RESEARCH ON SPEECH PERCEPTION

Technical Report No 2.

August 15, 1976

Department of Psychology

Indiana University

Bloomington, Indiana 47401

Supported by:

Department of Health, Education and Welfare
U.S. Public Health Service

National Institute of Mental Health
Grant No. MH-24027-03

and

National Institutes of Health
Grant No. NS-12179-02

THE STRUCTURE AND FLOW OF INFORMATION IN SPEECH PERCEPTION: EVIDENCE FROM SELECTIVE ADAPTATION OF STOP CONSONANTS

by

James R. Sawusch

Submitted to the Faculty of the Graduate School

in partial fulfillment of the requirements

for the degree Doctor of Philosophy

in the Department of Psychology

Indiana University

August, 1976

The Structure and Flow of Information in Speech Perception:

Evidence from Selective Adaptation of Stop Consonants

James R. Sawusch

Abstract

Recent experiments on selective adaptation with stop consonants have shown that this paradigm affects the relatively early stages of speech recognition. The experiments reported here dealt with two questions: How many levels of perceptual processing are affected and what some of the characteristics of these level(s) are.

Experiments 1 and 2 involved the systematic manipulation of the frequency (spectral) overlap between adapting and test stimuli. Adapting stimuli were generated that had formant frequencies which were separated from those in the test series by more than one critical bandwidth. These high adaptors produced consistent category boundary shifts in the low test series. When the adaptor was presented to one ear and test syllables to the other ear, 100% interaural transfer was found. In contrast, when the adaptor was an end-point of the test series only about 50% interaural transfer was found. These results indicate that two levels of processing are affected by selective adaptation: A frequency specific peripheral level and a central integrative level which does not require spectral overlap.

Experiments 3 and 4 further explored the characteristics of the integrative level. The stimuli from Experiments 1 and 2 were used in Experiment 3 along with a manipulation of the relative intensity of the adapting stimulus. This intensity manipulation was found to influence the category boundary shift at the integrative level with a louder

adapter causing a greater shift. This result indicates that the integrative level probably represents abstract auditory, as opposed to phonetic processing. Experiment 4 employed CV adaptors with different vowels and found adaptation for all of the adaptors, indicating that the integrative level is relatively insensitive to vowel context.

Experiments 5 and 6 focused on the adaptation of a CV series with VC and VC like stimuli. These experiments explored the sensitivity of processing at the peripheral and integrative levels to the position of the consonant within the syllable. No adaptation effect was found for a VC on a CV series (and vice versa) when the VC syllable contained transitions that were identical to one end of the CV series. However, when the first formant of the VC adaptor was changed an adaptation effect was found. These results indicate that first formant information is involved in place perception and that the adaptation effects at peripheral and integrative levels may cancel each other when a VC is used as an adaptor on a CV series. Other results from Experiment 6 indicate that a spectral change in frequency is necessary and that onset (offset) frequency alone will not yield selective adaptation effects.

On the basis of these results and previous investigations by other researchers an information processing model of the early stages of speech perception was implemented as a computer simulation. This model included the peripheral frequency specific and central integrative levels outlined above. The results of the simulation were compared to unadapted identification functions and adaptation results for a number of CV and VC series. The overall fit between observed and predicted functions was quite good. These simulation results offer strong support for the proposed two level model of selective adaptation and suggest that the simulation should be extended to cover features other than place in stop consonants.

Table of Contents

Page

Page

Acknowledgments

I would like to thank my adviser and chairman Professor David Pisoni for his patience and insightful comments on all phases of this dissertation. He introduced me to the area of speech perception and many hours have been spent with him in fruitful discussion. I am also grateful to the other members of my committee: Professor Richard Shiffrin, Professor Donald Robinson and Professor Stuart Shapiro for their advice, criticisms, encouragement and suggestions.

Special thanks are also due to the following people who helped in many ways: Jerry Forshee for his programming and help in setting up the OVE synthesizer and experimental control programs; Roger Ford for his help in running subjects; Professor Dennis Klatt of MIT for the use of his speech synthesizer in generating certain stimuli; Professor Kenneth Stevens of MIT for suggestions and comments on several experiments; and Professor Frank Restle for his advice and encouragement during this investigation.

Finally, I would like to thank Judy Hupp for the typing of the final draft of this manuscript and my wife Ann for her support, encouragement, suggestions and invaluable aid as a typist.

The research in this dissertation was supported in part by National Institute of Mental Health Grant No. MH-24027, National Institutes of Health Grant No. NS-12179, and Public Health Service Training Grant T01-MH-11219 to Indiana University.

List of Tables

List of Figures

Page

Page

CHAPTER I

Introduction

As a language user, every human being has the dual task of perception and production. The purpose of language is communication, to convey a message to another individual. Speech production may be thought of as the processes that convert meaning into sound, speech perception as the converse process of converting sound into meaning. However, despite the extensive interest in speech production and perception, relatively little is known about the actual functioning of these mechanisms. What is known suggests that speech perception and production involve a number of interrelated processes and a complex restructuring of the message. Specifically, the perceptual process of converting sound into meaning seems to require extensive recoding of the acoustic signal as well as the use of linguistic information not present in the acoustic waveform (cf. Chomsky & Halle, 1968).

The study of phonetics, the linguistic description of the sounds of language, provided the initial basis upon which much of the early work in speech perception was built. Linguists have described the sound structure of languages in terms of a set of 40 to 50 phonemes.[1] These units form a minimal description of the sounds of speech as they relate to the meaning of an utterance. Early work in speech analysis and synthesis began with this linguistic description and focused on the relationship between attributes of the speech signal and its phonetic representation. This early work involved the analysis of the speech waveform using the sound spectrograph, work on speech synthesis, and perceptual studies on intelligibility in an attempt to specify the minimal information necessary to cue a phonetic distinction. From this early work, a number of important issues in speech perception emerged: 1) the lack of invariance in the speech waveform; 2) the segmentation of the waveform into units of analysis; and 3) the role of various units of analysis in perception. These issues place certain constraints on models that could potentially describe the perception of speech. Since these issues have been central to most models of speech perception, they will be reviewed briefly here. Following the description of these issues several prominent models of speech perception will be discussed.

Issues in Speech Perception

Three basic and highly interrelated issues pertaining to speech perception may be identified. These issues deal with invariance, segmentation, and the units of perceptual analysis. The question of invariance came to light with the advent of the sound spectrograph (Koenig, Dunn & Lacy, 1946) and the first perceptual studies of speech using the pattern playback (Cooper, Delattre, Liberman, Borst & Gerstman, 1952). The basic problem was that the same acoustic cue could give rise to different percepts in different contexts and that different acoustic cues could give rise to the same percept in different contexts (see Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967, for a review). One classic example of the first of these two phenomena comes from an experiment on the unvoiced stop consonants [p, t, k] (Liberman, Delattre & Cooper, 1952). Bursts varying in frequency were added to the front of a series of English vowels. One particular burst, centered at 1400 Hz, gave rise to the perceived syllable [pi] when placed in front of the [i] vowel. However, in front of the vowel [a], this same burst was perceived as [ka]. Finally, in front of the vowel [u], the resulting percept was [pu]. Thus, one acoustic cue gave rise to different percepts, depending upon the vowel context.

The converse phenomenon, in which multiple cues give rise to the same percept, has also been observed. Liberman, Delattre, Cooper & Gerstman (1954) investigated the effects of initial formant transitions on the perception of consonants. In front of the vowel [i], a rising second formant was perceived as a "d." However, in front of the vowel [u], a falling second formant transition was necessary to hear a "d." Thus, in front of one vowel a rising transition yielded the same percept as a falling transition in front of a different vowel. A similar diversity of sound-to-phoneme relationships has been found in attempts to read sound spectrograms (Fant, 1962). In general, little one-to-one correspondence has been found between the acoustic cues in speech (or their visual analog, a sound spectrogram) and the phonetic representation of an utterance. The failure to find a simple one-to-one relationship between sound and phoneme is precisely what is meant by the term lack of invariance. This lack of invariance has not been solved by changing the unit of perceptual analysis either. The examples reviewed above all assume the phoneme as the basic unit of perceptual analysis. Using the syllable as the basic unit of perceptual analysis does resolve this problem but at the cost of treating [di] and [du] as separate entities. However, the problem of why the "d" in [di] sounds the same to a listener as the "d" in [du] still remains to be resolved.

There are phonemes which do have acoustic cues that are close to invariant. Some fricatives and affricates may be characterized by invariant intensity, frequency, or duration noise cues (Harris, 1958; Strevens, 1960). The problem of invariance is related to the diversity of cue to phoneme relationships. For some classes of sounds, such as the fricatives, the mapping from the acoustic signal to the phonetic representation is nearly one-to-one. For other classes, such as the stop consonants, a simple one-to-one mapping has not been found (although see Cole & Scott, 1974, for an alternative view). Most of the sounds of speech exhibit some context conditioned variability, although few fail to show the lack of invariance that characterizes the stop consonants. One reason for focusing on the invariance problem is that the listener hears the "d" in [di] and [du] as being the same. This is apparently accomplished in spite of the diversity of cues that characterize "d" in front of different vowels. The mapping of acoustic cue onto phonetic percept is therefore not a simple one-to-one relationship. Any proposed model of the speech perception system must somehow overcome this lack of invariance in order to retrieve the message from the acoustic signal.

The problem of segmenting the speech signal into units is intimately tied to the invariance question. Where invariant cues exist, the signal can also be segmented in the time domain. However, for most of the speech signal, segmentation according to acoustic criteria is exceedingly difficult. Speech is a continuously varying signal. Any particular stretch of this waveform carried information about a number of segments. This phenomena is generally referred to as contextual variability (see Fant, 1962; Liberman et al., 1967, for examples). The importance of this variability is in the processing that the perceptual system must go through to decode the message from the speech signal. Since the acoustic elements are not in a one-to-one linear mapping with phonemes or phonetic features, a complex restructuring process is involved in production and a complex decoding in perception.

In various attempts to solve the joint problems dealing with the lack of invariance and segmentation, different sized perceptual units have been proposed. The phonetic feature, phoneme and syllable have all been considered at one time or another by various investigators. Most of the work on invariance and segmentation cited above has been based on the perception of phoneme or phonetic segments. Evidence from a number of different tasks used to study various phenomena all indicate the psychological reality of the phoneme. Conrad (1964) and later Crowder (1971) found that the errors that subjects made in immediate memory tasks could easily be described as substitutions, deletions and additions of phoneme sized units. Miller & Nicely (1955) studied the perceptual confusions of stop consonant-vowel (CV) syllables. The errors that they found group well according to the similarity of the phonetic features in the syllables. Sheppard (1972) reanalyzed the Miller & Nicely (1955) results using a factor analytic technique and the results again fit well with a phonetic feature description.

Recently, the syllable has been proposed as the basic unit of perceptual analysis (Savin & Bever, 1970; Massaro, 1972). However, the description of speech in terms of syllabic units suffers from the same problems that were encountered for phonemes and phonetic features. The coarticulation phenomena that carry attributes of one phonetic feature or phoneme onto the succeeding phoneme are found for syllables also (Öhman, 1966). Coarticulation refers to the influence of the muscle movements necessary to produce one sound onto preceeding and succeeding muscle movements and their resulting acoustic manifestations. Thus, the production of a stop consonant is conditioned or influenced by the production of the adjacent vowel. The acoustic consequence of coarticulation is that one sound segment will carry information about a number of phonemes (or syllables). Coarticulation effects obscure the boundaries between all three potential units of analysis in speech: phonetic features, phonemes and syllables. As a result, syllabic units are very difficult to segment by acoustically defined criteria and show the same context conditioned variability (i.e., lack of invariance) exhibited by phonemes.

One resolution to the question of units of analysis has been offered by Studdert-Kennedy (1976). He has argued that the information in the speech signal that is necessary to identify a phoneme is distributed over an entire syllable such that the syllable is the carrier of phonetic information. The phoneme and the syllable are both basic perceptual units. Instead of asking what the basic unit is in perception, the proper question is what role each unit plays and how processing of that unit is carried out by the perceptual system. Models of the perceptual system must detail how the segmentation and invariance issues are resolved and what role the various units of analysis play in perceptual processing. Some of the models that will be reviewed in the next section have not been explicit on all of these points. However, their approaches to these issues of invariance, segmentation and units of analysis have provided a basic framework within which more recent models have often been formulated. A brief description of these models will provide the background for a more detailed description of an information processing model of speech perception.

Theories of Speech Perception

Various classes of theories have been proposed to deal with the issues of invariance, segmentation, and units of analysis. They may be broadly classed into those involving active mediation by higher level processes (active models) and those involving a straightforward analysis of the stimulus with no mediation by higher levels (passive models). This distinction, first made by MacKay (1956) has been primary to most of the models proposed for speech perception, especially phoneme perception.

Active models: Motor Theory. The motor theory of speech perception (Liberman, Cooper, Harris & MacNeilage, 1962; Liberman, Cooper, Harris, MacNeilage & Studdert-Kennedy, 1967) has played a central role in theoretical discussions of speech perception. Its basic tenet is that people perceive speech by reference to the motor commands necessary to produce it. However, strong empirical support for this assumption has not been forthcoming. Categorical perception, typically cited as support for the motor theory, has been adequately explained on other grounds (Pisoni, 1971, 1973; Pisoni & Lazarus, 1974). The appealing aspect of motor theory is that it resolves the invariance problem by reference to invariant articulatory commands (or their neural equivalent). However, evidence supporting the invariance of motor commands in speech production has not been found either (MacNeilage, 1970). The one argument left in favor of motor theory is that of parsimony. The processes of speech production and perception are both handled by one integrated mechanism. Although the parsimony argument is not trivial, it should not outweigh any strong empirical support that may be found for another model.

Analysis-by-Synthesis. Stevens (1960; Stevens & Halle, 1967) has proposed a more detailed, active theory of speech perception known as analysis-by-synthesis. In this model, the speech signal undergoes a preliminary form of processing. Any invariant features present are extracted at this point. However, since much of the signal does not possess these invariant characteristics, control is next passed to an active component. Here, a generative process makes use of contextual information and any invariant attributes found to construct a featural description of the input. This description is then converted by rule to an equivalent auditory pattern which is matched against the original input. If the match is sufficient, the message is considered decoded. If the mismatch between input and the generated pattern is unacceptable, a new featural description is generated and the comparison process is repeated.

Although this model is more detailed than motor theory, it suffers from the same lack of empirical support that motor theory suffers from. The problem is that although analysis-by-synthesis is far more specific than motor theory, it is still not precise enough to make empirically testable predictions. Without a number of additional assumptions, this model cannot account for any of the basic sets of data in speech perception, such as that from experiments on categorical perception, recognition masking or selective adaptation. Analysis-by-synthesis is simply an outline for the perceptual system. However, without specification of the details of operation at the various stages of feature extraction, generation, comparison, and the role of memory in this system, predictions about empirical data cannot be made.

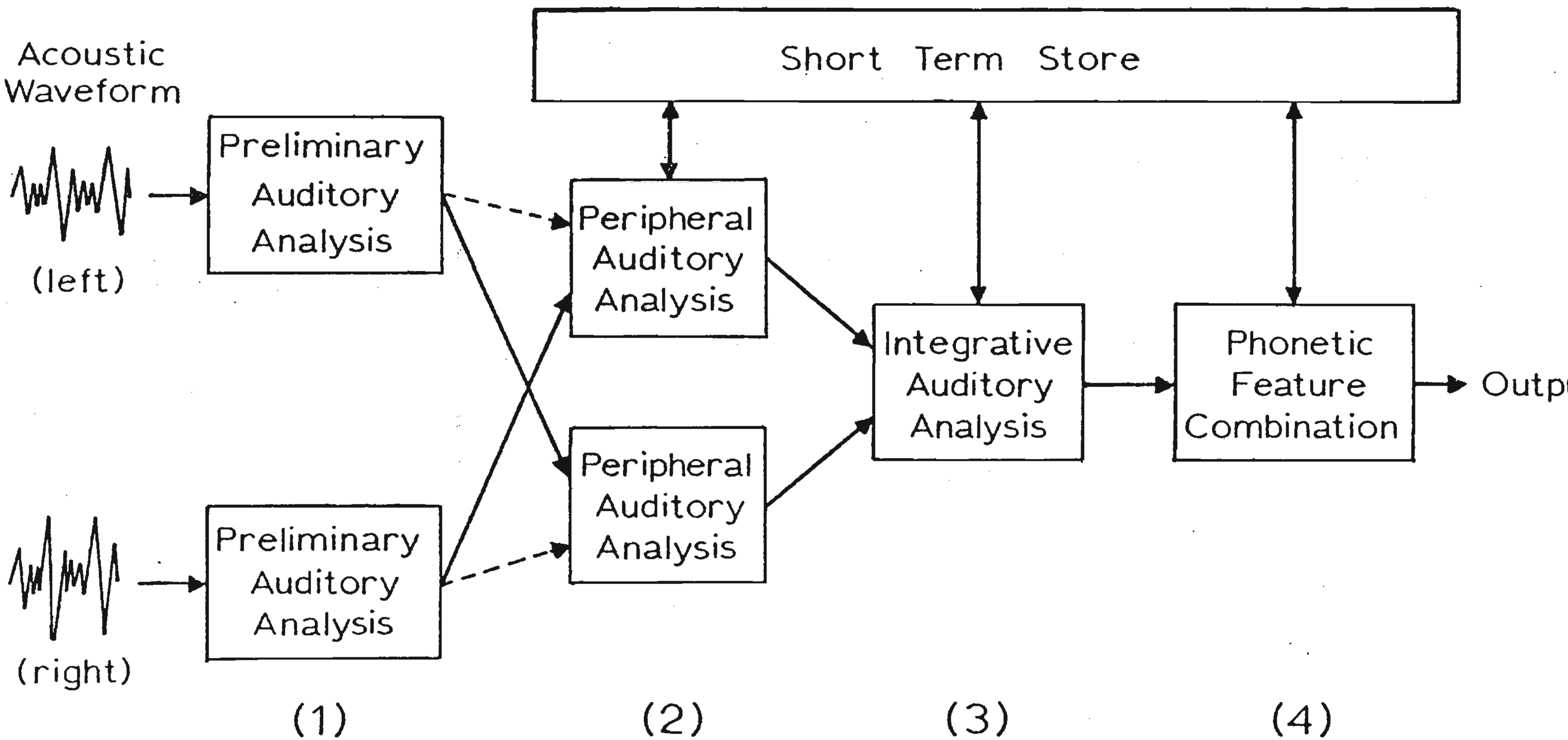
Passive models: Stage Theory. One recent formulation of speech perception, that of Bondarko et al. (1970) organizes speech perception into a hierarchy of stages. This formulation involves stages of auditory, phonetic, morphological, syntactic and semantic analysis. The auditory stage involves the extraction of "auditory features." These features are basically spectral in nature, such as formant frequency and amplitude. The second stage, phonetic analysis, involves a mapping of these auditory features into an abstract phonetic representation, such as distinctive features. The important point to note is that this mapping is not conceived of as one-to-one. Rather, many auditory features may map into one phonetic feature and one auditory feature may contribute to several phonetic features. The invariance problem is resolved, in part, by recourse to the analysis of multiple auditory features that are spread over time. Segmentation, in this model, is accomplished by higher levels of analysis and is not a property of these first two stages.

Many of the recent models of speech perception are similar to this outline proposed by Bondarko et al. (1970). These models all organize perception as a succession of partially overlapping stages of analysis. Details of the system, such as memory limitations and the coding processes and decision rules employed have been explicitly considered. It is to these models, formulated using the methods and goals of information processing, that we now turn.

Information Processing Models. A number of recent accounts of speech perception have emphasized that the perceptual process is divided into a hierarchy of stages (Studdert-Kennedy, 1974, 1975; Pisoni, 1975a; Pisoni & Sawusch, 1975; Tartter & Eimas, 1975; Cutting, 1976). In these models the early stages of perceptual processing are assumed to be automatic and normally not under conscious control. This is not to say that these models do not have an active control component or feedback systems. However, the active aspects of perceptual processing in these models are generally assumed to take place at higher levels of analysis such as phonological, morphological, syntactic and semantic stages (see Pisoni & Sawusch, 1975). A general model for these early stages of analysis is shown in Figure 1. The input to this model is the acoustic waveform at the ear. Its output is some form of phonetic or distinctive feature matrix. Between the input and output a succession of partly overlapping processing stages analyze the information in the waveform.

The first stage of analysis, Preliminary Auditory Analysis, involves a transformation of the acoustic waveform into its component frequencies and their intensities over time. This coding would be very similar to that produced by a sound spectrograph, except that certain temporal constraints and timing relations of the input are preserved. A number of auditory theories have been proposed to describe the peripheral encoding of acoustic waveforms at this level. For example, Licklider (1959) has outlined a model of auditory perception which provides the necessary spectral and temporal information for later stages. Cutting (1976) attributes sound localization in dichotic fusion studies to this first

Figure 1. Schematic diagram of an information processing model of the early stages of speech perception.



Short Term Store
Acoustic Waveform
(left)
(right)
Preliminary Auditory Analysis
Preliminary Auditory Analysis
Peripheral Auditory Analysis
Peripheral Auditory Analysis
Integrative Auditory Analysis
Phonetic Feature Combination
Outp
(1)
(2)
(3)
(4)

Figure 1.

level of analysis. The model of Licklider (1959) has just such mechanisms in it as would be necessary to account for these fusion results reported by Cutting (1976).

The second stage of analysis, Peripheral Auditory Analysis, is very similar to the auditory stage proposed by Bondarko et al. (1970). At this stage a number of spectral and temporal "features" are extracted from the output of the previous stage. Examples of such information include: 1) presence or absence of noise in the spectrum; 2) bandwidth of noise; 3) duration and intensity of noise; 4) onset relations of formants; 5) presence and frequency of the fundamental; 6) presence of rapid spectral changes; 7) direction and extent of spectral changes; and 8) global properties such as envelope amplitude. The extraction of these acoustic cues at the auditory level is done in a frequency specific manner. That is, a particular detector will be tuned to a specific frequency (or spectral) region. The detector will respond only if a particular acoustic cue is present in that frequency region. This will account for some of the dichotic masking results that have been found for CV syllables. When the vowels of the target syllable and the masking syllable are identical, more masking is found than when the vowels are not identical. For the case of dichotic masking using CV syllables, Pisoni (1975a; Pisoni & McNabb, 1974) used masking syllables with either [a], [ae], or [ɛ] vowels. The target syllable always had the [a] vowel. As the masker vowel moved from [a] to [ae] to [ɛ], the frequency separation between the formants of the target and the masker increased. As this separation increased, the amount of masking on the CV target decreased. These results indicate the presence of a frequency sensitive stage of processing in speech perception. In the present model, the Peripheral Auditory Analysis stage is just such a frequency specific level.

Up to this point, the perceptual processing that has been described may be distinguished by two factors. The first is that processing at the stages described is not specific to speech signals. Rather, these stages are general components of the human auditory system. The linguistic or specialized aspects of the speech perception system are built on top of these first two stages of analysis. A second distinguishing property of these two stages is that they are assumed to be bilaterally represented. The inputs from the two ears are, to a certain extent, kept separate through these first two stages of analysis. Evidence for this bilateral representation comes from studies of masking (Pisoni & McNabb, 1974; Pisoni, 1975a) and dichotic fusion (Cutting, 1976) as well as studies of the right ear advantage (Studdert-Kennedy & Shankweiler, 1970). Indicative of this is the data reviewed by Cutting (1976) where he shows that the fusions that are thought to occur at these first two levels of processing all show one particular property. More fusion is found for these stimuli when they are presented diotically (both stimuli to both ears) than when they are presented dichotically (one stimulus to each ear). On the other hand, the two fusions thought to occur at the higher levels of processing show the opposite trend (more dichotic than diotic fusion).

The third stage, Integrative Auditory Analysis, operates to extract patterns of features from those found by the peripheral auditory stage. Thus, multiple auditory features are mapped into single abstract features and one auditory feature may cue more than one abstract feature. The auditory property detectors proposed by Stevens (1973) are examples of the kinds of mechanisms that may be involved at this level. It should be noted, however, that the abstract features extracted here are not completely invariant: they still depend on context and the extraction of other abstract features to determine an invariant phonetic feature matrix. This is one reason why this stage of processing has been labeled abstract auditory as opposed to phonetic. The term phonetic is generally applied to a feature that is invariant and context independent.[2] However, in the present formulation, the resolution of the invariance problem occurs at both the integrative (abstract) auditory stage and the next stage of processing (Phonetic Feature Combination).

The abstract analysis stage is distinguished from earlier stages in that it is located centrally rather than peripherally. As mentioned earlier, there is more dichotic fusion than diotic fusion at this and succeeding levels which indicates that the abstract level is integrative, combining the separate information from the two ears. There is a second distinguishing feature of this abstract stage; it may represent linguistic processes that are part of a specialized speech system. It is very difficult, however, to distinguish between this view and an alternative view. It is possible that this level is only engaged by speech because speech is the only suitable, complex signal in the environment. For the moment, this question of whether the abstract level is linguistic or non-linguistic in nature will be left open. This issue will, however, surface again in the next chapter as a critical issue in recent experiments on selective adaptation.

The last stage, Phonetic Feature Combination, takes the outputs of the separate abstract auditory analyzers and combines them to form a feature matrix. In certain respects, the syllable is the unit of analysis here since, as noted by Studdert-Kennedy (1976), the acoustic information necessary to obtain an invariant feature matrix is distributed over the entire syllable. One possible conception of the operations that occur at this stage is in terms of a set of decision rules in which the outputs of the abstract auditory analyzers are mapped into distinctive feature format. The knowledge of the speaker about particular articulatory constraints may enter into these decisions (see Studdert-Kennedy, 1975; Pisoni, 1976). The phonological rules of the particular language also enter in at this point. In fact, the results of a number of phonological fusion studies (Day, 1968, 1970; Cutting, 1975; Cutting & Day, 1975) are the primary evidence in support of distinguishing between Phonetic Feature Combination and the earlier stages of analysis. In general, fusion will occur at much longer separation times between the two inputs for phonological fusion than for other types of fusion (Day, 1970; Cutting, 1975; or see Cutting, 1976, for a review). One other reason for separating this fourth stage from previous stages of analysis is that the phonetic feature is necessarily linguistic in nature and the decision rules used here are specific to language. As noted earlier, this is not the case for the first two stages of analysis and processing at the abstract (third stage) may or may not be linguistic in nature.

One last point to note about the model shown in Figure 1 concerns the role of memory in this system. All of Stages 2 thru 4 are connected to, and in certain respects considered a part of short term memory. The interaction between the processing system and memory is similar to that proposed by Craik (Craik & Lockhart, 1972) and Shiffrin (1975). As one stage of processing finishes with some form of output, it is maintained in short term memory for a certain period of time. How long it is maintained depends on the demands on memory for storing other information. While in short term memory, this information is available for further processing or, if appropriate, it is available for generating a response. Thus, memory is a limiting factor in the performance of this model. Accordingly, when the demand on short term memory due to a specific task is high (such as certain discrimination or masking tasks) performance is expected to decrease in a systematic fashion (see Pisoni, 1973, 1975a).

With this general framework for speech perception in mind, we are still left with the task of specifying what types of mechanisms are employed at these two auditory stages of processing. The idea of feature detectors has been proposed as one mechanism capable of resolving the basic information in the acoustic signal (Abbs & Sussman, 1971). Such detectors have been found in the visual system of the cat (Hubel & Wiesel, 1965) and the monkey (Hubel & Wiesel, 1968). These detectors respond to specific line orientations such as only vertical lines or only horizontal lines or particular patterns of line movement. Specialized detectors have also been found in the auditory cortex of the cat for changes in frequency (Whitfield & Evans, 1965) and for certain characteristics of two-tone frequency complexes (Whitfield, 1967). Although micro-electrode studies with human subjects are not possible, at least at the present time, a psychophysical paradigm involving adaptation (repeated presentation of a stimulus) has been used both in vision (Blakemore & Campbell, 1969; Blakemore & Sutton, 1969) and in audition (Kay & Mathews, 1972). These adaptation experiments are generally based on the supposition that if adaptation has a systematic effect on some aspect of perception, it was caused by the fatigue of a detector that is tuned to some information in the signal. Using this supposition, properties of the perceptual system can be investigated by systematic manipulation of the physical characteristics of the adapting and test stimuli.

The success of the adaptation paradigm when used as a part of certain psychophysical procedures makes it an attractive candidate for research on speech perception. The adaptation paradigm was first employed in speech perception research in an effort to look for mechanisms that are tuned to critical aspects of the speech signal (Eimas & Corbit, 1973; Bailey, 1975). It is to these recent developments and results using the adaptation paradigm that we turn in the next chapter.

Footnotes

[1] There are two different descriptions of phonetic units that are often used. The phoneme inventory referred to is that of the International Phonetic Association (1949). Alternatively, the sounds of language have been described in terms of a set of binary distinctive features (Jakobson, Fant & Halle, 1963). The term phonetic feature is also often used for this feature description of a speech sound. For the purposes of the present discussion, these two units are equivalent in that a phoneme represents a group of phonetic features.

[2] This comes from the criterion used in evaluating whether or not the difference between two speech sounds is phonetic. If the two sounds are part of otherwise identical morphemes that have different meanings, then the distinction is phonetic. The phoneme is traditionally thought to be the smallest linguistic unit that can cue a change in meaning.

CHAPTER II

Feature Detectors in Speech Perception: Evidence from Selective Adaptation Studies

Recently, a relatively new paradigm has been used to investigate various aspects of speech perception. This technique, known as selective adaptation, is similar to a paradigm previously used to study size perception in humans (Blakemore & Sutton, 1969). This paradigm consists of two phases: baseline determination and adaptation. The baseline measurement has usually been the identification of a set of syllables presented one at a time. Then, an adapting stimulus is presented repeatedly, in very quick succession. Following this adaptor, the same test simuli are presented again for identification. This procedure of adaptation and identification is repeated until a sufficient number of "adapted" responses have been collected for each of the baseline test stimuli.

Selective adaptation was first employed by Eimas and his coworkers (Eimas & Corbit, 1973; Eimas, Cooper & Corbit, 1973) to investigate the phonetic feature of voicing in stop consonants and concurrently by Bailey (1973, 1975) on the feature of place of articulation in stop consonants. The typical result has been that following adaptation, the phonetic boundary of the test series moves toward the adapting syllable. For example, Eimas & Corbit (1973) used synthetic [ba] and [pa] adaptors and tested on synthetic [ba]-[pa] and [da]-[ta] syllable series which varied voice onset time (VOT which refers to the interval between stop release and the onset of laryngeal pulsing). Following adaptation with the voiceless [pa] syllable, the phonetic category boundary of the [ba]-[pa] series shifted toward the [pa] end of the series. That is, subjects gave fewer [pa] responses to the [ba]-[pa] series following adaptation with [pa]. Similarly, the [da]-[ta] boundary shifted toward the [ta] (voiceless) end of its series. When the voiced [ba] syllable was used as an adaptor, the [ba]-[pa] series boundary shifted toward the [ba] (voiced) end of the series and the [da]-[ta] boundary shifted toward the [da] (voiced) end of its series. Bailey (1975) found similar results using a synthetic CV syllable series that ranged perceptually from [ba]-[da] along the feature of place of articulation. A [ba] adaptor shifted the [ba]-[da] phonetic category boundary toward the [ba] end of the series and a [da] adaptor shifted the boundary toward the [da] end of the test series.

Eimas & Corbit (1973) originally interpreted their results as reflecting the operation of a pair of linguistic feature detectors. One of these detectors operated to extract the voiced feature while the other registered the voiceless feature. Selective adaptation reduced the output of the detector that responded to the adaptor and thus shifted the region of overlap between these two detectors. The category boundary moved with this shift in the region of overlap between the two detectors.

Bailey (1975) interpreted his data in terms of fatigue at an auditory (non-linguistic) level of processing. In his data on the place feature, a [ba] adaptor had a larger effect on a [ba]-[da] test series than a [bɛ] adaptor (on the same test series). This result was interpreted by Bailey (1975) as indicating that adaptation had its effect prior to phonetic (linguistic) categorization because at a phonetic level of processing the "b" in [bɛ] should be equivalent to the "b" in [ba]. This issue, of where in the perceptual system adaptation is taking place, has been central to many of the experiments using selective adaptation and is part of the motivation of the present investigation.

Since these initial studies, a number of further studies have investigated the phonetic features of place, voicing and manner of production in consonants. These studies have generally been aimed at two issues: What aspects of the stimulus determine the adaptation effects and where in the perceptual system do these effects occur? In terms of the model sketched earlier, selective adaptation could be operating at any or all of four processing stages. It is also possible that selective adaptation operates subsequent to perceptual analysis, as a part of the response process. Before proceeding with an extensive review of the selective adaptation results to date, the question of whether selective adaptation affects the recognition system or later response organization stages will be considered.

Selective Adaptation and Response Organization

Evidence bearing on the question of whether selective adaptation is a perceptual or response organization phenomenon was first obtained by Eimas & Corbit (1973). As mentioned earlier, the labial voiceless stop [pa] had the same type of adaptation effect on an alveolar voicing series [da]-[ta] as it had on a labial voicing series [ba]-[pa]. This result indicated that the adaptation result was feature specific rather than syllable specific. However, the output of the recognition device is a feature matrix. The identification results of Eimas and Corbit (1973), Eimas et al. (1973) and Bailey (1975) do not rule out a response organization explanation. A number of theories of judgment, such as adaptation level theory (Helson, 1964) and range-frequency theory (Parducci, 1963,

1975) could potentially explain the selective adaptation results. These theories would take as their input the distinctive feature matrix that is the output of the recognition device. They would correctly predict that the effect of a repeated syllable (or feature) would be to move the category boundary toward the repeated syllable (feature).

A second result of Eimas & Corbit (1973) does indicate that category boundary shifts found in selective adaptation may not be due to response organization. Subjects were given an ABX discrimination task following adaptation with [pa]. The peak in discrimination for the [ba]-[pa] test series shifted toward [pa], mirroring the shift previously found in identification. Eimas & Corbit (1973) argued that this shift in the discrimination peak ruled out a simple response bias interpretation of selective adaptation. However, a direct test for the existence of response bias in the identification paradigm is necessary before any firm conclusions about the role of response bias in selective adaptation can be made.

In direct experimental tests of the judgment theories, Sawusch & Pisoni (1973) and Sawusch, Pisoni & Cutting (1974) tested the identification of a number of speech and non-speech dimensions under two experimental conditions. In the first condition, when a stimulus series was presented to a subject for judgment, each stimulus was presented equally often. This was done for tones varying in intensity, tones varying in frequency, a synthetic speech series that varied the nonlinguistic dimension of fundamental frequency and synthetic speech CV syllables that varied along the voicing or place dimensions. In all cases, subjects were to respond to the stimuli using one of two categories (e.g., loud

or soft for tones varying in intensity, [ba] or [pa] for a synthetic CV voicing series). In the second condition, the probability of occurrence of the stimuli within each series was changed. One of the end-point stimuli from the series occurred two or four times as often as each of the other stimuli from that series. The effect of this manipulation was to move the category boundary in the non-linguistic dimension series toward the more frequently occurring stimulus, relative to functions obtained in the equal probability control condition. This result is exactly as predicted by adaptation level and range-frequency theory and similar to results previously obtained with this paradigm on judgments of visual brightness (Helson, 1964) and weight (Parducci, 1963, 1975). However, for the two linguistic dimensions studied, place and voicing, no shift in the category boundary was found. These results indicate that response organization factors probably do not play an important role in the judgment of CV syllables varying along the phonetic features of place and voicing.

Another experiment by Sawusch and Pisoni (1976) extended this finding using the selective adaptation paradigm directly. Subjects in that experiment were tested on a [bi]-[di] place feature series. The middle stimulus from this series was ambiguously identified by subjects as either a [bi] or a [di] with about equal frequency. This middle stimulus was then used as an adaptor under two sets of instructions. One group was told that the repeated syllable they would hear was a [bi] whereas the other group was told that the repeated syllable was a [di]. The middle syllable was used as an adaptor because it would be compatible with both sets of instructions. Neither group showed a shift in their identification after adaptation, indicating that the instructions to subjects had no systematic effect during selective adaptation. Sawusch & Pisoni (1976) concluded that stimulus variables were the determining factor in selective adaptation and that response organization did not play a significant role. Thus, it seems reasonable to conclude that selective adaptation has its effect within the recognition system outlined previously and not at a later stage involving response organization processes. The question still remains, however, as to where in the recognition system selective adaptation has its effect.

Adaptation of the Voicing Feature

A number of selective adaptation experiments have studied the voicing distinction in stop consonants. The phonetic feature of voicing is carried by several acoustic cues. Among these are voice onset time (VOT), first formant onset frequency, extent of first formant transition and degree of aspiration. The primary cue for voicing appears to be what Lisker and Abramson (1964) termed VOT. This refers to the time interval between release of the consonantal burst and the onset of laryngeal pulsing. When voicing preceeds or is simultaneous with the release, the consonant is called voiced in English (e.g. [b,d,g]). When voicing lags behind release, the consonant is voiceless (e.g. [p,t,k]). VOT is not, however, a unitary cue since it consists of a set of smaller cues.

The acoustic cues to VOT. One of these cues is the relative delay in the onset of the first formant (hereafter referred to as cutback). This cue, which was first studied by Liberman, Delattre and Cooper (1958) is a sufficient cue for certain stop consonants. When the onset of the first formant lagged behind the higher formants by more than 30 msec the stops were identified by subjects as voiceless.

A second cue present in VOT is the duration of the first formant transition. Stevens & Klatt (1974) manipulated this cue separately from the cutback cue and investigated the trading relation between them. Their results indicated that the duration of the first formant transition changed the voiced-voiceless boundary between the syllables [da] and [ta]. For short duration first formant transitions, a 30 msec cutback was perceived as the voiceless stop [t]. When a longer first formant transition duration was used, a longer cutback lag was necessary to maintain the [t] percept. It should be emphasized that in the stimuli of Stevens & Klatt (1974), the total frequency transition for the first formant was constant and only the duration of this transition was manipulated.

Another cue to the voiced-voiceless distinction is the onset frequency of the first formant. In VOT, a voiced stop generally has a relatively low frequency onset for the first formant. Lisker (1975) found that when no first formant transition was present but the frequency of the first formant was changed, perception of the voicing feature changed. The lower in frequency that the first formant occurred, the more cutback was needed for the voiceless stop [k] to be heard.

In most experiments using synthetic speech that vary along the voicing dimension, the cues of cutback, extent of first formant transition and first formant onset frequency are confounded. As a further complication to this picture, the use of these multiple cues in perception seems to vary markedly with phonetic context. Summerfield & Haggard (1974) demonstrated this by varying the duration of transition and cutback cues. Both cues were found to be important before the vowel [a], with the transition cue being dominant. Before the vowel [i], however, there was no transition cue and the cutback cue dominated. Summerfield (1974, 1975) has proposed a model to handle this complex set of data on voicing. Two basic feature detectors extract acoustic cues from the waveform. One of these is a relative onset time detector. The second is a first formant onset frequency or transition detector. The outputs of these two cues are weighted and then combined to yield a decision about whether the input was voiced or voiceless. This model fits very nicely into the general approach outlined in the previous chapter. However, whether or not this model for voicing can account for the data from selective adaptation experiments is a separate question. The selective adaptation data on the voicing feature, and some of the models that have been proposed for these data are considered in the next section.

Selective adaptation to voicing. Eimas et al. (1973; Eimas & Corbit, 1973) originally interpreted their adaptation results on the voicing dimention as evidence for a linguistic feature detector system that was sensitive to VOT. This interpretation was based primarily on two main results. First, both a labial voicing series, [ba]-[pa], and an alveolar voicing series, [da]-[ta], exhibited adaptation effects to the same [ba] and [pa] adaptors. However, a "non-speech" adaptor which consisted of the first 50 msec of a voiced [da] syllable (referred to as a d-chirp) failed to produce a significant adaptation effect. Consequently, Eimas et al. argued that adaptation was not occurring at either an acoustic or a syllabic level but rather at a phonetic level. If adaptation were at the syllabic level, different test series (i.e., [ba]-[pa] vs. [da]-[ta]) should have shown a differential response to a single [ba] adaptor. As shown by Eimas & Corbit (1973) this was not the case. Their data showed that a [ba] adaptor had the same effect on both the [ba]-[pa] and the [da]-[ta] series. The same pattern of results was also found for the [pa] adaptor.

Secondly, the d-chirp, which contained the acoustic information for a voiced stop in initial position, did not produce a significant adaptation effect. The d-chirp contained a full first formant transition so it should have provided the essential acoustic cues for a voiced stop. The failure of this d-chirp to produce adaptation led Eimas et al. (1973) to conclude that the detectors were not operating at the level of acoustic cues, but at a later integrative phonetic level. This conclusion must, however, be modified somewhat. The null result of Eimas et al. does not indicate that adaptation is not taking place at an acoustic level. It does indicate that the d-chirp did not contain sufficient information to adapt any lèvel, auditory or phonetic. Since voicing is a relative timing cue, it could be argued that the d-chirp does not contain sufficient information to be registered by the appropriate auditory (or phonetic) detectors.

In another study involving the voicing dimension, Tartter and Eimas (1975) used a number of different non-speech adaptors. Of particular interest was their pae-F2F3 adaptor which consisted of the second and third formants of a voiceless [pae] syllable. This adaptor preserved the relative onset cue for voicelessness and eliminated any possible first formant cue to voicing. The pae-F2F3 adaptor had a significant effect on the [bae]-[pae] phonetic boundary. The boundary locus moved toward the voiceless [pae] end of the series. Thus, Tartter & Eimas

(1975) found a non-speech adaptor that produced a significant adaptation effect. This result is strong support for an auditory level of adaptation. The non-speech character of the pae-F2F3 adaptor should have precluded it from being analyzed by any phonetic level feature detection mechanisms.

One other result, from an experiment by Cooper (1974b) also indicates the presence of non-phonetic mechanisms in selective adaptation to voicing. Cooper alternated [da] and [ti] adaptors (also [di] and [ta]) and then tested subjects on two voicing continua: a [bi]-[pi] series and a [ba]-[pa] series. The results showed contingent adaptation effects which were vowel specific. That is, with [da] and [ti] adaptors, the [ba]-[pa] boundary shifted toward the voiced end of the series and the [bi]-[pi] boundary toward the voiceless end of the series. The converse shifts in the phonetic boundaries were found for [di] and [ta] alternating adaptors. This vowel contingent adaptation effect has since been replicated by Pisoni, Sawusch & Adams (1975) as part of a larger study using alternating syllables as adaptors.

The contingent adaptation effect on voicing requires one to postulate that the mechanisms involved in selective adaptation to voicing are at least partly vowel specific and hence frequency specific in nature. The implication of this is that if adaptation is occurring during phonetic feature processing, the phonetic level must also be vowel specific. As suggested earlier, it is the auditory level of speech processing that appears to be vowel and frequency specific, not the phonetic level. It thus seems reasonable to ascribe the contingent results of Cooper (1974b) to the adaptation of detectors at an auditory level of processing prior to phonetic categorization. This would also explain the adaptation effect found by Tartter & Eimas (1975) using the pae-F2F3 non-speech adaptor. In fact, it seems reasonable, at least in principle, to attribute all of the selective adaptation effects mentioned so far to auditory level detectors of the sort proposed by Summerfield (1974, 1975). That is, adaptation with a voiced stop affects a first formant detector. For different vowel contexts, where the first formant is in a different frequency region, different detectors would be affected. Adaptation with a voiceless stop has its effect upon a relative onset time detector. As with the first formant detectors, the channel for relative onset time is frequency sensitive.

At this point it seems reasonable to ask if there is any evidence favoring adaptation at a central, more abstract phonetic level. Two experimental results do favor, or at least make plausible, adaptation at a phonetic level, in addition to the earlier auditory level. The first of these is an experiment by Eimas et al. (1973). Selective adaptation was carried out using a [ta] adaptor under two conditions. In one of these, the adaptor and test syllables (a [da]-[ta] series) were both played binaurally (diotically). In the second condition, the adaptor was presented to one ear whereas the test syllables were presented to the other ear (dichotic presentation). The results revealed that when adaptation and testing were in separate ears, the resulting phonetic boundary shift was just as large as when the adapting and test syllables were presented to the same ear. This 100% interaural transfer effect indicates that the locus of selective adaptation to the voiceless end of a voicing series is central rather than peripheral.

According to the model outlined earlier, central effects are indicative of abstract integrative processing. The results of Eimas et al. (1973) thus indicate that the cues to voicelessness are extracted at a central site after binaural fusion. This is not, nowever, firm evidence that the adaptation effect is phonetic. In light of the adaptation effect found by Tartter & Eimas (1975) using the pae-F2F3 non-speech adaptor, it seems reasonable to conclude that the relative onset detectors in Summerfield's model for voicing (1974, 1975) operate at a central site, subsequent to binaural fusion. This result raises the question of whether the first formant detectors, which generally respond for voiced consonants, are located peripherally or centrally (or both). Unfortunately, Eimas et al. (1973) did not run an interaural transfer condition with a voiced (e.g. [da]) adaptor. On the basis of Summerfield's model, in which distinctly different auditory mechanisms underlie the voiced and voiceless attributes, an interaural transfer effect of considerably less than 100% for a voiced adaptor on a voicing test series would not be unexpected.

The most persuasive evidence for a phonetic level of adaptation to voicing comes from another study by Cooper (1974c). The adapting syllables for this study were two [da] syllables from Stevens & Klatt (1974). One, labeled [da]-long, had 40 msec transitions after the onset of voicing. The second, [da]-short contained only 10 msec transitions. Both of these adaptors had a VOT of +25 msec. The test series was a [ba]-[pa] series where VOT ranged from +5 to +55 msec in 5 msec steps. Formant transition duration for this series was 20 msec. The [da]-long adaptor had a significant effect and moved the voiced-voiceless boundary toward the voiced [ba] end of the series. The [da]-short had no significant effect on the test series. The critical question here was what aspects of the test series were being affected by the adaptor. The test stimuli around the phonetic boundary had no appreciable first formant transitions. Since the [ba]-[pa] test series had 20 msec duration transitions, no appreciable transition would be left for a +15 or +20 msec VOT. Thus, Cooper argued from these results that the adaptation effect caused by the [da]-long adaptor could not be due solely to detectors sensitive to first formant transition duration. Rather, a more integrative level which was sensitive to a number of auditory cues was adapted. As noted by Ades (1976), this argument rests on one critical assumption. Cooper (1974c) assumed that detectors sensitive to first formant information would not respond at all to the test stimuli with a +15 to +20 msec VOT. This assumption can be challenged on two grounds. First, the first formant detectors may also respond to first formant onset frequency (cf. Lisker, 1975) in addition to first formant transition duration. This could explain the adapting effect of the [da]-long adaptor. It is also possible that the first formant transition detectors do respond somewhat to the +15 to +20 msec VOT stimuli from Cooper's test series. This response would be greatly diminished, if not eliminated, following adaptation by the [da]-long syllable and the category boundary would shift toward the voiced end of the series. Consequently, the results of Cooper (1974c) should not be taken as firm evidence for an integrative level of adaptation for voicing.

In summary, an auditory level explanation of the voicing data seems to be tenable (cf. Ades, 1976). The adaptation results of Cooper (1974b) and Tartter & Eimas (1975) support the existence of an auditory level of adaptation. The question of whether adaptation also occurs at a more abstract integrative level is still open, although the data of Cooper (1974c) suggest that such a site for adaptation may exist. The main problem with most of the experiments on voicing, including the studies in selective adaptation, has been to carefully sort out and control the multitude of auditory cues that are present in the stimuli. With this control, selective adaptation can be used to investigate possible processing mechanisms for the multitude of acoustic cues to the voicing distinction. Without this control, it will be next to impossible to determine which cues are being adapted and whether an abstract, integrative (and possibly phonetic) level is involved in selective adaptation along the voicing dimension.

Adaptation of the Place Feature

The place of articulation feature in stop consonants, like the voicing feature, has been the subject of a number of selective adaptation experiments. Also, like voicing, the place feature is carried by a set of multiple acoustic cues. The three places of articulation that occur in English stops are bilabial (e.g. [b,p]), alveolar or dental (e.g. [d,t]) and velar (e.g. [g,k]). These labels refer to the point of maximal constriction in the vocal tract during articulation of the stop. Corresponding to these three values of the place feature is a complex set of acoustic cues involving formant transitions and a noise burst.

The acoustic cues to place. The dominant cues to place of production in initial stops are the direction and extent of the second and third formant transitions (Delattre, Liberman & Cooper, 1955; Liberman, Delattre, Gerstman & Cooper, 1956; Hoffman, 1958) and the presence and frequency location of an initial noise burst (Cooper, Delattre, Liberman, Borst & Gerstman, 1952; Hoffman, 1958). Generally, for syllable initial position, a low frequency burst followed by rising second and third formant (F2 & F3) transitions corresponds to the bilabial place feature. Similarly, a high frequency burst followed by falling F2 and F3 transitions usually corresponds to the alveolar feature. A mid-frequency burst followed by diverging F2 and F3 transitions generally corresponds to the velar feature. However, as noted in the previous chapter, these cues are not always invariant, especially in front of the extreme vowels. In a study on possible interactions and interdependence of the burst and transition cues, Hoffman (1958) varied the burst, F2 transition and the F3 transition orthogonally. Hoffman's results indicate that these three cues acted independently in perception, possibly in a fashion similar to vector addition.

Stevens (1973) has proposed an integrative mechanism that could use these three acoustic cues to place. Basically, he argues that the invariance problem for syllable initial stops can be resolved by using three integrative property detectors. One of these detectors would be sensitive to a general frequency rise in the spectrum and indicate the labial feature. A second would respond to a general fall in the spectrum and indicate the alveolar feature. The third property detector would respond to a divergence in the frequency spectrum from some middle value and correspond to the velars. These three property detectors could easily be operating in a manner similar to that suggested by Hoffman (1958). In terms of the model outlined earlier, the formant transitions and burst frequency are separately and independently extracted by frequency specific detectors at the peripheral auditory stage. Integrative property detectors at the abstract auditory stage would then respond to particular patterns of these auditory cues in a process much like vector addition. A system such as this would partly resolve the invariance problem since a set of multiple cues would be used rather than reliance on only one specific, isolated auditory cue.

This set of property detectors will not, however, completely resolve the invariance problem. The cues for stop consonants in syllable final position are the mirror images of those in initial position. The labial feature is characterized by falling transitions in final position (as opposed to rising in initial position) and vice versa for the alveolar feature. At least two solutions to this problem are possible. A separate set of integrative detectors for place in syllable final position could be proposed or the mapping of the integrative detectors into invariant place features could be resolved at a later stage (i.e., Phonetic Feature Combination in the model outlined earlier). At this later stage, information about syllable position would then be used to map the property detector outputs of Stevens (1973) into a set of invariant phonetic features.

One other approach to the recognition of the place feature for stop consonants has also been proposed by Stevens (Stevens & Blumstein, 1976). The frequency spectrum at onset (or offset) of a bilabial consonant may be characterized by a concentration of energy at relatively low frequencies. For alveolars, the concentration is at high frequencies whereas for the velars the concentration is at mid-frequencies. This frequency distribution is the same for both syllable initial and final positions.

Stevens & Blumstein (1976) have proposed that these spectral characteristics are extracted by the perceptual system. However, the problem of how these distributions are arrived at by the perceptual system has been left unresolved. In terms of the model outlined earlier, this recent formulation by Stevens & Blumstein may be a characterization of either the abstract auditory level or the phonetic feature combination level. An earlier stage of processing is still needed where the information necessary to determine the spectral frequency distribution is extracted. In order to distinguish among these possible conceptions of the perceptual processing of the place feature, we turn now to the selective adaptation data on place of production.

Selective adaptation to place. Following the work of Bailey (1973, 1975), a number of further studies using selective adaptation on the place feature have been reported. Cooper (1974a) and Cooper & Blumstein (1974) used a three category synthetic CV syllable series that ranged from [bae] through [dae] to [gae]. In the Cooper (1974a) study, one of the experiments involved adaptation with three syllables from the 13 element [bae]-[dae]-[gae] series. The adaptors were the first [bae], seventh [dae] and thirteenth [gae] syllables from the test series. The [bae] adaptor moved the [bae]-[dae] category boundary toward the [bae] end of the series while leaving the [dae]-[gae] boundary unaffected. Similarly, the [gae] adaptor moved the locus of the [dae]-[gae] category boundary toward the [gae] end of the continuum while leaving the [bae]-[dae] boundary relatively unaffected. The [dae] adaptor affected both boundaries, moving both toward the middle and thus narrowing the [dae] phonetic category. In a second experiment using ABX discrimination instead of identification, Cooper (1974a) found that the discrimination peaks moved in the same manner as the identification category boundaries. These results, as noted by Cooper, provide support for a three value place feature dimension and rule out a binary distinctive feature interpretation of selective adaptation in terms of features such as relative frontness or relative backness.

In other experiments, Cooper (1974a) and Cooper & Blumstein (1974) used a number of different adaptors in selective adaptation experiments. Among them were the synthetic vowel [ae] and real speech [bae], [p^hae], [bi], [mae], and [wae] syllables. All of these CV syllable adaptors share the bilabial place feature with the [bae] end of the [bae]-[dae]-[gae] test series. All of the real speech CV syllable adaptors except [wae] produced a significant effect on the [bae]-[dae] phonetic boundary. In each of these cases, the locus of the phonetic boundary moved toward the [bae] end of the series. The [wae] adaptor produced a small but non-significant shift in the category boundary toward [bae]. The [ae] adaptor had no systematic effect on either the [bae]-[dae] or the [dae]-[gae] boundary.

Cooper & Blumstein (1974) interpreted their results in terms of the similarity between the spectral cues (transitions) in the various adaptors and the [bae] end of the test series. The adaptation effects that they found were attributed primarily to an auditory level. However, the possibility of adaptation at an integrative, phonetic site was left open. Cooper (1974a) interpreted the adapting effect of [bi] on his [bae]-[dae]-[gae] test series as reflecting the operation of a second, phonetic level of adaptation.

This interpretation, that adaptation was taking place at a phonetic level in addition to a lower, auditory level, is questionable because of two factors. First, though the adaptors represent real speech with a range of consonants and one different vowel ([bi]), sound spectrograms of the real speech indicated that there was extensive overlap between the second and/or third formant transitions of the various adaptors and the [bae] end of the test series. Consequently, the acoustic information underlying the labial feature in each of the adaptors was very close to that in the test series. Secondly, the magnitude of the adaptation effect differed considerably for the various adaptors. The most adaptation was produced by the synthetic [bae] from the test series in Cooper's (1974a) study and by the real speech [bae] in the Cooper & Blumstein (1974) study. A comparison of the relative magnitudes of the effects produced by the various adaptors from both the Cooper (1974a) and the Cooper & Blumstein (1974) studies is shown in Table 1. Due to these two factors of spectral overlap and reduced effectiveness of alternative adaptors, selective adaptation could result from changes at either a frequency specific auditory stage or both auditory and integrative phonetic stages. In order to evaluate these possibilities, adaptation with non-speech stimuli and speech stimuli with a variety of different spectral properties is necessary. The interpretation in terms of adaptation at only an integrative phonetic level can be ruled out by the smaller adaptation effects found when irrelevant variations were made in the adaptors (such as the smaller adaptation effects of [bi] and [mae]). If adaptation were occurring only at the phonetic level, then adaptation with [bi] and [mae] should have been just as effective in moving the [b-d] phonetic category boundary as the [bae] adaptor.

Table 1

Shift in the [bae]-[dae] category boundary toward [bae]

as a function of various adapting syllables.

(Data from Cooper, 1974a; Cooper & Blumstein, 1974)

	Synthetic	Real Speech			Real Speech			
	[bae]	[bae]	[p^{h}ae]	[bi]	[bae]	[mae]	[vae]	[wae]
Shift (Stimulus Units)	1.46	.99	.61	.66	2.10	1.66	1.67	.38
Percent Shift	100%	68%	42%	45%	100%	79%	80%	18%

Several additional experiments have focused on the auditory level of processing and its role in selective adaptation. Tartter & Eimas (1975) used a variety of non-speech adaptors and tested subjects along a [bae]-[dae]-[gae] place series. Their adaptors included the first [bae] syllable and five non-speech adaptors made from this original [bae] syllable. The non-speech adaptors were the second and third formants of the [bae], the second formant alone, the third formant alone, the [bae]-chirp which consisted of the initial 45 msec of the syllable, and a "[bae]" syllable which had a falling first formant. All of these adaptors had a significant effect in moving the [bae]-[dae] phonetic boundary toward [bae]. However, none of the non-speech adaptors produced the same magnitude shift in the category boundary that was found with the entire [bae] syllable. If adaptation was affecting a frequency specific auditory level only, then the F2F3 pattern, which carries the essential place information, should have been as effective as the full syllable.

The results of Tartter & Eimas (1975) provide evidence for two conclusions. First, there is definitely an auditory component to selective adaptation. The effects of the isolated F2 and F3 adaptors on the category boundary appear to be quite strong. These patterns carry the same place information that is contained in the test syllables. However, these non-speech adaptors are not heard as speech and thus would not be expected to undergo any phonetic level processing. Consequently, the adaptation effects of these stimuli should be entirely confined to an auditory level of processing. Second, the smaller adaptation effects found with the F2F3 and [bae]-chirp adaptors seem to indicate the presence of a higher level component to adaptation. This higher level is apparently not adapted unless the adapting stimuli are hears as speech signals.

One other experiment using non-speech adaptors is also of interest. Pisoni & Tash (1975) used what they labeled a speech embedded chirp (SEC) as an adaptor. The test series in their experiment was a [ba]-[da] place series. The first [ba] and last [da] stimuli from this series were used as adaptors. The SEC adaptors were constructed from the speech adaptors by first deleting all of the steady state vowel following the transitions. Then, steady state formants were attached in front of the consonant transitions with center frequencies identical to the onset frequencies of the transitions. The resulting signal was a VC-like stimulus which was heard as speech-like by all subjects. It should be noted that the SECs are not VC syllables since the final F1 transition is rising and such a configuration is impossible for a speaker to produce.

Both of the SECs had a significant adapting effect on the [ba]-[da] test series. The b-SEC moved the locus of the phonetic category boundary toward [ba] and the d-SEC moved the boundary toward [da]. However, these SEC adaptors did not move the category boundaries nearly as much as the [ba] and [da] adaptors. Instead, they were only about 30% as effective as the end-point syllable adaptors. Thus, while the Pisoni & Tash (1975) study provides further support for an auditory level of adaptation, it suggests, as did the Tartter & Eimas (1975) study, that a more integrative level of adaptation may also be present.

Further support for an auditory component comes from the use of alternating adaptors on the place feature. As with the voicing feature (Cooper, 1974b), a contingent adaptation effect has also been found for the place feature. Pisoni & Sawusch (1976) tested on both a [bi]-[di] and a [ba]-[da] series. When [bi] and [da] adaptors were alternated, the [bi]-[di] series shifted toward the [bi] end while the [ba]-[da] series shifted toward the [da] end. A similar contingent effect was found for alternating [ba]-[di], [pi]-[ta] and [pa]-[ti] adaptation conditions. These results fit very well into a model of selective adaptation which incorporates a frequency specific auditory level. The second and third formant transitions for the two series used were generally in different frequency regions. Thus, rising [b] transition detectors could be fatigued in one frequency region and falling [d] transition detectors in another region. During testing, a CV series with a given vowel would only fall within one of these ranges. The net effect would be the contingent effect observed. Similar contingent adaptation results on the place feature have been reported by Miller & Eimas (1975) using a [bae]-[dae] place series and a [bi]-[di] series.

Ades (1974a) provided a different test for the presence of an auditory component in selective adaptation. Subjects were adapted and tested under two conditions. In one of these, all stimuli were delivered monaurally, with the adaptor and test stimuli going to the same ear. In the second (dichotic) condition, the adaptor was delivered to one ear and the test stimuli to the other ear. In all cases, the test series was a [bae]-[dae] place continuum and the [bae] and [dae] adaptors were the end-point stimuli from the test series. The dichotic adaptation conditions produced approximately 60% as much adaptation as that found in the monaural conditions. This result contrasts with the 95% transfer found previously by Eimas et al. (1973) on the voicing feature. The less than

100% interaural transfer is direct evidence for a bilaterally represented component in selective adaptation. The auditory level of the recognition system outlined previously would be just such a component. The important question is whether or not the 60% transfer of adaptation represents the processing of a central, phonetic level. Ades (1974a) offered two possible models, illustrated in Figure 2, that could account for these results. The model in panel A of Figure 2 is a two-tiered, auditory-phonetic model. If the auditory components are assumed to be separate and monaurally driven, then this model can account for the results of Ades' experiments. The 60% interaural transfer represents the common path through the central phonetic component of this model. By adapting in one ear and testing in the other, those auditory level detectors that were adapted have been circumvented. The model in panel B is an auditory, one level model. Each of the auditory components receives inputs from both ears. However, the input is stronger from one ear than the other. Models similar to this have been proposed to account for the right ear advantage found in dichotic listening tasks (Kimura, 1967; Studdert-Kennedy & Shankweiler, 1970). Since the contralateral pathways (i.e., from the left ear to the right hemisphere and vice versa) are stronger than the ipsilateral pathways, the interaural transfer results could be due to the difference in transmission pathways. Adapting in one ear would result in substantial adaptation in one analyzer via the contralateral pathway and lesser adaptation in the other analyzer. Testing in the opposite ear would send a strong signal to the weakly adapted analyzer and less than 100% transfer would result.

Figure 2. Two possible models for selective adaptation data. On the left (A) is a two tiered hierarchical model. On the right (B) is a one level bilateral model. (After Ades, 1974a)

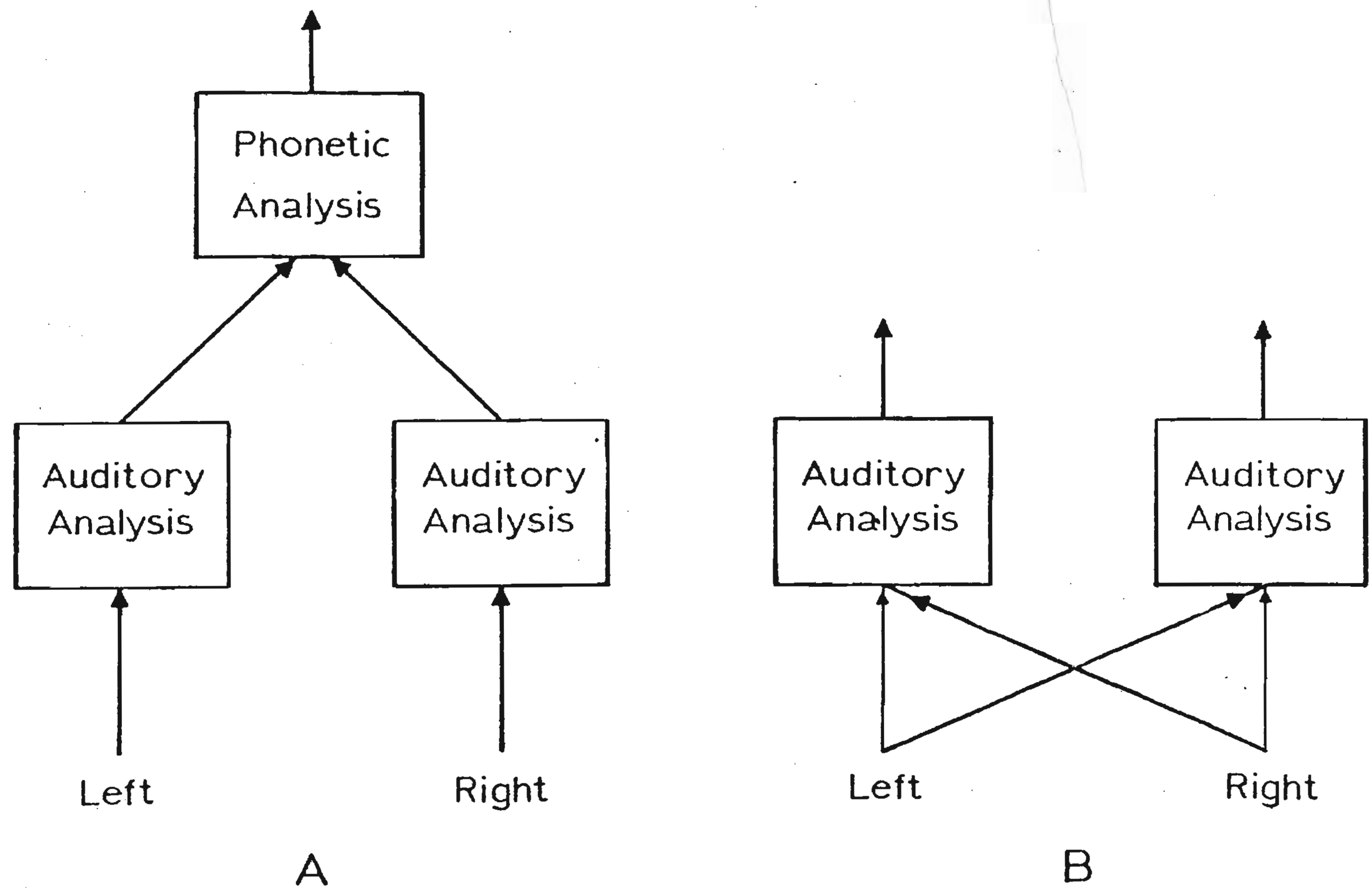

Phonetic Analysis
Auditory Analysis
Auditory Analysis
Left
Right
A
Auditory Analysis
Auditory Analysis
Left
Right
B

Figure 2.

In order to provide conclusive evidence for the two-tiered model, it would be necessary to find an adapting stimulus that produces the same adaptation effect regardless of which ear is adapted and which ear is tested. The bilateral, one-tiered model cannot nandle an instance of 100% interaural transfer. Thus, while the results of Ades (1974a) are suggestive, they do not offer conclusive evidence for the presence of a central, integrative (phonetic) level of adaptation.

None of the experiments reviewed so far have demonstrated conclusively that a central, integrative phonetic component is present in selective adaptation. An experiment by Ades (1974b) investigated the possibility of a VC syllable adapting a CV series and vice versa. The acoustic cues for the place feature in final position are the mirror image of those in initial position. Ades used two test series, a [bae]-[dae] series and an [aeb]-[aed] series where the stimuli were the mirror images of the [bae]-[dae] stimuli. Schematic spectrographs of these stimuli are shown in Figure 3. When a VC end-point was used as an adaptor and subjects were tested on the CV series, no adaptation was found. Conversely, the CV end-points failed to adapt the VC series. These results can be interpreted two ways. In one respect, these results would seem to rule out a strict phonetic level of adaptation. If adaptation were occurring at a level where phonetic features were invariantly coded such as the distinctive features of Chomsky & Halle (1968), then an [aeb] should have a b-adapting effect on a [bae]-[dae] series. However, the failure to find such a result does not absolutely rule out the possibility of such a level of adaptation. As mentioned previously, the acoustic cues for place in initial and final syllable position are mirror images.

Figure 3. Schematic spectrograms of the [aeb] (solid formants) and [aed] (dashed formants) endpoints and the [bae] (solid formants) and [dae] (dashed formants) endpoints from these two sets of syllables. (After Ades, 1974b)

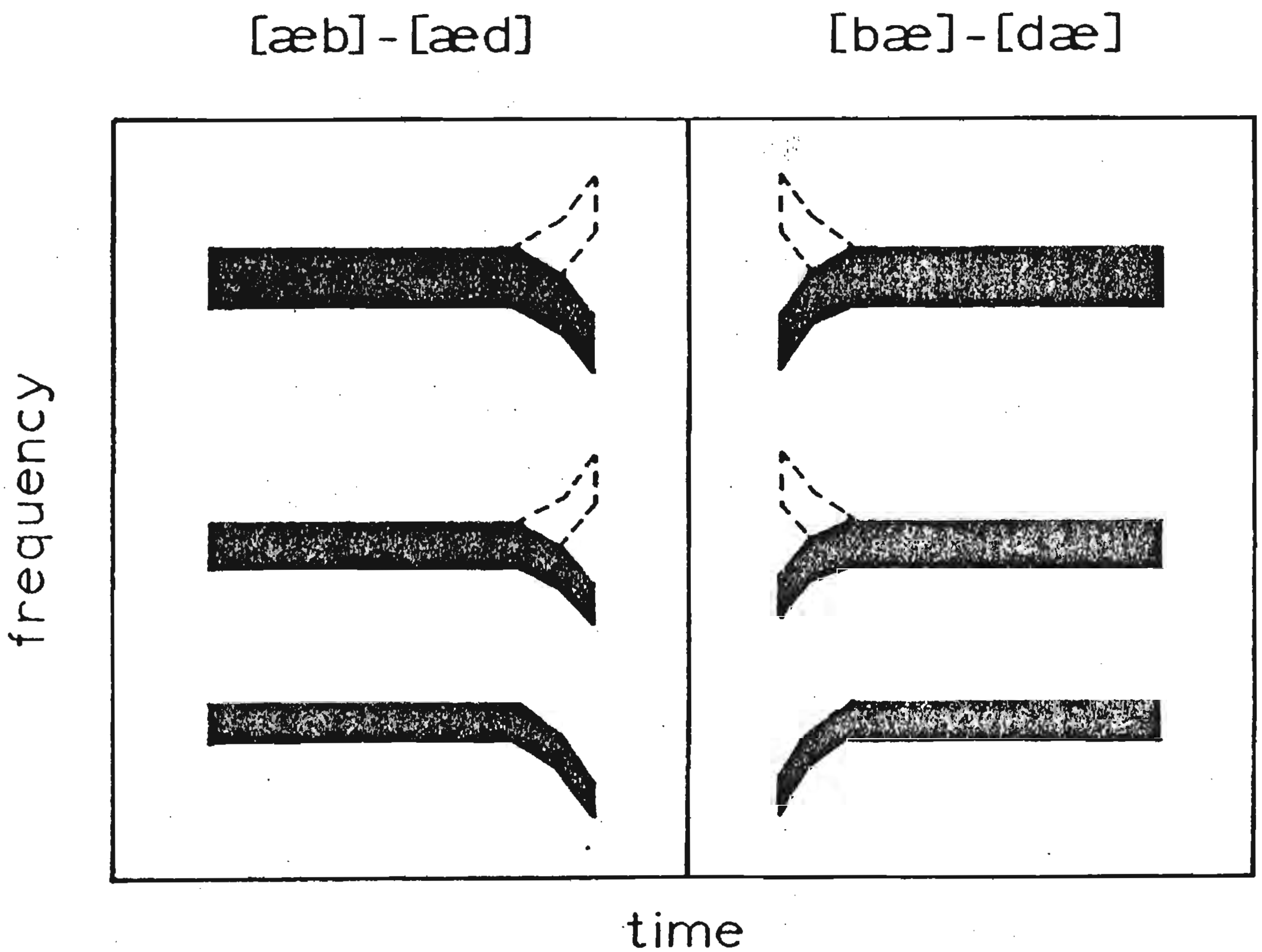


Figure 3.

At the acoustic level then, the falling transitions of an [aeb] could be expected to have a d-adapting effect on a [bae]-[dae] test series. This is precisely the type of result found by Pisoni & Tash (1975) as reviewed earlier. Consequently, if there is a phonetic component in selective adaptation, it may have cancelled the auditory component in the experiment by Ades (1974b). The question as to whether a feature in final position will adapt an analogous feature in initial position was also investigated by Pisoni (1975b).

The CV and VC stimuli used by Ades (1974b) had logarithmic formant frequency transitions (see Figure 3). Consequently, although the CV and VC syllables were mirror images of each other, the formant transitions followed different trajectories. The CV and VC stimuli employed by Pisoni (1975b) had linear formant frequency transitions. The final transitions for the [ab] and [ad] VC syllable adaptors were the exact reverse of the initial transitions for [ba] and [da]. The [ab] adaptor had a d-adapting effect on the [ba]-[da] test series as it moved the category boundary toward [da]. A converse, b-adapting effect for the [ad] syllable was also found. These adaptation effects followed the similarity of the formant trajectories between initial [b] and final [d] and vice versa and thus this adaptation effect seems to be taking place at an early auditory level. However, the b-adapting effect of [ad] was far smaller than the category boundary shift found with a [ba] adaptor. The d-adapting effect for [ab] was also much smaller than the shift found with the [da] adaptor. This would result if a larger, dominant auditory level effect were to cancel a smaller phonetic level effect in the opposite direction. Resolution of the question as to whether the place feature in final position will adapt an analogous feature in syllable initial position must await further research where the spectral composition of the CV and VC stimuli is manipulated systematically.

Manipulation of the acoustic cues in adapting and test syllables has been performed in a number of experiments. In attempts to provide an answer to the question of phonetic adaptation, three experiments have manipulated the spectral information in the adapting and test syllables (Blumstein & Stevens, 1975; Diehl, 1975; Ganong, 1975). Diehl used a test series that varied from [bɛ] to [dɛ] on the place feature. These synthetic stimuli were cued by formant transitions. One of the adapting syllables was a burst cued [tɛ] which had no transitions, only a high frequency initial burst. The burst cued [tɛ] produced a consistent, significant adaptation effect. The phonetic boundary shifted toward the [dɛ] end of the series following adaptation. This result fits very well with the two level model which predicts that the burst cued adaptor would fatigue the second abstract integrative level and thus cause a shift in the [bɛ]-[dɛ] boundary toward the [dɛ] category.

The experiment by Ganong (1975) was of a similar nature. Ganong tested subjects on a [bae]-[dae] series before and after presentation of four adaptors. The adapting syllables were the [dae] from the end of the test series, a [sae] with the formant transitions for a [dae] included, a [sae] without any formant transitions and a [tae] without formant transitions. Both the [sae] and [tae] without formant transitions shifted the phonetic boundary along the [bae]-[dae] series toward the [dae] end. The adaptors without transitions did not, however, produce as large a shift as the [dae] and [sae] adaptors with transitions. The one (auditory) level model cannot account for these results. At the auditory level, the transition and burst cues are assumed to be extracted separately (cf. Hoffman, 1958). As a consequence, a burst cued adaptor should have no effect on a transition cued test series.

Blumstein and Stevens (1975) performed a similar experiment in which subjects were tested on a transition cued [ba]-[da]-[ga] place series. Among the adaptors used were transition cued [da] and [ga], and [da] and [ga] syllables cued by both transitions and appropriate bursts. The syllables with both cues produced a larger adaptation effect than the transition cued adaptors. This result is also in agreement with the two level model. The one level auditory model cannot handle the increased adaptation due to the presence of an appropriate burst. The test stimuli did not contain bursts. Therefore, using the independent extraction assumption (Hoffman, 1958), the addition of a burst to a transition cued adaptor should not increase the adaptation effect at the auditory level. The increased adaptation found with the addition of the burst cue to the adaptor thus reflects adaptation at an integrative level where burst and transition cues are combined. The predictions of a one-level, auditory model are inconsistent with the results of Blumstein & Stevens (1975), Diehl (1975), and Ganong (1975).

It is possible, however, to save the one level model by making an additional assumption. Ades (1976) assumed that the transition detectors at the auditory level were also responding to the burst cue. When the burst frequency is in the immediate vicinity of the onset frequency for the transitions, this does indeed seem like a plausible argument. Thus, the results of Diehl (1975) could indeed have been due to transition detectors being triggered by the burst. In Diehl's experiment, the [tɛ] adaptor had a burst frequency of 2665 Hz. The starting frequency for the third formant transition was 2726 Hz for the [dɛ] end of the test series. This close proximity between the burst frequency and the onset frequency of the third formant transition for the [dɛ] end of the test series makes it very plausible that one auditory detector could have responded to both cues. The burst and transition frequencies for the Blumstein & Stevens (1975) experiment were not available.

Ganong's (1975) transitionless [sae] adaptor had a steady state third formant at 2400 Hz and initial 250 Hz bandwidth noise centered at 4500 Hz. In contrast, the starting frequency for the third formant transition at the [dae] end of the test series was 3000 Hz. The noise center frequency of the [sae] and the transitions for the [bae]-[dae] test series were separated by more than 1000 Hz. This large separation makes Ades' (1976) explanation of the adapting effect of the transitionless [sae] untenable. Any transition detector that responded to a burst that was over 1000 Hz away in frequency would be integrating information over a wide frequency range. This would be equivalent to blending the second and third stages of the recognition system (Figure 1) together. Since the distinction between the peripheral auditory and abstract auditory levels of the recognition system was motivated on experimental grounds other than selective adaptation, it seems useful to retain the distinction here. Consequently, a one level interpretation of Ganong's (1975) results is unreasonable, Ades (1976) not withstanding.

There are, however, a few experiments which seem to favor the one level model. These experiments concern the failure to obtain an adaptation effect when the adaptor and test were such that there should be an effect at the abstract auditory level. In one of these experiments, Bailey (1975) used two different [ba]-[da] place feature series. In the first, the place distinction was cued by the third formant transition. The second formant in this series was fixed as a slightly falling transition. In the second [ba]-[da] series, the place distinction was carried by the second formant transition. The stimuli in this series did not contain a third formant. Both the [ba]-fixed F2 and the [ba]-no F3 end-point syllables had an adapting effect on the [ba]-[da]-fixed F2 series. The [da]-fixed F2 adaptor did have a significant, although small, effect on the [ba]-[da]-no F3 test series. These results have been interpreted as supporting the one level auditory model of selective adaptation. However, they are not in conflict with the two level model. The [ba]-fixed F2 adaptor did not have a significant adapting effect on the [ba]-[da]-no F3 series. However, this would not be expected to produce a large adaptation effect on a [ba]-[da] place series lacking all F3 information. To explain this, we need to trace the effect of the [ba]-fixed F2 syllable as an adaptor according to the two level interpretation. The F3 component of the [ba]-fixed F2 adaptor would fatigue the rising transition detector in its particular frequency range. However, the fixed F2, since it is slightly falling, would be expected to affect the rise and fall detectors in this range about equally. If anything, it would affect the falling detector more than the rise detector. Thus, the information supplied to the integrative rise and fall detectors at the abstract level was somewhat ambiguous. The [ba]-fixed F2 adaptor would not affect any of the abstract, integrative level detectors to any great extent. When subjects were tested with a place series which was cued by F2 transitions (no F3) there would be essentially no effect. The detectors at the auditory level that are sensitive to the F2 transitions were hardly affected by the fixed F2 adaptor and no F3 was present in the test stimuli so there would not be much if any of an effect due to the auditory level detectors. The integrative level would similarly not reflect any shift as explained above. The small effect of the [da]-fixed F2 adaptor on the [ba]-[da]-no F3 series can be explained in a similar fashion. Consequently, the results of this experiment do not rule out a two level interpretation of selective adaptation to place.

A second experiment that has also been taken as evidence against the two level interpretation involved cross series adaptation between a [bi]-[di] and a [bu]-[du] place series. Bailey (1975) found that [bi] and [di] adaptors had no effect on a [bu]-[du] series and vice versa. According to the one level model, no adaptation would be expected for these cross series conditions. The formant transitions for the [bi]-[di] series were in a different spectral region from those of the [bu]-[du] series. However, before taking these results as evidence against adaptation occurring at an integrative level, a closer look at the stimuli used in Bailey's experiment is necessary. In the [bi]-[di] series, the second formant was always rising, even for the [di] syllables. Thus, the [di] syllable could not be expected to cause much adaptation at the integrative level because the pattern of formants does not fit that of either a general rising or general falling detector. Similarly, the [du] syllable would not affect the [bi]-[di] series. The transitions for the [du] are falling for both F2 and F3. However, the cues mediating the [bi]-[di] series do not follow this pattern. Since the second formant is always rising in the [bi]-[di] series, the adaptation of a general fall detector at the abstract, integrative level would not be expected to produce any adaptation effect on a [bi]-[di] series.

This explanation in terms of conflicting formant information will not suffice to explain the lack of effect with the [bi] and [bu] adaptors in the cross series conditions. Both the [bu] and [bi] syllables had rising second and third formant transitions. As a consequence, both syllables should have fatigued the integrative rising frequency detector at the abstract level.

Up to this point it has been implicitly assumed that the integrative detectors at the abstract auditory level respond uniformly to all inputs from the acoustic level detectors. However, this assumption need not hold. Rather, it seems plausible that the integrative detectors have a graded response depending upon the frequency range of the acoustic detectors from which they receive input. Thus, there is a distribution of response for the abstract level, based on the particular frequency range of the auditory level detectors from which input is received. In his experiment, Bailey (1975) used [bi] and [bu] adaptors. The second and third formant transitions for the syllable [bi] are at the upper end of the frequency range for an adult male voice. In contrast to this, the formant transitions for [bu], specifically the second and third formants, are at the lower end of the adult male voice frequency range. If these transitions represent the extremes over which an integrative rise detector responds, then perhaps neither would give rise to a maximal response. Thus, the adapting effect of either [bi] or [bu] should not be "maximal" at the abstract, integrative level. As a result, the cross series effects of a [bi] adaptor on a [bu]-[du] test series and vice versa would not be expected to be large. There would be no frequency specific, peripheral auditory level effect and a less than maximal integrative level effect in the cross-series conditions.

On the balance, the evidence for an abstract, integrative level of selective adaptation is suggestive but inconclusive. The data of Ganong (1975) seem to rule out a strict one level interpretation. However, the results of Bailey (1975) using the cross vowel series are very troublesome. They suggest that if adaptation is taking place at an integrative level, the effect is not large and does not generalize well across different vowels. Before a conclusion can be drawn as to how many levels of processing are affected by selective adaptation on the place feature further experimentation is necessary.

Adaptation of Other Phonetic Features in Consonants

In addition to place and voicing, several other features of various consonants have been studied using selective adaptation. Cole and Cooper (1975) investigated the voiced-voiceless distinction in a fricative series varying from [sa] to [za] to [da]. The primary acoustic cue that they investigated was the duration of friction. Their results were somewhat ambiguous because subjects had a hard time identifying stimuli as [za] and the [za] adaptor did not produce systematic category boundary shifts.

The manner of articulation feature has also received some attention. Several real-speech CV syllable series ranging from [m] to [b], [j] to [d], [v] to [b] and [f] to [b] were studied by Cole, Cooper, Singer and Allard (1975). They found selective adaptation when the end-point stimuli from their series were used as adaptors in all cases except one ([ba] did not adapt the [fa]-[ba] series), suggesting an auditory component in their results. Similarly, the syllables [di] and [ji] had appropriate adapting effects on a [ja]-[da] series. The one problem with the studies reported by Cole et al. (1975) is that natural speech stimuli were used. The CV continua studied were constructed by "cutting into" the initial consonant of an original spoken CV. However, no spectral analysis of these stimuli was reported. This makes a precise determination of adaptation at an auditory (spectral) level versus an abstract integrative level in their experiments impossible.

Cooper, Ebert & Cole (1976) and Diehl (1976) have also reported selective adaptation results for synthetic CV syllable series along a stop versus continuant dimension. Diehl (1976) used a series that ranged perceptually from [ba] to [wa]. The initial and final frequencies for the formant transitions for all stimuli in this series were identical. The duration of the initial formant transitions was varied from 6 to 66 msec. The [ba] adaptor moved the phonetic category boundary toward the stop ([ba]) end of the series and the [wa] adaptor moved the boundary toward the continuant ([wa]) end. The [ba] transitions and the [wa] transitions in isolation (no steady state vowel) were also used as adaptors and both produced the expected adaptation effect, although to a lesser degree than the full syllable adaptors. These results are very similar to those of Tartter & Eimas (1975) who did similar manipulations on the place dimension. As with the Tartter & Eimas study, the results of Diehl (1976) offer firm support for an early auditory level of adaptation but are only suggestive of adaptation at a later, more abstract level.

Cooper et al. (1976) used two synthetic series varying from [ba] to [wa] and [ga] to [ja] along the stop-continuant dimension. They found selective adaptation for both series when end-points from the test series were used as adaptors. Further, they used two [ga] adaptors on the [ba]-[wa] series. Both of these adaptors had formant transitions of 35 msec duration but differed in their envelope amplitudes. It should be noted that in the [ba]-[wa] series, a 35 msec duration of transition was identified as [wa] approximately 80% of the time. Despite this, both of the [ba] adaptors produced a significant movement of the category boundary toward [ba]. Cooper et al. (1976) argue that this result supports the involvement of a relatively abstract level of feature analysis in their results. The rate of transition seems to be analyzed in a place dependent fashion. The fact that the [ga] adaptors moved the category boundary toward [ba] despite the fact that the transition durations of the [ga] were the same as syllables identified as [wa] supports this conclusion. Thus, the adaptation effect for these [ga] syllables was occurring subsequent to some form of recalibration or retuning due to the place feature of the adaptor.

Care must be exercised, however, in interpreting the results of Cooper et al. (1976) and Diehl (1976). Their results could be due, in part, to the operation of some form of response bias mechanism (cf. Sawusch & Pisoni, 1973, 1976, as discussed earlier). Data on the possibility of response bias along the manner dimension of stop versus continuant is not yet available, despite the efforts of Cooper et al. (1976) to sort out sensory and response criterion effects in selective adaptation using a decision theory analysis. Cooper et al. (1976) used a TSD analysis on data from selective adaptation on a [ba]-[wa] series with the [ba] and [wa] end-point stimuli as adaptors. This analysis was based on the decision model for intensity proposed by Durlach & Braida (1969). The analysis revealed systematic changes in d' for various syllable pairings following selective adaptation. These changes were interpreted by Cooper et al. (1976) as reflecting basic sensory changes in the processing of the [ba]-[wa] series. However, before accepting this TSD analysis of selective adaptation or the conclusions of Cooper et al., the correspondence between the assumptions necessary for the Durlach & Braida (1969) model and the identification data for the [ba]-[wa] series should be checked. Some serious problems exist in their analysis of the data. For example, as part of their data analysis, Cooper et al. reported just noticeable difference (JND) estimates based on the d' values for the model. The JND for one group prior to adaptation was 160 msec of transition duration, based on the data for the first two stimuli. This extimate is clearly too large since stimuli 1 and 5, which differed by 20 msec of transition duration, were assigned reliably to different categories and hence should be near perfectly discriminable. The analysis used by Cooper et al. predicted a JND that was an order of magnitude too large. This discrepancy calls into question the applicability of the Durlach & Braida (1969) model to identification data from a speech series without additional evaluation. A check of the assumptions of the model, especially the equal variance, normal distribution assumptions, is necessary before accepting any of the conclusions of Cooper et al. (1976) on the role of response bias in selective adaptation.

The results from selective adaptation studies of features other than place and voicing have followed the same general pattern as that found for the place and voicing features. These data strongly indicate that adaptation affects a relatively early, auditory stage of processing. As yet, the data are inconclusive in determining whether or not adaptation occurs at an abstract, integrative level also. Further studies are needed in which the spectral properties of the adapting and test stimuli are systematically manipulated.

Where does selective adaptation occur: Overview.

The results of selective adaptation experiments on the place feature suggest a two level account of where adaptation occurs. The possible role of response bias in determining the place adaptation results seems to have been ruled out (Sawusch & Pisoni, 1976). Similarly, experiments on syllable position of the place feature (Ades, 1974b; Pisoni, 1975b) seem to rule out an abstract, position invariant phonetic interpretation based on something similar to the distinctive features of Chomsky & Halle (1968). The picture that has emerged is that of a frequency specific first stage (peripheral auditory analysis) followed by an integrative stage (abstract auditory analysis) which responds to particular patterns of auditory cues.

From the voicing results, however, a slightly different picture emerges. The evidence against a response bias account is similar to that for the place dimension. However, there is no firm evidence for an abstract, integrative site of adaptation. Rather, a model based on the acoustic cues (e.g., Summerfield, 1974, 1975) seems to be adequate to account for the present results.

The data on other cues are far from complete. Although the overall pattern of results for the manner feature is similar to that found for the place feature, there still exists the possibility that response bias may enter into the identification of consonantal features other than place and voicing. Further systematic research on features other than place and voicing is needed before a clear picture of where selective adaptation has its effect(s) will emerge.

CHAPTER III

The Role of Peripheral and Central Levels of Processing in Selective Adaptation on the Place Feature

The question of whether there is a central, integrative component in selective adaptation has been an important issue in much of the recent work using this paradigm. The experiments that bear directly on this issue have manipulated the spectral overlap and acoustic cue composition between adapting and test stimuli. With these types of manipulations, some attempt has been made to separate the contributions of two possible components in selective adaptation.

There is ample evidence for the presence of an auditory level in selective adaptation. Experiments using non-speech adaptors (Pisoni & Tash, 1975; Tartter & Eimas, 1975) have shown this beyond a reasonable doubt. The evidence for an abstract, integrative level is much less convincing. For the phonetic feature of voicing, only one experiment (Cooper, 1974c) is really suggestive of an integrative level of adaptation. Cooper (1974c) manipulated the duration and extent of the first formant transition between adaptor and test series. However, there is still the question as to whether all possible joint cues to voicing between the adaptor and test stimuli were eliminated in this experiment. Ades (1976), in a review of previous selective adaptation experiments, has attributed all of the voicing feature results (including those of Cooper, 1974c) to one early auditory level of processing and concluded that there was no strong evidence for two levels of adaptation on the voicing feature.

For the phonetic feature of place, the data are somewhat contradictory. On the one hand, experiments by Blumstein & Stevens (1975), Diehl (1975) and Ganong (1975) have manipulated the presence or absence of both burst and transition cues to place. Their results uniformly indicate that a burst cued adaptor will move the phonetic category boundary in a CV series that is cued only by transitions (Diehl, 1975; Ganong, 1975). Further, Blumstein & Stevens (1975) found that when both transition cues and appropriate bursts were combined together in an adaptor, the adapting effect was larger than when only transition cues were used. This was obtained for a CV test series which contained only transition cues, no bursts. Further, when inappropriate burst cues were added to syllables, (e.g. [g] burst to a transition cued [d]), the adaptation effect was greatly reduced. These results all suggest the presence of an abstract auditory level in selective adaptation, where the effects of bursts and transitions are integrated together (cf. Stevens, 1973).

However, it is possible to account for these results by using a one level model with one assumption (cf. Ades, 1976). The assumption is that bursts are processed by transition detectors. Consequently, both the bursts and the transitions would be processed by the same detectors. This explanation seems to be very plausible for the results of Diehl (1975). In this experiment, the burst cued [tɛ] adaptor had a burst frequency of 2665 Hz and the starting frequency for the third formant transition at the [dɛ] end of the [bɛ]-[dɛ] test series was 2726 Hz. The same detector could have been responding to the burst and the third formant transition. This argument is less plausible for the results of Ganong (1975) where the

the bursts and the formant transitions were widely separated in frequency (over 1000 Hz). Any detector that responded to both the burst and the transitions in Ganong's experiment would have to cover a frequency range of over 1000 Hz. This seems very unlikely for an early auditory level detector, especially in light of the fact that the critical bandwidth (for loudness summation) in this frequency region is on the order of 500 Hz (Zwicker, Flottorp & Stevens, 1957; Scharf, 1970). To insist that some auditory detector could indeed be responding to both the burst and the transitions is basically the same as postulating an integrative detector.

There are also results that are not consistent with an integrative level of selective adaptation. For example, Bailey (1975) performed several experiments in which the spectral overlap of the adapting and test stimuli was manipulated systematically. In one of these experiments two test series were used, a synthetic [bi]-[di] place series and a synthetic [bu]-[du] place series. The second and third formants for the [i] vowel series were very high in frequency, both over 2400 Hz. For the [u] vowel series, the steady state second and third formants were both under 2000 Hz. The initial formant transitions for the stops in these two series were also well separated in frequency. When the adaptor was an end-point of the test series, consistent category boundary shifts were found. The [bi] adaptor moved the [bi]-[di] category boundary toward [bi] and [di] adaptor moved the boundary toward [di]. Similarly, the [bu] adaptor moved the [bu]-[du] boundary toward [bu] and vice versa for the [du] adaptor. When the adaptor and test series had different vowels (i.e., [bi] adaptor, test [bu]-[du]), there was no adaptation effect. On the basis of these results, Bailey (1975) concluded that selective adaptation to place operated primarily in a spectrally specific fashion at an early auditory level of processing (i.e., where the speech input is represented as a neural spectrogram).

The present experiments were an attempt to resolve this conflicting set of results. On the one hand are the multiple cue experiments of Blumstein & Stevens (1975) and Ganong (1975) which support the operation of an abstract, integrative level during selective adaptation. The results of Bailey (1975), where the spectral overlap between adapting and test syllables was manipulated, support a one level interpretation of adaptation. Essentially no cross series adaptation was found when the adaptor and the test stimuli did not contain overlapping spectral properties.

In the present experiment, the spectral overlap between adapting and test syllables was systematically manipulated. The reason for manipulating the spectral overlap between adapting and test syllables in this experiment was to somehow bypass the peripheral auditory level and attempt to adapt the abstract integrative level. In order to accomplish this, the processing characteristics of the peripheral auditory level need to be well specified. In the model outlined in Chapter I (Figure 1), the processing at the peripheral auditory level was characterized as frequency specific. From work on auditory psychophysics, the concept of frequency specificity leads to the critical band (Scharf, 1970).

The critical band represents a frequency range over which a number of auditory phenomena take place. Within a critical bandwidth, the loudness of a complex sound (noise or multitone complex) is independent of frequency separation. However, once outside the critical band, loudness increases with increasing frequency separation (Zwicker et al., 1957). The critical band has also been implicated in studies of the masking of a tone by narrow band noise (Greenwood, 1961). As long as the noise lies completely within the same critical band as the tone, increasing the bandwidth of the noise makes detection of the tone more difficult. However, increases in the bandwidth of the noise masker outside of a critical bandwidth do not affect detection of the target tone (Greenwood, 1961). For a more extensive review of theoretical and emperical work on the critical band, the reader is referred to Scharf (1970).

The critical bandwidth has often been treated as the frequency bandwidth for early peripheral processing in the auditory system (see Scharf, 1970). As such, it may represent the bandwidth for intensity summation in the first stage of processing in Figure 1. If this is the case, then the early auditory detectors at the peripheral auditory level would receive as their input a matrix of intensity per critical bandwidth over time. Since the detectors at this level extract such basic information as the presence of formant transitions and this extraction is frequency specific, the critical bandwidth is a logical candidate for the frequency range over which these detectors respond. In the present experiments, several adapting stimuli were constructed such that all of their formant frequencies were separated from the formants of the test series by at least a critical bandwidth. Consequently, when one of these stimuli is employed as an adaptor, it should engage detectors at the frequency specific peripheral auditory level that are not engaged by the test series.

The test stimuli for the present experiment varied perceptually from [bae] to [dae]. Schematized spectrograms of the end-points of this series are shown in Figure 4. These two stimuli were also used as adaptors in the present study. Two other stimuli, illustrated with dashed lines in Figure 4, were also constructed. These stimuli had the same pattern of formant trajectories (for the first through third formants) as the end-points from the test series. However, the actual center frequencies for these stimuli were higher (by approximately one and one-half critical bandwidths) than their test series counterparts. Thus, these [bae] "high" and [dae] "high" adaptors should engage different frequency specific detectors than those engaged by the "low" frequency test series. Any adaptation found with the high adaptors should reflect an abstract, integrative level of processing that responds to a formant pattern (cf. Stevens, 1973) and not to particular absolute values of formant frequencies.

Experiment 1

The first experiment was performed using the stimuli outlined above to test for the presence of selective adaptation effects of the high stimuli on the low test series. Such a result would indicate that adaptation was occurring at an abstract, integrative level, subsequent to early, frequency specific auditory analysis. From the results of Bailey (1975) discussed earlier, there was some reason to doubt whether such adaptation would indeed occur.

Method

Subjects. The subjects for this experiment were 24 undergraduates at Indiana University enrolled in an introductory psychology course.

Figure 4. Schematic spectrograms of the endpoints of the [bae]-[dae] test series (solid black formants) and their corresponding "high" syllables (dashed outlines).

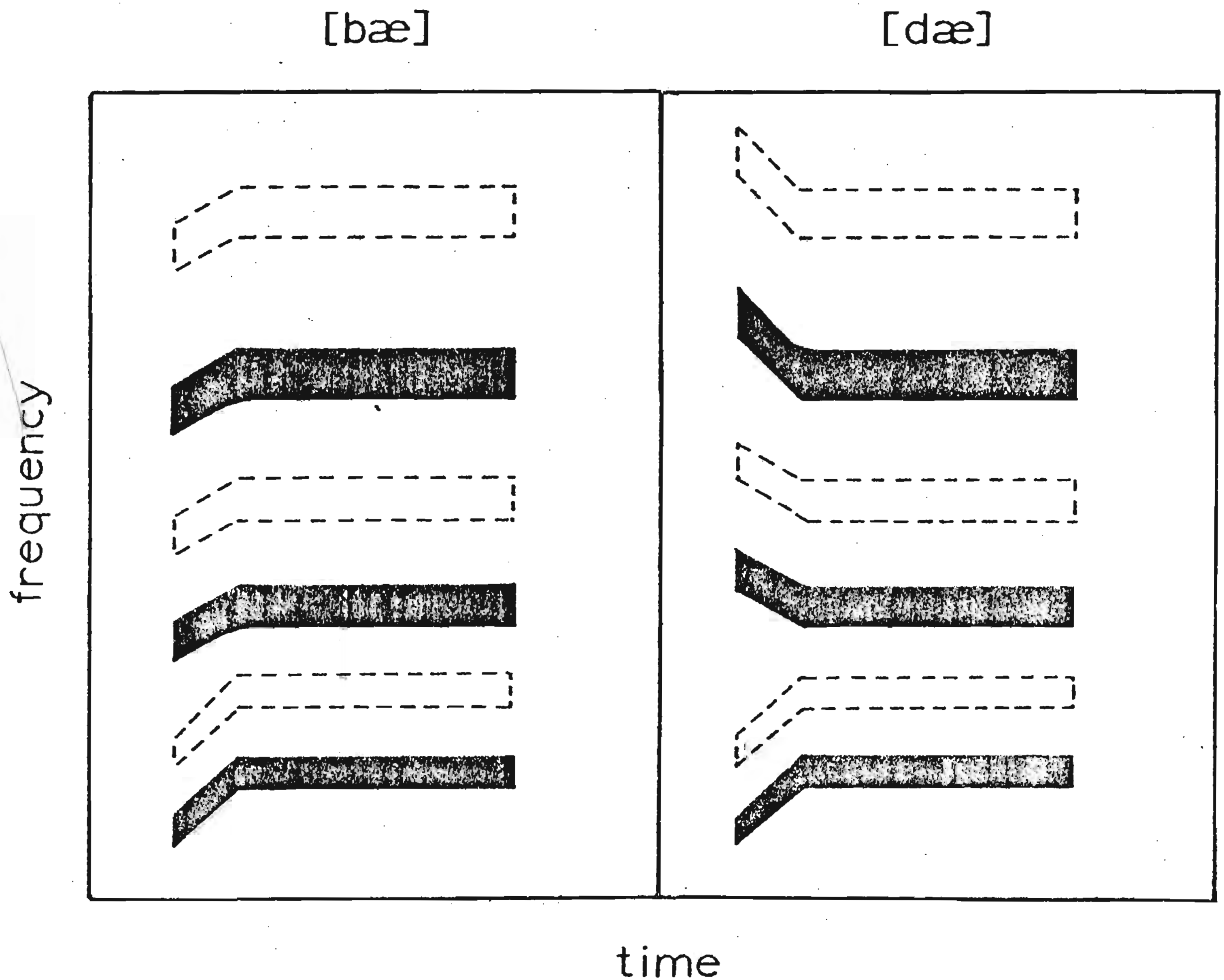


Figure 4.

They participated as part of a course requirement. All subjects were right-handed, native speakers of English with no known history of either speech or hearing disorder. They were divided into four groups of six subjects each.

Stimuli. The stimuli were five formant synthetic CV syllables that were prepared on the OVE IIId serial speech synthesizer at Indiana University. The test stimuli consisted of one series of ten CV syllables that ranged perceptually from [bae] to [dae]. These stimuli varied in the starting frequencies for the second and third formant transitions from 1234 Hz (F2) and 2263 Hz (F3) for the [bae] end to 1600 Hz (F2) and 2934 Hz (F3) for the [dae] end in nine approximately equal steps. The first formant transition had a starting frequency of 282 Hz for all ten stimuli. The duration of the formant transitions was 45 msec, followed by a 205 msec steady state vowel [ae]. The vowel had formant center frequencies of 672, 1425, and 2468 Hz for the first through third formants respectively. The fourth and fifth formants were fixed at 3500 Hz and 4000 Hz. This series will be referred to as the "low" series. The fundamental frequency was the same for all ten syllables and fell linearly from 154 Hz to 115 Hz over the 250 msec syllable duration.

Six other stimuli were also constructed with the same formant patterns as the end-point stimuli of the low test series. Two of these stimuli, referred to as [bae]-high and [dae]-high, had formant frequencies that were logarithmically scaled upward from the [bae]-low and [dae]-low end-point stimuli. The [bae]-high syllable had formant transitions with starting frequencies of 518 Hz, 1600 Hz, and 3108 Hz for F1 through F3 and the [dae]-high syllable had formant transition starting frequencies of 518 Hz, 2074 Hz, and 4031 Hz for F1 through F3 respectively. Again, all transitions were 45 msec in duration and followed by a 205 msec vowel [ae]. For the high syllables, the vowel had formant center frequencies of 924 Hz, 1848 Hz and 3390 Hz for F1 through F3 respectively. The fourth and fifth formants were also fixed at 3500 Hz and 4000 Hz respectively. The fundamental frequency contour was identical to that of the low series stimuli.

The last four stimuli that were prepared were "chirps" representing the formant transitions of the two high syllables and the end-points of the low syllable series. These chirps were prepared by removing the final 205 msec steady state vowel from each of the full syllables. These four chirps contained the full transitional information of the syllables, but no steady state vowel portion.

Procedure. All experimental events were controlled by a PDP-11 computer. The stored parameter codes for the OVE were synthesized in real time and presented binaurally through Telephonics (TDH-39) matched and calibrated headphones to subjects. The stimuli were presented at a level of 80 dB SPL for a steady state calibration vowel [ae].

The experiment consisted of two 1-hour sessions that were run on consecutive days. Subjects were run in groups of three. At the beginning of the first day, all subjects listened to a fifty trial practice sequence with each of the ten test syllables occurring five times in random order. Next, subjects listened to a 100 trial identification test, with each syllable occurring ten times in random order. Subjects were instructed that they would hear synthetic speech sounds approximating the syllables [bae] and [dae] and were to identify each syllable as either [bae] or [dae]. Following their identification response, subjects were instructed to enter a rating response, indicating how certain they were that they had identified the syllable correctly. A copy of the four point scale was present in front of subjects at all times and is also shown in Table 2. Subjects entered their responses by pushing the appropriate button on a response box in front of them. On the second day, a second 100 item identification series was presented.

Following the identification sequence(s) on each day, two adaptation sequences were presented. Eight different adapting stimuli were used: the [bae]-low and [dae]-low end-points from the test series and their associated chirps ([bae]-low chirp and [dae]-low chirp) and the [bae]-high and [dae]-high syllables and their associated chirps. The first group of subjects received the [bae]-low adaptor on one day and the [bae]-high adaptor on the other day. The order of presentation of adaptors was counterbalanced across days, half of the six subjects getting one order and half the other. The other three groups received a similar pairing of adaptors ([dae]-low & [dae]-high; [bae]-low chirp & [bae]-high chirp; [dae]-low chirp & [dae]-high chirp), again with the order of presentation across days counterbalanced within each group. In all cases, subjects were informed as to the nature of the repeated (adapting) stimulus that they would be listening to.

The adapting stimulus was presented for approximately 45 sec (75 repetitions with a 300 msec interrepetition interval). After each period of adaptation, eight of the test syllables were presented in random order for identification by subjects using the procedure outlined above. Twenty of these adaptation sequences were run on each day. Thus, by the end of

Table 2

Four point rating scale used by subjects in rating the accuracy of their identification responses

Response	Rating
+++	Positive response was correct
++	Probable response was correct
+	Possible response was correct
-	Guess

the experiment, each subject provided 16 adapted responses to each of the ten stimuli for each of their two adaptors.

Results and Discussion

Rating functions for each subject were obtained by expanding the two separate category responses with their four rating categories into an eight point rating scale. On this scale a 1 denotes a positive rating for a "B" response. On the other end, an 8 denotes a positive rating for a "D" response. The data for the two groups that received full syllable adaptors are shown in Figure 5. The phonetic boundary was determined by a computer program that located the point along the stimulus continuum corresponding to a rating of 4.5 by linear interpolation.

The data in the left hand panel of Figure 5 are for the [bae]-low and [bae]-high adapted group. Both adaptors produced a significant change in the category boundary toward the [bae] end of the series ($\underline{t}(5) = 6.35$, $\underline{p} < .001$ for the [bae]-low adaptor; $\underline{t}(5) = 3.77$, $\underline{p} < .01$ for the [bae]-high adaptor using one-tailed, correlated t-tests).[3] The difference between the two adapting syllables was also significant ($\underline{t}(5) = 2.31$, $\underline{p} < .05$). The group receiving the two [dae] syllable adaptors showed similar results (right hand panel, Figure 5). Both the [dae]-low and [dae]-high adaptors produced a significant shift in the category boundary toward [dae] and the difference between the adaptors was also significant ($\underline{t}(5) = 4.42$, $\underline{p} < .005$ for the [dae]-low adaptor; $\underline{t}(5) = 2.51$, $\underline{p} < .05$ for the [dae]-high adaptor; $\underline{t}(5) = 2.66$, $\underline{p} < .025$ for the difference between adaptors).

Figure 5. Unadapted (solid circles) and adapted rating functions for the two full syllable adapted groups. On the left are the [bae]-low (open circles) and [bae]-high (open triangles) adapted functions. On the right are the [dae]-low (open circles) and [dae]-high (open triangles) adapted functions.

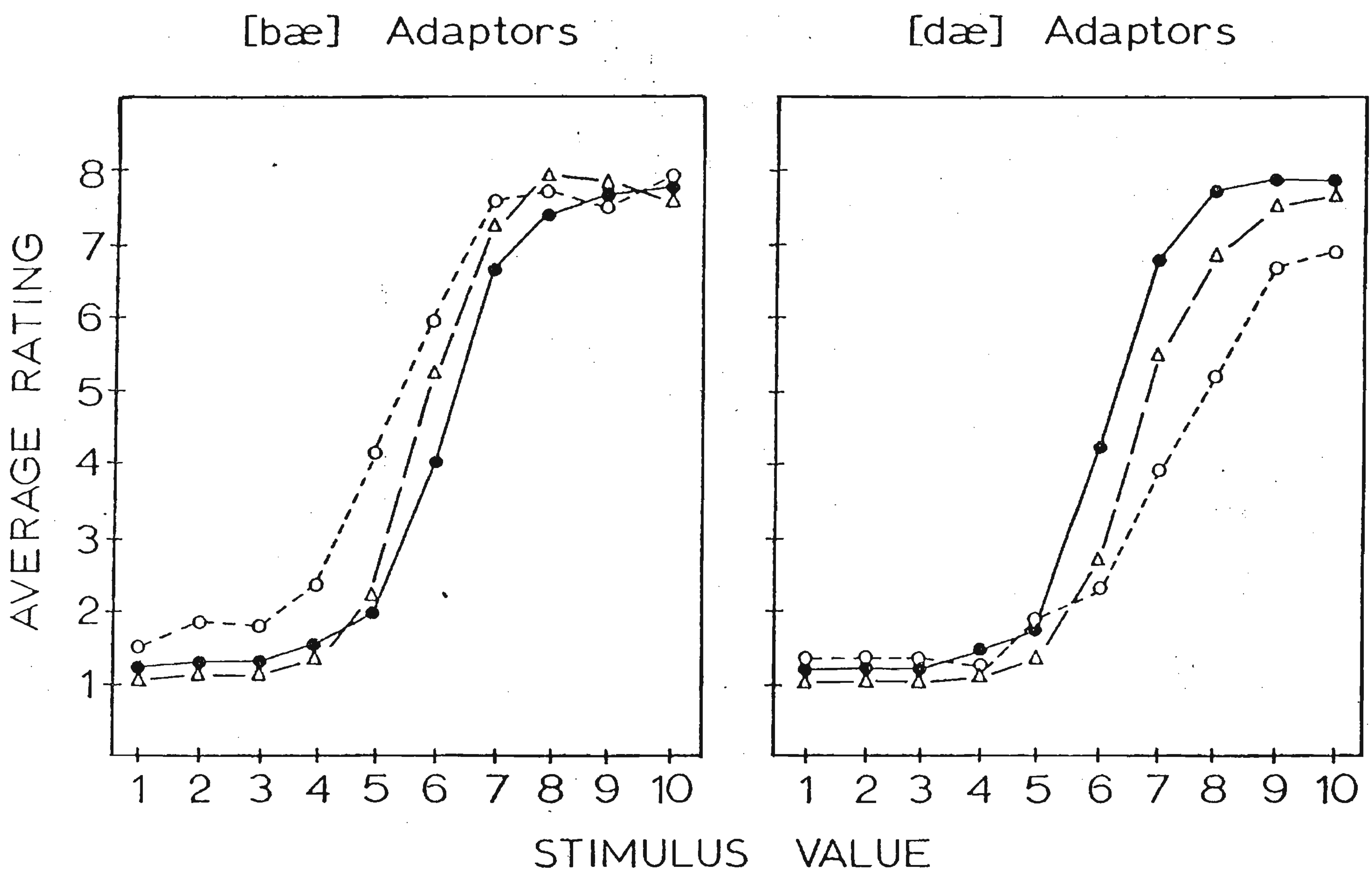

[bæ] Adaptors
[dæ] Adaptors
AVERAGE RATING
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
8
9
10
STIMULUS VALUE

Figure 5.

The data for the chirp adaptors were remarkably similar to that of the full syllables. The left side of Figure 6 contains the rating functions before and after adaptation with the [bae]-low and [bae]-high chirps. The [bae]-low chirp produced a significant shift in the phonetic category boundary toward [bae] ($t(5) = 3.73$, $p < .01$). However, the shift for the [bae]-high chirp failed to reach significance ($t(5) = 1.61$, $.05 < p < .1$). The difference between adaptors was also highly significant ($t(5) = 11.79$, $p < .001$). Both of the [dae] chirp adaptors produced a significant movement in the category boundary ($t(5) = 7.83$, $p < .001$ for the [dae]-low chirp and $t(5) = 4.41$, $p < .005$ for the [dae]-high chirp). The difference between adaptors was also significant ($t(5) = 4.11$, $p < .005$).

The high adaptors generally produced a substantial adapting effect on the low test series, though not as much as the low adaptors which were drawn from the test series. A comparison of the relative magnitudes of adaptation produced by the various adaptors appears in Table 3. Overall, the high adaptors were about 33-40% as effective as the low adaptors. The magnitude of this effect is in close agreement with the results of Ganong (1975) where the burst cued adaptors produced about 33% as much adaptation as the transition cued adaptors (on a transition cued test series). These results would seem to rule out a one level, auditory interpretation of selective adaptation. The adaptation effects of the high adaptors indicate the presence of an integrative level of processing where the particular frequency range of the transitions is irrelevant. At the same time, however, the consistently larger adaptation effects of the low syllables and chirps reinforce the evidence

Figure 6. Unadapted (solid circles) and chirp adapted rating functions. On the left are the [bae]-low chirp (open circles) and [bae]-high chirp (open triangles) adapted functions. On the right are the [dae]-low chirp (open circles) and [dae]-high chirp (open triangles) functions.

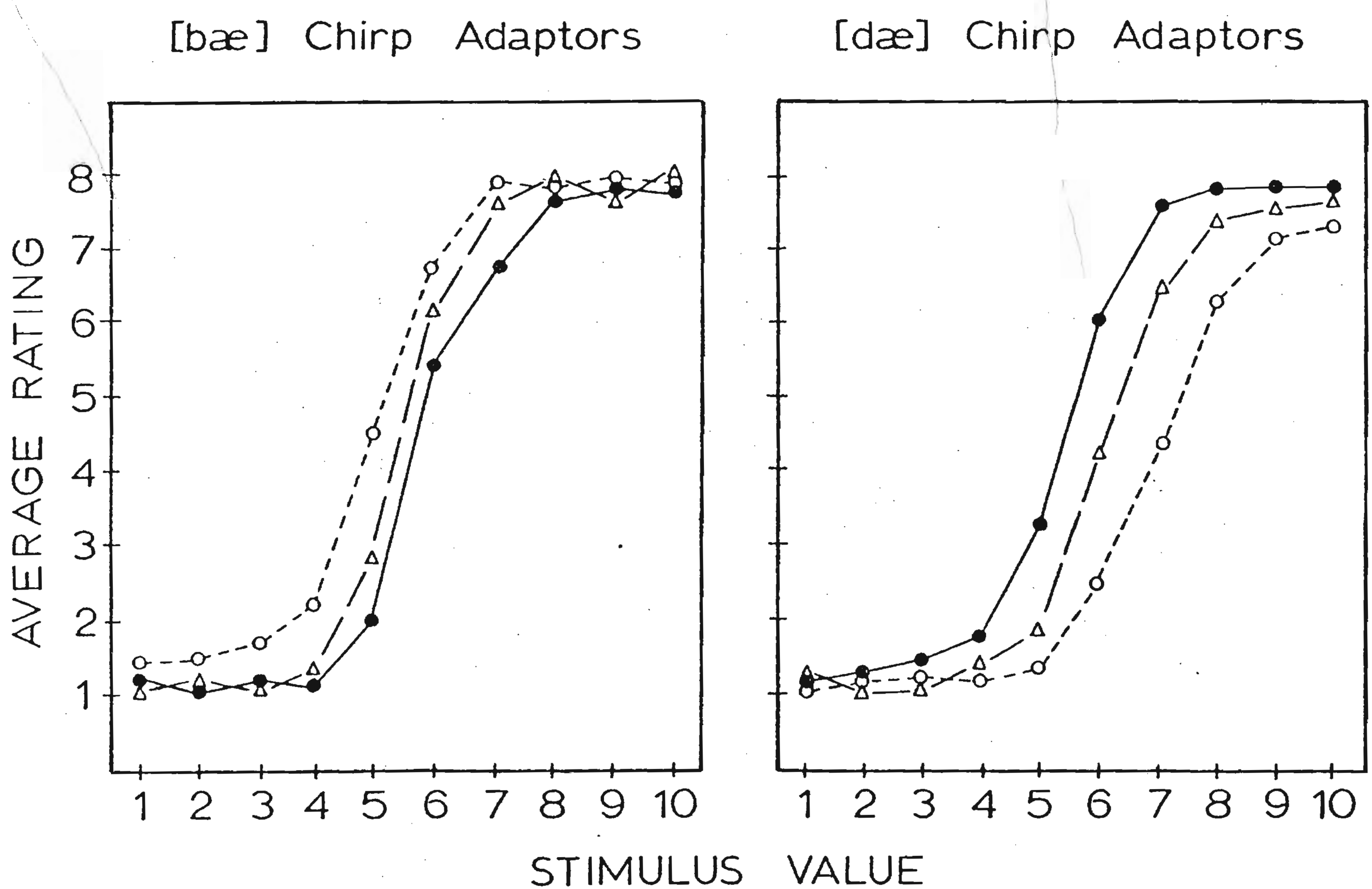

[bæ] Chirp Adaptors
[dæ] Chirp Adaptors
AVERAGE RATING
1
2
3
4
5
6
7
8
STIMULUS VALUE
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10

Figure 6.

Table 3

The absolute and relative phonetic boundary shifts for the eight adaptors of Experiment 1 (in stimulus units).

	[bae]-low	[bae]-low chirp	[bae]-high	[bae]-high chirp	[dae]-low	[dae]-low chirp	[dae]-high	[dae]-high chirp
Shift	1.13	.90	.56	.39	-1.52	-1.49	-.44	-.58
Percent	100%	79.6%	49.6%	34.5%	100%	98%	28.9%	38.2%

favoring a frequency specific, auditory level. Thus, the picture that emerges from these results is that of a two process system with both frequency specific and integrative components.[4]

One other aspect of the data should also be discussed here. The chirp adaptors generally produced as large of an adapting effect as their full syllable counterparts. None of the adaptation differences between a syllable and its chirp counterpart were significant (using two-tailed, independent t-tests). In one instance ([dae]-high), the chirp actually produced a slightly larger effect than the full syllable, in another case they were approximately equal ([dae]-low) and for the [bae] adaptors, the syllables generally produced slightly more adaptation. Thus, the vowel does not seem to contribute to the adaptation effect. It should be noted, however, that the subjects were told the exact nature of the chirps and no one reported any difficulty in listening to them as speech. The roughly equal effectiveness of chirp and full syllable adaptors indicates that the mechanisms being adapted are indeed tuned to the rapid formant transitions and appear to be unaffected by a steady state vowel. The question of whether the integrative aspects of the selective adaptation phenomena are linguistic and specific to speech cannot be answered at this point, however, because subjects reported listening to the chirps as speech. The next two experiments were designed, in part, to explore whether or not the integrative site of adaptation is indeed speech specific and where in the perceptual system (centrally or peripherally) this integrative level is located.

Experiment 2

The second experiment was designed as a further test of the properties of the frequency specific auditory stage versus a more abstract, integrative stage. The basic question pursued here was whether the integrative process was centrally or peripherally located. Previous experiments designed to answer this question have produced varied results. For example, Eimas et al. (1973) found that adaptation and testing in opposite ears (dichotic presentation) produced as much shift in a [da]-[ta] category boundary as adapting and testing in the same ear (monaural presentation). This 100% interaural effect was interpreted as evidence for a central site for adaptation on the phonetic feature of voicing.

Ades (1974a) used a similar procedure on the phonetic feature of place. However, Ades found approximately 55% interaural transfer. To account for this interaural transfer and other results, Ades (1974a) offered two models, both of which were previously shown in Figure 2 and have been discussed previously. Both of these models could, in principle, explain the data of Ades (1974a). The less than 100% interaural transfer results either from the adaptation of separate, monaural auditory channels in the two level model or from the difference in the contralateral and ipsilateral pathways in the bilateral, one level model. In both models, a syllable presented to one ear takes a different path through the processing system than a syllable presented to the other ear. Hence, presenting an adaptor in one ear and testing in the other will yield less adaptation than adapting and testing in the same ear. This is precisely the result Ades (1974a) found for a place CV series.

In order to distinguish between these two models, it is necessary to find an adapting stimulus which yields 100% interaural transfer from monaural to dichotic adaptation and testing. The two level model could readily account for 100% transfer by assuming that the adapting and test syllables engaged different peripheral auditory detectors but a common integrative detector. The bilateral model cannot account for 100% transfer since it has no central component. Experiment 2 was designed to test for the presence of a central component in selective adaptation. From the results of Experiment 1, the high syllables and chirps seem to offer an adaptation and test situation which could bypass the peripheral auditory analyzers engaged by the low test series. If this is the case and the integrative level is centrally located, then the high adaptors should produce interaural transfer of close to 100% in moving from monaural to dichotic presentation of adaptor and test series.

Method

Subjects. Subjects were 48 undergraduates at Indiana University who participated as part of a course requirement. All were right handed, native speakers of English with no known history of any speech or hearing disorder. None of the subjects had any previous experience with synthetic speech. They were divided into eight groups of six subjects each.

Stimuli. The same ten syllable [bae]-[dae] test series and the same eight adapting stimuli that were used in Experiment 1 were used in this experiment.

Procedure. The stimuli were recorded on magnetic tape using a PDP-11 computer to control all of the timing. These tapes were later reproduced on an Ampex AG-500 tape deck and presented to subjects through

Telephonics (TDH-39) matched and calibrated headphones. The stimuli were presented at a level of 70 dB SPL for a steady-state calibration vowel [ae].

The experiment consisted of two 1-hour sessions that were run on consecutive days. All test syllables were presented in the right ear. At the beginning of the first day, subjects were presented with a fifty trial practice test during which each syllable occurred five times in random order. Following that an identification tape with 100 syllables, each of the ten test syllables occurring ten times in random order, was presented. Subjects used the same two category identification response followed by a four point rating response that was employed in Experiment 1. Again, a copy of the rating scale was present in front of subjects at all times. Subjects wrote their responses to each syllable in specially prepared booklets.

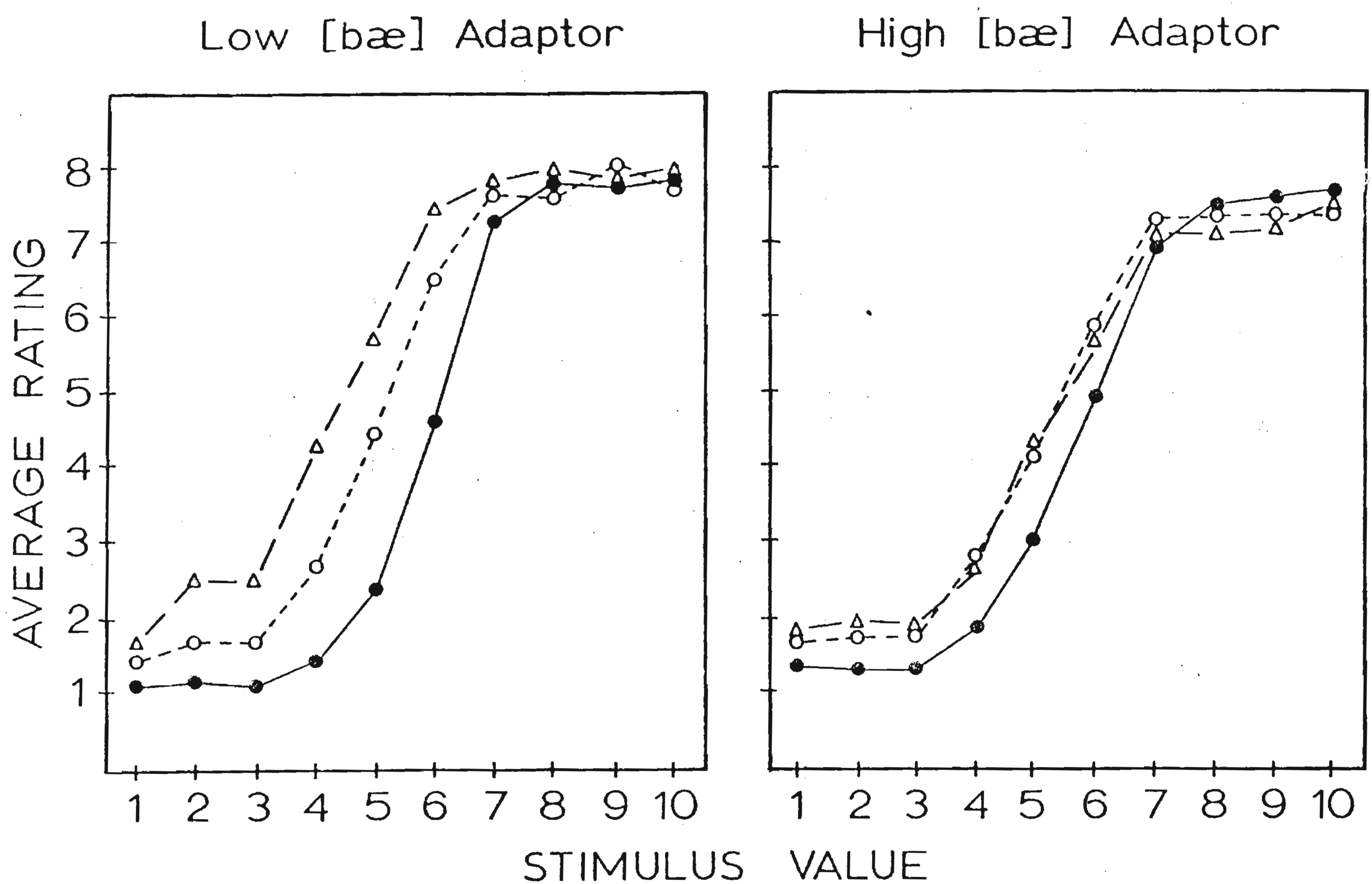
Following the identification sequence(s) on each day, two adaptation sequences were presented. Each of the eight groups listened to a different adapting syllable. Within each group, half of the subjects listened to the adaptor in the same (right) ear as the test syllables on the first day and in the opposite (left) ear on the second day. The other half received the reverse order of adaptor-ear presentation. In all cases, subjects were informed as to the nature of the adapting syllable and which ear it would be presented to. The number of repetitions of the adaptor and presentation of test stimuli were identical in all other respects to that used in Experiment 1.

Results and Discussion

The results for the low adaptors were very similar to those previously reported by Ades (1974a). The unadapted, monaurally adapted and dichotically adapted rating functions for the [bae]-low syllable adaptor are shown in the left hand panel of Figure 7. Both adaptation conditions produced a significant shift in the category boundary. The monaural condition produced a significantly greater effect than the dichotic condition ($\underline{t}(5) = 5.93$, $\underline{p} < .005$ for the monaural condition; $\underline{t}(5) = 5.00$, $\underline{p} < .005$ for the dichotic condition; $\underline{t}(5) = 3.25$, $p < .025$ for the difference). In contrast, the [bae]-high syllable produced virtually the same effect under both presentation conditions as shown in the right hand panel of Figure 7. The dichotic and monaural adaptation conditions both produced a significant shift in the category boundary, moving it toward [bae] ($\underline{t}(5) = 3.171$, $\underline{p} < .025$ for monaural adaptation and $t(5) = 5.864$, $\underline{p} < .005$ for dichotic presentation, and $t(5) = 0.03$, $\underline{p} > .4$ for the difference).

Similar patterns of results were found with the [dae]-low and [dae]-high syllable adaptors as shown in Figure 8 and with both sets of chirp adaptors as shown in Figures 9 and 10. With the [dae]-low syllable adaptor, both monaural and dichotic presentation produced a significant shift in the category boundary toward [dae] and the difference between them was also significant ($\underline{t}(5) = 3.161$, $\underline{p} < .025$ for monaural; $\underline{t}(5) = 3.72$, $\underline{p} < .025$ for dichotic; $\underline{t}(5) = 2.02$, $\underline{p} < .05$ for the difference). As with the [bae]-high syllable adaptor, the [dae]-high syllable produced nearly identical monaural and dichotic adaptation effects. Both adaptors had a significant effect in moving the [bae]-[dae] boundary toward [dae] ($\underline{t}(5) = 6.24$, $\underline{p} < .001$ for monaural; $\underline{t}(5) = 4.19$, $p < .005$ for dichotic; and $t(5) = -.61$, $\underline{p} > .25$ for the difference).

Figure 7. Unadapted (solid circles) and [bae] adapted rating functions. On the left are the [bae]-low dichotic (open circles) and monaural (open triangles) adapted functions. On the right are the [bae]-high dichotic (open circles) and monaural (open triangles) adapted functions.



Low [bæ] Adaptor
High [bæ] Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 7.

Figure 8. Unadapted (solid circles) and [dae] adapted rating functions. On the left are the [dae]-low dichotic (open circles) and monaural (open triangles) adapted functions. On the right are the [dae]-high dichotic (open circles) and monaural (open triangles) adapted functions.

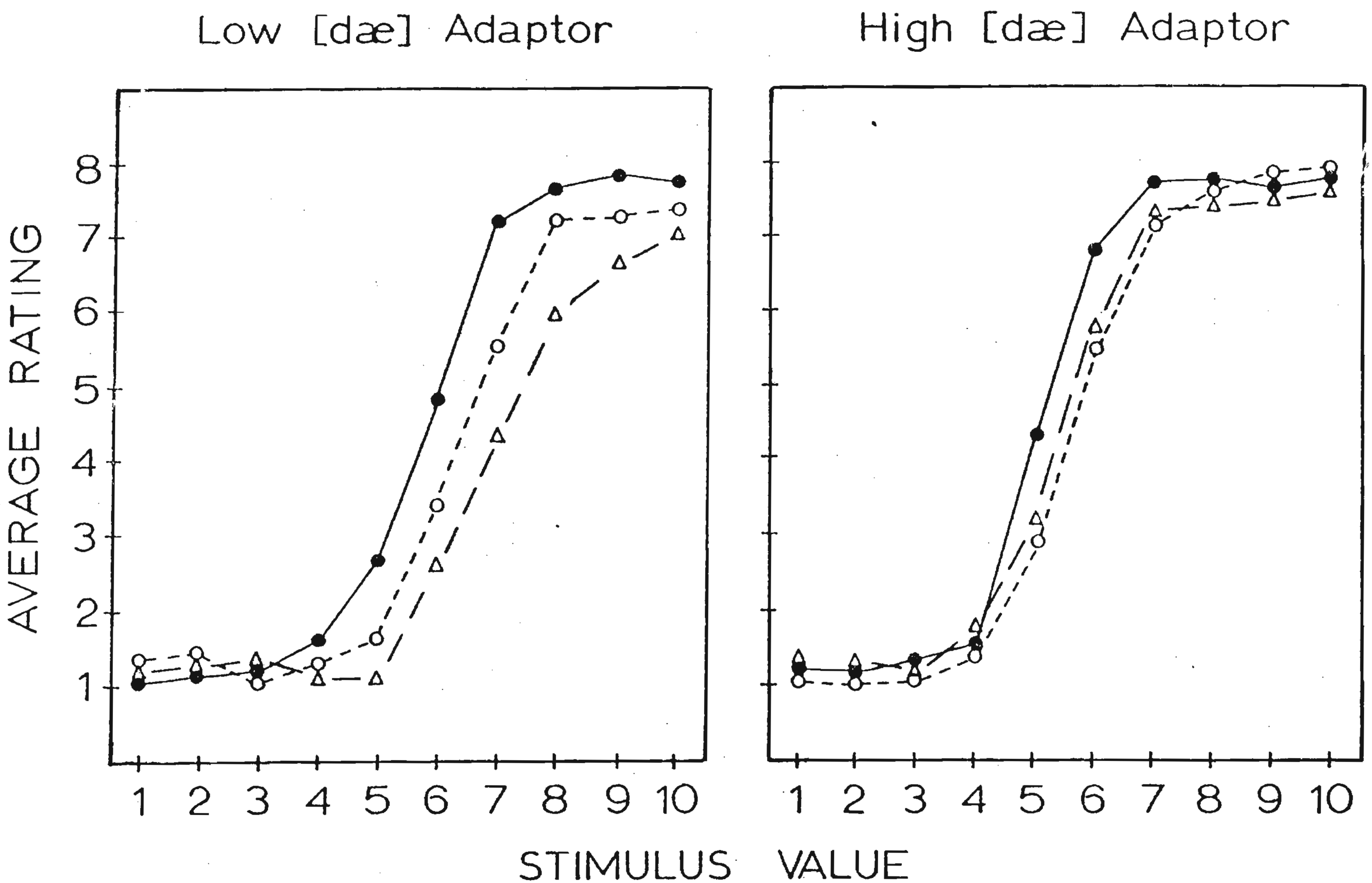

Low [dæ] Adaptor
High [dæ] Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 8.

Figure 9. Unadapted (solid circles) and [bae] chirp adapted rating functions. On the left are the low chirp dichotic (open circles) and monaural (open triangles) adapted functions. On the right are the high chirp dichotic (open circles) and monaural (open triangles) functions.

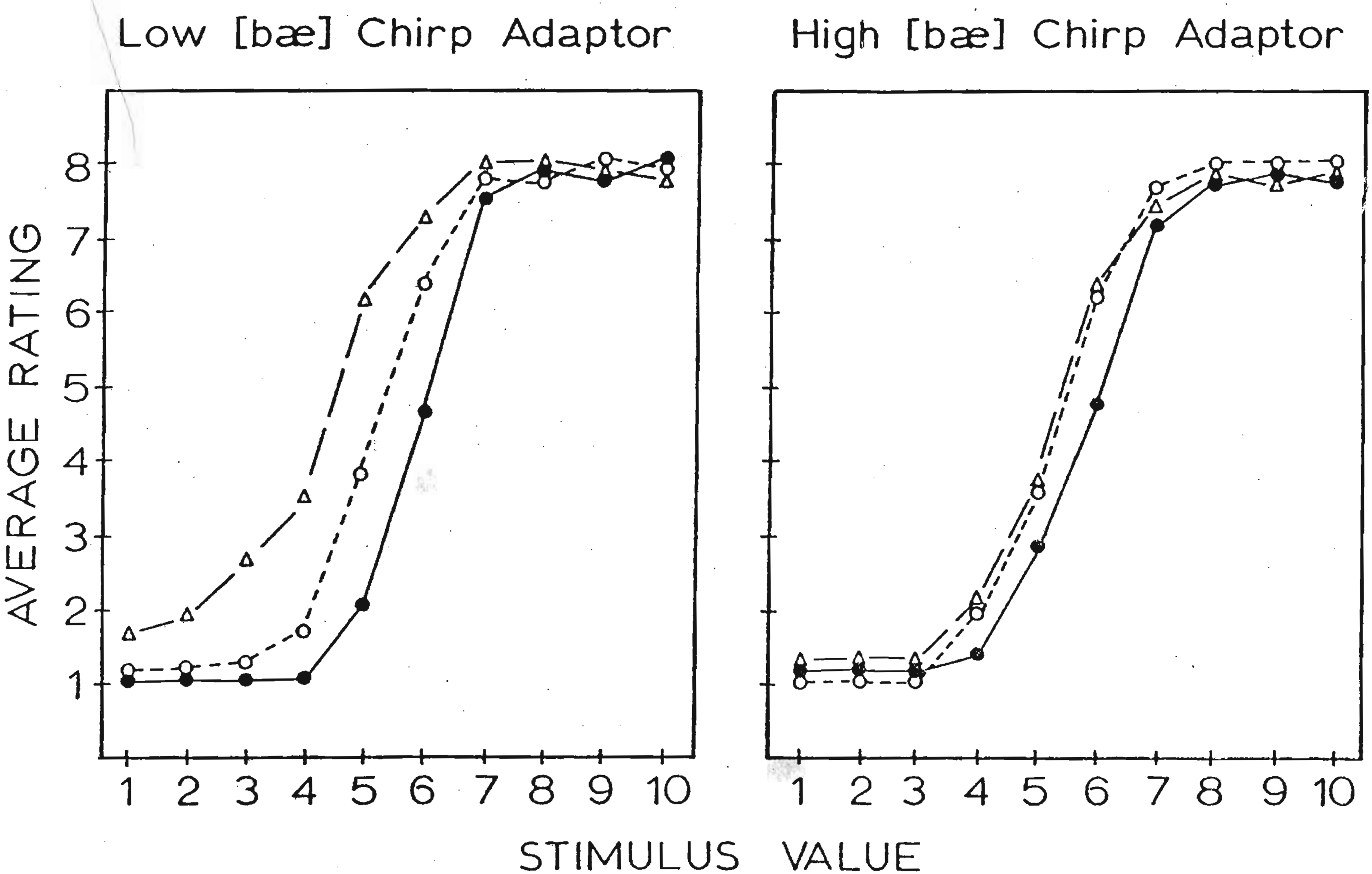
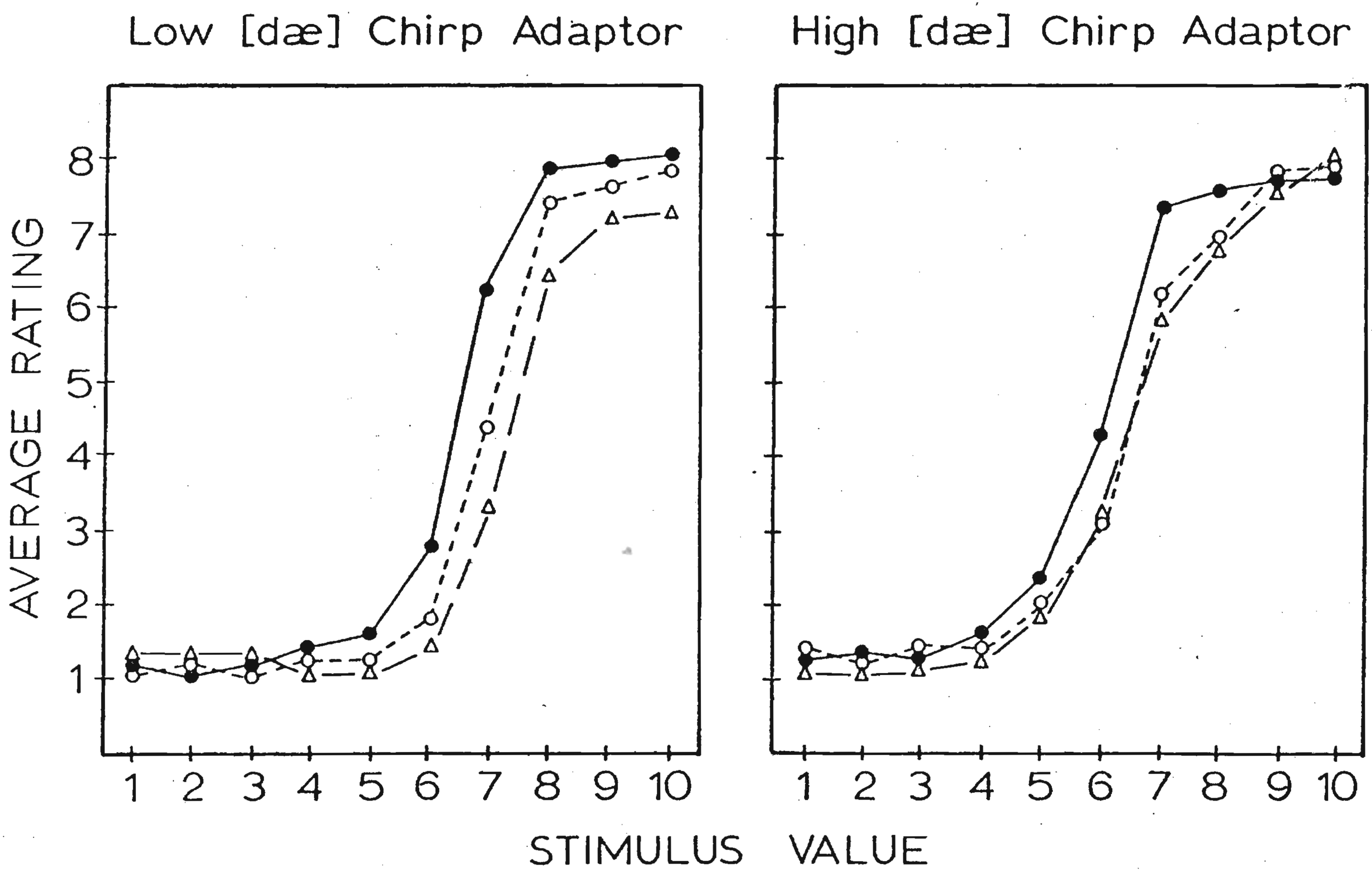

Low [bæ] Chirp Adaptor
High [bæ] Chirp Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 9.

Figure 10. Unadapted (solid circles) and [dae] chirp adapted rating functions. On the left are the low chirp dichotic (open circles) and monaural (open triangles) adapted functions. On the right are the high chirp dichotic (open circles) and monaural (open triangles) functions.



Low [dæ] Chirp Adaptor
High [dæ] Chirp Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 10.

The left hand panels of Figures 9 and 10 show the results for the [bae]-low chirp and [dae]-low chirp adaptors respectively. As with the full syllables, monaural presentation caused a larger boundary shift than dichotic presentation. All of the boundary shifts and the monaural-dichotic differences were significant ($t(5) = 5.08$, $p < .005$ for [bae]-low chirp monaural; $t(5) = 6.60$, $p < .001$ for [bae]-low chirp dichotic; $t(5) = 6.20$, $p < .001$ for their difference; $t(5) = 4.78$, $p < .005$ for [dae]-low chirp monaural; $t(5) = 2.49$, $p < .05$ for [dae]-low chirp dichotic; $t(5) = 2.33$, $p < .05$ for their difference).

The average rating functions for the [bae]-high chirp and the [dae]-high chirp adaptors are shown in the right hand panels of Figures 9 and 10. As with the full high syllable adaptors, there was complete transfer of adaptation from monaural to dichotic presentation. One adaptation condition, the [dae]-high chirp presented monaurally, failed to reach significance although the shift was in the expected direction. All other adaptation conditions produced a significant shift in the category boundary in the expected direction ($t(5) = 6.83$, $p < .001$ for the [bae]-high chirp monaural; $t(5) = 13.43$, $p < .001$ for the [bae]-high chirp dichotic; $t(5) = 1.19$, $p > .25$ for the difference; $t(5) = 1.76$, $.05 < p < .1$ for the [dae]-high chirp monaural; $t(5) = 4.80$, $p < .005$ for the [dae]-high chirp dichotic; and $t(5) = .143$, $p > .4$ for their difference).

The average boundary shifts for both monaural and dichotic adaptation conditions are shown in Table 4 for all eight adaptors. The percentage of interaural transfer is also given. There is a clear difference of 30% between the highest transfer for a low adaptor (the 58%

Table 4

Category boundary shifts (in stimulus units)

for both dichotic and monaural adaptation and percent interaural transfer

	Adaptor							
	[bae]-low	[bae]-low chirp	[dae]-low	[dae]-low chirp	[bae]-high	[bae]-high chirp	[dae]-high	[dae]-high chirp
Monaural Shift	1.65	1.71	-1.42	-.86	.82	.65	-.41	-.29
Dichotic Shift	.77	.64	-.82	-.39	.82	.57	-.49	-.27
Percent Transfer	47%	37%	58%	45%	100%	88%	120%	93%

of the [dae]-low syllable) and the lowest transfer for a high adaptor (the 88% of the [bae]-high chirp). Overall, the average interaural transfer for the low adaptors was 47%. This is slightly lower than the 55% found by Ades (1974a). However, Ades tested transfer from both left to right ears and right to left. The present study involved only left to right transfer. The average left to right transfer found by Ades was 49% which is in very close agreement with the 47% transfer found in the present study.

In contrast to the less than 50% transfer found with the low adaptors, the high adaptors yielded an average transfer of a full 100%. Consequently, the one level bilateral model proposed by Ades (1974a) can be safely rejected. The results of the present experiment, taken together with the findings of the first experiment, provide strong support for the presence of a centrally located integrative component in selective adaptation as well as a peripheral frequency specific component.

One other result from this experiment should also be noted. All four of the low adaptors in this experiment shifted subjects rating responses within the category of the adaptor as well as at the category boundary. For example, in Figure 7, the [bae]-low adapted function for monaural presentation shows a shift in the ratings for all of the [bae] syllables, as well as the boundary stimuli. For all of stimuli 1 through 7, the average rating responses showed a significant shift toward higher (more "d" like) ratings. This type of change in a rating response within a phonetic category has previously been reported by Sawusch (1976). The interesting point to note here is that none of the high adaptors produced a significant within series shift in rating responses. In contrast, all of the low series adaptors, when presented monaurally, produced a shift in the rating responses for the end of the test series from which the adaptor was drawn.[5] When the low adaptors were presented dichotically, the within category shifts were inconsistent and did not generally reach statistical significance.

This consistent difference between monaural and dichotic presentation and between the high and low adaptors may be taken as further evidence for the involvement of two separate and distinct levels of processing in selective adaptation to place. These results further suggest that the processing mechanisms at these two levels may undergo qualitatively different changes during selective adaptation. This topic of how selective adaptation affects a particular level of processing will be considered in more detail in a later chapter.

From Experiments 1 and 2, it is clear that there are two levels of processing that are affected by selective adaptation. The first is a frequency specific peripherally located component that will be referred to as the peripheral auditory level. The second component is centrally located and integrates over a broad frequency range. The terms "abstract auditory" or "integrative" will be used to describe this level. Some previous discussions of speech perception which have included this type of distinction have labeled the abstract level as phonetic and formulated this level in terms of a set of decision rules, rather than feature detectors (Pisoni, 1975; Pisoni & Sawusch, 1975). If the integrative level is indeed phonetic, then a formulation in terms of decision rules would seem appropriate. These rules would allow for various auditory features to be combined in a non-independent (non-additive) fashion. Such a set of rules does appear to be needed to resolve the invariance problem that was discussed earlier. An alternative formulation, essentially similar to that outlined in Chapter II, is to resolve the invariance problem over a series of stages. The integrative level would be sensitive to relative order of auditory cues within the syllable. For instance, a "b" in initial position would be extracted differently than a "b" in final position. The operations at this level could be formulated as abstract property detectors similar to those proposed by Stevens (1973). In this type of formulation, the integrative level does not process invariant distinctive features. Instead, the features processed at this level are still context dependent and require information about other features to be classified into a distinctive feature matrix at a later stage of analysis. Experiment 3 was designed, in part, to investigate whether the integrative level is indeed phonetic.

Experiment 3

If the central integrative level represents a phonetic level of processing, it might be expected to show a lack of sensitivity to variation in certain low level acoustic information. For example, manipulation of the relative intensity of the adapting and test stimuli might not affect this level. However, the formulation of the central integrative level in terms of an abstract set of property detectors offers a different view. According to this account, the integrative level should indeed be sensitive to adaptor and test intensity. An adaptor of high intensity would be expected to either fatigue more detectors, or the same detectors to a greater extent when compared to an adaptor of lower intensity. The relative intensities of adapting and test syllables have been manipulated by Hillenbrand (1975). When the adaptor was presented at an intensity of 15 dB greater than the test syllables, more adaptation was found than when the adaptor and test were presented at the same intensity. These results do not, however, bear directly on the issues raised previously. The adapting syllable used by Hillenbrand was the [ba] end of a [ba]-[da] place feature series. The increased adaptation found with louder adapting syllables could have been entirely due to increased adaptation at the (frequency specific) peripheral auditory level. Because the adaptor and test syllables were presented binaurally and the adaptor was a member of the test series, the locus of the intensity effect found in Hillenbrand's (1975) experiment is ambiguous.

The present experiment was an attempt to resolve this question using the general methodology and stimuli of Experiment 2. If the intensity effect found by Hillenbrand (1975) was occurring entirely at the peripheral auditory level, then no intensity effect should be found when the [bae]-high syllable is used to adapt the low [bae]-[dae] test series. Such a result would support a phonetic interpretation of the central integrative level found in Experiments 1 and 2. On the other hand, if the [bae]-high syllable does show an increased effectiveness as an adaptor when its intensity is raised, an abstract auditory level interpretation of the central component would be favored.

Method

Subjects. The subjects in this experiment were 20 Indiana University students who responded to an ad in the student newspaper. Some of these subjects had participated in previous experiments with synthetic speech. Subjects met the usual requirements and were paid at the rate of $2/hr.

Stimuli. The ten syllable [bae]-[dae] test series from Experiments 1 and 2 was employed again, along with one other stimulus, the [bae]-high syllable.

Procedure. All experimental events were controlled by a PDP-11 computer. The stored parameter codes for the OVE were synthesized in real time and presented to subjects through Telephonics (TDH-39) matched and calibrated headphones. The test stimuli were presented monaurally to the right ear at a level of 62 dB SPL for a steady state calibration vowel [ae].

The experiment consisted of two 1-hour sessions that were run on consecutive days. Subjects were run in groups of two or three. The first part of each session involved the determination of subjects baseline identification and rating of the [bae]-[dae] test series. The procedure was identical to that of Experiment 1 except for the intensity and ear of presentation (see above).

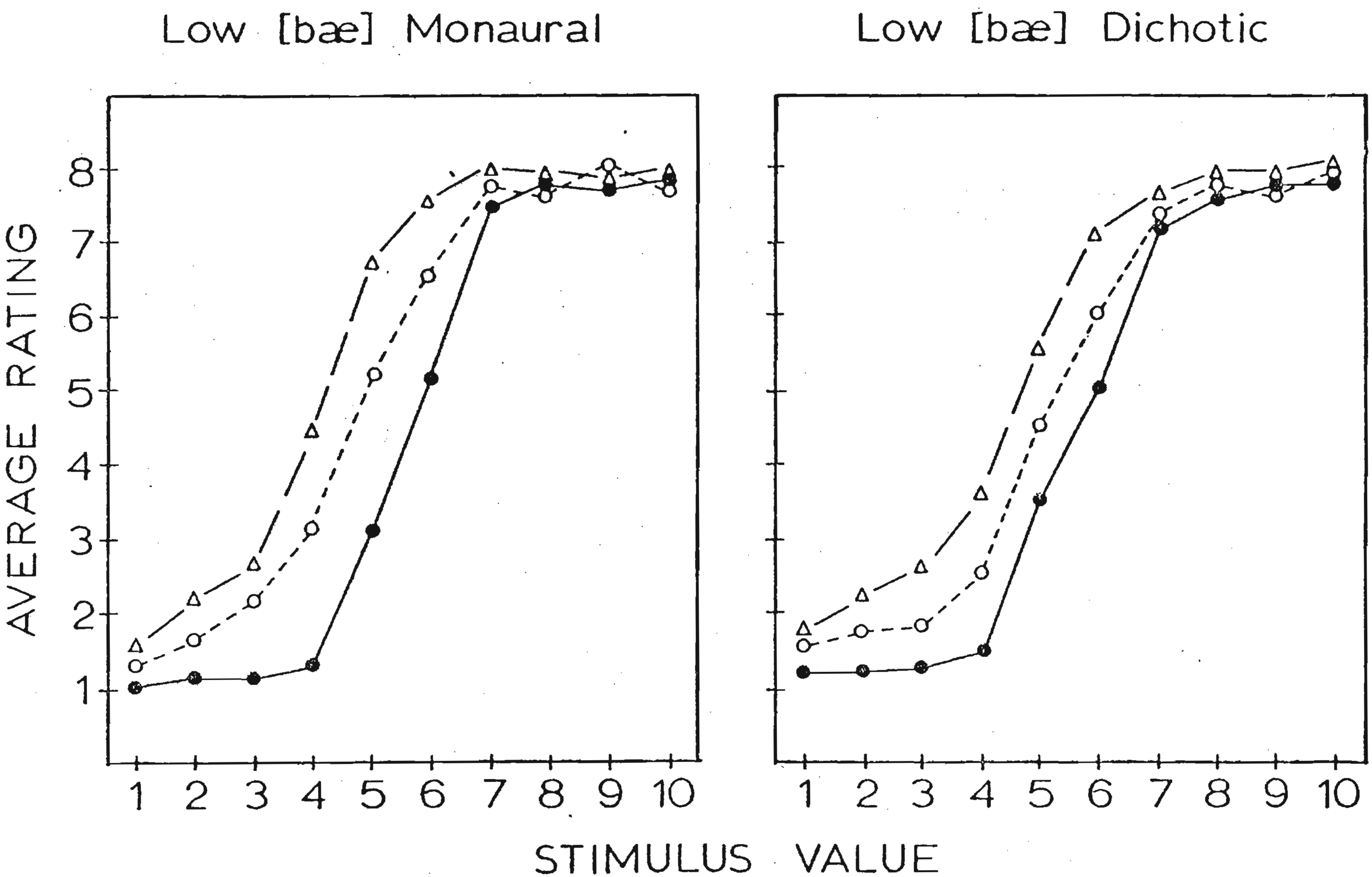
Following the identification sequence(s) on each day, two adaptation sequences were presented. Two of the four groups listened to the [bae]-high syllable adaptor and two groups listened to the [bae]-low syllable adaptor. Of the two groups who received the [bae]-high adaptor, one group heard the syllable in the same (right) ear as the test syllables. The other group heard the adaptor in the opposite (left) ear. The two groups receiving the [bae]-low syllable adaptor received a similar treatment. One group heard the adaptor in the left ear and test syllables in the right and the other group heard both adaptor and test syllables in the same (right) ear. Within each group, the adaptor was presented at a level of 54 dB SPL on one day and 70 dB SPL on the other day. (These levels were for a steady state vowel [ae].) For half of the subjects in each group the loud adaptor was presented on the first day and the soft adaptor on the second day. For the other half of the subjects, the reverse order was used. In all cases, subjects were informed as to the exact nature of the adapting syllable and which ear it would be presented to. The number of repetitions of the adaptor and presentation of test stimuli were identical in all other respects to Experiment 1.

Results and Discussion

The results for the low adaptor presented monaurally are shown on the left hand side of Figure 11. The [bae]-low adaptor had a significant effect on the category boundary for both adaptation intensities ($t(4) = 2.96$, $p < .025$ for the 54 dB adaptor; $t(4) = 10.88$, $p < .001$ for the 70 dB adaptor). In addition, the difference between intensities was also significant ($t(4) = 2.20$, $p < .05$). These results constitute a replication of the intensity effect originally found by Hillenbrand (1975).

The average ratings for the [bae]-low syllable adaptor presented dichotically are shown in the right hand panel of Figure 11. The same pattern of results found with the monaural presentation were also found for the dichotic presentation. The loud (70 dB) [bae]-low adaptor had a significant effect in moving the category boundary toward [bae] when presented in the opposite ear from the test syllables ($t(4) = 3.61$, $p < .025$). The difference between loud and soft adaptor was also statistically significant ($t(4) = 2.61$, $p < .05$) with the soft adaptor producing less adaptation. The boundary shift caused by the soft [bae]-low adaptor failed to reach significance ($t(4) = 1.13$, $p > .1$) although the shift was in the predicted direction. The average shifts in the category boundaries for both intensities for both groups appear in Table 5.

Figure 11. Unadapted (solid circles) and [bae]-low adapted rating functions. On the left are the 54 dB (open circles) and 70 dB (open triangles) monaurally adapted functions. On the right are the 54 dB (open circles) and 70 dB (open triangles) dichotic adaptation functions.



Low [bæ] Monaural
Low [bæ] Dichotic
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 11.

Table 5

Average category boundary shifts (in stimulus units) for adapting syllables presented at 54 dB and 70 dB relative to a 62 dB baseline test.

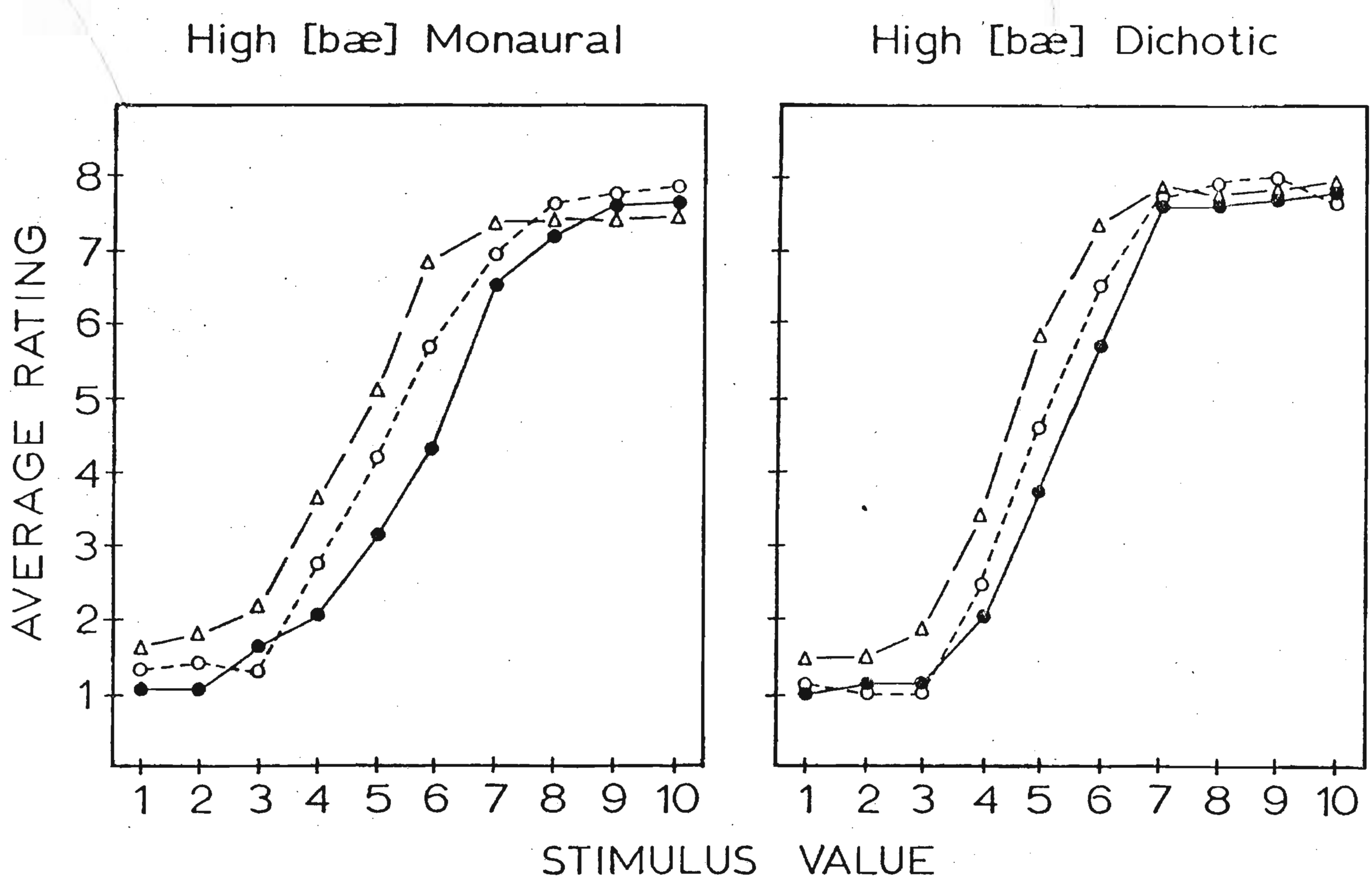
	[bae]-low monaural	[bae]-low dichotic	[bae]-high monaural	[bae]-high dichotic
70 dB	1.66	1.23	1.12	1.15
54 dB	.92	.35	.46	.45
Difference	.78	.88	.66	.70

The average rating functions, before and after adaptation, for the [bae]-high adaptor groups are shown in Figure 12. For the monaural presentation group, the loud adaptor produced a significant shift in the category boundary toward [bae] while the soft presentation condition produced a non-significant boundary movement toward [bae] ($\underline{t}(4) = 4.28$, $p < .01$ for the loud adaptor; $\underline{t}(4) = 1.34$, $\underline{p} < .1$ for the soft adaptor). The difference between the loud and soft adaptors was significant ($\underline{t}(4) = 3.02$, $\underline{p} < .025$) with the louder adaptor having the larger effect. The pattern of results for the [bae]-high syllable adaptor presented dichotically was essentially identical to that of the monaural presentation. Both adaptation conditions produced a significant movement of the category boundary toward [bae] ($\underline{t}(4) = 4.85$, $\underline{p} < .005$ for the loud adaptor; $\underline{t}(4) = 4.25$, $\underline{p} < .01$ for the soft adaptor). As with the other three groups, the difference between loud and soft conditions was also significant ($\underline{t}(4) = 5.18$, $\underline{p} < .005$). The average boundary shifts for both intensities for the two [bae]-high groups are also given in Table 5.

On the basis of these results, it appears that the central integrative level does respond to relatively low level auditory characteristics such as intensity. Both of the groups receiving the [bae]-high syllable adaptor showed a significant increase in adaptation following exposure to a 70 dB SPL adaptor as opposed to a 54 dB adaptor. The [bae]-low adaptor also showed less effectiveness in moving the category boundary when presented at a lower intensity.

The between groups comparison of interaural transfer of adaptation also replicated the findings of Experiment 2. From a comparison of the average category boundary shifts in Table 5, the [bae]-low adaptor shows

Figure 12. Unadapted (solid circles) and [bae]-high adapted rating functions. On the left are the 54 dB (open circles) and 70 dB (open triangles) monaural adaptation functions. On the right are the 54 dB (open circles) and 70 dB (open triangles) dichotic adaptation functions.



High [bæ] Monaural
High [bæ] Dichotic
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8 9 10
STIMULUS VALUE

Figure 12.

about 60% interaural transfer. This result comes from averaging the loud and soft adaptation shifts for dichotic presentation and comparing it to the same average for monaural presentation. The 60% transfer is somewhat greater than that found in Experiment 2, but it is still well below 100%. The same comparison between the [bae]-high monaural and [bae]-high dichotic conditions (see Table 5) reveals 100% interaural transfer. Thus, the results of the present experiment replicate those of Experiment 2 on the interaural transfer of adaptation with the [bae]-low and [bae]-high syllables.

Further analysis of the data revealed one additional important finding. From inspection of Table 5, it is apparent that the intensity manipulation had most of its effect at the central integrative level. The average increase in the category boundary shift for the [bae]-low adaptor presented monaurally was 0.78 stimulus units. The average increase for the two [bae]-high groups was 0.68 stimulus units. This provides additional support for the conclusion that the central, integrative level does respond to basic acoustic parameters. One possible reason for the apparent lack of an intensity effect at the peripheral auditory level is that this level may have already reached an adaptation ceiling. From previous experimental work (Hillenbrand, 1975), we know that the number of repetitions of an adaptor affects the amount of boundary shift. The greater the number of adapting syllables, the greater the magnitude of shift. In the present study, 75 repetitions of the adaptor were used before each sequence of test syllables. This relatively high number of repetitions may have induced essentially complete fatigue at the peripheral auditory level. Consequently, the apparent lack of any effect of intensity on the peripheral level may be the result of a ceiling effect. In order to investigate this, further experimentation manipulating the number of presentations of the adaptor is necessary.

On the basis of the results from Experiments 1 through 3, it is clear that there are indeed two components involved in selective adaptation. However, before any firm conclusions may be drawn from these data, they must be reconciled with the cross series adaptation results of Bailey (1975) mentioned earlier. In Bailey's cross series conditions, [bi] and [di] adaptors failed to induce category boundary shifts in a [bu]-[du] series and vice versa. Experiment 4 was run in order to check the previous adaptation results for extreme vowel CV series reported by Bailey (1975).

Experiment 4

The failure of the [di] adaptor to affect the [bu]-[du] series and the [du] adaptor to affect the [bi]-[di] series in Bailey's (1975) experiment can be accounted for on the basis of their qualitatively different formant transition trajectories, as discussed previously in Chapter II. However, the lack of a cross series adaptation effect for the bilabial [bi] and [bu] adaptors is in direct conflict with the adaptation results found for the high adaptors in Experiment 1, 2, and 3 where the spectral overlap between adapting and test syllables was varied yet category boundary shifts were still found.

The difference between the results of Bailey (1975) and the present experiments could be due to procedural differences between these experiments. In the present experiments subjects were always tested on one particular series of syllables. The testing procedure used by Bailey involved the random intermixing of both the [bi]-[di] and the [bu]-[du] test series. The same type of procedure has been used in the alternating adaptor experiments of Pisoni & Sawusch (1976), Ganong (forthcoming) and Ades (forthcoming). Ades (forthcoming) constructed two [b]-[d] place feature series which varied only in fundamental frequency. Subjects were adapted with alternating low pitch [b] and high pitch [d] syllables. The result was that the low pitched series shifted its category boundary toward [b] and the high pitched series shifted toward [d] (and vice versa for the [b]-high and [d]-low pair of adaptors). Ganong (forthcoming) used two [bo]-[go] series which varied in intensity and found a similar contingent effect. These contingent results, where the only difference between adaptor and test syllables is on an irrelevant dimension, suggest that either selective adaptation is specific to the particular syllables used as adaptor and test or that subjects can somehow keep the different test series selectively separated.

Pisoni & Sawusch (1976) ran subjects in tasks similar to those mentioned above with a slight modification in the procedure. Subjects were carefully instructed, in all conditions, to ignore any changes in pitch (or intensity). The results indicate that the adaptation in the alternating adaptors test conditions was the sum of the effects of each adaptor taken individually. No contingent effect was found for either intensity or fundamental frequency. This suggests that the results of Ades (forthcoming) and Ganong (forthcoming) may have been due to some form of perceptual streaming of the stimuli. Subjects could have assigned the syllables to different "sources" based on intensity or pitch differences in the stimuli. Stimuli from these different sources could have then been channeled to different sets of detectors. This type of explanation could possibly resolve the differential results of Pisoni & Sawusch with those of Ades and Ganong (see also Ades, 1976; Darwin, 1976, for further discussion and speculation on these results).

A source effect explanation could also account for the failure to find cross-series adaptation effects in Bailey's (1975) experiment. Subjects could have treated the two intermixed series of stimuli as originating from different sources and assigned them to different sets of detectors. The present experiment was designed to try to replicate the results of Bailey using a procedure where there was no intermixing of stimuli from different test series. As in the study by Bailey, a [bi]-[di] place feature series was used for testing. The adapting syllables were [bi] from the test series, a [bae] and a [bu]. Schematic spectrograms of all of the adaptors are shown in Figure 13. The test series was always the [bi]-[di] series, thus eliminating the possibility of subjects assigning different intermixed series to different sources. Thus, the outcome of this experiment should resolve whether or not the cross series adaptation results of Bailey (1975) were due to the particular procedure used.

<u>Method</u>

<u>Subjects</u>. The subjects in this experiment were eight undergraduate volunteers at Indiana University who responded to an ad in a student newspaper. All had some limited previous experience with synthetic speech. They met the usual requirements and were paid for their participation at the rate of $2/hr.

Figure 13. Schematic spectrograms of the stimuli from Experiment 4. On the left are the [bi] (solid formants) and [di] (dashed formants) endpoints of the test series. In the middle is the [bae] adaptor and on the right is the [bu] adaptor.

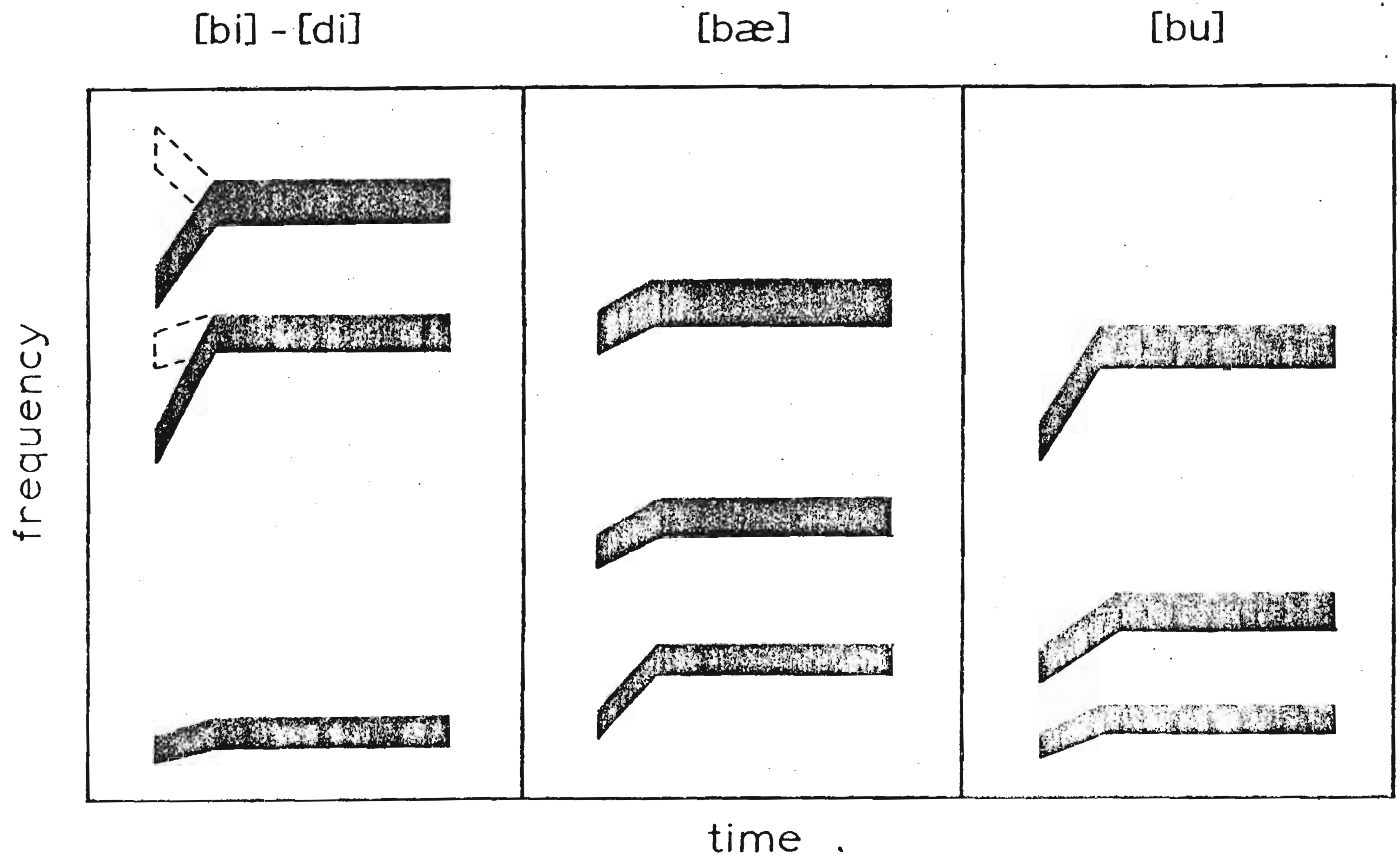


Figure 13.

Stimuli. A new set of five formant synthetic CV syllables was prepared on the OVE synthesizer at Indiana University. The test stimuli were eight syllables that ranged perceptually from [bi] to [di]. These stimuli varied in the starting frequencies of the second and third formant transitions from 1510 Hz (F2) and 2329 Hz (F3) for the [bi] end to 2135 Hz (F2) and 3591 Hz (F3) at the [di] end of the series in seven approximately equal steps. The first formant transition had a starting frequency of 200 Hz for all eight stimuli. The duration of the formant transitions was 40 msec, followed by a 210 msec steady state vowel [i]. The vowel had formant center frequencies of 266 Hz (F1), 2262 Hz (F2), and 3020 Hz (F3). The fourth and fifth formants were fixed at 3500 Hz and 4000 Hz.

Two other stimuli, to be used as adaptors, were also constructed. These stimuli also had rising second and third formant transitions, but in front of the vowels [ae] and [u]. The [bae] syllable had initial formant center frequencies of 282 Hz, 1270 Hz and 2262 Hz for the first through third formants respectively. These formants rose over a 40 msec time span to the 210 msec steady state [ae] vowel with center frequencies of 672 Hz (F1), 1425 Hz (F2), and 2540 Hz (F3). The [bu] syllable also had 40 msec duration transitions followed by a 210 msec steady state vowel. The formant transitions had starting frequencies of 200 Hz (F1), 616 Hz (F2), and 1554 Hz (F3). The steady state [u] vowel had formant frequencies of 299 Hz, 872 Hz, and 2016 Hz for the first through third formants respectively. As with all of the OVE stimuli, the fourth and fifth formants were fixed at 3500 Hz and 4000 Hz. The fundamental frequency was the same for all of these stimuli and fell linearly from 129 Hz to 94 Hz over the 250 msec syllable duration.

Procedure. As in Experiment 1, all experimental events were controlled by a PDP-11 computer. The stored parameter codes for the OVE were synthesized in real time and presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. To compensate for differences in overall intensity between syllables with different vowels, all syllables were presented at a level of 80 dB for their own steady state vowel (either [i], [ae], or [u]).

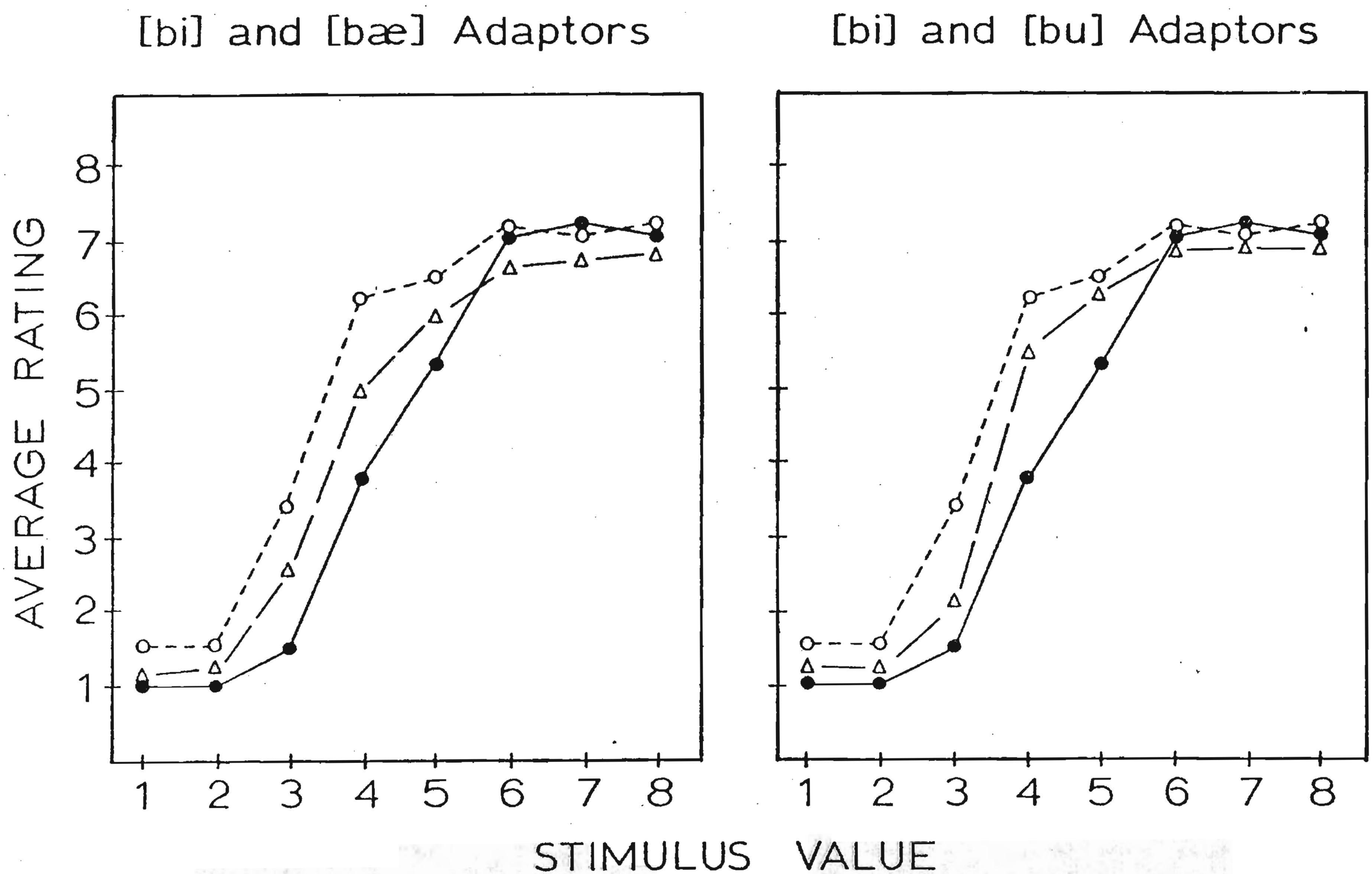
The experiment consisted of three 1-hour sessions that were run on consecutive days. Subjects were run in two groups of four subjects each. On each day, subjects listened to one identification sequence and two adaptation sequences. One of the two groups had the [bi] adaptor on the first day, [bae] on the second, and [bu] on the third day. The other group received the opposite order of adaptors. At the beginning of the first day only, all subjects listened to a 40 trial practice identification sequence (five occurrences of each of eight syllables in random order). Otherwise, every day started with an 80 trial identification sequence, each syllable occurring ten times in random order. The response procedure was essentially identical to that used in Experiment 1. Subjects first identified the syllable presented to them and then gave a confidence rating as to the accuracy of their identification response. All responses were entered on a response box in front of each subject.

The adaptation procedure was also very similar to that used in Experiment 1. The particular adaptor for the day was presented for 45 sec (75 presentations with a 300 msec inter-adaptor interval). After each period of adaptation, all eight syllables of the [bi]-[di] test series were presented in random order for identification. Twenty of these adaptation sequences were run on each day. By the end of the experiment, each subject had provided twenty adapted responses to each of the eight test syllables for each of the three adaptors. Subjects had also provided thirty unadapted responses (ten from each day) to each of the eight [bi]-[di] syllables. As in previous experiments, subjects were informed as to the exact nature of the repeated syllable that they were listening to on each day.

Results and Discussion

The average rating data for all eight subjects are shown in Figure 14. Both panels contain the unadapted function and the [bi] adapted function. The [bi] adaptor had a significant effect in moving the category boundary toward [bi] on the [bi]-[di] series ($t(7) = 6.03$, $p < .001$). The data of most interest, however, are the [bae] and [bu] adapted rating functions. These data appear in the left and right hand panels of Figure 14 respectively. Both adaptors had a significant effect in moving the [bi]-[di] category boundary toward the [bi] end of the series ($t(7) = 3.64$, $p < .005$ for the [bae] adapted function; $t(7) = 4.24$, $p < .005$ for the [bu] adapted function). One other point about this data should be noted. Every subject showed less adaptation with the [bu] and [bae] adaptors than with the [bi] adaptor. The difference in the category boundary shift between the [bi] adaptor and each of the others was significant ($t(7) = 2.78$, $p < .025$ for the [bi]-[bae] comparison; $t(7) = 2.59$, $p < .025$ for the [bi]-[bu] comparison). There was no significant difference between the effects of the [bae] and the [bu] adaptors. The little difference that does exist between the effects of the [bae] and [bu] adaptors can probably

Figure 14. Unadapted (solid circles) and various [b] adapted rating functions. On the left are the [bi] (open circles) and [bae] (open triangles) adapted functions. On the right are the [bi] adapted (open circles) and [bu] adapted (open triangles) functions.



[bi] and [bæ] Adaptors
[bi] and [bu] Adaptors
AVERAGE RATING
1
2
3
4
5
6
7
8
STIMULUS VALUE

Figure 14.

be attributed to the close proximity of the F2 transition at the [bi] end of the test series with the F3 transition in the [bu] adaptor.

These results are therefore in agreement with those found previously in Experiment 1. Since the [bi]-[di] series was very similar to that used by Bailey (1975) and the present [bu] adaptor was also similar to the [bu] used by Bailey, it seems reasonable to conclude that Bailey's failure to find a cross series adaptation effect was due to the particular testing procedure used rather than the absence of an adaptation effect. The results of Experiment 4 provide evidence against the previous interpretation of Bailey (1975) that selective adaptation of the place feature occurs only at the level of spectral (frequency) overlap. The pattern of results in Experiment 4 is entirely consistent with the two level interpretation of selective adaptation. Although the [bu] adaptor may have had some overlap with the [bi]-[di] test series, the [bae] adaptor had essentially no overlap. Thus, the adaptation effect induced by the [bae] on the [bi]-[di] series could only have been due to an integrative level which responds to all rising frequency transitions, regardless of spectral location.

General Discussion, Experiments 1, 2, 3 and 4

The results of Experiments 1 through 4 provide strong evidence for the operation of two distinct processes in selective adaptation to the place feature in speech. As with previous experiments, one of these is a relatively peripheral auditory level of processing. From the results of Experiments 1 and 2, this level is characterized by its frequency specificity. The information that feature detectors at this level respond to is confined within a narrow frequency band, approximately the width of a critical band (see Scharf, 1970, for bandwidth estimates).

The integrative level and its characteristics were the primary focus of the experiments in this chapter. Experiments 1 and 2 provide strong support for the existance of such a level in selective adaptation. The 100% interaural transfer found in Experiments 2 and 3 offers further support for the separate existence of an integrative level and demonstrates that this level is centrally located. Although this second level of adaptation is integrative and centrally located, it is not completely removed from low-level variations in the acoustic signal. The increased effectiveness of a more intense adaptor upon this level (Experiment 3) demonstrates this conclusively. Taken as a whole, these experiments offer strong support for the involvement of two separate levels in selective adaptation. However, previous experimenters who have advocated the two level approach have often termed the second level as being phonetic (Cooper, 1974b, 1975; Tartter & Eimas, 1975). The results of Experiment 3 would fit better, in our view, into a two process auditory model. The first level is peripheral and frequency specific, the second is central and integrates information over a fairly wide frequency range.

Further evidence on whether the integrative level is an abstract auditory or phonetic level of processing comes from a series of experiments by Cutting, Rosner & Foard (1975). These experimenters carried out a number of experiments, including selective adaptation, with a set of complex non-speech acoustic signals that varied in rise-time. These stimuli, called "plucks" and "bows" because they sound like either the plucking or bowing of a stringed instrument, have previously been found to exhibit categorical perception (Cutting & Rosner, 1974). Cutting et al.

(1975) tested subjects identification of these rise-time stimuli (under instructions to categorize them as either plucks or bows) both before and after adaptation. When the adapting stimuli were the end-points of the test series, large and consistent boundary shifts were found. The 0 msec rise-time (pluck) stimulus moved the category boundary toward the pluck end of the series and the 80 msec rise-time (bow) stimulus moved the category boundary toward the bow end of the series. In this case, all stimuli were 440 Hz sawtooth items varying only in rise-time. Of particular interest are the adaptation results for two other adapting stimuli. These stimuli also varied in rise-time. However, they were 294 Hz sinewaves. Consequently, they shared no frequency components in common with the 440 Hz sawtooth test series. Further, the 294 Hz and the 440 Hz stimuli are separated by more than one critical bandwidth in their frequency composition, making them comparable to the high and low stimuli used in Experiments 1, 2 and 3. These 294 Hz sinewave stimuli, when used as adaptors, produced boundary shifts on the 440 Hz sawtooth series. These shifts were in the predicted direction (according to rise-time) but they were much smaller than the within series shifts found with the 440 Hz sawtooth adaptors. Thus, from these results of Cutting et al. (1975) it appears that adaptation effects can be demonstrated at both frequency specific and integrative auditory levels of processing for non-speech adapting and test stimuli. This result argues very strongly that the integrative level tapped by the experiments in this chapter is part of man's general auditory system and is not a phonetic or linguistic mode of processing that is special to speech. However, the exact nature of the integrative level and whether it is better described in terms of feature detectors or decision rules is still unknown. Certain aspects of the processing at this integrative level may still be specific to speech, if only because speech is the only suitable, complex stimulus that naturally occurs in man's environment. It is to these issues about the nature of the processing at the integrative level and the earlier peripheral level that we turn in the next chapter.

Footnotes

[3] All t-tests, unless otherwise stated, were one-tailed tests for correlated measures. The one-tailed test was employed because of the directional nature of the expected changes in the locus of the category boundary.

[4] In pilot experiments a high test series, in which the two high syllable adaptors were the end-points, was also used. Both high and low end-point syllables were used as adaptors for this series. The same pattern of results that was found in Experiment 1 for the low test series was also found in the pilot studies on the high test series. Both high and low adaptors moved the category boundary for the high series. The high adapted rating functions exhibited about twice as much shift in the category boundary as that found with the low adaptors. Since the pattern of results was similar for both high and low test series in the pilot data, it was decided to focus on only the low series in Experiment 1 and succeeding experiments.

[5] For the [bae]-low syllable adaptor presented monaurally, the rating shifts for Stimuli 1 through 7 were all significant ($t(5) = 2.40$, $p < .05$ for Stimulus 1; $t(5) = 3.06$, $p < .025$ for Stimulus 2; $t(5) = 4.71$, $p < .005$ for Stimulus 3; $t(5) = 5.80$, $p < .005$ for Stimulus 4; $t(5) = 10.40$, $p < .001$ for Stimulus 5; $t(5) = 11.44$, $p < .001$ for Stimulus 6; and $t(5) = 3.63$, $p < .01$ for Stimulus 7). For the [bae]-low chirp adaptor presented monaurally, the rating shifts for Stimuli 2 through 7 were significant ($t(5) = 1.61$, $.05 < p < .1$ for Stimulus 1; $t(5) = 2.21$ $p < .05$ for Stimulus 2; $t(5) = 2.73$, $p < .025$ for Stimulus 3; $t(5) = 3.34$, $p < .025$ for Stimulus 4; $t(5) = 13.31$; $p < .001$ for Stimulus 5; $t(5) = 16.93$,

$p < .001$ for Stimulus 6; and $t(5) = 2.06$, $p < .05$ for Stimulus 7). The [dae]-low syllable adaptor presented monaurally caused significant rating shifts in Stimuli 5 through 9 ($t(5) = 2.62$, $p < .025$ for Stimulus 5; $t(5) = 6.07$, $p < .001$ for Stimulus 6; $t(5) = 3.99$, $p < .01$ for Stimulus 7; $t(5) = 2.62$, $p < .025$ for Stimulus 8; and $t(5) = 2.11$, $p < .05$ for Stimulus 9). The [dae]-low chirp adaptor also caused significant shifts in the rating responses for Stimuli 5 through 9 when it was presented monaurally ($t(5) = 2.02$, $p < .05$ for Stimulus 5; $t(5) = 3.08$, $p < .025$ for Stimulus 6; $t(5) = 2.37$, $p < .05$ for Stimulus 7; $t(5) = 2.79$, $p < .025$ for Stimulus 8; and $t(5) = 2.33$, $p < .05$ for Stimulus 9). All t-tests were one-tailed tests for correlated measures.

CHAPTER IV

Further Characteristics of Peripheral and Integrative Levels of Processing in Selective Adaptation to Place

Part of the appeal of a feature detector model for early speech processing is its apparent simplicity. The concept of a two level model in which frequency specific and auditory detectors feed an integrative second level set of detectors such as the model of Tartter & Eimas (1975) offers one possible way to resolve the invariance problem that has characterized speech perception research. Some recent experiments have, however, complicated this simple picture. Ades (1974b) has reported an experiment where CV syllables were used as adaptors on a VC test series and vice versa. His results indicate that a CV will not adapt a VC and neither will a VC adapt a CV. This result is important for two reasons. First, this result appears to rule out adaptation at a distinctive feature level, where a particular phonetic feature is assumed to be independent of context. If adaptation were affecting this type of processing, then a [bae] syllable should have moved the [aeb]-[aed] boundary toward [b]. No such result was found by Ades. Secondly, these results also appear to rule out a free running detector system where the same detector responds to initial and final position acoustic cues. Ades (1974b) interpreted his results as evidence for the presence of position specific detectors. The presence of this modification to feature detector systems makes them less attractive as the basis for a perceptual model. If one is not careful, the model ends up with a proliferation of feature detectors. This leads to a model of recognition with an infinite number of components (see Weinstein, 1973).

The conclusions drawn by Ades (1974b) must be tempered, however, due to the acoustic composition of his CV and VC syllables. The end-points of the CV and VC syllable series used by Ades (1974b) were reproduced previously as schematized spectrograms in Figure 3. As mentioned before, on an acoustic basis, the final [b] in [aeb] might be expected to show a [d] adapting effect on the [bae]-[dae] test series since both the [b] in [aeb] and the [d] in [dae] have falling transitions. Precisely such a result has been reported by Pisoni (1975b). It should be noted, however, that the final transitions of [aeb] and the initial transitions of [dae] do not overlap very extensively in frequency. Hence, little adaptation would be expected at a spectrally specific auditory level. The same can be said for all of the combinations of CV adaptor with VC test and vice versa in the experiments carried out by Ades (1974b).

If adaptation is not expected at a frequency specific level, would it be present at a slightly more abstract level? Prime candidates for this integrative type of detector are the auditory property detectors proposed by Stevens (1973). For stops in initial position, Stevens (1973) proposed a general spectral rise detector for the labials, a falling frequency detector for the alveolars and a diverging frequency detector for the velars. Such detectors would integrate burst and transition cues to yield a relatively invariant mapping of acoustic information onto phonetic features. These detectors are not, however, positionally invariant. For example, while falling transitions are a cue for alveolars in initial position, they are a cue for labials in final position. Ades' (1974b) results do indeed seem to rule out adaptation at a level where detectors respond in the manner described by Stevens (1973), unless the detectors are required to be positionally sensitive.

An experiment by Pisoni & Tash (1975) addressed itself to the question of whether the spectrally specific acoustic level detectors are positionally sensitive. This experiment was reviewed earlier, but the results are worth repeating. When the initial transitions for a [ba] syllable were moved to final position, this non-speech stimulus (labeled a b-SEChirp) had an adapting effect on a [ba]-[da] test series. The category boundary for the test series moved toward the [ba] end of the series. Similar, "d" adapting effects were found for the d-SEChirp. This adaptation effect can be attributed to a peripheral auditory level in the present model because the SEC adaptors contained the same transitions as the test series, yet the SEC adaptors were not consistently heard as speech (only speech like) by subjects in this experiment (Pisoni & Tash, 1975). On the basis of this result, the spectrally specific auditory component involved in selective adaptation does not appear to be positionally sensitive as suggested earlier by Ades.

Ades (1976), however, has offered another interpretation of the Pisoni & Tash (1975) results. He attributed the adapting effect of the SEChirps to a rapid intensity onset at the beginning of the initial steady state component. Since the initial steady state frequencies were the same as the starting frequencies of the transitions, a rapid onset could have triggered a set of initial position sensitive transition detectors and thus given rise to the adaptation effect. Ades (1976) was forced into such an interpretation because he tried to attribute all of the place adaptation results to a single frequency and position specific auditory level of processing. The following experiment was designed to distinguish between these alternative interpretations by systematically manipulating the spectral overlap between a CV place series and a VC place series.

Experiment 5

The present set of experiments was designed to resolve the conflicting interpretations of the experiments of Ades (1974b) and Pisoni & Tash (1975). In order to do so, separate sets of CV and VC test syllables were constructed so as to maximize the spectral overlap in formant transitions between one end of the CV series and one end of the VC series. The end-points of the two sets are shown as schematized spectrograms in Figure 15. The second and third formant transitions of the [bae] from the [bae]-[dae] series are identical with those of the [ʌd] from the [ʌb]-[ʌd] series. Thus, these stimuli include the spectral overlap of the SEChirp's used by Pisoni & Tash (1975) in a full VC context similar to that used by Ades (1974b). Using these stimuli, it was hoped that further light could be shed on the question of positional dependence of selective adaptation.

Method

Subjects. The subjects for this experiment were ten undergraduates at Indiana University who responded to an ad in the student newspaper. They met the usual requirements and were paid for their participation at the rate of $2/hr.

Stimuli. Two sets of synthetic five formant syllables were constructed on the OVE speech synthesizer at Indiana University. One series consisted of seven CV syllables that ranged perceptually from

Figure 15. Schematic spectrograms of the [ʌb] (dashed formants) and [ʌd] (solid formants) endpoints and the [bae] (solid formants) and [dae] (dashed formants) endpoints from the two syllable sets in Experiment 5.

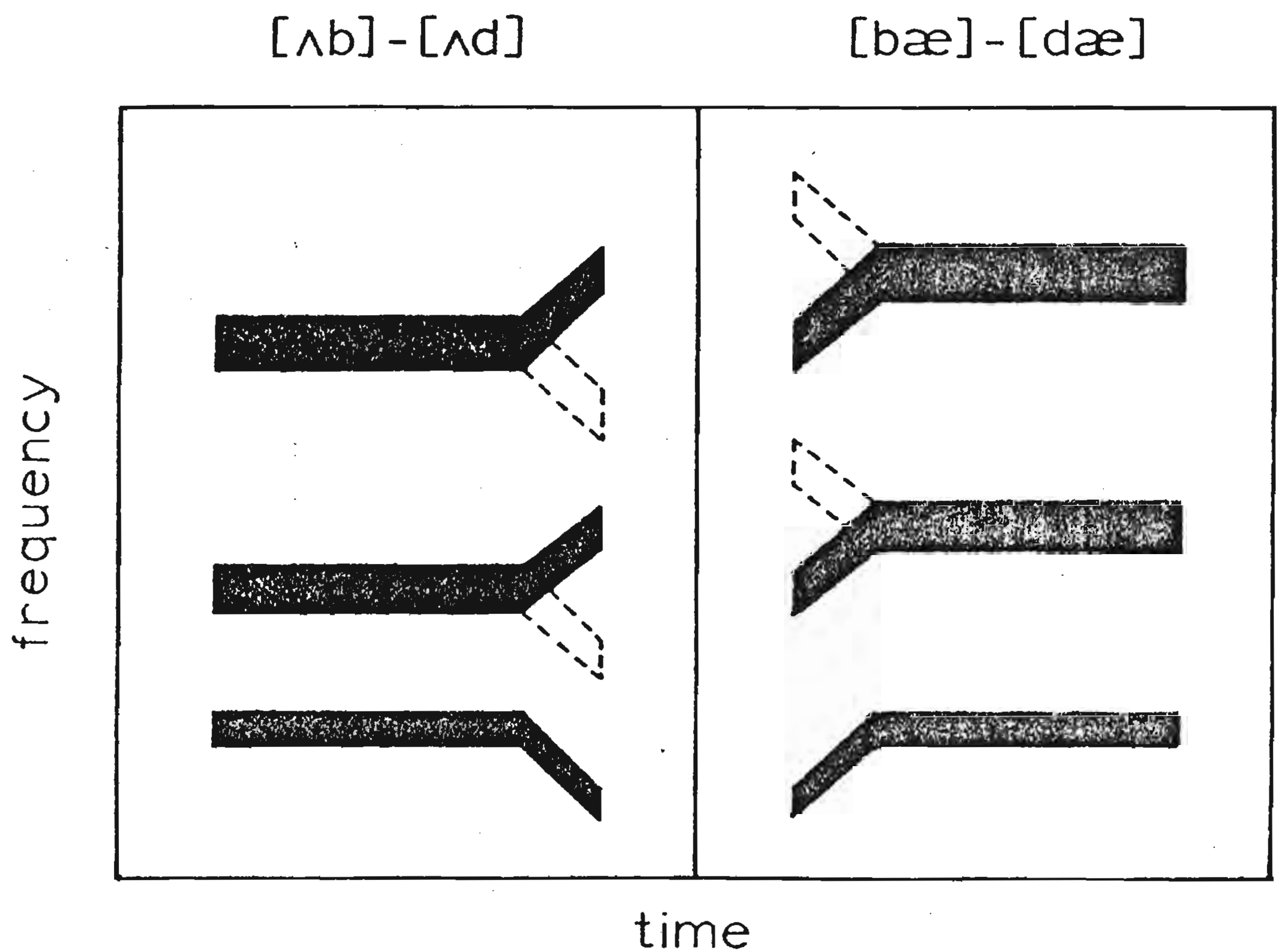


Figure 15.

[bae] to [dae]. These stimuli varied in the starting frequencies for the second and third formant transitions from 1198 Hz (F2) and 2262 Hz (F3) for the [bae] end of the series to 2135 Hz (F2) and 3489 Hz (F3) at the [dae] end in six approximately equal steps. The first formant transition had a starting frequency of 251 Hz for all seven stimuli. The transitions were followed by a steady-state vowel [ae] which had formant center frequencies of 654 Hz (F1), 1647 Hz (F2), and 2850 Hz (F3).

The second series contained seven VC syllables that ranged perceptually from [Λb] to [Λd]. These stimuli varied in their terminal frequencies for their second and third formant transitions. The [Λb] end of the series had terminal frequencies of 755 Hz (F2) and 1695 Hz (F3) while the [Λd] end of the series had terminal frequencies of 1647 Hz (F2) and 2850 Hz (F3). The five stimuli between these two ends had terminal frequencies that were approximately equally spaced. The first formant transition fell to 251 Hz for all seven stimuli. The formant transitions for all seven stimuli were preceeded by a steady-state vowel resembling [Λ] with formant center frequencies of 645 Hz, 1198 Hz and 2262 Hz. Stylized sound spectrographs of the end-points of these two series are shown in Figure 15. All formant transitions for both series had a duration of 45 msec and the vowel duration of each syllable was 205 msec. The fundamental frequency for all 14 syllables was held constant at 118 Hz. Finally, to minimize certain onset and offset transients, the amplitude of the seven CV syllables was ramped off over a period of 50 msec. Similarly, the amplitude of the seven VC syllables was ramped on over a 50 msec period.

One final stimulus, to be used as an adaptor, was also constructed. An F2F3 chirp was produced on a parallel resonance synthesized at MIT.[6] This two formant chirp had a duration of 45 msec and was constructed to match the second and third formant transitions of the [bae] end-point of the [bae]-[dae] series and the [Λd] end-point of the [Λb]-[Λd] series. This chirp had initial frequencies of 1198 Hz and 2262 Hz and final frequencies of 1646 Hz and 2850 Hz for its two formants. As in the CV and VC series, the fundamental frequency was constant at 118 Hz. This chirp was recorded on magnetic tape and later digitized and stored on the disc memory of a PDP-11 computer.

Procedure. All experimental events were controlled by a PDP-11 computer. The stored parameter codes for the OVE were synthesized in real time and presented binaurally to subjects through Telephonics (TDH-39) matched and calibrated headphones. The digitized F2F3 chirp was reconverted to analog form in real time and also presented binaurally to subjects. The stimuli were presented at a level of 80 dB SPL for a steady-state calibration vowel [ae].

The experiment consisted of three 1-hour sessions that were run on consecutive days. Subjects were run in two groups of five subjects each. One of these groups was always tested on the [bae]-[dae] series and the other was always tested on the [Λb]-[Λd] series. On the first day only, subjects were given a 35 trial practice sequence in which each of the seven syllables in their test series was presented five times in random order. Next, subjects listened to a 70 item identification test with each syllable occurring ten times in random order. Subjects were instructed that they would be hearing the syllables [bae] and [dae] ([ʌb] and [ʌd]) and were given identification and rating instructions as in the previous experiments.

Following the identification sequence, subjects listened to two adaptation sequences on each day. Three adapting stimuli were used: the [bae] end-point, the [ʌd] end-point, and the F2F3 chirp. A different adaptor was used each day and their order was the same for both groups: [bae] the first day, [ʌd] the second day, and the F2F3 chirp on the third day. No attempt was made to counterbalance the order of adaptor presentation across days. In all cases, subjects were informed as to the exact nature of the repeated sound that they would be listening to.

In a block of adaptation and testing, the adaptor was presented 75 times with 300 msec between presentations. Following this period of adaptation, seven stimuli were presented in random order for identification and rating by subjects. Twenty of these adaptation sequences were run on each day. Thus, by the end of the experiment, each subject had provided thirty unadapted responses and twenty adapted responses (for each adaptor) to each of the seven test syllables in their series.

Results and Discussion

For both the [bae]-[dae] and the [ʌb]-[ʌd] series, the end-point adaptor from the test series had a significant adapting effect. The [bae] adaptor moved the [bae]-[dae] category boundary toward [bae] ($\underline{t}(4) = 3.95$, $\underline{p} < .01$) and the [ʌd] adaptor shifted the [ʌb]-[ʌd] category boundary toward [ʌd] ($\underline{t}(4) = 6.70$, $\underline{p} < .005$). The F2F3 chirp also had a significant effect on both series, moving the [bae]-[dae]

boundary toward [bae] ($\underline{t}(4) = 2.79$, $\underline{p} < .05$) and the [Λb]-[Λd] boundary toward [Λd] ($\underline{t}(4) = 2.62$, $\underline{p} < .05$). The average rating function for the [bae]-[dae] test series group is shown in the left hand panel of Figure 16 along with the [bae] and the F2F3 chirp adapted functions. The left hand panel of Figure 17 shows the baseline, [Λd] adapted and F2F3 chirp adapted rating function for the [Λb]-[Λd] test series group.

In contrast to these within series adaptation conditions, no category boundary shift was found in the cross series conditions. The right hand panel of Figure 16 shows the [bae]-[dae] rating function before and after adaptation with [Λd]. The curves are virtually identical and no significant boundary shift was found. The same pattern of results can be seen in the right hand panel of Figure 17 for [bae] adaptation and testing on the [Λb]-[Λd] series. No significant category shift was found.

These results agree very well with those of Ades (1974b) in that no effect of a VC on a CV series (or vice versa) was found. This result was found by Ades (1974b) when the VC and CV syllables were mirror images and in the present study when the spectral overlap between the transitions for VC and CV syllables was maximized. The present results therefore make the interpretation of the SEChirp adaptation results of Pisoni & Tash (1975) somewhat difficult to handle. However, there were several minor procedural differences between the present study and that of Pisoni & Tash (1975) who used a greater number of repetitions of the adaptor. Since previous results have found that a greater number of presentations leads to more adaptation (Hillenbrand, 1975) it was decided to repeat, as closely as possible, the procedure of Pisoni & Tash (1975) using the cross series adaptation conditions of the previous study.

Figure 16. Unadapted (solid circles) and adapted rating functions for the [bae]-[dae] group. On the left are [bae] (open circles) and F2F3-chirp (open triangles) adapted functions. On the right are [ʌd] adapted functions for 75 (open circles) and 100 (open triangles) repetitions of [ʌd].

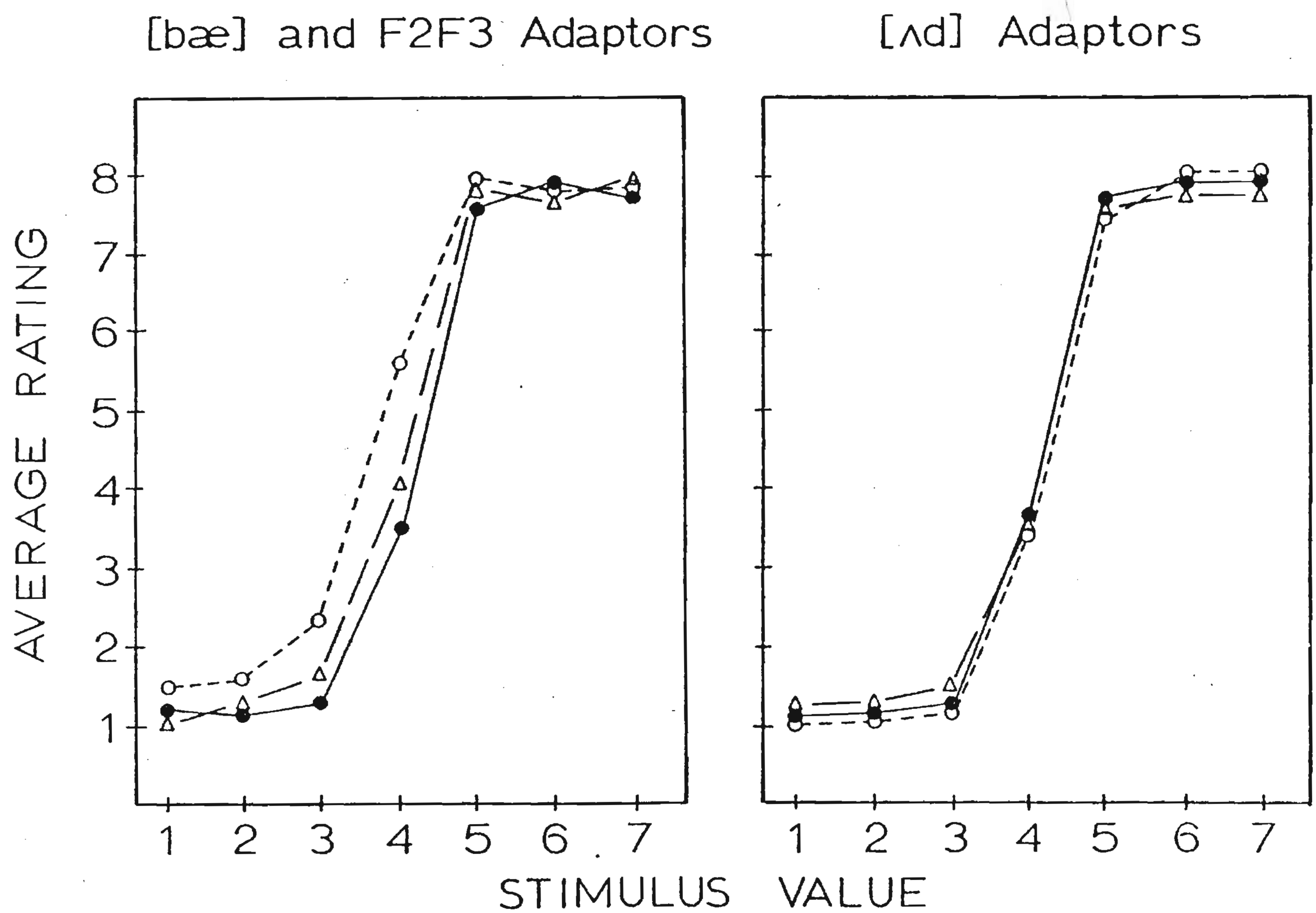

[bæ] and F2F3 Adaptors
[ʌd] Adaptors
AVERAGE RATING
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 16.

Figure 17. Unadapted (solid circles) and adapted rating functions for the [Λb]-[Λd] group. On the left are [Λd] (open circles) and F2F3-chirp (open triangles) adapted functions. On the right are [bae] adapted functions for 75 (open circles) and 100 (open triangles) repetitions of [bae].

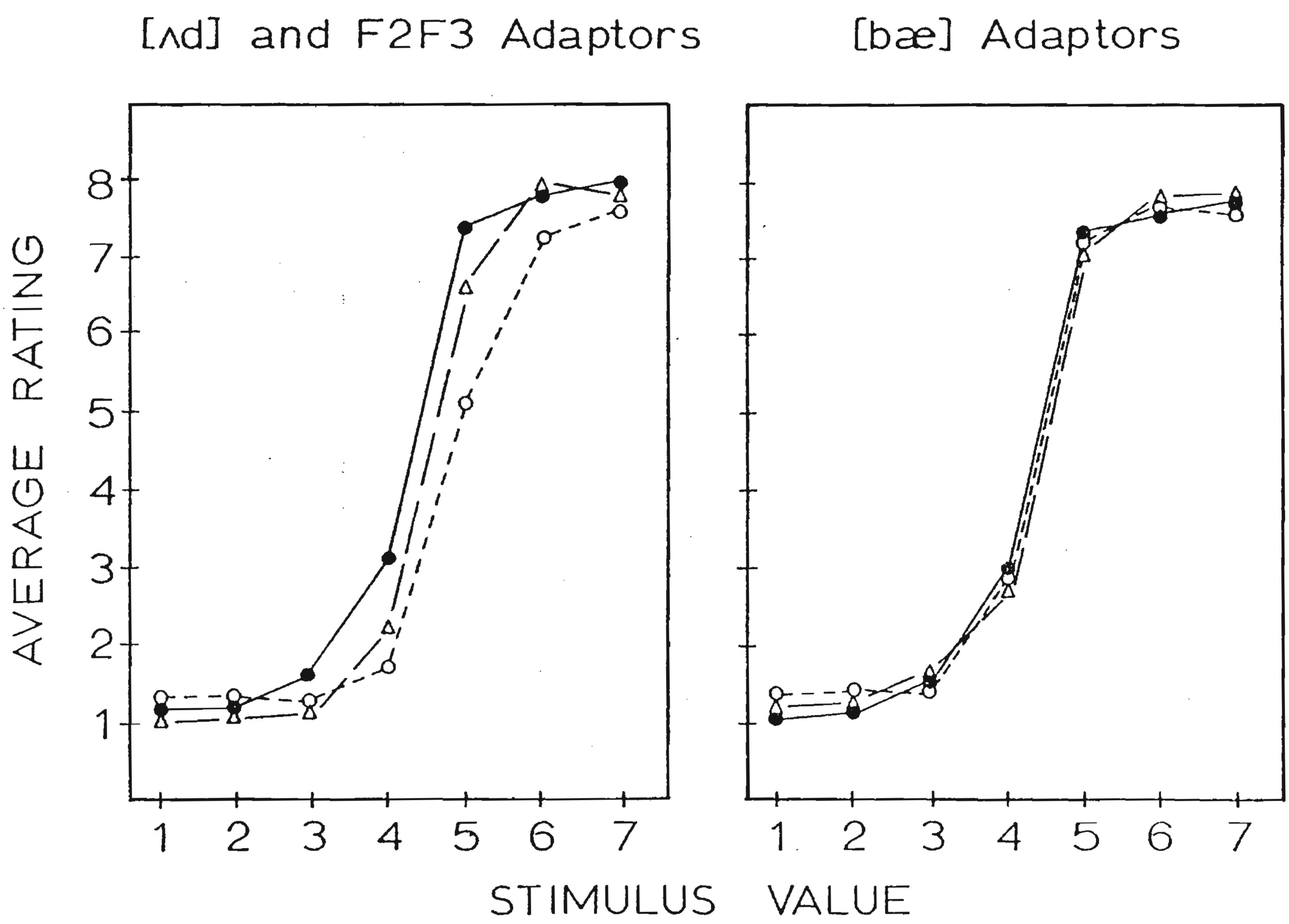

[ʌd] and F2F3 Adaptors
[bæ] Adaptors
AVERAGE RATING
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 17.

Experiment 5A

Method

Subjects. The same ten subjects who had participated in the previous study were asked to return for one more day of testing.

Stimuli. The same two sets of synthetic stimuli that were used in Experiment 5 were also used again here.

Procedure. The subjects were run in the same two groups of five with the same test series (either [bae]-[dae] or [Λb]-[Λd]) that they had listened to previously. They received two sequences of adaptation trials. Each sequence consisted first of 200 presentations of the adaptor ([bae] for the [Λb]-[Λd] test group, [Λd] for the [bae]-[dae] test group) which served as a warm-up. Following the warm up, 10 blocks of adaptation and testing were run. Each block consisted of 100 presentations of the adaptor with 300 msec between presentations. This was followed by seven syllables in random order for subjects to identify and rate as in the previous experiments. The use of 200 warm-up repetitions of the adaptor and 100 repetitions per block of testing is identical to the procedure used by Pisoni & Tash (1975).

Results and Discussion

The cross series adaptation results were virtually identical to those obtained in the main experiment. The [Λd] adaptor had no effect on the [bae]-[dae] test series and the [bae] adaptor had no effect on the [Λb]-[Λd] series. The average adapted rating data appear in the right hand panels of Figures 16 and 17 respectively. From these data, one can conclude that the difference between the results of Pisoni & Tash (1975) and the present findings is probably not the result of differences in procedure. One possible explanation of the adaptation results found by Pisoni & Tash (1975) has already been mentioned. Ades (1976) has suggested that the adaptation result arose from a rapid amplitude onset of the steady state formants, prior to the final transitions. Another explanation, however, is also possible. The adapting effects of the SEChirp could have been due to the first formant of the SEChirp. As mentioned previously, the first formant had a rising transition in final position. Thus, it mimicked the rising formants in initial position in the [ba]-[da] test series. This gives rise to yet another possible explanation of the adapting effect of the SEChirp. Since the first formant transition rose in final position, subjects may not have heard the SEChirp as a VC syllable (or even as speech-like). Since, on the basis of the experiments in Chapter III, adaptation takes place at two levels of speech processing, one frequency specific and the other integrative, and if an adaptor must be heard as speech to affect the second level, then the SEChirp may have adapted the auditory level only and resulted in the boundary shift. In contrast, the effect of an [ʌd] adaptor on a [bae]-[dae] series could be to adapt both levels, but in opposite directions. If the two levels of adaptation effectively canceled each other, no resultant adaptation effect would be observed. The next experiment was an attempt to differentiate which, if any, of these interpretations is the correct one.

Experiment 6

In order to investigate the question of whether acoustic level detectors are position sensitive and to resolve the conflicting interpretations of the SEChirp results, five new adapting stimuli were constructed. Four of them are shown as schematized spectrograms in Figure 18. A SEChirp, similar to that used by Pisoni & Tash, was constructed from the [bae] end of the [bae]-[dae] series used in Experiment 5. A second SEChirp, labeled F2F3-SEChirp, was constructed with a constant F1 and no final rising transition. To explore the explanation of Ades (1976) that the steady state onsets were causing the adaptation effect, two further vowel like stimuli were constructed. Both were three formant steady state patterns, the only difference being the center frequency of the first formant. Both of these stimuli had an abrupt amplitude onset similar to the amplitude onset of the [bae]-[dae] test series. A fifth adaptor, consisting of wide band noise with a 10 kHz upper frequency limit was also used to check for any overall change in the rating responses of subjects due to repeated presentation of any acoustic stimulus. These various adaptors therefore permit a strong test of the role of the first formant transition and the onset amplitude in determining selective adaptation effects.

Method

Subjects. The subjects were 30 new paid volunteers who responded to an ad in the student newspaper. Some of the subjects had limited previous experience with synthetic speech. All subjects met the usual requirements and were paid at the rate of $2/hr.

Stimuli. The seven step [bae]-[dae] series from Experiment 5 was also used in this study. Four additional adapting stimuli were also constructed on the OVE synthesizer. A b-SEChirp was constructed from the [bae] end of the test series by placing 205 msec duration steady-state formants in front of the frequency transitions. These formants

Figure 18. Schematic spectrograms of four of the adapting stimuli used in Experiment 6.

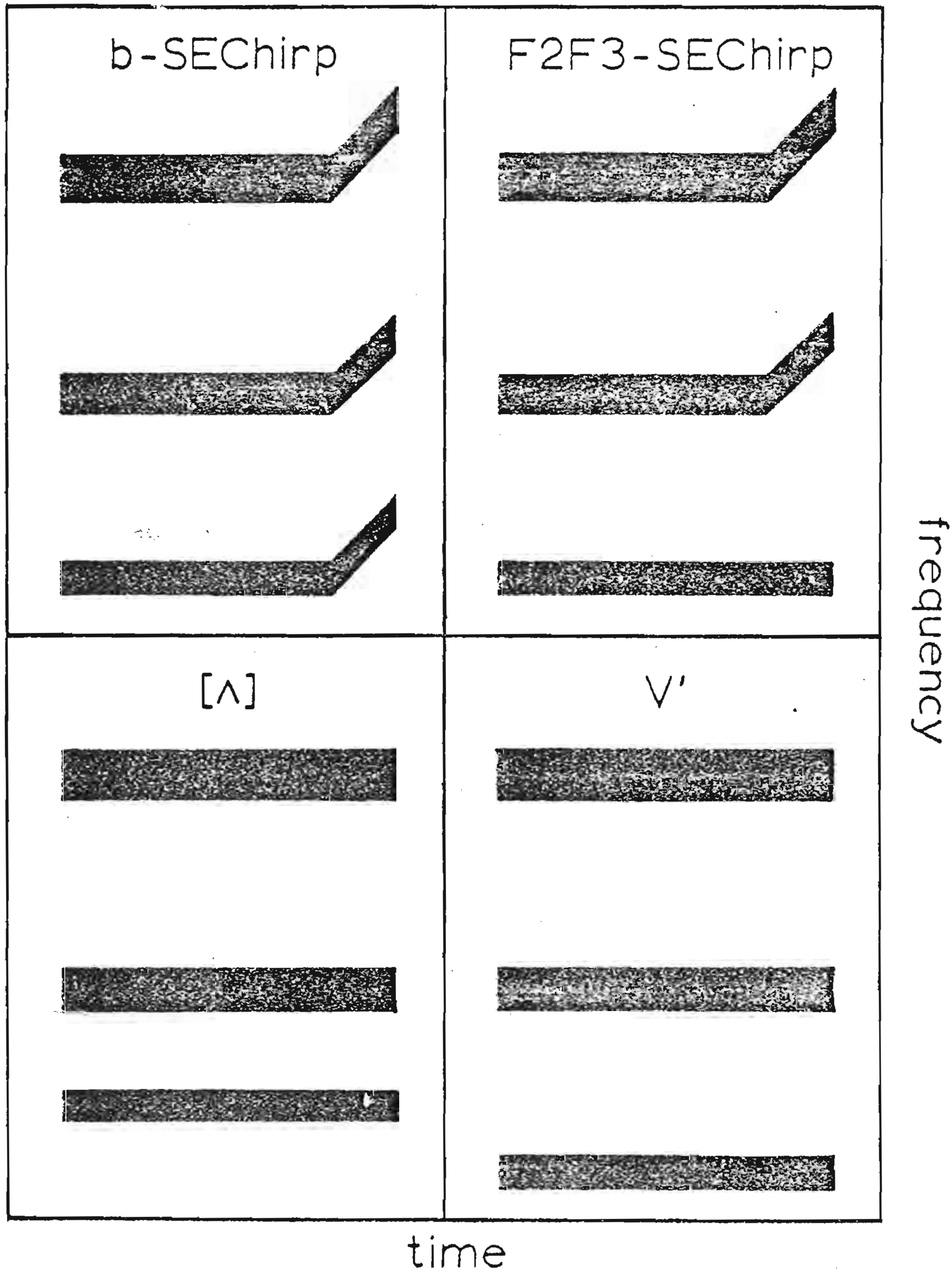


Figure 18.

had center frequencies of 251 Hz, 1198 Hz, and 2262 Hz. The steady state formants following the transitions were then deleted leaving a 250 msec stimulus. A second SEChirp, labeled the F2F3-SEChirp, was constructed from the b-SEChirp. The final formant transition for the first formant was removed and replaced by a steady state formant at 251 Hz. The amplitude onset for both of the SEChirps was gradually ramped on over the first 50 msec of the initial steady-state.

Two vowel like stimuli were also constructed from the [bae] end of the [bae]-[dae] series. One of these, labeled V' in Figure 18, had steady state values of 251 Hz, 1198 Hz and 2262 Hz for the first three formants. These values are identical to the starting frequencies of the three formant transitions for the [bae] end of the [bae]-[dae] series. When synthesized, this stimulus had a distinct tonal quality and sounded something like the vowel [u]. A second vowel, labeled [ʌ] in Figure 18, had formant center frequencies of 653 Hz, 1198 Hz, and 2262 Hz. These values matched the first formant steady state of the [ʌ] vowel to the first formant steady state of the [ae] portion of the [bae]-[dae] test stimuli. The second and third formant frequencies of the [ʌ] vowel, like the stimulus V', were identical to the onset frequencies of the second and third formant transitions at the [bae] end of the [bae]-[dae] test series. Both vowels had a duration of 250 msec and an amplitude onset similar to that of the [bae]-[dae] test series. All four of these adaptors had a constant fundamental frequency of 118 Hz.

The noise adaptor was constructed by repeated sampling of 250 msec segments from a noise generator (Grason-Stadler, Model 1724) with the upper frequency cutoff set at 10 kHz. Thus, the exact composition of the noise adaptor was random from presentation to presentation.

Procedure. All experimental events were controlled by a PDP-11 computer. Subjects were run in six groups of five each in one 1-hour session. Each group received a different adaptor (except the two groups who received the noise adaptor). The identification and test sequences were essentially identical to those of Experiment 5 except that only the [baê]-[dae] series was used for testing. Since subjects were run on one day only, the identification sequence was presented twice before the adaptation trials were begun.

Results and Discussion

The results of using the b-SEChirp as an adaptor are shown in the left hand panel of Figure 19. The b-SEChirp had a significant effect in moving the [bae]-[dae] boundary toward [bae] ($t(4) = 4.23$, $p < .01$). This result replicates the findings of Pisoni & Tash (1975) with one important difference. Since the steady state portion of this adaptor was gradually ramped on over a 50 msec period, it is unlikely that the adaptation effect was caused by the steady-state onset of the vowel. Rather, the adaptation effect seems to be the result of the presence of final transitions.

The F2F3-SEChirp had a marginally smaller adapting effect than the b-SEChirp. The rating data for this group appear in the right hand panel of Figure 19. The F2F3-SEChirp moved the [bae]-[dae] category boundary toward [bae] ($t(4) = 3.18$, $p < .025$). Further, the F2F3-SEChirp and the b-SEChirp did not differ significantly in the amount of shift they induced in the category boundary.[7] This indicates that the upturned first formant was not a primary contributor to the adaptation effect found by Pisoni & Tash (1975) and in the present

Figure 19. Unadapted (solid circles) and SEChirp adapted (open circles) rating functions for the two SEChirp adapted groups from Experiment 6.

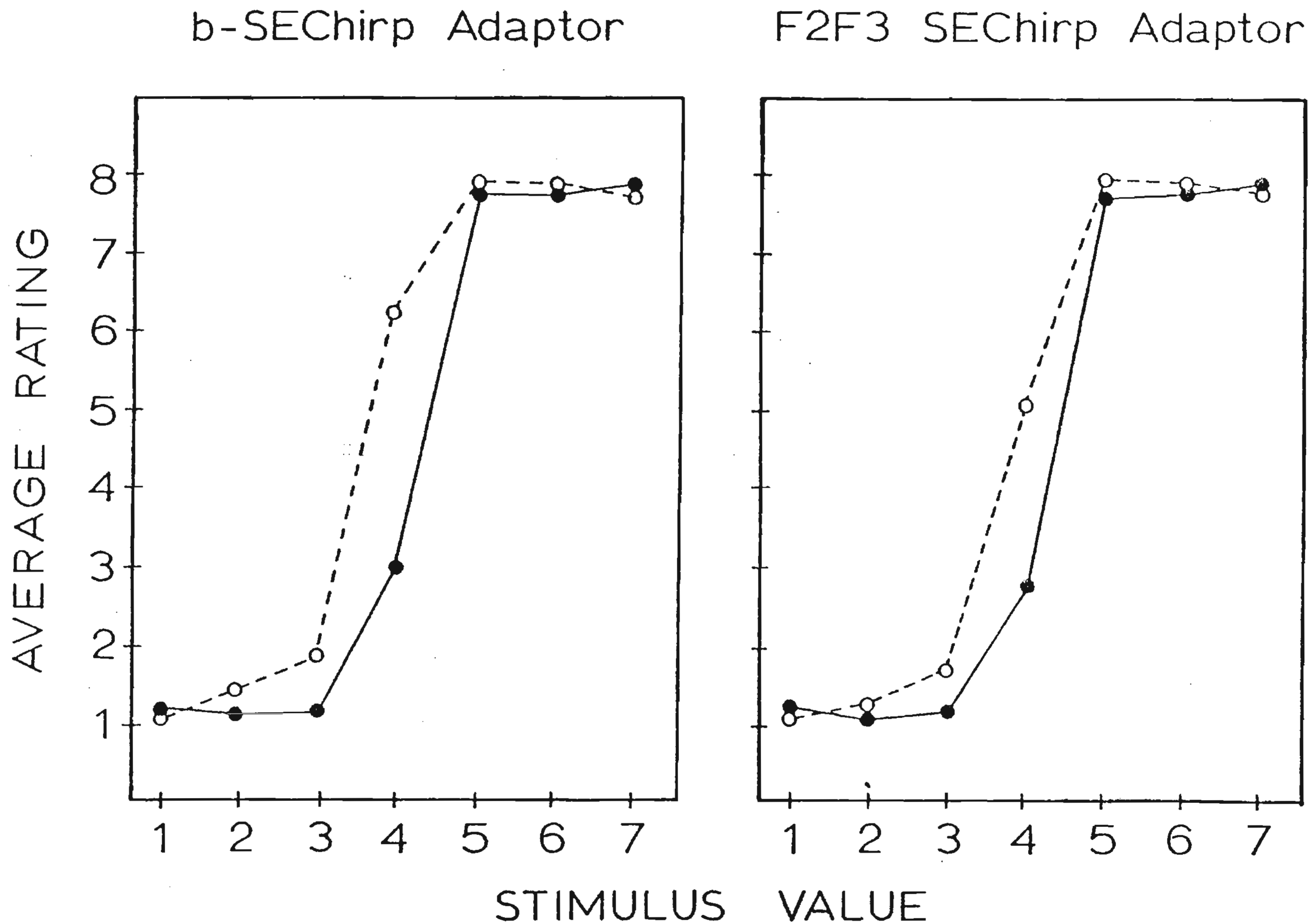

b-SEChirp Adaptor
F2F3 SEChirp Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 19.

experiment with the b-SEChirp. Instead, it seems to be the lack of a "normal" or falling first formant that determines whether adaptation effects take place. This point about the importance of the first formant will be considered again below.

The results of adaptation with the vowel-like stimuli V' and [ʌ] appear in Figure 20. Neither adaptor had an effect on the average rating responses of subjects to the [bae]-[dae] test series. This direct test permits a rejection of Ades' (1976) claim that the vowel onset of the SEChirp's could have caused the adaptation effects observed by Pisoni & Tash (1975). The formant transitions in final position are indeed the cause of the adaptation observed with SEChirps.

The lack of any adaptation effect with the vowel-like stimuli also has an interesting implication in connection with a recent account of place of articulation proposed by Stevens & Blumstein (1976). According to these authors the critical cue to place perception in stop consonants is the spectrum at onset (or offset). The bilabial stops [b,p] are characterized by a concentration of energy at relatively low frequencies. The alveolars [d,t] and the velars [g,k] are characterized by concentrations at high and middle frequency regions respectively. Detectors sensitive to onset (offset) would respond differentially to these relative energy concentrations. Thus, these detectors offer a potential resolution of the invariance issue for the stop-consonants. The V' adaptation condition in Experiment 6 was a direct test of this proposed detector system. The model of Stevens & Blumstein (1976) would characterize the V' adaptor from the present experiment as being roughly equivalent to the [bae] end-point syllable at onset. Both

Figure 20. Unadapted (solid circles) and vowel adapted (open circles) rating functions for the [bae]-[dae] series from Experiment 6.

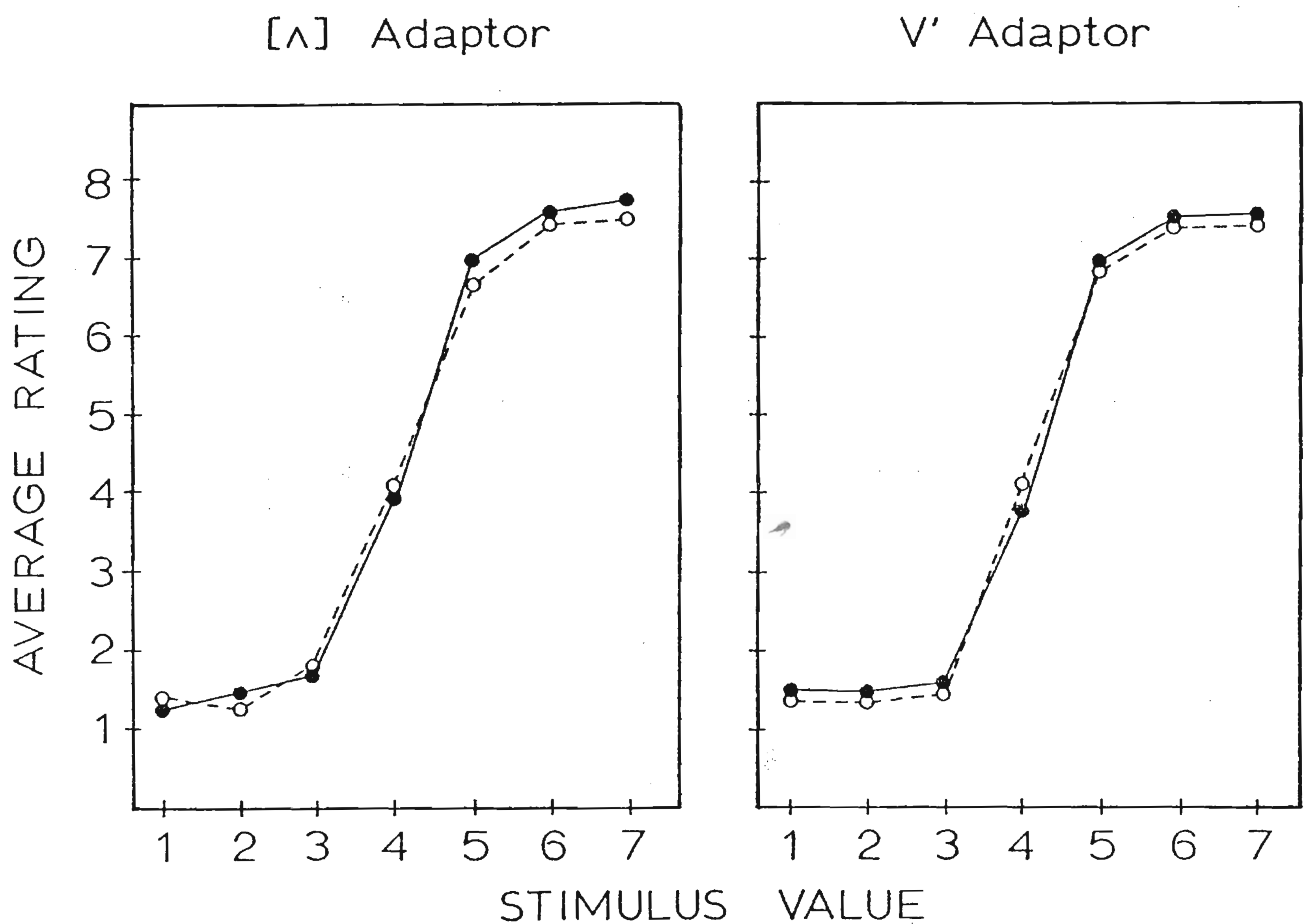
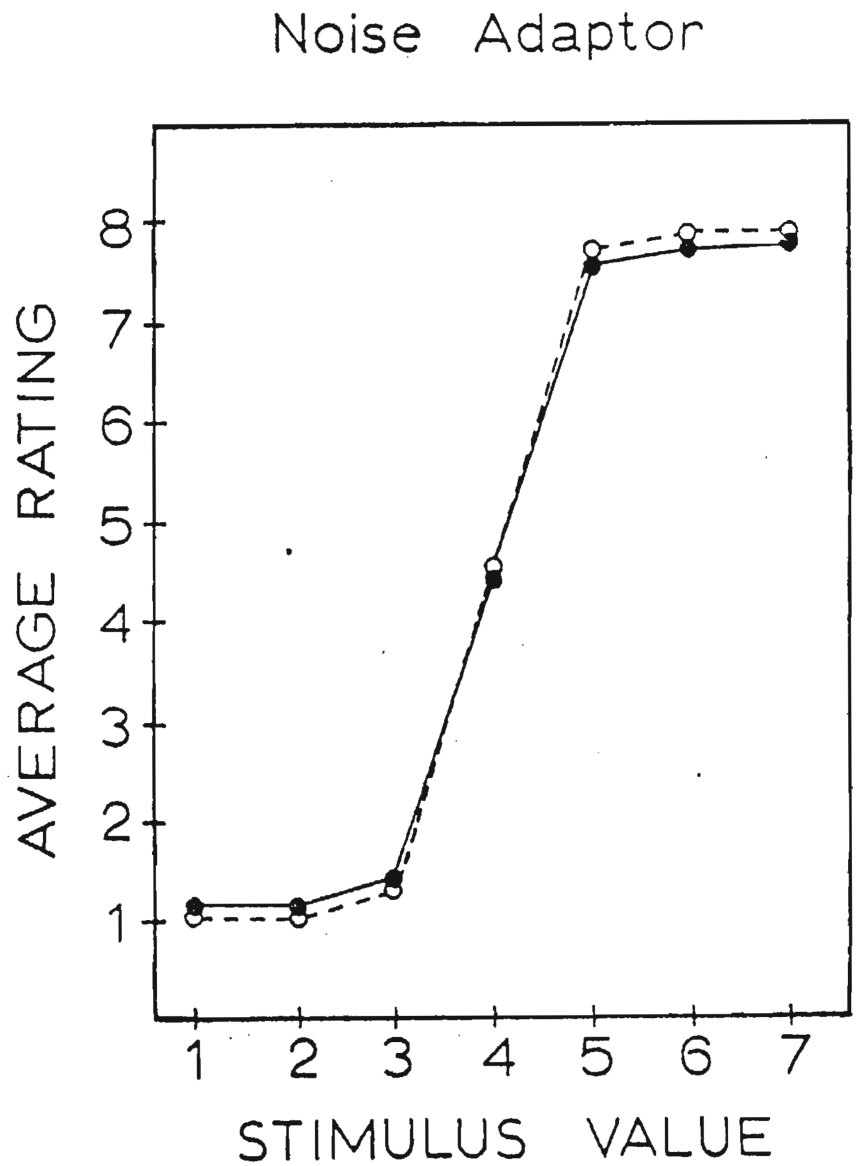

[ʌ] Adaptor
V' Adaptor
AVERAGE RATING
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 20.

of these stimuli have the same formant frequencies at onset and both have an abrupt amplitude onset. Thus, one particular onset-offset detector should respond to both V' and [bae]. However, the V' adaptor had no effect on the [bae]-[dae] series, indicating that the V' Stimulus and the [bae]-[dae] test stimuli apparently are processed by distinct and separate mechanisms through the peripheral and integrative auditory levels. Consequently, the approach taken by Stevens & Blumstein is not a description of the processing at either the peripheral or integrative levels. Their account could, however, be applicable to a still later stage of analysis, such as the decision rules at the phonetic feature combination level in the present formulation.

The final adaptation condition, where a wide band white noise burst was used as the adaptor also showed no adaptation. The unadapted and noise adapted identification functions for this group are shown in Figure 21. The noise bursts had no effect on the [bae]-[dae] rating responses either within categories or at the category boundary. If adaptation with the noise bursts had been observed, one could argue that transition detectors were simply summing energy over time within very broad frequency regions. The absence of any effect with the noise adaptor indicates that the detectors do not simply sum energy. Instead, they respond to a particular pattern of frequencies over time. The lack of an adaptation result with wide band noise suggests that these detectors may exhibit the inhibitory surround characteristics found for visual detectors (Hubel & Wiesel, 1965). The inhibitory surround means that only a particular pattern of frequencies over time will elicit a response from these detectors. Without the inhibitory surround, a band of noise would

Figure 21. Unadapted (solid circles) and 250 msec wide band noise burst adapted (open circles) rating functions for the [bae]-[dae] series.



Noise Adaptor
AVERAGE RATING
1
2
3
4
5
6
7
8
1
2
3
4
5
6
7
STIMULUS VALUE

Figure 21.

be registered by these detectors since it would contain the appropriate frequency components (as well as many adjacent, irrelevant components).

General Discussion: Experiments 5 and 6

The overall pattern of results found in Experiments 5 and 6 fits very well into a two level framework. From the SEChirp results of Experiment 6 and those of Pisoni & Tash (1975), it appears that the transition detector mechanisms at the peripheral auditory level are on "free-run." That is, they will respond whenever an appropriate frequency transition occurs, regardless of whether it is preceded by a steady-state (SEChirp results of Experiment 6), followed by a steady-state (from Experiment 5) or without a steady-state (F2F3 from Experiment 5). Further evidence that the adaptation with the SEChirps and the F2F3 chirp takes place at an auditory level comes from the directionality of the category boundary shifts. The F2F3 chirp shifted the [bae]-[dae] boundary toward [bae] but the [Λb]-[Λd] boundary moved toward [Λd]. These are precisely the directions expected at a frequency specific level of processing due to the spectral overlap between adaptor and test series. The shifts were in opposite phonetic directions in the two series, indicating that the source of the effect is definitely prior to phonetic categorization. A similar line of reasoning leads to the same conclusion regarding the effect of the SEChirps.

From the experiments in the previous chapter, it is clear that a central integrative level is present in selective adaptation. On the basis of Experiments 5 and 6, it is unclear whether adaptation at this level is sensitive to syllable position. One possible explanation of the failure of a CV to adapt a VC (and vice versa) is that the effects at the auditory level and the integrative level may cancel each other. From Experiments 1, 2, and 3 the magnitudes of the adaptation effects at the peripheral and central levels appear to be approximately equal. For a CV versus a VC, these levels could produce shifts in the opposite directions and the net effect of adapting on a CV and testing on a VC would be cancellation of any category boundary shift.

Another possible explanation of the failure of a CV to adapt a VC (and vice versa) is that no adaptation occurs at the integrative level between a CV and a VC. The reason that no auditory component of adaptation appears could be due to a differential weighting of the transitions as cues to various stops in final (as opposed to initial) position. The adaptation of the auditory level component would simply not be transmitted through the integrative level. A choice between these alternative explanations must await further research, possibly with stimuli like those used in Experiments 1 and 2 which apparently bypass the auditory component. Using such stimuli as VC adaptors on a CV test series it should be possible to directly test for whether the integrative component does somehow differentially respond in a CV versus a VC environment.

On the basis of these experiments, several conclusions seem warranted. The absence of adaptation with vowel-like stimuli whose formants corresponded to the starting frequencies of the consonantal transitions in the test series indicates that the adaptation effects observed with SEChirps are indeed due to the presence of a rapid spectral change, regardless of whether this change occurs in syllable initial or final position. Consequently, the auditory component of selective adaptation appears to be on free-run. These detectors respond regardless of where the transition information occurs within the syllable. The only distinguishing characteristic between the SEChirps and the [ʌd] syllable was the direction of the final first formant transition. However, the spectrally matched VC syllable failed to adapt a CV series, indicating that the integrative level is influenced by the direction of the first formant transition. However, the first formant is generally considered to be a cue to manner of production rather than place in stop consonants. Consequently, the integrative level may indeed represent a non-additive summation of a number of auditory cues. This brings the plausibility of a feature detector model at this integrative level into question. A better formulation of the integrative level found in Experiments 1 and 2 may simply be in terms of a set of decision rules. These rules would integrate a number of cues from the auditory level to form a complex, abstract auditory feature.

In summary, these experiments have demonstrated the involvement of two distinct levels of processing within selective adaptation to place. A number of the characteristics of these levels have been deduced on the basis of these and previous experimental results. These characteristics have been incorporated into a model of the perceptual processing of place information in stop consonants. In the following chapter, this model will be explained in detail and evaluated through the use of computer simulation.

Footnotes

[6] This two formant chirp was produced by a computer program designed to act as a parallel resonance synthesizer. This program was written by Dennis Klatt at MIT who assisted in the preparation of this stimulus. His help is gratefully acknowledged.

[7] $\underline{t}(8) = .519$, $\underline{p} > .5$ using a two-tailed t-test for independent measures.

CHAPTER V

A Model for Selective Adaptation on the Place Feature

The results of Experiments 1 through 4 offer convincing support for the involvement of two levels of processing in selective adaptation of the place feature in stop consonants. Experiments 5 and 6 explored the operation of these levels even further. On the basis of the results of these experiments, a picture of the operation of these two levels of processing has emerged. Both of these levels respond to the formant frequency transitions that have previously been shown to be sufficient cues for the various stop consonants. However, they respond to these transitions with separate and distinct characteristics. In the sections below these hypothesized stages and their processes will be outlined and then the implementation of this model as a computer simulation will be described.

Peripheral Auditory Analysis

The peripheral auditory stage of processing is distinguished by a number of characteristics. First, the location of this stage within the auditory system is assumed to be peripheral in that it receives input predominantly from one ear. The results of Experiments 1 and 2 also indicate that this level is frequency specific. A particular detector at this level would respond only to a formant transition or rapid spectrum change that passed through its frequency range. Experiments 1 and 2 further suggest that the width of this frequency range is on the order of one critical bandwidth. One other characteristic seems to distinguish this level. The results of Experiments 5 and 6 indicate that the transition detectors at this stage are on "free run." These detectors are not specific to syllable onset or offset. Transitions in final position will adapt those in initial position. Consequently, this peripheral level does not depend upon prior segmentation of the input in order to operate. All of these characteristics described so far fit rather well with a feature detector description of this level of processing.

Integrative Auditory Analysis.

Where the stage of processing just described was characterized as peripheral and frequency specific, the integrative level is just the opposite. This level reflects the operation of central mechanisms in speech perception. From the results of Experiments 1, 2, and 4, the two distinguishing characteristics of this level are that it integrates transitional information over a relatively wide frequency range and that it is centrally located within the auditory system.

This level has been described as auditory rather than phonetic for several reasons. The results of Experiment 3 indicate that this level responds to the intensity of the adaptor. This response to a relatively low level characteristic of the acoustic waveform does not fit well with the idea of an abstract phonetic level of processing that is independent of context. A second reason for describing this level as auditory is the failure of a CV to adapt a VC series and vice versa (Ades, 1974b). This result, taken together with the small adaptation effects of a VC on a CV series found by Pisoni (1975b) indicates that this integrative level is not an invariant level of processing. Instead, the processing at this level is somehow sensitive to syllable position. This conclusion is supported by the results obtained in Experiments 5 and 6 where the first formant transition was shown to be a factor in the adaptation effects found. These findings suggest that the presence of an appropriate first formant transition may be necessary to trigger the mechanism at this integrative level. The first formant transition is a cue to the class of stop consonants although it is not generally considered to be a cue to the place feature. Consequently, a description of the processing at this level in terms of phonetic features may be somewhat inappropriate.

It also seems inappropriate to characterize the processing at this level in terms of feature detectors, for much the same reasons already cited. An interpretation in terms of a set of decision rules, where such contextual factors as the presence of a first formant transition may be included seems to be more reasonable. In this respect, this level of abstract auditory coding is equivalent to what Pisoni & Sawusch (1975) have previously labeled Phonetic Feature Analysis.

These two levels of auditory coding in speech perception fit well into the framework outlined in the first chapter. Peripheral auditory analysis, Stage 2 in the model, represents the operation of feature detectors that extract basic spectral information such as formant transitions and steady-states. The integrative auditory level, Stage 3, represents the extraction of particular patterns from this set of auditory features. These patterns are then available for processing into a phonetic feature code at the phonetic feature combination stage. This formulation appears to be able to account for the selective adaptation results on the place feature, at least in a qualitative fashion.

Further investigation is needed, however, for both the place feature and for other features in consonants. The results of Experiments 5 and 6 suggest that the processing at the abstract integrative auditory level may be contingent upon the position of the transitions within the syllable. Further exploration of this possibility should be undertaken using either the criterion of interaural transfer of adaptation or a change in the frequency composition and overlap between adaptor and test syllables. These methods, which proved fruitful in Experiments 1 through 4, have not been extensively applied to the CV-VC adaptation question and are worthy of additional attention.

The use of interaural transfer and manipulation of the spectral composition of adaptor and test series should also be applied to other features besides place of articulation. As reviewed in Chapter II, Eimas et al. (1973) did adapt on VOT and found 90% interaural transfer. However, this was only obtained when the adaptor was the voiceless stop [t]. A voiced stop (e.g. [d]) was not used as an adaptor in their experiment. The use of a voiced stop adaptor in the interaural transfer paradigm offers a possible test of the two process account of voicing perception (Summerfield, 1974, 1975; Ades, 1976). The perception of voiceless stops is mediated by centrally located detectors that are sensitive to relative onset. In contrast, the perception of voiced stops is accomplished through detectors that are sensitive to the onset frequency and transitions of the first formant. From Experiments 1 through 4 it might be expected that both peripheral and central levels of processing would be found for transition detectors that registered first formant information. If a voiced adaptor exhibited less interaural transfer than that previously found for a voiceless adaptor, additional support would be found for the two process model of voicing perception.

The results of experiments on the stop versus semi-vowel distinction are also subject to several possible interpretations. Some experiments have sought to manipulate the spectral nature of the adapting and test syllables (Bailey, 1975; Cooper et al., 1976; Diehl, 1976). In general, these experiments offer strong support for a frequency specific level of adaptation and lesser, equivocal, support for an integrative level. Further research on this dimension, using the interaural transfer paradigm and careful spectral overlap manipulations involving the critical bandwidth is needed. However, based on the results of the present experiments, it would not be surprising if evidence for both a frequency specific (and possibly rate and duration specific) component and a separate integrative component was found for the stop versus semi-vowel distinction.

Computer Simulation of Place Perception in Stops

The model outlined in Chapter I and summarized above was programmed onto a PDP-11 computer for simulation purposes. This was done for a number of reasons. First, we hoped to be able to make quantitative predictions regarding the change in identification of a series of stimuli as a function of various selective adaptation conditions. Such predictions were carried out for the data already collected in Experiments 1, 2, 5 and 6. The simulation can also potentially be used to predict, in advance, the identification and adaptation of new sets of speech stimuli. Since the simulation is, in essence, a recognition system modeled after the human perceptual data, it could also be used for testing new synthetic speech stimuli prior to experimentation with subjects. The second and perhaps most important reason for implementing the simulation was to force a formalization of the sometimes vague ideas that have gone into this model and other similar models in speech perception. The present model is a rather complex set of verbal descriptions of processes that are thought to take place in speech perception. In implementing these verbal descriptions as a computer program, precise specification of the details of this model was necessary. One consequence of this is that it focuses attention on problems that are not yet adequately understood in speech perception, often revealing new problems.

The simulation program involved Stages 2 through 4 of the model outlined previously in Figure 1. Since neither filters nor a fast-Fourier transform program were available at the time, the first stage involving preliminary auditory analysis was not implemented as part of the simulation. Instead, the input was taken from a separate program which acted upon the stimulus parameter files for the OVE synthesizer. This program converted the OVE parameter files into a coding of intensity (in dB) per critical band over time using the Acoustic Theory of Speech Production (Fant, 1960). Since the OVE IIId synthesizer is built around the same principles (Liljencrants, 1968), this conversion of the OVE parameter files provides a fairly reasonable spectrographic description of the stimulus. The particular critical band center frequencies selected and their associated bandwidths are shown in Table 6. For convenience, the simulation was restricted to the frequency range from 50 Hz to 4 kHz since the formant information for the consonants used in these experiments is contained in this region. If detectors for burst information

Table 6

Critical band center frequency and width during preliminary auditory analysis.

Number	Center Frequency	Width	Lower Cutoff	Upper Cutoff
1	50	80	20	100
2	100	100	50	150
3	150	100	100	200
4	200	100	150	250
5	250	100	200	300
6	300	100	250	350
7	350	100	300	400
8	400	100	350	450
9	450	110	400	510
10	510	120	450	570
11	570	120	510	630
12	630	130	570	700
13	700	140	630	770
14	770	140	700	840
15	840	150	770	920
16	920	160	840	1000
17	1000	160	920	1080
18	1080	170	1000	1170
19	1170	190	1080	1270
20	1270	200	1170	1370
21	1370	210	1270	1480

Table 6, continued

Number	Center Frequency	Width	Lower Cutoff	Upper Cutoff
22	1480	230	1370	1600
23	1600	240	1480	1720
24	1720	250	1600	1850
25	1850	280	1720	2000
26	2000	300	1850	2150
27	2150	320	2000	2320
28	2320	350	2150	2500
29	2500	380	2320	2700
30	2700	400	2500	2900
31	2900	450	2700	3150
32	3150	500	2900	3400
33	3400	550	3150	3700
34	3700	600	3400	4000
35	4000	700	3700	4400
36	4400	800	4000	4800
37	4800	900	4400	5300
38	5300	1000	4800	5800
39	5800	1100	5300	6400
40	6400	1200	5800	7000
41	7000	1350	6400	7750
42	7750	1500	7000	8500
43	8500	1750	7750	9500

were to be added to the simulation, the frequency range would have to be extended to 8 or 10 kHz. An example of an input to the simulation in the form of intensity (in dB) per critical band over time is shown in Appendix C. This represents the low [bae] end of the [bae]-[dae] test series from Experiments 1 through 3.

The basic simulation. The model was simulated in two phases. The first phase concerned the operation of the model under normal unadapted conditions of identification. The second phase dealt specifically with the effects of selective adaptation. The entire simulation program, written in Fortran, appears in Appendix B. In the unadapted state, the spectral input is recoded through the stages of Peripheral Auditory Analysis, Abstract Auditory Analysis and Phonetic Feature Combination to yield a distinctive feature matrix. For the purposes of simulation, these three stages were treated as being successive and serial. This assumption is not, however, necessarily correct. The two auditory analysis stages could conceivably operate entirely in parallel. Processing at the phonetic level in this model could also begin as soon as the two prior auditory stages had provided some information about the input. The question of serial versus parallel processing is important. However, no direct empirical evidence as to whether the peripheral and abstract auditory levels operate in sequential or parallel fashion for stop consonants has been found yet.

At the Peripheral Auditory Analysis stage, five sets of feature detectors were implemented. These sets corresponded to rising, falling and steady-state formant detectors. Within each of these sets a particular detector was sensitive to relative amplitude peaks in a particular frequency region. The steady-state detectors responded to an amplitude peak at a constant frequency. For rising and falling transitions, the frequency change could be either rapid or gradual; these were implemented as separate detector sets. For a more general model, where the rate of transition could be an important cue, the detectors at this level could be broken down into a finer set of rate specific channels. Schematic illustrations of the patterns of amplitude x frequency x time relations that were registered by the various sets of detectors are shown in Figure 22.

A number of assumptions, such as the time duration over which these detectors operate, how many of these detectors are present (i.e., how closely they are spaced in the frequency domain), how frequently these detectors operate and the statistical distribution of their operation had to be resolved. The detectors were spaced in the frequency domain at half critical bandwidths. This provided an overlapping between adjacent detectors in their regions of sensitivity and partially resolved the distribution question raised above. This particular arrangement was chosen as a compromise between the proliferation of a relatively large number of detectors very closely spaced and the need for sufficient resolution in the frequency domain.

The detectors illustrated in Figure 22 operate over a 25 msec time span or window. The choice of 25 msec is somewhat arbitrary, but seemed to be in the correct range. It is sufficiently small to be able to resolve 15 msec formant transitions that are often present in natural speech. These detectors also scan the spectral input once every 10 msec. The choice of 10 msec between applications of these detectors to the input is

Figure 22. Examples of formant patterns that were registered by various peripheral level detectors in the computer simulation.

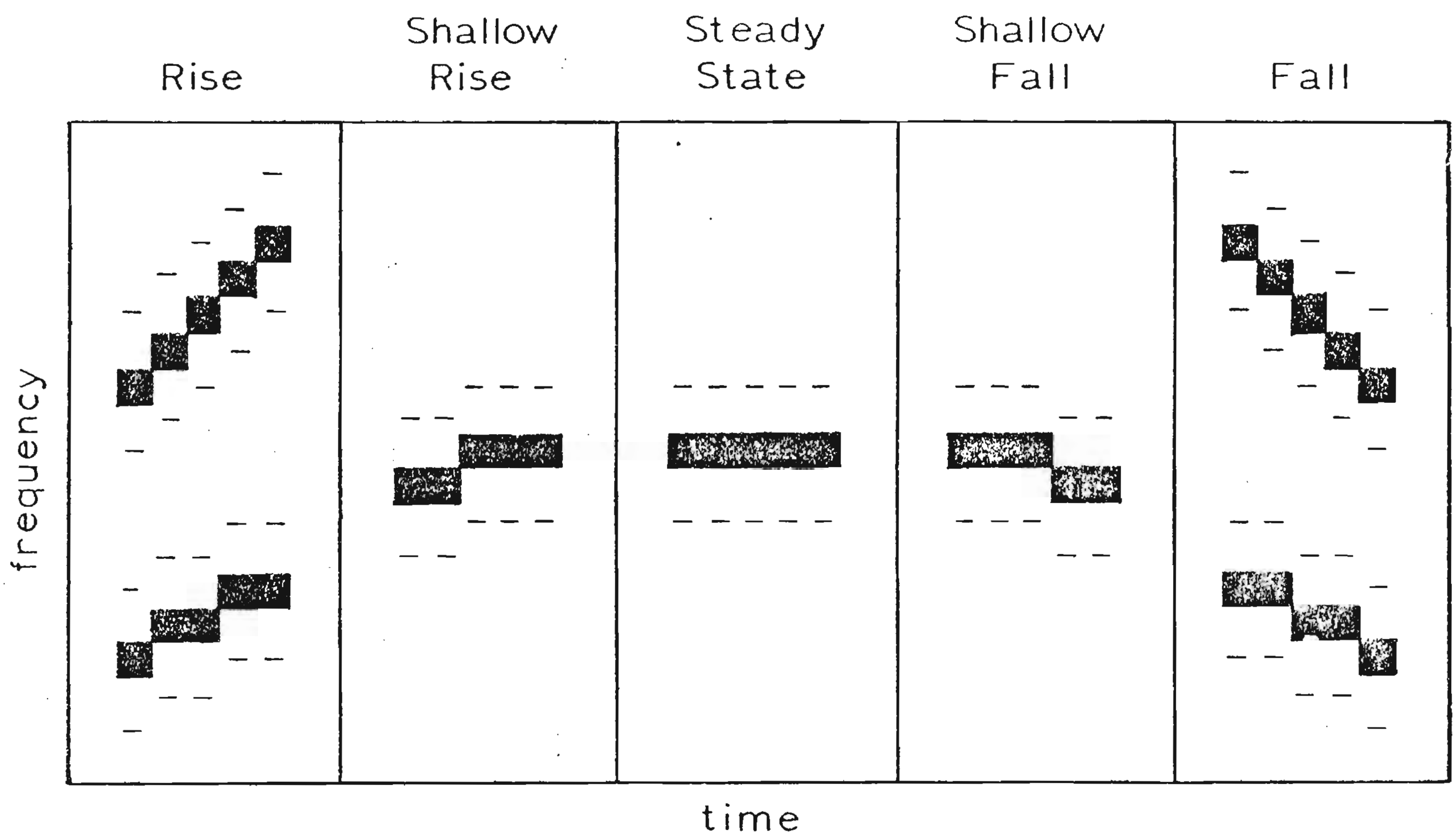

Rise
Shallow Rise
Steady State
Shallow Fall
Fall
frequency
time

Figure 22.

also somewhat arbitrary. A longer (15-20 msec) time between the scanning of the input by these detectors could be used, but at some loss in information retrieved. A shorter interval could also be used. However, the only gain would be to increase the redundancy of the matching process. In general, these two time constants of 25 msec for the window duration and 10 msec between samples were chosen for convenience because little information about such factors in the perception of complex auditory stimuli was currently available.

One other peripheral auditory level detector was implemented. A low frequency energy onset-offset (voicing) detector was also programmed. This detector registered the increase in energy (amplitude) in the lowest frequency bands between successive 15 msec samples. Again, as before, the choice of a particular time constant, in this case 15 msec, was somewhat arbitrary. A longer time constant could also be used. This onset-offset detector served as an indicator for the abruptness of onset and offset in the spectral waveform. It could also be potentially used to detect the presence of voicing in stop consonants.

The Abstract Auditory Level was implemented in the simulation in terms of a set of decision rules. Separate recognition rules were formulated for transitions in initial position and transitions in final position. For initial position, three rule systems were formulated: 1) rising transitions; 2) falling transitions; and 3) diverging transitions (high frequency rise and lower frequency fall). There were also three rule systems for final position: 1) falling transitions; 2) rising transitions; and 3) converging transitions (high frequency fall and lower frequency rise). The distinguishing factor between these two sets of detectors (initial versus final position) was the direction of any low frequency transition. If a rising low frequency transition was present, the initial position set of rules was activated. Conversely, if a falling low frequency transition was present, the final position rules were activated. This low frequency transition information corresponds to the presence of a rising first formant for CV syllables but a falling first formant for VC syllables. This particular interaction of the first formant transition with higher formant transitions (the traditional cues to place) was based on the results of Experiments 5 and 6 where the first formant transition was found to play a role in adaptation of a CV series with VC (and VC like) syllables.

The final level, Phonetic Feature Combination, converted the outputs from the two auditory levels into a rough articulatory phonetic feature matrix. For the stop consonant place dimension, a three valued feature was used which represented the bilabial ([b]), alveolar ([d]) and velar ([g]) places of articulation for stop consonants. It is at this level that the onset-offset detector information (mentioned under the Peripheral Auditory Level above) was used. This information was used to determine which set of general detectors from the abstract level were to be employed in making a final place feature decision. If an onset had been detected, the initial general rise output would indicate a labial stop, initial general fall would indicate an alveolar stop and initial divergence a velar stop. Conversely, an offset would indicate the use of the final position abstract outputs and a different correspondence rule. For final position, a general fall in spectral energy indicated a labial stop, general rise an alveolar stop and convergence the presence of a velar stop.

As with both of the auditory levels, the operations at the phonetic level were defined over a specific time frame and recurred at a constant sampling rate. For the phonetic level, a 40 msec window was used for the phonetic feature decision rules. This 40 msec window was subject to an exponential decay function that weighted the more recent abstract auditory outputs more heavily. The exponential decay function was included to simulate the decay from memory of the abstract auditory level outputs and to give a greater weight to more recent events. The process of comparing the abstract auditory outputs recurred at 20 msec intervals, providing an overlap between successive comparisons of 20 msec. As with the time constants at earlier stages, these two time constants are somewhat arbitrary. A longer window for the combination rules could be used without much change in the overall operation of the simulation.

This description of the simulation program has, of necessity, left out a number of implementation details. However, the critical aspects of the simulation around which the model was built have, hopefully, been adequately described. Before proceeding with a description of how selective adaptation produces its effects in the model, a demonstration of the model's basic capabilities and shortcomings is appropriate. The four synthetic syllable test series that were used in Experiments 1 through 6 were fed through the spectral conversion program and then input to the computer simulation. In order to obtain an output from the simulation that was comparable to the rating response that subjects provided in these experiments, one additional assumption was needed. It was assumed that subjects, in making a rating response, made use of residual precategorical information from the syllable. A number of investigators have assumed that subjects have such information available. For example, Fujisaki & Kawashima (1970) and Pisoni (1971, 1973) have used this type of precategorical memory as a part of their explanations of the ABX discrimination results. Miller (1975) and Sawusch (1976) have previously studied the presence of information from early stages of processing in selective adaptation experiments. This information for each of the alternative response categories would be combined together by a subject to yield a rating response. Within the simulation, this process was implemented in the Phonetic Feature Combination stage. As described above, the abstract auditory coding of the input was converted into a phonetic feature matrix. The exponentially weighted abstract outputs were retained and summed over the entire syllable. They were then converted into percentages (relative to the competing stop consonant place codings), and this was scaled linearly on a 1 to 8 rating scale. This provided an output from the simulation that was directly comparable to the subject rating data.

The predicted rating functions derived from the simulation for the [bae]-[dae] series from Experiments 1, 2, and 3 and the [bi]-[di] series from Experiment 4 are shown in Figure 23. A sample output from the simulation for the [bae] end of the [bae]-[dae] test series from Experiments 1-4 is shown in Appendix C. The observed subject rating functions for the same sets of stimuli from Experiments 1 and 4 are shown for comparison. The subject data for the [bae]-[dae] series in Figure 23 are the average unadapted rating functions for all 24 subjects from Experiment 1. The fit of the predicted rating function from the simulation and the observed subject rating data for the [bae]-[dae] series was very good.[8]

Figure 23. Comparison of predicted (solid circles) and obtained (open circles) unadapted rating functions for the [bae]-[dae] series (Experiment 1) and the [bi]-[di] series (Experiment 4).

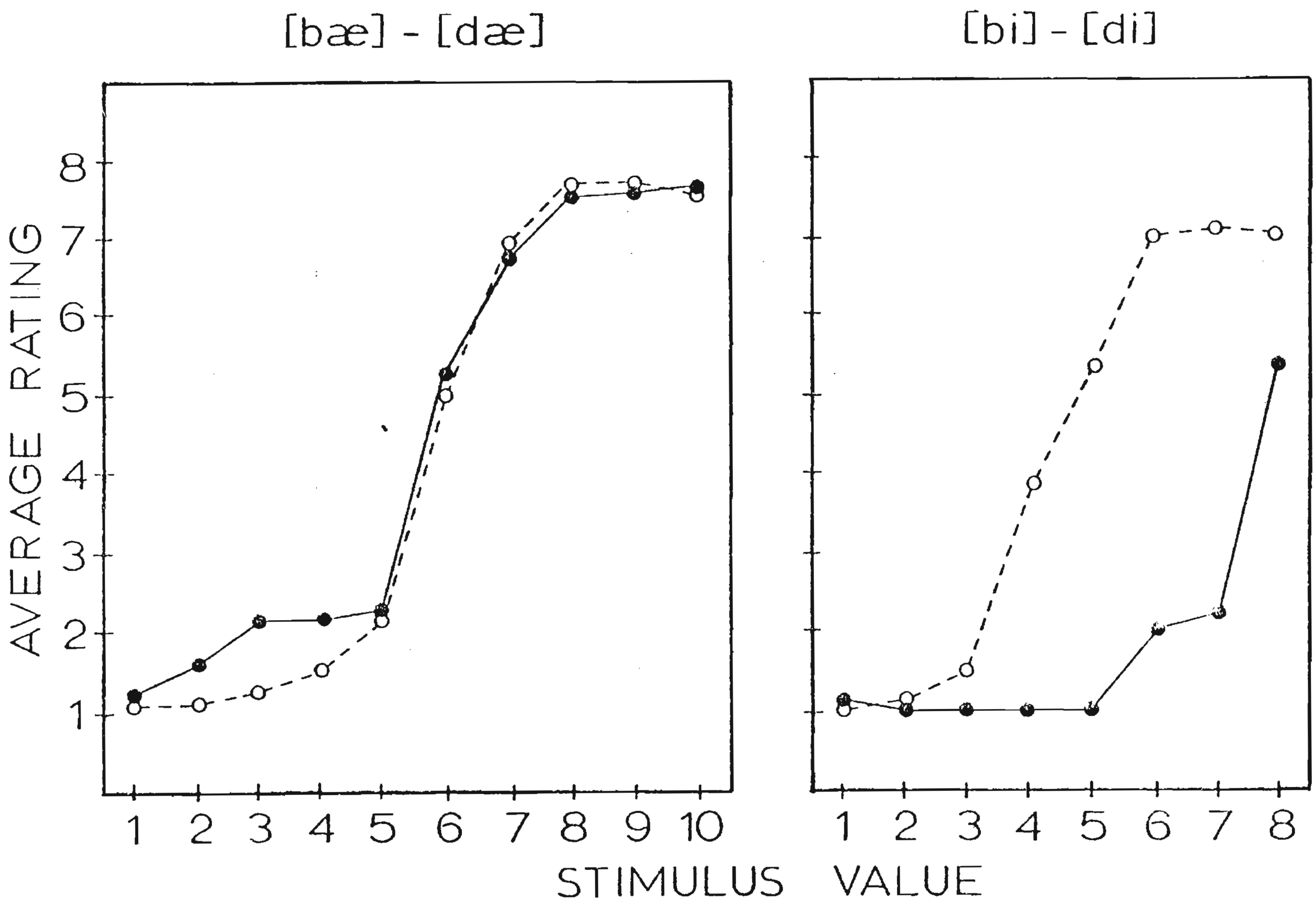

[bæ] - [dæ]
[bi] - [di]
AVERAGE RATING
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8 9 10
1 2 3 4 5 6 7 8
STIMULUS VALUE

Figure 23.

However, the fit between predicted and obtained functions for the [bi]-[di] series was very poor.[9] The reason for this poor fit is apparent from an examination of the schematized spectrograms of the [bi]-[di] end-point syllables in Figure 13. The [di] syllable is cued by a falling F3 transition and a rising F2 transition. Thus, conflicting information was presented to detectors at the abstract auditory level. This level must put together general fall and general rise information for the phonetic feature decision stage. This general rise and general fall information in the [di] syllables presented conflicting information for the model to distinguish among the syllables of the [bi]-[di] series from Experiment 4. It should be noted that the [di] syllable is one of the classical examples of the lack of invariance in speech sounds. Consequently, it is not at all surprising that the correspondence in the model between initial general rise and [b] and initial general fall and [d] should lead to erroneous recognition for this series. However, modification of the simulation is needed so that recognition of these syllables by the simulation will match subject data.

The predicted average rating functions and obtained functions for the two synthetic test series from Experiment 5 are shown in Figure 24. The fit between predicted and obtained data for both the [bae]-[dae] series and the [ʌb]-[ʌd] series was very good.[10]

As a final check on the simulation, two new series of synthetic syllables were constructed and presented to subjects for identification and rating. These series ranged perceptually from [ba] to [ga] and from [bu] to [du]. The [ba]-[ga] series varied in the direction and extent of the second formant transition. The same rising third formant

Figure 24. Predicted (solid circles) and obtained (open circles) rating functions for the [bae]-[dae] series and the [ʌb]-[ʌd] series from Experiment 5.

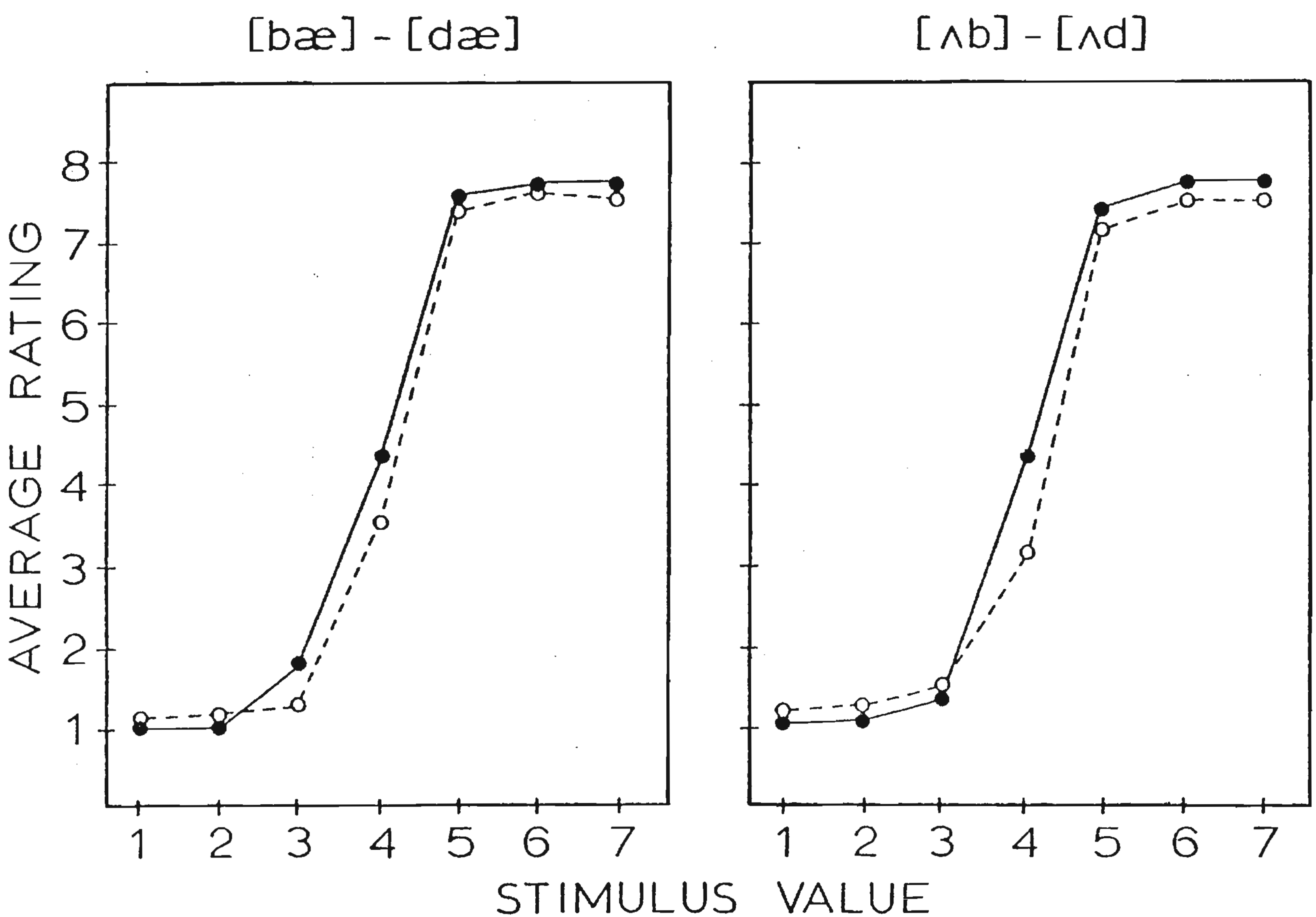


Figure 24.

transition was used for all of these stimuli. The [bu]-[du] series varied in the direction and extent of both second and third formant transitions. Table 7 contains the formant transition starting frequencies for both the [ba]-[ga] series and the [bu]-[du] series. The predicted rating functions for these two series are shown in Figure 25. The average rating responses for subjects to these stimuli are also shown in Figure 25. The pattern of fit between observed and predicted rating functions is roughly the same for both series. The fit is extremely good at the end-points of the series. However, in both cases, the simulation and subjects placed their category boundaries at different points within the two sets of stimuli.[11]

In summary, for the six synthetic CV series for which predicted and obtained rating functions were compared, the fit of the predicted to the obtained data was excellent in three cases. For two series the category boundary was displaced and for one series, the [bi]-[di] series, the obtained rating function was quite different from predicted. However, these discrepancies between predicted and obtained data could probably be resolved with one modification to the simulation. That modification would be to provide a set of weights for the abstract auditory level to use in summing the rise, fall, convergence, and divergence of formants from the peripheral level. At present, every peripheral auditory detector output is given a weight of 1 in this summation process. If, instead, a weight proportional to the frequency position of the peripheral level detector was employed, the predicted functions might better match the obtained data for those CV syllables with either extremely high frequency second and third formant transitions

Table 7

Starting frequencies in Hz of the second and third formant transitions for the [ba]-[ga] and [bu]-[du] syllable series.

	[ba]-[ga][a]		[bu]-[du][b]	
Stimulus	F2	F3	F2	F3
1	1007	2074	616	1554
2	1164	2074	733	1695
3	1307	2074	847	1902
4	1467	2074	979	2074
5	1599	2074	1099	2262
6	1744	2074	1233	2467
7	1902	2074	1384	2690
8	1958	2074	1510	2850

[a] Formant frequencies for the steady-state vowel [a] were: F1 = 733 Hz, F2 = 1099 Hz, and F3 = 2614 Hz. The first formant transition had a starting frequency of 287 Hz.

[b] Formant frequencies for the steady-state vowel [u] were F1 = 299 Hz, F2 = 872 Hz, and F3 = 2262 Hz. The first formant transition had a starting frequency of 200 Hz.

Figure 25. Predicted (solid circles) and obtained (open circles) unadapted rating functions for a [ba]-[ga] series and a [bu]-[du] series.

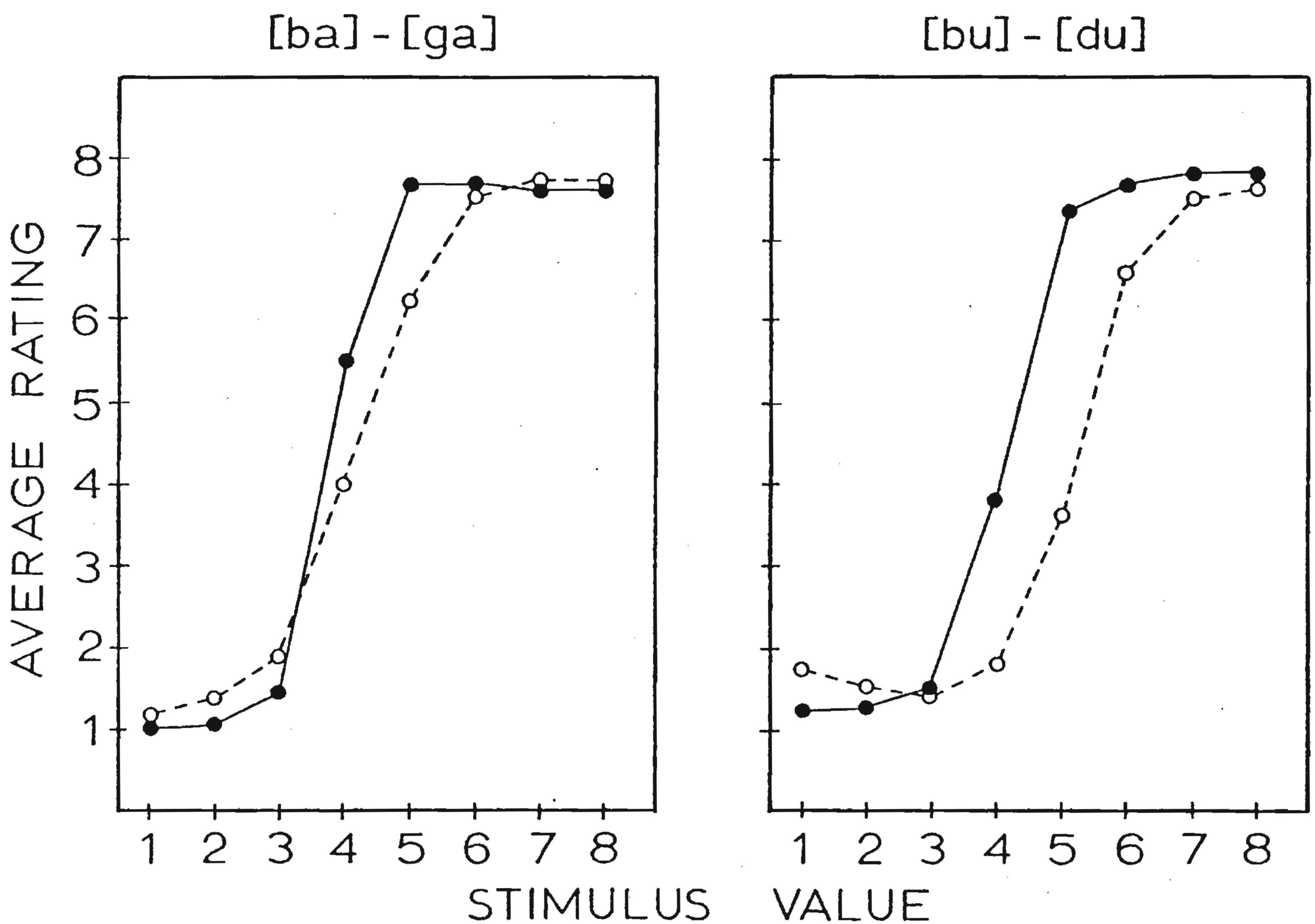

[ba] - [ga]
[bu] - [du]
AVERAGE RATING
8
7
6
5
4
3
2
1
1 2 3 4 5 6 7 8
1 2 3 4 5 6 7 8
STIMULUS VALUE

Figure 25.

(e.g., [bi], [di]) or low frequency transitions (e.g., [bu], [du], [ba], [ga]). For the high second and third formant frequency region, falling transitions would be given a greater weight than rising transitions. For the low second and third formant frequency region, the rising transitions would be given a greater weight. This change in the weights given to various peripheral level outputs at the integrative level would move the predicted category boundaries for the [bu]-[du], [bi]-[di] and [ba]-[ga] series closer to the obtained data.

Effects of selective adaptation. On the basis of Experiments 1 through 4, we proposed that selective adaptation has an effect on both the peripheral and the abstract auditory levels. As outlined earlier in Chapter III, it seems reasonable to suppose that selective adaptation has qualitatively different effects on each of these two auditory levels. On the one hand, there is evidence for a fatigue effect due to adaptation. The within phoneme category adaptation effects found in Experiment 2 and previously by Miller (1975) and Sawusch (1976) provide strong support for the fatigue notion. However, the within category adaptation effects have generally been found only for adapting syllables that overlap spectrally with the test series. Since this overlap is a necessary condition for adaptation at the Peripheral Auditory Level, as specified in the current model, it seems reasonable to conclude that the effect of adaptation on peripheral level detectors is one of fatigue. This would be reflected in a reduced response by these detectors over their entire range. This type of fatigue was incorporated into the simulation as a proportional decrease in the output response of peripheral level detectors. The particular proportion was left as a free parameter, to be entered by the user. In the unadapted state every detector operated with this proportion set to 1.0, indicating 100% of base line response. Following adaptation the simulation would apply the fatigue proportion entered by the user (ranging from 0.0 indicating complete adaptation to 1.0 indicating no adaptation) to those detectors that responded to the adapting stimulus. Thus, all further responding by these detectors would show a reduction in their output strength. This adapted state was maintained until a new adaptor was entered or the simulation was terminated.

The effect of selective adaptation at the abstract auditory level, however, seems to be more a matter of retuning than fatigue. The absence of any within category adaptation effect with the high series adaptors in Experiments 1, 2, and 3 may be used as support for this interpretation. These results indicate that adaptation affected only the boundary area between the general rise and general fall detectors and left the stimuli within the phonetic categories relatively unaffected. Within the simulation, this retuning at the abstract auditory level was programmed as a reweighting of the various peripheral level outputs during the operation of the general rise, fall, convergence and divergence detectors. The outputs of the shallow rise and shallow fall detectors, which were greatest for the boundary stimuli, were given a reduced weight by the general detectors following adaptation. As an example, if a [bae] was entered as an adaptor, then the weights associated with the shallow rise detectors from the peripheral level would be reduced, moving the category boundary for a series of synthetic [bae]-[dae] syllables toward [bae]. As with the peripheral level, the particular proportion of adaptation at the abstract level was a free parameter to be entered by the user. This retuning parameter was 1.0 in the unadapted state, indicating full weight was given to the appropriate shallow transition detectors. Following adaptation, the new weight (between 0.0 and 1.0) was maintained until a new adaptor was entered or the simulation was terminated.

Using the two adaptation parameters outlined above, an attempt was made to fit the adaptation data from Experiments 2, 5, and 6. The adaptation data for the [bi]-[di] series from Experiment 4 was not used because of the extremely poor fit of the simulation to the unadapted rating data for this series. For purposes of comparison, the monaural adaptation data from Experiment 2 was averaged with the binaural adaptation data from Experiment 1 for each of the eight adaptors. The average category boundary shifts that were obtained appear in Table 8. The simulation results for each of the eight adaptors are also shown in Table 8 as predicted category boundary shifts. The predicted shifts for the various [bae] adaptors are generally a little low while the predicted shifts for the various [dae] adaptors are generally a bit high. However, the same set of adaptation parameters was used for all eight adapting stimuli in the simulation. The peripheral auditory level output was set at 35% of baseline (0.35) and the abstract level had a retuning proportion of 0.55. No attempt was made to vary these parameters systematically to achieve a better fit for each of the individual adapting stimuli. If this had been done, the predicted boundary shifts for each individual adaptor might have been closer to the obtained shifts. However, given that only one set of parameters was used in the simulation

Table 8

Obtained category boundary shifts from Experiments 1 and 2 and the predicted shifts from the simulation (in stimulus units)

	Adaptor							
	[bae]-low	[bae]-low chirp	[bae]-high	[bae]-high chirp	[dae]-low	[dae]-low chirp	[dae]-high	[dae]-high chirp
Obtained Shift[a]	1.13	.90	.56	.39	-1.52	-1.49	-.44	-.58
Obtained Shift[b]	1.65	1.71	.82	.65	-1.42	-.82	-.41	-.29
Average	1.39	1.31	.69	.52	-1.47	-1.16	-.43	-.43
Predicted Shift[c]	1.17	1.17	.52	.52	-1.25	-1.25	-.55	-.55

[a] Obtained shifts for the eight adaptors from Experiment 1.

[b] Obtained monaural adaptation shifts from Experiment 2.

[c] Based on adaptation parameters of 0.35 at the peripheral level and 0.55 at the abstract level. See text for an explanation of these parameters.

for all eight of the adaptors, the predicted shifts are in very good agreement with the obtained data. The fit of the model to the data is even better when the variability of the experimental data is taken into account. The binaural adaptation data from Experiment 1 and the monaural data from Experiment 2 are also shown separately in Table 8 for comparison. For six of the eight adaptors, the predicted shift falls between the two obtained shifts. Consequently, the overall fit of the model to the adaptation data from Experiments 1 and 2 appears to be quite good and within the limits of measurement for the category boundary shifts.

The predicted and obtained category boundary shifts for the various adaptors from Experiments 5 and 6 appear in Tables 9 and 10 respectively. For the predicted boundary shifts, the peripheral level adaptation parameter was 35% (0.35) and the abstract level parameter was 0.55. These values are the same as those previously used in fitting the data from Experiments 1 and 2. The differences between observed and predicted shifts for all eight of the adaptors in Tables 9 and 10 were less than 0.1 stimulus units, indicating that the fit between the model and the data was very good indeed.

One further point about the fit between the adaptation data and the predictions from the model should be mentioned. The adapted rating functions for the [bae]-[dae] series in Experiment 2 showed a within category shift in the rating functions for the low adaptors presented monaurally (see Chapter III). These within category shifts occurred for the end of the test series from which the low adaptor was drawn. The predicted rating functions from the computer simulation exhibited this same within category rating change. As with the subject data from

Table 9

Obtained shifts in category boundaries for the rating data from Experiment 5 and predicted shifts from the simulation (in stimulus units)

	[bae]-[dae] test		[ʌb]-[ʌd] test	
	Adaptor			
	[bae]	[ʌd]	[bae]	[ʌd]
Predicted Shift[a]	0.59	0.0	0.0	-0.52
Obtained Shift	0.60	-0.09	0.04	-0.46

[a] The predicted shifts were based on adaptation parameters of 0.35 for the peripheral level and 0.55 for the abstract level. See text for an explanation of these parameters.

Table 10

Obtained shifts in category boundaries for the rating data from Experiment 6 and predicted shifts from the simulation (in stimulus units)

	Adaptor			
	b-SEChirp	F2F3-SEChirp	[ʌ]	V'
Predicted Shift[a]	0.40	0.40	0.0	0.0
Obtained Shift	0.49	0.41	0.02	-0.07

[a] The predicted shifts were based on adaptation parameters of 0.35 for the peripheral level and 0.55 for the abstract level. See text for an explanation of these parameters.

Experiment 2, the simulation did not show any within category rating shift for the high adaptors.

In contrast to the within category shifts found in Experiment 2, no such within category shifts were found in Experiments 5 or 6, even when the adaptor was an end-point of the test series. Within the simulation, no within category change in rating functions for either the [bae]-[dae] series or the [Λb]-[Λd] series was predicted. The computer simulation did not predict a within category shift because only one response category was receiving output for the [bae] end of the [bae]-[dae] test series (and the [Λd] end of the [Λb]-[Λd] test series). Consequently, although the [bae] end of the series received a reduced output from the peripheral level detectors following [bae] adaptation, no alternative coding of the stimulus was available and the ratings remained low (1.0). The peripheral fall and integrative fall detectors in the simulation simply did not respond to any stimuli within the [bae] end of the [bae]-[dae] series (or the [Λd] end of the [Λb]-[Λd] series). In summary, the simulation fit the major qualitative aspects of the adaptation data as well as quantitatively predicting the category boundary shifts with reasonable accuracy.

<u>Concluding Comments on the Model and Simulation</u>

The overall fit between the data and the predictions from the simulation was very encouraging. There were a few weak spots in the simulation, notably the large error in predicting the identification rating function for the [bi]-[di] series from Experiment 4. However, this problem does not appear to be inherent in the model. Rather, this mismatch between the data and the simulation appears to be due to certain simplifying assumptions that had to be made in order to implement the simulation. Specifically, the lack of a differential weighting of peripheral level outputs during processing at the abstract auditory level appears to be the main reason for the failure of the simulation on the [bi]-[di] series. Future work using the present model should focus on improving the decision rules at the abstract auditory level.

The model and its accompanying simulation seem to have done a good overall job in dealing with the present data. The implementation of the effects of selective adaptation as fatigue at the peripheral auditory level and retuning at the abstract auditory level was successful. The simulation was able to quantitatively predict the adaptation boundary shifts from Experiments 1, 2, 5 and 6 and recover the major qualitative aspects of the adaptation data. However, further work is needed, especially on two areas. Definite improvement is needed in the decision rules at the abstract auditory level. One possible modification, suggested above, is currently being implemented. The second major area for improvement is empirical instead of theoretical. Many assumptions had to be made in order to implement this simulation. Some of them, particularly those dealing with the time window over which various operations take place, were almost completely without empirical support. The problem here is that relatively little is known about these operations. Experimental work, possibly involving the exploration of short term memory and masking in these early stages of perceptual processing is needed.

The overall success of the simulation suggests that simulation in general and the present model in particular should be extended to features other than place of articulation in stop consonants. This extension will serve the dual role of forcing specification and refinement of current models and focusing attention on the gaps in our knowledge of how people perceive speech.

Footnotes

[8] Chi-square goodness-of-fit and mean squared error were used to evaluate the fit between the observed and predicted unadapted rating functions. For the [bae]-[dae] series from Experiment 1, $\chi^2(9) = 0.96$, $p > .995$. The mean squared error (MSE) was 0.228 between observed and predicted functions.

[9] For the [bi]-[di] series, $\chi^2(7) = 50.97$, $p < .001$ and MSE = 7.89 for the fit between predicted and observed (from Experiment 4) unadapted rating functions.

[10] For the [bae]-[dae] series, $\chi^2(6) = .39$, $p > .995$ and MSE = .19 while for the [Λb]-[Λd] series $\chi^2(6) = .55$, $p > .995$ and MSE = .23 for the fits between observed (Experiment 5) and predicted unadapted rating functions.

[11] For the [ba]-[ga] series, $\chi^2(7) = 1.06$, $p > .99$ and MSE = .71 while for the [bu]-[du] series $\chi^2(7) = 3.52$, $p > .90$ and MSE = 2.63 for the fits between observed and predicted unadapted rating functions.

CHAPTER VI

Summary and Concluding Remarks

Experimental Results

The basic thrust of the experiments reported here was to examine various levels of processing that are involved in the perception of place of articulation in stop consonants. These experiments also tested and examined some of the characteristics of selective adaptation and the level(s) that it affects in the perceptual process. Experiments 1 and 2 focused on the involvement of an abstract integrative level within selective adaptation. Adapting syllables were constructed that maintained the phonetic identity and overall formant pattern of the endpoints of a place feature test series, but these adaptors had no frequency components in common with the test series. The formants of these "high" adaptors were one and one-half critical bandwidths higher in frequency than those of the test series. The high adaptors produced sizeable and consistent category boundary shifts in the test series. However, these shifts were always smaller than those produced by the low adaptors.

These findings were refined somewhat in Experiment 2 where the interaural transfer of the adaptation effects was studied. Using the low adaptors, about twice as much adaptation was found when the adaptor and test were presented to the same ear as compared to when the adaptor and test syllables were presented to opposite ears. When the high frequency stimuli were used as adaptors, the adaptation effect was the same regardless of whether the adaptor and test were presented to the same or opposite ears. The 100% interaural transfer effects found for the high adaptors indicates that they affected a central level of processing. Further, because the high adaptors and low test series had no frequency components in common, the central level could be characterized as abstract in nature and not frequency specific. The 50% transfer for the low adaptors indicates that two levels of processing were involved with these stimuli. In addition to adapting the integrative level outlined above, an earlier, frequency specific, peripheral level was also affected. In particular, the frequency specificity of this level appears to be related to the notion of critical bandwidth.

The purpose of Experiment 3 was to explore whether the integrative level was sensitive to low-level acoustic variation in the speech signal. High and low syllable adaptors were presented at an intensity either 8 dB above the test syllables or 8 dB below the test syllables. It was found that the more intense adaptors produced larger category boundary shifts than the less intense adaptors. The increased adaptation for the high frequency adaptors was approximately equal to that found for the low adaptors, indicating that the effects of the intensity manipulation were at the integrative rather than peripheral level. Thus, the abstract integrative level is indeed sensitive to the relative intensity of adaptors and test. On the basis of these results, the abstract level was characterized as auditory rather than phonetic.

Experiment 4 was designed to resolve the conflicting results of Experiments 1, 2 and 3 and previous experiments that had failed to find adaptation when the adaptor and test syllables had no frequency components in common (Bailey, 1975). The results of Experiment 4 were entirely consistent with Experiments 1 through 3 and the absence of an adaptation effect previously reported by Bailey (1975) was attributed to the particular testing procedure that was used in his experiments.

Experiments 5 and 6 focused on the adapting effects of a CV on a VC series and vice versa. The spectral overlap between one end of a CV series and one end of a VC series was maximized. When these CV and VC syllables with maximal spectral overlap were used in a cross-series adaptation condition (CV adaptor with VC test and vice versa), no adaptation was found. On the basis of these results a new set of non-speech adaptors was generated that differed from the VC adaptor only in the first formant. These new VC like adaptors, labeled Speech Embedded Chirps (SEChirp), did produce a category boundary shift in the CV test series. The adaptation result of these SEChirps was attributed to the frequency specific auditory level because of the maximal spectral overlap between adaptors and test syllables in their second and third formant transitions. The adapting effect of the SEChirps suggests two conclusions. First, peripheral level detectors are not sensitive to whether a transition occurs at syllable onset or offset; these detectors are on "free-run." Secondly, the first formant was implicated as being involved in the operations of the integrative level of processing since only the first formant was changed in generating the SEChirps from the VC adaptor.

One final result from Experiment 6 was that when vowels with steady-state frequencies identical to the starting transition frequencies for the CV test series were used as adaptors, no adaptation was found. This result indicates that spectral change plays an important role in selective adaptation to place in consonants and that the onset or offset spectra of adaptor and test series are not by themselves sufficient for selective adaptation effects to be found.

Model and Simulation

On the basis of the results just outlined, an information processing model of place perception was constructed. This model was programmed as a computer simulation in order to formalize the model and make precise quantitative predictions. The model employed three levels of analysis. The first two levels corresponded to the frequency specific peripheral auditory level and the abstract integrative auditory level as outlined above. The third level involved a phonetic feature combination stage which transformed the outputs of the two auditory levels into a rough phonetic feature matrix. The input to this computer simulation was a spectral coding of a stimulus in terms of intensity per critical band over time. The peripheral auditory level operated with frequency specific formant detectors (rise, fall and steady-state) that operated on 25 msec segments of the input. The outputs from these detectors went through general formant pattern detectors at the abstract level. These general detectors, roughly characterized by rising, falling, converging and diverging formant patterns, are similar to the auditory property detectors proposed recently by Stevens (1973). The outputs from these abstract detectors were converted, in conjunction with onset-offset and first formant information, into a phonetic feature matrix at the phonetic feature combination stage.

The effect of adaptation within this model was twofold. At the peripheral level, adaptation operated to fatigue a detector and reduce its output to some percentage of its unadapted state. At the abstract level, adaptation caused a reweighting or retuning of the peripheral level outputs as they were assembled by the general auditory detectors. This particular implementation of adaptation was based on the differential effectiveness of the various adaptors in changing the perception of the good exemplars of a phonetic category. When the adaptor and test series shared frequency components (spectral overlap) the adaptor often caused a decrease in the rating response for the end of the test series from which the adaptor was drawn. However, when the adaptor and test series did not contain any spectral overlap, the adaptation effect was entirely confined to stimuli at the category boundary.

The simulation was evaluated by comparing the predicted identification of various sets of synthetic stop consonant syllables to subject data both before and after adaptation. In general, the simulation predicted the identification data very well with the exception of one CV series ([bi]-[di]). Both the quantitative and qualitative fit between the adaptation data and the predictions from the simulation were excellent. The good overall fit between the predictions from the model and the data offers strong support for the proposed model of the earliest stages of speech perception.

Future Empirical and Theoretical Directions

A number of new empirical and theoretical questions were raised by the present experiments and the model. In particular, further work is needed on the characteristics of the integrative level. A blending of the interaural transfer paradigm with spectral manipulations of CV and VC syllables may help to elucidate whether the integrative level is sensitive to the temporal position of consonantal information within the syllable. Further work is also needed on how the various cues from the peripheral auditory level interact at the integrative level.

The process of implementing the model as a computer simulation raised a number of important theoretical questions. Among these is whether the peripheral and integrative auditory levels operate in serial or parallel. Other questions concern the role of memory within this model and the time course or integration period over which the processes at these various levels operate. One final theoretical consideration is the extension of the present model to other features in consonants such as voicing and manner of articulation. Models of the perception of voicing currently exist which are compatible with the model proposed here (Summerfield, 1975). Extending the present model and simulation to other features would make precise prediction possible and extend its generality. Computer simulation would make precise specification of models for the perception of other features necessary and highlight any inadequacies in current formulations.

One final remark about the adaptation paradigm should be made here. This experimental paradigm has proven to be extremely useful in exploring the relatively early stages of speech perception. However, as our knowledge of these processes increases, the new questions that are raised will be difficult, if not impossible, to answer within a single experimental paradigm. In this respect, selective adaptation should be treated as a tool, to be used in conjunction with other paradigms such as dichotic listening, reaction time and masking. Hopefully, future research will make further use of both computer simulation and the blending of multiple experimental paradigms to advance our understanding of the complex processes involved in speech perception.

References

Abbs, J. H. & Sussman, H. M. Neurophysiological feature detectors and speech perception: A discussion of theoretical implications. Journal of Speech and Hearing Research, 1971, 14, 1, 23-36.

Ades, A. E. A bilateral component in speech perception: Journal of the Acoustical Society of America, 1974, 55, 610-616. (a)

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 61-67. (b)

Ades, A. E. Adapting the property detectors for speech perception. Manuscript in preparation, 1976.

Ades, A. E. Source effects in speech perception. (forthcoming)

Bailey, P. Perceptual adaptation for acoustical features in speech. Speech Perception: Report on Research in Progress in the Department of Psychology. Northern Ireland: The Queen's University of Belfast, 1973, 2.2, 29-34.

Bailey, P. Perceptual adaptation in speech: Some properties of detectors for acoustical cues to phonetic distinctions. Unpublished doctoral thesis, 1975, University of Cambridge, Cambridge, England.

Blakemore, C. & Campbell, F. W. On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. Journal of Physiology, 1969, 203, 237-260.

Blakemore, C. & Sutton, P. Size adaptation: A new after effect. Science, 1969, 166, 245-247.

Blumstein, S. E. & Stevens, K. N. Property detectors for bursts and transitions in speech perception. Paper presented at the 89th meeting of the Acoustical Society of America, April, 1975, Boston.

Bondarko, L. V., et al. A model of speech perception in humans. Working Papers in Linguistics No. 6, 1970, Computer & Information Science Research Center, Ohio State University, Columbus, Ohio. Technical Report 70-12.

Chomsky, N. & Halle, M. Sound Pattern of English. New York: Harper & Row, 1968.

Cole, R. A. & Cooper, W. E. Perception of voicing in English affricates and fricatives. Journal of the Acoustical Society of America, 1975, 58, 1280-1287.

Cole, R. A., Cooper, W. E., Singer, J. & Allard, F. Selective adaptation of English consonants using real speech. Perception & Psychophysics, 1975, 18, 227-244.

Cole, R. A. & Scott, B. Toward a theory of speech perception. Psychological Review, 1974, 81, 348-374.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-83.

Cooper, F. S., Delattre, P. C., Liberman, A. M., Borst, J. M. & Gerstman, L. J. Some experiments on the perception of synthetic speech sounds. Journal of the Acoustical Society of America, 1952, 24, 597-606.

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974, 56, 617-627. (a)

Cooper, W. E. Contingent feature analysis in speech perception. Perception & Psychophysics, 1974, 16, 201-204. (b)

Cooper, W. E. Selective adaptation for acoustic cues of voicing in initial stops. Journal of Phonetics, 1974, 2, 303-313. (c)

Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), Cognitive theory: Vol. I. Hillsdale, N.J.: Erlbaum Associates, 1975.

Cooper, W. E. & Blumstein, S. A "labial" feature analyzer in speech perception. Perception & Psychophysics, 1974, 15, 591-600.

Cooper, W. E., Ebert, R. R. & Cole, R. A. Perceptual analysis of stop consonants and glides. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 92-104.

Craik, F. I. M. & Lockhard, R. S. Levels of processing: A framework for memory research. Journal of Verbal Learning and Verbal Behavior, 1972, 11, 671-684.

Crowder, R. G. The sound of vowels and consonants in immediate memory. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 587-590.

Cutting, J. E. Aspects of phonological fusion. Journal of Experimental Psychology: Human Perception and Performance, 1975, 1, 105-120.

Cutting, J. E. Auditory and linguistic processes in speech perception: Inferences from six fusions in dichotic listening. Psychological Review, 1976, 83, 114-140.

Cutting, J. E. & Day, R. S. The perception of stop-liquid clusters in phonological fusion. Journal of Phonetics, 1975, 3, 99-113.

Cutting, J. E. & Rosner, B. S. Categories and boundaries in speech and music. Perception & Psychophysics, 1974, 16, 564-570.

Cutting, J. E., Rosner, B. S. & Foard, C. F. Rise time in nonlinguistic sounds and models of speech perception. Status Report on Speech Research, 1975, SR-41, 71-94.

Darwin, C. J. The perception of speech. In E. C. Carterette and M. P. Friedman (Eds.), Handbook of perception: Vol. VII. New York: Academic Press, 1976.

Day, R. S. Fusion in dichotic listening. Unpublished doctoral dissertation, Stanford University, 1968.

Day, R. S. Temporal-order perception of a reversible phoneme cluster. Journal of the Acoustical Society of America, 1970, 48, 95. (Abstract)

Delattre, P. C., Liberman, A. M. & Cooper, F. S. Acoustic loci and transitional cues for consonants. Journal of the Acoustical Society of America, 1955, 27, 769-773.

Diehl, R. L. The effect of selective adaptation on the identification of speech sounds. Perception & Psychophysics, 1975, 17, 48-52.

Diehl, R. Feature analyzers for the phonetic dimension stop vs. continuant. Perception & Psychophysics, 1976, 19, 267-272.

Durlach, N. I. & Braida, L. D. Intensity perception I. Preliminary theory of intensity resolution. Journal of the Acoustical Society of America, 1969, 46, 372-383.

Eimas, P. D., Cooper, W. E. & Corbit, J. D. Some properties of linguistic feature detectors. Perception & Psychophysics, 1973, 13, 247-252.

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

Fant, G. Acoustic theory of speech production. The Hague, Netherlands: Mouton & Co., 1960.

Fant. G. Descriptive analysis of the acoustic aspects of speech. Logos, 1962, 5, 3-17.

Fujisaki, H. & Kawashima, T. Some experiments on speech perception and a model for the perceptual mechanism. Annual Report of the Engineering Research Institute, Faculty of Engineering, University of Tokyo, 1970, 29, 207-214.

Ganong, W. F. An experiment on "phonetic adaptation." RLE Progress Report, 1975, 116, 206-210.

Ganong, W. F. An acoustic component of phonetic adaptation (forthcoming).

Greenwood, D. D. Auditory masking and the critical band. Journal of the Acoustical Society of America, 1961, 33, 484-502.

Harris, D. S. Cues for the discrimination of American English fricatives in spoken syllables. Language and Speech, 1958, 1, 1-7.

Helson, H. Adaptation level theory. New York: Harper & Row, 1964.

Hillenbrand, J. M. Intensity and repetition effects on selective adaptation to speech. Research on Speech Perception Progress Report, Indiana University, Bloomington, Indiana, 1975, 2, 56-137.

Hoffman, H. S. Study of some cues in the perception of the voiced stop consonants. Journal of the Acoustical Society of America, 1958, 30, 1035-1041.

Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurophysiology, 1965, 28, 229-289.

Hubel, D. H. & Wiesel, T. N. Receptive fields and functional architecture of monkey striate cortex. Journal of Physiology, 1968, 195, 215-243.

International Phonetic Association. Principles of the International Phonetic Association. Revised edition. London: International Phonetic Association, 1949.

Jakobson, R., Fant, C. G. M. & Halle, M. Preliminaries to speech analysis. Cambridge, Mass.: MIT Press, 1963.

Kay, R. H. & Mathews, D. R. On the existence in human auditory pathways of channels selectively tuned to the modulation present in frequency-modulated tones. Journal of Physiology, 1972, 225, 657-677.

Kimura, D. Functional assymmetry of the brain in dichotic listening. Cortex, 1967, 3, 163-178.

Koenig, W., Dunn, H. K. & Lacy, L. Y. The sound spectrograph. Journal of the Acoustical Society of America, 1946, 17, 19-49.

Liberman, A. M., Cooper, F. S., Harris, K. S. & MacNeilage, P. F. A motor theory of speech perception. Proceedings of the Speech Communication Seminar, Stockholm, 1962. (Stockholm: Royal Institute of Technology, 1963, D3.)

Liberman, A. M., Cooper, F. S., Harris, K. S., MacNeilage, P. F. & Studdert-Kennedy, M. Some observations on a model for speech perception. In W. Wathen-Dunn (Ed.), Models for the perception of speech and visual form. Cambridge, Mass.: MIT Press, 1967.

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

Liberman, A. M., Delattre, P. C. & Cooper, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. American Journal of Psychology, 1952, 65, 497-516.

Liberman, A. M., Delattre, P. C. & Cooper, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. Language and Speech, 1958, 1, 153-167.

Liberman, A. M., Delattre, P. C., Cooper, F. S. & Gerstman, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. Psychological Monographs, 1954, 68, 1-13.

Liberman, A. M., Delattre, P. C., Gerstman, L. J. & Cooper, F. S. Tempo of frequency change as a cue for distinguishing classes of speech sounds. Journal of Experimental Psychology, 1956, 52, 127-137.

Licklider, J. C. R. Three auditory theories. In S. Koch (Ed.), Psychology: A study of a science. New York: McGraw-Hill, 1959.

Liljencrants, J. C. W. A. The OVE III speech synthesizer. IEEE Transactions on Audio and Electroacoustics, 1968, AU-16, 137-140.

Lisker, L. Is it VOT or a first formant transition detector. Journal of the Acoustical Society of America, 1975, 57, 1547-1551.

Lisker, L. & Abramson, A. S. A cross language study of voicing in initial stops: Acoustical measurements. Word, 1964, 20, 384-422.

MacKay, D. M. The epistemological problem for automata. In C. E. Shannon & J. McCarthy (Eds.), Automata studies. Princeton, N.J.: Princeton University Press, 1956.

MacNeilage, P. F. Motor control of serial ordering of speech. Psychological Review, 1970, 77, 182-196.

Massaro, D. W. Preperceptual images, processing time and perceptual units in auditory perception. Psychological Review, 1972, 79, 124-145.

Miller, G. A. & Nicely, P. E. An analysis of perceptual confusions among some English consonants. Journal of the Acoustical Society of America, 1955, 27, 338-352.

Miller, J. L. Properties of feature detectors for speech: Evidence from the effects of selective adaptation on dichotic listening. Perception & Psychophysics, 1975, 18, 389-397.

Miller, J. L. & Eimas, P. D. Studies on the selective-tuning of feature detectors for speech. Developmental Studies of Speech Perception, 1975, 3, 103-116. (Progress report of the Walter S. Hunter Laboratory of Psychology, Brown University.)

Öhman, S. E. G. Coarticulation in VCV utterances: Spectrographic measurements. Journal of the Acoustical Society of America, 1966, 39, 151-168.

Parducci, A. Range-frequency compromise in judgement. Psychological Monographs, 1963, 77, 1-50.

Parducci, A. Contextual effects: A range-frequency analysis. In E. C. Carterette & M. P. Friedman (Eds.), Handbook of perception, Vol. II. New York: Academic Press, 1975.

Pisoni, D. B. On the nature of categorical perception of speech sounds. Supplement to: Status Report on Speech Research, Haskins Laboratories, November, 1971.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 1973, 13, 253-260.

Pisoni, D. B. Dichotic listening and processing phonetic features. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman & D. B. Pisoni (Eds.), Cognitive Theory: Volume I. Hillsdale, N. J.: Erlbaum Associates, 1975. (a)

Pisoni, D. B. Stages of processing in speech perception: Feature analysis. Paper presented at the Eighth International Congress of Phonetic Sciences, August, 1975, Leeds, England. (b)

Pisoni, D. B. Speech perception. In W. K. Estes (Ed.), Handbook of learning and cognitive processes. Hillsdale, N. J.: Lawrence Erlbaum Associates, 1976 (in press).

Pisoni, D. B. & Lazarus, J. H. Categorical and noncategorical modes of perception along the voicing continuum. Journal of the Acoustical Society of America, 1974, 55, 328-333.

Pisoni, D. B. & McNabb, S. D. Dichotic interactions and phonetic feature processing. Brain & Language, 1974, 1, 351-362.

Pisoni, D. B. & Sawusch, J. R. Some stages of processing in speech perception. In A. Cohen and S. G. Nooteboom (Eds.), Structure and process in speech perception. New York: Springer-Verlag, 1975.

Pisoni, D. B. & Sawusch, J. R. Simple and contingent adaptation effects on the place feature in stop consonants. Manuscript in preparation, 1976.

Pisoni, D. B., Sawusch, J. R. & Adams, F. T. Simple and contingent adaptation effects in speech perception. Paper presented at the 90th meeting of the Acoustical Society of America, San Francisco, November, 1975.

Pisoni, D. B. & Tash, J. B. Auditory property detectors and processing place features in stop consonants. Perception & Psychophysics, 1975, 18, 401-408.

Savin, H. B. & Bever, T. G. The nonperceptual reality of the phoneme. Journal of Verbal Learning and Verbal Behavior, 1970, 9, 295-302.

Sawusch, J. R. Selective adaptation effects on end-point stimuli in a speech series. Perception & Psychophysics, 1976, 20, 61-65.

Sawusch, J. R. & Pisoni, D. B. Category boundaries for speech and non-speech sounds. Paper presented at the 86th meeting of the Acoustical Society of America, November, 1973, Los Angeles.

Sawusch, J. R. & Pisoni, D. B. Response organization and selective adaptation to speech sounds. Submitted to Perception & Psychophysics, 1976.

Sawusch, J. R., Pisoni, D. B. & Cutting, J. E. Category boundaries for linguistic and nonlinguistic dimensions of the same stimuli. Paper presented at the 87th meeting of the Acoustical Society of America, April, 1974, New York.

Scharf, B. Critical bands. In J. V. Tobias (Ed.), Foundations of modern auditory theory. New York: Academic Press, 1970.

Shepard, R. N. Psychological representation of speech sounds. In E. E. David & P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill, 1972.

Shiffrin, R. M. Short-term store: The basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), Cognitive theory: Volume I. Hillsdale, N. J.: Erlbaum Associates, 1975.

Stevens, K. N. Toward a model for speech recognition. Journal of the Acoustical Society of America, 1970, 32, 47-55.

Stevens, K. N. The potential role of property detectors in the perception of consonants. Paper presented at the Symposium on Auditory Analysis and Perception of Speech, Leningrad, USSR, August, 1973.

Stevens, K. N. & Blumstein, S. E. Context-independent cues for place of articulation in stop consonants. Paper presented at the 91st meeting of the Acoustical Society of America, April, 1976, Washington, D.C.

Stevens, K. N. & Halle, M. Remarks on analysis by synthesis and distinctive features. In W. Wathen-Dunn (Ed.), Models for the perception of speech and visual form. Cambridge, Mass.: MIT Press, 1967.

Stevens, K. N. & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 1974, 55, 654-659.

Strevens, P. Spectra of fricative noise in human speech. Language and Speech, 1960, 3, 32-47.

Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), Current trends in linguistics, Volume 12. The Hague, Netherlands: Mouton, 1974.

Studdert-Kennedy, M. The nature and function of phonetic categories. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman & D. B. Pisoni (Eds.), Cognitive theory, Vol. I. Hillsdale, N.J.: Erlbaum Associates, 1975.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), Contemporary issues in experimental phonetics. New York: Academic Press, 1976.

Studdert-Kennedy, M. & Shankweiler, D. Hemispheric specialization for speech perception. Journal of the Acoustical Society of America, 1970, 48, 579-594.

Summerfield, Q. Towards a detailed model for the perception of voicing contrasts. Speech Perception, 1974, 3, 1-26. (Progress report, Psychology Department, Queen's University of Belfast, Northern Ireland).

Summerfield, Q. Cues, contexts and complications in the perception of voicing contrasts. Speech Perception, 1975, 4, 99-129. (Progress report, Psychology Department, Queen's University of Belfast, Northern Ireland).

Summerfield, Q. & Haggard, M. P. Perceptual processing of multiple cues and contexts: Effects of following vowel upon stop consonant voicing. Journal of Phonetics, 1974, 2, 279-295.

Tartter, V. C. & Eimas, P. D. The role of auditory and phonetic feature detectors in the perception of speech. Perception & Psychophysics, 1975, 18, 293-298.

Weisstein, N. Beyond the yellow-Volkswagen detector and the grandmother cell: A general strategy for the exploration of operations in human pattern recognition. In R. L. Solso (Ed.), Contemporary issues in cognitive psychology. Washington, D. C.: V. H. Winston, 1973.

Whitfield, I. C. The auditory pathway. London: Arnold, 1967.

Whitfield, I. C. & Evans, E. F. Responses of auditory cortical neurones to stimuli of changing frequency. Journal of Neurophysiology, 1965, 28, 655-672.

Zwicker, E., Flottorp, G. & Stevens, S. S. Critical band width in loudness summation. Journal of the Acoustical Society of America, 1957, 29, 548-557.

Appendix A

A Description of the OVE IIId Control Program: OVEXEC

The OVE IIId serial speech synthesizer control program (OVEXEC) is organized as a main program and subroutines. The main program accepts commands, and calls the appropriate subroutine(s) to execute those commands. A description of the commands and their usage is contained in the OVEXEC User's Guide.

The OVEXEC program executes in an overlay environment. The main program and the TALK subroutine are in the root segment and they reside in core memory while the program is being run. The commands are divided into five subroutines, organized according to the type of operation they perform. The organization is as follows:

I. Cassette tape input/output and directory listing commands.

II. Commands to manipulate entire lines of synthesis parameters. (A line consists of all 15 parameter values needed to specify the synthesizer output at one instant in time.)

III. Commands to manipulate values of a single parameter, such as replacing, repeating or interpolating.

IV. Commands to input and output files of parameter values from the disc. These commands include input, output, deletion and renaming of disc files of parameter values.

V. Synthesis commands. These commands control the actual synthesis of an utterance. These commands include repeated and single synthesis of a set of parameters and varying the rate of synthesis.

When a command is entered by the user, the subroutine that will execute the command is read into the overlay section of core memory from the disc. If this subroutine is already in core, a new read from the disc is not required. In order to minimize the continual reading of new subroutines into the overlay section of memory, the commands were grouped into the sections described above. Similar commands, that are often used together, were built into the same subroutine. The overlay structure is used to minimize the amount of core memory that is necessary to run OVEXEC. Using overlays, about 14K (14000 words) of core memory is necessary. Without overlays, about 24K of core memory would be necessary.

File structure of OVE parameter files.

Access to OVE parameter files is accomplished through a directory. The directory is a disc file which contains the size of a parameter file (if it exists) and a mnemonic label (1-4 characters) which is used as an identifier for the contents of the file. The directory has a capacity for listing up to 500 parameter files.

The cassette tape input-output is provided for making a copy of the entire OVE directory and its parameter files. In this manner, separate libraries of stimuli can be kept. If a particular library is required, the directory and parameter files can be loaded from cassette onto the disc.

Synthesis capabilities.

Parameter values are output to the synthesizer through an analog interface. In order to synthesize a stimulus, the parameters must be in core memory (either created there or read from the disc). As mentioned previously, the utterance may be synthesized either once or repeatedly. If the synthesis is being done repeatedly, the rate of synthesis (time between output of consecutive lines of parameters) may be changed from one repetition of the utterance to the next. The maximum duration of an utterance is approximately 2.5 seconds when the time between lines is 5 milliseconds.

There is also a provision in the program for synthesizing a series of different stimuli. Any set of 24 or fewer stimuli may be synthesized with about 1 second between stimuli. This feature allows the user to make comparisons among a set of stimuli (such as a set of vowels or CV syllables).

Parameter manipulation and visual display.

As mentioned previously, parameter commands fall into two categories: those that manipulate entire lines and those that treat the parameters individually. The commands that manipulate entire lines include the repetition and deletion of lines and zeroing the parameters of a line. Two commands are provided for treating parameters individually. One of these allows the user to enter the parameter's values for a series of lines directly. The second is an interpolation command. The user specifies initial and final values, and the form of interpolation (linear or logarithmic). The program calculates the new interpolated parameter values.

Parameter display is accomplished with one of two commands. The first displays 20 lines of parameters at a time on the CRT display. The user may request more lines or terminate the display at any time. The second command is intended for hard copy. The entire set of parameters is dumped, line by line, onto a lineprinter. The mnemonic label associated with the utterance (from the directory) is printed as a header.

The input for parameter values and the output via the display commands may be done in one of two modes: code or Hz/dB. In the code mode, input/output is done using the octal OVE parameter codes (see the OVE IIId manual). This permits exact specification by the user of the parameter values. However, since users are more likely to be familiar with intensity and frequency scales, it is often more convenient to enter parameter values as hertz and/or decibels. The program automatically converts these Hz/dB entries into the nearest octal code value available. In the Hz/dB mode, it is an easy matter to take published values of stimuli and implement them for the OVE.

<u>Summary</u>

OVEXEC provides versatile and easy to use control of the OVE IIId speech synthesizer. Using the options such as Hz/dB input/output, interpolation and the synthesis of an entire range of stimuli, even beginning users of the program can manipulate the OVE's output to suit their needs.

Appendix B

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:07:53                PAGE 001

           C      MAIN PROGRAM FOR SPEECH RECOGNITION SIMULATION
           C      FIRST SET UP ARRAYS IN COMMON FOR VARIOUS PARAMETERS
           C      INPUT PARAMETERS IN BLKI
0001              COMMON /BLKI/ IFIRST,ISR,IDRTN,NSMPL,ISPCH(36,70)
           C      RESULTS OF PERIPHERAL AUDITORY ANALYAIS IN BLKA
0002              COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
                 $ ISS(35,35),NOFSET(35)
           C      RESULTS OF ABSTRACT AUDITORY ANALYSIS IN BLKP
0003              COMMON /BLKP/ IIRZ(35),IIFL(35),IGDVG(35),ILF1(35),IMF1(35),
                 $ IHF1(35),ILF2(35),IMF2(35),IHF2(35),IGCNV(35),IFRZ(35),IFFL(35)
           C      RESULTS OF PHONETIC FEATURE COMBINATION IN BLKD
0004              COMMON /BLKD/ IPFM(3,25)
           C      ARRAYS HOLDING ADAPTATION PARAMETERS IN BLKF
0005              COMMON /BLKF/ FTGP(4,35),FTGA(6)
           C      START UP
           C      INPUT BASIC INSTRUCTIONS & INITIALIZE MEMORY
0006              CALL SADP
           C      IF CURRENT INPUT IS TO BE AN ADAPTOR, TYPE A 1
0007          200 WRITE (5,90)
0008              READ (5,100) IADAPT
0009              IF (IADAPT.EQ.1) CALL SADP
           C      INPUT PARAMETER THAT INDICATES TYPE OF OUTPUT
0011              WRITE (5,98)
0012              READ (5,100) IO
           C      INPUT TIME FRAME AT WHICH INPUT IS TO START (>1)
0013              WRITE (5,99)
0014              READ (5,100) IFIRST
0015              IF (IFIRST.LE.0) CALL EXIT
           C      ZERO ARRAYS AND STORAGE VARIABLES
0017              DO 300 J = 1, 70
0018                    DO 300 I = 1, 36
0019                    ISPCH(I,J) = 0
0020          300       CONTINUE
0021              DO 1000 J = 1, 35
0022                    DO 400 I = 1, 35
0023                    IRZ(I,J) = 0
0024                    IFL(I,J) = 0
0025                    ISZ(I,J) = 0
0026                    ISF(I,J) = 0
0027                    ISS(I,J) = 0
0028          400       CONTINUE
0029              NOFSET(J) = 0
0030              IIRZ(J) = 0
0031              IIFL(J) = 0
0032              IFRZ(J) = 0
0033              IFFL(J) = 0
0034              IGDVG(J) = 0
0035              IGCNV(J) = 0
0036              ILF1(J) = 0
0037              IMF1(J) = 0
0038              IHF1(J) = 0
0039              ILF2(J) = 0
0040              IMF2(J) = 0
0041              IHF2(J) = 0
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:07:53                PAGE 002

0042    1000 CONTINUE
0043         PAF = 1.0
0044         AAA = 1.0
       C    INPUT A SYLLABLE
0045         CALL PREADA(NAME,MN1,MN2)
0046         IF (ISR.LE.0) CALL EXIT
0048         ILAST = IFIRST + NSMPL + 4
       C    ANALYSE THE SYLLABLE
       C     PERIPHERAL AUDITORY ANALYSIS
0049         DO 2000 I = IFIRST, ILAST, 2
0050         CALL AUDANL(I)
0051    2000 CONTINUE
       C     ABSTRACT (INTEGRATIVE) AUDITORY ANALYSIS
0052         DO 3000 I = IFIRST, ILAST, 2
0053         CALL PHNANL(I / 2)
0054    3000 CONTINUE
       C     ROUTINE TO GENERATE ADAPTATION
0055         IF (IADAPT.EQ.1) CALL ADPEVL(PAF,AAA)
       C     DEBUG PRINT OUT ROUTINE, PRINTS COMMON ARRAYS
0057         CALL DBGOUT(IO,ILAST,NAME)
       C     PHONETIC FEATURE COMBINATION
       C     DECIDE ON IDENTITY AND PRINT OUT MATRIX
0058         WRITE (6,101)
0059         WRITE (6,109) NAME,MN1,MN2
0060         IF (IADAPT.EQ.1) WRITE (6,115) PAF,AAA
0062         CALL PHONM
       C     OUTPUT RUNNING PHONETIC FEATURE MATRIX
0063         WRITE (6,114) (I, I=40,340,20)
0064         WRITE(6,110) (IPFM(1,I),I=1,25)
0065         WRITE(6,111) (IPFM(2,I),I=1,25)
0066         WRITE(6,112) (IPFM(3,I),I=1,25)
0067         ENDFILE 6
       C     GO BACK AND START OVER
0068         GO TO 200
0069      90 FORMAT (1H ,'ADAPTATION (0-NO,1-YES)-->',$)
0070      98 FORMAT (1H ,'FULL OUTPUT (1-YES,0-NO)-->',$)
0071      99 FORMAT (1H ,'FIRST LINE-->',$)
0072     100 FORMAT (I5)
0073     101 FORMAT (1H1,'SPEECH RECOGNITION SIMULATION')
0074     109 FORMAT (1H0,'STIMULUS (OVE) FILE:',I4,'   MNEMONIC: ',2A2)
0075     110 FORMAT (1H ,'STOP CONSONANT PLACE:',T25,25(I3,1X))
0076     111 FORMAT (1H ,'VOWEL HEIGHT:',T25,25(I3,1X))
0077     112 FORMAT (1H ,'VOWEL PLACE:',T25,25(I3,1X))
0078     113 FORMAT (1H0)
0079     114 FORMAT (1H0,'TIME (MSEC FROM ONSET)  ',25(I3,1X))
0080     115 FORMAT (1H ,'** ADAPTOR, PERIPH: ',F5.3,'  CENTRAL: ',F5.3)
0081         END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:08:48                PAGE 001

      C      INPUT ROUTINE, PRELIMINARY AUDITORY ANALYSIS
0001         SUBROUTINE PREADA(NAME,MN1,MN2)
0002         COMMON /BLKI/ IFIRST,ISR,IDRTN,NSMPL,ISPCH(36,70)
      C      INPUT OF INTENSITY PER CRITICAL BAND OVER TIME
0003         WRITE (5,100)
0004     100 FORMAT (1H ,'FILE NAME ',$)
      C      INPUT FROM DISC FILE
0005         CALL ASSIGN(2,'FILE.DAT',-7,'RDO')
0006         DEFINE FILE 2 (60,36,U,KREC)
      C      INPUT VARIABLES INDICATING TIME FRAME BETWEEN SAMPLES AND
      C       MNEMONIC FOR INPUT
0007         READ (2'1) ISR,NAME,MN1,MN2
0008         J = IFIRST - 1
0009      40 J = J + 1
      C      INPUT INTENSITY PER CRITICAL BAND OVER TIME
0010         READ (2'KREC) (ISPCH(I,J),I=1,36)
0011         IF (J.GT.69) GO TO 80
0013         IF (ISPCH(1,J).NE.-777) GO TO 40
0015         ISPCH(1,J) = 0
0016      80 NSMPL = J - IFIRST + 1
0017         IDRTN = NSMPL * ISR
0018         CALL CLOSE(2)
0019         RETURN
0020         END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:08:59              PAGE 001

       C      PERIPHERAL AUDITORY ANALYSIS STAGE ROUTINE
       C       CURRENT AND NEXT TIME SLICE ARE BLENDED TOGETHER AT THIS STAGE
0001          SUBROUTINE AUDANL(ICTM)
0002          COMMON /BLKI/ IFIRST,ISR,IDRTN,NSMPL,ISPCH(36,70)
0003          COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
             $ ISS(35,35),NOFSET(35)
0004          COMMON /BLKF/ FTGP(4,35),FTGA(6)
0005          L = ICTM / 2
0006          M = ICTM - 4
       C      STEADY STATE DETECTORS
0007          DO 1000 I = 4, 33
       C      DEBUG OUTPUT STATEMENTS
       C      WRITE (5,101) I,ICTM
       C 101 FORMAT (1H ,'I ',I4,', ICTM',I4)
0008          ISS(I,L) = (IGM(I,I,M) + IGM(I,I,M+1)) / 2
0009     1000 CONTINUE
       C     RISE AND FALL DETECTORS
0010          DO 2000 I = 4, 31
       C   RISE
0011          ISZ(I,L) = ((IGM(I,I+1,M) + IGM(I,I+1,M+1)) / 2) * FTGP(1,I)
0012          IRZ(I,L) = (IGM(I,I+2,M) + IGM(I,I+3,M) + IGM(I,I+4,M)
             $ + IGM(I,I+2,M+1) + IGM(I,I+3,M+1) + IGM(I,I+4,M+1))
             $ * FTGP(2,I)
0013     2000 CONTINUE
0014          DO 3000 I = 6, 34
       C   FALL
0015          ISF(I,L) = ((IGM(I,I-1,M) + IGM(I,I-1,M+1)) / 2) * FTGP(3,I)
0016          IFL(I,L) = (IGM(I,I-2,M) + IGM(I,I-3,M) + IGM(I,I-4,M)
             $ + IGM(I,I-2,M+1) + IGM(I,I-3,M+1) + IGM(I,I-4,M+1))
             $ * FTGP(4,I)
0017     3000 CONTINUE
       C     ONSET - OFFSET OF LOW FREQUENCY ENERGY (VOICING) DETECTOR
0018          I1 = ISPCH(1,ICTM-2) + ISPCH(1,ICTM-1) + ISPCH(2,ICTM-2) +
             $ ISPCH(2,ICTM-1) + ISPCH(3,ICTM-2) + ISPCH(3,ICTM-1) +
             $ ISPCH(1,ICTM-3) + ISPCH(2,ICTM-3) + ISPCH(3,ICTM-3)
0019          I2 = ISPCH(1,ICTM) + ISPCH(1,ICTM+1) + ISPCH(2,ICTM) +
             $ ISPCH(2,ICTM+1) + ISPCH(3,ICTM) + ISPCH(3,ICTM+1)
             $ + ISPCH(1,ICTM+2) + ISPCH(2,ICTM+2) + ISPCH(3,ICTM+2)
0020          NOFSET(ICTM / 2) = (I2 - I1) / 9
0021          RETURN
0022          END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:09:21                    PAGE 001

          C  GEOMETRIC MEAN FUNCTION
          C       THIS FUNCTION DOES ACTUAL PATTERN MATCHING FOR PERIPH AUD ANL.
0001              INTEGER FUNCTION IGM(I1,I2,ICTM)
0002              COMMON /BLKI/ IFIRST,ISR,IDRTN,NSMPL,ISPCH(36,70)
0003              IGM = 0
          C       IF INITIAL OR FINAL VALUES ARE NOT LOCAL INTENSITY MAXIMA,
          C        PATTERN DOES NOT MATCH
0004              IF (ISPCH(I1,ICTM).LE.ISPCH(I1-2,ICTM).OR.ISPCH(I1,ICTM).LE.
                $ ISPCH(I1+2,ICTM)) RETURN
0006              L4 = ICTM + 4
0007              IF (ISPCH(I2,L4).LE.ISPCH(I2-2,L4).OR.ISPCH(I2,L4).LE.
                $ ISPCH(I2+2,L4)) RETURN
          C       R IS RETAINING VARIABLE FOR GEMETRIC MEAN
0009              R = 2 * (ISPCH(I1,ICTM) + ISPCH(I2,L4)) - ISPCH(I1-2,ICTM)
                $ - ISPCH(I2-2,L4) - ISPCH(I1+2,ICTM) - ISPCH(I2+2,L4)
          C       SET UP LOCAL CONTROL VARIABLES
0010              L = I2 - I1
0011              M = 1
0012              IF (L.LT.0) M = -1
0014              L = M * L + 1
0015              XP = 2.0
0016              IP1 = I1
0017              IP2 = 0
          C       MAIN LOOP, DO ONCE FOR EACH OF 3 TIME SLICES BETWEEN INITIAL
          C        AND FINAL
0018              DO 7000 J = 1, 3
0019              JC = ICTM + J
0020              IFLAG = 0
0021              IR = 0
          C       INNER LOOP, SCAN TIME SLOT FOR A LOCAL MAXIMA
0022              DO 6000 I = 1, L
0023              K = I1 + I * M - M
0024              IF (ISPCH(K,JC).LE.ISPCH(K-2,JC).OR.ISPCH(K,JC).LE.ISPCH(K+2,
                $ JC)) GO TO 6000
          C       HAVE LOCAL MAXIMA, IF IT IS NOT CONTINUUS WITH THAT OF
          C        PREVIOUS TIME SLOT, GO TO 4000
0026              IF (K.NE.IP1.AND.K.NE.IP1+M) GO TO 4000
0028              IF (IFLAG.NE.0) GO TO 3000
0030              R = R * (2 * ISPCH(K,JC) - ISPCH(K-2,JC) - ISPCH(K+2,JC))
0031              XP = XP + 1.0
0032              IFLAG = K
0033              GO TO 5000
0034       3000   IF (IP2.EQ.0) GO TO 4250
          C       IF LOCAL MAXIMA IS NOT CONTINUUS WITH SECONDARY LOCAL
          C        MAXIMA FROM PREVIOUS LOCAL MAXIMA, GO TO 6000
0036       4000   IF (K.NE.IP2.AND.K.NE.IP2+M) GO TO 6000
0038       4250   R = R * (2 * ISPCH(K,JC) - ISPCH(K-2,JC) - ISPCH(K+2,JC))
0039              XP = XP + 1.0
0040              IF (IFLAG.NE.0) GO TO 4500
0042              IFLAG = K
0043              IP2 = 0
0044              GO TO 5000
0045       4500   IP2 = K
0046       5000   IR = 1
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:09:21                 PAGE 002

          C      IF NO CONTINUUS LOCAL MAXIMA FOUND, RETURN
0047       6000 CONTINUE
0048             IF (IR.EQ.0) RETURN
0050             IP1 = IFLAG
0051       7000 CONTINUE
0052             IF (IP1.NE.I2.AND.IP2.NE.I2.AND.IP1+M.NE.I2.AND.IP2+M.NE.I2)
                $ RETURN
          C      TAKE GEOMETRIC MEAN
0054             IGM = IFIX(R ** (1.0 / XP))
0055             RETURN
0056             END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:10:04                PAGE 001

      C  ABSTRACT AUDITORY LEVEL --STOPS AND VOWELS.
            SUBROUTINE PHNANL(ICTM)
            COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
           $ ISS(35,35),NOFSET(35)
            COMMON /BLKP/ IIRZ(35),IIFL(35),IGDVG(35),ILF1(35),IMF1(35),
           $ IHF1(35),ILF2(35),IMF2(35),IHF2(35),IGCNV(35),IFRZ(35),IFFL(35)
            COMMON /BLKF/ FTGP(4,35),FTGA(6)
      C   VOWEL HEIGHT DETECTORS F1 FREQUENCY REGION
      C     LOW F1
            ILF1(ICTM) = ISS(5,ICTM) + ISS(6,ICTM) + ISS(7,ICTM)
           $ + ISS(8,ICTM)
      C     MIDDLE F1
            IMF1(ICTM) = ISS(8,ICTM) + ISS(9,ICTM) + ISS(10,ICTM) +
           $ ISS(11,ICTM)
      C     HIGH F1
            IHF1(ICTM) = ISS(11,ICTM) + ISS(12,ICTM) + ISS(13,ICTM) +
           $ ISS(14,ICTM)
      C   VOWEL POSITION (FRONT-BACK) DETECTORS F2 AND F3 REGION
      C     LOW F2 & F3
            ILF2(ICTM) = ISS(14,ICTM) + ISS(15,ICTM) + ISS(16,ICTM) +
           $ ISS(17,ICTM) + ISS(18,ICTM) + ISS(19,ICTM) + ISS(28,ICTM)
           $ + ISS(27,ICTM)
      C     MIDDLE F2 & F3
            IMF2(ICTM) =  ISS(20,ICTM) +ISS(21,ICTM) + ISS(22,ICTM) +
           $ ISS(23,ICTM) + ISS(24,ICTM) + ISS(29,ICTM) + ISS(30,ICTM)
      C     HIGH F2 & F3
            IHF2(ICTM) = ISS(25,ICTM) + ISS(26,ICTM) + ISS(27,ICTM) +
           $ ISS(28,ICTM) + ISS(31,ICTM) + ISS(32,ICTM)
      C    INTEGRATIVE RISE, FALL, DIVERGENCE, AND CONVERGENCE DETECTORS
            ISMRF1 = 0
            ISMFF1 = 0
      C     RISING AND FALLING F1 DETECTOR
            DO 500 I = 3, 12
            ISMRF1 = ISMRF1 + IRZ(I,ICTM)
            ISMFF1 = ISMFF1 + IFL(I+2,ICTM)
        500 CONTINUE
      C     ONSET AND OFFSET INDICATORS (RAPID INTENSITY CHANGE)
      C      1 = ONSET: -1 = OFFSET
            IST = 0
            DO 700 I = 1, 9
            IF (NOFSET(ICTM + 5 - I).GE.9) IST = 1
            IF (NOFSET(ICTM + 5 - I).LE.-9) IST = -1
        700 CONTINUE
      C     ADD UP SHALLOW RISE, SHALLOW FALL, STEEP RISE & STEEP FALL
            IRZS = 0
            IFLS = 0
            IZ = 0
            IL = 0
            IZS = 0
            ILS = 0
      C     SUMM OVER ENTIRE F2 & F3 REGION
            DO 1000 I = 13, 32
      C     SHALLOW RISE
            IRZS = IRZS + ISZ(I,ICTM)
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:10:04                PAGE 002

        C      STEEP RISE
0032           IZS = IZS + IRZ(I,ICTM)
        C      RETAIN CRITICAL BAND NUMBER OF HIGHEST STEEP RISE
0033           IF (IRZ(I,ICTM).GT.0) IZ = I
        C      SHALLOW FALL
0035           IFLS = IFLS + ISF(47-I,ICTM)
        C      STEEP FALL
0036           ILS = ILS + IFL(47-I,ICTM)
        C      RETAIN CRITICAL BAND NUMBER OF LOWEST STEEP FALL
0037           IF (IFL(47-I,ICTM).GT.0) IL = 47 - I
0039      1000 CONTINUE
        C      DEBUG OUTPUT STATEMENTS
        C       WRITE (5,100) IL,IZ
        C   100 FORMAT (1H ,'LOW FALL: ',I3,'      HIGH RISE: ',I3)
        C      IF NO RISING F1 OR NO ONSET, SKIP INITIAL INTEGRATIVE DETECTORS
0040           IF (ISMRF1.LE.ISMFF1.OR.IST.NE.1) GO TO 5000
        C     TEST FOR DIVERGENCE.  IF FOUND GO TO 2000
0042           IF (IL.GT.0.AND.IZ.GE.IL-1.AND.IZ.LE.IL+8) GO TO 2000
        C      NO DIVERGENCE, SUM RISE AND FALL INTO INTEGRATIVE RISE AND
        C       FALL DETECTORS
0044           IIRZ(ICTM) = IZS + SQRT(FLOAT(IRZS)) * FTGA(1)
0045           IIFL(ICTM) = ILS + SQRT(FLOAT(IFLS)) * FTGA(2)
        C      SHALLOW DIVERGENCE
0046           IF (IZ.GE.IL-1.AND.IL.GT.0) IGDVG(ICTM) = SQRT(FLOAT((
              $ IRZS + IFLS) / 2)) * FTGA(3)
0048           RETURN
        C      STRONG DIVERGENCE FOUND
0049      2000 CONTINUE
0050           IIRZ(ICTM) = SQRT(FLOAT(IRZS)) * FTGA(1)
0051           IIFL(ICTM) = SQRT(FLOAT(IFLS)) * FTGA(2)
0052           IGDVG(ICTM) = SQRT(FLOAT((IRZS + IFLS) / 2)) * FTGA(3)
0053           IF (ILS.GE.IZS+IRZS.OR.IFLS.LE.5) GO TO 3000
        C      IF SHALLOW FALL, INTEGRATIVE RISE GETS A BOOST
0055           IIRZ(ICTM) = IIRZ(ICTM) + IZS
0056           RETURN
0057      3000 CONTINUE
        C      ADD RISE AND FALL TOGETHER FOR DIVERGENCE
0058           IGDVG(ICTM) = IGDVG(ICTM) + IZS + ILS
0059           RETURN
        C      FINAL POSITION INTEGRATIVE DETECTORS
0060      5000 CONTINUE
        C      IF NO FINAL F1 FALL OR NO OFFSET, SKIP THIS AREA
0061           IF (ISMFF1.LE.ISMRF1.OR.IST.NE.-1) RETURN
        C      TEST FOR CONVERGENCE
0063           IF (IZ + 4.LE.IL.AND.IZ.GT.0.AND.IZ+12.GE.IL) GO TO 6000
        C      NO CONVERGENCE, SUMM RISE AND FALL INTO INTEGRATIVE DETECTORS
0065           IFRZ(ICTM) = IZS + SQRT(FLOAT(IRZS)) * FTGA(5)
0066           IFFL(ICTM) = ILS + SQRT(FLOAT(IFLS)) * FTGA(4)
        C      SHALLOW CONVERGENCE
0067           IF (IZ+4.LE.IL.AND.IZ.GT.0) IGCNV(ICTM) = SQRT(FLOAT((IRZS
              $ + IFLS) / 2)) * FTGA(6)
0069           RETURN
0070      6000 CONTINUE
        C      STEEP CONVERGENCE FOUND
```

```
                IFRZ(ICTM) = SQRT(FLOAT(IRZS)) * FTGA(5)
                IFFL(ICTM) = SQRT(FLOAT(IFLS)) * FTGA(4)
                IGCNV(ICTM) = SQRT(FLOAT((IRZS + IFLS) / 2)) * FTGA(6)
                IF (IZS.GE.ILS+IFLS.OR.IRZS.LE.5) GO TO 7000
      C         IF SHALLOW RISE, INTEGRATIVE FALL GETS A BOOST
                IFFL(ICTM) = IFFL(ICTM) + ILS
                RETURN
          7000 CONTINUE
      C         ADD RISE AND FALL TOGETHER FOR CONVERGENCE
                IGCNV(ICTM) = IGCNV(ICTM) + ILS + IZS
                RETURN
                END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:11:25               PAGE 001

C      DEBUG OUTPUT ROUTINE
       SUBROUTINE DBGOUT(IO,LAST,NAME)
       COMMON /BLKI/ IFIRST,ISR,IDRTN,NSMPL,ISPCH(36,70)
       COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
     $ ISS(35,35),NOFSET(35)
       COMMON /BLKP/ IIRZ(35),IIFL(35),IGDVG(35),ILF1(35),IMF1(35),
     $ IHF1(35),ILF2(35),IMF2(35),IHF2(35),IGCNV(35),IFRZ(35),IFFL(35)
       COMMON /BLKF/ FTGP(4,35),FTGA(6)
C      IS A PRINT OUT OF THE INPUT REQUESTED (IO > 1000)
       IF (IO.LT.1000) GO TO 1100
C       OUTPUT THE INPUT CODING OF INTEN X FREQ X TIME
       WRITE (6,100) NAME
       WRITE (6,103) (I,I=1,36)
       DO 1000 I = IFIRST-1, LAST
       WRITE (6,101) I,(ISPCH(J,I),J=1,36)
  1000 CONTINUE
C      IS A PRINTOUT OF THE RESULTS OF PERIPHERAL AUDITORY ANALYSIS
C       WANTED (HUNDREDS DIGIT OF IO IS NON-ZERO)
  1100 IF (MOD(IO,1000).LT.100) GO TO 2100
       WRITE (6,102) NAME
       WRITE (6,103) (I,I=1,35)
       DO 2000 K = IFIRST, LAST, 2
       I = K / 2
       WRITE (6,101) K,(ISS(J,I),J=1,35),NOFSET(I)
       WRITE (6,103) (ISZ(J,I),J=1,35)
       WRITE (6,103) (IRZ(J,I),J=1,35)
       WRITE (6,103) (ISF(J,I),J=1,35)
       WRITE (6,103) (IFL(J,I),J=1,35)
  2000 CONTINUE
C      IS A PRINTOUT OF THE RESULTS OF ABSTRACT AUDITORY ANALYSIS
C       WANTED (TENS DIGIT OF IO IN NON-ZERO)
  2100 IF (MOD(IO,100).LT.10) GO TO 3100
       WRITE (6,105) NAME
       DO 3000 K = IFIRST, LAST, 2
       I = K / 2
       WRITE (6,104) K,IIRZ(I),IIFL(I),IGDVG(I),IFRZ(I),IFFL(I),
     $ IGCNV(I),ILF1(I),IMF1(I),IHF1(I),ILF2(I),IMF2(I),IHF2(I)
  3000 CONTINUE
C      IS A PRINT OUT OF ADAPTATION PARAMETERS WANTED (ONES DIGIT OF
C       IO IS NON-ZERO)
  3100 IF (MOD(IO,10).LT.1) RETURN
       WRITE (6,106) NAME
       WRITE (6,103) (I,I=1,35)
       DO 4000 I = 1, 4
       WRITE (6,107) I,(FTGP(I,J),J=1,35)
  4000 CONTINUE
       WRITE (6,108) (FTGA(I),I=1,6)
       RETURN
   100 FORMAT (1H1,'PRELIMINARY AUDITORY ANALYAIS (INPUT): ',I3)
   101 FORMAT (1H ,'TIME: ',I2,1X,35(',',I2),',',I3)
   102 FORMAT (1H1,'PERIPHERAL AUDITORY FEATURE ANALYSIS: ',I3)
   103 FORMAT (1H ,9X,36(',',I2))
   104 FORMAT (1H ,'TIME: ',I2,2X,12(',',I5))
   105 FORMAT (1H1,'ABSTRACT AUDITORY FEATURES: ',I3/' ',12X,' IRZ   IFL'
```

```
            $ ,'   DVG    FRZ    FFL   CNVG    LF1    MF1    HF1    LF2    MF2    HF2')
0047    106 FORMAT (1H1,'FATIGUE PARAMETERS: ',I3)
0048    107 FORMAT (1H ,2X,I3,4X,35F3.1)
0049    108 FORMAT (1H0,'  IRZ    IFL   DVG    FFL    FRZ  CNVG',/,2X,
            $ 6(F5.3,1X))
0050        END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:11:54                PAGE 001

      C     SUBROUTINE TO EVALUATE AN INPUT AS AN ADAPTOR & APPLY ADAPTATION
      C       ADAPTATION PARAMETERS
              SUBROUTINE ADPEVL(PF,AF)
              INTEGER IS(6)
              COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
             $ ISS(35,35),NOFSET(35)
              COMMON /BLKP/ IIRZ(35),IIFL(35),IGDVG(35),ILF1(35),IMF1(35),
             $ IHF1(35),ILF2(35),IMF2(35),IHF2(35),IGCNV(35),IFRZ(35),IFFL(35)
              COMMON /BLKF/ FTGP(4,35),FTGA(6)
      C       INPUT 2 ADAPTATION PARAMETERS, PERIPHERAL FATIGUE (PF) AND
      C         ABSTRACT RETUNING (AF)
              WRITE (5,100)
          100 FORMAT (1H ,'ENTER PERIPHERAL AND ABSTRACT FATIGUE PARAM(2F6.3)')
              READ (5,101) PF,AF
          101 FORMAT (2F6.3)
      C       EVALUATE PERIPHERAL FATIGUE
              DO 2000 I = 1, 34
              IS(1) = 0
              IS(2) = 0
              IS(3) = 0
              IS(4) = 0
      C       SUM RISE AND FALL
              DO 1000 J = 1, 35
              IS(1) = IS(1) + ISZ(I,J)
              IS(2) = IS(2) + IRZ(I,J)
              IS(3) = IS(3) + ISF(I+1,J)
              IS(4) = IS(4) + IFL(I,J)
         1000 CONTINUE
      C       IF BOTH SHALLOW RISE AND SHALLOW FALL OCCUR, FATIGUE BOOTH
              IF (IS(1).EQ.0.OR.IS(1).NE.IS(3)) GO TO 1200
              FTGP(1,I) = ((1.0 - PF) / 2.0) + PF
              FTGP(3,I+1) = FTGP(1,I)
      C       APPLY FATIGUE TO ANY STEEP DETECTOR THAT RESPONDS AND ITS
      C         ASSOCIATED SHALLOW DETECTOR
         1200 IF (IS(1).GT.IS(3)) FTGP(1,I) = PF
              IF (IS(3).GT.IS(1)) FTGP(3,I+1) = PF
              IF (IS(2).LE.0) GO TO 1400
              FTGP(1,I) = PF
              FTGP(1,I+1) = PF
              FTGP(2,I) = PF
         1400 IF (IS(4).LE.0) GO TO 2000
              FTGP(3,I) = PF
              FTGP(3,I-1) = PF
              FTGP(4,I) = PF
         2000 CONTINUE
      C         EVALUATE ABSTRACT RETUNING
              DO 3000 I = 1, 6
              IS(I) = 0
         3000 CONTINUE
      C       SUM ABSTRACT OUTPUTS
              DO 4000 I = 1, 35
              IS(1) = IS(1) + IIRZ(I)
              IS(2) = IS(2) + IIFL(I)
              IS(3) = IS(3) + IGDVG(I)
```

```
0047          IS(4) = IS(4) + IFFL(I)
0048          IS(5) = IS(5) + IFRZ(I)
0049          IS(6) = IS(6) + IGCNV(I)
0050     4000 CONTINUE
        C     APPLY RETUNING TO MAXIMUM OUTPUT INITIAL DETECTOR AND MAX OUTPUT
        C      FINAL DETECTOR
0051          DO 5000 I = 1, 3
0052          IF (IS(I).GT.0.AND.IS(I).EQ.MAX0(IS(1),IS(2),IS(3))) GO TO 4500
0054          IF (IS(I+3).LE.0.OR.IS(I+3).NE.MAX0(IS(4),IS(5),IS(6)))
             $ GO TO 5000
0056     4500 FTGA(I) = AF
0057          FTGA(I+3) = AF
0058     5000 CONTINUE
0059          RETURN
0060          END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:12:24              PAGE 001

      C  INITIALIZE FATIGUE ARRAYS: ROUTINE TO SETUP ADAPTATION ARRAYS
      C       WITH NO ADAPTATION
            SUBROUTINE SADP
            COMMON /BLKF/ FTGP(4,35),FTGA(6)
            DO 1000 I = 1, 4
                  DO 1000 J = 1, 35
                  FTGP(I,J) = 1.0
       1000 CONTINUE
            DO 2000 I = 1, 6
            FTGA(I) = 1.0
       2000 CONTINUE
            RETURN
            END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:12:41                    PAGE 001

      C   PHONETIC FEATURE COMBINATION STAGE
0001        SUBROUTINE PHONM
0002        COMMON /BLKA/ IRZ(35,35),IFL(35,35),ISZ(35,35),ISF(35,35),
          $ ISS(35,35),NOFSET(35)
0003        COMMON /BLKP/ IIRZ(35),IIFL(35),IGDVG(35),ILF1(35),IMF1(35),
          $ IHF1(35),ILF2(35),IMF2(35),IHF2(35),IGCNV(35),IFRZ(35),IFFL(35)
0004        COMMON /BLKD/ IPFM(3,25)
      C     IPFM(1,I)   PLACE FEATURE FOR STOPS
      C                       1 :  BILABIAL
      C                       2 :  ALVEOLAR-DENTAL
      C                       3 :  VELAR
      C    IPFM(2,I)  HEIGHT FEATURE FOR VOWELS
      C                       1 :  HIGH
      C                       2 :  MEDIUM
      C                       3 :  LOW
      C   IPFM(3,I)  PLACE FEATURE FOR VOWELS
      C                       1 :  FRONT
      C                       2 :  MIDDLE
      C                       3 :  BACK
      C    ZERO MEANS NO VALUE
      C   CLEAR PHONETIC FEATURE MATRIX
0005        DO 1000 I = 1, 3
0006              DO 1000 J = 1, 25
0007              IPFM(I,J) = 0
0008   1000 CONTINUE
0009        T1 = 0.0
0010        T2 = 0.0
0011        T3 = 0.0
0012        U1 = 0.0
0013        U2 = 0.0
0014        U3 = 0.0
0015        J = 1
0016        DO 3000 K = 4, 34, 2
      C   PLACE FEATURE: RWM IS A REAL WEIGHTING FUNCTION
      C      OUTPUTS OF ABSTRACT ANALYSIS FOR PREVIOUS 40 MSEC ARE
      C      COMBINED AND WEIGHTED
0017        B = RWM(IIRZ(K),IIRZ(K-1),IIRZ(K-2),IIRZ(K-3))
0018        C = RWM(IFRZ(K),IFRZ(K-1),IFRZ(K-2),IFRZ(K-3))
0019        A = RWM(IIFL(K),IIFL(K-1),IIFL(K-2),IIFL(K-3))
0020        D = RWM(IFFL(K),IFFL(K-1),IFFL(K-2),IFFL(K-3))
0021        V = RWM(IGDVG(K),IGDVG(K-1),IGDVG(K-2),IGDVG(K-3))
0022        W = RWM(IGCNV(K),IGCNV(K-1),IGCNV(K-2),IGCNV(K-3))
0023        M = 0
      C     TABULATE ONSET AND OFFSET
0024        DO 2000 I = 1, 7
0025        IF (NOFSET(K + 4 - I).GE.9) M = 1
0027        IF (NOFSET(K + 4 - I).LE.-9) M = -1
0029   2000 CONTINUE
      C     IF ONSET, USE INITIAL ABSTRACT OUTPUTS
0030        IF (M.GT.0) IPFM(1,J) = IMXCNV(B,A,V)
      C      IF OFFSET, USE FINAL ABSTRACT OUTPUTS
0032        IF (M.LT.0) IPFM(1,J) = IMXCNV(D,C,W)
      C      TABULATE RATINGS
0034        IF (IPFM(1,J).EQ.0) GO TO 2500
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:12:41                PAGE 002

0036         IF (M.LT.0) CALL RATING(D,C,W,U1,U2,U3)
0038         IF (M.GT.0) CALL RATING(B,A,V,T1,T2,T3)
0040    2500 CONTINUE
      C   VOWEL HEIGHT
0041         B = RWM(ILF1(K),ILF1(K-1),ILF1(K-2),ILF1(K-3))
0042         A = RWM(IMF1(K),IMF1(K-1),IMF1(K-2),IMF1(K-3))
0043         V = RWM(IHF1(K),IHF1(K-1),IHF1(K-2),IHF1(K-3))
0044         IPFM(2,J) = IMXCNV(B,A,V)
      C   VOWEL PLACE
0045         B = RWM(ILF2(K),ILF2(K-1),ILF2(K-2),ILF2(K-3))
0046         A = RWM(IMF2(K),IMF2(K-1),IMF2(K-2),IMF2(K-3))
0047         V = RWM(IHF2(K),IHF2(K-1),IHF2(K-2),IHF2(K-3))
0048         IPFM(3,J) = IMXCNV(V,A,B)
0049         J = J + 1
0050    3000 CONTINUE
0051         WRITE (6,100) T1,T2,T3
0052         WRITE (6,100) U1,U2,U3
0053     100 FORMAT (1H0,'AVE(1): ',F7.3,'  AVE(2): ',F7.3,'  AVE(3): ',F7.3)
0054         RETURN
0055         END
```

```
FORTRAN IV      V01C-03A    THU 05-AUG-76 00:13:15                 PAGE 001

      C     MAXIMUM CONVERSION FUNCTION, FINDS MAX OF 3 REAL NUMBERS
             INTEGER FUNCTION IMXCNV(V1,V2,V3)
             IF (AMAX1(V1,V2,V3).EQ.V1) IMXCNV = 1
             IF (AMAX1(V1,V2,V3).EQ.V2) IMXCNV = 2
             IF (AMAX1(V1,V2,V3).EQ.V3) IMXCNV = 3
             IF (V1.EQ.V2.AND.V1.GE.V3) IMXCNV = 0
             RETURN
             END
```

```
FORTRAN IV      V01C-03A   THU 05-AUG-76 00:13:20                PAGE 001

      C     EXPONENTIAL REAL WEIGHTING FUNCTION, EXPONENTIAL DECAY
      C       WEIGHTING OVER 40 MSEC WINDOW
              REAL FUNCTION RWM(I1,I2,I3,I4)
              RWM = I1 + I2 * EXP(-0.5) + I3 * EXP(-1.0) + I4 * EXP(-1.5)
              RETURN
              END
```

```
      C      TABULATE RATINGS
             SUBROUTINE RATING(A1,A2,A3,R1,R2,R3)
              R1 = R1 + A1
              R2 = R2 + A2
              R3 = R3 + A3
             RETURN
             END
```

Appendix C

PRELIMINARY AUDITORY ANALYAIS (INPUT): 264

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36
TIME: 5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 6	29	30	30	29	29	28	25	24	22	21	20	19	18	18	18	17	18	18	22	26	16	14	13	12	12	13	17	22	11	8	7	9	15	9	4	1
TIME: 7	29	30	30	29	28	28	29	25	23	22	20	19	18	18	18	17	17	17	19	34	18	14	13	12	12	13	17	22	11	8	7	9	15	9	4	1
TIME: 8	29	30	30	29	27	27	28	27	24	22	21	19	18	19	18	17	18	17	19	34	18	14	13	12	12	12	14	32	13	9	8	9	16	9	4	1
TIME: 9	29	30	30	29	27	26	26	27	28	23	21	20	19	19	18	17	17	17	18	27	21	15	13	12	12	12	14	33	13	9	8	9	16	9	4	1
TIME: 10	29	30	30	29	27	26	26	26	28	26	22	20	19	19	18	17	17	17	17	22	29	16	14	12	12	12	12	21	17	10	8	9	16	9	4	1
TIME: 11	31	32	32	31	29	28	27	27	27	30	27	23	21	21	20	19	19	19	19	24	31	18	16	14	14	14	14	23	19	12	10	11	18	11	6	3
TIME: 12	31	32	32	31	29	28	27	26	26	27	31	26	22	21	20	19	19	19	19	21	35	20	16	15	14	14	14	23	19	12	10	11	18	11	6	3
TIME: 13	31	32	32	31	29	28	27	26	26	26	28	30	23	22	20	19	19	19	19	21	35	20	16	14	14	13	13	17	31	13	10	11	18	11	6	3
TIME: 14	31	32	32	31	29	28	27	26	25	25	25	28	28	23	21	20	19	19	18	20	26	24	17	15	14	14	14	17	31	13	10	11	18	12	6	3
TIME: 15	31	32	32	31	29	28	27	26	25	25	25	28	28	23	21	20	19	19	18	20	26	24	17	15	14	14	14	17	31	13	10	11	18	12	6	3
TIME: 16	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 17	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 18	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 19	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 20	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 21	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 22	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 23	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 24	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 25	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 26	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 27	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 28	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 29	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 30	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 31	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 32	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 33	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 34	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 35	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 36	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 37	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 38	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 39	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 40	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 41	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 42	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 43	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 44	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 45	33	34	34	33	31	30	29	28	27	27	27	30	30	25	23	22	21	21	20	22	28	26	19	17	16	16	16	19	33	15	12	13	20	14	8	5
TIME: 46	31	32	32	31	29	28	27	26	25	25	25	28	28	23	21	20	19	19	18	20	26	24	17	15	14	14	14	17	31	13	10	11	18	12	6	3
TIME: 47	31	32	32	31	29	28	27	26	25	25	25	28	28	23	21	20	19	19	18	20	26	24	17	15	14	14	14	17	31	13	10	11	18	12	6	3
TIME: 48	29	30	30	29	27	26	25	24	23	23	23	26	26	21	19	18	17	17	16	18	24	22	15	13	12	12	12	15	29	11	8	9	16	10	4	1
TIME: 49	27	28	28	27	25	24	23	22	21	21	21	24	24	19	17	16	15	15	14	16	22	20	13	11	10	10	10	13	27	9	6	7	14	8	2	-1
TIME: 50	25	26	26	25	23	22	21	20	19	19	19	22	22	17	15	14	13	13	12	14	20	18	11	9	8	8	8	11	25	7	4	5	12	6	0	-3
TIME: 51	25	26	26	25	23	22	21	20	19	19	19	22	22	17	15	14	13	13	12	14	20	18	11	9	8	8	8	11	25	7	4	5	12	6	0	-3
TIME: 52	23	24	24	23	21	20	19	18	17	17	17	20	20	15	13	12	11	11	10	12	18	16	9	7	6	6	6	9	23	5	2	3	10	4	-2	-5
TIME: 53	21	22	22	21	19	18	17	16	15	15	15	18	18	13	11	10	9	9	8	10	16	14	7	5	4	4	4	7	21	3	0	1	8	2	-4	-7
TIME: 54	21	22	22	21	19	18	17	16	15	15	15	18	18	13	11	10	9	9	8	10	16	14	7	5	4	4	4	7	21	3	0	1	8	2	-4	-7
TIME: 55	19	20	20	19	17	16	15	14	13	13	13	16	16	11	9	8	7	7	6	8	14	12	5	3	2	2	2	5	19	1	-2	-1	6	0	-6	-9
TIME: 56	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
TIME: 62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

```
PERIPHERAL AUDITORY FEATURE ANALYSIS: 264
          , 1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,
TIME:  6  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 29
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME:  8  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  9
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 10  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,14, 0, 0, 0, 0, 0, 0, 0,17, 0, 0, 0, 0,12, 0, 0,  1
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7,14, 0, 0, 0, 0, 0, 0,10,16, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,19, 0, 0, 0, 0, 0, 0, 0,20, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 12  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 8, 8, 0, 0, 0, 0, 0, 0,14, 0, 0, 0, 0,12, 0, 0,  1
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3,13, 0, 0, 0, 0, 0, 0, 9,14, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 9,10,11, 0, 0, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,23, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 5, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 14  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,16, 0, 0, 0, 0, 0, 0, 8,14, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 9,15, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0,12,12, 0, 0, 0, 0, 0, 0, 0, 0, 0,18, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,12, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 16  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 3, 0, 0, 0, 0, 0, 0, 0, 0,12, 0, 0, 0, 0, 0, 0, 5,20, 0, 0, 0,12, 0, 0,  2
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 3, 0, 0, 0, 0, 0, 0, 0, 9,12, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0,11,14, 0, 0, 0, 0, 0, 0, 0, 0,19, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,12, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 18  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 20  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 22  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 24  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 26  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 28  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
```

```
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,-0, 0, 0, 0, 0, 0, 0,
TIME: 30  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 32  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 34  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 36  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 38  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 40  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 42  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 44  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,   0
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 46  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  -2
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 48  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  -4
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 50  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  -4
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 52  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  -4
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 54  , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 6, 7, 0, 0, 0, 0, 0, 0, 0,11, 8, 0, 0, 0, 0, 0, 5,21, 0, 0, 0,12, 0, 0,  -9
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 0, 0, 0, 0, 0, 0, 0, 0,11, 0, 0, 0, 0, 0, 0,13, 0, 0, 0, 0,11, 0,
          , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

```
TIME: 56 , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,-21
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 58 , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, -6
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 60 , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  0
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
TIME: 62 , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,  0
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
         , 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0,
```

ABSTRACT AUDITORY FEATURES: 264

		IRZ	IFL	DVG	FRZ	FFL	CNVG	LF1	MF1	HF1	LF2	MF2	HF2
TIME: 6	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0
TIME: 8	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0
TIME: 10	,	45,	0,	0,	0,	0,	0,	0,	0,	0,	17,	14,	17
TIME: 12	,	42,	2,	0,	0,	0,	0,	0,	0,	0,	14,	16,	14
TIME: 14	,	24,	3,	0,	0,	0,	0,	0,	0,	0,	8,	30,	8
TIME: 16	,	24,	3,	0,	0,	0,	0,	0,	0,	3,	5,	32,	5
TIME: 18	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 20	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 22	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 24	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 26	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 28	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 30	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 32	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 34	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 36	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 38	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 40	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 42	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 44	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 46	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 48	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 50	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 52	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 54	,	0,	0,	0,	0,	0,	0,	0,	0,	13,	5,	40,	5
TIME: 56	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0
TIME: 58	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0
TIME: 60	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0
TIME: 62	,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0,	0

```
FATIGUE PARAMETERS: 264
        , 1, 2, 3, 4, 5, 6, 7, 8, 9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24,25,26,27,28,29,30,31,32,33,34,35,
   1    1.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.0
   2    1.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.0
   3    1.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.0
   4    1.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.01.0

  IRZ   IFL   DVG   FFL   FRZ  CNVG
1.000 1.000 1.000 1.000 1.000 1.000
```

SPEECH RECOGNITION SIMULATION

STIMULUS (OVE) FILE: 264   MNEMONIC: BAE

AVE(1): 69.294 AVE(2): 2.000 AVE(3): 0.000

AVE(1): 0.000 AVE(2): 0.000 AVE(3): 0.000

TIME (MSEC FROM ONSET)	40	60	80	100	120	140	160	180	200	220	240	260	280	300	320	340									
STOP CONSONANT PLACE:	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
VOWEL HEIGHT:	0	0	3	3	3	3	3	3	3	3	3	3	3	3	0	0	0	0	0	0	0	0	0	0	0
VOWEL PLACE:	0	2	2	2	2	2	2	2	2	2	2	2	2	2	0	0	0	0	0	0	0	0	0	0	0