RESEARCH ON SPEECH PERCEPTION

Progress Report No. 2

September 1974 - August 1975

David B. Pisoni
Principal Investigator

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-24027-02

and

National Institute of Health
Grant No. NS-12179-01

Contents

INTRODUCTION

This is the second report of research activities on speech perception conducted in the Department of Psychology at Indiana University. The main purpose of this report is to summarize our research activities over the past year and make them available to interested colleagues. Some of the papers contained in this report are extended manuscripts that have been prepared for publication as journal articles or book chapters. Other papers are short reports of research that has been presented at professional meetings during the past year or brief progress reports on the state of on-going research projects in our laboratory. Because of the lag in journal publications and our interest in following the most recent work of other colleagues interested in speech perception, we would be most grateful if you would send us copies of your own reprints, preprints and progress reports. Please address all correspondence to:

Professor David B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47401
U.S.A.

EXTENDED MANUSCRIPTS

Some Stages of Processing in Speech Perception

David B. Pisoni and James R. Swausch

Department of Psychology

Indiana University

Bloomington, Indiana 47401

U.S.A

This paper was presented at the "Symposium on Dynamic Aspects of Speech Perception: held at the Instituut voor Perceptie Onderzoek, Eindhoven, The Netherlands, August 4, 5, and 6, 1975, and appears in A. Cohen and S. Nooteboom (Eds.), Structure and Process in Speech Perception. Heidelberg: Springer-Verlag, 1975.

Some Stages of Processing in Speech Perception

David B. Pisoni and James R. Sawusch

Indiana University
Bloomington, Indiana 47401 U.S.A.

Introduction

In this paper we consider a number of theoretical issues that have occupied the attention of investigators concerned with human speech perception. Most extant theories of speech perception have been quite general and vague and for the most part not terribly well developed. It is our view that many of the theoretical issues have not been dealt with adequately in the past primarily because of the failure of these theories to make explicit and precise statements that can be tested empirically. Indeed, most of the current theories are formulated in such broad terms that the scope of the theory cannot be delimited in any satisfactory way to deal with existing phenomenon or predict new findings.

The present contribution is not an empirical paper and we do not plan to report any new experimental findings. Rather, we feel that an attempt to deal with some of the fundamental issues in speech perception within an organized framework is warranted at this time, especially at a conference such as this. We consider speech perception in both the broad sense, encompassing syntactic and semantic analysis, and, in a more narrow sense, involving only auditory and phonetic analysis. However, our major emphasis will be on the earliest stages of perceptual analysis of speech. The contents of this chapter are organized into four major sections. In the first section, we briefly summarize what we see as the major theoretical issues in speech perception. Several of the traditional issues involving segmental analysis will be touched upon briefly here such as the lack of invariance and segmentation, the role of articulation in speech perception and units of perceptual analysis. In the second section, a distinction between taxonomic and dynamic models of speech perception is considered along with a brief review of several prominent speech perception theories. In the third section, we describe a fairly general information processing model of speech perception. Although the major emphasis in this discussion is on phonetic perception, the model relies heavily on syntactic and semantic processes which are used to guide earlier stages of perceptual analysis. In the fourth section, we briefly examine the role of higher-level linguistic processes in speech perception.

Our major purpose in preparing this contribution has been two-fold. First, we believe that it is useful to theorize in a general way about the process of speech perception outside the context of the particular experimental paradigms. For the present it seems like a reasonably good strategy to focus on general questions in speech perception rather than particular issues. The second purpose is our hope that the approach we have taken here will prove to be useful not only in thinking about speech perception but also in stimulating further theoretical discussion in the future.

Some Theoretical Issues in Speech Perception

Invariance and Segmentation

Most investigators agree that the two most important problems for speech perception theory deal with the lack of invariance and segmentation of the physical signal. The earliest observations showed that there was generally a lack of one-to-one relation between information in the acoustic signal and segments resulting from linguistic analysis. Examination of spectrographic patterns of real speech indicated that there were, in general, no segments in the speech signal that corresponded uniquely to segments in the message. The early perceptual experiments with synthetic speech indicated that a single segment of the acoustic signal carried information about several successive phonetic segments. Other studies showed that very different acoustic signals were often perceived as identical linguistic segments and identical acoustic signals were often perceived as different linguistic segments depending on the context. The acoustic cues to a given phonetic segment have been shown to be numerous and quite variable depending on the phonetic context, stress, speaking rate and speaker, which are natural consequences of the dynamics of the articulatory apparatus. For our purposes, the most important implication of these findings is that they place fairly rigid constraints on the types of perceptual theories that might be proposed. For example, filter and template matching theories would seem to be poor candidates for phonetic perception since the segments for analysis cannot be defined exclusively by properties of the physical signal. Based on these considerations the only acceptable class of theories are those that employ higher level linguistic information in the earliest stages of recognition.

Articulation in Speech Perception

Theoretical accounts of speech perception have often emphasized some type of articulatory-motor involvement during perceptual processing. There seem to be two primary reasons for this concern with articulation. The first is basically historical. Acoustic phonetics and much of the experimental research on speech perception were guided by the traditional articulatory descriptions developed by phoneticians. Thus, the acoustic cues underlying the perception of different classes of sounds were related to the articulatory gestures and dimensions that distinguished these sounds in production. Early work indicated that speech sounds could be specified precisely in terms of a few simple and relatively independent dimensions which provided the same distinctions in perception that articulatory dimensions provided in production and suggested that perceptual and articulatory processes may be intimately linked.

A second reason for assuming some type of interaction between perception and production is that speech sounds are produced by a sound source which has well defined acoustic constraints for the listener (Fant, 1960; Flanagan, 1972). Recent perceptual work indicates a close articulatory-perceptual match between changes in vocal tract shape and the acoustic signal. For example, Stevens (1972) has found that in certain parts of the vocal tract, a small change in the position of the tongue constriction resulted in a large variation of the formant frequencies; whereas in other regions of the vocal tract which correspond to the "natural" articulatory positions, a large variation in the position of the major tongue constriction had only a small effect on the formant frequencies. Stevens argues that the articulatory positions for particular phonetic segments, especially the consonants, have been chosen at locations where large deviations in articulation will produce only minimal changes in the acoustic signal. The existence of "Quantal" relations of this type between articulation and sound output may place constraints on the possible inventory of articulatory and acoustic properties that may be used as the basis for phonetic features in natural languages. Thus, according to Stevens, an acoustic property that would be a suitable candidate for a phonetic feature would be one that could be generated by the vocal tract without requiring very precise positioning or manipulating of the articulatory apparatus. Sounds produced in these "natural" regions might represent both articulatory and acoustic correlates of a particular phonetic feature. Thus, if knowledge of articulatory constraints is used in perception, the important questions for any theory of speech perception become <u>where</u> and <u>how</u> this knowledge is put to use by the perceptual system. In our view, the failure of the Motor Theory and Analysis by Synthesis approaches as psychological theories stems from their vagueness in specifying how knowledge of articulation would be used in perception.

Units of Perceptual Analysis

A central issue in speech perception has always been the choice of a unit of perceptual analysis. Numerous investigators have argued for the primacy of the feature, phoneme, syllable or word as their candidate for the basic perceptual unit. Other investigators, motivated chiefly by developments in generative linguistic theory, have proposed much larger units for perceptual analysis such as clauses or sentences. Many of the problems about the choice of a perceptual unit could be resolved if a distinction were made as to the level of perceptual analysis under consideration. It should be obvious that the size of the processing or structural unit will vary as a function of the level of processing in the linguistic system. There is now sufficient psychological evidence available to argue for the simultaneous existence of all units of linguistic analysis from features to syllables to clauses. To argue that there is one basic unit in speech perception is to acknowledge that language exists primarily in one form rather than another. However, the major fact about human language is that it exists in many forms, most of which are inaccessible to conscious inspection. Thus, we normally do not listen to phonetic details in perceiving speech. We recognize words, not sounds, and find it difficult to separate our observations from their interpretations. For example, consider the distinction between phonemes and syllables. Several investigators have argued that the syllable is a more basic perceptual or linguistic unit than the phoneme. The major claim here is that the phoneme is more "abstract" than the syllable because syllables exist independently as articulatory or acoustic units and phonemes do not. But most of these arguments have failed to take note of the symbiotic function of phonemes and syllables. As Studdert-Kennedy points out, the syllable serves as the carrier of phonetic information; its purpose is to create contrast for phonetic perception so that the listener can detect sound segments and features for the subsequent operations of the perceptual system. By choosing the syllable as a basic unit, some investigators have assumed that the invariance issue could be resolved. But there are several problems with this line of reasoning too. First, there are difficulties in defining the syllable acoustically. Second, syllables are not indivisible units; they have a complex internal structure which at some point must be recovered by the perceptual system. And finally, the acoustic cues for syllables are affected by context in much the same way as the cues for phonemes. One way to resolve the issue of units is to simply assume that the perceptual mechanism requires information distributed over at least a whole syllable in order to derive minimal information for a reliable phonetic interpretation. We should note, as Fodor, Bever and Garrett (1974) have remarked, that this strategy does not necessarily reflect on the size of the perceptual unit but only on the type of information that this analysis requires.

Syntax and Semantics

It is well known that words presented in sentential context are more intelligible than the same words presented in isolation. The traditional explanation of this finding was that syntax and semantics played only a minor role in perception by narrowing down or constraining the number of possible alternatives available to the listener. This view of speech perception assumed that the process was organized serially and that phonetic and phonemic characteristics of an utterance were recognized more or less directly from a sequence of physical properties in the sound wave. However, much of the research on the perception of connected speech indicates that processing does not proceed serially on the basis of a sequence of independent decisions about individual segments. Indeed, a good part of the information used in the perceptual process is not determined exclusively by the physical properties of the signal. There is good evidence, for example, that it is extremely difficult if not impossible for a linguist to discover the correct phonemic representation of an utterance without the use of higher-level syntactic information (see for example, Shockey & Reddy, 1974). And the limited success of recent speech understanding systems has been brought about by the explicit deployment of syntactic, semantic and conceptual information in the recognition process. This suggests that syntactic and semantic information must play an active and intimate role in any theory of speech perception since this information guides the early stages of phonetic and phonological processing. As we note below, most of the current theories of speech perception have failed to consider syntax and semantics seriously.

Prosody and Speech Perception

Perhaps one of the most neglected topics of research in speech perception is prosody. As Svensson (1974) has recently observed, there is a very wide gap between the work on isolated segments and features and the role of syntactic factors in speech perception. For the most part, prosody and grammar have been ignored in speech perception theories. The way prosodic information might be used in the recognition process has not been considered in any detail. This is surprising since there is some evidence that prosodic features such as pitch, for example, may provide cues to segment the speech stream into constituents that would be suitable candidates for further syntactic analysis (Lea, 1973). We feel that this is an important problem that must be considered in any theoretical account of speech perception. Prosody may serve as the interface between low-level segmental information and higher-levels of grammatical structure in speech.

Paradigms in Speech Perception Research

Several remarks need to be made about paradigms in speech perception. Much of the speech perception research that has been carried out over the last twenty-five years developed around specific experimental paradigms instead of being focused on general issues. Thus, most of the experimental questions that investigators devoted their attention to were defined by the particular types of experiments that were currently in vogue rather than being derived from some specific theoretical perspective. This is, of course, not surprising since relatively few specific predictions could be derived from the currently available theories. For example, numerous categorical perception experiments were carried out, the results of which were used for a long time as the major empirical support for the Motor Theory of Speech Perception. More recently, there has been a strong interest in the use of dichotic listening techniques to study hemispheric differences in speech perception and to detail some of the processes underlying speech perception. Although the dichotic listening findings may have important implications for understanding some aspects of brain function in speech perception, we feel that most of the results have only limited theoretical value. Indeed, we would be willing to argue that much of the dichotic listening work has detracted from the main goal of research in this area, namely, the development of a theory of speech perception. Finally, there has been a great deal of work in the last year or two using selective adaptation procedures to study the functional organization of the speech processing system (see Cooper, 1975 for a review). The early work with this technique raised the issue of the possible existence of feature or property detectors which mediate the perception of various phonetic distinctions. The specific issues are intricate and the arguments complicated. But there seems to be fairly good evidence for the existence of property detectors which respond to certain types of acoustic information in the signal. The evidence for specialized detectors that respond to more abstract phonetic information is somewhat weaker. Nevertheless, work using this technique has proceeded at an extremely rapid rate. It is, however, still too early to tell whether this line of research will produce any long range payoffs. To summarize, we feel that there is clearly a need to focus on basic issues rather than paradigms in speech perception research if we are to make any advances in the field and if we are to develop a strong theoretical framework in which to pursue these questions.

Models of Speech Perception

Taxonomic vs. Dynamic

Historically models of speech perception were derived from the taxonomic principles of segmentation and classification. It was assumed that linguistic structure, especially the phonological level, could be discovered by the use of these procedures applied automatically to a corpus of data. Thus, according to this view, a separate level of phonological structure could be described independently of higher-level grammatical or syntactic considerations. More recent models, developed within the framework of generative linguistic theory, however, have stressed a more active or dynamic view of the process. For example, Chomsky and Halle (1968) have argued quite convincingly that speech perception is an "active process" in which acoustic information is utilized to form hypotheses about the structure of sentences. These hypotheses which are then based in part on expectations and knowledge of the language are subsequently used to generate structures which are compared against the original input. In this way, the listener uses phonological principles to determine the phonetic properties of the stimulus. As Chomsky and Halle have put it, what the listener "hears" is what is internally generated by the rules for the hypothesized structure. This view is, of course, a version of analysis by synthesis which we will discuss below. Although it is a very abstract theory, it emphasizes an important point that has often been overlooked. A listener's interpretation of a speech stimulus may not be determined exclusively by the properties of the physical signal. Since our conscious experience of language almost always follows the semantic intent of the message and not its form, it is difficult to become aware of the physical properties of the stimulus. Similarly, the final semantic interpretation may involve information which has no physical correlate. Thus, what is finally perceived depends on the listener's knowledge of his language as well as numerous extragrammatical factors that determine what the listener expects to hear.

In the remainder of this section we briefly review several of the current theories of speech perception. We do not intend this review to be exhaustive in any sense and only those theories that have played a prominent role in recent research will be considered.

Motor Theory of Speech Perception

The basic assumption of the Motor Theory is that "speech is perceived by processes that are also involved in its production (Liberman, et al., 1967)." This view of speech perception was motivated by the observation that the perceiver is also a speaker and it would be unparsimonious to assume that the speaker-listener might use two entirely distinct processes for language perception and production. As we mentioned earlier, one of the central problems in speech perception and perhaps the major reason for postulating a motor theory is the issue of "encoding" or "lack of invariance" between the acoustic signal and phonemic perception. One possible way of resolving the invariance problem was to assume that the same perceptual response to widely differing acoustical patterns arises because the pattern would be produced by the same articulation or underlying motor commands to the articulators. Similarly, different perceptual responses to fairly similar patterns arise from different articulations or different motor commands.

Although the motor theory has occupied an overwhelmingly dominant place in contemporary perceptual theory, the link between empirical data and theory has not been very strong. Indeed, much of its early support was based on recurrent demonstrations of the same experimental outcome using near identical techniques and a very restricted set of stimuli. For example, the primary data used to support the early versions of the Motor Theory came from categorical perception experiments with synthetic stop consonants and steady-state vowels. Although this particular issue has received a great deal of attention over the last few years we will not elaborate on these findings because of space limitations. Most of the current evidence seems to suggest that the categorical perception results are more a function of coding processes in short-term memory than differences in production between various classes of speech sounds (Pisoni, 1973).

If we set aside the categorical perception findings, there seems to be very little direct empirical support for some active mediation of articulatory knowledge or information during perceptual processing. In our opinion, most of the current arguments for Motor Theory rest on parsimony, logic, or faith rather than a firm empirical foundation. The one exception is the recent finding of Cooper (1974) on perceptuo-motor adaptation who showed that listening to repetitions of a /pi/ utterance produced correlated shifts in the subject's VOT values in speech production. Although the magnitude of the effect was small, it is, nevertheless, the first direct support for the hypothesized link between speech perception and production.

From our point of view, the most serious problem for Motor Theory rests on the failure to specify the level of perceptual analysis where articulatory knowledge is employed in recognition. In the information processing model to be described below, it is possible to specify in detail where articulatory information is employed. However, because the model is not completely dependent on the overlap of perceptual and productive processes and only uses articulatory knowledge in a heuristic way, we think that the model is more general than traditional Motor Theory approaches.

Analysis-by-Synthesis

The Analysis-by-Synthesis model proposed by Stevens and his colleagues is much more explicit than the Motor Theory (Stevens, 1960; Stevens & Halle, 1967; Stevens & House, 1972). The basic assumption of the model is similar to the Haskins position; there exist close ties between the processes of speech production and perception, and there are components and operations common to both. The perceptual process begins initially with peripheral processing of the speech signal to yield a description in terms of auditory patterns. In cases where phonetic features are not strongly context-dependent, the auditory pattern will provide a relatively direct mapping of those features during preliminary analysis. The output of preliminary analysis is a rough matrix of phonetic segments and features which is then transferred to the control system. Thus, recognition of some features is thought to take place by relatively direct operations on the acoustic information output from peripheral analysis. However, when there is no invariant attribute to identify the phonetic feature, additional processing is required. An hypothesis concerning the representation of the utterance in terms of phonetic segments and features (i.e., an abstract feature matrix) is constructed. This representation then forms the input to a set of generative rules which produce candidate patterns which are compared against the original patterns. Results of this match are then sent to a control component. The phonetic description is transferred to higher stages of linguistic analysis. Analysis-by-synthesis is simply a more carefully specified version of the Motor Theory except that the invariance problem is presumably resolved at the neuroacoustic level rather than the neuromotor level. The Analysis-by-Synthesis Theory, like the Motor Theory, is quite abstract and there has been little direct experimental evidence to support this view.

Fant's Auditory Theory

Fant's theory is not terribly well developed. His major objection to the "active" motor-type theories is simply that the evidence used to support these theories is not conclusive (Fant, 1967). We agree with this statement. Fant claims that all the arguments brought forward in support of the motor theory would fit just as well into sensory-based theories where the decoding process proceeds without postulating the active mediation of speech motor centers. The basic idea in Fant's model is that the motor and sensory functions become more and more involved as we proceed from the peripheral to central stages of analysis. He assumes that the final destination is the concept of a "message" which comprises brain centers common to both perception and production. According to Fant, there are separate sensory (auditory) and motor (articulatory) branches, although he leaves the possibility of interaction between these two blocks open. Auditory input is first processed by the ear and subject to primary auditory analysis. These incoming auditory signals are then submitted to some kind of direct encoding into distinctive auditory features. Auditory features are then combined together in some as yet unspecified way to form phonemes, syllables, morphemes and words. Although much of Fant's concern has been on continued acoustical investigations of distinctive features his model obviously leaves much to be desired. It is at present much too gross to be tested and the problems of invariance and segmentation which we noted earlier still remain to be resolved.

<u>Bondarko's Stage Theory</u>

In line with a number of recent formulations, Bondarko et al., (1970) have proposed a model that has a series of hierarchically organized stages of perceptual analysis between the acoustic signal and intended message. The following stages are proposed: (a) auditory analysis, (b) phonetic analysis, (c) morphological analysis, (d) syntactic analysis and (e) semantic analysis. Although the model is quite general, it has several properties that are similar to our own views. We consider only the first two stages in their model. The first stage involves auditory analysis of the signal and provides a description of the stimulus in terms of "auditory features." Some type of spectral analysis is assumed to take place at this stage but the presence of possible auditory property detectors which may respond differentially to specific acoustic attributes is also considered. Thus, some type of low-level auditory feature analysis may take place relatively close to the periphery. The second stage involves phonetic analysis. The output of this stage is abstract and may be either an acoustic or articulatory representation of the speech event. This representation is thought to be based on phonetic segments or distinctive features. Provision is also made for the use of various types of normalization routines so that some of the variability from different talkers can be reduced for subsequent processes of recognition. The invariance problem can be resolved in this model by assuming that multiple decisions are made on the basis of outputs from a number of auditory feature detectors. Thus, this aspect of the model is designed not to focus on isolated auditory features (i.e., acoustic cues) but to examine multiple cues simultaneously and in parallel. Segmentation is accomplished at higher levels of analysis. The importance of this approach to speech perception should be emphasized. The model is based, in part, on a number of assumptions about the flow of information within the organism. Memory and coding processes are given strong emphasis and decisions are assumed to take place at many levels of perceptual analysis. The model, in its simplest form, incorporates many of the ideas that have proved fruitful in recent information processing approaches to perception and memory, to which we now turn.

Information Processing in Speech Perception

In recent years the study of speech perception has begun to adopt the aims and methods of information processing models that have been employed in the study of visual and auditory perception. The basic approach views perception as a hierarchically organized sequence of events involving structures for storage and processes for the transformation of information over time. During these stages information is transformed, reduced, elaborated, stored, recovered and used to make various types of decisions.

Three Assumptions

There are three basic assumptions in current information processing models. First, perception is not immediate but is the outcome of distinct operations distributed over time. One goal of information processing models is to attempt to specify in as much detail as possible the operations which occur from the onset of a stimulus to the response of an observer. A flow chart is typically used to represent the various stages of information processing. The second assumption is that there are "capacity limitations" at various stages of analysis during the processing of this information. Because the nervous system cannot maintain all aspects of sensory stimulation there are limits on the capacity to store and process sensory data which require that information be recoded into different forms. One goal of research in information processing is to identify where these capacity limitations occur and the nature of the recoding operations. Finally, the third assumption is that perception necessarily involves various types of memorial processes since recoding and retention of information must occur at all stages of information processing. The study of speech perception, therefore, necessarily entails the study of perceptual memory.

Stages of Processing

Several recent accounts of speech perception have begun to emphasize process and to divide this process into a hierarchy of stages: auditory, phonetic, phonological, syntactic, semantic, etc. (see for example; Studdert-Kennedy, 1974; 1975).

Figure 1 shows some of the processes which are assumed to take place between the initial acoustic signal and its final conceptual representation in the mind of the listener. According to this view, the speech signal undergoes a series of successive transformations whereby information is recoded into more and more abstract forms of representation. On one hand, the stages of processing are thought to be partially successive since spoken language is inherently a temporal phenomena. However, in real-time operation, decisions at various stages must also occur in parallel to permit information from higher more abstract levels of analysis to be employed in processes at lower levels. Figure 2 shows the type of approach we have in mind here. Phonological, syntactic and semantic decisions can all take place simultaneously in order to derive a phonological representation for a sentence. Rather than elaborate on this formulation, it seems sufficient to remark that it is consistent with the types of arguments we made earlier about the role of higher-level linguistic processes in perception. We will return to this again below.

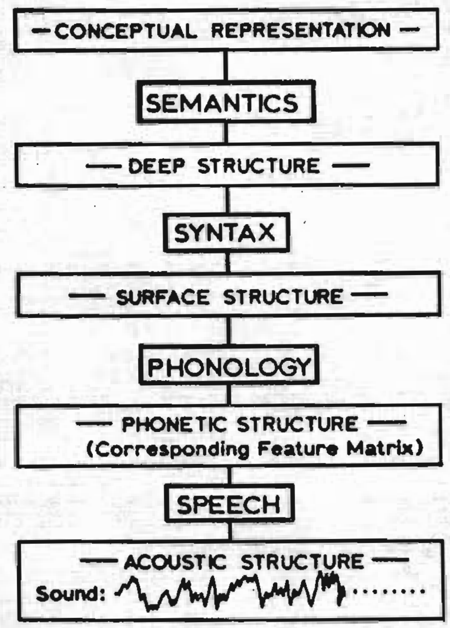


Fig. 1 Stages of Processing in Speech Perception. (Adapted from Liberman, 1970)

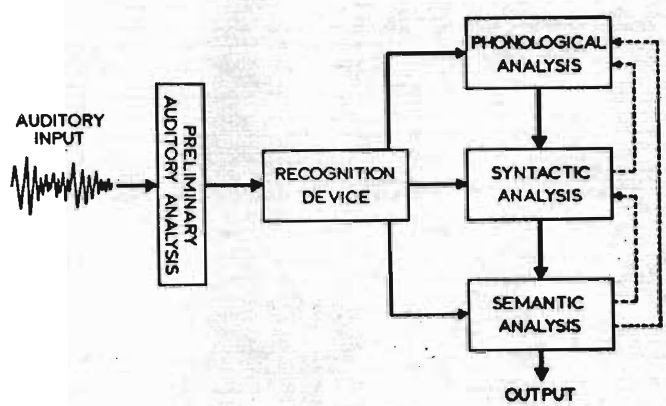


Fig. 2 Functional Organization of Speech Perception Processes.

An Information Processing Model

In this section we briefly sketch the structure of an information processing model, particularly the earliest stages of analysis that deal with recognition. Figure 3 shows a block diagram of the components. Auditory input enters the system

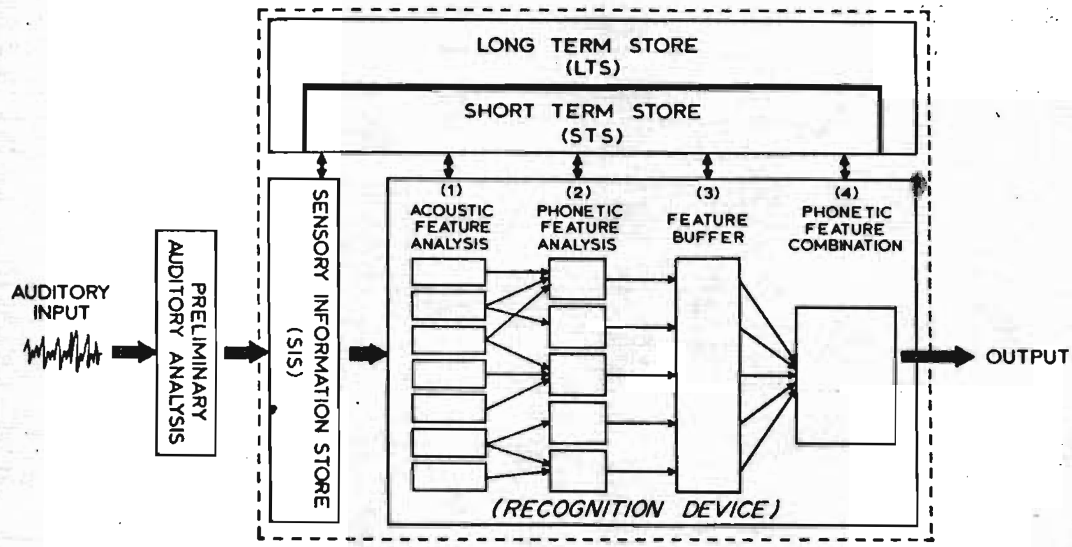


Fig. 3. Components and Organization of the Speech Recognition Process.

and is processed in progressive stages. The output of Preliminary Auditory Analysis is assumed to be some type of spectral display in terms of frequency, time and intensity. Sensory input is processed automatically (i.e., not under the control of the subject) through several levels of analysis without the operation of conscious selective attention. Sensory information is maintained in a relatively gross unanalyzed form in the Sensory Information Store (SIS). Information is further processed by the "recognition device" which is shown as four distinct stages in Figure 3. Information from any or all of these stages of processing is placed in Short-term Store (STS) where the subject can selectively rehearse, encode, or make conscious decisions about it. Information in Long-term Store (LTS) is also assumed to be used in the recognition process. In keeping with recent formulations, STS may be thought of as simply that portion of LTS that is temporarily activated (see Shiffrin, 1975; Bjork, 1975).

Automatic processing by the recognition device is assumed to take place as follows. In Stage 1, Acoustic Feature Analysis, we suggest that auditory features of the speech signal are recognized by a system of individual auditory property detectors (Stevens, 1973). For example, consider the case of a simple CV syllable. We assume that there are detectors which will respond selectively to at least some of the following types of auditory information in the signal: (1) presence or absence of the fundamental frequency from the beginning of the stimulus; (2) abrupt rise in the frequency of the fundamental at the transition from consonant to steady-state vowel; (3) presence or absence of noise from the beginning of the syllable; (4) frequency of the spectral maximum of burst; (5) bandwidth of the noise; (6) duration and intensity of the noise; (7) presence or absence of a rapid change in the spectrum; (8) direction, extent and duration of a change in the spectrum.

In Stage 2, Phonetic Feature Analysis, we propose that a set of decision rules is used to map the multiple auditory features extracted from Stage 1 into phonetic distinctive features. These decision rules can be thought of as having knowledge of the articulatory constraints on speech production in as much as they affect the nature of the acoustic signal although it is not necessary for the model. We would also rather not assume that a "phonetic processor" exists as a distinct physiological mechanism. Instead, we would prefer to describe its function simply in terms of decision rules at the present time. We assume that decisions about features and segments within a syllable, for example, are made on the basis of information distributed across the whole syllable. As we remarked earlier, this does not commit us to arguing for some special status for the syllable.

These features are subsequently maintained in Stage 3, the Feature Buffer which is a storage mechanism to maintain decisions about the feature composition of a particular syllable. The reason for a feature buffer is two-fold. First, not all phonetic features are assumed to be processed (i.e., recognized) at the same rate. Secondly, some type of storage is needed to preserve feature information more or less independently for subsequent stages of phonological analysis. Feature information is then used in Stage 4, Phonetic Feature Combination, where individual features can be recombined to form discrete phonetic segments. Information about the feature specification of these phonetic segments is in the form of a rough distinctive feature matrix which is then made available to higher levels of processing for phonological and syntactic analysis.

One may ask at this point whether the approach we have advocated here has any relevance to the perception of connected speech in dynamic context. We think it does for several reasons. First, although much of the early research on speech perception used isolated CV syllables, these studies were aimed at specifying the minimal and sufficient cues for perception of a specific phonetic distinction. Thus, these cues may be thought of as representing the lower bound on the information contained in the waveform which is necessary for a decision assuming no higher level information was available. We would anticipate better or more accurate performance on a given task if higher-level grammatical information were available. Second, we feel that the characterization of these minimal acoustic cues that were found in isolation must be in part correct. Various synthesis by rule programs for generating connected speech have been written which formalize the use of these isolated cues in a systematic way and which produce fairly intelligible speech

(Mattingly, 1968; Kuhn, 1973).

To summarize briefly, low-level analyses are assumed to be relatively automatic processes of the perceptual system. Higher-level syntactic and semantic processes are assumed to operate in a heuristic way in perception. One implication of this approach is that the perceptual system makes available to higher stages of processing only a "crude" analysis of the input. We assume that there is considerable ambiguity at these early stages of analysis and only very tentative decisions are made about the phonological structure of the input. Before proceeding, several remarks should be made about the role of the memory system in information processing approaches.

Sensory, Short-term and Long-term Memory

A basic concept of information processing analyses is the notion of an iconic or preperceptual store (Sperling, 1960; Massaro, 1972). This is thought of as a very temporary storage medium which preserves all of the stimulus information in a literal or veridical form for several hundred milliseconds so that pattern recognition processes can operate on the input. During this time period, the sensory information is converted into more persistent and abstract memory codes for representation in short-term storage. Short-term store (STS) or "working memory" is thought to have a very limited capacity from which information is rapidly lost unless active control processes are operating such as rehearsal. The contents of STS may be thought of as a person's conscious memory. Long-term store (LTS) on the other hand, is assumed to be a permanent repository for information with a relatively unlimited capacity. However, the availability of information in LTS is constrained by various search and retrieval operations (Shiffrin & Atkinson, 1969). LTS receives information from STS. Rehearsal of information in STS is considered to be an "active" process which regenerates decaying memory traces and also causes information in STS to be entered into LTS. Recognition is assumed to be a process whereby the sensory input or some derived version of it "makes contact" with a stored representation in long-term memory or some type of representation that has been constructed or synthesized by rules in long-term memory. The information present in short-term memory is thought to consist of a combination of information from both the sensory input and information currently activated from long-term memory. Information that is in the sensory information store is not simply transferred to STS but is "recoded" while still being maintained at the earlier stage of processing. It is generally assumed that the earliest stages of the recognition process occur automatically and without conscious control by the subject. A good part of the information from the earliest stages of processing is lost by decay and only a relatively abstract representation of the input is maintained in STS. Although SIS, STS, and LTS are typically referred to as "structural" properties of the memory system, no claim is made as to whether they correspond to distinct neurological systems. Indeed, recent formulations simply represent different memory stores as different phases of activation of a single neurological system (see Shiffrin, 1975).

Coding Processes

In this section we briefly consider the nature of coding processes in speech perception. One of the problems with current information processing models is that they say very little about the exact form in which information is represented at various stages of processing or the types of recoding operations that are assumed to take place. Thus, for example, no claims are made with regard to the physiological reality of any coding scheme. For the most part, this is an arbitrary question. However, with regard to speech perception we think the situation is somewhat different. Speech sounds are not an arbitrary set of acoustic signals in the environment. That is, they are not symbols but rather signs and they form a class of natural categories in much the same way as other sign stimuli in the environment (Mattingly, 1972). It seems reasonable to suppose then that the non-arbitrariness of speech sounds derives in part from the production system and the way in which they are produced. Thus, some aspects of the coding processes in memory may ultimately be linked to articulation. Although this has been a topic of considerable debate, its resolution does not seem to be very close in hand. It is sufficient to note here that the properties of memory codes for speech sounds and the nature of recoding processes would appear to be central theoretical issues in speech perception. It seems to us that contemporary information processing models have failed to come to grips with these types of issues.

Higher Level Processing

In this section we consider some aspects of higher-level analyses in speech perception. Three stages of processing will be distinguished; phonological, syntactic and semantic. These are not conceived of as separate entities but rather as labels for the type of encoding and form of information from long-term memory that is used in processing. Together, these stages form an integrated system whose net product is comprehension of the acoustic input. However, before treating the integration of these three processes, they should be examined separately.

Phonological processing consists of integrating segmental and prosodic information from the recognition device. The result of this analysis is a tentative string of phonetic segments, including some word and morpheme boundaries. This tentative string is fed to the syntactic processor, along with prosodic information directly from the recognition device.

The purpose of the syntactic processor is to construct a preliminary parse tree of the input. Traditionally, this process has been conceived of as the application of a sequence of rules, the net result of which is a parse tree or parse graph. We propose that the theoretical syntactic processor follows real-time limitations on memory and processing capacity that are known to occur in humans (see Fodor, Bever & Garrett, 1974). One approach to this appears to be the augmented transition network (ATN) proposed recently by Woods (1970, 1973). The ATN is neutral with regard to the type and number of rules used in syntactic processing. The choice of rules can follow transformational grammar (Chomsky, 1965) or systemic grammar (Halliday, 1961; 1967) or any other system of rules. What is important is that the ATN makes specific use of memory in syntactic processing and provides a processor for applying syntactic rules, whatever their origin (see Kaplan, 1972).

The semantic processor works on information from the syntactic processor in conjunction with prosodic information from the recognition device. The semantic processor does not have to wait for the construction of a full parse tree since it is intimately involved with all aspects of the syntactic processing. We assume that the semantic processor draws upon knowledge of context, specific lexical meaning, and other forms of semantic information stored in long-term memory. This is not the place to propose a specific model of semantic information processing. For the present, we simply note that there are more than a few such theories available (Anderson & Bower, 1973; Bierwisch, 1971; Kintsch, 1974; Quillian, 1968; Rumelhart, Lindsay & Norman, 1972; Schank, 1972; Winograd, 1972). However, the psychological evidence distinguishing these theories is still inconclusive.

Up to this point, we have not mentioned lexical insertion. In this present framework, it is not necessary to conceive of a specific stage of lexical insertion. Instead, each of the three stages may look-up and insert correct lexical entries as necessary. Lexical search and insertion is a property of higher level analysis rather than a specific stage of processing. Another point about the organization of these three processes should be noted. These are not sequential stages, one following the other. They operate basically in parallel. Phonological analysis operates on smaller segments of the output from the recognition device than the syntactic or semantic processors. As a result, some aspects of its processing are faster than those of syntax and semantics. The net result is that, while information is processed in parallel, it gradually converges upon more abstract processors.

Two final points should be mentioned about higher level processing. The first of these concerns feedback from higher levels of processing to lower levels and the second concerns the existence of a central processor or "executive" to oversee the whole process. As shown in Figure 2, there are feedback routes from semantic and syntactic levels. However, there is no direct feedback to the recognition device. As we mentioned earlier, the recognition device is assumed to operate automatically and it provides only a crude analysis of the acoustic input. However, the higher levels of processing, including phonological, syntactic and semantic, must be capable of exchanging information and revising previous decisions.

In the present conception, there is no central executive that oversees the operations of the total system. Rather, each level operates on information as it becomes available. The semantic processor has the final responsibility of accepting or rejecting the results of previous analysis. Since the goal of processing an acoustic input is to recover the speaker's intended message, the speech perception process must ultimately be directed by a semantic system.

Concluding Remarks

We began this paper by considering a number of contemporary theoretical issues in speech perception. Our main purpose in reviewing these problems was to show how they place constraints on the possible class of theories in speech perception. We then described several of the prominent theories and attempted to point out some of their strengths and weaknesses. We tried to show that most of the current theoretical efforts have failed to consider seriously the role of higher-level linguistic processes in perception, particularly syntax and semantics and the influence of prosodic factors. In our view, it seems that most of these theories have tacitly endorsed the traditional taxonomic assumption that an autonomous level of phonological structure can be derived from an input without the contribution of higher-order linguistic information.

We also described some aspects of the information processing approach that underlies our own thinking about speech perception and sketched the type of model that we have in mind. One of the points that we emphasized was the role of coding processes in speech perception, especially the types of recoding operations and the properties of memory codes at different stages of analysis. In the final section, we briefly examined some aspects of higher-level analyses in speech perception and how these might form an integrated system. We tried to emphasize that the earliest stages of perceptual processing provide only a rough and tentative analysis of the acoustic input while larger units are constructed and subsequently made available for syntactic and semantic analysis.

As we mentioned earlier, one of our goals in this paper was to focus on general questions in speech perception rather than particular issues in narrowly defined experimental contexts. One reason for doing this was our feeling that work in speech perception might eventually have implications for deciding theoretical questions in linguistics, particularly phonetic and phonological issues (see Lieberman, 1970). Thus, for example, issues dealing with the possible inventory of universal phonetic features or various types of phonological processes might be motivated on perceptual grounds as opposed to strictly formal linguistic criteria (see for example, Bever and Langendoen, 1971). Indeed, several investigators have recently argued that phonetics and phonology should play a more predictive role in linguistic theory rather than its current interpretative function (Lindblom, 1971; Ohala, 1971).

However, these sorts of goals may never be reached until the theoretical work in speech perception becomes formalized within some unified framework. We hope that some of the issues we have considered in this paper will stimulate further theoretical discussion in the future.

Acknowledgments

Preparation of this paper was supported in part by USPHS grants MH-24027-02, NS-12179-01 and MH-11219-06 to Indiana University.

References

Anderson, J. R. & Bower, G. H. Human associative memory. Washington, D. C.: V. H. Winston, 1973.

Bever, T. G. & Langendoen, D. T. A dynamic model of the evolution of language. Linguistic Inquiry, 1971, 2, 433-463.

Bierwisch, M. On classifying semantic features. In D. D. Steinberg & L. A. Jakobovits (Eds.), Semantics. Cambridge, G. B.: Cambridge University Press, 1971.

Bjork, R. A. Short-term storage: the ordered output of a central processor. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.) Cognitive theory: Volume I. Potomac, Maryland: Erlbaum Associates, 1975.

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

Chomsky, N. Aspects of a theory of syntax. Cambridge, Mass.: MIT Press, 1965.

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

Cooper, W. E. Perceptuomotor adaptation to a speech feature. Perception & Psychophysics, 1974, 16, 2, 229-234.

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

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

Fant, G. Auditory patterns of speech. In W. Wathen-Dunn (Ed.) Models for the perception of speech and visual form. Cambridge, Mass.: MIT Press, 1967.

Flanagan, J. L. Speech analysis, synthesis and perception. (Second Edition) New York: Academic Press, 1972.

Fodor, J. A., Bever, T. G. & Garrett, M. F. The psychology of language. New York: McGraw-Hill, 1974.

Halliday, M. A. K. Categories of the theory of grammar. Word, 1961, 17, 241-292.

Halliday, M. A. K. Notes on transitivity and theme in English. (Part I). Journal of Linguistics, 1967, 1, 37-81.

Kaplan, R. M. Augmented transition networks as psychological models of sentence comprehension. Artificial Intelligence, 1972, 3, 77-100.

Kintsch, W. The representation of meaning in memory. Hillsdale, N. J.: Erlbaum Associates, 1974.

Kuhn, G. M. A two-pass procedure for synthesis-by-rule. Journal of the Acoustical Society of America, 1973, 54, 339 (A).

Lea, W. A. An approach to syntactic recognition without phonemics. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, No. 3, 249-258.

Liberman, A. M. The grammars of speech and language. Cognitive Psychology, 1970, 1, 301-323.

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

Lieberman, P. Towards a unified phonetic theory. Linguistic Inquiry, 1970, 1, 3, 307-322.

Lindblom, B. E. F. Numerical models in the study of speech production and speech perception: some phonological implications. Papers from the Institute of Linguistics, University of Stockholm, October, 1971, Publication No. 7.

Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. Psychological Review, 1972, 79, 2, 124-145.

Mattingly, I. G. Synthesis by rule of general American English. Supplement to Haskins Laboratories Status Report on Speech Research, April, 1968.

Mattingly, I. G. Speech cues and sign stimuli. American Scientist, 1972, 60, 3, 327-337.

Ohala, J. J. Physical models in phonology. Paper presented at the VII International Congress of Phonetic Sciences, Montreal, 1971.

Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. Perception & Psychophysics, 1973, 13, 2, 253-260.

Quillian, M. R. Semantic memory. In M. Minsky (Ed.), Semantic information processing. Cambridge, Mass.: MIT Press, 1968.

Rumelhart, D. E., Lindsay, P. H. & Norman, D. A. A process model for long-term memory. In E. Tulving & W. Donaldson (Eds.), Organization of memory. New York: Academic Press, 1972.

Schank, R. C. Conceptual dependency: a theory of natural language understanding. Cognitive Psychology, 1972, 3, 552-631.

Shiffrin, R. M. Short-term store: the basis for a memory system. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), Cognitive theory: Volume 1. Potomac, Maryland: Erlbaum Associates, 1975.

Shiffrin, R. M. & Atkinson, R. C. Storage and retrieval processes in long-term memory. Psychological Review, 1969, 76, 179-193.

Shockey, L. and Reddy, R. Quantitative analysis of speech perception: results from transcription of connected speech from unfamiliar languages. Paper presented at the Speech Communication Seminar, Stockholm, Sweden, August 1-3, 1974.

Sperling, G. The information available in brief visual presentations. Psychological Monograph, 1960, 74, (11, Whole No. 498).

Stevens, K. N. Toward a model for speech recognition. Journal of the Acoustical Society of America. 1960, 32, 47-55.

Stevens, K. N. The quantal nature of speech: evidence from articulatory-acoustic data. In E. E. David, Jr. and P. B. Denes (Eds.), Human communication: A unified view. New York: McGraw-Hill, 1972.

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

Stevens, K. N. & House, A. S. Speech perception. In J. Tobias (Ed.), Foundations of modern auditory theory: Volume II. New York: Academic Press, 1972, 1-62.

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

Studdert-Kennedy, M. Speech perception. In Lass, N. J. (Ed.), Contemporary issues in experimental phonetics, Springfield, Illinois: C. C. Thomas, 1975.

Svensson, S. G. Prosody and grammar in speech perception. Monographs from the Institute of Linguistics, University of Stockholm, 1974, No. 2, Institute of Linguistics, Stockholm, Sweden.

Woods, W. A. Transition network grammars for natural language analysis. Communications of the ACM, 1970, 13, 10, 591-606.

Woods, W. A. An experimental parsing system for transition network grammars. In R. Rustin (Ed.), Natural language processing. New York: Algorithmics Press, 1973.

Winograd, T. Understanding natural language. New York: Academic Press, 1972.

Simple and Contingent Adaptation Effects in Speech Perception

David B. Pisoni, James R. Sawusch, and Frank T. Adams

Department of Psychology

Indiana University

Bloomington, Indiana 47401

Abstract

Experiments on selective adaptation have shown that the locus of the phonetic boundary between two segments shifts after repetitive listening to an adapting stimulus. In the present paper we report three experiments that employed a contingent adaptation procedure to study the possible interactions between auditory and phonetic processes in the perception of stop consonants. In the contingent procedure, two alternating stimuli are presented as adaptors and their combined effects on perception can be studied. The first experiment examined the voicing feature and the results replicated the previous findings of Cooper who showed that adaptation of voicing takes place in a vowel dependent manner. The second and third experiments extended the contingent procedure to an analysis of the place feature. In the second experiment simple and contingent adaptation were studied in order to show that the simple adaptation effects obtained with single adaptors can be canceled by the presentation of a contingent sequence containing identical vowels. Finally, in the last experiment the vowel context of the contingent adaptors and test series was manipulated. Differential adaptation effects were observed for two separate place series. The direction of the shifts in each series was governed by the vowel context of the adapting syllables suggesting that selective adaptation of the place feature is also carried out in a vowel dependent manner. The opposing shifts obtained for both voicing and place features in the contingent adaptation experiments suggest that a large component of the selective adaptation effect takes place at an auditory stage of perceptual processing prior to phonetic analysis.

Simple and Contingent Adaptation Effects in Speech Perception[1]

David B. Pisoni, James R. Sawusch, and Frank T. Adams

Indiana University

Bloomington, Indiana

Several recent studies have reported evidence for the existence of specific feature detector mechanisms in speech perception (Eimas & Corbit, 1973; Eimas, Cooper & Corbit, 1973; Cooper, 1974a). In these studies a selective adaptation paradigm has been used in which repetitive presentation of a particular adapting stimulus alters the perception of a sequence of synthetic test stimuli. For example, Eimas and Corbit (1973) showed that adaptation with the syllable [ba] caused the locus of the phonetic boundary between [ba] and [p^ha] to shift toward the [ba] end of the continuum. Stimuli near the boundary which were identified as [ba] when the listener was in an unadapted state were subsequently labeled as [p^ha] after adaptation with [ba]. Similar findings were obtained when [p^ha] was used as an adaptor; the locus of the phonetic boundary shifted toward the [p^ha] end of the stimulus continuum. Eimas and Corbit also showed that these results were not specific to the syllables or phonemes in the test series but rather were due to the presence of a component or feature of the consonants. For example, adaptation with the voiceless bilabial stop [p^h] produced approximately equivalent effects on the identification functions for a series of alveolar stop consonants (i.e., [d] and [t^h]) as it did for a series of bilabial stops ([b] and [p^h]). In both cases, the locus of the phonetic boundary shifted toward the voiceless end of the stimulus continuum.

Eimas and Corbit interpreted these initial findings as support for the hypothesis that the selective adaptation effects have a phonetic rather than auditory basis. They argued that selective adaptation was due to the fatiguing of detectors that are specialized for processing abstract linguistic features as opposed to simply processing the auditory information that may underlie these distinctions. According to Eimas and his colleagues (see also Eimas, Cooper, & Corbit, 1973) the perception of the voicing feature involves two distinct types of feature detectors that are organized as opponent pairs, a voiced detector (+V) and a voiceless detector (-V). Each detector is selectively tuned to a restricted range of partially overlapping VOT values. Repeated presentation of a stimulus containing a specific feature is assumed to fatigue that detector and therefore reduce its sensitivity. Thus, after adaptation the opponent or unadapted detector will provide more information to the decision process than the adapted detector and will ultimately result in a shift in the locus of the phonetic boundary in identification (see Cooper, 1975a, for a review).

The selective adaptation studies have generated a great deal of research in speech perception over the last few years and have provided numerous insights into some aspects of the organization of the speech perception system. However, several important questions still have not been adequately resolved. For example, although Eimas and his colleagues have argued that the adaptation results have their locus at the phonetic feature stage of perceptual analysis, several other accounts have also been proposed recently. We briefly consider two of these, an auditory account and a combined, multi-component view of adaptation. The auditory account attributes the adaptation effects to interaction at the auditory stage of perceptual analysis prior to phonetic categorization. Results showing adaptation when the adaptor and test series have spectral components in common would be consistent with this interpretation. On the other hand, the multi-component view assumes that adaptation takes place at several stages of analysis and that both auditory and phonetic components contribute to the observed effects. Experiments showing adaptation effects when the adaptor and test series differ acoustically and have no simple invariant acoustic relation between them would be consistent with this interpretation as well as a phonetic account.

Auditory Account of Adaptation

The auditory explanation has received strong support in a number of studies that have examined the feature of place of articulation in stop consonants. For example, Bailey (1973) showed that the adaptor and test series must have common spectral components for adaptation to take place. Ades (1974) has shown that consonants in final position in a syllable will not produce adaptation on the same consonants in initial position. Thus, at least in this experiment, the adaptation could not be due to fatiguing of detectors at a phonetic level since phonemes are assumed to be identical linguistically in any position. Recently, we have obtained adaptation effects when the adapting stimuli were speechlike VC syllables that were spectrally similar to the CV test series (Pisoni & Tash, 1975). Finally, Cooper and Blumstein (1974) have also reported strong evidence in favor of an auditory account when the adaptors shared the same place features as the test series but differed in their manner of articulation.

There is also evidence for strong auditory effects in the analysis of the voicing feature. For example, Cooper (1974b) has shown that the analysis of voicing is carried out in a vowel dependent manner. In this study,

Cooper used a contingent adaptation paradigm in which listeners were presented with an alternating sequence of two different adaptors, [da] and [t^hi] and then tested along two separate continua, a [ba] to [p^ha] series and a [bi] to [p^hi] series. As we shall see below, the results of this study are of interest because if the adaptation effects occurred exclusively at a phonetic feature level, independently of vowel context, no adaptation would be anticipated since the opponent effects of the voicing features in the adaptors [d] and [t^h] should cancel each other out. However, Cooper found just the opposite results, namely, a different effect on each of the two identification functions. Thus, for the [ba] to [p^ha] series the boundary shifted toward the voiced end of the stimulus continuum whereas for the [bi] to [p^hi] series it shifted toward the voiceless end. The adaptation effects for each series were a function of the vowels in the adapting syllables rather than the phonetic features in the consonants. Although Cooper's results were small, they indicate that adaptation of voicing occurs at an auditory level which appears to be partially dependent on vowel context. In another study, Cooper (1974c) has also reported some evidence for a more complex integrative mechanism in the processing of voicing information. In this experiment, the VOT cue for voicing was placed in opposition to another cue for voicing, the presence or absence of formant transitions following voice onset (Stevens & Klatt, 1974). By using variations in both voicing cues, Cooper was able to obtain a differential adaptation effect that could not be attributed exclusively to either the VOT cue or the formant transition cue. These results appear to be due to either an integrative acoustic cue analyzer at the auditory stage that processes voicing cues or alternatively to a phonetic mechanism that responds to voicing information regardless of the specific acoustic properties of the cues. In any case, there is evidence for adaptation effects occurring at one or possibly two stages of auditory analysis.

Auditory and Phonetic Account

Although the evidence in favor of an auditory interpretation to the selective adaptation effects is strong, several findings seem to point to the presence of an additional phonetic component to adaptation. For example, Cooper (1974a) has proposed that selective adaptation is a multi-component process that operates at several levels of perceptual analysis in the speech processing system. According to Cooper, one component of adaptation may be attributable to fatiguing of mechanisms at an auditory level whereas another component may reflect the contribution of adaptation at a more abstract phonetic feature level. The observed magnitude of the adaptation effect would therefore be due to two distinct components, auditory and phonetic. In support of this multi-component view, Cooper (1974a) reported several experiments on the place feature that used adaptors which contained different acoustic information but identical phonetic information. In one experiment, Cooper manipulated the vowel context of the adapting stimulus while keeping the consonant segment the same. In another experiment, Cooper used a cross-consonant procedure in which the adaptor had the same place feature but differed from the test series in voicing. In each case, the magnitude of the shift obtained with a crossed-series adaptor was smaller than the shift obtained with adaptors drawn from the same test series. Cooper attributed these effects to adaptation at a phonetic feature level since acoustic commonality could not be easily found by an examination of some simple invariant property of the stimuli.

More recently, Diehl (1975) has also studied adaptation of the place feature in a [bɛ] to [dɛ] continuum. In this study, however, the adapting stimuli were selected so that although they shared identical phonetic features with the test series, they had little in common with them acoustically. The adaptors Diehl used were the voiceless stops [pɛ] and [tɛ] in which the place feature was cued by a burst rather than formant transitions. The adaptation results showed a significant shift in the predicted direction for six subjects with the burst cued [tε] but only four of the six subjects showed the expected shift for the burst cued [pε]. Diehl argued that since the burst cued adaptors had virtually nothing in common acoustically with the formant transition cues specifying place in the test series, at least some component of the adaptation effects must occur at the phonetic feature level.

Cooper (1975) has also reported a number of studies showing perceptuo-motor adaptation effects. After repetitive listening to an adaptor, Cooper has observed small but systematic changes in the acoustic properties of utterances produced by the listener. Because of space limitations we will not review the details of these studies. However, we simply note here that it would be extremely difficult to maintain a strict auditory interpretation of the selective adaptation effects in light of these findings.

In almost all of the perceptual adaptation experiments a simple adaptation procedure has been employed in which a single stimulus is used as an adaptor. As we noted earlier, the only exception is Cooper's (1974b) contingent analysis of voicing. Although the results obtained with the simple adaptation procedure have provided important information about some of the types of factors that produce adaptation in speech perception, it has not been possible to study the interaction of auditory and phonetic features directly. In this regard, Cooper's contingent procedure is of interest since it provided a fairly direct way to study the separate contribution of auditory and phonetic components to adaptation and the possible rivalry between the two. In the present paper, we report three perceptual experiments that employed a contingent adaptation procedure. The first experiment consists of a replication of Cooper's original study on the voicing feature. We felt that a replication was necessary before proceeding to the other experiments because the results were relatively small when compared to those obtained in simple adaptation. Moreover, they may have been due simply to an order effect present during the experiment.

In the second and third experiments we applied the contingent procedure to study place of articulation in stop consonants. It is well known that the acoustic cues for place of articulation in consonants show a great deal of context conditioned variability as a function of the preceding and following vowel context (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). Although the cross-consonant and cross-vowel experiments on the place feature with simple adaptors have suggested the possibility of adaptation effects at a phonetic level, it is also quite reasonable to suppose that adaptation of place like voicing is also not carried out independently of vowel context. Thus, if we found vowel contingent effects with the place feature, we would have additional support for a strong auditory component to selective adaptation. Accordingly, in these experiments we were interested in the degree to which adaptation of place is a function of vowel context of the adaptors and how auditory and phonetic features interact when they are placed in opposition in a contingent adaptation procedure.

Experiment I

In this experiment we replicate the results of Cooper's (1974b) contingent analysis of voicing with a slightly different experimental procedure. Although Cooper found a small adaptation effect in which the adaptor influenced only the test series containing the same vowel, this result might have been due to the order in which the test stimuli were presented to his subjects. Cooper tested only one stimulus series on each of two consecutive days with the same adapting sequence, [da] and [t^hi]. Thus, the shifts obtained might have been due simply to the subject's utilization of a strategy that favors a particular test series. In the present study, we used both test series, [ba]-[p^ha] and [bi]-[p^hi], on each day in order to eliminate the possibility that Cooper's contingent results were due to an order effect or subject bias. Since we planned to use the contingent procedure in other experiments we thought that a replication was in order.

Method

Subjects. Twenty paid volunteers served as Ss. All were right-handed native speakers of English with no known history of a hearing or speech disorder. Ss responded to an advertisement in the student newspaper and were paid at the rate of $2.00 per hour. Ss were divided into two groups of ten Ss each.

Stimuli. The stimuli were three-formant synthetic speech sounds that were prepared on the parallel resonance synthesizer at Haskins Laboratories. All stimuli were recorded on magnetic tape for later playback. The test stimuli consisted of two series of thirteen CV syllables that varied in voice onset time (VOT). One series ranged perceptually from [ba] to [p^ha] whereas the other ranged from [bi] to [p^hi]. In each series, the stimuli varied from one another in equal steps of 5 msec from 0 msec VOT to +60 msec VOT. Formant transitions were 50 msec in duration and were appropriate for the bilabial consonants [b] and [p^h]. Except for differences in the formant frequencies and amplitudes of the vowels, all other acoustic parameters were equivalent for the two test series. The duration of each stimulus was 300 msec.

In addition to the two test series, four adapting stimuli were also prepared, [da], [di], [t^ha], and [t^hi]. The VOT values for each of these stimuli were identical to the end point stimuli from each test series, 0 msec VOT for the voiced stops and +60 msec for the voiceless stops. These stimuli were also 300 msec in duration. Except for the differences in formant transitions, all acoustic parameters were identical to the test stimuli containing the same vowel.

Procedure. The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the earphones equivalent to 80 dB SPL for a steady-state calibration vowel [a]. Measurements were made on a VTVM (Hewlett Packard Model 1051) before the presentation of each tape. Ss were tested in groups of five in a small testing room used for speech perception experiments.

The experiment was conducted on two consecutive days. At the beginning of each day, Ss received two identification tapes, one for each test series. Each tape contained a different randomization of ten replications of each of the 13 test stimuli in a series. Ss were told that they would hear synthetic speech sounds approximating the syllables [ba], [p^ha], [bi], or [p^hi] and they were to respond by writing a "B" or a "P" on the prepared response sheets. After two days, each S contributed a total of 20 responses to each of the thirteen stimuli in each test series.

Immediately following the identification tapes an adaptation test was presented. One group of Ss received the adapting syllables [di] and [t^ha] whereas the other group received [da] and [t^hi]. Each syllable was presented three times in succession before a change over to the other syllable. The interstimulus interval for repetitions was 300 msec. After each minute of adaptation, a sequence of nine stimuli was presented in random order for identification. On half the trials the nine test stimuli were drawn from stimuli 1 through 9 on the stimulus continuum, whereas on the other half they were drawn from stimuli 5 through 13. The test series also alternated from trial to trial so that on each day half of the stimuli came from the [ba]-[p^ha] series and the other half came from the [bi]-[p^hi] series. Two different tapes were prepared for each adapting sequence so that a different tape could be used on each day. By the end of the experiment, each S provided 18 responses to the five middle stimuli in each test series (5, 6, 7, 8, 9) and nine responses to the four end stimuli in each series (1, 2, 3, 4, and 10, 11, 12, 13).

Results and Discussion

The individual subject's data for both adaptation conditions are presented in Table 1 in the form of phonetic boundary loci. Phonetic boundaries

----

Insert Table 1 about here

----

were determined by means of a computer program that performed a linear interpolation between the two closest points on either side of the crossover in the identification functions. The group data are also shown in Figure 1.

Insert Figure 1 about here

The results are quite similar to those reported by Cooper (1974a). In both conditions, the voicing boundary of the [a] vowel series occurs at a smaller VOT value than the boundary of the [i] vowel series. The baseline boundaries for the two test series differ significantly from each other $t\ (19) = 6.90, p < .001$. Of more interest, however, are the results from the contingent adaptation conditions. As shown in Figure 1, these also replicate Cooper's findings. Adaptation with [da]-[t^hi] or [di]-[t^ha] produced shifts in the loci of the phonetic boundaries in both test series in the direction of the vowel context of the adaptor. Thus, in the [da]-[t^hi] condition, the [a] test series shifted toward the voiced end of the stimulus continuum whereas the [i] series shifted in the opposite direction toward the voiceless end. Analogous results were obtained in the [di]-[t^ha] condition. Both of the shifts in the [a] series stimuli were highly significant by t tests, $p < .005$ for the [da]-[t^hi] group and $p < .001$ for the [di]-[t^ha] group. However, the shifts in the [i] stimulus series did not reach significance. These findings are comparable to Cooper's, who also found a much smaller shift in the [i] series stimuli although his result was just marginally significant.

The results of this experiment replicate the basic findings reported by Cooper. We also conclude with him that there is some evidence that the analysis of the voicing feature is carried out in a vowel dependent manner. At least some component of the selective adaptation effect appears to take place at an auditory level of perceptual analysis. This must be the case, because if the locus of adaptation were exclusively at a phonetic level where features are presumably independent of vowel context, the combined effects of the voicing features in the two contingent adaptors should have cancelled each other out. Instead, the phonetic boundaries shifted in opposite directions as a function of the vowel. Although there is evidence for a dependency on the vowel, we should point out that the observed contingent effects in Cooper's experiment and the present study are not nearly as large as those found with simple adaptors in the earlier adaptation studies on voicing. This suggests the possibility that, in fact, some phonetic adaptation may have taken place and the shifts we found could be due to the combined operation of both phonetic and auditory factors. Thus, at the phonetic level, the differential effects of the voicing features may well have been cancelled out and therefore contributed only a negligible amount to the observed adaptation effect. Moreover, if we assume a somewhat larger contribution to adaptation at an auditory level which most of the literature seems to suggest, we might not only expect to find opposite shifts in the phonetic boundaries of the two test series as a function of vowel context but also somewhat smaller shifts than those obtained with simple adaptors. In the case of a single adaptor, adaptation may take place at both auditory and phonetic levels, whereas in the contingent situation, the contribution of phonetic information may be canceled while the auditory information remains the same. It turns out that this explanation would be consistent with Cooper's (1974b) multi-component view of adaptation which assumes adaptation effects at both auditory and phonetic stages. Thus, having replicated the basic findings of Cooper's contingent adaptation of voicing, we turn to our main interest which is an examination of the place feature in stop consonants.

Experiment II

In this experiment we applied the contingent adaptation procedure to study the feature of place of articulation. It is well established that the place feature shows a great deal of context conditioned variability and therefore it was of some interest to determine whether selective adaptation of place would also show a dependency on vowel context like that observed for voicing. Moreover, since the magnitude of the effects obtained with simple adaptors were somewhat larger for place than voicing, this provided another opportunity to study the interaction of auditory and phonetic features in selective adaptation. In the present experiment, we examined both simple and contingent adaptation effects for a place series when the vowels in the adaptors were the same as the test series. If the contingent effects are dependent on vowel context, we should be able to eliminate any contingent effect by using the same vowel in both adaptors and test series. In the third experiment, we performed the reverse manipulation by specifically varying the vowels in the adaptors in anticipation of a differential effect on each test series.

Method

Subjects. The Ss were 24 undergraduate students enrolled in introductory psychology. They received 2 hours of credit toward a course requirement and met the same requirements as those Ss used in Experiment I. The Ss were divided in four groups of six Ss each.

Stimuli. The test stimuli were also generated on the speech synthesizer at Haskins Laboratories and consisted of seven three-formant patterns that varied in place of articulation between [ba] and [da]. The stimuli were 300 msec in duration and differed from each other in the direction and starting frequencies of the second- and third-formant transitions. These values are presented in Table 2.

Insert Table 2 about here

Four adapting stimuli were also prepared, [ba], [da], [p^ha], and [t^ha]. The voiced stops were selected from the endpoints of the stimulus series. The voiceless stops were identical to the voiced endpoints on all acoustic parameters except VOT, which was set at +60 msec.

Procedure. The experiment was conducted on two separate days. At the beginning of each day, Ss received two 70-item identification tests. Each test contained ten presentations of each of the seven test stimuli in a random order. After baseline, Ss were presented with an adaptation test. Simple adaptation was presented on the first day followed by contingent adaptation on the second day. On the first day, each of the four groups received a different adaptor ([ba], [da], [p^ha], or [t^ha]) which was presented repetitively for 1½ minutes. Ss then received a randomized sequence of the five middle stimuli from the test series for identification. There were 2 sec of silence between adaptation and the presentation of the test series, 3 sec between test stimuli, and 5 sec between adaptation trials. Ss received two presentations of the adaptation tape which contained ten different randomizations of the five test stimuli for a total of 50 trials. On the second day, each group received a contingent adaptation test consisting of either a [ba] and [da] or [p^{h}a] and [t^{h}a] adapting sequence. The two groups that received the simple voiced adaptors, [ba] or [da], on the first day, received the voiced adaptors [ba] and [da] on the second day. The analogous procedure was followed with the voiceless group which received the [p^{h}a] and [t^{h}a] sequence on the second day. The contingent adaptors were presented in an alternating sequence fc 1½ minutes. Each syllable was presented three times in succession before a change over to the other syllable. The interstimulus interval was 300 msec. As in the simple adaptation test, after each adaptation sequence, a different randomization of the five middle stimuli from the test series was presented for identification. Ss received two repetitions of the same adaptation tape. By the end of the experiment each S provided a total of 20 responses to each of the five test stimuli. All other test conditions were the same as those used in the previous experiment. Ss were instructed to identify the test stimuli as "B" or "D". For adaptation, they were told which stimuli were to be presented as adaptors.

Results and Discussion

As in the previous experiment, phonetic boundary loci for baseline and adaptation conditions were obtained by means of a computer program. The individual and mean loci for Ss in each of the four conditions are given in Table 3. The group identification functions for simple and contingent adaptation

----

Insert Table 3 about here

----

are shown in Figures 2 and 3 for voiced and voiceless adaptors, respectively.

Insert Figures 2 and 3 about here

For simple adaptation, in each condition the phonetic boundary shifted in the expected direction towards the adaptor relative to baseline. The shifts in all four conditions were all highly significant ($p < .005$) by correlated t tests. However, in the contingent adaptation conditions none of the boundary shifts were significantly different from the baseline condition. Thus, when the vowel context of both contingent adaptors is the same as the test series, the adaptation effects can be effectively canceled. These results provide some tentative support for the idea that adaptation of the place feature is also carried out in a vowel-dependent manner.

We also performed an analysis of variance on the difference scores between baseline and adaptation conditions in order to test for main effects in the experiment and the presence of any interactions. The difference between voiced and voiceless adaptors was highly significant [F (1, 20) = 37.36, $p < .001$]. In each case, the voiced adaptors produced larger shifts than the voiceless ones. This result replicates the findings of Cooper (1974b) and Cooper and Blumstein (1974) who showed smaller shifts on a [bɛ], [dɛ], [gɛ] place series with the voiceless adaptor [p^hɛ].

There was no main effect for place of articulation of the adaptor (i.e., bilabial vs. alveolar) nor were any of the higher order interactions significant. As expected, the main effect of adaptation (i.e., simple vs. contingent) was highly significant [F (1, 20) = 35.57, $p < .001$].

The outcome of this experiment on the place feature is quite clear. The adaptation effects obtained with simple adaptors will cancel in a contingent procedure when the vowel context is the same. Coupled with the previous findings on the voicing feature, these results provide strong support for the idea that at least one component of adaptation occurs at an auditory level prior to phonetic categorization. Furthermore, the results imply that the analysis of the consonant features requires acoustic information about the vowel context of the syllable.

Although the results of this experiment suggest that the analysis of the place feature is dependent on the vowel context, this conclusion is based on demonstrating that the simple adaptation effects will cancel under specified conditions. Stronger support would be obtained by demonstrating differential adaptation effects on place when the vowel context is manipulated directly. The next experiment was carried out to study this possibility.

Experiment III

In this experiment we use the contingent procedure again to study the place feature. However, now the vowel context of the adaptors is manipulated and we test on two place series in order to demonstrate differential adaptation effects as a function of vowel context.

Method

Subjects. Five paid volunteers served as Ss. They were obtained in the same manner as those used in Experiment I and met the same requirements.

Stimuli. Two new three-formant place series ranging between [b] and [d] were synthesized at Haskins Laboratories. Each series contained nine stimuli which varied in the direction and extent of the second and third formant transitions. One series contained the vowel [a] and the other the vowel [i]. All stimuli were 300 msec in duration. The starting frequencies of both series are given in Table 4. The adapting stimuli, [ba] and [di], were the end points

---

Insert Table 4 about here

---

from each respective test series.

Procedure. The procedure was similar to that used in Experiment I. The experiment was conducted on two days. On each day, Ss were first presented with the baseline identification tests for the [ba]-[da] and [bi]-[di] series separately. Each test contained 90 trials in which the nine stimuli were presented ten times each in a random order. Adaptation consisted of listening to an alternating sequence of [ba] and [di] and then identifying a different randomized sequence of nine stimuli selected from one of the test series. On half the trials the stimuli were drawn from the [ba]-[da] series and on the other half they were drawn from the [bi]-[di] series. At the end of the experiment, each S provided 20 identification responses to each stimulus in the baseline conditions and 18 responses to each stimulus in the adaptation conditions.

Results and Discussion

The individual and mean phonetic boundaries for each test series are presented in Table 5. The results indicate the presence of differential shifts

----

Insert Table 5 about here

----

in the locus of the phonetic boundary for each place series. Adaptation with the [ba] and [di] alternating sequence produced a shift in the [ba]-[da] series toward the [b] category and an opposite shift in the [bi]-[di] series toward the [d] category. Although there were a small number of Ss, t tests established that both shifts were significant ($p < .025$). Thus, adaptation of the place feature is also carried out in a vowel-dependent manner.

Additional support for the dependence of place adaptation on vowel context was provided in another experiment. Three Ss were presented with the voiceless adaptors [p^hi] and [t^ha] in exactly the same conditions as the main study. The results showed differential shifts in both test series in the predicted direction as governed by the vowel context of the adaptor. Thus, after adaptation with the contingent adaptors [p^hi] and [t^ha] the boundary of the [ba]-[da] series shifted toward the [da] end of the continuum whereas the [bi]-[di] series shifted toward the [bi] end. Moreover, as in Experiment II, the contingent shifts obtained on the place series with voiceless adaptors were somewhat smaller than those obtained with voiced adaptors.

The contingent effects on place obtained in the present experiment have a number of implications for specifying the locus of the adaptation effects in the speech perception system. We suggested earlier that the contingent adaptation procedure provided a more direct way to study the potential interaction of auditory and phonetic components than simple adaptation. Indeed, as we noted, the smaller magnitude of the adaptation effects obtained in the contingent procedure might have been due to the cancellation of opposing effects at the phonetic level. This seemed like a reasonable explanation of the voicing results found by Cooper (1974b) and replicated in our first experiment. However, the magnitude of the contingent effects found in the present experiment is roughly comparable to the magnitudes of the simple effects found in Experiment II and suggests that most if not all of the adaptation effects can probably be attributed to interaction at an auditory level prior to phonetic analysis. Obviously, this is an important result since it bears directly on the issue of whether there are specific phonetic adaptation effects which cannot be traced uniquely to some common albeit complex acoustic property that adapting and test stimuli share. It is also clear that the issue dealing with the existence of auditory versus phonetic feature detectors cannot be resolved very easily from the results of the present study on contingent effects for place and we might be in a better position by adopting at the present time, the multi-component view of adaptation proposed by Cooper (1974b, 1975). The results of the present experiment and specifically the apparent absence of any phonetic effects are still, however, somewhat difficult to handle with this model. For example, if adaptation at a phonetic level canceled we might have anticipated smaller shifts in the contingent condition since its contribution would be based exclusively on the adaptation taking place at an auditory level. Instead, we have obtained shifts that are comparable to those found with simple adaptors thus suggesting that opponent phonetic effects of the contingent adaptors may not have contributed to the observed adaptation.

Summary and Conclusions

The results of these three experiments indicate that selective adaptation of both place and voicing features is carried out in a vowel-dependent manner. The presence of these contingent adaptation effects as a function of vowel context has two implications. First, these results provide additional support for an auditory account of the selective adaptation findings based on an interaction of the acoustic or spectral properties of the stimuli rather than a phonetic account based on the interaction of abstract phonetic features at a later stage of processing. Second, the contingent effects provide further support for the idea that the analysis of consonant features involves the use of acoustic information distributed over the whole syllable. With regard to the analysis of stop consonants this implies that the acoustic cues for the consonant and vowel segments appear to be interdependent and that there may be few if any invariant acoustic properties in the physical stimulus that are associated directly with a specific linguistic segment resulting from perceptual analysis.

In contrast to simple adaptation, the contingent procedure appears to provide a way to study the interaction of auditory and phonetic processes simultaneously. Obviously, additional experiments are required to specify the relative contributions of these components to adaptation and the exact nature of their interactions. Additional work on these problems is currently being carried out in our laboratory.

In summary, these experiments have demonstrated vowel dependent selective adaptation effects for voicing and place features in stop consonants. The first experiment replicated the results reported by Cooper on voicing. The second and third experiments demonstrated that the feature of place of production is also dependent on the vowel context of the syllable. Both sets of findings are consistent with an account of selective adaptation that has its locus at an auditory stage of perceptual analysis.

References

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 1, 61-66.

Bailey, P. Perceptual adaptation for acoustical features in speech. Speech Perception Series 2, 1973, No. 2, Department of Psychology, Queen's University of Belfast. Pp. 29-34.

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, & D. B. Pisoni (Eds.), Cognitive Theory, Volume I. Potomac, Maryland: Erlbaum Associates, 1975.

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

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

Eimas, P. D., Cooper, W. E., & Corbit, J. D. Some properties of linguistic feature detectors. Perception & Psychophysics, 1973, 13, 2, 247-252.

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

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

Pisoni, D. B., & Tash, J. Auditory property detectors and processing place features in stop consonants. *Perception & Psychophysics*, 1975, in press.

Stevens, K. N., & Klatt, D. H. The role of formant transitions in the voiced-voiceless distinction for stops. *Journal of the Acoustical Society of America*, 1974, *55*, 653-659.

Footnotes

[1]This research was supported by NIMH Research Grant MH-24027 and NIH Research Grant NS-12179 to Indiana University. We wish to thank Dr. Frank S. Cooper for making the facilities of Haskins Laboratories available to us for the preparation of the stimulus materials.

Table 1

Individual and Mean Phonetic Boundary Loci

(in msec of VOT)

	[da]-[t^hi] Adaptors					[di]-[t^ha] Adaptors			
	[ba]-[p^ha] Series		[bi]-[p^hi] Series			[ba]-[p^ha] Series		[bi]-[p^hi] Series	
S	Baseline	Adapt	Baseline	Adapt	S	Baseline	Adapt	Baseline	Adapt
01	22.3	19.5	30.0	31.3	11	27.5	32.2	33.4	41.7
02	25.5	22.5	22.2	25.0	12	31.0	34.4	40.7	34.2
03	26.3	22.5	29.5	28.5	13	27.9	31.5	38.8	40.7
04	26.1	23.2	33.8	27.5	14	22.7	23.2	30.5	28.4
05	28.9	27.7	33.2	32.5	15	22.9	27.7	31.5	27.2
06	25.9	24.5	33.8	35.0	16	27.7	31.3	32.3	35.0
07	28.2	24.5	32.9	35.0	17	27.9	30.0	35.9	36.7
08	25.7	23.5	34.1	40.0	18	24.5	28.1	33.0	34.2
09	28.8	27.3	33.0	32.5	19	22.8	24.1	41.7	39.3
10	30.9	32.2	37.5	45.0	20	28.2	31.6	28.5	26.3
Mean	27.0	24.4	32.5	32.6	Mean	26.4	29.5	33.1	33.0

Table 2

Starting Frequencies of the Second- and Third-Formant Transitions for the Synthetic CV Test Stimuli

	Starting Frequencies (in Hz)	
Stimulus	F2	F3
1	996	2180
2	1075	2348
3	1155	2525
4	1232	2694
5	1312	2862
6	1386	3026
7	1465	3195

Table 3

Individual and Mean Phonetic Boundary Loci

(in Stimulus Units)

VOICED ADAPTORS					VOICELESS ADAPTORS				
GROUP I					GROUP III				
S	Baseline	Ba Adapt	Baseline	Ba-Da Adapt	S	Baseline	P^ha Adapt	Baseline	P^ha-T^ha Adapt
01	3.69	2.22	3.90	4.00	13	3.67	2.69	3.43	2.71
02	3.82	2.40	4.31	4.38	14	3.59	3.00	3.56	3.44
03	3.57	2.56	4.00	5.30	15	4.62	3.67	4.11	3.71
04	4.47	2.67	4.42	5.73	16	3.91	3.23	3.50	3.56
05	3.43	2.60	3.47	3.91	17	3.17	2.57	3.40	3.10
06	5.42	2.64	5.46	4.33	18	3.57	3.17	3.38	3.42
Mean	4.07	2.51	4.26	4.60	Mean	3.75	3.05	3.56	3.32
GROUP II					GROUP IV				
S	Baseline	Da Adapt	Baseline	Ba-Da Adapt	S	Baseline	T^ha Adapt	Baseline	P^ha-T^ha Adapt
07	3.77	5.18	3.91	4.25	19	3.54	4.25	4.41	4.17
08	3.59	5.01	3.78	4.10	20	3.71	4.47	3.62	3.71
09	3.53	5.56	3.67	4.10	21	4.38	5.42	4.23	4.53
10	4.09	5.77	4.40	4.22	22	4.00	4.23	4.47	3.91
11	3.59	5.33	3.71	4.86	23	3.77	4.38	4.47	4.50
12	3.91	5.00	3.83	3.82	24	4.64	5.00	4.67	4.62
Mean	3.75	5.31	3.88	4.23	Mean	4.01	4.62	4.31	4.24

Table 4

Starting Frequencies in Hz of the Second- and Third-Formant Transitions for the [ba]-[da] and [bi]-[di] Test Stimuli

	[ba]-[da][a]		[bi]-[di][b]	
Stimulus	F2	F3	F2	F3
1	1075	2348	1465	2180
2	1155	2525	1541	2348
3	1232	2694	1620	2525
4	1312	2868	1695	2694
5	1386	3026	1772	2862
6	1465	3195	1845	3026
7	1541	3363	1920	3195
8	1620	3530	1996	3363
9	1695	3698	2078	3530

[a]Formant frequencies for the steady-state vowel [a] were: F1 = 769 Hz, F2 = 1232 Hz, F3 = 2525 Hz.

[b]Formant frequencies for the steady-state vowel [i] were: F1 = 286 Hz, F2 = 2307 Hz, F3 = 3026 Hz.

Table 5

Individual and Mean Phonetic Boundary Loci

(in Stimulus Units)

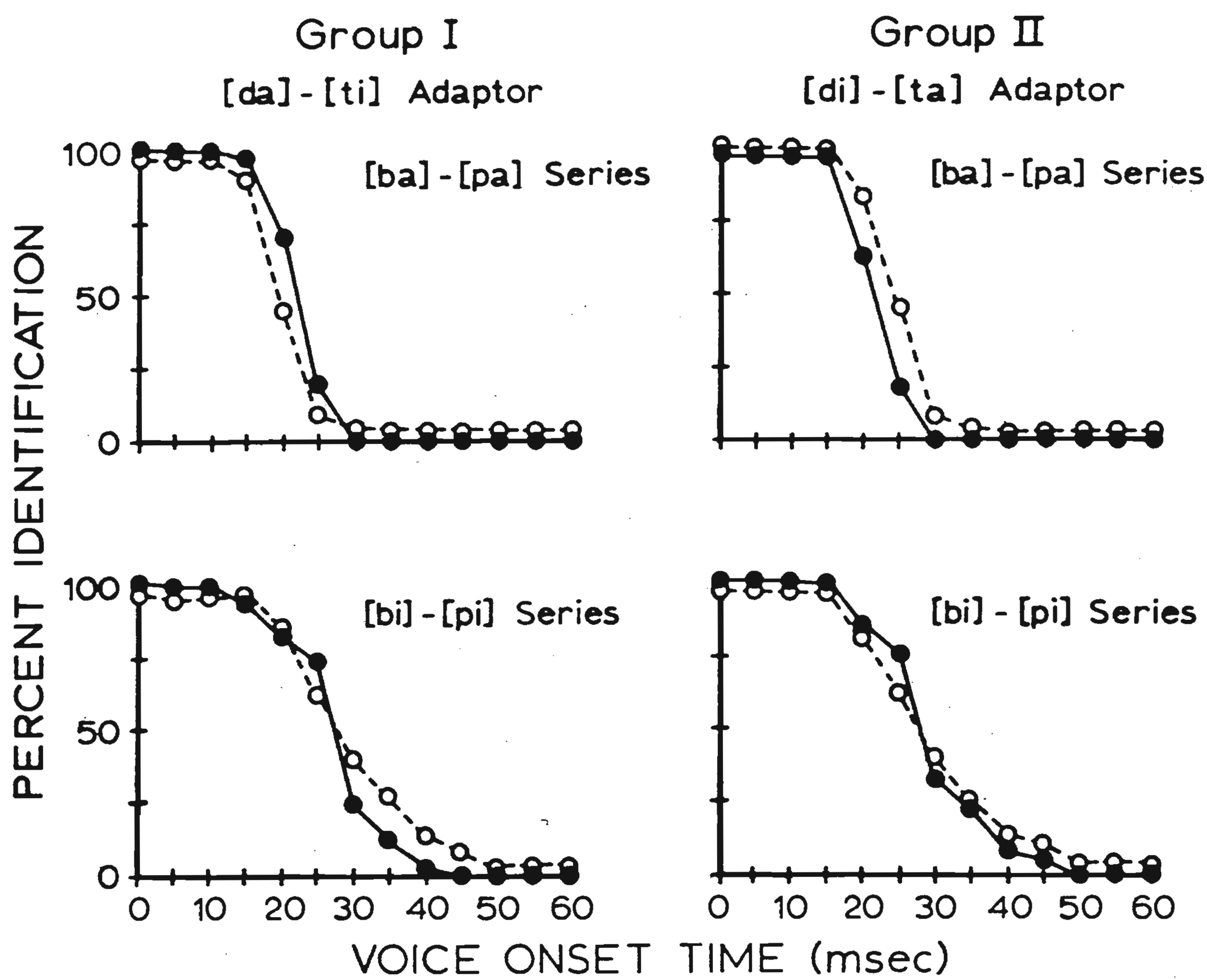
	[ba]-[di] Adaptors			
	[ba]-[da] Series		[bi]-[di] Series	
S	Baseline	Adapt	Baseline	Adapt
01	4.11	2.00	5.40	5.50
02	4.00	1.38	5.88	6.36
03	2.83	3.00	6.00	7.54
04	3.89	2.33	5.20	6.67
05	3.25	2.30	4.75	6.60
Mean	3.54	2.18	5.42	6.56

Figure Legends

Figure 1. Identification functions showing the percentage of "B" responses before adaptation (solid lines) and after contingent adaptation (dashed lines) for each group of ten Ss. Group I received the [da]-[t^hi] adaptors whereas Group II received the [di]-[t^ha] adaptors.

Figure 2. Identification functions showing the percentage of "B" responses before adaptation (solid lines) and after adaptation (dashed lines) with voiced stops. The top functions show simple adaptation with [ba] or [da] whereas the bottom functions show the results of contingent adaptation with [ba] and [da].

Figure 3. Identification functions showing the percentage of "B" responses before adaptation (solid lines) and after adaptation (dashed lines) with voiceless stops. The top function shows simple adaptation with [p^ha] or [t^ha] whereas the bottom functions show the results of contingent adaptation with [p^ha] and [t^ha].



Group I
[da] - [ti] Adaptor
Group II
[di] - [ta] Adaptor
[ba] - [pa] Series
[bi] - [pi] Series
PERCENT IDENTIFICATION
100
50
0
0 10 20 30 40 50 60
VOICE ONSET TIME (msec)

Figure 1.

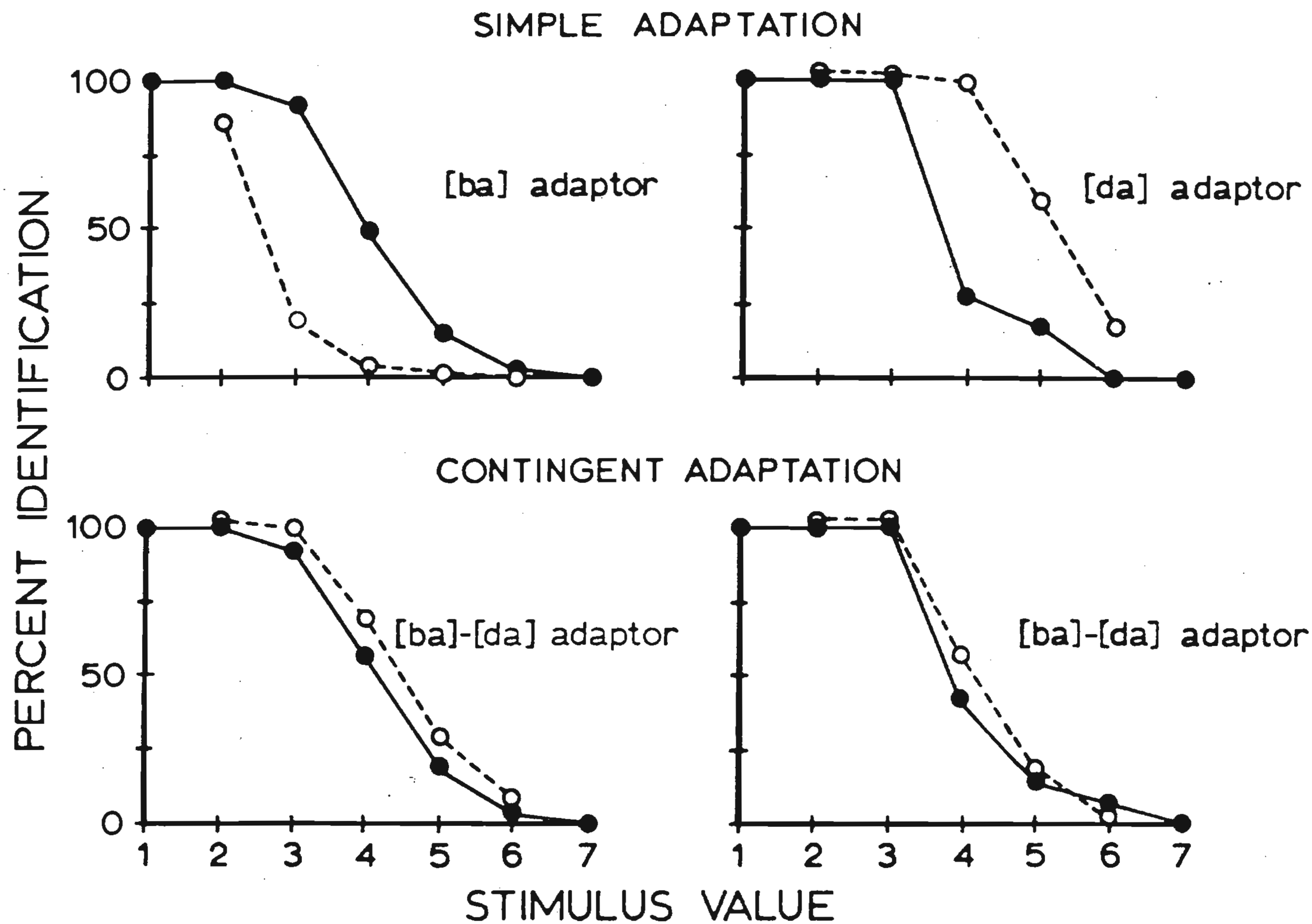

SIMPLE ADAPTATION
[ba] adaptor
[da] adaptor
CONTINGENT ADAPTATION
[ba]-[da] adaptor
[ba]-[da] adaptor
PERCENT IDENTIFICATION
100
50
0
1
2
3
4
5
6
7
STIMULUS VALUE

Figure 2.

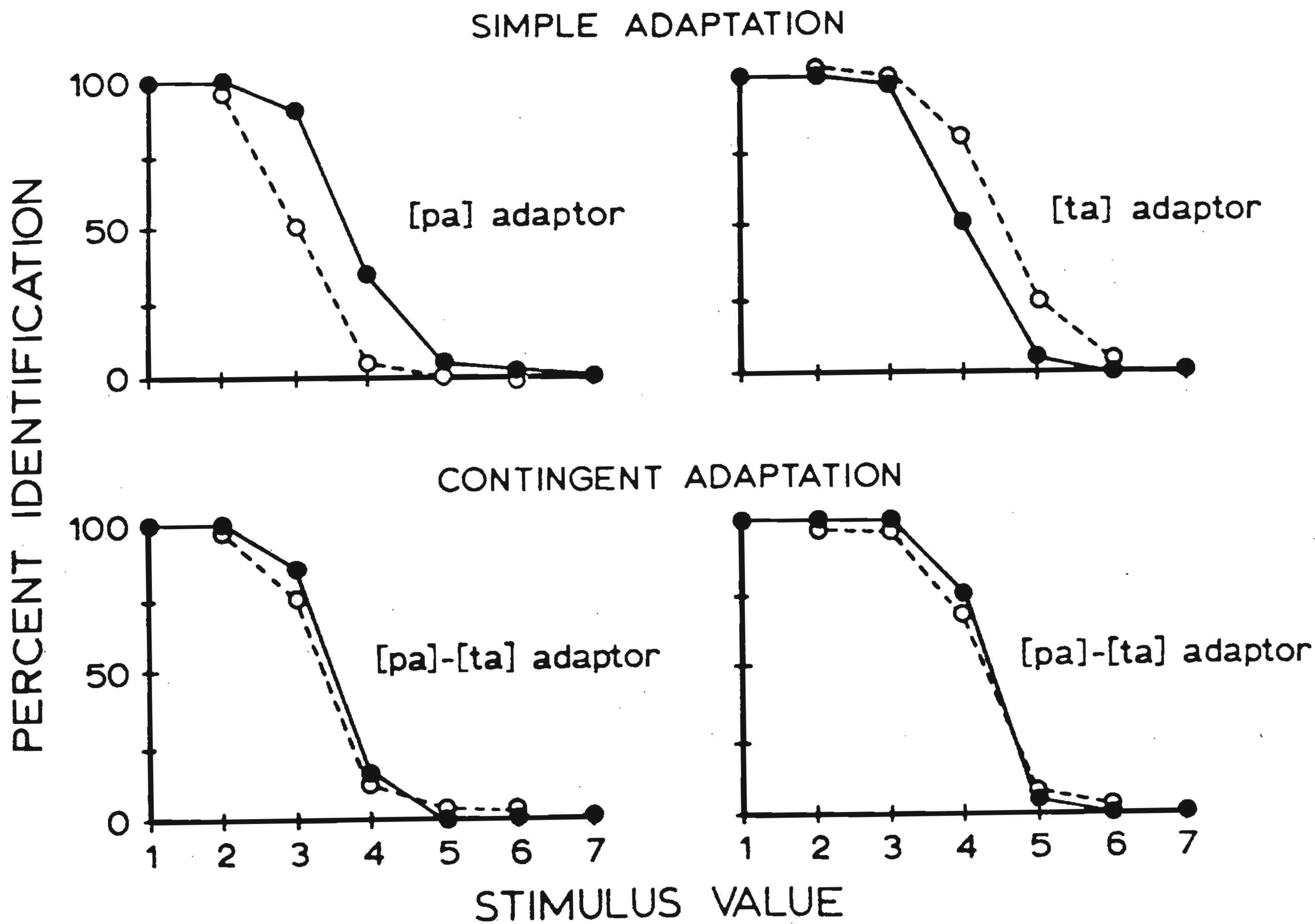

SIMPLE ADAPTATION
[pa] adaptor
[ta] adaptor
CONTINGENT ADAPTATION
[pa]-[ta] adaptor
[pa]-[ta] adaptor
PERCENT IDENTIFICATION
100
50
0
100
50
0
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 3.

INTENSITY AND REPETITION EFFECTS
ON SELECTIVE ADAPTATION TO SPEECH

BY

JAMES M. HILLENBRAND

Submitted to the faculty of the Graduate School
in partial fulfillment of the requirements
for the degree Master of Arts in the
Department of Speech and Hearing
Sciences, Indiana University
August, 1975

ACKNOWLEDGEMENTS

I would like to express my deep appreciation to Dr. David Pisoni, director of this thesis, Dr. Gordon Schuckers, chairman of the thesis committee and Dr. David Ratusnik who served as a member of the thesis committee. Their criticisms, suggestions and encouragement were of inestimable value to every aspect of the design of the study and preparation of the thesis.

I also wish to thank Dr. N. John Castellan, Dr. Donald Robinson, James Sawusch, Sandra MacNabb, Chuck Arnold, Dennis Ingrisano and Jeffrey Tash, who contributed their time and valuable ideas to this work.

In addition, I would like to express my appreciation to Dr. David Pisoni for providing the funds (NIMH Grant MH-24027 and NIH Grant NS-12179) and making available the facilities of the Speech Perception Laboratory at the Indiana University Psychology Department which made this investigation possible.

Above all, I wish to express my sincere gratitude to Dr. Robert Milisen and Mrs. Gladys Devane who introduced me to the study of communication and who provided both encouragement and inspiration throughout my academic career at Indiana University.

J.M.H.

CHAPTER I

INTRODUCTION

Recent theoretical accounts have begun to view the perception of speech as a complex process involving several hierarchically organized stages of perceptual analysis (Stevens, 1960; Stevens and Halle, 1967; Fant, 1967; Stevens and House, 1972; Studdert-Kennedy, 1974; Pisoni and Sawusch, 1975). According to these stage models, the acoustic signal undergoes a series of transformations in which information is recoded into successively more abstract forms. The information processing model of Pisoni and Sawusch (1975) exemplifies the basic aspects of stage models (Fig. 1).

Insert Fig. 1 about here

At the earliest stage of analysis, Preliminary Auditory Analysis, the signal is transformed into a spectral display in terms of frequency, intensity and time. The output of this stage is then processed by a four-stage Recognition Device. The spectral display is first analyzed in terms of acoustic features, such as formant transitions, burst cues and the presence or absence of the fundamental frequency from the beginning of the stimulus. The output of the Acoustic Feature Analysis stage then undergoes Phonetic Feature Analysis where a set of decision rules maps multiple

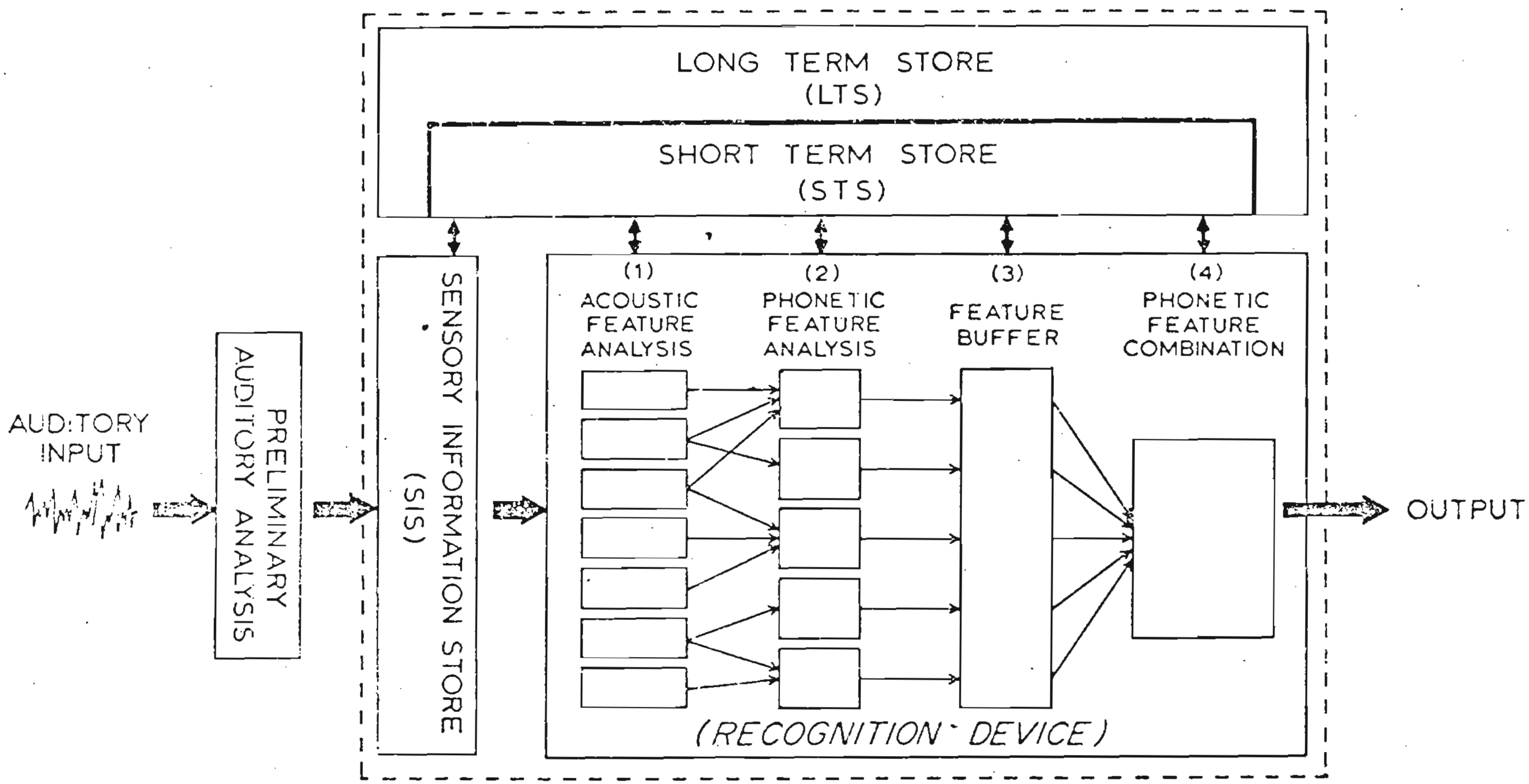


Figure 1. Information processing model of speech perception (Pisoni and Sawusch, 1975).

acoustic features into phonetic distinctive features. The individual distinctive features extracted at this level are stored in the Feature Buffer and are subsequently recombined to form discrete phonetic segments through processing by the Phonetic Feature Combination stage. The output of the Recognition Device is available to be operated upon by the higher levels of phonological, syntactic and semantic analysis.

While other stage models may differ from the Pisoni and Sawusch model in many specifics, most recent approaches have incorporated the distinction between acoustic and phonetic levels of processing (Stevens and House, 1967; Fant, 1967; Bondarko, 1970). This distinction has received experimental support from a variety of sources (cf. Studdert-Kennedy and Shankweiler, 1970; Studdert-Kennedy, Shankweiler and Pisoni, 1972).

A growing body of experimental evidence suggests the possibility that feature or property detectors are operating at one or more of the stages of perceptual analysis (cf. Studdert-Kennedy, 1974; Cooper, 1975). Considerable controversy exists, however, as to whether these feature detectors are responsible for the extraction of relatively low-level acoustic features or whether they function to extract the relatively more abstract phonetic features of the signal.

The possibility also exists that feature detectors operate at both acoustic and phonetic levels of processing. The present study will address itself to this acoustic/phonetic issue.

CHAPTER II

REVIEW OF THE LITERATURE

The notion of feature detectors, cells selectively responsive to a restricted range of values of a stimulus on some dimension, was first demonstrated in the visual ganglion cells of the frog by Lettvin, Maturana, McCulloch and Pitts (1959). They demonstrated the existence of specialized neural receptors selectively sensitive to restricted ranges of light intensity, movement and contour. Four classes of feature detectors were observed: edge detectors, movement detectors, dimming detectors and convex edge detectors.

More complex detector systems were reported by Hubel and Wiesel (1962, 1965) in the visual cortex of the cat. Using a technique which allowed recording from single cells, they provided evidence for the existence of feature detectors selectively responsive to edges, the orientation of lines, and movement in a specific direction. Based on this work, the authors further suggested that the extraction of features from the visual signal involves several hierarchically organized stages, with each stage extracting successively more abstract features.

Feature detector systems have also been demonstrated in auditory perception. Cells responsive to specific gradients of frequency and intensity change have been reported in the cortex and inferior collibulus of the cat (Whitfield and Evans, 1965; Evans and Whitfield, 1964; Nelson, Erulkar and Bryan, 1966). Evidence has also been reported for feature detectors involved in acoustic signaling systems, such as the mating call of the male bullfrog (Frishkopf and Goldstein, 1963; Capranica, 1965) and the "isolation peep" of the squirrel monkey (Wollberg and Newman, 1972).

The first suggestion that feature detectors mediate the extraction of the features of speech sounds was made by Abbs and Sussman (1971). The authors observed that feature detectors tend to be found in animals in both vision and audition (1) when the rapid processing of particular stimuli is important for survival and (2) when the speed of processing is at the practical limits of the organism's neuronal capabilities. Arguing that the perception of speech satisfies these criteria, Abbs and Sussman postulated feature detectors that respond to "physical parameters that are composed of several different aspects of the signal, i.e. frequency, intensity, rate of frequency change, rate of intensity change, and durational characteristics of these attributes" (p. 24).

Evidence bearing directly on the feature detector hypothesis was first obtained by Warren and Gregory (1958) using an adaptation paradigm. Adaptation involves the presentation of a stimulus repeatedly or continuously for a relatively long period of time. Following prolonged stimulation the organism is typically found to be less sensitive to that stimulus and relatively more sensitive to an adjacent or related stimulus. Prolonged stimulation is assumed to fatigue neurological units selectively responsive to some feature represented in the stimulus (Studdert-Kennedy, 1974).

In Warren and Gregory's study, subjects listened to a word presented repeatedly once or twice a second and were instructed to report any changes that they heard in the word. Subjects reported a large number of changes or "verbal transformations" to other meaningful words, although they were not always related phonetically to the original word.

In a refinement of this technique, Goldstein and Lackner (1974) repeatedly presented real-speech CV, V, and VC nonsense syllables monaurally to subjects who reported hearing changes generally in a single phone and confined to one or two distinctive features. The authors reported that significantly more transformations occured in the right ear than the left for consonants but not for vowels, and further that transformations followed various phonological constraints of English.

Although the verbal transformation experiments suggest the existence of feature detectors involved in the processing of speech sounds, the use of real as opposed to synthetic speech renders it impossible to assign the perceptual effects to specific acoustic features of the stimulus (Cooper, 1975).

In order to examine the adaptation phenomenon in greater detail, Eimas and Corbit (1973) constructed two continua of synthetic CV syllables varying from [ba] to [p^ha] and [da] to [t^ha] in equal steps of the acoustic cue voice onset time (VOT).

Voice onset time is a major cue to voicing distinctions in stops in initial position and differentiates the voiced stops [b], [d] and [g] from the voiceless stops [p^h], [t^h] and [k^h]. Acoustically, VOT refers to the interval between the time of onset of the second and third formants relative to the time of onset of the first formant. The amount of delay in VOT required for a stop to be perceived as voiceless is about 30 to 40 msec. for the bilabial series, although it varies with place of production (Lisker and Abramson, 1964).

Adaptation consisted of repeatedly presenting stimuli from either the extreme voiced end of the continuum ([ba] or [da]) or the extreme voiceless end of the continuum ([p^ha] or [t^ha]). Identification functions for these stimuli were obtained for listeners in the unadapted state and then following adaptation. The loci of the phonetic boundaries were calculated by the method of least mean squares.

Identification functions obtained following adaptation showed a significant shift in the locus of the phonetic boundary in the direction of the voicing category of the adapting stimulus. That is, more responses were assigned to the unadapted category after adaptation. For example, after adaptation with either of the voice stops, [ba] or [da], shifts were obtained in the phonetic boundary locus for both the [ba] - [p^ha] series and the [da] - [t^ha] series in the direction of the voiced end of the continuum. Adaptation with either of the voiceless stops, [p^ha] or [t^ha], resulted in shifts toward the voiceless end of either continuum. The magnitude of these shifts typically ranged from 5 to 15 msec. in VOT.

The strong presence of crossed-series shifts (e.g. shifts in the [da] - [t^ha] continuum following adaptation with [ba]) was regarded by Eimas and Corbit as a particularly important effect. Although crossed-series shifts were only about 80% the magnitude of same-series shifts, their presence indicated that the adaptation effect operates primarily on some property of the consonant, rather than the consonant as a unit. The assumption upon which this is based is that the presence of the crossed-series shifts indicates the sharing of a common property between the adapting stimulus and the adapted members of the test series. Eimas and Corbit claimed that this common property was voicing, represented acoustically by variations in VOT.

Basing their conclusions primarily on the presence of crossed-series shifts, Eimas and Corbit proposed the existence of two classes of feature detectors that extract information dealing with the abstract phonetic feature of voicing. The authors proposed that each detector is sensitive to a restricted range of VOT values. Some VOT values, however, excite both detectors and the phonetic boundary locus corresponds to the stimulus that excites both detectors equally often. The authors assumed that repeated presentation of a stimulus containing a specific voicing feature would reduce its sensitivity. Thus, after adaptation the opponent or unadapted detector provides more information to the decision process than the adapted detector and subsequently results in a shift in the locus of the phonetic boundary between the two segments.

The property of this interpretation that has drawn the greatest attention is the idea that the adaptation effect operates on a mechanism that extracts phonetic (i.e. voicing) rather than acoustic information (i.e. VOT).

Since publication of the Eimas and Corbit study, numerous studies have been directed toward specifying the nature of the adaptation effect in greater detail. Most of this research attempts to answer two fundamental questions. The first question concerns the neural site of adaptation (see Cooper, 1975). Studies in this area have basically attempted to determine whether the adaptation effect operates on fused information from both ears or information extracted monaurally. The second, and perhaps more complicated question, concerns the function that the adapted component of the system performs in the perception of speech. Although Eimas and his collaborators have attributed the selective adaptation effects to abstract phonetic feature detectors in the Chomsky and Halle (1968) sense, there is now a growing body of evidence that indicates that these results may have an acoustic basis at a much earlier stage of perceptual analysis. It has been argued that the proposed feature detectors do not necessarily respond to the phonetic features of a segment but may, in fact, respond to acoustic features corresponding to the phonetic units that are being adapted. The presence of crossed-series shifts, upon which Eimas and Corbit based their claim for phonetic feature detectors, is ambiguous since the adapting stimulus and the adapted members of the test series share both the phonetic feature of voicing and the acoustic feature of VOT. As such, the feature detectors may be acoustic rather than phonetic in nature.

Neural site of Adaptation

In a second series of experiments by Eimas, Cooper and Corbit (1973) an attempt was made to determine the location of the adaptation effect by testing for interaural transfer of adaptation. In this experiment, the adapting stimulus [t^ha] was presented to one ear while the [da] - [t^ha] test series was presented to the other ear. It was reasoned that if adaptation occured under this condition, then the effect could be attributed to a binaurally driven mechanism, since the effects of a monaurally driven mechanism would be confined to the ear to which the adapting stimuli were presented. Significant phonetic boundary shifts were obtained under the monaural condition which were approximately 95% of the magnitude of shifts obtained for the same subjects under binaural presentation. These results suggest that adaptation operates primarily on fused information from both ears, and as such is central in location, rather than peripheral.

Although the interaural transfer effect can be attributed to a mechanism that operates primarily on information subsequent to binaural fusion, these results do not preclude the possibility that a component of the effect operates on monaural information. An experiment by Ades (1974a) sought to test this possibility. The stimuli used in this experiment varied along the phonetic dimension place of articulation. In English, place of articulation serves to differentiate between the voiced stops [b], [d] and [g] and the voiceless stops [p^h], [t^h] and [k^h]. In terms of articulation, this dimension refers to the point in the vocal tract where the breath stream is constricted or temporarily interrupted. Second and third formant transitions (i.e. rapid formant frequency changes) are the major acoustic cues to place of articulation (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967).

The test stimuli used in the Ades study were constructed to range perceptually from [bæ] to [dæ] by systematically varying the extent and direction of the second and third formant transitions. The adapting stimuli [bæ] and [dæ] were presented simultaneously, one to each ear. Identification functions for a [bæ] - [dæ] continuum were obtained before and after adaptation separately for each ear. Ades reports that listeners exhibited small but statistically significant phonetic boundary shifts following adaptation. For example, after adaptation with [bæ] to the right ear and [dæ] to the left ear, shifts were obtained in the direction of [bæ] when tested in the right ear and [dæ] when tested in the left ear. These results suggest that some component of the adaptation effect may operate on unfused information from each ear, leaving open the possibility of a peripheral component.

In a second experiment, however, Ades (1974a) provided further support for the notion that the adaptation effect operates primarily on fused information from both ears. The adapting syllable [bæ] was presented to listeners in fused and unfused conditions. In the fused condition, the first formant transition was presented to one ear while the second and third formant transitions were simultaneously presented to the other ear. In the unfused condition, the same transitions were presented in an alternating sequence. Significant phonetic boundary shifts were obtained in the fused condition but not in the unfused condition, providing evidence for a binaurally driven mechanism.

Cooper (1975) indicated that these studies provide evidence in support of both binaurally and monaurally driven sites of adaptation and proposed that "both sites account for the total effect during regular binaural testing. Since all other relevant experiments conducted in the meantime have involved binaural testing only, we are in the unfortunate position of being unable to assign the results of such tests to either or both of the adaptation sites" (p. 31). Further research in binaural-monaural testing is clearly needed before the vast amount of data in binaural testing can be adequately interpreted in terms of the site of adaptation.

Function of the Adapted Component

The question that has generated the most research in selective adaptation concerns the issue of whether the adapted mechanism is responsible for the extraction of abstract phonetic features, as Eimas and Corbit (1973) argued, or whether it is responsible for the extraction of relatively low-level acoustic features.

In an attempt to corroborate their earlier phonetic interpretation, Eimas, Cooper and Corbit (1973) conducted an experiment designed to determine whether the feature analysis of speech operates on general auditory information or only information that is contained in a linguistic context. In this experiment, adaptation with [da] was compared to adaptation with a truncated syllable consisting of the first 50 msec. transition of [da]. It was reasoned that all the acoustic (VOT) information necessary to make a determination on voicing was available in the first 50 msec. transition. This "slice" of the syllable, however, does not sound like speech but rather is perceived by most listeners as a nonspeech "chirp" (Mattingly, Liberman, Syrdal and Hawles, 1971). Eimas et al. argued that if adaptation were obtained with [da] but not with the "d-chirp", it could be concluded that adaptation does not operate as a part of the general auditory system, but rather operates as part of a specialized speech processor which engages only when sounds are recognized as speech.

Adaptation with [da] produced the predicted shifts in the phonetic boundary of the [da] - [t^ha] continuum, but adaptation with the d-chirp produced no significant shifts. The authors interpreted these results as evidence for feature detectors which operate on auditory information only to the extent that it is embedded in a linguistic context (i.e. part of a speech signal).

This provided indirect support for the phonetic interpretation since it seemed that adaptation was not the result of the fatigue of general auditory feature detectors but rather detectors that were dependent upon linguistic or phonetic information as input. Several investigators, however, have taken issue with the authors basic assumption that the acoustic information contained in the d-chirp is identical to that contained in the full [da] syllable with respect to voicing distinctions (Studdert-Kennedy, 1974; Wood, 1973). It has been argued that VOT as a cue to voicing operates by signalling the time of onset of the second and third formants relative to the time of onset of the first formant. It is possible, then, that insufficient acoustic information is available in the d-chirp to make this relative judgement since the duration of the formants is so brief. If this is the case, then the Eimas et al. assumption that the acoustic information contained in the d-chirp is identical to that contained in the full [da] syllable with reference to voicing decisions is questionable.

The phonetic interpretation of the Eimas et al. data has also been questioned by Cooper (1975) who has noted that although no significant shifts were obtained with the d-chirp, five of the seven listeners exhibited small but nonsignificant shifts in the direction of the [d] category. Cooper also cites an unpublished replication by Ades in which listeners exhibited "small, but in this case, statistically significant shifts using comparable VOT stimuli" (p. 31). Both results suggest the possibility of an acoustic account of adaptation.

The two studies by Eimas and his colleagues examined the effects of adaptation on the phonetic feature of voicing signalled by changes in VOT. Recently, investigators have attempted to examine the effect of adaptation for place of articulation. Place of articulation was of particular interest for two reasons. First, the place dimension serves to distinguish three categories (e.g. [b], [d], and [g]) as opposed to voicing which distinguishes only two categories (e.g. [b] and [p]). It was unclear whether adaptation with [b], for example, would result in shifts in a [b] - [d] continuum only or whether shifts would be obtained in a [d] - [g] continuum as well. The second reason for the special interest in the place dimension is that, unlike the feature of voicing, there exists no invariant range of acoustic cues which signal place distinctions. Rather, the acoustic cues for place (second and third formant transitions) are highly variable depending upon the phonetic environment of the consonant. The familiar example of [di] and [du] illustrates this point (Liberman, 1970). In the syllable [di], the cue for the perception of the consonant [d] is a rising second formant transition, beginning at about 2200 Hz and rising to about 2600 Hz. In the syllable [du], on the other hand, the second formant transition falls from about 1200 Hz and terminates in a steady-state vowel at about 700 Hz. Thus, the same perception is cued, in different phonetic environments, by radically different acoustic features.

The first series of experiments in selective adaptation using a place of articulation continuum were conducted by Cooper (1974a). Cooper constructed a test series consisting of 13 synthetic syllables ranging from [bæ] to [dæ] to [gæ]. Cooper sought to determine: (1) whether adaptation operates only on acoustically invariant information, such as VOT or whether it operates on contextually variable cues as in place information, and (2) whether adaptation with the endpoints [bæ] or [gæ] would produce shifts in the single continuum which they bordered or whether shifts would be obtained in both continua.

Adaptation with [bæ] produced the predicted shifts in the [bæ] - [dæ] continuum, but the [dæ] - [gæ] continuum remained relatively stable. Similarly, adaptation with [gæ] produced a shift in the [dæ] - [gæ] boundary but did not affect the [bæ] - [gæ] boundary. These results indicate that adaptation can, indeed, be demonstrated with contextually variable information. Cooper suggested that adaptation is operating on relatively stabile feature modes, such as bilabial, alveolar and velar since adaptation with the end-point stimuli affected only the continua which they bordered. Adaptation, then, could not be seen as affecting a purely relative feature analyzing system (see Cooper, 1974a).

Another important finding of this study was that adaptation with the middle category, [dæ], produced shifts in both the [bæ] - [dæ] continuum and the [dæ] - [gæ] continuum. Cooper interpreted this result as providing tentative support for the phonetic interpretation. The assumption underlying this claim is that the adapting syllable [dæ] contains slightly falling second and third formant transitions, yet it affected stimuli near the [bæ] - [dæ] boundary which contain rising transitions as well as stimuli near the [dæ] - [gæ] boundary which contain falling transitions. Cooper concluded that this effect could not be accounted for by an explanation based on the fatiguing of acoustic feature detectors which are selectively responsive to either the absolute starting frequency or the direction of second and third formant transitions.

In a second experiment, Cooper sought to determine the effect of variations of the adapting stimulus on phonetic boundary shifts for the same [bæ] - [dæ] - [gæ] continuum. In one experimental condition, the real speech syllable [bæ] was compared to the synthetic syllable [bæ] in its effect on the [bæ] - [dæ] - [gæ] continuum. Results indicated that adaptation with both stimuli resulted in a shift in the [bæ] - [dæ] boundary, but the synthetic [bæ] resulted in a significantly greater shift than its real speech counterpart. Cooper speculated that this difference was due to the fact that the real speech syllable "contained more acoustic information that was irrelevant to the task of perceiving phonetic distinctions based solely on differences in the second- and third-formant transitions" (p. 624).

In a second experimental condition, Cooper tested for the effect of crossed-vowel adaptation. In the crossed-vowel condition, adaptation with the real speech syllable [bi] was compared in its effect on the [bæ] - [dæ] - [gæ] boundary with adaptation with the real speech syllable [bæ]. Results indicated that both syllables were effective in shifting phonetic boundary loci. However, the magnitude of the shifts obtained with [bi] were significantly less than those obtained with [bæ]. This result is of interest since the acoustic cues specifying place in the [bi] syllable are different from those in the syllable [bæ]. The finding that some adaptation was obtained with [bi] suggests that the sharing of the phonetic feature of place, rather than the acoustic features corresponding to that distinction, was responsible for the obtained effect. On the other hand, the finding that [bi] was a less effective adaptor suggests that this phonetic interpretation may be too simplistic. To explain these results, Cooper proposed a multi-component view of adaptation in which adaptation occurs at more than one level of perceptual analysis, such that part of the effect is attributable to the fatiguing of feature detectors at the acoustic level of analysis and part of the effect is attributable to the fatiguing of feature detectors at the phonetic level. Thus, [bi] was a less effective adaptor because it served to fatigue feature detectors only at the phonetic level, whereas adaptation with [bæ] resulted in fatigue at both acoustic and phonetic levels.

This multi-component interpretation, however, falls short of explaining the results of Cooper's third experimental condition in which he tested for crossed-consonant adaptation. In this condition, adaptation with the real speech syllable [pʰæ] was compared to adaptation with the real speech syllable [bæ]. This condition was of interest because the syllables [bæ] and [pʰæ] share both the acoustic and phonetic information with reference to place distinctions. According to the multi-component account, then, the [bæ] and [pʰæ] syllables should be equally effective adaptors since both syllables should result in fatigue at acoustic and phonetic levels. The results, however, do not support this prediction. Adaptation with [bæ] resulted in significantly greater shifts than those obtained with [p^hæ]. Clearly, an additional component is necessary to account for this difference. Cooper proposed a higher level "phonetic unit" component of adaptation "which operates not on individual distinctive features but on the consonantal sound as a unit..." (p. 624). Thus, [bæ] was a more effective adapting stimulus because its consonant was represented in the test series, whereas the consonant /p/ of [p^hæ] was not.

The crossed-vowel and crossed-consonant results, then, provide some evidence for feature detectors operating to extract both acoustic and phonetic information in a feature-specific manner, and further, suggest the possibility of detector mechanisms which operate on the consonant sound as a unit. In order to obtain further information about the organizational properties of the feature-specific component of adaptation, Cooper and Blumstein (1974) conducted an experiment involving a number of crossed-consonant conditions. Identification functions were first obtained for a [bæ] - [dæ] - [gæ] continuum. Following this, listeners were adapted with one of five adapting syllables: [bæ], [p^hæ], [mæ], [væ] and [wæ]. All of the adapting syllables share the place feature in that they all have an initial labial segment. In relation to the control syllable [bæ], however, they differ on either the feature of voicing ([p^hæ]) or manner ([mæ], [væ] and [wæ]). The authors reasoned that if adaptation operated in a feature-specific manner, i.e. irrespective of voicing or manner information, all of the labial adapting stimuli should result in phonetic boundary shifts in the [bæ] - [dæ] continuum. The results partially supported this prediction. Adaptation with [bæ] and [p^{h}æ] produced shifts in [bæ] - [dæ] boundary toward [bæ], similar to those obtained in the Cooper (1974a) study. This finding indicates that place information is extracted independently of voicing information. Adaptation with [mæ] and [væ] also produced the predicted shifts, indicating that place information is extracted independently of manner information. The adapting syllable [wæ], however, produced only a slight, nonsignificant shift in the [bæ] - [dæ] boundary, although it was in the predicted direction. Cooper and Blumstein suggested that since [wæ] is a semivowel, this finding may indicate that the defining limit of the "labial" analyzer is that it extracts place information only for the class of true consonants.

In a more recent study, Cooper (1974b) examined the influence of vowel context on selective adaptation. Two test series were constructed, one varying from [ba] to [p^{h}a] and the other varying from [bi] to [p^{h}i]. Adaptation consisted of presenting the syllables [da] and [t^{h}i] in alternating groups of three. Cooper predicted that if adaptation operates in a vowel-independent manner, such that phonetic features are extracted independently of their vowel environment, then no adaptation should occur in this condition. The reasoning behind this prediction is that the adapting stimuli represent opposing values on the phonetic feature of voicing. If voicing is extracted independently of vowel environment, the effect of adaptation with [da] (+voiced) should be cancelled by the effect resulting from adaptation with [t^hi] (-voiced). In other words, voiced and voiceless detectors should be fatigued equally and no perceptual shifts should be observed in either test continuum. However, his results indicated that the extraction of features is considerably more complex than this. Adaptation with the alternating sequence of [da] and [t^hi] resulted in shifts toward [b] in the [ba] - [p^ha] series and toward [p^h] in the [bi] - [p^hi] series. Thus, the perception of the [ba] - [p^ha] series was primarily influenced by adaptation with [da], and conversely, the perception of the [bi] - [p^hi] series was primarily influenced by adaptation with [t^hi]. Cooper suggested that adaptation does not operate by extracting phonetic features in any simple feature-specific manner, but rather operates in a vowel-contingent manner. It is therefore difficult to account for these results by proposing feature detectors which operate to extract phonetic distinctive features in the Chomsky-Halle (1968) sense.

Some investigators have argued that the results favoring a feature-specific component may reflect the operation of complex acoustic detector mechanisms sensitive to the acoustic cues underlying phonetic distinctive features (cf. Pisoni and Tash, 1975). The crossed-consonant and crossed-vowel results obtained by Cooper (1974a) and Cooper and Blumstein (1974) do not clearly differentiate between phonetic cues and the acoustic cues that underlie these distinctions. Although adaptation in these studies was obtained using adapting stimuli that differed acoustically from the adapted members of the test series, the magnitude of that difference may have been overstated. DeLattre, Liberman and Cooper (1955) suggested that although varying vowel context changes the absolute direction of formant transitions, each consonant has a characteristic frequency position toward which it "points". Thus, it may be this abstract frequency locus which accounted for the crossed-vowel shifts, rather than phonetic similarity. The crossed-consonant shifts may also be accounted for in terms of acoustic detectors. For example, in the Cooper (1974a) and Cooper and Blumstein (1974) studies, it was found that [p^hæ] was a less effective adaptor than [bæ]. This can be accounted for in acoustic terms by noting that [p^hæ] contains relatively weak second and third formant transitions, whereas [bæ] and the adapted members of the test series contain second and third formant transitions that are relatively strong in energy. Cooper and Blumstein also found that [mæ] and [væ] were effective adapting stimuli, whereas [wæ] resulted in only small nonsignificant shifts. This too can be explained in acoustic terms by noting that [mæ], [væ] and the adapted members of the test series share sharply rising formant transitions, which differ from the gradually rising transitions in [wæ]. It is therefore difficult to determine from the data presented thus far, whether adaptation operates on phonetic features, acoustic features or both.

In order to distinguish between acoustic and phonetic interpretations of selective adaptation, Diehl (1975) conducted an experiment which examined the effect of adapting with a stimulus cued for place by a burst of noise on a place continuum cued by variations in formant transitions. Diehl argued that the evidence presented in favor of phonetic feature detectors by Cooper (1974a) and Cooper and Blumstein (1974) rests on the questionable assumption that the crossed-vowel and crossed-consonant conditions represent situations in which the adapting stimuli and adapted members of the test series share phonetic features but do not share acoustic features. Diehl argued that in order to test the phonetic hypothesis, an adapting stimulus must be presented "which shares a phonetic feature with some of the test stimuli but which has virtually nothing in common with them acoustically" (p. 49). If such a stimulus produced phonetic boundary shifts, the phonetic hypothesis would be supported. On the other hand, if no adaptation were observed, the phonetic hypothesis could be rejected. In order to provide such a test, a series of test stimuli was constructed ranging from [bɛ] to [dɛ] by varying the starting frequency and direction of the second and third formant transitions. The adapting stimuli consisted of the two endpoint stimuli, [bɛ] and [dɛ], and the voiceless stops [p^hɛ] and [t^hɛ]. The place value of these four stimuli was cued in the same manner as the test stimuli, i.e., by variations in second and third formant transitions. Two other adapting stimuli were constructed which were phonetically identical to the [p^hɛ] and [t^hɛ] syllables. These stimuli, however, were synthesized by preceding the steady-state vowel portion with a 15 msec. burst of noise at a certain frequency and 15 msec. of silence rather than by varying formant transitions. Diehl argued that these stimuli shared phonetic features with some of the test stimuli but had virtually nothing in common with them acoustically. Presumably, shifts obtained in the transition-cued [bɛ] - [dɛ] continuum following adaptation with the burst-cued [p^hɛ] or [t^hɛ] could be accounted for only by phonetic similarity. Results provide support for at least some component of adaptation operating at the phonetic level of processing. Adaptation with [bɛ], [dɛ] and the transition-cued [p^hɛ] produced the predicted shifts in the [bɛ] - [dɛ] boundary. The transition-cued [t^hɛ] produced no reliable shifts. Four of the six subjects, however, perceived the stimulus as [p^hε] instead of [t^hε]. These subjects showed reliable shifts in the [bε] - [dε] boundary toward [bε]. This interesting, though accidental result, seems to indicate that the perceived phonetic character of an adapting stimulus has a greater effect on the direction of boundary shifts than its actual acoustic character. A similar result was obtained with the burst-cued [p^hε] which produced no significant boundary shifts. Two of the subjects did not hear this stimulus as [p^hε] and showed shifts toward [dε]. The four subjects who perceived the stimulus as [p^hε], however, showed reliable shifts toward [bε]. The burst-cued [t^hε] was perceived in the expected manner by all the subjects as [t^h] and resulted in reliable shifts toward [dε].

Diehl concluded that since the burst-cued [p^hε] was acoustically dissimilar to the adapted members of the test series which were transition-cued, the phonetic boundary shifts could be accounted for only by feature detectors which extract phonetic information. Recently, however, Pisoni and Tash (1975) have taken issue with Diehl's assumption that the burst-cued adaptors have virtually nothing in common acoustically with the transition-cued test series. Pisoni and Tash maintain that Diehl's results could be accounted for by assuming that both bursts and formant transitions indicate the initiation of rapid spectral change at a particular frequency (cf. Stevens, 1973). In this way, bursts and transitions are seen not as widely different acoustic events but as somewhat analogous in the sense that they both involve rapid spectral change.

Diehl's phonetic conclusion, however, is further supported by the accidental finding that the subject's assignment of a stimulus to a phonetic category had an effect on the direction of phonetic boundary shifts which seemed to override the actual acoustic character of the stimulus. This finding should be interpreted with some caution, however, since it was not the result of controlled manipulation of variables by the experimenter.

Although Diehl's results may provide evidence for some component of adaptation operating at the phonetic level of processing, an experiment by Ades (1974b) indicates that the application of a phonetically-based feature detector model is clearly limited. Ades examined the effect of adaptation with a CV syllable on a VC test series, as well as adaptation with a VC syllable on a CV test series. Two test series were constructed, one ranging from [bæ] to [dæ] and the other ranging from [æb] to [æd]. The endpoint syllables, [bæ], [dæ], [æb] and [æd], were presented as adapting stimuli. Since the information specifying place in initial consonants is acoustically quite different from that specifying place in final consonants, any crossed-series adaptation could be attributed only to the fatigue of phonetic detectors. Therefore, if adaptation operates at a phonetic level, the repeated presentation of any syllables containing, for example, the consonant [b], should result in shifts toward [b] in any continuum containing that segment, regardless of its position in the syllable. Thus, adaptation with [æb] should result in phonetic boundary shifts in the direction of [b] in a [bæ] - [dæ] continuum as well as an [æb] - [æd] continuum. The results, however, did not support this prediction. Adaptation with either [bæ] or [dæ] produced the predicted shifts in the [bæ] - [dæ] continuum, but did not affect the perception of the [æb] - [æd] continuum. Similarly, adaptation with either [æb] or [æd] resulted in shifts in the [æb] - [æd] continuum but not in the [bæ] - [dæ] continuum. Although not entirely incompatable with a phonetically-based model, this result clearly specifies the limits of such a model. If adaptation is operating at the phonetic level, as is suggested by Diehl's study, place analyzers must process information for initial and final consonants separately (Cooper, 1975).

Another possibility is that the Ades' results reflect the operation of feature detectors which are sensitive to individual acoustic features rather than phonetic distinctive features. Since the acoustic information specifying place in a CV syllable is markedly different from that specifying place in a VC syllable, no adaptation of acoustic feature detectors would be expected. Results of an earlier experiment by Ades (1973), however, cannot be accounted for by acoustic detectors alone. Ades examined the effect of adaptation with the first 38 msec. segment of the syllables [bæ] and [dæ] on a [bæ] - [dæ] test series. Although Eimas, Cooper and Corbit (1973) were unable to demonstrate perceptual shifts on a voicing series following adaptation with a d-chirp, Ades pointed out that place distinctions, unlike voicing, do not rely on the relative onset times of fairly sustained formants. Ades maintained that all of the acoustic information specifying place is contained in the initial formant transitions. Adaptation with the b- and d-chirps resulted in significant phonetic boundary shifts in the [bæ] - [dæ] boundary, although these shifts were significantly less than those typically obtained with the full syllables. However, listeners found it relatively easy to identify the chirps as speech-like, suggesting that they may have been processed to a phonetic stage of analysis. It was therefore unclear whether the adapted mechanism operated on acoustic or phonetic information. A more stringent test was provided by an additional condition in which adaptation was attempted with stimuli which sounded even less like speech. These stimuli consisted of only the second and third formant transitions with the first formant omitted entirely. These stimuli, which sound like "tweets" rather than chirps, represent exactly that information necessary to make place distinctions. Significant phonetic shifts were also observed with these stimuli, although the effect was considerably less than that obtained with either syllables or chirps.

These results clearly do not suggest that detectors are operating to extract acoustic features exclusively. Since each stimulus contained identical acoustic information with respect to place distinctions, adaptation resulting from the fatigue of acoustic detectors should have been equal for syllables, chirps and tweets. A particularly difficult finding to account for by means of an acoustic model is the greater effectiveness of chirps as compared to tweets. The only difference between these stimuli was the absence of the first formant in the tweet. Since the first formant presumably adds little information with regard to place, the smaller effects with the tweet cannot be accounted for by a decrement in acoustic feature detector fatigue. There are two possible ways of viewing this finding. First, it could be assumed that adaptation is exclusively phonetic, and further, that phonetic adaptation will increase as the phonetic feature contained in the adapting stimulus approximates that of the adapted members of the test series. Viewing phonetic features as relative events accounts for Ades' results since syllables, chirps and tweets may be seen as representing, respectively, strong, moderate and weak exemplars of the feature. An exclusively phonetic model, however, is in conflict with much of the preceding data (Cooper, 1974b; Ades, 1974b) which suggest that some component of adaptation operates at the acoustic level. Perhaps Ades' results can best be accounted for by viewing adaptation as a multi-component process (Cooper, 1974a) in which part of the adaptation effect reflects fatigue at the acoustic level while another part reflects fatigue at the phonetic level. Adaptation with syllables, then, can be seen as reflecting fatigue of both acoustic and phonetic detectors. Since the tweets were the least speech-like stimuli, they may be seen as reflecting the fatigue of acoustic detectors exclusively. Since subjects perceived the chirps as b-like or d-like, adaptation with these stimuli may reflect acoustic adaptation and some phonetic adaptation. It is also possible that the greater effectiveness of the syllable, compared with the chirp, reveals the operation of a higher level "phonetic unit" detector, suggested by Cooper (1974a) which operates on the consonant sound as a unit rather than on individual phonetic features.

Several recent studies support the notion that some component of adaptation operates at the acoustic level of analysis. An experiment by Bailey (1973) tested for the effects of adaptation when the information specifying place in the adapting stimulus differed from that in the test syllables. Two sets of syllables were constructed ranging from

[ba] to [da]. In one series, the second formant was fixed and all information specifying place was carried by the third formant transition. In the other series, no third formant was present and all information specifying place was carried by the second formant transition. Bailey tested for crossed-series adaptation under two conditions: (1) adaptation with [ba] or [da] cued by second formant transitions and tested on a [ba] - [da] series cued by third formant transition and (2) adaptation with [ba] or [da] cued by third formant transitions and tested on a [ba] - [da] series cued by second formant transitions. Results showed that adaptation with the second formant-cued syllables shifted phonetic boundaries for the series cued by third formant transitions. However, no crossed-series adaptation was obtained when the adaptor contained third formant place cues and the test series contained second formant cues. The finding that crossed-series adaptation was obtained in only one direction argues strongly for an acoustic interpretation of selective adaptation and suggests that phonetic boundary shifts may be predicted by the amount of spectral similarity between the adaptor and the test series.

This notion finds support from the results of a recent study by Tartter (1975). In this study the effect of adaptation was examined on both place and voicing continua when the adapting stimuli contained varying amounts of acoustic information relevant to the particular phonetic distinction being examined. Results indicated that the fewer the acoustic cues shared between the adapting and test stimuli, the smaller the adaptation effect.

A recent study by Pisoni and Tash (1975) also supports the notion that at least part of the adaptation effect is the result of feature detector fatigue at the acoustic level. In order to test for adaptation at the acoustic level, stimuli were constructed in which the same acoustic information was represented in different serial positions in the adapting and test stimuli. The adapting stimuli consisted of the first 50 msec. transitions of either [ba] or [da] (i.e., a chirp) preceded by a steady-state vowel fixed equal to the starting frequencies of the transitions. It was reasoned that phonetic adaptation was unlikely with the "speech-embedded chirps". This assumption was based on Ades' (1974b) finding that adaptation with a phonetic segment in the final position was unable to produce phonetic boundary shifts on a test series containing the same phonetic segment in the initial position. Any adaptation, then, would be seen as reflecting fatigue at the acoustic level only. Adaptation with the speech-embedded chirps resulted in small but significant phonetic boundary shifts toward the phonetic category from which the adaptor's formant transitions were originally obtained. Thus, since the adapting stimuli preserved the acoustic properties specifying place but did not supply any relevant phonetic information, the shifts were seen as the result of fatigue at the acoustic level. Another interesting finding was that the magnitude of the shifts obtained with the speech-embedded chirps was similar to that obtained by Ades (1974b) with tweets. Recall that these stimuli were also seen as producing adaptation at the acoustic level, suggesting that the two stimulus types are operated upon by the same perceptual mechanism.

Although the acoustic-phonetic issue is far from resolved, it is possible to draw some preliminary conclusions. Strong support for at least some component of adaptation operating at the acoustic level has been obtained in a number of studies (Cooper, 1974b; Cooper and Blumstein, 1974; Ades, 1973, 1974b; Bailey, 1973; Tartter, 1975; Pisoni and Tash, 1975). Additional support was obtained for adaptation at the phonetic level (Cooper, 1974a; Diehl, 1975; Ades, 1974b). Several investigators, notably Cooper (1974a) and Ades (1974b), have suggested that much of the seemingly contradictory data on selective adaptation to speech can be accounted for by assuming that part of the adaptation effect results from the fatigue of feature detectors operating to extract acoustic information and part of the effect results from the fatigue of feature detectors operating to extract phonetic information. Although it is not possible at present to rule out the possibility that selective adaptation effects are exclusively the result of the fatigue of acoustically-based detectors, the data at present seem to suggest that adaptation is a multi-component process which reflects the fatigue of detectors operating to extract both acoustic and phonetic features.

The present study was conducted in order to provide some information about the effect of low-level auditory variables on selective adaptation to speech sounds. Phonetic boundary shifts were obtained as a function of the intensity and number of repetitions of the adapting stimulus. If variations in the intensity of the adaptor influence the magnitude of the obtained shift, additional support would be provided for an account of adaptation effects in terms of low-level energy relations between adaptor and test stimuli. Conversely, the absense of an intensity effect would tend to support a phonetic interpretation based only on a interaction at an abstract feature level. Since variations in intensity do not alter any of the phonetic features of the adapting stimulus, a phonetically-based detector model would not predict corresponding variations in the magnitude of phonetic boundary shifts.

The number of repetitions of the adaptor was manipulated in order to determine the range over which adaptation effects could be obtained. Effects due to the number of repetitions could be the result of adaptation operating at both auditory and phonetic levels. An experiment by Bailey (1974) examined the relationship between the number of adaptor repetitions and the magnitude of phonetic boundary shifts. Six subjects listened to the adapting syllables [ba], [da] or [ga] repeated 8, 16, 24 or 32 times per adaptation sequence. Adaptation sequences were presented in blocks of 50 yielding total number of repetitions per block of 400 (8 per sequence), 800 (16 per sequence), 1200 (24 per sequence) and 1600 (32 per sequence). Subjects received two blocks of trials in each of the four number of repetitions condition during an adaptation session. Adapting syllables were presented to one ear and test syllables were presented for identification to the other ear with one test syllable following each adaptation sequence. Results indicated that there was no systematic relationship between the number of adaptor repetitions and the magnitude of phonetic boundary shifts.

These results, however, present some problems in interpretation since all four conditions of repetition were presented during a single adaptation session. Although each trial block (representing a number of repetitions condition) was separated by a rest period, the time course of adaptation to speech is unknown. Thus, it is possible that adaptation effects are additive from one trial block to another and further, that the effects in Bailey's study were maximal after the first few trial blocks. If this were the case, adaptation on one trial block would not be independent of adaptation on previous trial blocks. It would therefore be difficult to determine the effect of one trial block independently of those which preceded it.

The repetitions variable was examined in the present study in order to determine the relationship between the number of adaptor repetitions and the magnitude of phonetic boundary shifts (1) with independent groups of subjects (so as to eliminate the possibility of an order effect), (2) with a larger number of subjects, (3) with a wider range between the lowest and highest number of repetitions conditions and (4) with both adapting and test stimuli presented binaurally.

In summary, the purpose of this study was to examine the effect of low-level auditory variables on selective adaptation to speech sounds. Specifically, the following questions were asked:

1. Does the magnitude of phonetic boundary shifts increase as a function of increases in the intensity of the adapting stimulus?

2. Does the magnitude of phonetic boundary shifts increase as a function of increases in the number of repetitions of the adapting stimulus?

3. What is the range over which significant phonetic boundary shifts can be obtained?

CHAPTER III

GENERAL METHODOLOGY

Stimuli

All of the stimuli were three-formant synthetic CV syllables prepared on the Haskins Laboratories' parallel resonance speech synthesizer and recorded on audio tape for later playback.

The test stimuli consisted of a series of seven CV syllables, 300 msec. in duration, ranging perceptually from [ba] to [da]. The stimuli differed from one another in the direction and extent of second and third formant transitions. The starting frequencies for the seven syllables are presented in Table 1. The first formant started at 412 Hz and was the

Insert Table 1 about here

same for all seven syllables. All transitions were 50 msec. in duration and linear. The final 250 msec. of the syllables consisted of three steady-state formants appropriate for the English vowel [a]. These formants were centered at 769 Hz, 1232 Hz and 2525 Hz.

The adapting stimulus was either the endpoint [ba] or the endpoint [da] of the test series.

TABLE 1

Starting Frequencies of the
Second- and Third-Formant Transitions
for the Synthetic CV Test Stimuli

	Starting Frequencies (Hz)	
Stimulus	F2	F3
1	996	2180
2	1075	2348
3	1155	2525
4	1232	2694
5	1312	2862
6	1386	3026
7	1465	3195

Apparatus

All stimuli were recorded on audio tape and reproduced on an Ampex AG-500 two-track tape recorder and presented binaurally through matched and calibrated Telephonics (TDH-39) headphones. The adapting and test stimuli were recorded on separate channels of the tape and were calibrated separately. For the test stimuli, the gain of the tape recorder playback was adjusted to yield a voltage across the headphones equivalent to 80 dB SPL (re: .0002 dynes/cm^2) for a vowel-like [a] calibration signal. The adapting stimuli were varied by means of decade attenuators (Daven; Models 2511, 2513) to yield voltages equivalent to either 65, 80 or 95 dB SPL. All measurements were made using a Hewlett-Packard VTVM (Model 400) prior to the presentation of each tape.

CHAPTER IV

PILOT INVESTIGATIONS

In order to determine if intensity and repetition effects could be obtained when each variable was manipulated separately, four pilot studies were conducted. Two experiments examined the effect of a decrease in the intensity of the adaptor on phonetic boundary shifts, one using the adapting stimulus [ba] and the other using [da]. Two other experiments examined the effect of a decrease in the number of repetitions of the adaptor on phonetic boundary shifts. One study employed [ba] as the adaptor and the other used [da].

Subjects

Sixteen Indiana University undergraduates served as subjects. Subjects were paid $2.00 per hour for their participation in the experiment. All subjects were right-handed native speakers of English with no reported history of a speech, hearing or language disorder. None of the subjects had had any previous experience with synthetic speech or the selective adaption procedure.

Procedure

Baseline identification functions were first obtained for all four groups of listeners in the unadapted state by presenting subjects with 20 presentations of each of the seven test stimuli in random order. There were three seconds separating the syllables. Subjects were instructed to identify each syllable as "B" or "D" by writing the appropriate letter on prepared response sheets.

Two of the four groups then listened to the adapting syllable [ba] presented 100 times at 80 dB. The syllables in the adaptation sequence were separated by 225 msec. The adaptation sequence was followed by two seconds of silence and then the presentation of one of 14 different randomized orders of the seven test stimuli. The stimuli in the test sequence were separated by three seconds. After the seventh syllable was presented for identification, five seconds intervened before the next adaptation sequence was presented. This cycle of adaptation followed by the presentation of seven syllables for identification was repeated a total of 28 times for each session.

To summarize, an adaptation sequence consisted of 100 presentations of the syllable [ba] followed by the presentation of a random order of the seven test syllables for identification. The adaptation sequence was repeated a total of 28 times per session. Thus, the adapting syllable [ba] was presented to subjects a total of 2800 times.

The procedure for the remaining two groups of subjects was identical except that the adapting stimulus was [da] instead of [ba].

Subjects then returned 48 hours later for a second day of baseline identification and adaptation. For two of the groups, the intensity of the adapting stimulus (either [ba] or [da]) was presented at 65 dB instead of the original 80 dB. For the remaining two groups, the adapting stimulus was presented at 80 dB again, but the number of repetitions of the adaptor was decreased from the original 2800 times (100 per adaptation sequence) to 700 times (25 per adaptation sequence).

In summary, four groups of subjects with four subjects in each group participated in two-day experiments. Adaptation sequences for the four groups were as follows:

Group 1:	Day 1 -	2800 [ba] at 80 dB
	Day 2 -	2800 [ba] at 65 dB
Group 2:	Day 1 -	2800 [da] at 80 dB
	Day 2 -	2800 [da] at 65 dB
Group 3:	Day 1 -	2800 [ba] at 80 dB
	Day 2 -	700 [ba] at 80 dB
Group 4:	Day 1 -	2800 [da] at 80 dB
	Day 2 -	700 [da] at 80 dB

Results

Tables 2-5 show the individual and mean phonetic bound-

Insert Tables 2-5 about here

TABLE 2

Individual and Mean Phonetic Boundary
Loci for Group 1

2800 [ba] at 80 dB

Subject	Without Adaptation	Adaptation with [ba] at 80 dB	Shift
1	3.47	3.27	.18
2	4.88	3.47	1.41
3	3.53	2.71	.82
4	3.63	2.82	.83
$\overline{X}$	3.88	3.07	.81

2800 [ba] at 65 dB

Subject	Without Adaptation	Adaptation with [ba] at 65 dB	Shift
1	3.71	3.50	.21
2	3.23	2.70	.53
3	3.47	3.39	.08
4	3.47	3.32	.15
$\overline{X}$	3.47	3.23	.24

TABLE 3

Individual and Mean Phonetic
Boundary Loci for Group 2

2800 [da] at 80 dB

Subject	Without Adaptation	Adaptation with [da] at 80 dB	Shift
1	3.60	4.61	1.01
2	3.53	5.00	1.47
3	3.83	4.65	.82
4	4.31	6.24	1.92
$\overline{X}$	3.82	5.13	1.31

2800 [da] at 65 dB

Subject	Without Adaptation	Adaptation with [da] at 65 dB	Shift
1	3.67	4.47	.80
2	3.56	4.62	1.06
3	3.77	4.47	.70
4	4.40	5.53	1.13
$\overline{X}$	3.85	4.77	.92

TABLE 4

Individual and Mean Phonetic
Boundary Loci for Group 3

2800 [ba] at 80 dB

Subject	Without Adaptation	Adaptation with 2800 [ba]	Shift
1	4.50	3.86	.64
2	3.59	2.88	.73
3	3.71	3.50	.21
4	3.58	3.13	.45
$\overline{X}$	3.85	3.34	.51

700 [ba] at 80 dB

Subject	Without Adaptation	Adaptation with 700 [ba]	Shift
1	4.00	3.44	.56
2	4.47	4.13	.34
3	3.53	3.38	.15
4	3.50	3.07	.43
$\overline{X}$	3.88	3.51	.37

TABLE 5

Individual and Mean Phonetic Boundary Loci for Group 4

2800 [da] at 80 dB

Subject	Without Adaptation	Adaptation with 2800 [da]	Shift
1	3.73	6.09	2.36
2	3.93	5.26	1.25
3	3.52	4.58	1.06
4	4.54	6.31	1.97
$\bar{X}$	3.93	5.56	1.66

700 [da] at 80 dB

Subject	Without Adaptation	Adaptation with 700 [da]	Shift
1	4.00	5.46	1.46
2	4.18	5.50	1.32
3	3.59	4.38	.79
4	5.09	6.36	1.27
$\bar{X}$	4.22	5.43	1.21

ary loci for listeners in the unadapted state, following adaptation and the difference between boundary loci before and after adaptation. Boundary loci were computed by linear interpolation, yielding an estimate of the point along the stimulus continuum which would receive 50% [ba] responses and 50% [da] responses.

Subjects in all conditions showed a shift in the locus of the phonetic boundary of the [ba] - [da] continuum in the predicted direction. Subjects in Groups 2 and 4 who listened to the adapting stimulus [da] exhibited greater shifts ($\overline{X}$ = 1.28) than subjects in Groups 1 and 3 who listened to the adapting stimulus [ba] ($\overline{X}$ = .48). Subjects in Group 1 (Table 2), for which intensity was manipulated, exhibited greater phonetic boundary shifts when the adapting stimulus [ba] was presented at 80 dB ($\overline{X}$ = .81) than when the stimulus was presented at 65 dB ($\overline{X}$ = .24). This finding was consistent for all subjects except subject number 1. An intensity effect was also found for subjects in Group 2 (see Table 3). All four subjects showed greater shifts when adapting stimulus [da] was presented at 80 dB ($\overline{X}$ = 1.31) than when it was presented at 65 dB ($\overline{X}$ = .92).

The data for Groups 3 and 4 indicate that the magnitude of phonetic boundary shifts also changes as a function of the number of repetitions of the adaptor. All subjects in Group 3 showed greater phonetic boundary shifts when the adapting stimulus [ba] was presented 2800 times ($\overline{X}$ = .51) than whe it was presented 700 times ($\overline{X}$ = .37). Similarly, subject in Group 4 showed greater shifts when [da] was presented 2800 times ($\overline{X}$ = 1.66) than when it was presented 700 time ($\overline{X}$ = 1.21). This was consistent for all but subject numb 3 who showed a slightly greater shift at 700 repetitions.

Discussion

The data from the four pilot studies give a prelimi indication that both of the relatively low-level stimulus energy variables, intensity and number of repetitions, af the magnitude of phonetic boundary shifts. In all but one instance, subjects showed greater phonetic boundary shift in the condition in which the adapting stimulus was prese at a higher intensity, and in all but one instance, subje showed greater shifts when the number of repetitions of th adaptor was greater. However, since the conditions in wh smaller phonetic boundary shifts were predicted always oc on the second day of adaptation, it is impossible to rule an order effect. That is, it is possible that the decrem in phonetic boundary shifts in the lower stimulus energy c ditions reflects a greater resistance to adaptation on the second day merely as a function of exposure to the adapta procedure on the first day. In order to eliminate the possibility of an order effect and to study these variable in greater detail, a more extensive experiment was conducted involving nine independent groups of subjects who listened to the adapting stimulus [ba] presented either 280, 700 or 2800 times at either 65, 80 or 95 dB.

CHAPTER V

MAIN EXPERIMENT

Subjects

One hundred thirty-five Indiana University undergraduates served as subjects. Participation in the experiment fulfilled a requirement for a course in introductory psychology. In addition, subjects were paid $1.00 for their participation. The subject selection criteria were the same as in the previous experiments.

Procedure

Procedures for baseline identification and adaptation were identical to those for the pilot studies. Nine separate groups of subjects listened to the adapting stimulus [ba] presented 280 (10 per sequence), 700 (25 per sequence) or 2800 times (100 per sequence) at either 65, 80 or 95 dB. Thus, the three levels of adapting stimulus intensity combined with the three levels of adaptor repetitions yielded nine experimental conditions. Fifteen subjects were run in each condition. Subjects were run in a small room in groups ranging from two to eight subjects.

Results

Table 6 shows percent identification of [b] for the

Insert Table 6 about here

TABLE 6

Percent Identification of [b] by Stimulus for Each of the Nine Experimental Groups in the Baseline Condition

Experimental Group	Stimulus 1	2	3	4	5	6	7
280 at 65 dB	99	98	94	59	4	1	0
280 at 80 dB	100	98	98	52	3	1	1
280 at 95 dB	99	98	89	46	13	3	1
700 at 65 dB	99	98	96	56	9	1	0
700 at 80 dB	100	99	96	25	18	0	0
700 at 95 dB	100	98	94	62	19	0	0
2800 at 65 dB	99	97	90	46	6	2	1
2800 at 80 dB	99	100	99	40	7	6	2
2800 at 95 dB	100	99	98	59	4	2	1

seven test stimuli for each of the nine groups in the unadapted state. This table indicates that subjects as a group generally identified stimuli 1, 2 and 3 as [b] and 5, 6 and 7 as [d] with the crossover from [b] to [d] occuring approximately at stimulus 4. Subjects as a group showed the relatively sharp identification functions typical of stop consonant synthetic continua.

Phonetic boundary loci were computed prior to and following adaptation with [ba]. Correlated t-tests were used to compare baseline phonetic boundary loci with boundary loci following adaptation for each of the nine groups. Table 7 shows pre- and post-adaptation phonetic boundary

---

Insert Table 7 about here

---

loci and the results of correlated t-tests for each of the nine groups. Pre- and post-adaptation boundary loci were found to differ significantly for all groups, indicating that significant adaptation effects were obtained in all conditions of intensity and number of adaptor repetitions.

Phonetic boundary shifts were calculated for each subject by subtracting the phonetic boundary locus following adaptation from the boundary locus prior to adaptation. Table 8 shows the mean phonetic boundary shifts and standard

TABLE 7

Tests of Significance for Shifts in Phonetic Boundary Loci Between Baseline and Adaptation Conditions for Each of the Nine Experimental Groups

Experimental Group	Baseline	Adaptation	t-value	p less than
280 at 65 dB	4.15	3.92	4.60	.001
280 at 80 dB	4.03	3.75	4.45	.001
280 at 95 dB	4.01	3.60	4.90	.001
700 at 65 dB	4.09	3.64	3.22	.005
700 at 80 dB	3.69	3.28	5.79	.001
700 at 95 dB	4.29	3.65	5.22	.001
2800 at 65 dB	3.95	3.49	4.88	.001
2800 at 80 dB	3.81	3.01	6.95	.001
2800 at 95 dB	4.16	3.14	10.72	.001

NOTE: All tests were one-tailed. Degrees of freedom for each contrast were 14.

Insert Table 8 about here

deviations for each of the nine groups. It can be seen that the general trend is for phonetic boundary shifts to become larger as the intensity of the adaptor increases and as the number of adaptor repetitions increases. A two-way analysis of variance for independent groups revealed significant main effects for both intensity ($F\ (2,126) = 7.85,\ p < .001$) and number of repetitions ($F\ (2,126) = 16.59,\ p < .001$). No significant interactions were found, indicating that the two variables were essentially operating independently to influence the magnitude of phonetic boundary shifts. A Tukey multiple range test indicated that levels one, two and three of both factors were significantly different from one another beyond the .05 level of confidence.

Figure 2 displays mean phonetic boundary shifts for

Insert Figure 2 about here

the three intensity conditions as a function of the number of adaptor repetitions. With the exception of the 65 dB condition, which showed virtually no change from 700 to 2800 repetitions, all groups showed an increase in phonetic boundary shifts as the number of adaptor repetitions increased.

TABLE 8

Mean Phonetic Boundary Shifts and Standard Deviations for Each of the Nine Experimental Groups

		Intensity (dB)		
		65	80	95
Number of Repetitions	280	.21 (.20)	.28 (.24)	.41 (.34)
	700	.45 (.54)	.41 (.28)	.64 (.47)
	2800	.46 (.34)	.80 (.44)	1.02 (.37)

NOTE: Each mean is based on the difference between baseline and adaptation phonetic boundary loci averaged over 15 subjects. Standard deviations are in parentheses.

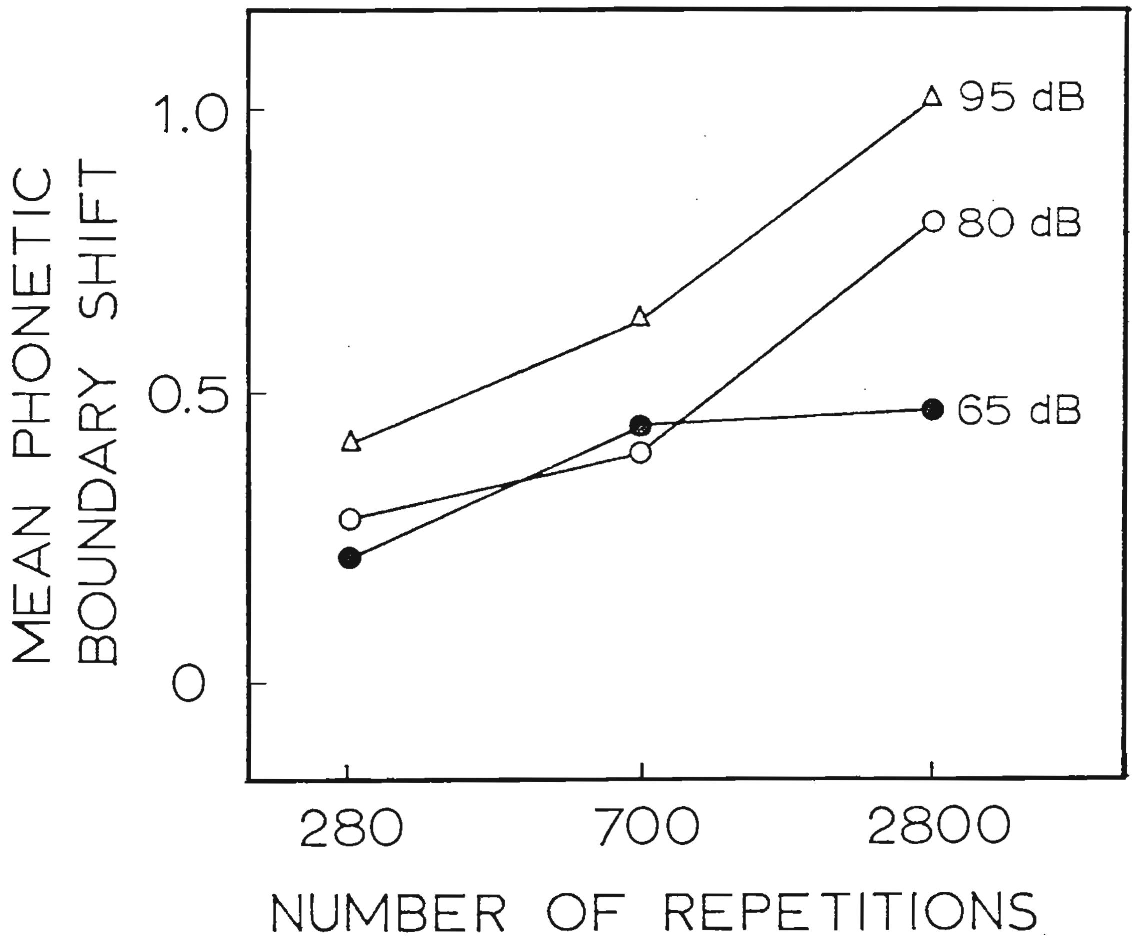


Figure 2. Mean Phonetic Boundary Shifts as a Function of the Number of Adaptor Repetitions.

Figure 3 displays mean phonetic boundary shifts for

Insert Figure 3 about here

---

the three conditions of repetition as a function of the intensity of the adaptor. This graph reveals that with the exception of the 700 repetitions condition, which showed a slight decrease from 65 to 80 dB, all groups showed an increase in phonetic boundary shifts as the intensity of the adaptor increased.

Discussion

The results of the main experiment were essentially in agreement with the findings of the pilot studies. The trend found in the pilot studies toward larger phonetic boundary shifts with greater values of intensity and number of repetitions was strongly confirmed in the main experiment. Further, since the experimental groups in the main experiment were independent, it is possible to rule out an order effect based on prior exposure to the adaptation procedure. The implications of these findings will be discussed in detail later, but in general they tend to support an account of adaptation based on low-level auditory variables, rather than an account based on the interaction of more abstract phonetic variables.

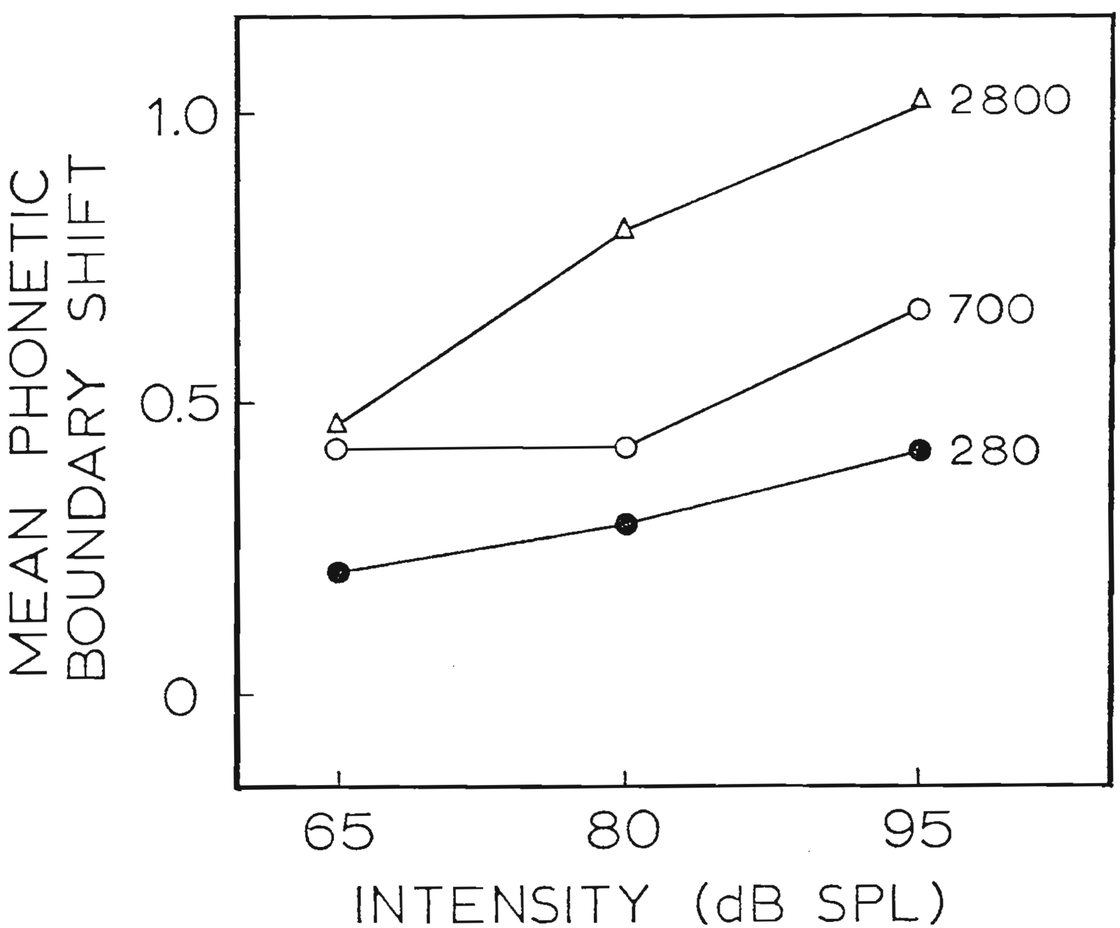


Figure 3. Mean Phonetic Boundary Shifts as a Function of Adapting Stimulus Intensity.

The finding that statistically significant phonetic boundary shifts occured in all conditions of intensity and repetition indicates that reliable adaptation effects can be obtained with as few as 280 adaptor repetitions, or 10 presentations per adaptation sequence. Although the shifts which occured in this condition were relatively small, they were highly significant ($p < .001$), suggesting that adaptation was not operating at a minimum. This finding strongly suggests the possibility that reliable adaptation effects could be obtained with many fewer adaptor repetitions; perhaps, as Bailey (1974) suggests, as few as one adaptor presentation per adaptation sequence. The repetition effects found in this study, however, are not in agreement with Bailey's (1974) finding that the magnitude of phonetic boundary shifts was unrelated to the number of adaptor repetitions. This discrepancy is somewhat difficult to interpret since there were several methodological differences between Bailey's study and the present experiment. First, adaptation sequences in Bailey's experiment were presented in blocks of 50, with one test syllable following each sequence. This differs from the procedure followed in the present study in which adaptation sequences were presented in blocks of 28, with 7 test syllables following each sequence.

Second, subjects in Bailey's experiment received adaptation under all four conditions of repetition in a single session. As was discussed previously, it is possible that adaptation on one trial block was affected by adaptation on previous trial blocks. In light of the possibility that adaptation effects are additive from one trial block to another, it is possible that the result of this would be to artificially increase phonetic boundary shifts on later trials. This would tend to obscure a repetitions effect. Since each subject in the present study received adaptation under only one condition, an order effect of this type can be ruled out.

Another possible source of the discrepancy is the fact that adapting stimuli in the Bailey study were presented to one ear and test syllables were presented to the other. In the present study, both adapting and test stimuli were presented binaurally. As will be discussed in detail later, it is possible that at least one component of both the intensity and repetition effects reflects the operation of a component which operates on unfused information from both ears. If this is the case, then monaural presentation would not be sensitive to the effect. Since the effects of a component which operates on unfused information would be restricted to the ear to which the adapting stimuli are presented, these effects could not be tested with the adapting and test stimuli presented to opposite ears.

Another important finding of the main experiment was the absence of a significant interaction between intensity and repetitions. This finding indicates that the two variables were essentially operating independently to influence the magnitude of phonetic boundary shifts. The independence of the two variables is a possible indication that intensity and repetitions may operate to influence the magnitude of phonetic boundary shifts in a "trade-off" relationship based on total stimulus energy. In other words, the magnitude of phonetic boundary shifts is dependent, in part, upon total adapting stimulus energy which may be increased (1) by increasing intensity or (2) by increasing the number of adaptor repetitions. Further, given a boundary shift of a certain magnitude, obtained x repetitions at y intensity, a trade-off relationship would imply that a shift of similar magnitude could be obtained by increasing x by a certain amount and decreasing y by a certain amount, or vice versa. However, since the repetitions effect may be the result of increases in the number of presentations of a phonetic feature, rather than (or perhaps in addition to) increases in total stimulus energy, it is not possible to conclude that the basis of the proposed trade-off is total stimulus energy.

To summarize, the principal findings of the main experiment were:

1. The magnitude of phonetic boundary shifts tended to become larger as adapting stimulus intensity increased.

2. The magnitude of phonetic boundary shifts tended to become larger as the number of adaptor repetitions increased.

3. Significant phonetic boundary shifts were found under all conditions of intensity and repetitions.

4. Intensity and number of repetitions were found to operate independently to influence the magnitude of phonetic boundary shifts.

CHAPTER VI

GENERAL DISCUSSION

The finding that the relatively low-level stimulus energy variables, intensity and number of repetitions, affected the magnitude of phonetic boundary shifts argues for at least one component of adaptation operating at an early stage of perceptual analysis, prior to phonetic processing. These results, then, provide support for an auditory or acoustic account of adaptation, rather than an account based on the interaction of more abstract phonetic features. The intensity effect is especially difficult to explain by a phonetic account. Since varying the intensity of the adaptor alters only its acoustic characteristics and none of its phonetic characteristics, an account of adaptation based on the extraction of phonetic distinctive features would not predict variations in the magnitude of phonetic boundary shifts with variations in intensity. In this regard, the intensity effect is somewhat analogous to Ades' (1974b) finding that adaptation with a segment in final position was unable to produce phonetic boundary shifts on a test series containing the same phonetic segment in the initial position, and vice versa. The purpose of Ades' study was to vary acoustic features of the adaptor while keeping phonetic features constant. A phonetic account does not predict variations in adaptation effects since no phonetic features have been varied; an acoustic account predicts variations. The intensity effect is similar in principle to the Ades' finding since variations in adaptation effects were found to result from variations in auditory variables in the absence of variation of phonetic features.

The repetitions effect, however, is somewhat ambiguous on this point since increasing the number of adaptor repetitions could be seen as increasing adaptation at both auditory and phonetic levels of processing. It could be argued that increasing the number of adaptor repetitions resulted in larger phonetic boundary shifts by increasing the total amount of adapting stimulus energy, indicating an auditory account. However, if one views adaptation as operating at more than one level of perceptual analysis, as Cooper (1974a) suggests, it is possible to interpret the repetitions effect as the result of variations in the number of times a particular phonetic feature was presented, rather than variations in total stimulus energy. This view also suggests that part of the repetitions effect may be due to variations in total stimulus energy, affecting an auditory or acoustic component, and part may be due to variations in the number of presentations of a phonetic feature, affecting a higher level phonetic component.

Another problem in interpreting the repetitions effect, apart from the acoustic-phonetic issue, is that increases in the number of adaptor repetitions were inevitably accompanied by an increase in the amount of time required for an adaptation sequence. For example, an adaptation session for subjects who received 280 repetitions lasted approximately 22 minutes, while a session for subjects who received 2800 repetitions lasted approximately 42 minutes. It is difficult to determine what systematic effect longer presentation times would have, independent of greater number of repetitions, but its possible confounding effect cannot be ruled out in this study.

Another issue which concerns the interpretation of both the intensity and repetition effects concerns the problem, discussed earlier, of whether adaptation operates peripherally, centrally or both. Recall that while strong support was provided for one component of adaptation operating centrally (Eimas, Cooper and Corbit, 1973; Ades, 1974a), equally convincing evidence was provided indicating that part of the adaptation effect operates on unfused information from both ears (Ades, 1974a). This finding leaves open the possibility of a peripheral component. It is quite plausible, then, that the intensity and repetition effects may reflect variations in the amount of adaptation operating peripherally. That is, increasing intensity, for example, may result in an increase in the fatigue of peripheral auditory receptors, rather than central acoustic processors. In fact, the effects presented here are consistent with peripheral adaptation effects using simple non-speech stimuli, such as pure tones and wide-band noise (cf. Small, 1963). Studies in this area typically use subjects' changes in loudness perception following prolonged presentation of an auditory stimulus as a measure of adaptation. There is strong evidence to indicate that perceptual shifts become larger as a function of increases in either adapting stimulus intensity (Hood, 1950; Jerger, 1957) or increases in either the duration of a continuous sound (Egan, 1955) or the number of presentations of an intermittant sound (Carterette, 1955). These results suggest that the amount of sensory fatigue is related to the total amount of stimulus energy reaching the receptor which can be increased either by increasing the intensity of the adaptor or its duration. If the repetitions variable of the present study is accepted as being analogous to the duration variable of the peripheral adaptation experiments, at least a surface similarity emerges between intensity and duration effects with non-speech stimuli and the intensity and repetition effects found in this study. It is tempting to speculate, then, that the intensity and repetition effects reported here reflect a component of adaptation which operates peripherally, in a trade-off relationship, as a function of total stimulus energy.

Although this is perhaps a plausible explanation, the relationship between shifts in loudness perception used as a measure of adaptation in experiments with non-speech sounds, and shifts in identification responses used as a measure of adaptation in experiments with speech sounds, is unclear at this point. The basic problem here is to account for changes in identification responses as a result of changes in the sensitivity of auditory receptors. A possible explanation may be derived by examining the acoustic characteristics of the adapting and test stimuli. It was noted earlier that distinctions along a place dimension may be made, in part, as a result of abstract frequency loci characteristic of certain consonants (DeLattre, Liberman and Cooper, 1955). Inspection of Table 1 reveals that this frequency locus is lower for the adapting stimulus [ba] (stimulus 1) than for the other stimuli. Repeated presentation of [ba], then, may result in a decrease in the sensitivity of auditory receptors tuned to these relatively low frequencies. Further, since peripheral adaptation effects spread to adjacent frequencies (Thwing, 1955), receptors sensitive to intermediate frequency loci might also be less sensitive. Therefore, when an ambiguous test stimulus is presented, i.e., one with an intermediate frequency locus, more information will be available from the unadapted, higher frequency receptors. This would have the effect of decreasing the probability that the stimulus would be identified as [b] and, relatively, increasing the probability that it would be identified as [d].

A possible test of a peripheral account of the intensity and repetition effects would be to replicate the present study with adapting stimuli presented to one ear and test stimuli presented to the other ear. Since peripheral effects should be localized to the ear to which the adapting stimuli are presented, intensity and repetition effects should be much less apparent or not present at all under the monaural condition. Such a finding would support an account of these effects based on adaptation operating either peripherally or at some very early stage, prior to binaural fusion.

Regardless of the outcome of such a study, the strong presence of an intensity effect supports the notion that at least some component of adaptation operates at an early stage of perceptual analysis, prior to phonetic processing. It should be noted, however, that some of the data on selective adaptation to speech cannot be accounted for solely on the basis of low-level acoustic detectors. Especially difficult to explain on the basis of an exclusively acoustic interpretation are the study by Ades (1973) and a recent experiment by Cooper (1974c) on perceptuo-motor effects. Cooper showed that perceptual adaptation with [p^hi] produced shifts in subjects' characteristic VOT production values.

As well as providing some preliminary support for the hypothesized link between perception and production, this study presents evidence that is difficult to explain solely on the basis of passive detector mechanisms which operate in any simple input-output fashion to extract relatively low-level acoustic features (see also Cooper and Nager, 1975).

CHAPTER VII

SUMMARY AND CONCLUSIONS

Numerous recent studies, using the selective adaptation paradigm, have reported evidence for the possible existence of "feature" or "property" detectors in the perception of speech sounds. Although the original studies by Eimas and his colleagues, using this procedure, were interpreted as evidence for feature detectors operating to extract phonetic features in the Chomsky-Halle sense, a growing body of evidence indicates that these results may have an auditory or acoustic basis at a much earlier stage of perceptual analysis. In the present study, the intensity and number of repetitions of the adapting stimulus were varied in order to provide additional information about the locus and nature of selective adaptation to speech sounds.

Nine separate groups of subjects, with 15 subjects per group, identified syllables from a synthetically produced [ba] - [da] continuum before and after adaptation with the endpoint [ba]. The adapting syllable was presented 280, 700 or 2800 times at either 65, 80 or 95 dB. The results showed that the magnitude of the shift increases as the intensity of the adaptor increases and as the number of repetitions of the adaptor increases.

The intensity effect was interpreted as evidence in favor of at least one component of adaptation operating at an early stage of perceptual analysis, prior to phonetic processing. The repetitions effect, however, was seen as somewhat ambiguous in the sense that it could be explained in terms of either variations in total stimulus energy (supporting an auditory interpretation) or in terms of variations in the number of presentations of a phonetic feature (supporting a phonetic interpretation). It was noted that the intensity and repetition effects found in this study are comparable to intensity and duration effects in peripheral adaptation experiments using non-speech sounds. An explanation of these effects was offered based on the fatigue of peripheral auditory receptors in particular frequency loci.

It was further pointed out that the results of this study do not preclude the possibility of adaptation operating at a phonetic level of processing, as well as an auditory or acoustic level.

REFERENCES

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

Ades, A.E. Some effects of adaptation on speech perception. Quarterly Progress Report of the Research Laboratory of Electronics, M.I.T., 1973, 211, 121-129.

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 and Psychophysics, 1974, 16, 61-66. (b)

Bailey, P. Perceptual adaptation for acoustical features in speech. Speech Perception: Report on Research in Progress in the Department of Psychology, The Queen's University of Belfast, Northern Ireland, 1973, 2.2, 20-34.

Bailey, P.J. Procedural variables in selective adaptation to speech. Speech Perception, 2, Spring, 1974, Department of Psychology, The Queen's University of Belfast.

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

Capranica, R.R. The Evoked Vocal Response of the Bullfrog. Cambridge: M.I.T. Press, 1965.

Carterette, E.C. Perstimulatory auditory fatigue for continuous and interrupted noise. Journal of the Acoustical Society of America, 27, 103-111.

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 and Psychophysics, 1974, 201-204. (b)

Cooper, W.E. Perceptuomotor adaptation to a speech feature. Perception and Psychophysics, 1974, 16, 229-234. (c)

Cooper, W.E. Selective adaptation to speech. F. Restle, R.M. Shiffrin, N.J. Castellan, H. Lindman, and D.B. Pisoni (Eds.), Cognitive Theory, Volume I. Potomac, Maryland: Erlbaum, 1975.

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

Cooper, W.E. and Nager, R.M. Perceptuo-motor adaptation to speech: An analysis of bisyllabic utterances and a neural model. Journal of the Acoustical Society of America, 1975. (In Press).

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

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

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

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

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

Evans, E.F. and Whitfield, I.C. Classification of unit responses in the auditory cortex of the unanaesthetized and unrestrained cat. Journal of Physiology, 1964, 17, 476-493.

Fant, G. Auditory patterns of speech. In W. Watson-Dunn (Ed.), Models for the Perception of Speech and Visual Form, Cambridge: M.I.T. Press, 1967, 111-125.

Frishkopf, L. and Goldstein, M. Response to acoustic stimuli in the eighth nerve of the bullfrog. Journal of the Acoustical Society of America, 1963, 35, 1219-1228.

Goldstein, L.M. and Lackner, J.R. Alterations of the phonetic coding on speech sounds during repetition. Cognition, 1974, 2, 279-297.

Hood, J.D. Studies in auditory fatigue and adaptation. Acta Oto-Laryngologica, 1950, Supplement No. 92.

Hubel, D.H. and Wiesel, T.N. Receptive fields, binocular interaction and functional architecture in the cat's visual cortex. Journal of Physiology, 1962, 160, 106-154.

Hubel, D.H. and Wiesel, T.N. Receptive fields and functional architecture in two nonstriate visual areas (18 and 19) of the cat. Journal of Neurophysiology, 1965, 28, 228-289.

Jerger, J.F. Auditory adaptation. Journal of the Acoustical Society of America, 1957, 29, 357-363.

Lettvin, J.Y., Maturana, H.R., McCulloch, W.S. and Pitts, W.H. What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers, New York, 1959, 47, 1910-1951.

Liberman, A.M. The grammars of speech and language. Cognitive Psychology, 1970, 1, 301-323.

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

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

Mattingly, I.G., Liberman, A.M., Syrdal, A.K. and Hawles, T. Discrimination in speech and nonspeech modes. Cognitive Psychology, 1971, 2, 131-157.

Nelson, P.G., Erulkar, S.D. and Bryan, S.S. Response units of the inferior colliculus to time-varying acoustic stimuli. Journal of Neurophysiology, 1966, 29, 834-860.

Pisoni, D.B. and Sawusch, J.R. Some stages of processing in speech perception. In A.S. Cohen and S. Nooteboom (Eds.), Proceedings of the Symposium on Dynamic Aspects of Speech Perception, Springer-Verlag, 1975, (In Press).

Pisoni, D.B. and Tash, J. Auditory property detectors and processing place features in stop consonants. Perception and Psychophysics, 1975, In Press.

Small, A.M. Auditory adaptation. In J. Jerger (Ed.), Modern Developments in Audiology, 1963, Academic Press.

Stevens, K.N. Toward a model for speech recognition. Journal of the Acoustical Society of America, 1960, 32, 47-55.

Stevens, K.N. Potential role of property detectors in the perception of consonants. Quarterly Progress Report, 110, R.L.E., M.I.T. Cambridge, Massachusetts, 1973, 155-168.

Stevens, K.N. and Halle, M. Remarks on analysis by synthesis and distinctive features. In W. Watson-Dunn (Ed.), Models for the Perception of Speech and Visual Form, Cambridge: M.I.T. Press, 1967, 88-102.

Stevens, K.N. and House, A.S. Speech Perception. In J.V. Tobias (Ed.), Foundations of Modern Auditory Theory, New York, Academic Press, 1972, 1-62.

Stevens, K.N. and Klatt, D.H. Role of formant transitions in the voiced-voiceless distinction of stops. Journal of the Acoustical Society of America, 1974, 55, 653-659.

Studdert-Kennedy, M. Speech Perception. In N.J. Lass (Ed.) Contemporary Issues in Experimental Phonetics, C.J. Thomas, 1974.

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

Studdert-Kennedy, M., Shankweiler, D. and Pisoni, D.B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. Cognitive Psychology, 1972, 3, 455-466.

Tartter, V.C. Selective adaptation of acoustic and phonetic detectors. Unpublished M.A. Thesis, Brown University, June, 1975.

Thwing, E.J. Spread of perstimulatory fatigue to neighboring frequencies. Journal of the Acoustical Society of America, 27, 741-748.

Warren, R.M. and Gregory, R.L. An auditory analogue of the visual reversible figure. American Journal of Psychology, 1958, 71, 612-613.

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

Wollberg, Z. and Newman, J.D. Auditory cortex of squirrel monkey: Response patterns of single cells to species specific vocalizations. Science, 1972, 175, 212-213.

Wood, C.C. Levels of processing in speech perception: Neurophysiological and information-processing analyses. Unpublished Ph.D. dissertation, Yale University, 1973.

SHORT REPORTS AND WORK IN PROGRESS

Selective Adaptation Effects on End-point Stimuli

James R. Sawusch

Indiana University

Selective Adaptation Effects on End-point Stimuli

Recent models of the speech perception system have been proposed within an information processing framework. These models have generally proposed successive levels of auditory and phonetic analysis of the speech waveform. (Cooper, 1975; Pisoni, 1975; Pisoni & Sawusch, 1975; Tartter & Eimas, 1975). Although there has been some consensus on the distinction between auditory and phonetic levels of processing, the precise nature of the perceptual mechanism at each of these levels is still under intensive investigation. Abbs and Sussman (1971) proposed that models of speech perception be phrased in terms of selectively tuned feature detectors. Specifically, they suggested that these feature detectors would be tuned to such types of information as a change in frequency, high energy concentration at low frequencies, and so forth (Abbs & Sussman, 1971). These types of information have been found to be sufficient cues for the identification of certain speech sounds (see Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967 for review).

Evidence implicating the existance of feature detectors was first reported by Eimas and his co-workers (Eimas, Cooper & Corbit, 1973; Eimas & Corbit, 1973). Their studies, using the selective adaptation paradigm, demonstrated the presence of some form of feature detector in the perception of the voicing feature. Since these initial results, further research has implicated detector mechanisms in mediating the perception of the feature of place of production (Cooper, 1974a; Cooper & Blumstein, 1974; Pisoni & Tash, 1975; Tartter & Eimas, 1975) as well as the voicing feature (Cooper, 1974b; 1974c; Tartter & Eimas, 1975).

Concurrent with these new experimental results, models of the perceptual system have been proposed to account for these findings. Among the models proposed, those of Cooper and Nager (1975) and Tartter and Eimas (1975) are fairly specific. Both models propose successive stages of auditory and phonetic analysis. Both models also assume that selective adaptation operates to fatigue detectors operating at both auditory and phonetic levels. The evidence for selective adaptation affecting detectors at an auditory level of analysis is fairly strong (Ades, 1974; Baily, 1973; Pisoni & Tash, 1975; Tartter & Eimas, 1975). The evidence for selective adaptation operating at a phonetic level has not been as strong and interpretations of this evidence have diverged accordingly (Cooper, 1975; Pisoni & Tash, 1975; Tartter & Eimas, 1975).

Another assumption of both models is that selective adaptation is assumed to affect the entire range of a feature detector. However, except for a recent study by Miller (1975) there has been no direct support for this assumption. The general result of selective adaptation studies is that following adaptation, the phonetic boundary shifts toward the category of the adaptor. This shift in identification is mirrored by a shift in discrimination (Cooper, 1974a). However, the shifts for both identification and discrimination are confined to the phonetic boundaries. No systematic changes have been observed at the end-points of a test series for a standard adaptation paradigm.

Miller (1975) reasoned that no shift had been found near the endpoints of the test series because the identification paradigm was too insensitive. The _S_s task in a standard identification paradigm is simply to categorize the stimuli using one of two or three phonetic labels. Miller (1975) used a dichotic listening paradigm and investigated feature competition among good exemplars of stop CV syllables both before and after selective adaptation. The results indicated that features were less effective in competing after adaptation. Thus, these results indicate that selective adaptation does effect the processing of good exemplars of a category and not merely boundary stimuli.

If Miller's explanation of the failure of identification studies to pick up end-point shifts in selective adaptation is correct, then it should be possible to pick up end-point effects with a more sensitive identification paradigm. Glanzman and Pisoni (1973) used confidence ratings in conjunction with ABX and 4IAX discrimination paradigms. Their results indicated that Ss assignment of ratings was directly related to their accuracy of discrimination. In using confidence ratings, Glanzman and Pisoni tapped a source of information that previous discrimination experiments had neglected. Their results demonstrated that Ss could provide information about the stimuli beyond the simple discrimination response. From these results, it seemed plausible that the use of a rating scale, rather than a two category judgment, would be a more sensitive measure in an identification paradigm.

The purpose of the present experiment was twofold. First, could Ss use a rating scale in an identification task consistently. If Ss could use the rating scale, then would other stimuli, besides those at the phonetic boundary, show an effect due to selective adaptation.

Method

Subjects. Fourteen paid volunteers served as Ss. All were right-handed native speakers of English with no known history of a hearing or speech disorder. Ss responded to an advertisement in the student newspaper and were paid at the rate of $2.00 per hour. Ss were divided into two groups of seven Ss each.

Stimuli. The stimuli were three formant synthetic CV syllables that were prepared on the parallel resonance synthesizer at Haskins Laboratories. All stimuli were recorded on magnetic tape for later playback. The test stimuli consisted of one series of nine CV syllables that ranged perceptually from [bi] to [di] on the place feature. These stimuli varied in their starting frequencies for the second and third formant transitions. These initial frequencies are displayed in Table 1. The duration of the formant transitions was 50 msec, followed by a 250 msec steady state vowel ([i]). The vowel had formant center frequencies of 287, 2307, and 3026 hertz for the first through third formants.

Procedure. The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give a voltage across the headphones equivalent to 80 dB SPL for a steady state callibration vowel [a].

The experiment was conducted on two consecutive days. At the beginning of each day, Ss listened to two identification tapes. Each tape contained a different randomization of ten replications of each of the nine stimuli. Ss were told that they would hear synthetic speech sounds approximating the syllables [bi] and [di]. They were instructed to respond to the stimuli using a six point rating scale. A copy of the response scale was present in front of each S at all times. The rating scale is presented in Table 2. Ss recorded their rating response in prepared response booklets.

Immediately following the identification tapes an adaptation test was presented. One group of Ss received the adapting syllable [bi] and the other group received the syllable [di]. The adapting stimuli were the end-

Table 1

Starting frequencies of the second and third formant transitions for the synthetic CV series [bi] - [di]. The fixed steady-state formants were centered at: 286 Hz(F_1), 2307 Hz(F_2), and 3026 Hz(F_3).

Stimulus No.	Starting Frequencies (in Hz)	
	F_2	F_3
1	1465	2180
2	1541	2348
3	1620	2525
4	1695	2694
5	1772	2862
6	1845	3026
7	1920	3195
8	1996	3363
9	2078	3530

Table 2

Six point rating scale that _S_s used in responding to the test stimuli.

Response	Rating
1	Positive stimulus was a "b"
2	Stimulus sounded like a "b"
3	Unsure, stimulus could have been a "b"
4	Unsure, stimulus could have been a "d"
5	Stimulus sounded like a "d"
6	Positive stimulus was a "d"

point syllables from the test series. The adapting syllable was presented for one minute (100 repetitions with a 300 msec interstimulus interval). After each minute of adaptation, the nine test syllables were presented in random order for identification by Ss using the rating scale. Nine of these adaptation trials were run on each day. Thus, at the end of the experiment, each S had furnished 18 adapted responses to each of the nine stimuli.

Results & Discussion.

The data from the first identification test tape on the first day was treated as a practice run for Ss to accustom themselves to the rating scale. This data was eliminated from further analysis for all Ss.

The identification data for both groups is pooled together in Figure 1. Ss systematically assigned low ratings to the stimuli at the [bi] end of the series (stimuli 1-4) and high ratings to the stimuli at the [di] end of the series (stimuli 6-9). Stimulus 5 received an ambiguous rating of approximately 3.50. One other point to note is the upturn in the ratings at stimulus 1. It appears that the best exemplars of the [bi] category are stimuli 2 and 3. Stimulus 1 received a significantly higher rating than stimulus 2 ($t(13) = 2.654$, $p < .01$ for a one-tailed, correlated t-test). Thus, it appears that Ss can provide a reliable rating response in an identification paradigm. Further, the rating response appears to be more sensitive to differences between stimuli from the same phonetic category than the standard identification paradigm.

In order to compare the adaptation data from the present experiment with previous selective adaptation experiments, the six point rating scale data was collapsed into a two category scale. Responses 1, 2 and 3 were treated as "bi" responses. Similarly, responses 4, 5 and 6 were treated

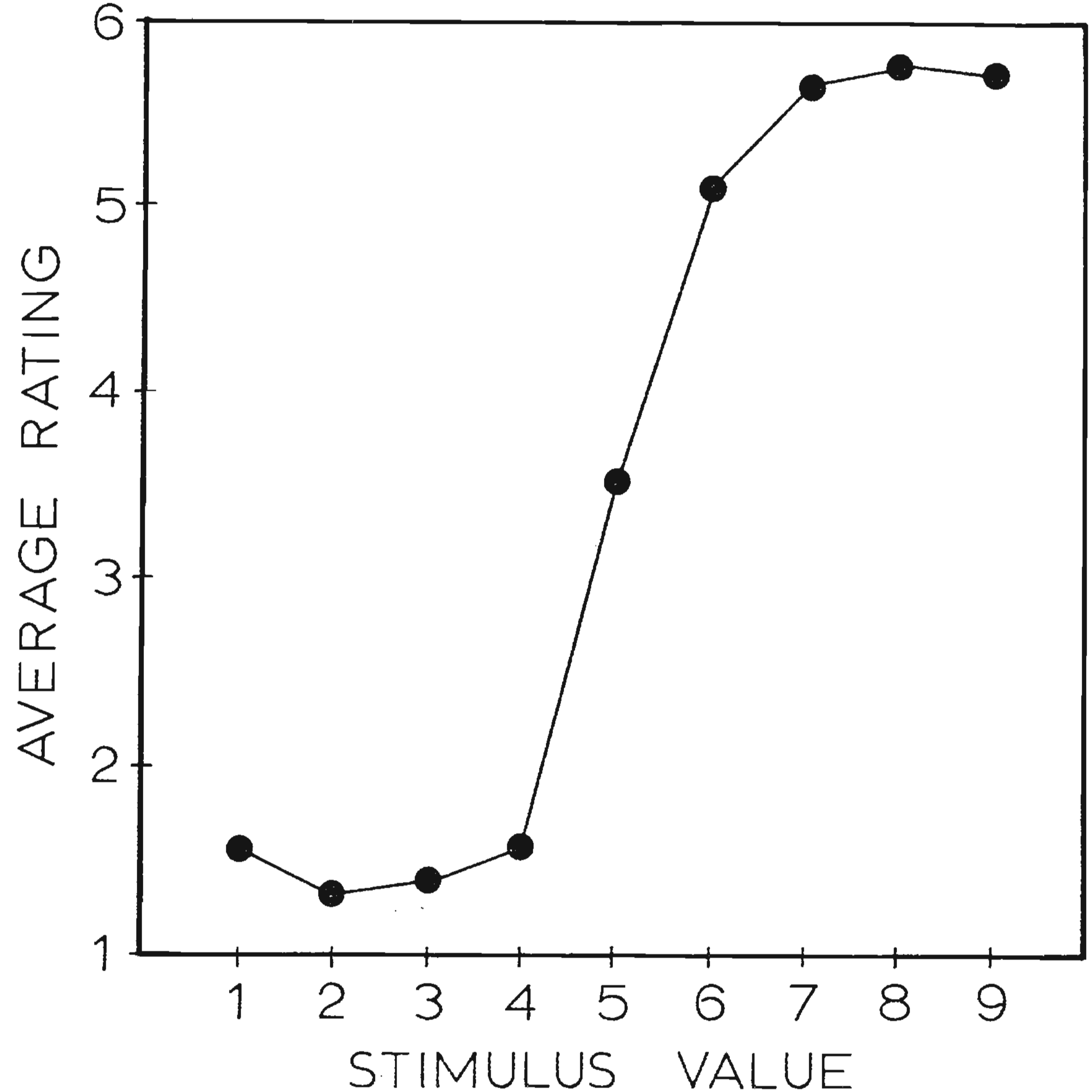


Figure 1. Average rating responses of all 14 Ss from groups one and two for the unadapted [bi]-[di] series.

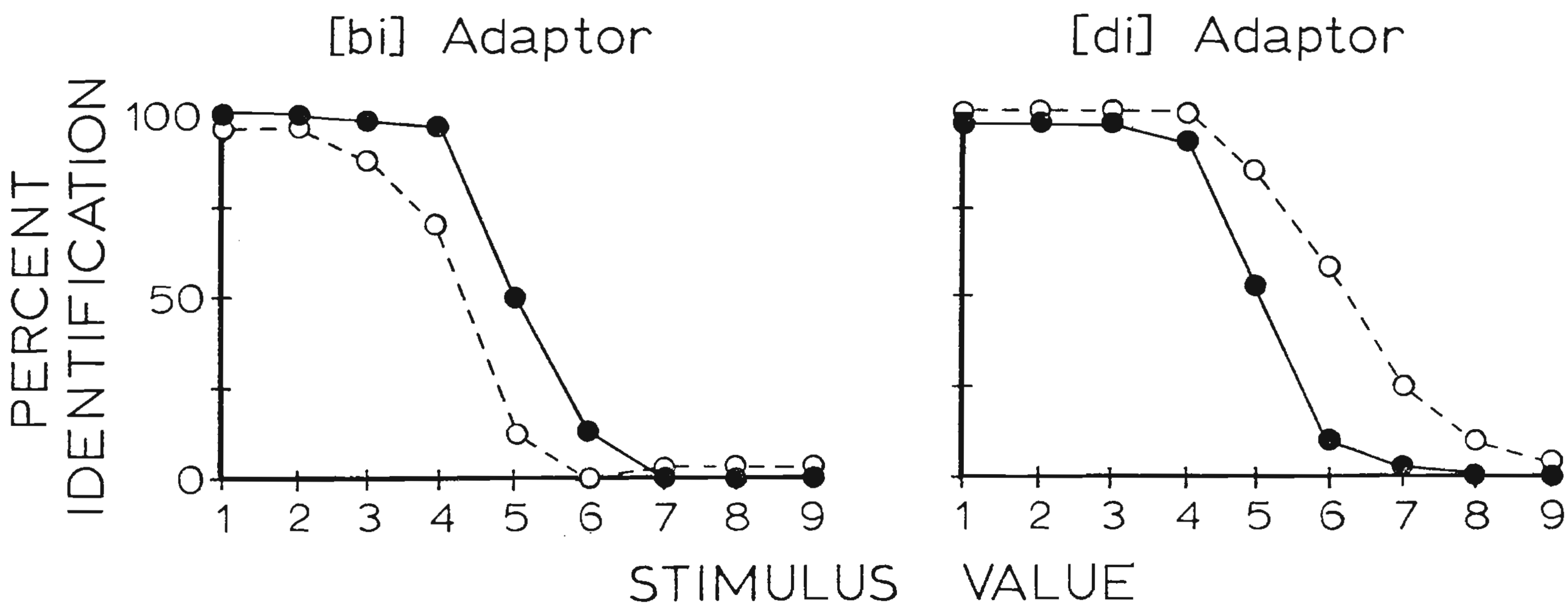


Figure 2. Unadapted (solid circles) and adapted (open circles) identification functions for the [bi]-[di] series. The six rating scale responses were colapsed into two category responses.

as "di" responses. The data for both groups, before and after adaptation, is plotted in Figure 2 as per cent of "bi" responses. The shifts in the phonetic category boundaries for both groups were highly significant ($t(6) = 5.14$, $p < .005$ for the [bi] adapted group; $t(6) = 7.45$, $p < .001$ for the [di] adapted group using one-tailed, correlated t-tests). These shifts are similar to those found previously for selective adaptation on a place series (Cooper, 1974a; Pisoni & Tash, 1975; Tartter & Eimas, 1975).

The average rating response for each of the nine stimuli before and after [bi] adaptation is displayed in the left hand panel of Figure 3. Since individual Ss showed the same trend as the group, only the group data is shown. The shift in the average rating for each of the first five stimuli was significant ($t(6) = 2.21$, $p < .05$ for stimulus 1; $t(6) = 2.02$, $p < .05$ for stimulus 2; $t(6) = 3.18$, $p < .01$ for stimulus 3; $t(6) = 3.32$, $p < .01$ for stimulus 4; and $t(6) = 5.00$, $p < .005$ for stimulus 5, all one-tailed, correlated t-tests). The shifts in average rating for each of stimuli 6 through 9 were not significant, although they were in the same direction as the shifts for stimuli 1 to 5.

The right-hand panel of Figure 3 displays the average rating responses before and after adaptation for the [di] adapted group. The shifts in the average rating for each of the last six stimuli were significant ($t(6) = 2.45$, $p < .05$ for stimulus 4; $t(6) = 4.36$, $p < .005$ for stimulus 5; $t(6) = 7.09$, $p < .001$ for stimulus 6, $t(6) = 6.39$, $p < .001$ for stimulus 7; $t(6) = 5.61$, $p < .001$ for stimulus 8; and $t(6) = 4.16$, $p < .005$ for stimulus 9, all one-tailed, correlated t-tests). The shifts in the average rating for each of stimuli 1, 2 and 3 were not significant, although they were in the same direction as the shifts for stimuli 4 through 9.

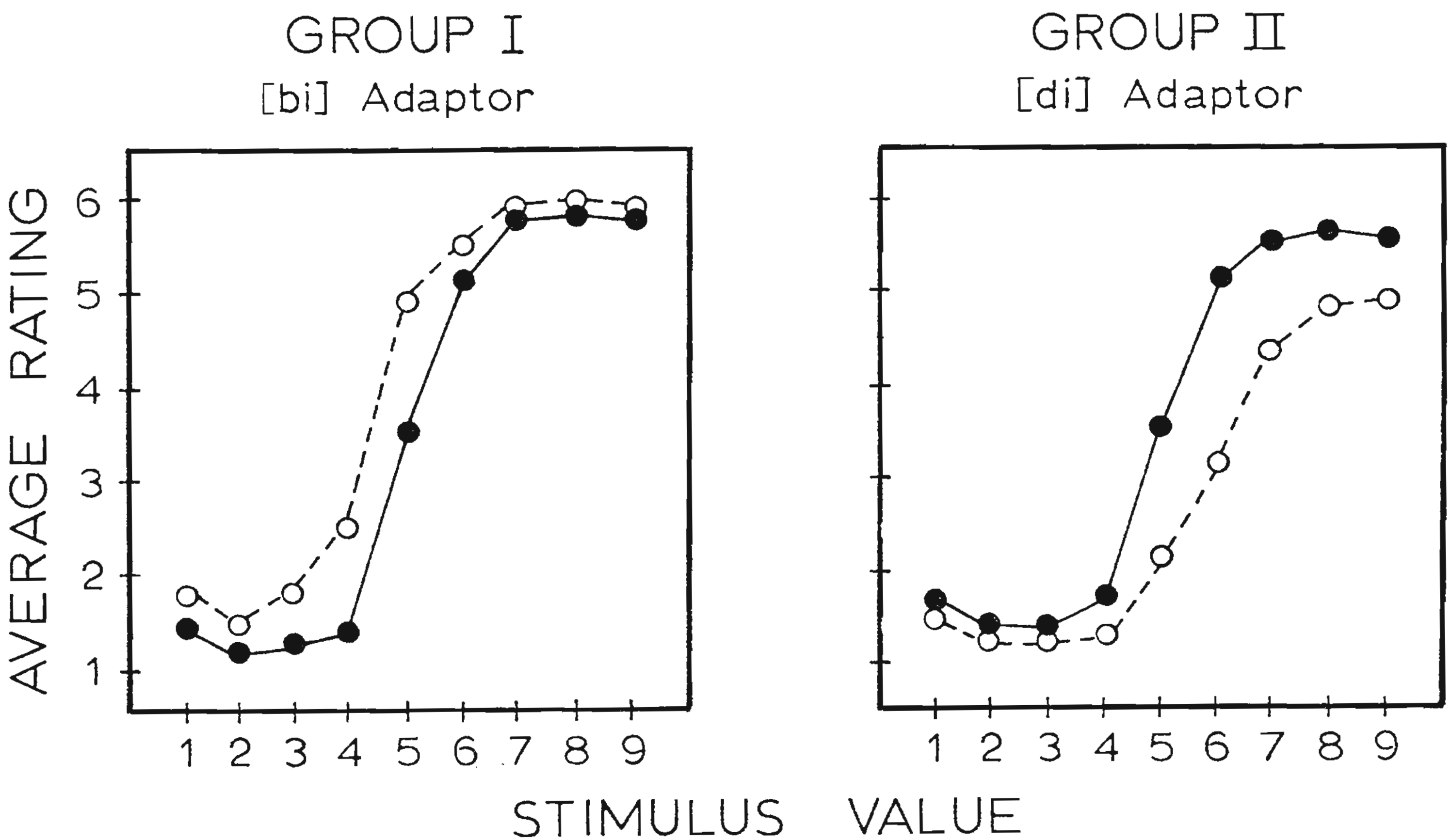


Figure 3. Unadapted (solid circles) and adapted (open circles) rating response functions. Group I (N=7) received the [bi] adaptor and Group II received the [di] adaptor.

These results clearly indicate that selective adaptation effects the entire category from which the adapting stimulus is drawn, and not just the boundary stimuli. Further, because the opposite category was relatively unaffected, two inferences can be drawn. First, any detector operating for the opposing category was unaffected by selective adaptation. Secondly, opposing detectors do not extend significantly into each others range. Rather, their response range seems to be sharply delimited. If there were a significant degree of overlap between opposing detectors, then the rating responses should have changed in the unadapted, as well as the adapted categories. The failure to find such a change in the unadapted category indicates that the opposing detectors do not overlap extensively.

This experiment does not answer the question as to the level at which selective adaptation takes place. The changes in rating response in the present experiment could be taking place entirely within an auditory level of processing (Pisoni & Tash, 1975) or they could reflect changes at both auditory and phonetic levels (Cooper, 1975; Tartter & Eimas, 1975). However, regardless of the processing level at which the adaptation effect takes place, information about the relative outputs of the detector mechanisms operating is preserved and transmitted to higher levels. If this type of information were not retained and used, then Ss would not have shown a change in their rating response as a function of adaptation and the rating responses for all stimuli from within the same phonetic category would have been identical.

In summary, the following conclusions seem justified in light of these results. First, Ss can effectively use a rating scale for responding in an identification task. Further, their responses indicate that all stimuli from within the same phonetic category are not treated equally. Information differentiating stimuli within a phonetic category is available to Ss and they use this information in making a rating response. Secondly, the present results agree with those of Miller (1975). Selective adaptation effects the entire category from which the adapting stimulus is drawn. The opposing category is relatively unaffected by the adaptation process. This is reflected in the change in Ss ratings of stimuli from the same phonetic category as the adaptor. Thus, the models of Cooper and Nager (1975) and Tartter and Eimas (1975) which assumed that adaptation affected the entire range of a detector, have been supported. Hopefully, further use of the rating scale in conjunction with selective adaptation paradigms will further illuminate the nature of the selective adaptation phenomena and the speech perception processes.

Acknowledgments

This research was supported by NINDS Research Grant No. 12179 to Indiana University. I would like to thank David B. Pisoni for his insight and suggestions during all phases of this work and Dr. F. S. Cooper for making the facilities at Haskins Laboratories available for stimulus preparation.

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. How phonetic is selective adaptation? Experiments on syllable position and vowel environment. Perception & Psychophysics, 1974, 16, 61-67.

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.

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. 1. Potomac, Maryland: Erlbaum Associates, 1975.

Cooper, W. E. & Blumstein, S. A "labial" feature analyzer in speech perception. Perception & Psychophysics, 1974, 15, 590-600.

Cooper, W. E. & Nager, R. M. Perceptuo-motor adaptation to speech: an analysis of bisyllabic utterances and a neural model. Journal of the Acoustical Society of America, 1975, 58, 256-265.

Eimas, P. D., Cooper, W. E. & Corbit, J. D. Some properties of linguistic feature detectors. Perception & Psychophysics, 1973, 13, 2, 247-252.

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

Glanzman, D. L. & Pisoni, D. B. Decision processes in speech discrimination as revealed by confidence ratings. Paper presented at the 85th meeting of the Acoustical Society of America, Boston, April, 1973.

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

Miller, J. L. Properties of feature detectors for speech: Evidence from the effects of selective adaptation on dichotic listening. _Perception & Psychophysics_, 1975, in press.

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_. Potomac, Maryland: Erlbaum Associates, 1975.

Pisoni, D. B. & Sawusch, J. R. Some stages of processing in speech perception. Paper presented at the Symposium on Dynamic Aspects of Speech Perception, Eindhoven, Netherlands, August, 1975.

Pisoni, D. B. & Tash, J. B. Auditory property detectors and processing place features in stop consonants. _Perception & Psychophysics_, 1975, in press.

Tartter, V. C. & Eimas, P. D. The role of auditory and phonetic feature detectors in the perception of speech. _Cognitive Psychology_, 1975, in press.

STAGES OF PROCESSING IN SPEECH PERCEPTION: FEATURE ANALYSIS

David B. Pisoni

Department of Psychology

Indiana University

Bloomington, Indiana 47401

U. S. A.

This paper was presented at the Eighth International Congress of Phonetic Sciences, Leeds, England, August 17-23, 1975 and is to be published in the Proceedings of the Congress.

STAGES OF PROCESSING IN SPEECH PERCEPTION: FEATURE ANALYSIS*

David B. Pisoni

Indiana University

Bloomington, Indiana

U. S. A.

Recently a number of investigators have begun to use the information processing approach to study speech perception. A basic assumption of this approach is that perception is viewed as a hierarchically organized sequence of events involving stages of storage and various transformations of information over time.

Figure 1 shows the components of a model of the earliest stages of

---

Insert Figure 1 about here

---

speech perception. Auditory input enters the system and is processed in progressive stages. The output of Preliminary Auditory Analysis is assumed to be a neural representation of the stimulus in terms of frequency, time and intensity. Sensory input is processed automatically without the operation of conscious selective attention. Sensory information is maintained in a relatively gross unanalyzed form in the Sensory Information Store (SIS). Information is further processed by separate stages involving acoustic and phonetic feature analysis. In Stage 1, Acoustic Feature Analysis, we assume that auditory features of the speech signal are recognized by a system of individual auditory feature detectors.

* This research was supported by grants from the National Institute of Mental Health (MH-24027) and the National Institutes of Health (NS-12179) to Indiana University.

The output of acoustic feature analysis is some set of acoustic cues or auditory features which forms the input to the next higher stage of processing.

In Stage 2, Phonetic Feature Analysis, we assume that a set of decision rules is employed to map multiple auditory features into more abstract phonetic features. At this stage there is a many-to-one mapping where several different auditory features provide information about a particular phonetic feature. The output of phonetic feature analysis is a set of phonetic features in the form of a distinctive feature matrix. Information from these stages is placed in short-term store where encoding, rehearsal and decision processes can operate on it and where it is made available for higher levels of linguistic processing.

The general purpose of the present study was to examine some of the properties of the auditory and phonetic feature analysis stages outlined above. In two experiments a selective adaptation procedure was used in which the repetitive presentation of an adapting stimulus alters the perception of a sequence of test stimuli (Eimas & Corbit, 1973, and Cooper, 1975 for a review). It has been assumed that the perception of a specific auditory or linguistic feature is mediated by feature detectors that are organized as opponent pairs. Repeated presentation of a stimulus containing a specific value of the feature is assumed to fatigue that detector and reduce its sensitivity. After adaptation the opponent or unadapted detector provides more information to the decision process than the adapted detector and therefore results in a shift in the locus of the phonetic boundary in identification.

Schematized versions of the synthetic stimuli used in these experiments are shown in Figure 2.

Insert Figure 2 about here

The test stimuli always consisted of a seven step CV continuum ranging from [ba] to [da] in which the formant transitions were varied systematically. The adaptors in the two experiments reported here consisted of [ba], [ab] and bTRAN as shown in Figure 2. The bTRAN stimulus had the same identical formant transitions as the [ba] adaptor except that they were in final position of a VC syllable. However, the [ab] stimulus had acoustically different transitions than the [ba] even though the phonetic segments [b] are the same. Previous results on selective adaptation with speech stimuli suggested the possibility of adaptation at an abstract phonetic feature level (Eimas & Corbit, 1973; Eimas, Cooper & Corbit, 1973). In this study we asked which way adaptation would go on a CV continuum when acoustic and phonetic features were placed in conflict in VC adaptors. A previous study by Ades (1974) failed to find any adaptation effects on a CV continuum when the adaptors were VCs. However, these null results are somewhat ambiguous since adaptation may have occurred at both auditory and phonetic levels and therefore canceled any observed effects on identification.

The results of the two experiments are shown in Figure 3. The top graph

----

Insert Figure 3 about here

----

in each panel shows baseline identification functions before adaptation and the identification functions after adaptation with a [ba] adaptor. In both experiments, adaptation with a [ba] produced a shift in the identification function toward the adaptor indicating that fewer test stimuli were identified in the [b] category. The bottom graphs in each panel show the results with VC adaptors. In the left hand graph adaptation with the bTRAN adaptor produced a small but reliable shift in the boundary toward the [b] category

($p < .025$). In contrast, adaptation with [ab] produced a reliable shift in the opposite direction toward the [d] category ($p < .05$). Thus, the direction of the shift in the locus of the phonetic boundary appears to be more a function of spectral similarity of adaptor and test series than commonality of their phonetic feature composition. If adaptation occurred only on a phonetic feature level, adaptation with [ab] should have produced shifts in the same direction as those obtained with the [ba] adaptor, namely, toward the [b] category. Likewise, adaptation with the bTRAN adaptor which has formant transitions which are actually appropriate for [d] in final position should have produced shifts in the direction of the [d] category.

From these results we conclude that the locus of adaptation effects in speech perception occurs at an auditory stage of processing prior to phonetic feature analysis. The perceptual mechanisms that are adapted appear to respond to the spectral rather than phonetic properties of the adapting stimuli. Although a strictly phonetic account of adaptation would be ruled out by these findings, there are several pieces of evidence which point to adaptation at either a more complex level of auditory analysis or phonetic analysis (see Cooper, 1975).

To sum up, selective adaptation effects have been found on a CV continuum when the adaptors are VCs. The direction of the shifts in the phonetic boundaries is consistent with an auditory account based on the spectral similarity of adaptor and test series. The main results are consistent with an information processing approach that assumes specialized auditory feature detectors for processing speech sounds. The extent to which there are specialized phonetic feature detectors in speech perception remains to be seen from the outcome of future research.

REFERENCES

Ades, A. E. 1974 "How Phonetic is Selective Adaptation? Experiments on Syllable Position and Vowel Environment," Perception & Psychophysics, 16, 1, 61-66.

Cooper, W. E. 1970 "Selective Adaptation to Speech," In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), Cognitive Theory: Volume I. Hillsdale, New Jersey: Lawrence Erlbaum Associates, pp.

Eimas, P. D. and Corbit, J. D. 1973 "Selective Adaptation of Linguistic Feature Detectors," Cognitive Psychology, 4, 99-109.

Eimas, P. D., Cooper, W. E. and Corbit, J. D. 1973 "Some Properties of Linguistic Feature Detectors," Perception & Psychophysics, 13, 2, 247-252.

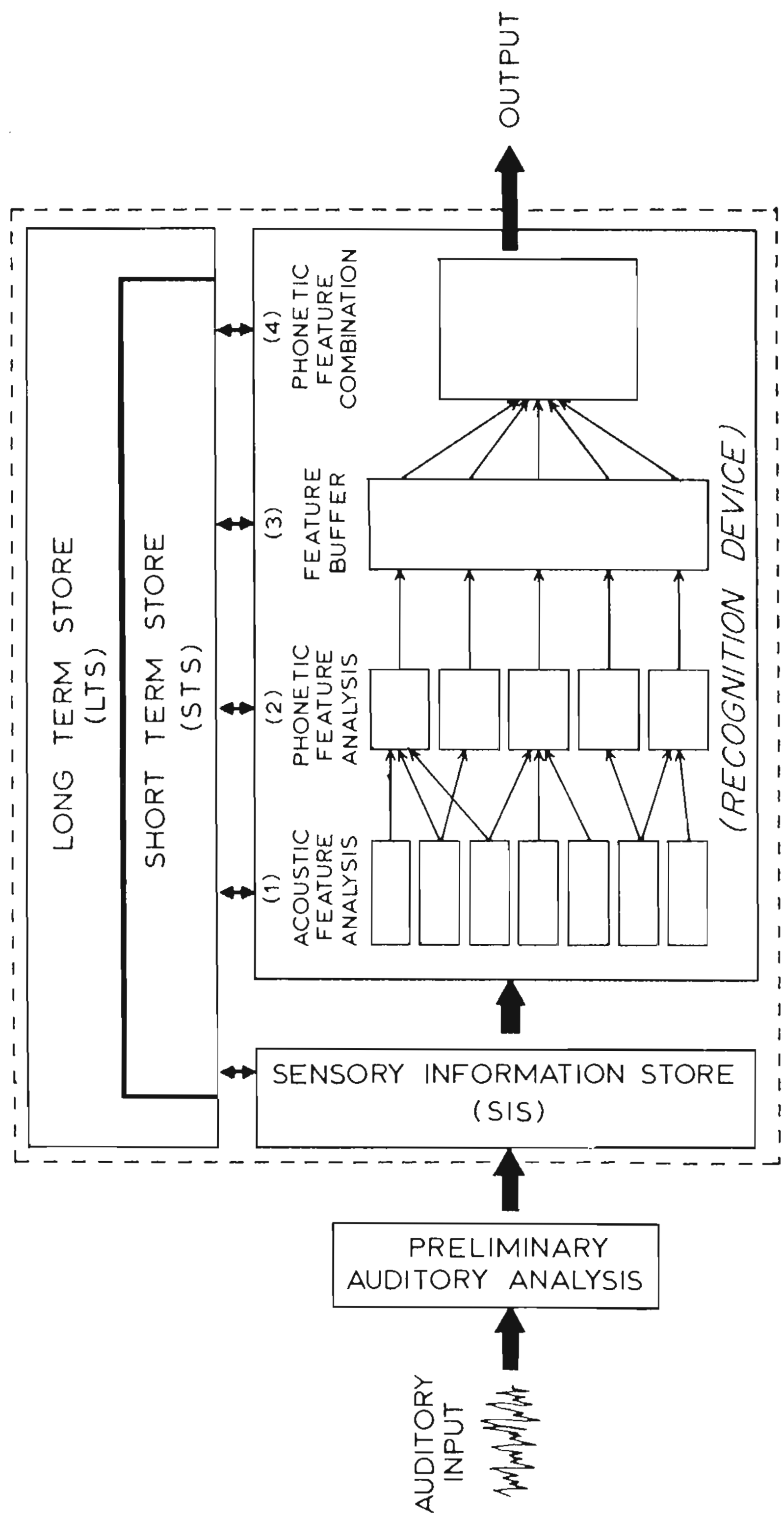
Pisoni, D. B. and Sawusch, J. R. 1975 "Some Stages of Processing in Speech perception" In A. Cohen and S. Nooteboom (Eds.), Proceedings on Dynamic Aspects of Speech Perception. Heidelberg: Springer-Verlag.

Figure Legends

Figure 1. Information processing model of the earliest stages of speech perception (After Pisoni and Sawusch, 1975).

Figure 2. Schematized spectrographic patterns of synthetic test series [ba] and [da] and adapting stimuli [ab] and bTRAN used in the perceptual experiments.

Figure 3. Identification functions showing the percentage of "B" responses before adaptation (solid lines) and after adaptation (dashed lines) for each experiment. Experiment I is based on data from 8 listeners. Experiment II is based on 5 listeners. Each listener provided 20 responses to each stimulus in a given experimental condition.



OUTPUT
LONG TERM STORE
(LTS)
SHORT TERM STORE
(STS)
(1)
ACOUSTIC
FEATURE
ANALYSIS
(2)
PHONETIC
FEATURE
ANALYSIS
(3)
FEATURE
BUFFER
(4)
PHONETIC
FEATURE
COMBINATION
(RECOGNITION DEVICE)
SENSORY INFORMATION STORE
(SIS)
PRELIMINARY
AUDITORY ANALYSIS
AUDITORY
INPUT

Figure 1

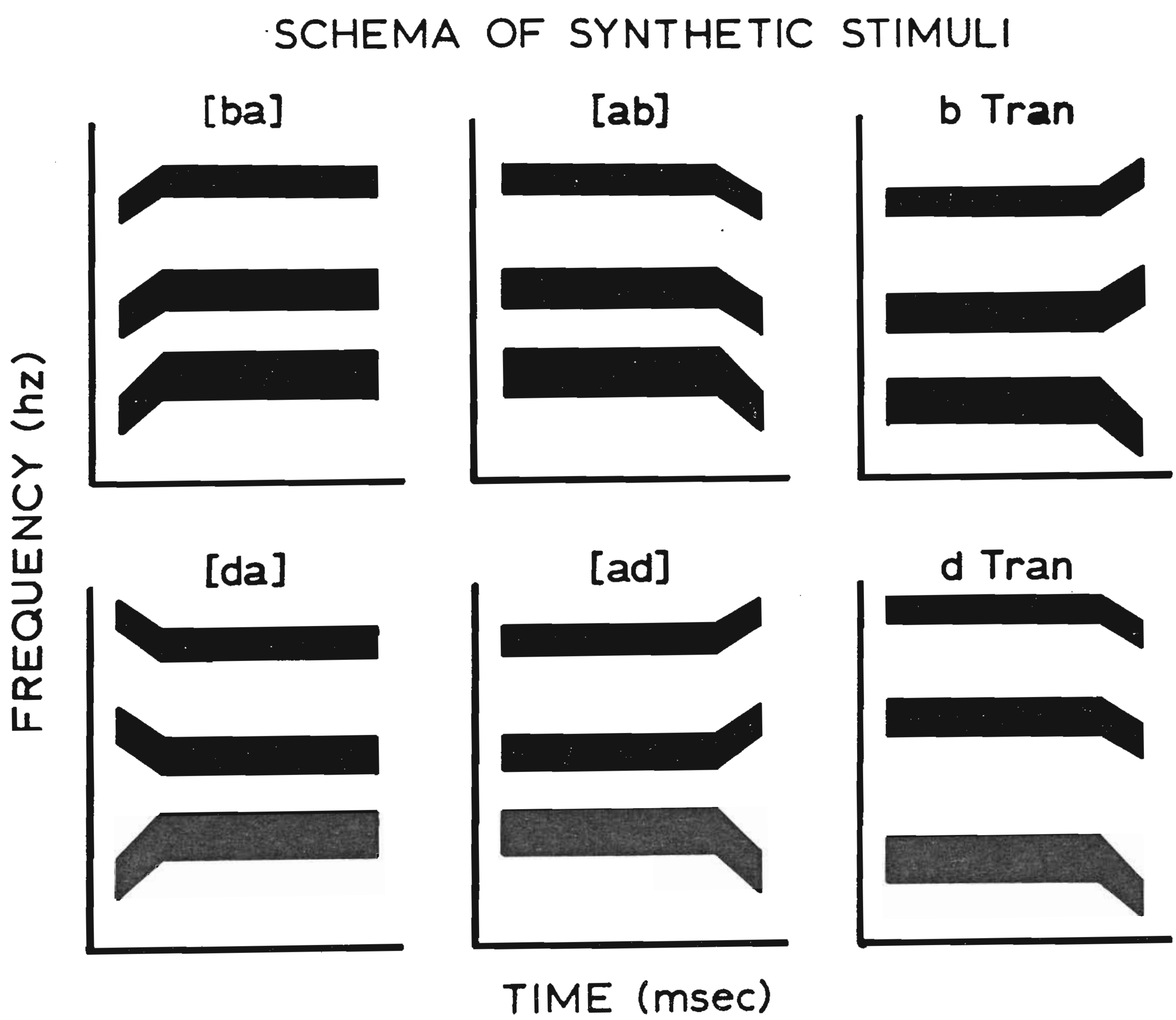

SCHEMA OF SYNTHETIC STIMULI
[ba]
[ab]
b Tran
[da]
[ad]
d Tran
FREQUENCY (hz)
TIME (msec)

Figure 2

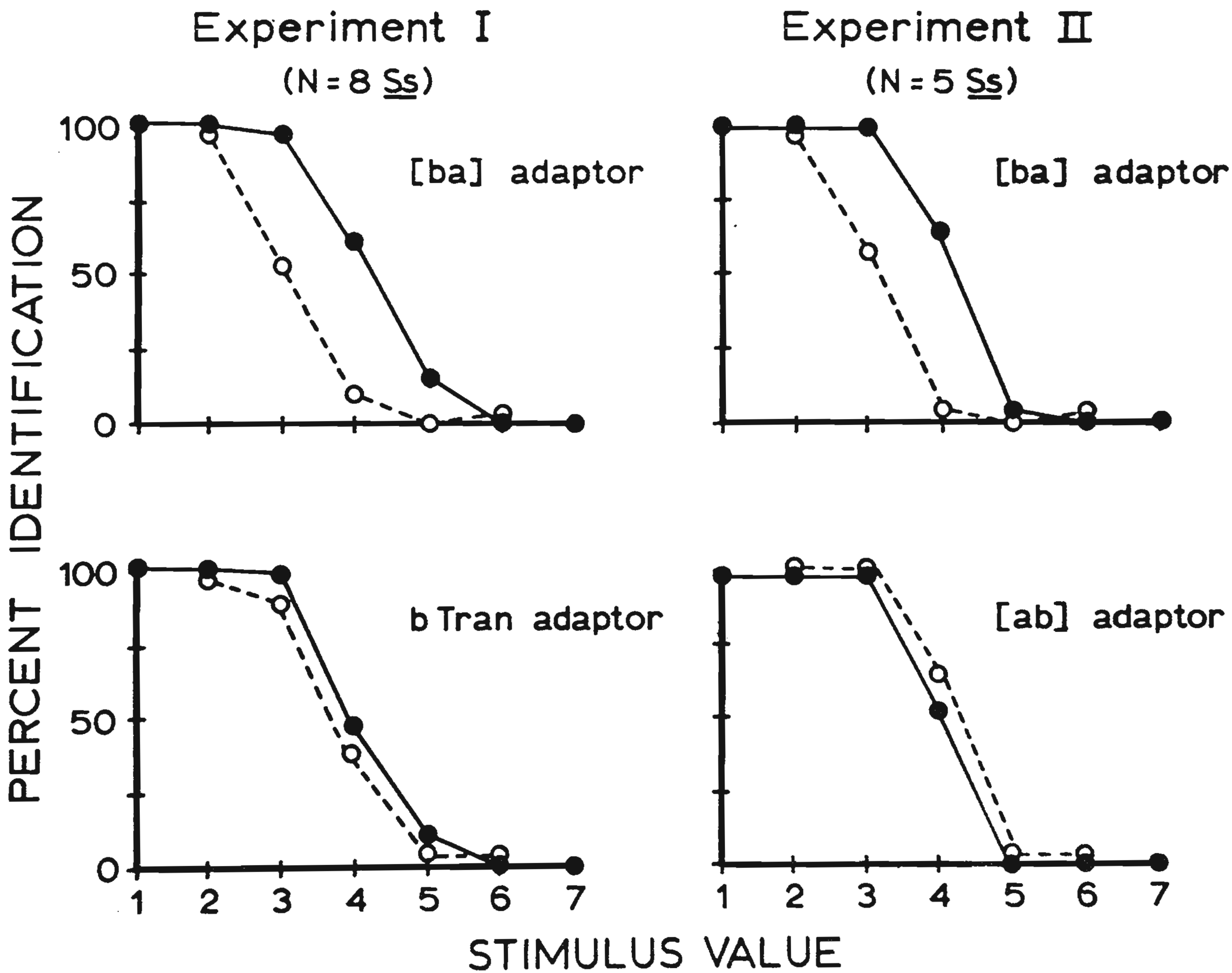

Experiment I
(N = 8 Ss)
Experiment II
(N = 5 Ss)
[ba] adaptor
[ba] adaptor
b Tran adaptor
[ab] adaptor
PERCENT IDENTIFICATION
100
50
0
100
50
0
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 3

Must the Output of the Phonetic Feature Detector be Binary?

Sandra D. McNabb

Indiana University

Features and feature detectors have played an important part in the development of psychology. Neurophysiological studies have found evidence for cells, that selectively, respond to certain light patterns (Lettvin, Maturana, McCulloch, and Pitts, 1959), and to cells which are sensitive to orientation of lines and movement in specific directions (Hubel and Wiesel, 1962). There has long been evidence from memory and speech perception studies that suggests that at some stage of processing, speech sounds are analyzed in terms of its features. Studies in consonant confusions (Miller and Nicely, 1955; Wickelgren, 1966), backward recognition masking (Pisoni and McNabb, 1974), and confusions during dichotic listening (Cole and Scott, 1972) all indicate that phonetic features are available to the subject. Eimas and Corbit (1973) used a selective adaptation procedure to demonstrate that this feature analysis was carried out via phonetic feature detectors that had properties similar to visual feature detectors. They hypothesized two detectors, each sensitive to a restricted range of values on the voice onset time continuum. They demonstrated that the sensitivity of a detector was decreased when stimuli within its range were repeatedly presented. The result of this loss of sensitivity was to move the phonetic boundary towards the adapting stimulus.

In the Eimas and Corbit study they also plotted a discriminability function of adjacent stimuli before and after adaptation. They found that the peak in the function, which usually denotes the phonetic boundary, moved as well as the boundary itself. They concluded that the categorical nature of speech was due to the output of the feature detector.

The present paper addresses the question: Must the output of the phonetic feature detector be binary? In order to account for the peaked discriminability functions, Eimas and Corbit assumed that the output was indeed binary: "... no distinction is made when the same detector is excited by two different values of VOT, then the peaked discriminability functions are readily accounted for (1973)". However, here Eimas and Corbit have assumed that the phonetic decision is the same as the output of the phonetic feature detector. Recent results from Cooper and Blumstein (1974) can be taken as evidence for the distinction between the output of the phonetic feature detector and the phonetic decision. The adapting stimuli in the Cooper and Blumstein study were /bæ/, /p^{h}æ/, /mæ/, and /væ/ and the test series was the 13 step continuum /bæ/ - /gæ/. They found relatively strong adaptation with the syllables /bæ/, /mæ/, and /væ/, and much less with /p^{h}æ/ and /wæ/. They concluded 1) because a shift was found with the labiodental fricative /væ/, then the bilabial detector must extract more general feature information; 2) the feature detector must be sensitive only to true consonants since adaptation with the bilabial semiconsonant /wæ/ did not produce a consistent shift. It is possible that these results reflect subtle differences between the output of the phonetic detector and the phonetic decision. The phonetic feature label of /bæ/, /p^{h}æ/ and /wæ/ may be identical, but at the same time, the bilabial phonetic feature detector may not respond equally to all tokens. The syllable /wæ/ may not be as good an adapting syllable as /bæ/ simply because of the initial sensitivity differences. Likewise, /væ/ might be as good an adapting stimulus as /bæ/ because the detector is initially equally sensitive to both.

In the present experiment in order to decide whether or not the output of the phonetic feature detector is binary, it is necessary to have stimuli that are acoustically distinct yet phonetically equivalent. In this experiment the first three stimuli of the seven-step continuum of /ba/ - /da/ were chosen as the adapting stimuli, and the test series was the /ba/ - /da/ continuum. According to Eimas and Corbit, any stimulus within the range of the detector would be an equally good adapting stimulus. However, if the output of the detector was not binary and the binary phonetic decision came at a later stage, then the initial differences in sensitivity would be reflected in the output, and the three adapting stimuli should produce different shifts. If the bilabial detector was most sensitive to stimulus one, the endpoint stimulus, then repeated presentation of stimulus one should cause the greatest adaptation resulting in the greatest decrease in sensitivity. Likewise, if the bilabial detector is less sensitive to those stimuli close to the phonetic boundary, then adaptation with a syllable close to the boundary (i.e., stimulus three) should have little effect.

Confidence ratings were introduced as a second dependent variable because it was felt that a simple BA, DA response could not capture these subtle differences.

Method

Subjects. The subjects were 28 introductory psychology students who received credit towards their course requirement. All subjects were right-handed, native speakers of English, and they reported no history of any speech or hearing defects. No subject reported ever having been in any selective adaptation experiment. All subjects participated in three 45-minute sessions of the experiment.

Stimuli. All of the stimuli in this experiment were three-formant synthetic speech syllables. They were constructed on the parallel resonance synthesizer at Haskins Laboratories and recorded on magnetic tape.

The test stimuli consisted of seven CV syllables that ranged perceptually from /ba/ to /da/. The first formant remained constant in all seven stimuli while the second and third formants differed in the starting frequencies. The syllables were 300 msec in duration; the first 50 msec

Insert Table 1 about here

were linear transitions and the final 250 msec consisted of the steady-state formants for the English vowel /a/. The center frequency for these formants were 769 Hz (F_1), 1232 Hz (F_2), and 2525 Hz (F_3). The three adapting syllables corresponded to syllables 1, 2, and 3 from the test series.

Apparatus. All stimuli were recorded on high quality tape and reproduced on an Ampex AG-500 tape recorder. The stimuli were presented diotically through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder was adjusted to give voltage across the headphones equivalent to 75 dB SPL re 0.0002 dynes/cm^2.

Procedure. Each 45 minute session consisted of two parts: Baseline performance for that session and adaptation. In order to determine the baseline for that session, ten random sequences of the seven test stimuli

Table 1

Starting Frequencies of the Second- and Third-Formant Transitions for the Synthetic CV Test Stimuli

	Starting Frequencies (in Hz)	
Stimulus	F_2	F_3
1	996	2180
2	1075	2348
3	1155	2525
4	1232	2694
5	1312	2862
6	1386	3026
7	1465	3196

(ISI = 3 sec) were presented to the subjects. The subjects were required to give two responses per syllable, first to identify it as /ba/ or /da/ (by filling in the appropriate space on the answer sheet), and secondly, to qualify his identification by using the following rating scale:

1. very sure BA
2. sure BA
3. not very sure BA
4. not very sure DA
5. sure DA
6. very sure DA.

No subject reported any difficulty in using the rating scale.

The adaptation part of the session consisted of 28 adaptation trials with a rest break after 14 trials (when the tape was rewound and played again). Each trial consisted of 100 repetitions of the adapting stimulus, a pause, and the seven test syllables. The subject had to respond to each syllable both with its label (/ba/ or /da/) and with a value from the rating scale. The seven test syllables were presented in a different random order for each trial. At the end of the session, each test syllable had been presented for identification 28 times.

Only one adapting syllable was used in a single session. Since the experimental variable was entirely within-subject, each subject participated in three sessions. The subjects were run in groups of four or five, and there were six groups in all representing the possible permutations of the three conditions.

Results

Three different analyses of the data have been performed. The first concerns the binary "BA" "DA" response, the second concerns the confidence ratings, and the third concerns the confidence ratings converted to a TSD measure, P(A). First of all since there were twenty-eight subjects in this experiment, it is not feasible to plot each subject's data. In Figure 1 the probability of responding "BA" to each test stimulus is plotted as a function of each adapting stimulus. The ID function is also plotted. These data do follow the usual shift pattern: the phonetic boundaries under the adaptation conditions have moved towards the adapting stimuli. In this experiment the subject's boundary is defined as that point along the continuum which would receive, by extrapolation, 50% of the "BA" responses and 50% of the "DA" responses. A single factor, repeated measures analysis of variance was performed on the difference between the pre-adaptation and post-adaptation boundaries. This measure was not significant [$F(2.81) = .592$, $p < .556$], indicating no differences among the mean shifts for each adapting syllable. Because of the no difference finding, the data from the three conditions can be collapsed and a correlated t-test performed to test whether or not this shift is different from 0. The collapsed data do show that the shift is significant, $t(83) = 20.22148$, $p < .001$.

On the basis of the simple binary response, it must be concluded that the acoustically different adapting stimuli all had the same effect.

However, the confidence ratings give a different picture. The proportion of times subjects gave a "very sure BA" response for each test

---

Insert Figure 2 about here

FIGURE 1

PROBABILITY OF RESPONDING "BA" TO EACH TEST STIMULUS

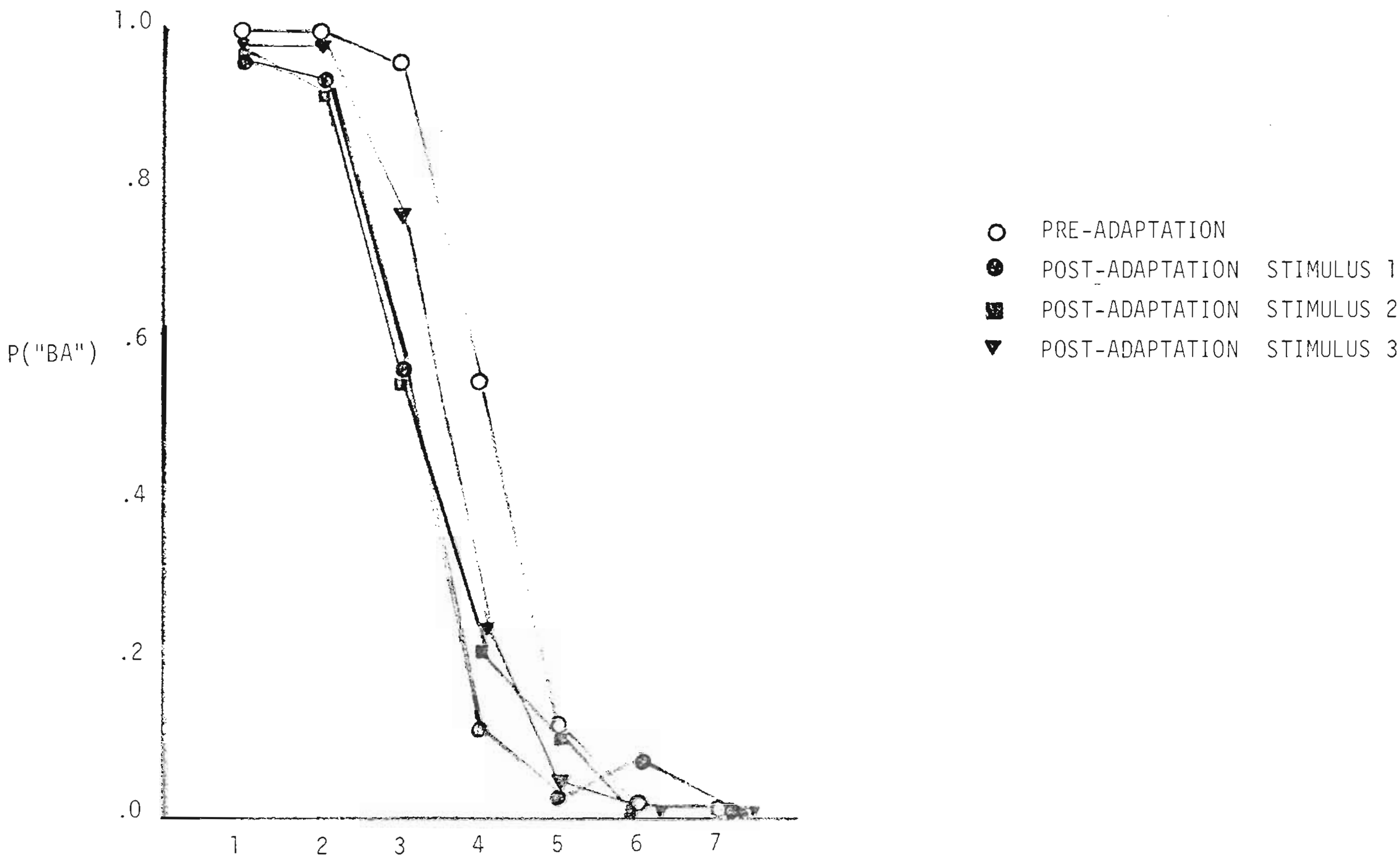
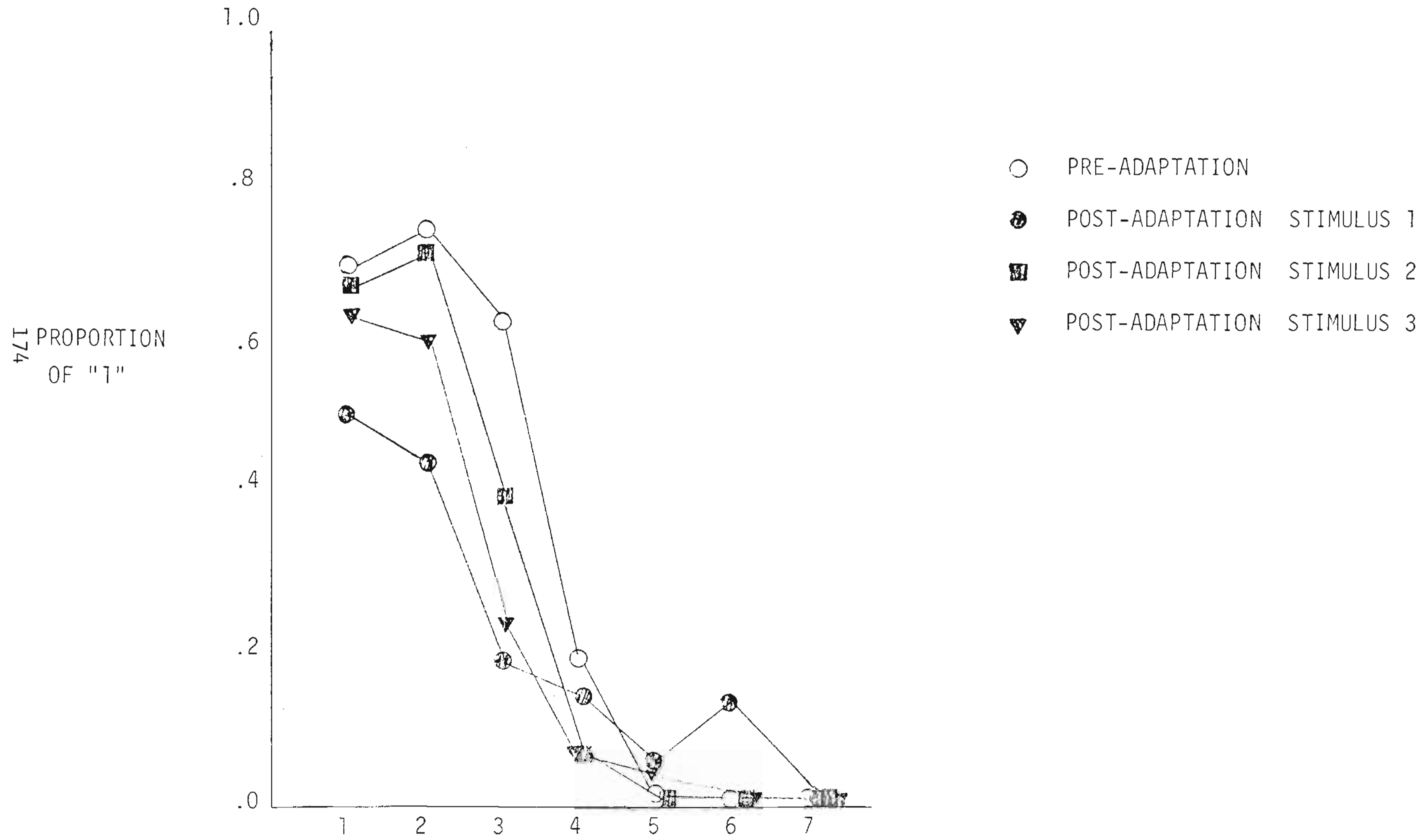


FIGURE 2

PROPORTION OF "VERY SURE BA" RATINGS FOR EACH TEST STIMULUS



syllable is plotted in Figure 2. Here it is very clear that the different adapting stimuli did have a different effect. Subjects are much less sure that stimulus 1 is a /ba/ after adaptation with stimulus 1 than adaptation with stimulus 3.

Finally, the confidence rating data can be converted to hit and false alarm rates and a ROC curve can be constructed. In order to avoid all assumptions about underlying distributions and/or criteria, the area under the curve, P(A), was computed (see Green and Swets, 1966). As the sensitivity or ability to discriminate /ba/'s from /da/'s increases, the measure P(A) increases. The post-adaptation curve was subtracted from the pre-adaptation curve in order to measure the change in sensitivity as a function of the different adapting stimuli. An analysis of variance revealed a significant difference among the three mean changes [$F(2,81) = 48.872$, $p < .001$]. In Figure 3, the ROC curves for the three conditions are plotted. In summary, the results indicate that with the traditional

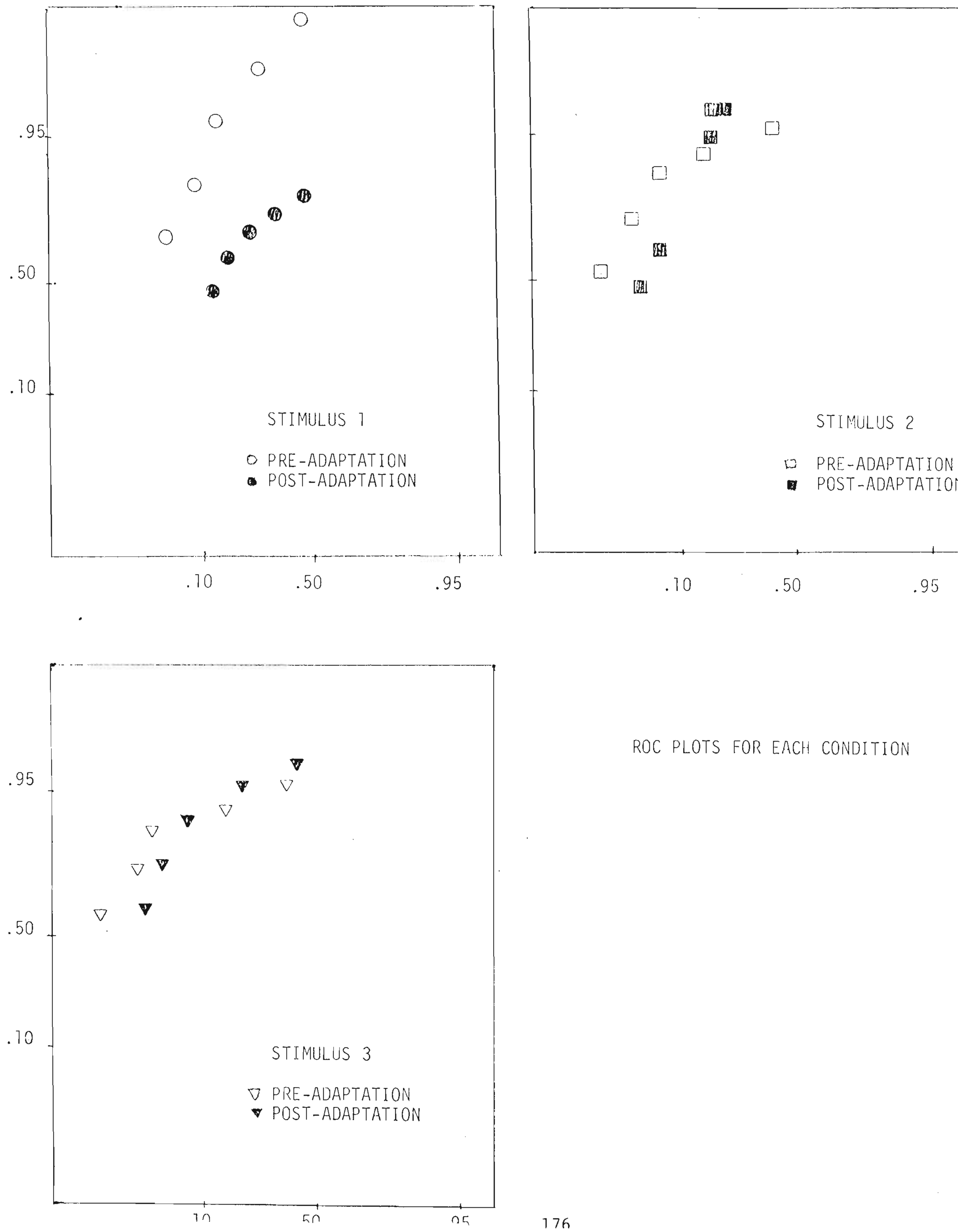
----
Insert Figure 3 about here
----

binary responses, no difference among the adapting stimuli were found. Further analyses using confidence ratings and P(A) do indicate consistent differences in the selective adaptation.

Discussion

The major result of this study was that acoustically different stimuli did not produce comparable effects. It is possible to reject Eimas and Corbit's assumption that the phonetic decision is the same as the output of the phonetic feature detector. Instead, on the basis of these data, it is better to assume that the output of the phonetic feature detector

FIGURE 3



ROC PLOTS FOR EACH CONDITION

reflects its sensitivity, and that at a later stage of processing a phonetic decision is made. It is still possible to get categorical perception and peaked discrimination functions because the response of one feature detector exceeds that of others. At the decision stage it is necessary only to decide which detector is responding the most. The phonetic boundary corresponds to that point where both feature detectors respond equally.

This explanation is consistent with Cooper and Blumstein's (1974) study with labial feature detectors. As was mentioned earlier in this paper, initial differences in sensitivity would cause corresponding differences in adaptation and boundary shifts. It is possible that the bilabial detector is less sensitive to /wæ/ than to /bæ/ and, therefore, /wæ/ would cause less adaptation. Ades (1974) recently found that the repeated presentation of a CV syllable had an adapting effect on a CV continuum, but not on the VC continuum, and vice versa. Ades concluded that "adaptation could not be phonetic in a truly linguistic sense..." since the position of the phoneme was crucial in determining the adaptation. However, it is entirely possible that the decision as to what constitutes a [b] in final position is based on different information than what constitutes a [b] in initial position. The decision is the same, but it is based on different information. It is entirely possible that the adaptation of the phonetic feature detector has different effects on the phonetic decision in both conditions. The phonetic detector versus the phonetic decision is an important distinction that has long been overlooked.

Finally, the choice of the dependent variable was crucial in this study. It is important in all speech discrimination and identification studies to have as much information from the subject as possible. It was shown in this study that the simple "BA" "DA" response could not capture all of the information that the subject had about the test syllable. Further research should include confidence ratings to insure valid conclusions.

References

Ades, A. E. How phonetic is selective adaptation? Experiments on syllable positions and vowel environment. Perception & Psychophysics, 1974, 16, 57-62.

Cole, R. A. & Scott, B. Distinctive feature control of decision time: Same-different judgments of simultaneously heard phonemes. Perception & Psychophysics, 1972, 12, 91-94.

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

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

Green, D. M. & Swets, J. A. A Signal Detection Theory and Psychophysics, New York: Wiley, 1966.

Hubel, D. H. & Wiesel, T. N. Receptive fields, binocular interaction and functional architectures in the cat's visual cortex. Journal of Physiology (London), 1962, 160, 106-154.

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S. & Pitts, W. H. What the frog's eye tells the frog's brain. Proceedings of the IRE, 1959, 47, 1940-1951.

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

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

Wickelgren, W. A. Distinctive features and errors in short-term memory for English consonants. Journal of the Acoustical Society of America, 1966, 39, 388-398.

BOOK REVIEWS

The Psychology of Language: An Introduction to Psycholinguistics and Generative Grammar. By J.A. Fodor, T.G. Bever and M.F. Garrett New York: McGraw-Hill, 1974, Pp. xvii, 537.

Reviewed by David B. Pisoni, *Indiana University, Bloomington*

To appear in *Language*

The psychology of language: an introduction to psycholinguistics and generative grammar. By J. A. Fodor, T. G. Bever, and M. F. Garrett. New York: McGraw-Hill, 1974. Pp. xvii, 537.

Reviewed by David B. Pisoni, Indiana University, Bloomington

Fodor, Bever and Garrett (FBG) tell us that one of the major goals of this book is an attempt to develop an "experimental mentalism," a psychology of the mental processes that underlie complex behavior. According to FBG, psycholinguistics may be thought of as a test case for the efficacy of an experimental mentalism. The richness and complexity of language and language behavior is assumed to be the outcome of complex underlying mental processes most of which are inaccessible to direct observation. The authors argue that an account of these processes can only be captured by theories which acknowledge elaborate and complicated mental processes. This book is more or less a "progress report" on research in psycholinguistics over the past decade, particularly work dealing with the encoding and decoding of the syntactic structures of sentences. The book is not simply a review and summary of work in psycholinguistics as one might anticipate from the title but is FBG's interpretation of a fairly select body of research which assumes or specifically tests a transformational approach to syntax. This is, of course, not surprising since most of the contemporary work in psycholinguistics grew out of, or was in response to, developments in transformational generative grammar. The book is not, in my view, an introduction to the psychology of language since it presupposes a reasonable amount of background in linguistics on the part of the reader. The book is fine for a graduate course in psycholinguistics. Indeed, I used it just this past spring semester. But I would suspect that anyone who attempted to use it as a text in an undergraduate course might have his work cut out for him.

One of the interesting things about the book is that it presents a very specific point of view about the relationship between grammars and psychological models of the mental processes involved in sentence recognition. In the early days of psycholinguistics immediately following the publication of "Some Psychological Studies of Grammar" by George Miller (Miller, 1962) most people assumed a fairly direct and intimate relationship between properties of the grammar and the psychological or mental operations employed by listeners in sentence processing. FBG have taken great pains to show that much of the early experimental findings were simply fortuitous and that in actuality there is a much more abstract relation between grammars and psychological processes. As the reader will see, this is an important issue to consider since it means that psycholinguistics has a subject matter of its own which is separate and distinct from formal linguistics.

Many of the questions that are raised throughout the book have implications not only for psycholinguistics, in particular, but for cognitive psychology more generally. It may come as a surprise to many linguists but there have been virtually no attempts to integrate the vast body of literature in experimental psycholinguistics that has accumulated since the early 1960's. Although there have been a score of anthologies reprinting some of the "classic" articles, numerous conference proceedings and a small handful of introductory paperbacks in psycholinguistics, there have been few if any attempts to put it all together in the form of a full length book that could serve as a text, particularly in an advanced course. If nothing else, FBG have filled a need that has existed for much too long. However, looked at in a somewhat different way, the book is a refreshing change from the standard textbooks that have appeared in ever increasing proclivity in experimental psychology over the last few years. Although many readers will probably disagree with some of FBG's proposed solutions, there is no doubt that most will concur with the issues that are raised in the book which is no small feat these days in any field of psychology let alone linguistics proper.

In the introduction, FBG try to place the study of language into the broader context of cognitive psychology. Thus, psycholinguistics may be thought of as a specific example of how organisms use abstract concepts to organize their perceptions and guide behavior. According to FBG, much of the domain of cognitive psychology concerns how organisms apply concepts to what they do and there is good reason to suspect that much of the observable behavior of an organism is mediated by a complex series of internal processes and events.

One of the central problems in cognitive psychology appears frequently in the book, namely, the problem of "perceptual constancy" or "stimulus equivalence." Consider for a moment the concept of a chair. How do we know that something is a chair when we see it? Do we compare it against a stored image in memory or do we extract critical features from the stimulus complex that define an object as a chair? Most writers agree that our perception of an object like a chair is a function of a number of factors including sensory information, our knowledge of the concept (i.e., chair), and our expectations about what we should or should not see given a particular context. Numerous examples of perceptual constancies occur at all levels of language structure from phonetics all the way through syntax and semantics. These form a thread throughout the book. It is FBG's hope that whatever is learned about these sorts of problems within the context of language will have some implication for other areas of cognitive psychology. More specifically, with regard to the study of language, FBG consider two questions. First, what is the nature of the internal representation of sentences? And second, what is the nature of the encoding and decoding systems used to convert the internal representations of sentences into acoustic objects and vice versa. FBG see these as the most important problems in psycholinguistics and therefore devote a great deal of attention to them in the book.

Chapter 2, 'Psycholinguistics of taxonomic grammar' is a brief history of structural linguistics before the 1960's and a discussion of some of the first psycholinguistic implications that followed from the joint collaboration of linguists and psychologists. In the first part of this chapter, FBG give a fairly concise description of the assumptions of pre-Chomskyan taxonomic linguistics. For example, they describe two conditions that they claim formed the foundations of structural linguistics. The first, the "taxonomic condition," assumed that entities postulated at one level must be derived from sequences of smaller entities at a previous level of description. One result of this condition was the ordering of levels of linguistic description into a hierarchy: phonetic, phonemic, morphological, syntactic. These levels were assumed to be independent with no interaction of higher levels or lower levels. The second condition FBG describe is the "operationalist condition." By the application of well-defined "discovery procedures", an analysis of the structure at each level of linguistic description could be obtained automatically. The primary goal of taxonomic linguistics was to produce a taxonomy of some corpus of utterances of a particular language. FBG then go on to describe several taxonomic procedures including complementary distribution, free variation, and the minimal pair test. This section gives a fairly good account of structural linguistics and goes into sufficient detail so that even a person who had no previous background in linguistics could understand the basic assumptions.

In the second part of the chapter, FBG try to present a picture of the type of psychological theories that were popular during the 1950s and how these were compatible with the goals and methodology of structural linguistics. To take one example, there was widespread agreement among psychologists at this time that language was a form of behavior much like other behaviors and was amenable to the same basic laws of learning that could characterize simpler behaviors. For example, take the problem of serial ordering in behavior. One of the basic assumptions was that the problems of serial ordering in language could be dealt with by the notion of transitional probability which presumably reflected the associative habits of mature language users. FBG then go into Osgood's mediational approach which assumed associative connections between both overt and covert elements. Finally, FBG compare associative relations with various types of grammatical relations found in language in order to show that the basic associative principles cannot account for the serial order and hierarchical structures found in natural languages.

In chapter 3, 'Generative grammars,' FBG begin with a discussion of the goals of grammatical description. They consider observational, descriptive and explanatory adequacy and then go on to point out some of the limitations of taxonomic grammars. For the most part this is a pretty clear exposition of Chomsky's arguments in _Syntactic Structures_ (1957) and _Aspects of the Theory of Syntax_ (1965). In the next section, FBG consider transformational grammars. First there is a brief historical diversion to the work of Zelig Harris on co-occurrence grammar which the authors note was the forerunner of transformational generative grammar. The remainder of the chapter is taken up with explicating a version of the "standard" theory of transformational grammar which is pretty much based on _Aspects_. For example, FBG state that: (1) transformations do not change meaning and the syntactic information necessary for semantic interpretation is represented in deep structure, (2) grammatical relations are defined over deep structures, (3) transformations do not introduce structure - - rather the effect of successive transformations is to degrade the trees to which they apply, (4) selectional restrictions, constraints on cooccurrence of specific lexical items, are stated in the base and therefore prior to the operation of any transformations other than rules for lexical insertion. FBG then provide some examples of how the following syntactic phenomena are handled within the standard theory: affix movement, passivization, relative clauses, complementization, and lexical insertion. In the final section of the chapter, FBG briefly consider the role of phonology in generative grammar. They note that the phonological component simply has an interpretative function in the grammar, to map the surface syntactic structure of a sentence onto sequences of phonetic segments. Only at the tail end of the chapter do FBG give us an inkling of some of the strong psychological claims made by Chomsky. For example, FBG say that the grammar claims to give an account of a speaker-hearer's knowledge of the structure of his language. To the extent that this is an accurate reflection of this knowledge, it would therefore represent information that would be employed in language production, comprehension and acquisition. Although there is no discussion of the role of semantics within a generative grammar in this chapter, it is taken up in the latter half of the next chapter. For the most part this chapter is an excellent introduction to the basic ideas of transformational grammar as they were formulated in the middle 1960s.

Chapter 4, 'Semantics,' consists of two relatively independent sections, one dealing with signs and symbols and the other dealing with the role of semantics in generative grammar. In the first section, FBG take up the problem of meaning and, in particular, the relationship between symbols and their referents. FBG point out that American psychologists who claimed to being studying word meaning were actually dealing with the problem of reference. This is not surprising since it was assumed that words acquired their meanings by being associated with the objects or events in the environment. FBG point out several problems with assuming that meaning derives from the particular referent named. For example, not all words in the lexicon name things; the nature of the referent is not given upon uttering the word; not only do words that name referents have varied referents, but meaning also varies with context. Thus, FBG argue that naming is not the central function that words perform nor is it the major function of language. Related to the problem of naming and reference is the traditional distinction between signs and symbols. That is, whether things get their meaning naturally or by convention. The major point here is that human languages are systems of symbols rather than signs and that words have only an arbitrary relation to their referents. This is, of course, one of the characteristics that distinguishes human languages from animal communication systems. FBG then focus on several traditional ways of dealing with the problem of reference and the distinction between signs and symbols.

In the second section, FBG deal with the role of semantics in a generative grammar. The focus of this section is on the way in which sentences are used to perform "speech acts" such as assertion, asking questions, giving orders, promising, inferring, etc. FBG argue that speech acts are the basic functions of verbal behavior. The linguistic vehicle for performing a speech act is the sentence. They cite two preliminary goals of a semantic theory. First, a semantic theory should specify the semantic properties that sentences can exhibit. That is, a semantic theory should enumerate the various types of speech acts that sentences can be used to perform and the necessary conditions for their performance. Second, a semantic theory should provide some type of semantic representation which specifies how the semantic properties of particular sentences are determined by their lexical content and syntactic structure. Unfortunately, FBG do not provide us with even a glimpse of what the formal properties of their semantic theory might look like. They do note, however, that there has been a tradition of employing some version of first-order predicate logic as a system for specifying the semantic relations for sentences. According to FBG, although one can translate a sentence into some logical system, this translation cannot be claimed to be a representation of the logical force of the sentence. That is, the logical representation may not specify the commitments made by a speaker when he undertakes to use a sentence to perform a particular type of speech act. While it is not apparent from FBG's discussion of semantics, numerous investigators have attempted to employ some type of propositional system in developing semantic representations. Indeed, a recent book by Anderson and Bower (1973) specifically employed a propositional notation system to represent the structure of sentences stored in long-term memory (see also Kintsch, 1974).

Of all of the chapters in the book, this was perhaps the most disappointing. Given the recent interest by linguists in semantics and semantic theory and the re-direction of research in psycholinguistics to semantic problems, one would have hoped for a more systematic treatment of the issues in semantics. Instead, we get a lot of hand waving and circumlocutions about what a proper semantic theory ought to look like.

In chapter 5, 'Psychological reality of grammar,' FBG provide a summary of the experimental work dealing with the relation between the grammar and the mental operations employed in perception and comprehension. The basic conclusion that FBG come to after reviewing this work is that the structures postulated by linguistic theory have a "psychological reality" but the processes or grammatical operations (transformations, re-write rules, etc.) do not. This conclusion is based on the observation that the structural distinctions marked in the grammar such as surface structure and deep structure are reflected in subject's behavior in various types of experimental tasks. However, there is a good deal less correspondence between presumed grammatical operations and psychological or mental processes. The early psycholinguistic experiments used traditional verbal learning and memory paradigms that were available from experimental psychology in an attempt to determine whether the distinctions made by linguistic theory were reflected in the subject's behavior in laboratory tasks. Numerous experiments showed that sentences were not merely lists of unrelated words but that they had an internal structure as organized perceptual units. The well known "click" studies were aimed at demonstrating that structure was imposed upon sentences as a consequence of grammatical analysis rather than simply being marked objectively in the physical stimulus. Other studies showed that there was good evidence for the distinction between deep and surface structure of sentences. Indeed, the prompted recall experiments of Blumenthal (1967) and Wanner (1968) provided strong evidence that the internal representation of a sentence in memory corresponded to the basic grammatical relations marked in deep structure. There was also good evidence for the "coding hypothesis," that a sentence was stored in memory in some abstract form independently of its transformational history. However, numerous people began to worry about the types of coding strategies that subjects adopted in these experiments and how these strategies might influence the nature of the representation of sentences in memory. Although it was suspected for some time, there is now very strong evidence that the long-term memory representation of sentences is quite abstract and more closely related to the subject's interpretation of the sentence than its original syntactic form (Fillenbaum, 1966; Sachs, 1967). Put another way, subjects remember the meaning of a sentence rather than its form. The studies of Bransford and Franks (1971) suggested that subjects will even integrate several related sentences into an organized conceptual structure. As a consequence, information about the form of the original input sentences is lost rapidly and cannot be recovered. These sorts of results have brought about a revival of interest by psychologists on work in human memory processes and, in particular, on the processes by which linguistic information is encoded, stored in memory, and retrieved for some psychological operation. It might be worthwhile to point out here that although FBG argue that the evidence for the psychological reality of grammatical operations is weak, other investigators, notably Gough (1971) have challenged their interpretations of these results.

Chapter 6, 'Sentence perception', consists of two parts, phonetics and syntax recognition. FBG argue that a necessary step in sentence perception and comprehension involves the recovery of the structural description of a sentence. They say that any proposed model of a sentence recognizer will have to be constrained in at least three ways. First, there are formal constraints; the model must be compatible with an optimal grammar whatever that may turn out to be. Second, there are psychological constraints involving the compatibility of the sentence recognition model with known psychological functions some of which will involve cognitive and neurological considerations. Finally, there are empirical constraints dealing with how listeners process sentences in real-time.

In the first part of this chapter, FBG consider various perceptual models of phonetics. They consider and then correctly reject what they call the "naive" view of speech perception (see for example Massaro, 1972; Cole and Scott, 1974). This is the view that phonetic perception can be carried out by means of straightforward template matching or filtering of the acoustic waveform against stored patterns in memory. They then proceed to a discussion of the invariance and linearity conditions and show how these pose serious problems for any theory of phonetic perception. In this section, FBG briefly review some of the classic work in acoustic phonetics along with a discussion of the various types of perceptual units that have been postulated. The perceptual constancy problem for phonemes is also brought up here. As I noted earlier, this is a general problem in other areas of perception although it seems more exaggerated within the context of speech perception. FBG devote a fair amount of attention to the supposed relation between production and perception. Active perceptual theories such as the "Motor Theory of Speech Perception" and "Analysis by Synthesis" are considered to be the best candidates for perceptual models of phonetics.

In the second part of this chapter, FBG deal with syntax recognition. One of the fundamental facts that psycholinguists have had to work with is that some types of sentences are more difficult to understand than others. Obviously if this were not the case, we would not have an interesting psychological problem to study. FBG state that at some level of perceptual analysis the deep structure of a sentence must be recovered. They consider two recognition procedures: analysis by analysis and analysis by synthesis. In analysis by analysis, the grammar is simply run backwards from the bottom up starting with words and then intermediate structures and finally ending up with a sentence. One problem with this procedure is how to recover elements that were deleted by transformations. The most prominent theory has been some version of analysis by synthesis whereby the input is compared to structures that have been generated from some internal representation. This is the approach originally advocated by Chomsky (Chomsky & Miller, 1963). Of course, as FBG point out, there are serious problems with this approach too. For example, the size of the search space must increase as the length of the input increases.

FBG then consider an alternative to analysis by synthesis which involves the use of "heuristic strategies". The argument here is that analysis by analysis and analysis by synthesis approaches have assumed that the grammar plays a fairly intimate and direct role in the recognition process. In contrast, the perceptual heuristics strategy postulates a more abstract relation between the grammar and the psychological operations employed in sentence recognition. According to FBG, listeners are assumed to use heuristic strategies to make direct inferences about the relations in base structures. For example, by means of clausal analysis, listeners group words together which are members of a common sentoid. These clauses act as units to provide information about deep sentences. Related to clausal analysis is the "canonical sentoid strategy" which involves imposing an SVO ordering on any NP - V - NP sequence. FBG also propose a "lexical analysis" strategy whereby inferences about the deep structure configuration of a sentence can be made based on the syntactic features of the lexical items. Thus, FBG stress the importance of the verb in a sentence since it constrains the types of possible grammatical structures a sentence can have. Finally, FBG consider surface structure elements as cues to various deep structure configurations. For example, the presence of relative pronouns or a tense marker facilitates the segmentation of a sentence into appropriate units for further processing. What FBG are talking about here are parsing strategies and it was surprising to see how they overlooked the fairly extensive literature that has been accumulating in artificial intelligence and computational linguistics over the past few years. In particular, the work of Woods (1970) and Kaplan (1972) on augmented transition networks presents a strong alternative to the traditional transformational linguistic approach that FBG advocate.

One of the main points that FBG make in this chapter is that psycholinguistics has a subject matter of its own. That is, the study of the relationship between grammars and the psychological processes involved in sentence recognition is distinct from linguistic theory and requires its own formal apparatus. This, FBG tell us,is the "most important result of the last 10 years of research."

Chapter 7 is entitled 'Sentence production' but it actually deals with two relatively independent topics: language and thought and sentence production. In the first section, FBG discuss the Whorfian hypothesis and the relationship between language and thought. They describe some of the classic studies on codability and then try to show how the radical Whorfian view is unrealistic given the types of findings obtained by Berlin and Kay (1969) on basic color terms and the findings of Eimas (1971) on speech perception in infants. For continuity, this material might have been better in a separate chapter.

In the second section, FBG deal with the production of sentences. The study of speech production happens to be an area of research in psycholinguistics that has not received very much attention by research workers. This is, of course, an unfortunate state of affairs since most of the theoretical work has been based on results obtained in perception and comprehension tasks. While it has tacitly been assumed that perception and production are symmetrical processes, there is very little empirical evidence one way or the other on the issue. FBG deal briefly with structural variables and surface structure constraints on sentence production. For the most part, sentence production appears to be a top-down, left-to-right process. Although FBG consider hesitations in speech and breathing and various types of speech errors in production, it is clear that our knowledge of the types of processes involved in speech production is very limited at the present time. It is also obvious that a full account of the sentence production process is not going to emerge from an analysis of only one class of empirical data. Moreover, there is good reason to suppose that an understanding of the various components, operations and strategies of sentence production will only come about when we gain a better understanding of the nature of semantic representations within some type of formal system. Although there are numerous unresolved problems in sentence perception and comprehension, the topic of speech production represents a virtual gold mine of possibilities for future research in psycholinguistics. I suspect that we will see a number of people re-directing their research interests to speech production over the next few years.

Chapter 8, 'First language learning,' is the last substantive chapter in the book. In the first section, FBG focus on some of the recent work which has attempted to teach language to chimpanzees, specifically, the studies of Washoe by the Gardners and Sara by Premack. FBG raise several important questions about the relevance of these sorts of comparative experiments to the study of language. While many linguists would probably not find the chimpanzee work terribly interesting, a number of psychologists have viewed these findings as strong evidence against the claims of species-specificity for language that have been advocated by Chomsky (1972) and Lenneberg (1967) and others. FBG consider four basic questions about the chimpanzee work. The first deals with what this line of work has shown us about chimpanzees. In reply, FBG say that it has shown us that chimpanzees are smart and intelligent and that they can do things that we didn't think they could. The second question concerns what these experiments have shown us about the chimpanzee's ability to learn human languages. FBG contend that neither Washoe nor Sara has mastered a productive system that is comparable to human language. They claim that neither chimp shows any evidence of constituent structure nor anything remotely resembling the use of transformational rules. FBG's third question is concerned with what the chimpanzee studies have shown us about the species-specificity for language learning in humans. FBG argue that even if the chimpanzee studies turned out to be successful, the outcome would not be at all relevant to the question since human language is innate. That is, language is represented in the genetic code. The real question in their view is whether chimpanzees learn to talk in the same way as humans do and the answer to this is obviously no. Finally, FBG's last question concerns what the chimpanzee studies have shown us about language. FBG's reply to this is a resounding "nothing." They argue that when humans learn language they exploit a genetic endowment that other organisms, in this case, chimpanzees, do not have. While I happen to be convinced by their arguments against the chimpanzee work, I think it might have been better to put them in a separate chapter on communication systems rather than at the beginning of the chapter on language learning.

In the second section, FBG consider some "theories" of language development. This is a fairly standard account of Chomsky's and McNeill's views on language acquisition. There is nothing new here that couldn't be found in original sources of five to ten years ago.

The third and final section is entitled 'some empirical studies of language acquisition.' This is a very sketchy account of some of Lois Bloom's work which has attempted to analyze the functions of early language within a transformational framework. The results of Carol Chomsky and Brown and Hanlon are also described briefly. There is a noticable disinterest by FBG in semantic development and the role of parental speech to children, two areas that have received intensive investigation over the last five years.

In the last section of the book, FBG consider how well their mentalistic program has done within the context of psycholinguistics. They reiterate the central theme that has appeared throughout the book, namely, that the objective properties of the stimulus environment have their effects mediated via the internal states of the organism. FBG state here that their theory of the internal structure of the speaker-hearer has two components. The first is called "conceptual analysis" and involves a characterization of the structure of the internal representation of the stimulus at each "psychologically" real level of analysis. The second component consists of a "theory of mental operations" which specifies the mapping rules by which one gets from one level of description to the next. These may be thought of as "computational routines" or psychological processes. For the most part, FBG conclude that the mentalist program in psycholinguistics has been successful. The behavior of the speaker-hearer is, to a large extent, dependent on the abstract internal representation of the physical stimulus. Moreover, the transformational approach has provided one means of characterizing these types of internal representation. In general, the claims for the psychological reality of the major structural relations postulated by transformational grammar have received fairly good experimental support. However, the claims for the psychological reality of grammatical processes and operations has not fared as well. The primary goal of future work in psycholinguistics, FBG tell us, is to provide an account of precisely how the psychological processes speaker-hearers employ are related to the linguistic structures formalized in the grammar. According to FBG, many of these problems eventually will be focused on questions about language acquisition and it is here, FBG speculate, that many of the solutions to these problems will be found. While it is no doubt true that language learning has presented numerous complex problems for contemporary learning theories in psychology and much remains to be done, I find FBG's appeal to the field of language acquisition as the panacea for our difficulties a bit distressing. A theory of the psychological (i.e., mental) operations in sentence recognition may be quite distinct from a theory of acquisition.

In final summary, this is a book that presents a very definite point of view about psycholinguistics. It is uniquely representative of the brand of psycholinguistics that has developed over the last ten years or so within the Chomskyan framework at M.I.T. Although not everyone will agree with FBG, in my view, the book is essential reading for both the linguist and psychologist who might want to delve into some of the basic issues in psycholinguistics.

REFERENCES

Anderson, J. R. and Bower, G. H. 1973. Human associative memory. Washington, D. C.: V. H. Winston.

Berlin, B. and Kay, P. 1969. Basic color terms: Their universality and evolution. Berkeley: University of California Press.

Blumenthal, A. L. 1967. Prompted recall of sentences. Journal of Verbal Learning and Verbal Behavior, 6, 203-206.

Bransford, J. D. and Franks, J. J. 1971. The abstraction of linguistic ideas. Cognitive Psychology, 2, 331-350.

Chomsky, N. 1957. Syntactic Structures. The Hague: Mouton.

Chomsky, N. 1965. Aspects of a theory of syntax. Cambridge, Mass.: MIT Press.

Chomsky, N. 1972. Language and Mind. New York: Harcourt Brace Jovanovich.

Chomsky, N. and Miller, G. A. 1963. Introduction to the formal analysis of natural languages. In R. D. Luce, R. R. Bush and E. Galanter (Eds.) Handbook of Mathematical Psychology, New York: John Wiley, Pp. 269-321.

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

Eimas, P., Sigueland, E., Jusczyk, P., and Vigorito, J. 1971. Speech perception in infants. Science, 171, 303-306.

Fillenbaum, S. 1966. Memory for gist: Some relevant variables. Language and Speech, 9, 217-227.

Gough, P. B. 1971. Experimental Psycholinguistics. In W. O. Dingwall (Ed.) A Survey of Linguistic Science. College Park, Maryland: Linguistics Program, University of Maryland, Pp. 252-292.

Kaplan, R. M. 1972. Augmented transition networks as psychological models of sentence comprehension. Artificial Intelligence, 3, 77-100.

Kintsch, W. 1974. The Representation of Meaning in Memory. Hillsdale, N. J.: Erlbaum Associates.

Lenneberg, E. H. 1967. Biological Foundations of Language. John Wiley and Sons.

Massaro, D. W. 1972. Preperceptual images, processing time, and perceptual units in auditory perception. Psychological Review, 79, 2, 124-145.

Miller, G. 1962. Some psychological studies of grammar. American Psychologist, 17, 748-762.

Sachs, J. S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. Perception and Psychophysics, 2, 437-442.

Wanner, E. 1968. On remembering, forgetting, and understanding sentences: A study of the deep structure hypothesis, Unpublished doctoral dissertation, Harvard University.

Woods, W. A. 1970. Transition network grammars for natural language analysis. Communications of the ACM, 13, 10, 591-606.

INSTRUMENTATION

SPEECH PERCEPTION RESEARCH LABORATORY: THE STATE OF THE COMPUTER SYSTEM

Jerry C. Forshee

Indiana University

This paper presents a description of the laboratory as it has developed during the past year. During this period the DEC computer system was delivered, two interfaces were constructed and put into service (one to process speech and one to interface to human subjects). A basic set of real time functions was also implemented for the on-line processing of speech.

Computer Configuration.

The configuration of the PDP11 computer system (see Figure 1) began as a packaged OEM (original equipment manufacture) system, the PDP11e/05. This packaged system offered the basic configuration that was needed, and purchasing through an OEM offered the prospect for a substantial savings. The system consisted of the processor (KD11B), with three integral options: 16K of 900 μsec. core memory (2 MM11-Ls), a line frequency clock (KW11L), and a console terminal port (Serial Communications Line). This packaged system also included several peripheral options: a 30 character/sec. hard copy terminal (LA30), a dual drive cassette tape system (TA11) which has a storage capacity of 92K bytes/drive, and a random access moving head disk system (RK11) with a capacity of 1.2 M words/drive. Also included in the system was a ROM bootstrap for the disk (BM792YB).

As the development of the laboratory progressed during the year, hardware features were added as needed by the development of our own hardware and software. Three parallel digital I/O modules (DR11-C), each consisting of 16 bits of digital input and output plus several control and status bits, were added to support the special purpose hardware that we developed. One module was added to connect the PDP11's Unibus with the subject Response

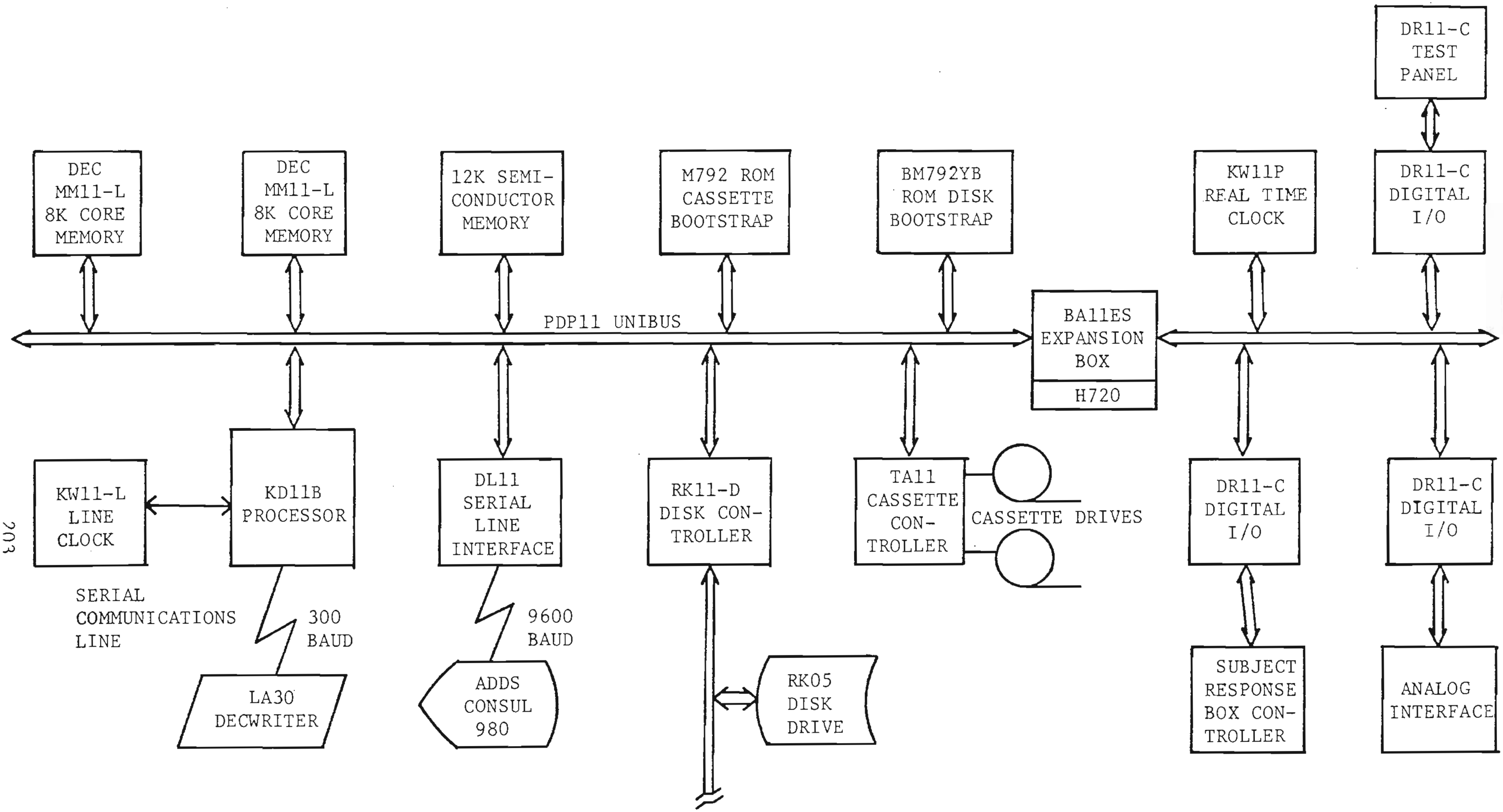


Figure 1. PDP11 COMPUTER SYSTEM CONFIGURATION

Box Controller, which interfaces up to 16 human subject stations. Another module was added for the Analog Subsystem Controller, which manages the speech processing functions on the system. The third was added to support a special digital I/O test panel which is used in testing the special purpose hardware and interfaces as they are developed.

To increase the throughput of the system, primarily for text editing and secondarily for increasing the speed of the dialog between the console terminal (CPU) and the user, a high speed CRT terminal (ADDS Consul 980) has replaced the LA30 as the console terminal. The addition of an asynchronous serial line interface (DL11) allows this CRT terminal to be connected to the computer. All user computer dialog and text editing is done on the high speed (9600 baud) CRT terminal. Trial listings from FORTRAN and assembly programs are also done on the high speed terminal. When hard copy is desired a special program is called up to spool out a listing file from the disk to the 300 baud hard copy terminal. This is done as a foreground job and allows background activity such as text editing, program assemblies or compilations to proceed seemingly without interruption.

The collection of subject response latencies to the resolution of milliseconds necessitated a clock with greater speed than the KW11L. A programmable real time clock (KW11P) was added for this function. In addition to timing the latencies for up to 16 subject stations, this clock is also used to precisely synchronize the presentation of stimuli in a dichotic presentation where there is a temporal lag between stimuli presented to the two ears. Several other miscellaneous timing functions are served by this clock: inter-trial intervals, inter-stimulus intervals, inter-block intervals, etc.

For reasons of both convenience and speed a ROM (M792), which we programmed for a cassette bootstrap, was added to the system. All of the diagnostic programs supplied by DEC were delivered to us on cassette tapes, and the addition of a cassette bootstrap helps speed up the procedure whenever it is necessary to run diagnostics.

We found that the size of our real time experimental control programs with the necessary buffers for storing speech waveforms was greater than the available 16K of core memory. To alleviate this problem and expand the system to its maximum memory size of 28K, a 12K semiconductor memory system (Monolithic Systems) was added to the system. Not only has this been a very cost effective solution, but we have found 28K of memory to be sufficient for all of our activities.

This gradual expansion of the system one or two options at a time soon came to exceed the power and physical space available in the CPU mounting box. It was necessary to add an expansion box (BA11ES) and power supply (H720) to house several of these options. This expansion system allows for the addition of a total of 24 additional interface boards (Small Peripheral Controllers).

<u>Response Box Controller</u>.

The Response Box Controller (RBC) serves to interface the human subject to the computer. It provides two essential services. First, it provides output of cue and feedback information to subjects. Second, it codes and buffers the responses made by the subjects. The RBC has the capacity to control from 1 to 16 subject stations. Each of the several stations may be equipped, for output, with from 1 to 16 discrete output event lights and, for input, with from 1 to 16 discrete pushbutton switches or an ASCII keyboard. The response manipulanda option can be varied easily by choosing one of several different response boxes. Each of the several different boxes is designed with a particular experimental paradigm in mind, i.e., two button for two choice discrimination, six button for confidence rating, ASCII keyboard for recall, etc.

The RBC logic is divided into three major functional sections, (a) output to response boxes, (b) input from response boxes, and (c) front panel control logic. To operate a given event light on a response box the program puts certain information on the output word of the DR11-C. This information consists of a 4 bit box address, a 4 bit light address and a 1 bit function code (on or off). Two additional lines are provided to minimize operations. A control line to select all boxes can be activated which then overrides the box address and operates the lights addressed according to the function line specification. Also, a line to select all lights can be activated to override the 4 bit light address, thus allowing all lights of the addressed box to be operated simultaneously according to the function line. The RBC receives the information transmitted by the program, decodes the box address and enables the selected box, and passes along the light address and function codes (see Figure 2). This information received by the response box is then acted upon. The light address is decoded and the latch of the light(s) selected are set or cleared according to the function line specification (see Figure 3).

Input to the RBC from the response boxes proceeds as follows: the response box codes the button or key press into a 6 bit binary code. This information along with a strobe signal is transmitted to the RBC where the information is stored until read by the computer. When

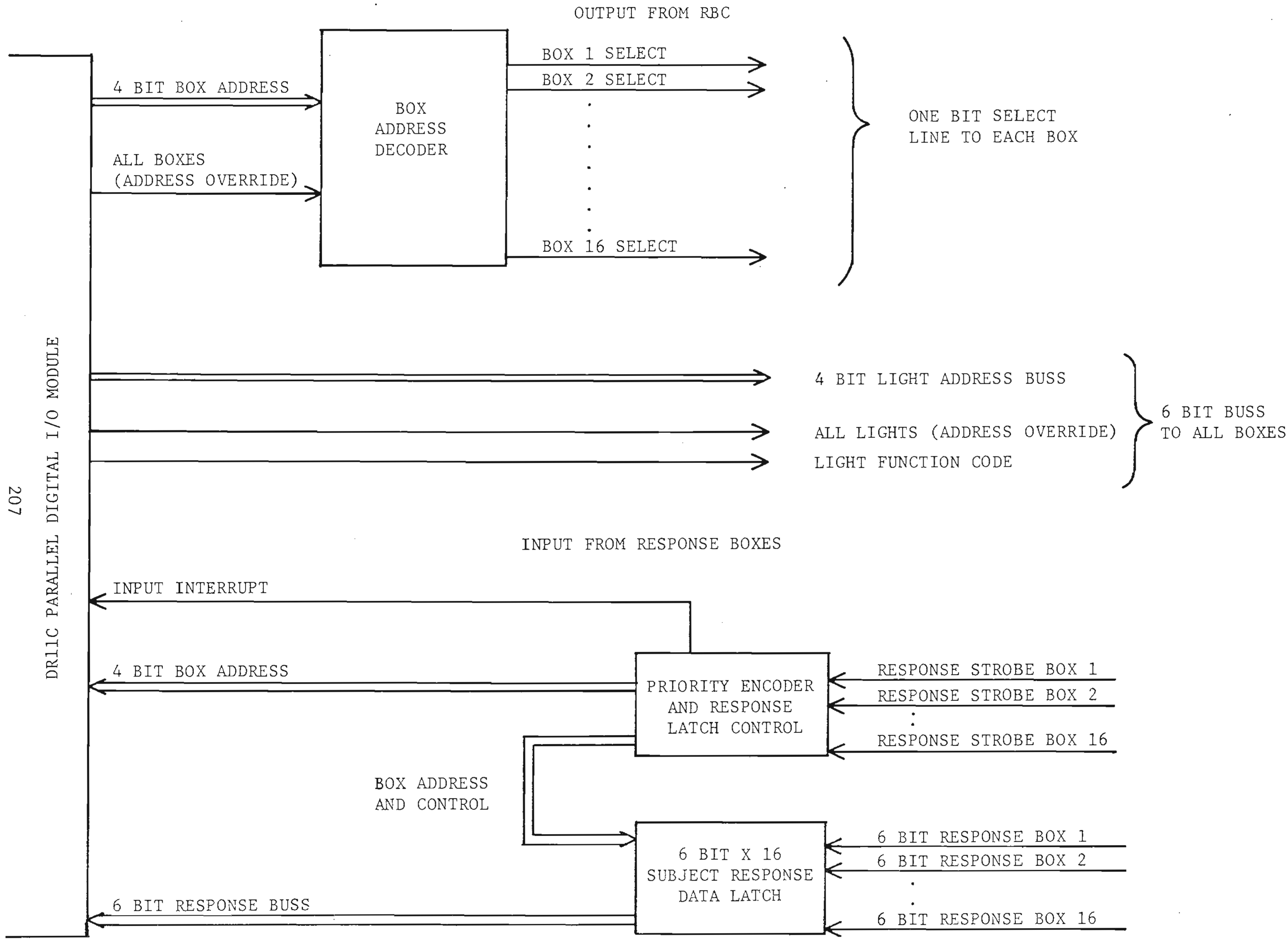


Figure 2. BASIC BLOCK DIAGRAM OF RESPONSE BOX CONTROLLER (RBC)

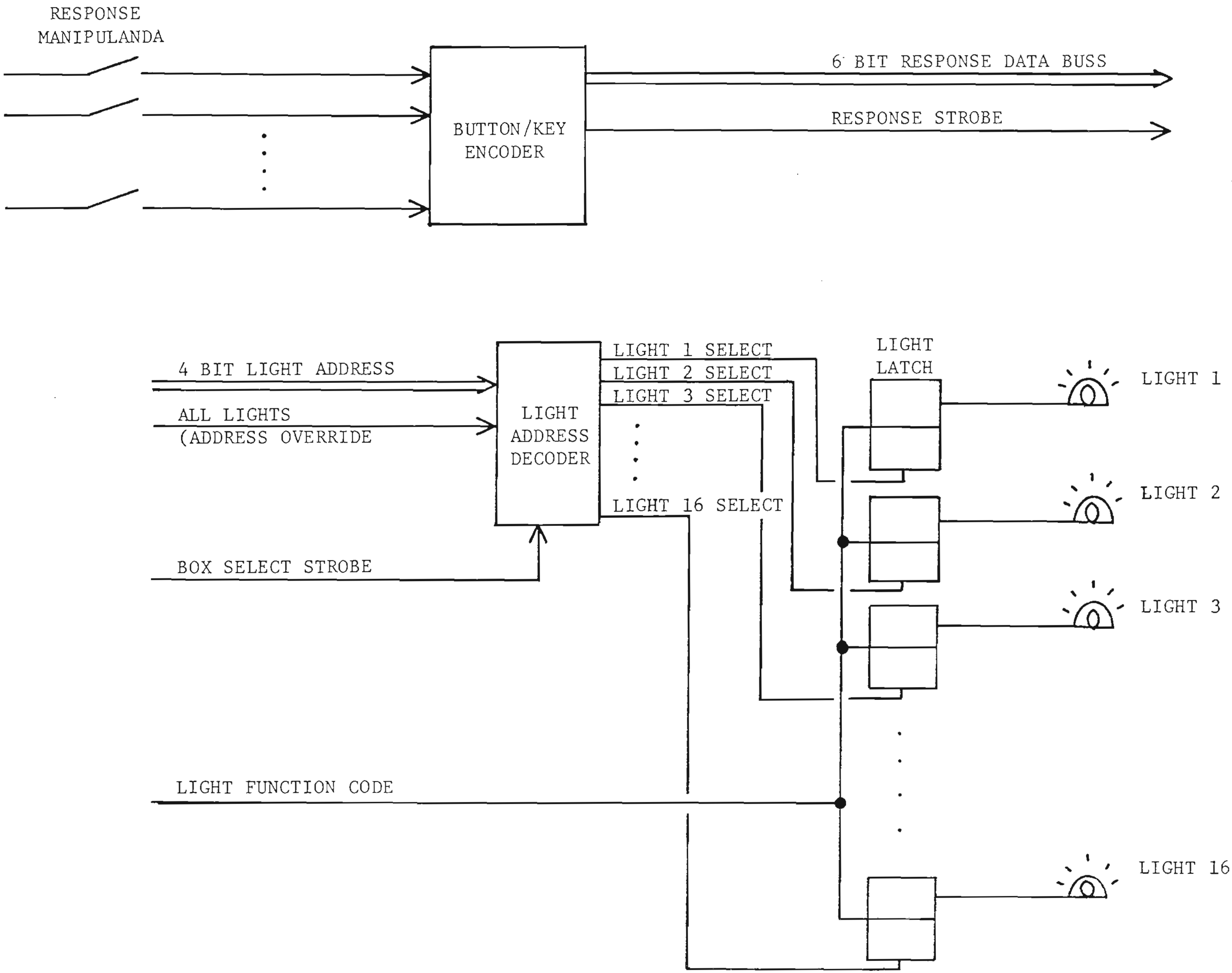


Figure 3. BASIC BLOCK DIAGRAM OF RESPONSE BOX LOGIC

commanded by the RBC control program, the RBC codes the box address and then transmits the box address and coded response to the computer (see Figures 2 and 3).

The front panel logic of the RBC is a very powerful self-diagnostic tool. It operates in either an off-line or an on-line mode. In the off-line mode the front panel controls can be used to manually perform all functions of the RBC. Outputs to any box and/or any light can be addressed and turned on or off, also the all lights and all boxes functions can be used. This allows both the output devices at the subject stations and the output address system to be completely checked without the need of the computer. Inputs from the response boxes can also be verified. The user can move from station to station pushing each button. When a button is pressed, the response data is displayed on the front panel and remains displayed until cleared manually at which time the next button can be verified. In the on-line mode the front panel is very useful in visually monitoring the action of subjects. A given subject station or all stations can be selected for monitoring. When a selected station has observed a response, the data is not only transmitted to the computer, but is also displayed at the front panel, where it can be observed until the next response is recorded and displayed.

Analog Subsystem Controller.

The analog interface implements several distinct functions combined physically as a single integral unit. The Analog Subsystem Controller consists of the following functional units: (a) a two channel 12 bit digital to analog (D/A) converter, (b) a one channel analog to digital (A/D) converter, (c) a buffered interface for the OVE IIId Speech Synthesizer,

(d) a two channel audio mixer and amplifier, and (3) two programmable attenuators.

The interrelationship of these five major functional sections of the Analog Subsystem Controller is schematically presented in Figure 4. The first step in processing an analog signal is to select the analog (audio) inputs to which the system is to attend. Each channel (referred to as A and B) has an input mixer which can select any combination of its ten inputs. The analog signals assigned to these inputs are: (1) the synthesizer output, (2) D/A_1 output, (3) D/A_2 output, (4) audio tape recorder channel A output, (5) audio tape recorder channel B output, (6) tone generator 1, (7) tone generator 2, (8) noise generator 1, (9) noise generator 2, and (10) microphone. After the analog input(s) have been selected they may be digitized or passed through the programmable attenuators and directly out to the subject earphones.

The two programmable attenuators, one for each channel, may be set to any level of attenuation from 0 to 63 dB in 1 dB steps. In addition to being under program control by the computer, the attenuators can be set manually from the front panel of the interface with thumbwheel switches. Mounted next to these switches each attenuator has a two-digit display (seven segment LED type) which displays the selected attenuation level, in dBs. This display is updated whenever the attenuation level is changed, either by program or manually.

When activated, the set of analog inputs selected by the A channel (A channel only) input mixer may be digitized. The A/D converter takes samples at the rate of 20 khz, one sample every 50 μseconds. The A/D converter is a Datel model 12B2D3, a 12 bit converter with a throughput rate of 20

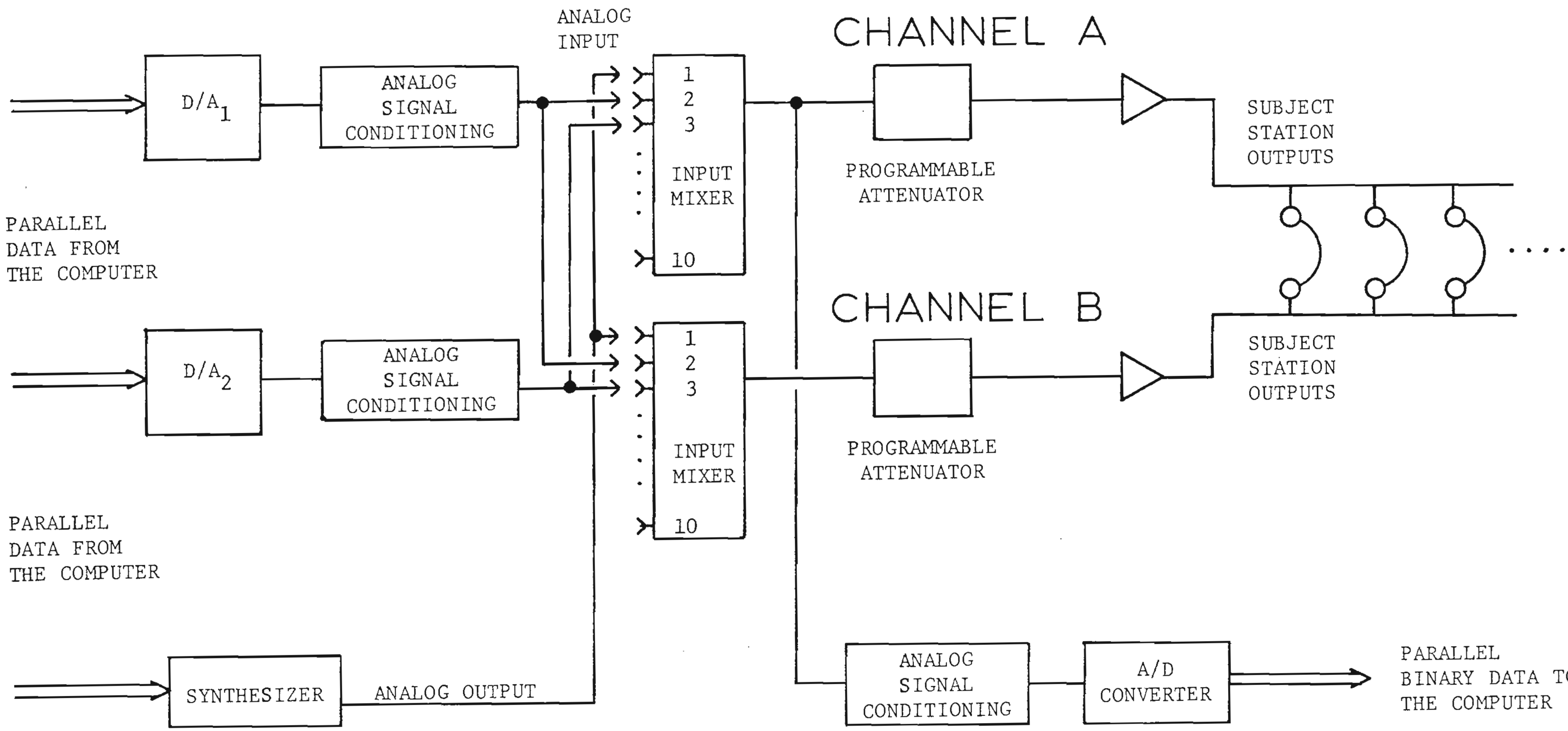


Figure 4. BLOCK DIAGRAM OF THE ANALOG SUBSYSTEM CONTROLLER

μseconds. Prior to being digitized an analog signal is first filtered through a 10 khz low pass filter. It is then passed through a preemphasis amplification circuit to selectively amplify the higher frequencies so as to increase the signal to noise ratio in these regions. The filters used are 7 pole, 10 khz low pass filters supplied by White Instruments which have an amplitude roll off of greater than 85 dB/octave. The preemphasis circuit was originally developed at MIT's Lincoln Laboratory and subsequently modified at Haskins Laboratories. This modified circuit was provided to us by Mr. David Zeichner and has been implemented here.

Three of the devices providing analog inputs to the input mixers produce analog signals under computer control. These devices are the speech synthesizer and the two D/A converters. The synthesizer is an OVE IIId parallel driven serial model speech synthesizer. The computer outputs a set of values every synthesis time period, one for each of the 15 synthesizer input parameters. The 15 digital parameters to be sent to the synthesizer are first sent to the interface one by one where they are buffered. The synthesizer does not have a provision for loading all 15 parameters and then commanding the synthesis to start; rather, as each parameter value is received at the synthesizer that new value immediately contributes to the complex synthetic speech waveform output. Since we want to minimize this gradual shift in the synthesizer output, all 15 parameter values are first loaded into a first-in-first-out (FIFO) buffer. The time to format and transfer all 15 values takes on the order of 800 μseconds. Then the interface is given the command to dump the FIFO to the synthesizer as fast as it can take the values (approximately 1/μsecond).

The D/A converters are used to playback previously digitized synthetic or real speech utterances which have been stored previously in digital form on the random access disk memory. The D/A converters are Datel's model L12B2D3. This D/A playback system follows the same sampling rate and signal conditioning protocols as for the digitization process. Also, the signal is reconditioned to match the original signal that was digitized by use of a deemphasis circuit that complements the preemphasis circuit used at digitization time. This deemphasis circuit serves to attenuate the higher frequencies to their original amplitude level. The circuit was also developed at Haskins Laboratories.

Care was taken in the construction of the D/A and A/D modules because of the 20 khz data rate and the importance that there be no cumulative time skew or lost samples. To insure an accurate sampling rate, a 100 khz quartz crystal clock circuit was implemented with appropriate logic to generate a 20 khz, 50% duty cycle clock. It is this clock that controls the timing of critical events. For the D/A conversion this involves loading of the converter with another digital value and for the A/D conversion it is the initiating of the sample taking. It was necessary to build into the interface a certain amount of buffering since the 1 word per 50 μseconds is close to the speed of the PDP 11/05 under direct program control for the operations necessary in that time span. Data transfers are interrupt driven and output is accomplished by moving data pointed to by a program location and then incrementing that location. This process may closely approach the maximum time available depending on the exact programming sequence necessary. To help ease the time strain problem each device (the two D/A's and the A/D) has its own hardware FIFO buffer of 64 words. These buffers are built around the Fairchild 3341, 4 bit by 64 word MOS FIFO buffer chip, which features completely independent inputs and outputs. This makes it ideal for an application, such as this, where the input and output of data will often be asynchronous. In both D/A and A/D cases this asynchronous buffering allows the buffer to be accessed precisely every 50 µseconds by the clock driven activity (output for D/A and input for A/D) while being accessed asynchronously, as fast as possible by the computer program. This allows the program to dump the entire contents of the FIFO once the context switch to it has been made and its tables and pointers are set up. Using this technique there has been no problem in achieving a continuous 20 khz sampling rate under direct program control.

Program Supported Research Activities.

Considerable progress has been made during this first year of development. We have been successful in implementing the full range of activities that allow for the on-line generation of synthetic speech, the digitization of speech waveforms, the presentation of these speech stimuli in real time to human observers and subsequent collection and analysis of the responses of these observers.

The first step in this process is the generation of synthetic speech. To generate a synthetic utterance the user interacts with the computer system by means of the OVE control program, OVEXEC. The waveform produced by the OVE synthesizer, at any instant, is the synthesis resulting from the values input for each of its 15 parameters. These parameters form the basis of the speech synthesis. The OVEXEC program allows the user to generate speech utterances by constructing files of parameter values which drive the OVE during synthesis. The commands of the program have been divided into four major categories:

1. Parameter and line modification. These commands allow the user to create a table of parameter values. Parameter values can be entered by lines; values or lines can be inserted, deleted or modified. For a series of lines one or several values can be interpolated between an initial and terminal value.

2. Line and file display. The contents of the current file can be output to the CRT terminal 20 lines at a time or the entire file can be printed out on the hardcopy terminal. There is also a command in this category to look up and output the OVE stimulus file directory.

3. I/O Operations. The commands in this category are used to operate on disk and cassette files. Stimulus parameter files may be input from cassette tape to the disk or output from the disk to cassette tape. Files on the disk may be input to the program for editing, appended to the current parameter buffer in the program, deleted, named or renamed. Once a parameter set is created in the program buffer, it can be output to the disk, catalogued and made permanent.

4. Generating Synthetic Speech. These commands are used to output the parameter files through the synthesizer. A single file can be output either repeatedly or a single time. A series of parameter file names can be entered and then output to the synthesizer as a continuous series (with appropriate ISIs). This allows the user to listen to an experimental sequence of stimuli. The output of the OVE can be directed to an Ampex (AG-500) tape recorder, and an audio tape can be made of the stimuli for use in off-line experiments, or for later playback into the analog interface for digitization.

The second major step in preparation for running an on-line speech perception experiment is to make the stimuli available in digital form. The source of the waveform to be digitized is usually audio tape, although any analog signal is acceptable, i.e., human speaker, synthesizer, tone generator, noise generator, etc. When digitizing from audio tape a noise sampling program is used to determine the level of ambient noise on the audio tape. The data on the tape's noise level is then input to the digitization program. This information is necessary because the program must be able to distinguish between the tape's ambient noise level and the beginning of the stimulus to be digitized. The 20 khz data rate and the speed of the disk mechanics, along with the software overhead, make disk buffering and simultaneous digitizing a marginally compatible activity. The digitizing program continuously samples the input, putting the samples into a ring buffer, constantly monitoring the digitized values to determine when the input transitions from noise to stimuli. When this transition is sensed the position is noted. Sampling is then continued until that position (less one location) is again reached in the ring buffer. Thus, the operation of synchronizing the digitization process and the mechanical starting of the tape recorder is eliminated. All that is necessary is to start the digitization program prior to starting the motion of the tape recorder; and the program then synchronizes the sampling of the actual stimulus.

Once the stimulus has been digitized, it is represented on the system as a disk file which can be edited. The editing of the digital waveform is necessary to eliminate the few samples of noise at either end of the actual stimulus. The process of editing the digital waveform is somewhat laborious at present since no graphical output of the waveform is yet available. Thus, the samples have to be printed out and examined individually by hand. This process is verified by re-outputting the waveform through the D/As and examining the analog waveform on a memory oscilloscope.

The third step in carrying out an experiment on the system is to write the FORTRAN program that actually conducts the on-line experiment. The programmer, for this task, has available to him a large library of assembly language subroutines which perform all the necessary real time functions to (a) present stimuli to subjects, (b) time stimulus intervals, and (c) collect subject responses and their latencies.

Stimulus information may be presented to subjects, either as acoustic stimuli or lights. Subroutines are available for presenting a digital waveform either once or several times in succession. These stimuli can be modified in real time by selecting different levels of attenuation or mixing it with a signal from a continuous source, i.e., noise generator, tone generator. Pure tones can also be generated at any desirable frequency and output by using the D/A converter. The subjects' stimulus light panel can be turned on or off individually by subject station or across all subject stations simultaneously.

The program interface to the subject stations is also provided by a set of subroutine calls. These calls are divided logically into two groups: routines which are called only once and those that are called every trial. The routines which are called only once are used prior to the very first trial to initialize the system and provide any information that may be unique to the particular experiment session, i.e., the number of subjects present, occupied stations and specification of unique subject numbers.

The subroutines that are used on every trial are used to activate/deactivate the response manipulanda, start/stop latency intervals, fetch responses and latencies from the system, etc. In the following example, from a selective adaptation paradigm, 9 stimuli are presented in random order for 10 repetitions, with each repetition being preceded by an adaptation sequence of 100 repetitions of the adapter stimulus.

```
10        ITRIL = 0
20        DO 1000 IREP = 1, 10
30        DO 2000 IADPT = 1, 100
40        CALL UTTER(N)
50        CALL MRKTIM(300)
60   2000 CONTINUE
70        CALL MRKTIM(IAI)
80        DO 1000 ISTIM = 1, 9
90        ITRIL = ITRIL + 1
100       CALL ALLOW(IALL)
110       CALL RTSTRT(IALL)
120       CALL UTTER(IORDER(ITRIL))
130       CALL MRKTIM(ISI)
140       CALL IGNOR(IALL)
150       CALL RTSTOP(IALL)
160       DO 1000 ISUB = 1, 6
170       CALL RESP(ISUB,IDATA(ISUB,ITRIL,1))
180       CALL LATEN(ISUB,IDATA(ISUB,ITRIL,2))
190  1000 CONTINUE
```

The DO loop at line 20 sets up the repetition iteration. Lines 30 through 60 present the adapter stimulus. The call to UTTER at line 40 outputs the stimulus whose digital waveform is in stimulus file N. Line 50 serves to space out the repetitions of the adapter stimulus by producing a 300 millisecond delay. Line 70 provides for an additional delay to separate the presentation of the adapter stimulus from the test stimuli. Line 80 sets the iteration for one repetition of all 9 stimuli. Line 90 simply increments the cumulative trial counter. Lines 100 and 110 start a response interval, ALLOW activates the response manipulanda and RTSTRT is used to initialize the latency interval; the parameter IALL is defined to indicate that the desired function is to be carried out for all subjects. Line 120 outputs a stimulus from a previously randomized list. MRKTIM at line 130 provides the inter-stimulus-interval. Lines 140 and 150 are used to terminate the response interval. Lines 160 and 190 are used to fetch from the system each subject's response and latency which is stored in a program array.

With the complicated program sequencing and sophisticated interrupt servicing (for D/A transfers and subject responses) provided by assembly language subroutines, we have found FORTRAN to be a most suitable language for controlling experiments. FORTRAN has the advantage of being almost universal, and once the new laboratory user becomes familiar with the subroutine library it becomes quite easy to program any particular experiment.

To date we have implemented several standard paradigms used in speech perception work. These paradigms include: identification, adaptation, discrimination, same-different reaction time, recognition masking and training. These paradigms are fully implemented including data recovery and summary statistical procedures.

Future System Development.

Several expansions of the present computer configuration are planned. We have already become conscious of the limited processing power of the PDP11/05 in speech perception research. To date this has been only an awareness not a serious problem. However, as the set of supported activities expands, we expect this to become more noticeable. A KE11-A, Extended Arithmetic Option, has been ordered to enhance the processor. This option adds multiple shifts and integer multiplication and division to the instruction set. This additional power will facilitate the development of simultaneous synthesis and digitization, digital synthesis and digital filtering during the coming year.

In order to increase our capacity to process and store speech waveforms both on-line and off-line, we are planning the addition of a second disk and a magnetic tape drive (industry compatible). This disk will be used to store our library of acoustic stimuli, i.e., real and synthetic speech, tones, clicks, and other hybrid stimuli. The magnetic tape will be used to provide an easy transfer onto the system of all the private files that belong to a user as he logs onto the system. This tape will also be used to store long passages of speech that are too voluminous for the disk storage.

Expansions are also planned to add hardcopy and graphics output facilities to the computer system. The addition of a VR14, point plot scope, will allow the user to view the digital waveform he is working with. This feature will be especially useful for editing digitized waveforms and for viewing waveforms produced by digital algorithms. Hardcopy of graphics output will be possible with the addition of a Versatec printer-plotter. When the Versatec is not being used as a plotter, it may be used as a high-speed line printer for program listings, data dumps, analysis summaries, and any other printed output requirement.

A Description of the OVE IIId Control Program: OVEXEC

James R. Sawusch

Indiana University

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 the analog interface which is described elsewhere in this volume by Forshee (1975). 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 the DECwriter (30 line per minute teletypewriter). 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 (such as the vowels of Peterson and Barney, 1952) and implement them for the OVE.

Summary

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.

REFERENCES

Peterson, G. H. & Barney, H. L. Control methods used in a study of the vowels. _Journal of the Acoustical Society of America_, 1952, _24_ 175-184.

Forshee, J. C. Speech perception research laboratory: The state of the computer system. _Research on Speech Perception Progress Report_, Indiana University, Bloomington, Indiana, 1975, _2_,

Publications:

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

Pisoni, D. B. Review of "Acoustic cues for constituent structure" by Robert Scholes. Linguistics, 1975, 146, 104-106.

Pisoni, D. B. Auditory short-term memory and vowel perception. Memory & Cognition, 1975, 3, 1, 7-18.

Pisoni, D. B. Contemporary issues in speech perception. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.), Cognitive Theory: Volume I. Hillsdale, N. J., 1975, 1-4.

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., 1975, 79-102.

Manuscripts to be published:

Pisoni, D. B. Information processing and speech perception. In G. Fant (Ed.) Proceedings of the 1974 Speech Communication Seminar. New York: John Wiley, 1975 (In Press).

Pisoni, D. B. Review of "The Psychology of Language" by J. A. Fodor, T. G. Bever & M. F. Garrett. Language (In Press).

Pisoni, D. B. & Sawusch, J. R. Some stages of processing in speech perception. In A. Cohen & S. Nooteboom (Ed.), Structure and Process in Speech Perception. Heidelberg: Springer-Verlag, 1975 (In Press).

Pisoni, D. B. & Tash, J. Auditory property detectors and processing place features in stop consonants. Perception & Psychophysics, 1975 (In Press).