 to 120, 120 to 210, and 500 to 700 ms. There was a main effect of electrode ($F=9.39$; $df=8,72$; $p<0.01$), showing that the MMN had a clearly defined scalp distribution.

Insert Figure 5 about here

Comparison of the N1, SP, and MMN Scalp Distributions. Figure 6 shows the response to the standard stimulus from ERP Condition 1 (/ba/-standard) mapped on an outline of the head viewed from above.

Insert Figure 6 about here

The analysis compared the scalp distributions of the N1, SP and the MMN. Despite the fact that the scalp distribution of the MMN did not differ significantly between 60-120, 120-210, and 500-700 ms, we chose not to collapse across time epochs but to compare the scalp

distribution of the MMN at distinct time ranges with N1 and SP. Since the N1 had an average peak latency of 121 ms, this allowed us to examine the relationship of the scalp distribution of the N1 to the MMN occurring before and after N1 peak latency. The MMN from 500 to 700 ms was not further analyzed since as Table 3 shows the converted amplitudes all hovered around the same value, meaning that there was less of a distinct scalp distribution at this latency range.

The first analysis was a Wave (N1, SP and MMN from 60-120) x Electrode ANOVA. This yielded a significant Wave x Electrode interaction ($F=3.02$; $df=16,144$; $p<0.05$). This was followed with factorial comparisons that compared each wave to the other in 2 x 9 (Wave x Electrode) ANOVAs. There was a main effect of electrode from the ANOVA on N1 and SP ($F=56.21$; $df=8,72$; $p<0.01$), and from the ANOVA on N1 and the MMN ($F=35.52$; $df=16,144$; $p<0.05$). This indicates that the N1 and the SP have a similar scalp distribution and that the N1 and the MMN occuring from 60-120 ms have a similar scalp distribution. The comparison of the SP and the MMN (60-120) yielded a significant Wave x Electrode interaction ($F=4.03$; $df=8.72$; $p<0.05$) indicating that they have scalp distributions that differ in some way. Other interactions did not reach significance.

The second analysis compared the scalp distributions of the MMN from 120-210 ms, the N1 and the SP. This yielded a significant Wave x Electrode interaction ($F=3.93$; $df=16,144$; $p<0.01$). This was followed by factorial comparisons of each wave with the other. These comparisons yielded a significant Wave (N1 vs MMN) x Electrode

interaction ($F=3.60$; $df=8,72$; $p<0.05$) and a significant Wave (SP vs. MMN) x Electrode interaction ($F=4.91$; $df=8,72$; $p<0.01$) indicating that the MMN from 120 to 210 ms had a scalp distribution different from both the N1 and the SP.

Post-hoc comparisons showed that as evident in Table 3, the N1 and SP were maximal fronto-centrally (Figure 6), whereas the MMN (Figure 5) from 60 to 120 and 120 to 210 ms was maximal centro-parietally.

Laterality effects. Newman-Keuls tests showed that the N1 amplitude at T3 was greater than at T4. The SP was larger at C3 and T3 than at C4 and T4. The MMN from 60 to 120 was larger at T3 than at T4. And the MMN from 120 to 210 ms did not have any significant laterality effects.

Part II: Button Press Results

Identification Performance

Test 1. Figure 3 shows how stimuli along the /ba/ to /da/ continuum were identified in Test 1 in which subjects were presented with all 9 stimuli at equal probabilities and identified stimuli as either /ba/ or /da/. The categorical boundary stimulus was stimulus 5. The stimuli used in the ERP experiment were 1, 4 and 7 in Condition 1 and 3, 6 and 9 in Condition 2. As Figure 3 shows, stimuli 1 and 4 were usually identified by subjects as /ba/ and stimulus 7 as /da/. Stimuli 1 and 4 did not differ in the frequency of % /ba/ responses (Chi square = 0.83; p=0.36). Subjects identified stimulus 9 as /da/ 100% of the time, and stimulus 3 as /ba/ 100% of the time. The categorical perception of stimulus 6 was not as strong. Subjects identified stimulus 6 as /da/ 86% of the times it was presented and as /ba/ 14% of the time, which differed significantly from the responses to the prototypical /da/ stimulus 9 (Chi square=12.32; $p<0.01$). Thus, stimuli 1, 4 and 7 satisfied one of the requirements for being categorically perceived, that stimuli along a physical continuum be given one label on one side of a category boundary and another label on the other side. Stimuli 3, 6 and 9 also largely satisfied this condition although stimulus 6 was less strongly categorized than the other two stimuli.

Test 2. Repetition of the prototypical /ba/ and /da/ caused

significant shifts in the identification functions. Figure 3 shows how repetition of stimulus 1 caused a shift of the identification curve towards the /ba/ end of the continuum meaning that more of the stimuli were identified as /da/. After repetition of stimulus 1, the boundary stimulus, stimulus 5 shifted significantly from 49% /ba/ response to 17% /ba/ response (Chi square=13.35; $p<0.01$). After repetition of stimulus 9 there was an opposite shift; subjects perceived more of the stimuli as /ba/ and stimulus 5 shifted significantly to a 91% /ba/ response (Chi square = 21.60; $p<0.01$).

Test 3. Figure 7 shows the results of Test 3 in which either stimulus 4 or stimulus 6 had a probability of 60% and the remaining stimuli a probability of 5%. Repetition of stimulus 4 or stimulus 6 did not result in significant shifts in the identification functions. Repetition of stimulus 4 did not cause the % /ba/ response to this stimulus to change from the % /ba/ response in the equiprobable stimulus condition (Chi square=2.50; $p=0.11$). Similarly, repetition of stimulus 6 did not significantly change the % /da/ response to this stimulus to change from responses in the equiprobable condition (Chi square=3.08; $p=0.08$).

Insert Figure 7 about here

Test 4. Adaptation caused some shifts in boundaries for the stimuli used in the ERP experiment. Figure 8 shows the results of Test 4 in which subjects listened to runs containing either stimuli

1, 4 and 7 or 3, 6 and 9 at probabilities of 10, 80 and 10%. Stimulus 4 was identified as /ba/ 78% of the time. This differed significantly from the % /ba/ response to stimulus 1 (Chi square=16.51; $p<0.01$). Similarly, the standard stimulus 6 was identified as /da/ only 66% of the time. Stimuli 6 and 9 differed in the % /ba/ response (Chi square=18.86; $p<0.01$). Thus stimulus 6 was very close to being on the category boundary which is defined as a stimulus with 50% /da/ response.

Insert Figure 8 about here

Reaction Times

The button press reaction times for each test are summarized in Appendix III. These values are based on the median button press RT for each subject. Figure 9 shows how identification reaction times were slower around the category boundary stimuli.

Insert Figure 9 about here

Reaction time histograms from Test 4 are shown in Figure 10.

Insert Figure 10 about here

Table 4 provides a quantification of grand-mean histograms for Tests 4 and 5. This table shows that in Tests 4 and 5, subjects

responded faster to across-category deviant stimuli than to within-category deviants. However these differences in reaction time did not reach statistical significance in the ANOVAs of median reaction times. Subjects' median RTs from Tests 4 and 5 were analyzed in a 2 x 2 (Series x Within/Across Category) ANOVA. There were no significant main effects or interactions.

Insert Table 4 about here

Correlation of BP Responses with MMN Amplitude

The relationship between how stimuli were categorized in Test 4 and the size of the MMN amplitude was examined with a Pearson Product-Moment Correlation for the 5 subjects for whom identification results existed. The MMN amplitude from 120 to 210 ms at Cz was used for this analysis since a MMN was present for both the /ba/-standard and the /da/-standard conditions at this time range. The identification performance values used in the correlation were the absolute differences between a subject's % of /ba/ response to the deviants and to the standard stimulus. The larger the resulting value, the greater the difference in how a standard and deviant were categorized. The amplitude of the MMN was significantly related to categorical identification of stimuli ($r=-0.56$; $df=18$; $p<0.01$). The larger the difference between the percentage of /ba/ responses to a deviant and to a standard, the larger the MMN amplitude.

DISCUSSION

In this study subjects heard a train of standard speech sounds interspersed with two infrequent deviant speech sounds, which were acoustically equidistant from the standard in the extent of the slope of the second formant transition. The MMN was obtained by subtracting the ERP waveform to standard stimuli from the ERP waveform to one or the other deviant stimuli. When the standard stimuli were drawn from the /ba/ end of the continuum, a MMN evoked by across-category deviants was significantly larger than that evoked by within-category deviants. These results support the experimental hypothesis that the MMN reflects categorical processing of the speech stimuli. In the condition where the standard stimuli were drawn from the /da/ end of the continuum there were no significant differences between the MMN to the within-category and across-category deviant stimuli. However, regardless of phonemic condition, MMN amplitude had a positive correlation with categorical identification. Thus the more that subjects identified a deviant as belonging to a different category from the standard, the larger the MMN to this deviant in the ERP experiment.

The grand mean data in the /da/-standard condition (Figure 4 and Table 2) suggest that there may have been a small MMN for both within-category and across-category deviant stimuli. This could support the hypothesis that stimuli were processed along an acoustic continuum. The very small and variable amplitudes may have prevented

statistical confirmation that a MMN existed. Thus, although the grand mean data suggest the presence of a small MMN to both within and across-category deviants, the statistical analyses do not confirm this.

The identification results suggest another explanation for the MMN findings in the /da/-standard condition. Stimulus 6, the standard stimulus, was not as strongly categorized as a /da/, as stimulus 4, the standard stimulus in the /ba/-standard condition, was categorized as a /ba/. As Figure 3 shows, when stimuli were presented equiprobably, stimulus 6 had a significantly lower percentage /da/ response than the prototypical /da/ stimulus 9. In Test 4, which used the same stimulus presentation and probabilities as the ERP experiment, stimulus 6 was categorized as a /ba/ 34% of the time. These findings suggest that stimulus 6 was too close to the category boundary to test our hypothesis. Stimulus 6 could have activated both /ba/ and /da/ receptors in the cortex, so that neither deviant would elicit much of a MMN. It is also possible that the variation in how the standard stimulus was perceived resulted in a very small MMN to both within-category and across-category deviants. In a recent study Winkler, Paavilainen, Alho, Reinikainen and Näätänen (in press) showed that when a repetitive standard tone varied within a range of intensity, the MMN to deviant tones still occurred but with reduced amplitude. Winkler suggested that the trace for the standard could include the variation in the standard. Thus, both deviant stimuli in the da-standard condition may have

evoked small inconsistent MMNs that reflected this variation in how the standard stimulus 6 was perceived.

Another possible explanation is that the /da/ stimuli are not as effective as the /ba/ stimuli in setting up a categorical trace. The sounds /ba/ and /da/ are part of the /ba-/da-ga/ continuum that are produced by movements of the lips and the tongue at the front, middle and back of the mouth. Although /ba/ is truly at an end-point of the perceptual continuum, /da/ is actually mid-way along this continuum. Humans may be more sensitive to the end-points than the middle point.

Even the /ba/-standard condition did not totally agree with our hypothesis. Based on pilot data we had expected that all deviant stimuli would elicit a MMN, and that the MMN to deviants that lay across a category would be larger than to deviants that lay within a category. However, the results from the /ba/-standard condition showed that the within-category deviants did not elicit consistent MMNs in most subjects. It would appear that deviant stimuli from within the same category as the standard do not elicit the neuronal mismatch process.

There have not been extensive studies of ERPs and categorical perception. Molfese (1980b) and Molfese et al. (1985) reported significant ERP changes with categorical perception. However, it is difficult to relate their waveforms to those described in this paper. Our findings contrast with those of Aaltonen et al. (1987) who found that the MMN reflected acoustic and not categorical differences

between vowel sounds. However, the vowel continua are not as strongly categorically perceived as consonants are.

The Nature and Scalp Distribution of the Responses

The MMN in the /ba/-standard condition was present at about 60 to 120 ms after stimulus onset. Since the second formant frequency transition itself lasts about 50 ms, this suggests that the categorical processing of speech occurs very quickly and that the MMN process was evoked by differences between stimuli in the second formant transition. The MMN was sustained for several hundred ms after the stimulus and it did not change its scalp distribution over time. The usual duration of the MMN to simple acoustic differences in tones is about 200 ms (Näätänen, 1988). This suggests that speech stimuli evoke activity that is sustained for a longer period of time than the activity evoked by tones.

The differences between scalp distributions of the N1, SP, and the MMN were quite subtle. The N1 and the SP were maximal fronto-centrally whereas the MMN had a more centro-parietal distribution. These results agree with previous findings on the scalp distribution of the N1, but the findings on SP differ slightly from previously reported studies using tone stimuli. We did not find as has been previously reported that the SP had a scalp distribution anterior to that of the N1 (Picton et al., 1978; Scherg et al., in press). Instead, as Hillyard and Woods (1979) reported, the SP appeared to spread laterally in response to speech stimuli.

The SP was the most strongly lateralized of the waves. It was significantly larger over the left than the right hemisphere at both temporal and central electrodes. The N1 was larger at left than at right temporal electrodes. The grand mean MMN waveforms suggest that the MMN was slightly larger at left than right hemisphere central and temporal electrodes. But the only significant MMN laterality effect was a larger amplitude from 60 to 120 ms at T3 than at T4. The slightly larger amplitude of N1, SP, and the MMN over the left hemisphere electrodes agree with previous findings (Näätänen and Picton, 1987; Scherg et al., in press) that these components are usually larger contralateral to the stimulated ear. It would be necessary to compare the responses to tones and speech stimuli delivered to each ear to investigate if the larger amplitudes over the left hemisphere constitute any special speech-related effects.

To pursue questions this study has raised it would be worthwhile to use stimuli drawn from the entire /b-d-g/ continuum to see whether the MMN reflects early categorical processing at the /g/ endpoint of the continuum as it does at the /b/ endpoint. Another line of investigation would be to use other speech sounds that are categorically perceived, such as voice-onset time to see if the categorical MMN phenomenon is generalizeable to other categorically perceived speech sounds. Another area to investigate would be to study adults as they learn a new language or as they develop better skills in categorical perception. For example, if subjects were trained to identify the stimuli in the /da/-standard condition more

effectively, would the stimuli elicit a categorical MMN? How would the MMN change over time as subjects developed their categorical perception skills?

Summary

To summarize, there is clear support for a categorically evoked MMN in the /ba/-standard condition. Although the ERP results of the /da/-standard condition could be interpreted as reflecting processing along an acoustic continuum, the identification results suggest that the standard stimulus may have been too close to the category boundary to evoke a consistent categorical mismatch. The correlation between categorical identification and MMN amplitude support this suggestion. The standard stimulus in the /da/-standard condition was less strongly categorized and there were correspondingly weaker MMN effects. Thus the results suggest that categorical processing of speech sounds occurs independently of attention at an echoic memory stage.

REFERENCES

Aaltonen, O., Niemi, P., Nyrke, T., & Tuhkanen, M. (1987). Event-related brain potentials and the perception of a phonetic continuum. Biological Psychology, 24, 197-207.

Ades, A.E. (1977). Vowels, consonants, speech and nonspeech. Psychological Review, 84, 524-530.

Borden, G.J. & Harris, K.S. (1984). Speech science primer (2nd ed.). Baltimore: Williams & Wilkins.

Chapman, L.J. & Chapman, J.P. (1987). The measurement of handedness. Brain and Cognition, 6, 175-183.

Darwin, C.J. (1976). The perception of speech. In E.C. Carterette & M.P. Friedman (Eds.), Handbook of perception, Vol. 7 : Language and speech (pp. 175-226). New York: Academic Press.

Dixon, W.J. (Ed.) (1981). BMDP Statistical Software 1981. Berkeley, CA: University of California Press.

Eimas, P.E. & Corbit, J.D. (1973). Selective adaptation of linguistic feature detectors. Cognitive Psychology, 4, 99-109.

Eimas, P.E., Miller, J.L. & Jusczyk, P.W. (1987). On infant speech perception and the acquisition of language. In S. Harnad (Ed.), Categorical perception (pp. 161-195). Cambridge: Cambridge University Press.

Foreit, K.G. (1977). Linguistic relativism and selective adaptation for speech: A comparative study of English and Thai. Perception & Psychophysics, 21, 347-351.

Fujisaki, H., & Kawashima, T. (1969). On the modes and hearing mechanisms of speech perception. Annual Report of the Engineering Research Institute (Vol. 28, pp. 67-73). Tokyo: Faculty of Engineering, University of Tokyo.

Fujisaki, H. & Kawashima, T. (1971). A model for the mechanisms of speech perception: Quantitative analysis of categorical effects in discrimination. Annual Report of the Engineering Research Institute (Vol. 30, pp. 59-68). Tokyo: Faculty of Engineering, University of Tokyo.

Hari, R., Hämäläinen, M., Ilmoniemi, R., Kaukoranta, E., Reinikainen, K., Salminen, J., Alho, K., Näätänen, R., & Sams, M. (1984). Responses of the primary auditory cortex to pitch changes in a sequence of tone pips: Neuromagnetic recordings in man. Neuroscience

Letters, 50, 127-132.

Harnad, S. (1987). Psychophysical and cognitive aspects of categorical perception: A critical overview. In S. Harnad (Ed.), Categorical perception (pp. 1-25). Cambridge: Cambridge University Press.

Hillyard, S.A. & Woods, D.L. (1979). Electrophysiological analysis of human brain function. In M.S. Gazzaniga (Ed.), Handbook of behavioural neurobiology Vol. 2. (pp. 345-378). New York: Plenum Press.

Jasper, H.H. (1958). Report of the committee on methods of clinical examination in electroencephalography. Electroencephalography and Clinical Neurophysiology, 10, 370-375.

Keppel, G. (1973). Design and analysis: A researcher's handbook. Englewood Cliffs, N.J.: Prentice-Hall.

Keppel, G. & Saufley, W.H. (1980). Introduction to design and analysis. San Francisco: WH Freeman.

Kuhl, P.K. (1987). The special-mechanisms debate in speech research: Categorization tests on animals and infants. In S. Harnad (Ed.), Categorical perception (pp. 355-386). Cambridge: Cambridge University

Press.

Lawson, E.A. & Gaillard, W.K. (1981). Mismatch negativity in a phonetic discrimination task. Biological Psychology, 13, 281-288.

Liberman, A.M., Cooper, F.S., Shankweiler, D. & Studdert-Kennedy, M. (1967). Perception of the speech code. Psychological Review, 74(6), 431-461.

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(5), 358-368.

Locke, S & Kellar, L. (1973). Categorical perception in a non-linguistic mode. Cortex, 9, 355-369.

Lovrich, D., Novick, B. & Vaughan, H.G. (1988). Topographic analysis of auditory event-related potentials associated with acoustic and semantic processing. Electroencephalography and Clinical Neurophysiology, 71, 40-54.

Massaro, D.W. (1987). Categorical partition: A fuzzy-logical model of categorization behavior. In S. Harnad (Ed.), Categorical perception (pp.254-286). Cambridge: Cambridge University Press.

McCallum, W.C., & Curry, S.H. (1979). Hemisphere differences in event related potentials and CNVs associated with monaural stimuli and lateralized motor responses. In D. Lehmann & E. Callaway. (Eds.), Human Evoked Potentials: Applications and Problems (pp. 235-250). New York: Plenum.

McCallum, W.C. & Curry, S.H. (1980). The form and distribution of auditory evoked potentials and CNVs when stimuli and responses are lateralized. In H.H. Kornhuber & L. Deecke. (Eds.), Progress in Brain Research, Volume 54. Motivation, Motor and Sensory Processes of the Brain: Electrical Potentials, Behaviour and Clinical Use (pp. 767-775). Amsterdam: Elsevier.

McCarthy, G. & Wood, C. (1985). Scalp distributions of event-related potentials: An ambiguity associated with analysis of variance models. Electroencephalography and clinical Neurophysiology, 62, 203-208.

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A.M., Jenkins, J.J. & Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [l] by native speakers of Japanese and English. Perception & Psychophysics, 18, 331-340.

Molfese, D.L. (1978). Neuroelectrical correlates of categorical speech perception in adults. Brain and Language, 5, 25-35.

Molfese, D.L. (1980a). Hemispheric specialization for temporal information: Implications for the perception of voicing cues during speech perception. Brain and Language, 11, 285-299.

Molfese, D.L. (1980b). The phoneme and the engram: Electrophysiological evidence for the acoustic invariant in stop consonants. Brain and Language, 9, 372-376.

Molfese, D.L. (1983). Event related potentials and language processes. In A.W.K. Gaillard and W.R. Ritter (Eds), Tutorials in event-related potential research: Endogenous components (pp. 345-367). Amsterdam: North-Holland.

Molfese, D.L. (1987). Electrophysiological indices of categorical perception for speech. In S. Harnad (Ed.), Categorical perception (pp.421-443). Cambridge: Cambridge University Press.

Molfese, D.L., Linnville, S.E., Wetzel, W.F. & Leicht, D. (1985). Electrophysiolgical correlates of handedness and speech perception contrasts. Neuropsychologia, 23(1), 77-86.

Näätänen, R. (1984). In search of the short-duration memory trace of a stimulus in human brain. In L. Pulkkinen and P. Lyytinen (Eds.), Essays in Honour of Martti Takala. Jyväskylä Studies in Education, Psychology and Social Science (pp. 22-36). Jyväskylä,

Finland: University of Jyväskylä.

Näätänen, R. (1988). Implications of ERP data for psychological theories of attention. Biological Psychology, 26, 117-163.

Näätänen, R., Paavilainen, P. Alho, K. Reinikainen, K., & Sams, M. (1987). Inter-stimulus interval and the mismatch negativity. In C. Barber, & T. Blum (Eds.), Evoked potentials III (pp. 392-397). London: Butterworths.

Näätänen, R. & Picton, T.W. (1986) N2 and automatic versus controlled processes. In W.C. McCallum, R. Zappoli & F. Denoth (Eds.), Cerebral psychophysiology: Studies in event-related potentials (EEG supple. 38) (pp. 169-186). Amsterdam: Elsevier.

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(4), 375-425.

Neisser, U. (1966). Cognitive psychology. New York: Appleton-Century-Crofts.

Pastore, R.E. (1987). Categorical perception: Some psychophysical models. In S. Harnad (Ed.), Categorical perception (pp. 29-52). Cambridge: Cambridge University Press.

Picton, T.W. (1987). The recording and measurement of evoked potentials. In A.M. Halliday, W.R. Butler & R. Paul (Eds.), A textbook of clinical neurophysiology (pp. 23-40). Toronto: John Wiley & Sons.

Picton, T.W. & Stuss, D.T. (1984). Event-related potentials in the study of speech and language: A critical review. In D. Caplan, A.R. Lecours & A. Smith (Eds.), Biological perspectives on language (pp. 303-360). Cambridge, MA: MIT Press.

Picton, T.W., Woods, D.L., & Proulx, G.B. (1978a). Human auditory sustained potentials. I. The nature of the response. Electroencephalography and Clinical Neurophysiology, 45, 186-197.

Picton, T.W., Woods, D.L., & Proulx, G.B. (1978b). Human auditory sustained potentials. II. Stimulus relationships. Electroencephalography and Clinical Neurophysiology, 45, 198-210.

Repp, B. H. & Liberman, A.M. (1987). Phonetic category boundaries are flexible. In S. Harnad (Ed.), Categorical perception (pp.89-112). Cambridge: Cambridge University Press.

Rosen, S. & Howell, P. (1987). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. Harnad (Ed.),

Categorical perception (pp. 113-160). Cambridge: Cambridge University Press.

Sams, M., Alho, K., & Näätänen, R. (1984). Short-term habituation and dishabituation of the mismatch negativity of the ERP. Psychophysiology, 21(4), 434-441.

Sams, M., Hamalainen, M., Antervo, A., Kaukoranta, E., Reinikainen, K., & Hari, R. (1985). Cerebral neuromagnetic responses evoked by short auditory stimuli. Electroencephalography and Clinical Neurophysiology, 61, 254-266.

Samuel, A.G. (1982). Phonetic prototypes. Perception & Psychophysics, 31, 307-314.

Scherg, M., Vajsar, J., & Picton, T.W. (in press). A source analysis of the human auditory evoked potentials. Journal of Cognitive Neuroscience.

Scherg, M., & von Cramon, D. (1985). Two bilateral sources of the late AEP as identified by a spatio-temporal dipole model. Electroencephalography and Clinical Neurophysiology, 62, 32-44.

Scherg, M., & von Cramon, D. (1986). Evoked dipole source potentials of the human auditory cortex. Electroencephalography and Clinical

Neurophysiology, 65, 344-360.

Seldon, H.L. (1985). The anatomy of speech perception. In A. Peters & E.G. Jones (Eds.), Cerebral cortex (Vol. 4): Association and auditory cortices (pp. 273-327). New York: Plenum Press.

Stevens, K.N. (1981). Constraints imposed by the auditory system on the properties used to classify speech sounds: Data from phonology, acoustics and psycho-acoustics. In T.F. Myers, J. Laver, & J. Anderson (Eds.), The cognitive representation of speech (pp. 61-74). Amsterdam: North Holland.

Studdert-Kennedy, M., Liberman, A.M., Harris, K.S., & Cooper, F.S. (1970). Motor theory of speech perception: A reply to Lane's critical review. Psychological Review, 77(3), 234-249.

Werker, J.F. & Tees, R.C. (1984). Cross-language speech perception: Evidence for perceptual reorganization during the first year of life. Infant Behaviour and Development, 7, 49-63.

Winkler, I., Paavilainen, P., Alho, k., Reinikainen, K. & Näätänen, R. (in press). The effect of small variation of the frequent auditory stimulus on the event-related brain potential to the infrequent stimulus. Psychophysiology.

Wolpaw, J.R., & Penry, J.K. (1975). A temporal component of the auditory evoked response. Electroencephalography and Clinical Neurophysiology, 39, 609-620.

Wolpaw, J.R., & Penry, J.K. (1977). Hemispheric differences in the auditory evoked response. Electroencephalography and Clinical Neurophysiology, 43, 99-102.

Wolpaw, J.R. & Wood C.C. (1982). Scalp distribution of human auditory evoked potentials I: Evaluation of reference electrode sites. Electroencephalography and Clinical Neurophysiology, 54, 15-24.

Wood, C.C., & Wolpaw, J.R. (1982). Scalp distribution of human auditory evoked potentials. II. Evidence for multiple sources and involvement of auditory cortex. Electroencephalography and Clinical Neurophysiology, 54, 25-38.

Table 1

Latencies and Amplitudes for P1, N1, P2, and SP to Standard Stimuli. Mean and standard deviation at Cz.

	/Ba/-standard		/Da/-standard	
	Latency (ms)	Amplitude (uV)	Latency (ms)	Amplitude(uV)
P1	65 ± 9.2	1.1 ±0.7	64 ± 4.4	1.1 ±1.0
N1	121 ±14.0	-2.0 ±1.3	118 ±10.4	-2.1 ±1.1
P2	175 ±17.2	-0.4 ±1.3	174 ±17.3	-0.2 ±1.3
SP		-2.6 ±1.1		-2.5 ±1.1

Table 2
MMN Mean Amplitude (uV) and Standard Deviations at Cz for /Ba/-standard (1) and /Da/-standard (2) Conditions.

Time Epoch (ms)	Condition	Within-Category	Across-Category
0-60	1	.11 ± .43	.06 ± .47
	2	-.21 ± .38	.10 ± .43
60-120	1	.35 ± .44	-.44 * ± .58
	2	-.26 ± .68	.13 ± .66
120-210	1	.17 ± .35	-.61 ** ± .60
	2	-.30 * ± .45	-.15 ± .67
210-350	1	-.07 ± .38	-.64 * ± .86
	2	-.52 * ± .77	-.25 ± .92
350-500	1	.06 ± .65	-.64 * ± .86
	2	-.20 ± .58	-.22 ± .98
500-700	1	.37 ± .57	-.45 * ± .60
	2	.23 ± .63	-.09 ± .63

Measurements significantly different from zero at * $p<0.05$ or ** $p<0.01$.

Table 3

Scalp Distributions. Means and standard deviations of normalized amplitudes for the /ba/-standard condition. N1 and SP measurements are for the standard stimulus. MMN measurements are for the across-category deviant stimulus.

	Fz	Cz	Pz	C3	C4	T3	T4	M1	M2
N1	.92	94	.61	.88	.79	.52	.17	.29	.13
	±.13	±.09	±.24	±.07	±.14	±.25	±.20	±.20	±.15
SP	.95	.97	.59	.90	.76	.44	.10	.19	.07
	±.07	±.04	±.20	±.08	±.11	±.23	±.09	±.25	±.14
MMN	.61	.75	.84	.90	.63	.59	.32	.29	.28
(60-120)	±.30	±.25	±.16	±.15	±.28	±.30	±.21	±.36	±.35
MMN	.58	.78	.83	.82	.64	.30	.40	.16	.26
(120-210)	±.37	±.22	±.18	±.14	±.21	±.23	±.39	±.20	±.27
MMN	.52	.65	.66	.70	.68	.46	.45	.55	.57
(500-700)	±.37	±.32	±.36	±.21	±.27	±.37	±.23	±.23	±.34

Table 4

Summary of Button Press Reaction Times (ms) in Tests 4 and 5 to Within-category (W) and Across-category (A) deviants

Test	Stimulus	Mean	SD	Range	Median	Mode
4a	1 (W)	552	155	245 - 950	540	500
	7 (A)	537	95	355 - 1065	510	505
4b	9 (W)	531	130	170 - 935	550	585
	3 (A)	524	95	345 - 775	520	555
5a	1 (W)	667	160	395 - 1050	665	685
	7 (A)	619	140	360 - 1040	605	610
5b	9 (W)	680	175	390 - 1085	660	680
	3 (A)	643	135	425 - 975	625	740

In Test 4 Ss identified stimuli as "ba" and "da". In Test 5 Ss were asked to identify only the infrequent "ba" and "da" and not to respond to the frequent standard stimulus. Tests 4a and 5a used stimuli 1, 4 and 7. Tests 4b and 5b used stimuli 3, 6 and 9.

Figure 1. A sound spectrogram of /ba/ and /da/ showing how the acoustic characteristics of these speech sounds change over time. The sounds were produced by a male speaker. The vertical lines are produced when the vocal cords vibrate (at a rate of about 110 Hz). The darkness of the trace indicates the energy level of the spectral components. A spectral peak, such as one made by a formant, produces a dark area at a point along the vertical axis corresponding to the formant's frequency. When the formant frequency is unaltered during vowel production this produces a dark horizontal bar. When the formant frequency is increased, the dark bar bends upward; when it is decreased the bar bend downward. The second formant transition is the initial change in frequency in the second lowest band of energy. For /ba/ the second formant transition shows as an upward bend in the formant. For /da/ the second formant transition shows as a downward bend in the formant.

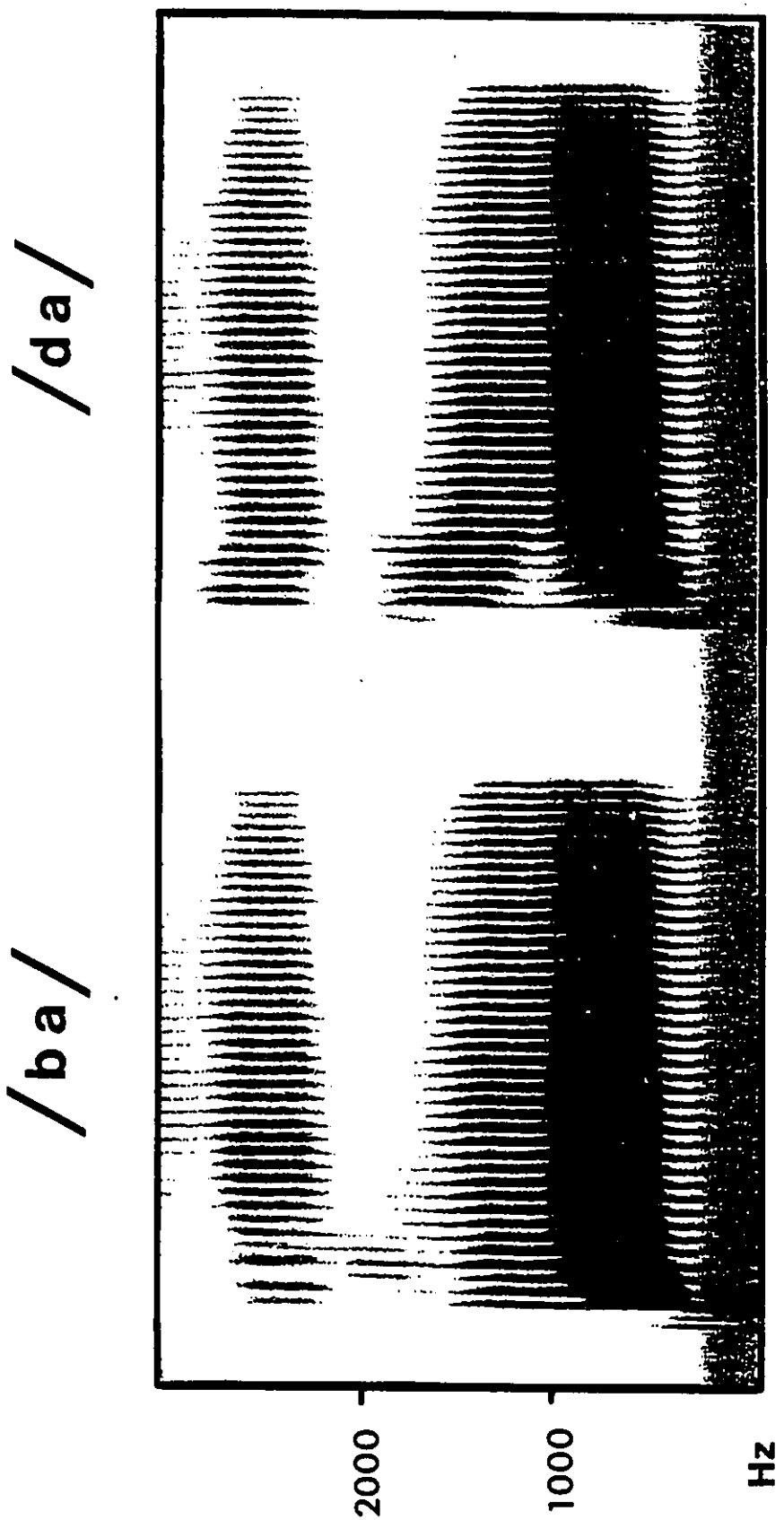

/ba/
/da/
2000
1000
Hz
100 ms

Figure 2. Sound spectrogram of the /ba/ to /da/ continuum, showing the graded changes in the second formant transition that result in listener's perception of speech sounds /ba/ or /da/. A male speaker produced the sounds /ba/ and /da/ and then a computer constructed the intervening sounds. Stimuli are number 1 to 9 for the experiment.

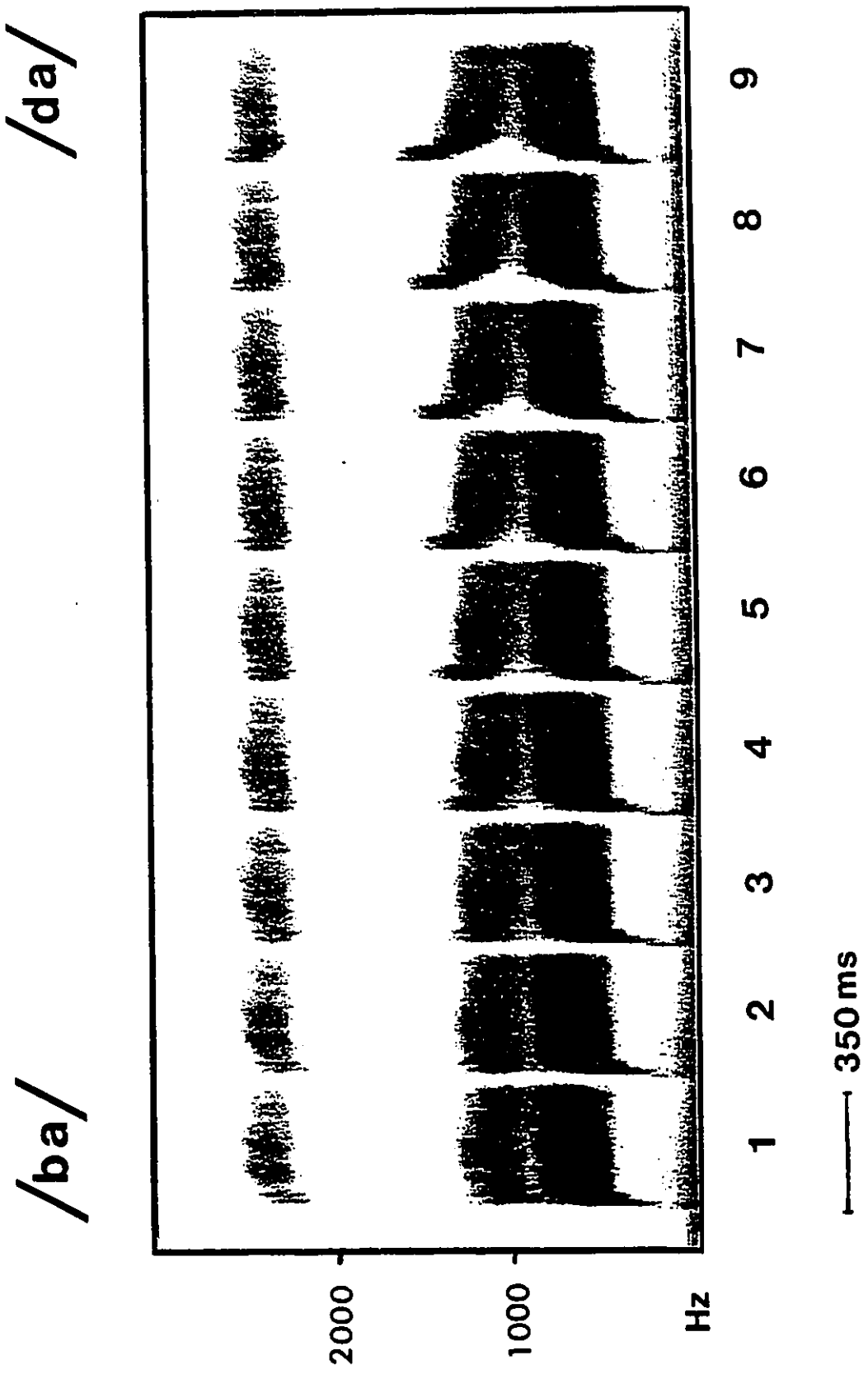

/da/
/ba/
1
2
3
4
5
6
7
8
9
2000
1000
Hz
350 ms

Figure 3. Identification functions for the /ba/ to /da/ continuum. Stimuli 1 to 9 vary along the continuum of the second formant transition shown in Figure 1. The figure shows the average percentage of times 5 subjects identified a given stimulus as /ba/. When stimuli (N=200) were randomly presented at equal probabilities (squares) stimuli 1 to 4 were perceived as "ba" and stimuli 6 to 9 were perceived as "da". The cross-over from the /ba/ to the /da/ category at stimulus 5 is known as the phoneme boundary. A higher probablity of presentation of a stimulus from the /ba/ (circles) or /da/ (triangles) end-point of the continuum caused shifts in the boundary towards that end of the continuum.

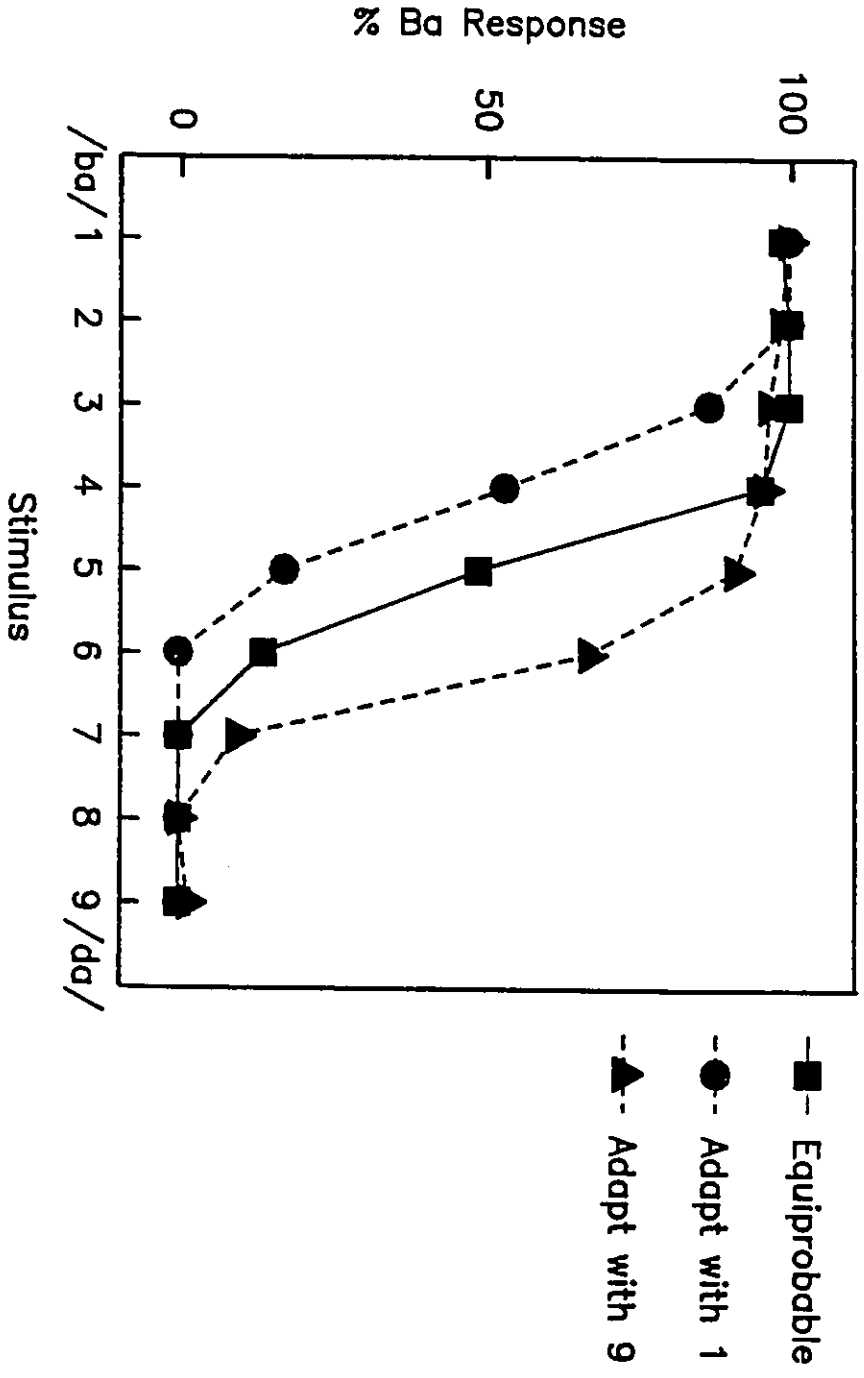

% Ba Response
0
50
100
/ba/ 1 2 3 4 5 6 7 8 9 /da/
Stimulus
Equiprobable
Adapt with 1
Adapt with 9

Figure 4. ERP responses to speech stimuli in the /ba/-standard and /da/-standard conditions. These tracings are average waveforms from 10 Ss recorded from Cz using a sternovertebral reference. Negativity at the vertex is plotted upward. The thick continuous line shows the P1-N1-P2-SP response to standard stimuli from the /ba/ and /da/-standard conditions. The dotted lines on these tracings are the responses to across-category or within-category deviant stimuli. Below these waveforms are shown the difference waveforms, the MMNs, obtained by subtracting the response to standard stimuli from the response to deviant stimuli. The duration of the stimulus is shown by the rectangle on the time scale.

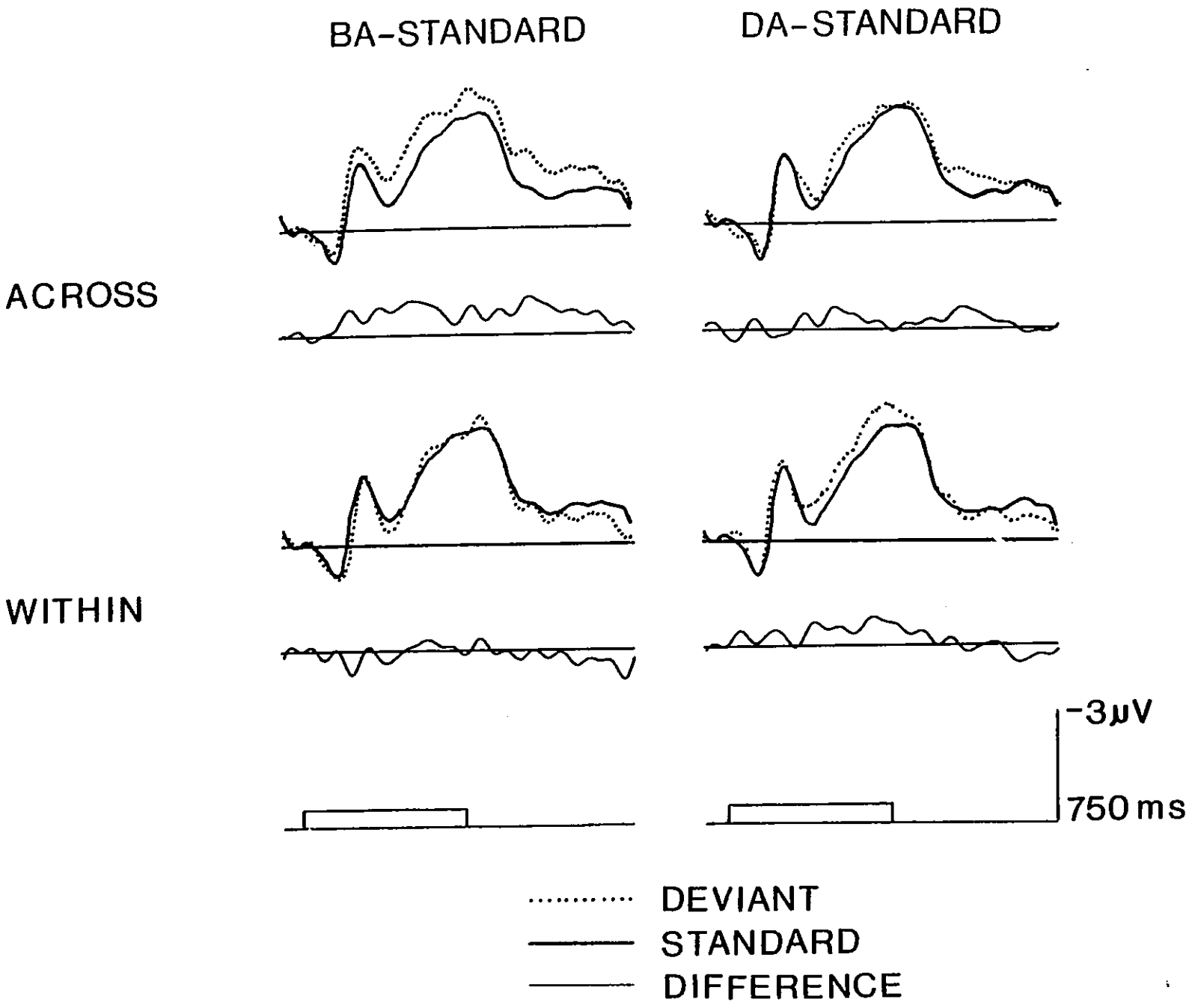

BA-STANDARD
DA-STANDARD
ACROSS
WITHIN
-3μV
750 ms
DEVIANT
STANDARD
DIFFERENCE

Figure 5. Scalp distribution of the MMN to the across-category deviant in the /ba/-standard condition. These tracings represent the average waveforms from 10 subjects recorded with the same electrode montage as in Figure 5.

7 - 4
ACROSS-CATEGORY MMN

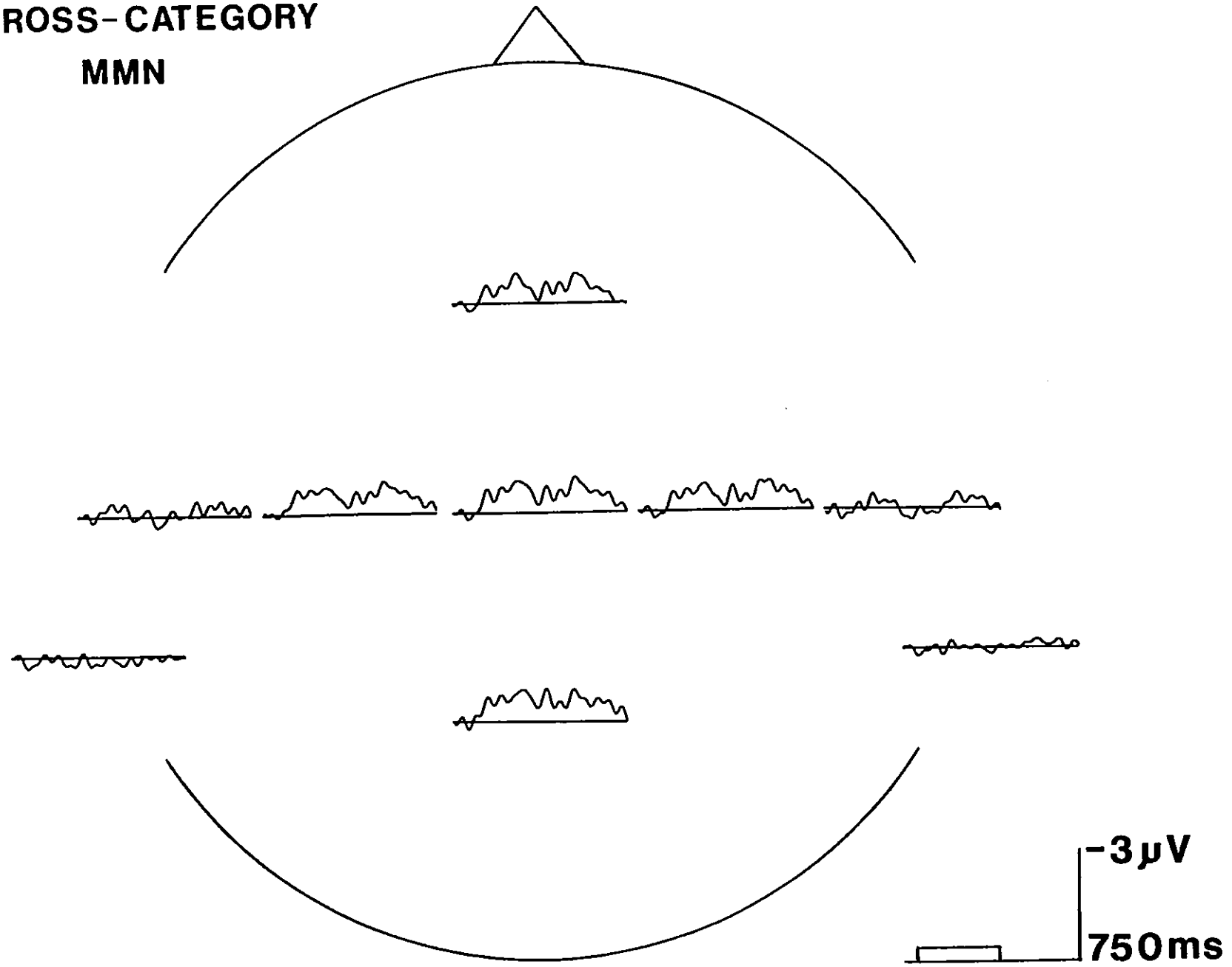


Figure 6. Scalp distribution of the P1-N1-P2-SP response to standard stimuli in the /ba/-standard condition. The data represent the average waveforms from 10 subjects, recorded at Fz, Cz, Pz, C3, C4, T3, T4, M1, and M2 using a sternovertebral reference with negativity at the scalp shown as an upward deflection.

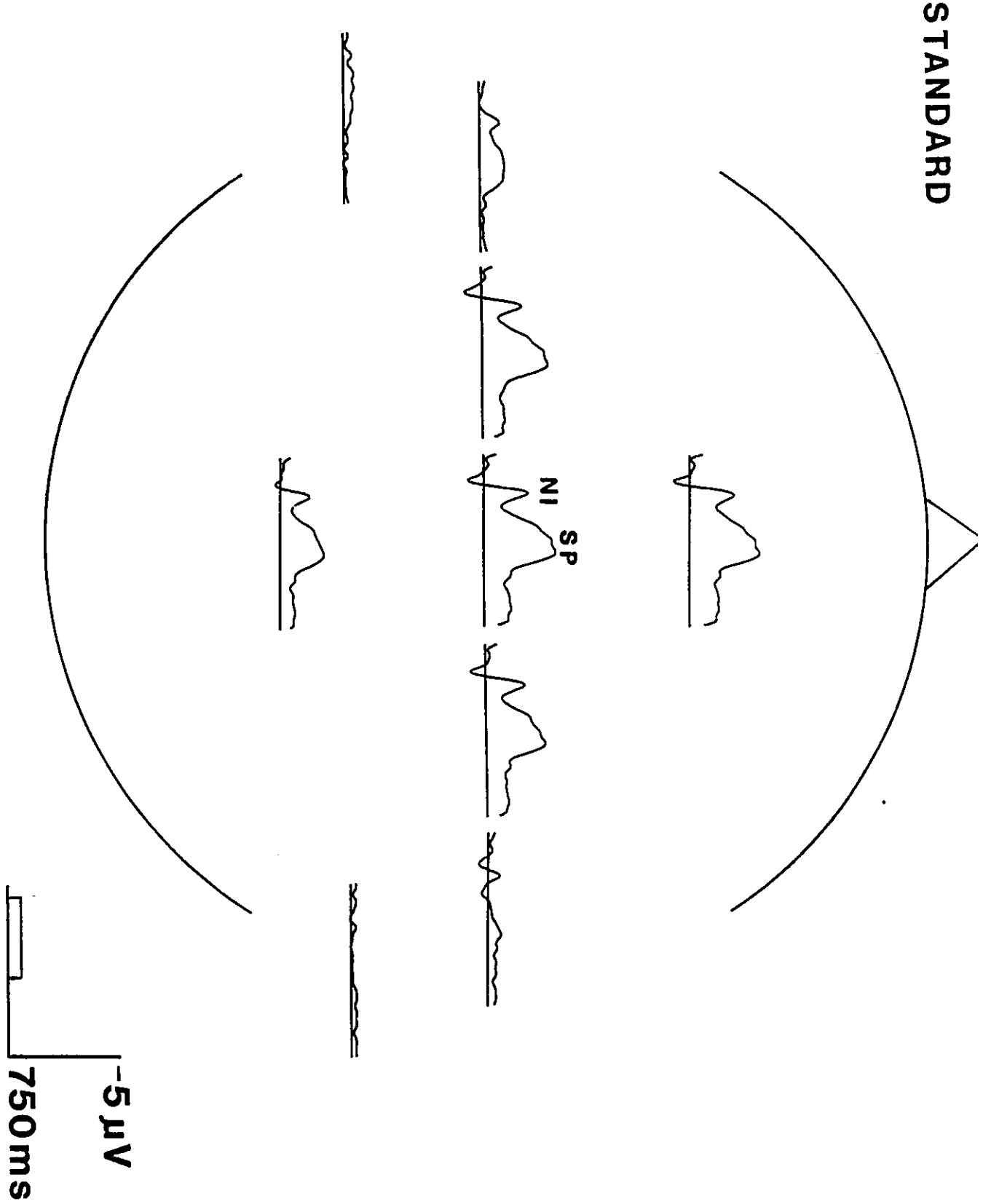

STANDARD
N1
SP
-5 μV
750ms

Figure 7. Adaptation with stimulus 4 or 6. Identification functions for the /ba/ to /da/ continuum, showing the average percentage of times 5 subjects identified a stimulus as /ba/, when all stimuli were randomly presented at equal probabilities (squares), when stimulus 4 was presented 60% of the time (circles), and when stimulus 6 was presented 60% of the time (triangles). This figure illustrates that there were no significant shifts in the identification functions with repetition of stimulus 4 or 6.

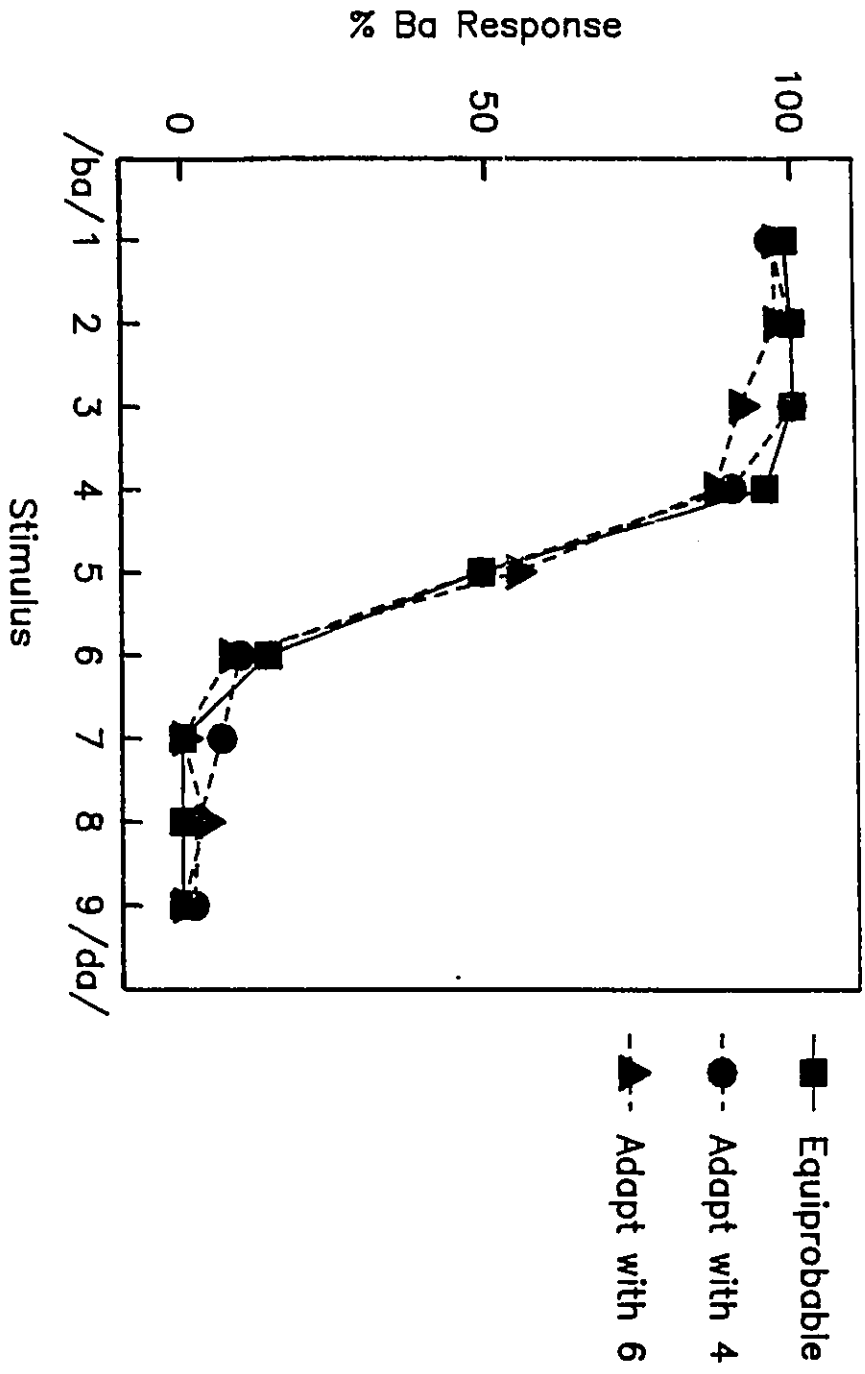

% Ba Response
0
50
100
/ba/ 1 2 3 4 5 6 7 8 9 /da/
Stimulus
Equiprobable
Adapt with 4
Adapt with 6

Figure 8. Identification functions for the stimuli used in the ERP experiment. This figure shows the average percentage of times 5 subjects identified stimuli as /ba/. The triangles mark responses to random presentation of stimuli 1, 4, and 7 at 10, 80 and 10% probabilities (N=100). The squares show responses to presentation of stimuli 3, 6 and 9 at 10, 80 and 10% probabilities. The dotted line represents the identification function when the stimuli were presented equiprobably (Figure 7). Stimulus 4, the /ba/-standard in the ERP experiment was categorized as /ba/ 78 % (± 26) of the time. Stimulus 6, the /da/-standard in the ERP experiment was categorized as /da/ only 66% (±36) of the time.

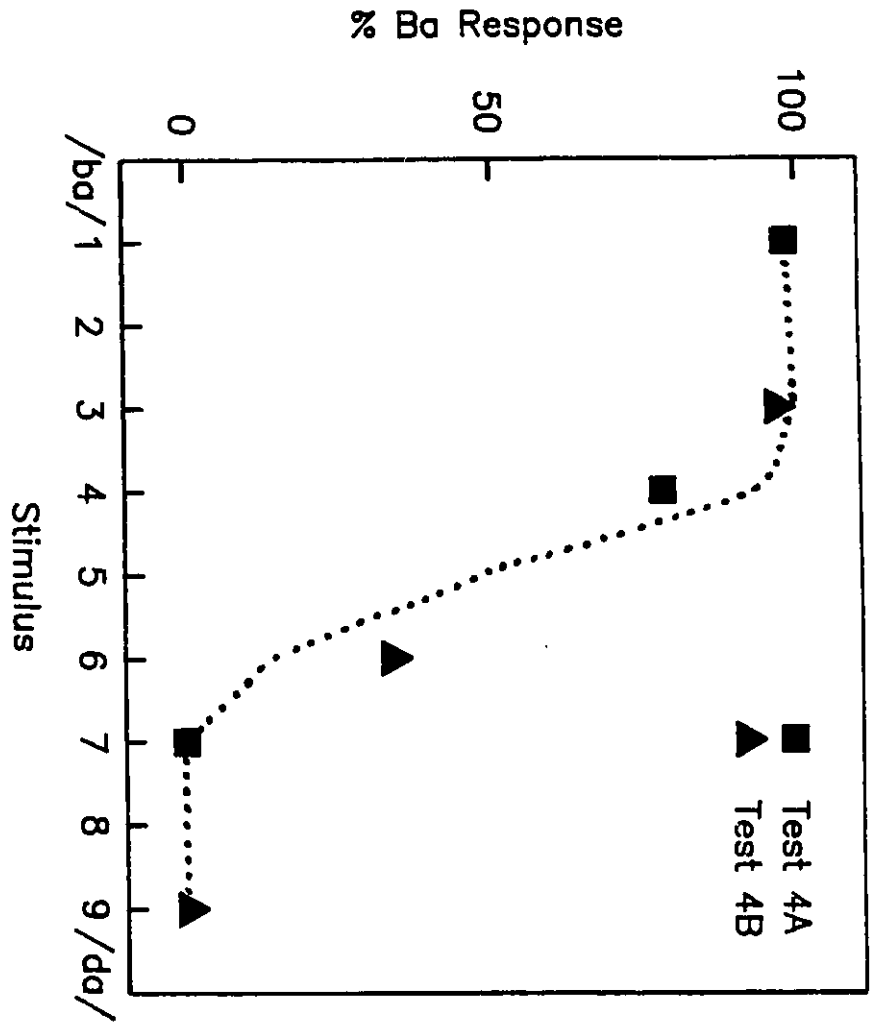

% Ba Response
0
50
100
/ba/ 1
2
3
4
5
6
7
8
9 /da/
Stimulus
Test 4A
Test 4B

Figure 9. Reaction times for identification of stimuli along the /ba/ to /da/ continuum. Stimuli (N=200) were randomly presented at equal probabilities. The mean reaction times are based on the median reaction times of 5 subjects. The circles indicate stimuli identified as /ba/, and the squares indicate stimuli identified as /da/. This figure illustrates the peak in RT around the phoneme boundary.

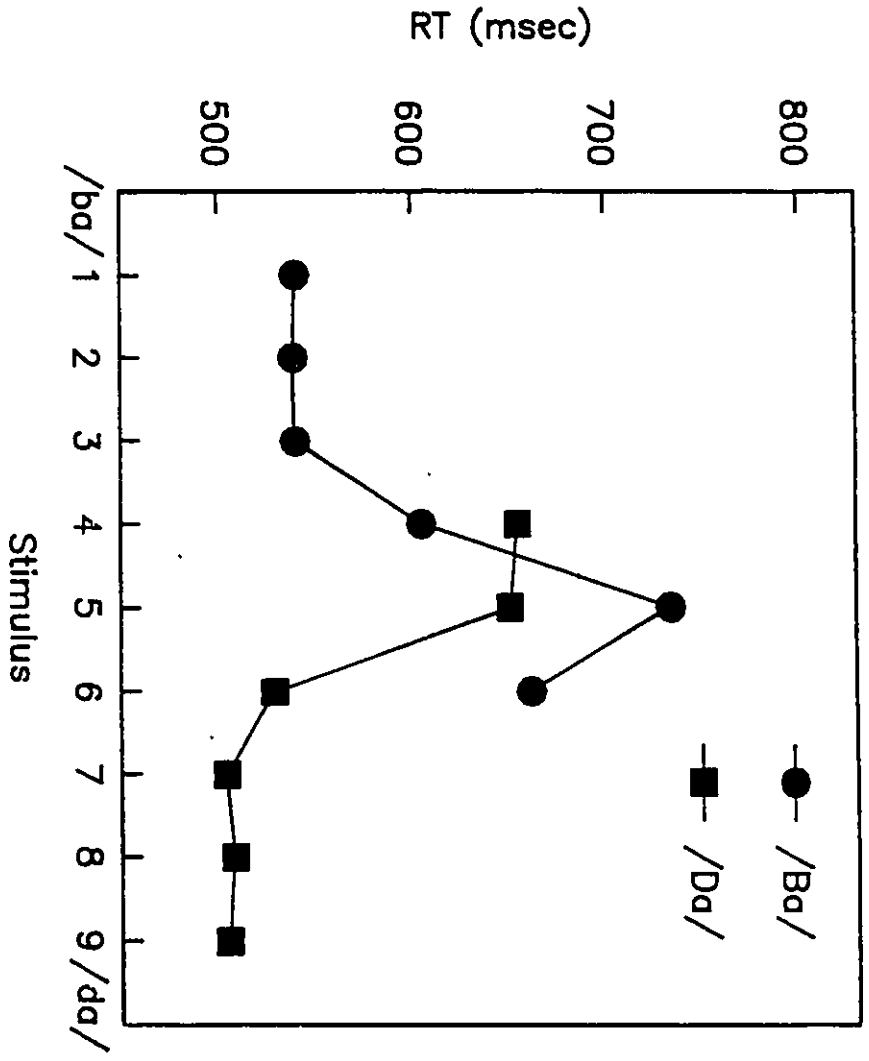
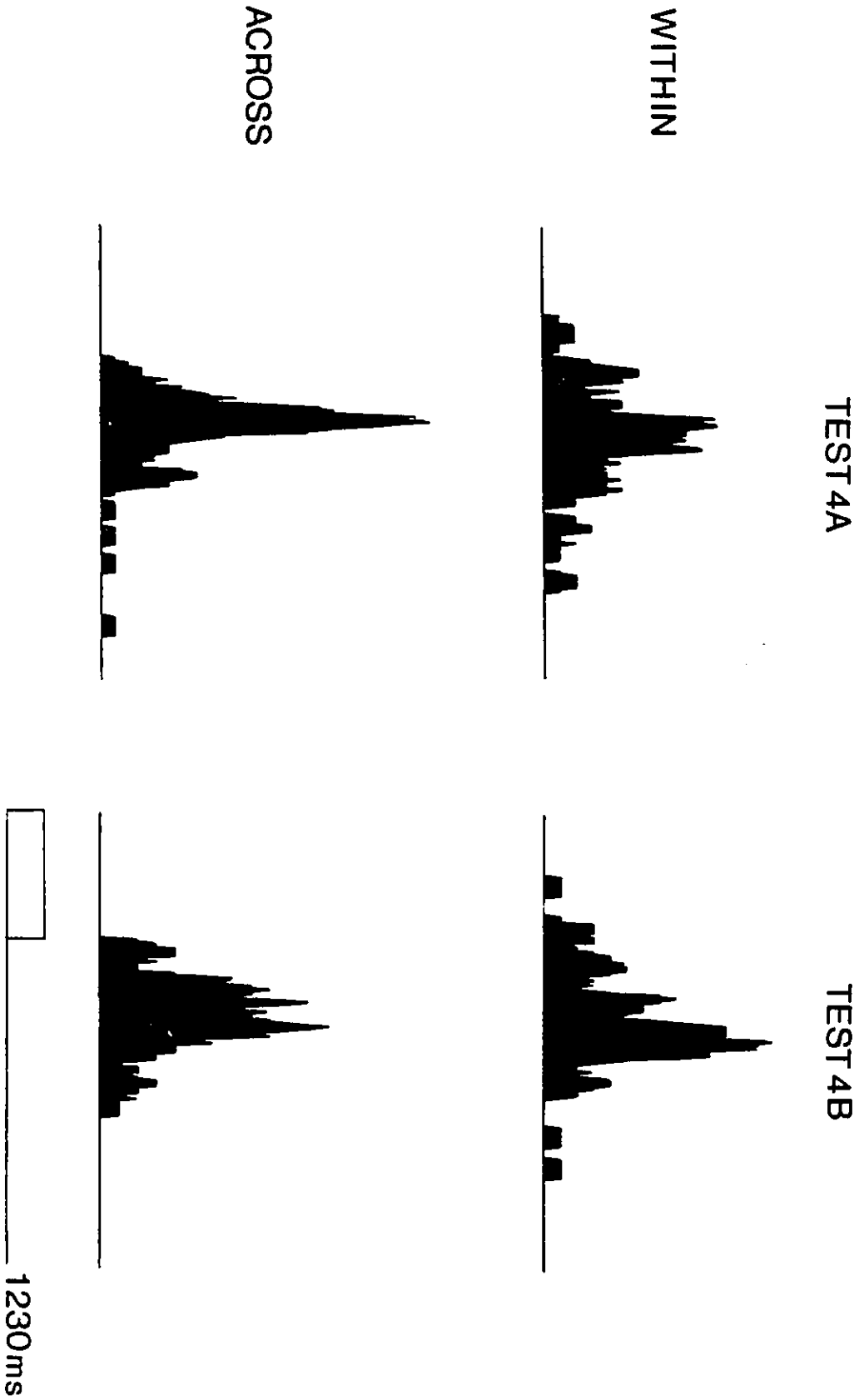

RT (msec)
500
600
700
800
Stimulus
/ba/ 1
2
3
4
5
6
7
8
9 /da/
/Da/
/Ba/

Figure 10. Button press reaction time histograms from Test 4. These are the grand mean histogram for 5 subjects. In Test 4a, Ss heard stimuli 1, 4 and 7. The within-category responses were the /ba/ respsonses to stimulus 1 and the across-category responses were the /da/ responses to stimulus 7. In Test 4b, Ss heard stimuli 3, 6 and 9. The within-category responses were /da/ responses to stimulus 9 and the across-category repsonses were /ba/ responses to stimulus 3. The figure shows that subjects tended to respond faster to across-category than to within-category stimuli.


WITHIN
ACROSS
TEST 4A
TEST 4B
1230ms

SUMMARY OF THESIS

This thesis investigated how the human auditory system processes the brief frequency changes that occur in speech sounds. Human speech contains frequency transitions during the production of stop-consonants that signal changes in meaning. The first study therefore examined the processing of frequency changes using frequency-modulated tones as stimuli and recording steady state and transient responses. The first study thus consisted of three experiments evaluating ERPs to non-speech frequency changes. The experiments evaluated steady state and transient auditory evoked potentials to tones that were sinusoidally modulated in frequency and to tones that alternated between two frequencies with a linear ramp. The linear ramp was equal in duration to frequency transitions occuring in speech and the tones were presented at modulation rates typical of average syllable production. The steady state responses to sinusoidal FM were small and difficult to record at both the first and second harmonics. Ramp FM evoked larger and more consistent second harmonic steady state responses than the sinusoidal FM. Only the ramp FM stimuli elicited transient EPs and these only at low modulation rates. This is in agreement with formulations by Clynes (1969) and Kohn et al. (1978) that in order to elicit an N1-P2, there must be a constant frequency portion preceding a ramp change. These responses were larger to upward ramps than to downward ramps suggesting that FM neurons specifically activated by upward changes

in frequency were contributing to the response. The response to two simultaneously presented ramp FM tones differed from the sum of responses to the individual tones indicating some interaction in the processing of the two stimuli.

Since the first study found that the largest and most reliable responses were to simultaneously presented ramp frequency changes, the second study used speech sounds containing discrete frequency changes. In the second study of the thesis computer-modified speech sounds from the /ba/ to /da/ continuum were presented to reading subjects as a train of standard speech sounds interspersed with two types of infrequent deviant speech sounds. One deviant stimulus lay within the same category as the standard and the other lay across the categorical boundary from the standard, but both were acoustically equidistant from the standard in terms of the second formant transition. When the standard stimulus was drawn from the /da/ end of the continuum, the MMN to across-category deviants was not larger than the MMN to within-category deviants. Grand mean waveforms suggested that both deviant stimuli elicited small MMNs. Although this may indicate processing along an acoustic continuum, the results of the psychophysical tests suggest that the standard stimulus in this condition was too close to the category boundary for the deviants to evoke a consistent categorical mismatch. When the standard stimulus was drawn from the /ba/ end of the continuum, the across-category deviants elicited a clear mismatch negativity (MMN) in the auditory event-related potential whereas the within-category

deviants did not. This MMN began at about 60-120 ms after stimulus onset and was present for several hundred ms. In addition, regardless of phonemic condition, MMN amplitude had a positive correlation with categorical identification. These results suggest that categorical processing of speech sounds occurs independently of attention at an early echoic memory stage. The fact that this processing is categorical lends support to the idea that speech perception is unique and differs from the processing of other auditory stimuli.

APPENDIX 1: HAND USAGE QUESTIONNAIRE

Please indicate below which hand you ordinarily use for each activity.

With which hand do you:

1.	draw?	1. Left 2. Right 3. Either
2.	write?	1. Left 2. Right 3. Either
3.	use a bottle opener?	1. Left 2. Right 3. Either
4.	throw a snowball to hit a tree?	1. Left 2. Right 3. Either
5.	use a hammer?	1. Left 2. Right 3. Either
6.	use a toothbrush?	1. Left 2. Right 3. Either
7.	use a screwdriver?	1. Left 2. Right 3. Either
8.	use an eraser on paper?	1. Left 2. Right 3. Either
9.	use a tennis racket?	1. Left 2. Right 3. Either
10.	use scissors?	1. Left 2. Right 3. Either
11.	hold a match when striking it?	1. Left 2. Right 3. Either
12.	stir a can of paint?	1. Left 2. Right 3. Either
13.	On which shoulder do you rest a bat before swinging?	1. Left 2. Right 3. Either

Scoring is as follows:

Left = 3 point

Right = 1 point

Either = 2 points

To be termed Right-handed a subject must have a score between 13 and to 17.

APPENDIX 2

HÔPITAL GÉNÉRAL D'OTTAWA
OTTAWA GENERAL HOSPITAL
501 CHEMIN SMYTH ROAD, OTTAWA K1H 8L6

CONSENTEMENT À DES RECHERCHES MÉDICALES
CONSENT FOR MEDICAL RESEARCH

Nom de l'étude-Name of research project

_____

Chercheur-Researcher

_____

Je consens à la participation de:

_____
(patient)

à l'étude précitée et j'autorise par la présente le(s) médecin(s) et chercheur(s) à procéder aux examens et/ou à dispenser les traitements suivants:

I consent to the participation of:

_____
(patient)

in the above study and hereby authorize the physician(s) and investigator(s) to proceed with the following examinations and/or to administer these treatments:

_____

_____

J'ai reçu une explication détaillée de tous les effets secondaires et des risques connus ayant trait aux examens et aux traitements.

J'ai également reçu une description de tous les avantages à attendre de ces examens et de ces traitements. On m'a fait connaître d'autres formes d'examens et de traitements.

J'ai eu l'occasion de poser des questions au sujet de ces examens et de ces traitements et on y a bien répondu.

On m'a dit que je pouvais retirer mon consentement et suspendre ma participation à l'étude à n'importe quel moment et pour quelque motif que ce soit.

En toute connaissance de cause, je consens volontairement à participer à cette étude.

I have had a detailed description of all the known side effects and risks related to the examinations and treatments.

I have also received a description of any benefits that may be expected from these examinations and treatments. Alternative forms of examinations and treatments have been disclosed to me.

I have been given an opportunity to ask questions concerning the examinations and treatments involved and the questions which I have asked have been adequately answered.

I have been told that I can withdraw my consent and stop my participation in the study at any time and for any reason.

With full knowledge of this, I voluntarily consent to participate in the study.

Nom-Name _____ Signature _____ Lien de parenté / Relationship _____

PATIENT OU PERSONNE LÉGALEMENT RESPONSABLE / PATIENT OR PERSON LEGALLY RESPONSIBLE

1. TÉMOIN/WITNESS: _____ _____ DATE _____
Nom-Name Signature

2. TÉMOIN/WITNESS _____ _____ DATE _____
Nom-Name Signature

Evoked Potential Studies of Suprathreshold Hearing

Patient Information Sheet

During this experiment we shall record the electrical responses of your brain to sounds. Our goal is to see whether we can use these recordings to measure how well you distinguish between different sounds.

Prior to the recording, several electrodes will be attached to your scalp. The skin is rubbed slightly and the electrodes are attached with a sticky paste. During the recording, various sounds will be presented through earphones or through speakers. A computer will analyze the electrical responses recorded from your scalp. You may also be asked to press buttons in response to certain sounds.

The recordings that we take are similar to those of routine clinical "electroencephalography" (brain wave-recording). The electrodes are sometimes mildly uncomfortable. The experiment can be boring because the sounds have to be repeated many times.

The results of these experiments will not provide any direct benefit to you. However, these results may help us in designing new tests of hearing that can be used to evaluate the hearing abilities of young children or the functioning of hearing aids.

APPENDIX 3: Identification Reaction Times (ms)
(based on median RTs from each subject)

TEST 1: EQUIPROBABLE STIMULI

Stimuli Identified as /ba/	1	2	3	4	5	6	7	8	9
Mean	540	539	540	605	734	662			
S.D.	132	126	112	120	81	137			
Stimuli identified as /da/									
Mean				655	651	529	504	508	505
S.D.				129	150	116	153	161	119

ADAPTATION WITH STIMULUS 1

Stimuli Identified as /ba/	1	2	3	4	5	6	7	8	9
Mean	487	491	499	580	"21				
S.D.	97	90	89	117	389				
Stimuli Identified as /da/									
Mean			569	523	510	487	468	464	485
S.D.			350	33	53	78	87	101	84

ADAPTATION WITH STIMULUS 9

Stimuli Identified as /ba/	1	2	3	4	5	6	7	8	9
Mean	492	517	515	509	577	622	757		
S.D.	115	101	77	112	79	56	141		
Stimuli Identified as /da/									
Mean		345	325	522	528	556	446	455	450
S.D.		0	0	465	193	202	102	93	95

ADAPTATION WITH STIMULUS 4

Stimuli Identified as /ba/	1	2	3	4	5	6	7	8	9
Mean	544	529	535	566	634	567	732	365	525
S.D.	107	94	100	123	141	195	70	0	0
Stimuli Identified as /da/									
Mean				608	668	560	495	519	491
S.D.				164	81	107	106	84	99

ADAPTATION WITH STIMULUS 6

Stimuli Identified as /ba/	1	2	3	4	5	6	7	8	9
Mean	513	508	528	563	661	691		710	
S.D.	60	73	61	44	118	106		0	
Stimuli Identified as /da/									
Mean		715	483	643	461	482	454	469	448
S.D.		0	168	256	66	84	67	89	105

STIMULI 1,4,7 FROM TEST 4A (identify "ba" and "da")

Stimuli Identified as /ba/	1	4	7
Mean	537	558	
S.D.	59	70	
Stimuli Identified as /da/			
Mean		709	503
S.D.		88	45

STIMULI 3,6,9 FROM TEST 4B (identify "ba" and "da")

Stimuli Identified as /ba/	3	6	9
Mean	533	681	
S.D.	77	25	
Stimuli Identified as /da/			
Mean	590	598	500
S.D.	0	184	114

STIMULI 1 AND 7 FROM TEST 5A (identify only deviant stimuli)

	1	7
Mean	646	597
S.D.	100	61

STIMULI 3 AND 9 FROM TEST 5B (identify only deviant stimuli)

	3	9
Mean	630	673
S.D.	93	102