ATTENTION DEMANDS OF PROCESSING PHONETIC INFORMATION IN THE PERCEPTION OF DICHOTIC SPEECH

Thesis by

David Saul Isenberg

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1977

(Submitted February 25, 1977)

ACKNOWLEDGEMENTS

The author extends sincere thanks to:

Dr. Sheila Blumstein, Dr. James Bonner, Dr. Harold Goodglass and Dr. Edgar Zurif for their continuing and patient support throughout all phases of this work.

Mr. Errol Baker for help with statistics.

Dr. E. A. Boling for help with software and hardware.

Ms. Cheryl Clark and Mr. Brian Glynn for help running subjects.

Dr. David Pisoni, Dr. Bruno Repp and one anomymous reviewer for insightful comments on earlier phases of this work.

Dr. Alvin Liberman for making the facilities of Haskins Laboratories available, and Ms. Cecelia Dewey and Dr. Terry Halwes for instruction in the use of those facilities.

The entire office staff of the Division of Biology.

The following people who helped make this work easier in many diverse ways; Dr. Laird Cermak, Dr. Charles R. Hamilton, Dr. Michael Trachtenberg, Ms. Dahlia Zaidel and Dr. Eran Zaidel.

All those who served as subjects in the present experiments or related work for their fortitude and endurance.

This work was supported by a USPHS Grant (GM-00086) awarded to California Institute of Technology, by NIH Grants (MS-11408 and MS-06209) awarded to Boston University School of Medicine, and by an NIH contract (N01-HD-1-2420) awarded to Haskins Laboratories.

ABSTRACT

We know intuitively and from dichotic shadowing studies that we must actively listen for a message carried by speech to enter consciousness. Is such active listening necessary to process phonetic information? Theories of speech perception which have been developed to account for certain facts of acoustic phonetics - notably the lack of invariant or segmented acoustic forms corresponding to phonemes - make implicit or explicit assumptions that cognitive processes are involved in the mental encoding of phonetic information which are thought to require attention. On the other hand, mental encoding operations which have been studied appear to proceed automatically.

In order to explore this question, two studies employed dichotic listening in conjunction with a secondary digit memory task to investigate claims that phonetic distinctive features of stop consonants require capacity in short-term memory (STM) in dichotic speech perception. Experiment I found no interference of a dichotic two-ear identification task upon STM contingent upon number or type of feature contrast of the dichotic pair. Interference with STM was found in a dichotic discrimination task for pairs which contrast on place alone. In the absence of such differences for the identification task, these results could not be interpreted to reflect demands of perceptual processing. Experiment II - designed to rule out certain artifacts - replicated the negative results of the identification task in Experiment I.

Experiment III used a probe reaction-time task to assess demands upon limited capacity during a dichotic one-ear stop consonant identification task. No effect of number or type of feature contrast upon probe reaction-time was found for non-identical dichotic pairs. A difference in probe reaction-time between identical and non-identical pairs was attributed to the necessity of response selection in the latter case. Experiments I, II and III, taken together, demonstrate that attention is not necessary for processing phonetic information in speech perception.

Automatic processing of stop consonants was demonstrated in Experiments IV and V. A dichotic phoneme monitoring task was employed to direct attention to one ear, and selective adaptation along the voicing dimension was used to measure processing contingent upon the phonetic contents of the non-attended ear. Large effects of the non-attended channel upon selective adaptation were interpreted to reflect automatic speech-related processing of that channel.

To the extent that active theories of speech perception may be construed to predict attentive processing, the present studies are taken as disconfirmation of such theories. Expansion of the search for acoustic-phonetic invariants and exploration of the interaction of higher linguistic levels with phonetic processing are proposed as two avenues of approach toward a viable passive theory of speech perception.

An appendix explores several different dichotic feature effects found in the present studies in terms of processing differences contingent upon type of feature contrast.

TABLE OF CONTENTS

I. INTRODUCTION

We know from our own intuition and from studies of attentional processes in listening to running speech (e.g. Cherry, 1953) that the perception of speech generally requires our attention - we must actively listen rather than passively hear for a message carried by speech to enter consciousness. The present discussion is motivated by the question of the role attention - the deployment of mental effort which often accompanies certain cognitive operations - plays in the process of speech perception, and more specifically, whether or not attention is devoted to the recovery of phonetic information from the acoustic signal of speech. In normal situations when we listen to speech we are not directly aware of the phonetic elements of the conversation. Instead, we are primarily aware of meaning (Cutting & Pisoni, 1975). However, when we look at the complex and intricate way that phonetic information is represented in the acoustic speech signal (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967), we see that the abstraction of such information from that signal must be a highly complex process. Furthermore, the major theories of speech perception (Liberman et al, 1967; Stevens, 1972) contain explicit or implicit assumptions of processes such as hypothesis testing and response selection which are thought to require attention (Neisser, 1967; Kerr, 1973).

Before we can begin to explore the question of attention in the perceptual processing of phonetic information, though, we should examine more closely how phonetic information is represented in the acoustic speech signal and what theories have evolved to account for its perception. By the same token, we should also survey how attention is deployed in various cognitive tasks and what its role is thought to be in human cognition.

The problem of acoustic representation of phonetic information.

The source of the acoustic speech signal is the flow of air from the lungs which produces a periodic sound when the glottis is constricted and a non-periodic turbulance when constrictions are produced by the articulators (tongue, teeth and lips). Relatively pure examples of these types of sounds are the vowel /a/ and the fricative consonant /s/, respectively. As a general rule, however, running speech contains a successive array of periodic and non-periodic components. The higher frequencies of these components are modulated by the shape of the oral cavity. The sounds of speech may be represented visually as an oscillogram, but they are usually represented by a frequency-intensity plot across time called a spectrogram. The spectrogram has been the major tool in acoustic phonetics for over three decades (Koenig, Dunn & Lacy, 1946). It conveniently represents the time-varying spectrum of speech to the eye as dark areas in those frequencies where acoustic energy is concentrated. Where the energy is concentrated around certain frequencies over some period of time the dark bands in the spectrogram are called formants.

The development of a machine called the pattern playback allowed speech spectrograms to be reconverted to auditory signals (Cooper, Delattre, Liberman, Borst & Gerstmann, 1952). This development made it possible to study which parameters of the acoustic signals carry phonetic information. Simplified spectrograms were constructed and varied systematically in a search for the acoustic properties necessary and sufficient to perceive specific speech sounds. It was discovered that the perception of certain phonemes, such as vowels, was cued by formant patterns that were relatively invariant across different contexts (Delattre, Liberman, Cooper

& Gerstmann, 1952). Other phonemes were found to be carried on an acoustic signal that varied greatly as a function of the neighboring segments. A simple example of this phenomenon is illustrated in Fig. 1. This figure shows the simplified spectrograms sufficient to produce the consonant-vowel (CV) syllables /di/ and /du/ on the pattern playback. The steady state portions of the stimuli are sufficient to produce the vowels /i/ and /u/. However, nowhere is there to be found an acoustic event that uniquely corresponds to the percept /d/. If only the initial formant transitions of these acoustic signals are played they are heard as non-speech pops or clicks. Similarly, if only the second formant transitions are played in isolation they sound like rapid respectively rising and falling whistle sounds or "chirps" (Liberman et al, 1967).

Thus, equivalent phonetic percepts may be based on radically different acoustic information. Different acoustic events that are categorized by the same phoneme may be said to be allophones of that phoneme. For example, the initial sounds in the English words "deed" and "dude" are allophones of the same phoneme. That is to say that even though the acoustic structure differs, the differences are not phonetically relevant, for in both cases the speaker intends and the listener perceives the same initial sound. The words "bead" and "deed" on the other hand, are both acoustically and phonetically different in their initial segments.

The fact that there is no invariant or isolable acoustic segment corresponding to the perceptually isolable phoneme /d/ in the (now famous) example of /di/ and /du/ has been an important insight of acoustic phonetics, and it now constitutes a major problem for the theory of speech perception (Liberman et al 1967; Stevens, 1960, 1972; Studdert-Kennedy, 1975).

Figure 1.1

Formants sufficient to produce /di/ and /du/ on the pattern playback. (The figure is redrawn from Figure 1 of Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967).

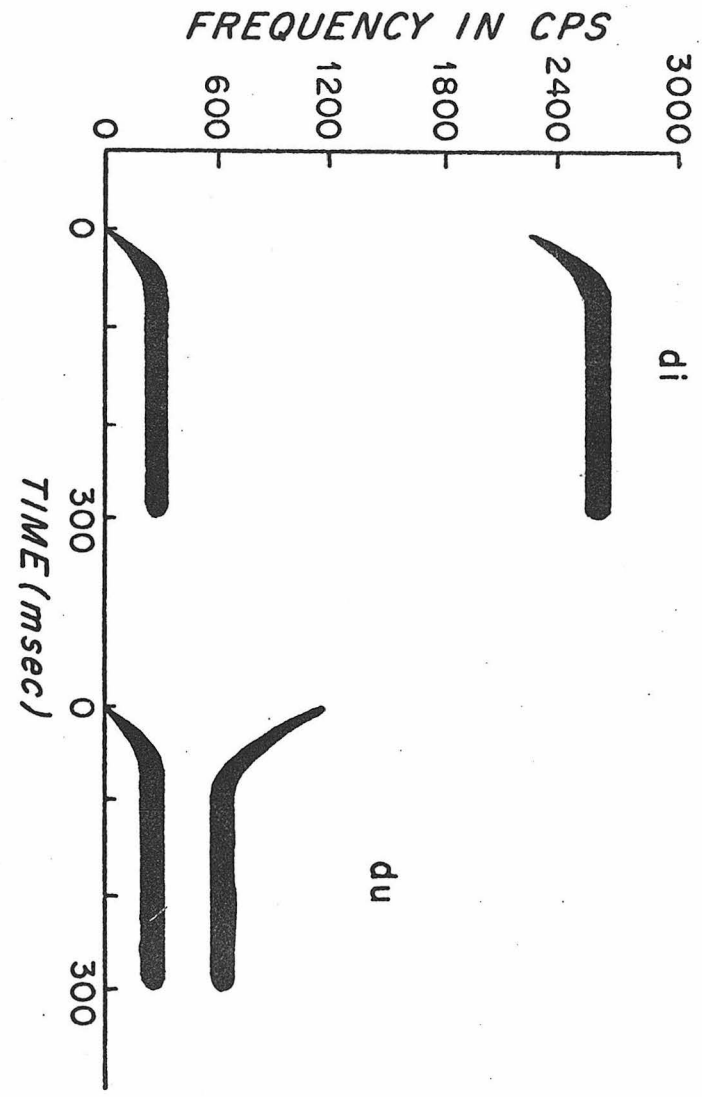

FREQUENCY IN CPS
3000
2400
1800
1200
600
0
di
du
0
300
0
300
TIME (msec)

FIGURE 1.1

The literature abounds with other examples of the phenomenon of lack of invariance. The syllable /id/ can be produced by the mirror image of the spectrogram in Fig. 1 for /di/ (Liberman, Delattre, Cooper & Gerstman, 1954). In the latter the transition of the second formant rises over time and in the former it falls over time, but the percept /d/ remains.

In fact, it is difficult to find general examples of phonetic-acoustic invariance in speech (but cf. Cole & Scott, 1974). Possibly the only candidates are the fricatives and the stressed vowels (Studdert-Kennedy, 1974). But in running speech the acoustic forms of even the vowels change as a function of their neighboring consonants (Lindblom, 1963).

"Thus, in general, the acoustic cues for successive phonemes are intermixed in the sound stream to such an extent that definable segments of sound do not correspond to segments at the phoneme level. Moreover, the same phoneme is most commonly represented in different phonemic environments by sounds that are vastly different. There is, in short, a marked lack of correspondence between sound and perceived phoneme. This is a central fact of speech perception." (Liberman et al, 1967, p. 432).

How then, does man perceive the phoneme as perceptually invariant when it is carried on a physically variable signal, and as a discrete entity when there are no corresponding discrete acoustic units?

Theories of speech perception.

One hypothesis (Wickelgren, 1969) assumes that there are central phoneme detectors consisting of one or more neurons for every allophone in every phonemic context. While elegant in its simplicity, this notion is exceedingly uneconomical, especially when one considers that phonemic

segments can influence the acoustic form of other segments which are not their immediate neighbors (Kozhevnikov & Chistovich, 1965). Halwes & Jenkins (1971) calculate that if such coarticulation effects exist two phonemes removed from the target there must be as many context-sensitive allophones as there are neurons in the brain. This observation alone renders the theory untenable.

Another attempt at a passive theory of speech perception is based on the notion that the syllable (Massaro, 1972) or syllabic nucleus (Cole & Scott, 1974), rather than the phoneme or context-sensitive allophone, is the basic unit of speech perception. Such an assumption appears at first glance to be a valid way to reduce the number of invariant recognition units. Coarticulation effects, however, exist between syllables as well as within them (Treon, 1970). Furthermore syllables, like phonemes, do not exist as discrete acoustic segments (Mermelstein, 1975). Still further, the acoustic form of the vowels, the phonetic correlates of syllabic nucleii, vary as a function of rate of speech (Lindblom, 1963) and across speakers (Fant, 1966). Thus, a syllable-oriented passive speech perception theory appears to have the same types of drawbacks as a passive phoneme-oriented theory. Clearly a template-matching type approach to speech perception is inadequate.

The fact that the acoustic signal of speech is produced by articulation and the observation that, "When articulation and sound wave go their separate ways.... perception always goes with articulation," (Liberman, 1957, p. 121) gave strength to long-held theories (e.g. de Cordemoy, 1668, cited by Cooper, 1974) that "speech is perceived by processes that are

also involved in its production" (Liberman et al, 1967, p. 452). The theory assumes that "at some level or levels of the production process there exist neural signals standing in one-to-one correspondence with various segments of the language - phoneme, word, phrase, etc.[1] Perception consists in somehow running the process backward, the neural signals corresponding to the various segments being found at their respective levels." (ibid, p. 454). Crucial to the theory is the concept of encoding. The consonant /d/ is thought to be acoustically encoded into /i/ in the spoken syllable /di/ and therefore must enter a specialized decoding mechanism in order to be decoded into its original abstract segments. The vowel /i/ on the other hand, is thought to be less encoded, being more invariant, and therefore has less need of the hypothetical specialized mechanism. "The level at which the encoding process is entered for the purposes of perceptual decoding may.... determine which (acoustic) shapes can and cannot be detected in raw perception" (ibid p. 454).

The motor theory's main postulate is that articulatory knowledge is used in perception to account for the lack of acoustic-phonetic invariance. Articulatory knowledge appears the only straightforward way to make sense of the "temporally scattered and contextually variable patterns of

---

[1] While it is clear that, in running speech, these higher levels substantially interact with phonemic processing (Miller, Heise & Lichten, 1951; Savin & Bever, 1970; Foss & Swinney, 1973; McNeill & Lindig, 1973; Marslen-Wilson, 1975), the experiments described within this dissertation are based on the assumptions that; (1) phonemic processing can be investigated in isolation by using stimuli with no higher-level linguistic relevance, (2) that it is necessary in the processing of higher segments, and (3) that such investigation is illuminating in its own right. Thus, the present discussion does not deal adequately with the perception of words, phrases, etc., and instead deals with speech as though the syllable were the highest level of analysis. The interaction of higher levels with phoneme processing is discussed further in Chapter V.

speech" (Studdert-Kennedy, 1974). However, it fails to specify how such articulatory knowledge is used. How, for example, are the neural signals corresponding to the various segments found in the neuromotor system? How is it decided whether to employ "raw perception" or to involve the specialized mechanism in determining the final percept? The motor theory, at first glance, appears to merely push the invariance-segmentation problem one step further out of the grasp of experimental analysis and back into the realm of the abstract.

Another more explicit theory of speech perception was evolving in parallel with the motor theory which also invoked articulatory knowledge as a solution to the invariance-segmentation problem. The analysis-by-synthesis theory originated as an attempt at designing a "phonetic type-writer," a machine that could accept speech input and transform it into a discrete series of phonetic symbols (Stevens, 1960). Thus, the theory was forced to deal concretely and specifically with the operations and mechanisms to be employed in the proposed automaton, albeit at the potential expense of veridical description of human perceptual processes.

Stevens' (1960) first attempt at modeling the flow of information during speech perception proposed two distinct processing loops - one to compute an articulatory description corresponding to the incoming acoustic speech signal, and another to match phonetic segments to the articulatory description. Auditory signals input to the first loop are stored in a buffer for later comparison and also are analyzed into basic auditory components (e.g. power spectrum, amplitude envelope). The auditory properties specified by this analysis are then used to generate hypotheses about the

articulatory behavior necessary and sufficient to produce such auditory properties. When an articulatory description has been found which produces auditory properties which deviate minimally from those of the input signal, the articulatory description is then passed to the second loop. This loop contains a device with operational knowledge of rules relating how phonetic segments correspond to articulation. It generates a sequence of phonetic segments to match the incoming articulatory description. The phonetic symbols generated by the device are transformed into an articulatory description which is compared to the incoming articulatory description, the output of the first loop. When phonetic symbols have been found which generate an articulatory description deviating minimally from the incoming articulatory description, the phonetic symbols are output from the system.

Later versions of the model (Stevens, 1972) abandoned the moment-to-moment computation of an articulatory description in favor of a fast, passive abstraction of certain acoustic properties which have been found to be relatively invariant.

These properties do not correspond to the phoneme, but are closer, instead, to the distinctive feature. Distinctive feature systems have long been used in the structural analysis of language (Jakobson, 1962), and were originally developed as a system of classifying the articulatory and acoustic dimensions of phonemic segments (Jakobson, Fant & Halle, 1952).

Such systems, while varying in particulars, all use phonetic classification matrices with several articulatory dimensions each of which can take on a limited number of values. An example of such a system (Blumstein, 1974) might have the dimensions consonant (which can take on the values

consonant or vowel), manner of articulation (with the values stop, nasal, fricative, affricate, liquid, glide), place of articulation (bilabial, alveolar, velar) and voicing (voiced, voiceless). In this system the phoneme /l/ would be classified as consonantal, liquid (manner), alveolar (place) and voiced. The phoneme /d/ would contrast with the phoneme /l/ only in the manner feature (stop rather than liquid).

Certain acoustic invariances have been associated with many of the distinctive features. For example, the presence of the characteristic resonance of the fixed nasal cavity always signals the presence of the nasal manner feature. The presence of wide bands of noise signals fricatives and the presence of silence preceding or following rapid format transitions signals stops (Jakobson, Fant & Halle, 1952).

In the newer version of the analysis by synthesis model (Stevens, 1972) these types of acoustic cues are extracted by a preliminary analysis process and used as bench marks in the forming and testing of hypotheses relating acoustics and phonetics through articulation. Specifically, the acoustic information of the signal is passed through the preliminary analysis process discussed above which in turn passes its output to a control process. The control process is equipped to generate hypotheses about the features and phonemes in the utterance. These hypotheses are then passed to a set of quasi-articulatory generative rules which transforms them into information which contains the relevant aspects of the acoustic forms of the hypothesized phonetic utterance. This information then is compared to the relevant aspects of the original auditory input and the discrepancies are again passed to the control process. The control process can then decide

to generate a new hypothesis if the discrepancies are large or output the phonemic string if they are acceptably small.

The new analysis-by-synthesis theory has considerable advantages over both the older model (Stevens, 1960) and the motor theory (Liberman et al, 1967). It is a specific and plausible account of a speech recognition device which solves the invariance-segmentation problem by the utilization of articulatory knowledge in perception.

Furthermore, the theory acknowledges certain acoustic invariances that exist on the distinctive feature level and uses them to simplify and speed the process by using them as fixed spectral and/or temporal reference points, allowing the abandonment of a moment-to-moment articulatory computation stage, and narrowing the possible alternative phonemic hypotheses. In other words the theory shows explicitly how the data of "raw perception" (Liberman et al, 1967, p. 454) might be used.

The role of attention in perceptual processing.

The theories of speech perception mediated by articulatory knowledge are intimately connected to other theories which postulate efference (Festinger, Ono, Burnham & Bamber, 1967), corollary discharge (Sperry, 1950) reafference, (von Holzt, 1954) or feedback (MacKay, 1965) as necessary in perceptual experience. The common element of all these postulated processes is that they can be characterized as active, or information-determined, rather than passive, or stimulus-determined, (MacKay, 1965).

Attention has been considered isomorphic with the comparison of incoming information with the current status of an active system (MacKay, 1965; Miller, Galanter & Pribram, 1960), but since a simple thermostatically

controlled home heating system is such an active information-determined system, it is obvious that such identity is not warranted. The possibility remains that such a comparison may be necessary for attentive processing, though clearly not sufficient.

Neisser, in his formative book Cognitive Psychology (1967), treats active processing in quite a different way. In his view human perception is "....assumed to have two stages, of which the first is fast, crude, wholistic and parallel and the second is deliberate, attentive, detailed and sequential." (ibid, p. 10). The first stage consists of passive or "preattentive" mechanisms while the second stage is an active process of construction, which Neisser asserted was "itself the mechanism of auditory attention" (ibid, p. 213).

Neisser's (1967) theory of speech perception is an analysis-by-synthesis theory not materially different from Stevens' (1972) theory. "....To 'follow' one conversation in preference to others is to synthesize a series of linguistic units which match it successfully. Irrelevant, unattended streams of speech.... fail to enjoy the benefits of analysis-by-synthesis. As a result they are only analyzed by the passive mechanisms which might be called 'preattentive processes' " (Neisser, 1967, p. 213).

Let us look at Stevens' (1972) model with respect to Neisser's (1967) assumptions about attention. The acoustic signal first is analyzed into its spectral components, etc., and passed through a preliminary analysis process which abstracts or recognizes properties which are associated with phonemic distinctive feature invariants. Neisser assumes such processing to be passive or non-attended. The output of the preliminary analysis process is passed to a control process which forms hypotheses about the phonemic structure of the signal and passes these hypotheses to other processes which

generate acoustic features, compare these to the "heard" signal and pass some measure of error or mismatch back to the control process. This operation would be the attention process, according to Neisser.

Thus, the Neisser-Stevens model implies that the more hypotheses that must be generated by the control process for a given phonemic segment, the more attention it would take. In other words, attention to a phoneme would vary as a direct function of its encodedness, or inversely with its relative acoustic invariance.

Other cognitive theorists have started their analysis of attention with the observation that man can only process a limited amount of information at one time (Broadbent, 1958, 1971; Keele, 1973; Kahneman, 1973). This observation, when coupled with the common subjective report that attention to one stimulus, thought or activity precluded or considerably diminishes the ability to attend to something else, led these psychologists to identify attention with man's limited capacity for processing information.[2] Thus, attention could be measured by the degree to which one task interfered with another (Kerr, 1973). By cleverly manipulating the structure of the main task and the interfering task, psychologists have

---

[2]The distinction between structural (e.g. Keele, 1973) and capacity (e.g. Kahneman, 1973) models of attention will be largely ignored here (cf. Kerr, 1973, for an excellent review). The structural theorists tend to speak of limited capacity mechanisms, while the capacity theorists deal in terms of the allocation of a limited capacity, but for the purposes of this discussion both terms are identical and will be used interchangeably. Both theories assume that the interference of one task upon another is a valid measure of attention.

been able to construct a rough taxonomy of mental operations which do or do not require attention.[3]

Encoding is one of the operations most extensively studied. The word encoding is used differently in the literatures of acoustic phonetics and cognitive psychology. Instead of referring to the acoustic representation of phonetic information, here it refers to reception of the proximal stimulus and the subsequent contacting or activation of its representation(s) in memory. In other words, it refers to the coding of the proximal physical stimulus into the internal mental system of the subject. This sense of the word will henceforth be denoted as "mental encoding".

The time course of mental encoding in a letter matching task was studied by Posner & Boies (1971). One letter was presented followed at some variable time by another letter. The subject (S) had to respond same

---

[3]Two caveats must be exercised when applying the dual task paradigm to the study of attention. First, one must design both tasks such that if interference occurs, it will be central and not "structural" interference (Kerr, 1973). For example, if the primary task were standing up and the secondary task were sitting down, interference between these two tasks could not be interpreted as central or due to capacity limitations. The use of different sensory and motor modalities or different stimulus and response characteristics have been used to avoid this problem, though no specific rules have been formulated (cf. Kerr, 1973, p. 405). Second, care must be exercised that the interpretation of the results is based on thorough analyses of both tasks. For example, if in condition A there is decrease in performance on both tasks and in condition B there is a greater decrease in performance on the secondary task accompanied by a lesser decrease on the primary task, an interpretation that the secondary task reflected greater attentional demands at condition B than at condition A would be suspect. The ideal situation in this paradigm is when performance decrements on one task occur in the face of unchanged performance on the other (Kerr, 1973). When this configuration is not present, the results may still be interpreted if requisite caution is applied (Kerr, 1976, personal communication).

or different. It was found that response time (RT) was at a minimum when the interstimulus interval (ISI) was about 500 msec, and this was taken to reflect the time necessary to mentally encode the first letter. This minimum was the same whether the task was to match two physically identical letters (e.g. AA), two nominally identical letters (Aa) or to decide whether two letters belonged to the class of consonants or vowels (AE), even though absolute RT varied systematically across tasks.

Then, in another experiment, the letter matching task with a fixed ISI of one second was employed with a secondary task of responding to a probe (the onset of white noise) presented unpredictably at one of several times in the trial. It was found that reaction time to the probe was actually fastest in the 500 msec following the onset of the first letter. This result has been confirmed repeatedly (Posner & Klein, 1973; Comstock, 1973) and is interpreted to indicate that mental encoding does not require processing capacity or attention. Furthermore, several other studies using different paradigms have reinforced this conclusion, suggesting that the mental encoding of a visual stimulus is an automatic process and is not confined by capacity limitations (e.g. Beller, 1970; Keele, 1972; Posner & Boies, 1971).

To the extent that the mental encoding of a speech stimulus involves contacting its <u>phonemic</u> representation in memory, this interpretation is at odds with that construed from Neisser (1967) and Stevens (1972) in that it would predict no attentional involvement in the mental encoding of phonetic information. On the other hand, since we know that speech has a very complicated relationship between proximal stimulus and percept, it could be that speech is encoded differently from visual stimuli. If the

Neisser-Stevens model were correct, automatic encoding would only occur for distinctive features associated with acoustic invariants.

One process in the Posner & Boies (1971) experiment was isolated as requiring attention. RT's to the probe noise began to increase significantly 500 msec after onset of the first letter. Since the encoding period had already ended and since there was no analogous increase during preparation for the first letter, the authors attributed the increased RT to generation or maintenance of the distinctive visual features of the first letter.

Another mental operation, response selection, was isolated as requiring attention by Noble, Trumbo & Fowler (1967; also cf. Trumbo & Noble, 1970). Their experiment used visuo-motor tracking as a primary task and one of several verbal secondary tasks. In the no response condition, subjects had to listen to a sequence of spoken numbers to learn them. In the anticipatory response condition subjects had to anticipate which number came next in the sequence. In the free response condition, subjects had to say numbers in any sequence they wanted, and in the same response condition subjects had to repeat the number they just heard. Error on the tracking task was analyzed on the basis of the type of secondary task. The same response and no response conditions were not different from the control condition, in which tracking task was performed alone. The free response and anticipatory response tasks, however, did cause significant interference with tracking. The results are interpreted as indicating that the response selection stage was the locus of the interference. If responding were that locus, then the same response condition would also

have shown interference. However, only the two conditions with a selection requirement had an effect.

The Noble, Trumbo & Fowler (1967) experiment has a paradoxical message for this discussion. Stevens (1972) postulates selection of a phonemic response as an explicit state of his model. Furthermore, much selection, generation and testing of distinctive articulatory and auditory features precedes prior to response selection. Noble et al (1967) and Posner & Boies (1971), as well as the earlier theorists (e.g. Miller, Galanter & Pribram, 1960), would say that these stages require attention. Yet in the Noble et al (1967) study, perception of the speech stimulus in the no response condition did not cause interference. Though we must exercise caution in accepting the null hypothesis, it is important to note that the attention hypothesis was not confirmed in a situation which was potentially an appropriate test of it.

More evidence on the role of attention in speech processing comes from the shadowing paradigm, introduced by Cherry in 1953. Cherry (1953) found that when he asked his subjects to repeat, or shadow, one channel of a stereophonic tape with unrelated spoken prose messages on each channel, that subjects could not report the verbal content of the ear they were not shadowing, or even that the language of the non-shadowed ear changed from English to French or consisted of reversed speech. On the other hand, subjects were able to correctly report that a man's or woman's voice was used or that the non-shadowed "message" was a pure tone.

Treisman (1964) attempted to quantify Cherry's finding by examining shadowing performance on the attended channel as a function of the

composition of the stimulus material in the non-attended channel. She found that when a woman's voice spoke the shadowed message, the least interference was produced when a man's voice read the non-shadowed message. Much smaller, but nevertheless significant interference with shadowing was found when the shadowed and non-shadowed channels contained semantically similar messages and when the subject was familiar with foreign languages presented in the rejected channel.

Cherry's (1953) and Treisman's (1964) results suggest that only the gross acoustic features of an unattended message are analyzed. However, several experiments have amplified Treisman's finding of small but significant effects of the semantic content of the non-attended ear (Lewis, 1970; MacKay, 1973; Corteen & Wood, 1972; but cf. Wardlaw & Kroll, 1976 for a failure to replicate) suggesting that at least some processing of the non-attended speech signal must be occurring. These findings have been interpreted as consistent with the notion that encoding does not require attention (Keele, 1973; Lewis, 1970; Posner and Snyder, 1975) whether the stimulus is visual or spoken.

The problem of the role of attention in phonemic processing.

This brief survey has illustrated how the data of acoustic phonetics have led to active models of speech perception. Several of the processes in these active models, either implied or explicitly stated, have been shown by cognitive psychologists to require attention. At the same time, these psychologists have effectively demonstrated that mental encoding of a visual or to some extent a speech stimulus does not require attention, but rather proceeds automatically and in parallel with other processes. Thus,

acoustic phonetics predicts that the mental encoding of speech requires attention while cognitive psychology predicts that such encoding is automatic. The aim of the following chapters is to combine the approach of the acoustic phonetician with that of the cognitive psychologist to determine the role of attention in the mental encoding of spoken phonemic segments.

II. THE ROLE OF SHORT-TERM MEMORY IN SPEECH PERCEPTION

Articulation and perception find a common theoretical ground in the notion of distinctive features (Jakobson, Fant & Halle, 1963). While features have long had obvious descriptive value as target loci in articulation or as relatively invariant acoustic and articulatory patterns (Stevens, 1972), evidence of their perceptual "reality" has been much slower in being generally accepted. Studies of subjective scaling of similarity of phonemes (Greenberg & Jenkins, 1964), perceptual confusions of phonemes filtered through various pass-bands and in several levels of noise (Miller & Nicely, 1955; Shepard, 1972) or across speakers and languages (Singh, 1966) and perceptual transformation of a repeated syllable (Goldstein & Lackner, 1974) can all now be parsimoniously explained within a distinctive feature model (Studdert-Kennedy, 1974).

The finding of intrusion errors in recall from short-term memory (STM) which differed systematically from target items only in one or two distinctive features (Wickelgren, 1966; Conrad, 1964) suggested that features serve as a code common to several phases of the speech perception process. In other words, perception, articulation and memory were thought to be linked to the same system of internal representation - distinctive features.

The fact that the distinctive feature effect in recall from STM obtains for visually presented verbal stimuli (Conrad, 1964) as well as auditorily presented speech (Wickelgren, 1966) suggests that the locus of these effects is a general purpose central short-term store rather than an auditory sensory memory which is functionally equivalent to the proximal acoustic stimulus (Massaro, 1972).

Two findings from dichotic listening research motivated even stronger claims about how distinctive features in STM function in speech perception. The findings were: 1) The probability of correct identification of a dichotically presented pair of consonant-vowel (CV) sounds increases with the number of distinctive features shared by members of the pair, even when acoustic similarity is greatly reduced by using different vowels in the two members of the pair (Studdert-Kennedy, Shankweiler & Pisoni, 1972). 2) Blend errors are found in conjunction with this feature sharing effect in which all feature values in a dichotic pair are preserved but "local sign", or information concerning which features go with which, is lost (Studdert-Kennedy & Shankweiler, 1970). Thus, a blend error for the dichotic pair /ta/ - /ba/ would be the response /pa/ - /da/.

The latter finding suggests a common locus for feature processing or storage because blend errors would only occur by chance if the feature composition of each input were processed and stored separately. Several authors have assumed - explicitly (Blumstein, 1974; Blumstein & Cooper, 1972; Oscar-Berman, Zurif & Blumstein, 1975; Sawusch & Pisoni, 1974; Pisoni & Tash, 1974) or implicitly (Fodor, Bever & Garrett, 1974) - that this common locus of storage is STM. Thus, the feature sharing advantage has been considered to arise as follows: Inputs contrasting on two features require not only the storage of more features, but also the storage of their local signs. When the inputs contrast on only one feature, not only are there fewer feature values to store, but, at least in the present paradigm, local sign is no longer necessary since the two values of the contrasting feature both go with the same value of the matching feature (cf. Blumstein

& Cooper, 1972, p 212). Thus, in the latter case, capacity requirements may be reduced (cf. Pisoni, 1975, p. 97).

This hypothesis is closely related to the Neisser-Stevens model in two respects. First, STM is clearly limited in capacity (Miller, 1956). Second, information in STM is actively maintained by a rehearsal process, which itself places demands upon the limited capacity system (e.g. Shulman & Greenberg, 1971). It is unclear whether or not such a hypothesis considers STM to be isomorphic with the control process of Stevens' (1972) model or whether it only serves as an adjunct storage device to the control process. Some recent views consider STM to be the output of attentive processing (Bjork, 1975), while other views consider that selection, recoding and rehearsal of certain stimulus traces is the attention process itself (Shiffrin, 1975). Despite which view one takes, it can be clearly seen that the concepts of attention and STM are closely related. Thus, the claim that features are processed or stored in STM is construed in the present discussion to mean that features are potentially available for rehearsal and other types of attended processing, as well as that they demand capacity.

Blumstein & Cooper (1972) have also demonstrated that when same-different discrimination judgments, rather than identification responses, are required for dichotically presented CV sounds, feature similarity makes the task more, rather than less difficult. Accordingly, they suggest "...in the identification task the subject must analyze the auditory information into its linguistic components and hold them in short-term memory long enough to encode his response" (ibid, p. 212) while in the same-different task, "...storage of information is not a factor. Instead the subject need

only judge the relative similarity between competing stimuli" (ibid, p. 212). In other words, by the Blumstein & Cooper (1972) hypothesis features need enter STM only when organization of a phonemic response is required.

The two dichotic tasks provide a potentially powerful analytic tool to dissociate those stages which require capacity from those which do not. For example, if features are stored or processed in STM during perception, as opposed to response organization, one would expect that regardless of the primary task, the fewer features shared by the stimulus pair (i.e., the more separate feature values present), the greater the interference would be with a memory task. If, on the other hand, feature storage or processing in STM occurs when organization of a phonemic response is required as Blumstein and Cooper suggest, then one might expect to see such interference when the primary task was identification but not when a same-different judgement was required. The first experiment was designed to test these alternatives. A dual task paradigm (Kerr, 1973) was used to assess demands of feature processing upon capacity in STM. The primary tasks were the two dichotic listening tasks - two-ear identification and same-different discrimination. Since STM is defined in common sense terms as that memory in which one holds an unfamiliar telephone number from the time it is looked up until the time it is dialed, the secondary task was remembering a string of seven random digits.

EXPERIMENT I

Method

Subjects

Twelve subjects between the ages of 18 and 30 were recruited from the Woods Hole periscientific community, and were paid two dollars an hour for their time. All reported that they were right handed, native English speakers, and had no known hearing or neurological deficits. No subject had had any previous experience with dichotic listening. No subject's data were discarded for any reason.

Stimuli

The stimuli for the dichotic task were 80 pairs of different CVs where the consonants were drawn from the set /ptkbdg/ and all were followed by the vowel /a/. The stimuli were natural speech spoken by a trained female phonetician, synchronized for onset, matched for intensity, and 350-375 msec in duration. The dichotic stimuli were identical to those used by Blumstein & Cooper (1972), and were obtained through the courtesy of the senior author.

Of the 80 pairs, 32 contrasted on place, 16 contrasted on voicing, and 32 contrasted on both voice and place. The stimuli for the memory task were 80 nonidentical permutations of 7 single syllable numbers drawn from the set (1, 2, 3, 4, 5, 6, 8, 10) and randomly assigned to each CV pair. These strings were recorded on tape in a male voice at one digit per second such that the last digit ended 2 sec before the onset of the dichotic stimulus.

Procedure

Subjects were tested individually in a quiet room with a Tandberg 1200X tape recorder. The output of Koss Pro 4AA headphones were matched by means of a 1000 Hz calibration tone measured at 80 dB on a General Radio sound meter (Type 1565Z).

Each subject performed two different dichotic listening tasks: a same-different discrimination task and an identification task.

In the same-different discrimination condition subjects heard the memory set, heard the dichotic pair, reported on the dichotic pair by saying "same" or "different" and simultaneously pointing to S or D on a card in front of them, and then recalled the memory set. The identification condition was precisely the same except that when reporting the dichotic stimulus, subjects said two CV syllables and simultaneously pointed to two appropriate letters on a card with P, T, K, B, D, and G printed on it. The simultaneous pointing was especially important here to eliminate perceptual errors by the experimenter. For the same-different task subjects were instructed to report "same" when they heard the same initial phoneme in each member of the dichotic pair, and otherwise to report "different". Examples of what constituted a "same" pair and a "different" pair were given as follows: "when you hear /ba/ in this ear and /ba/ in that ear, you should respond 'same'." In the identification task subjects were told that in the present session all the pairs of CVs contained different initial phonemes and their job was to report which two phonemes they heard. The experimenter recorded responses to both tasks on paper tape with an ASR-33 teletype. Each S received the same-different task and the identification task, in that order and on two separate days. The same-different task

was always presented first to eliminate the possibility that subjects would discover that all pairs were different.

Within each day subjects were given 80 trials and then the headphones were reversed to balance any unknown asymmetries between channels and another 80 trials were presented. Which channel went to which ear first was balanced between subjects. Also, the tape was divided into two halves, each containing equal numbers of pairs sharing place, voice, and neither place nor voice, and which half was heard first was also balanced between subjects.

On the first day subjects were given at least 20 practice trials. They were given practice trials until they gave at least five "same" and five "different" responses. On the second day subjects were given at least 10 practice trials and were practiced until the experimenter felt that the subject understood the new task.

Results

Dichotic Tasks

The data from the two dichotic tasks are summarized in Table 2.1. Within subjects analyses of variance were performed on the percentage of trials correctly reported by each subject in each condition.

The first analysis of the identification task was a one-way analysis of the effect of feature sharing on those trials where both CVs were reported correctly. There was a significant feature effect ($F(2,22) = 5.77$, $p < .01$) reflecting the usual identification advantage accruing to pairs distinguished by only a single feature. That is, the opposing inputs which

Table 2.1

Percentage correct in the two dichotic tasks of Experiment I as a function of feature contrasts.

	Place Contrast	Voicing Contrast	Double Contrast
Identification Task (both ears correct)	48.44	46.61	31.25
Same-Different Task (different responses)	32.29	53.65	67.45

contrasted on either place or voice were reported correctly more often than pairs that contrasted on both features. This was significant in a Newman-Keuls test of multiple comparisons ($t = 3.38$, $k = 2$, $p < .05$). Voicing contrasts were not significantly different from place contrasts ($t < 1.0$). Another analysis was performed on those trials where at least one ear was reported correctly. The usual right ear advantage for speech emerged ($(R-L)/(R+L) = .074$: $F(1,11) = 12.60$, $p < .005$) as well as a feature effect similar to that for both ears correct ($F(2,22) = 16.85$, $p < .001$).

Analysis of the same-different task showed that the feature effect was significant ($F(2,22) = 15.87$, $p < .001$), and in line with that found by Blumstein & Cooper (1972) where a two-feature contrast elicited better performance (i.e. more "different" responses) than single contrasts ($t = 4.50$, $k = 2$, $p < .005$). However, here voicing contrasts were significantly different from place contrasts ($t = 3.40$, $k = 2$, $p < .05$) but not significantly different from double contrasts ($t = 2.20$, $k = 2$, n.s.). This is in line with Blumstein and Cooper's (1974) findings and will be discussed further in the Appendix.

Every subject made a considerable number of "same" or error responses. As many as 126 and no fewer than 43 "same" responses were emitted during the 160 trials. Since there was a reciprocal relationship between "same" and "different" responses, further analysis here would have yielded no new information.

Thus, the dichotic performance shows the two opposed feature effects clearly and the usual right ear advantage for speech. These effects, while not surprising, must necessarily obtain to be able to interpret the results of the secondary task.

Memory Task

The memory data were scored strictly by position. For example, if a stimulus "1234568" evoked a response "2345698", only the last digit would be correct. In those rare instances where a subject reported more or less than 7 digits, the response string was truncated at 7 or the subject was asked to guess until 7 digits had been reported. The memory data were sorted by the feature relationships of the co-occurring dichotic pair, the correctness of the pair, the ear of correct report, and by serial position, where appropriate. Within subjects analyses of variance were performed on percentage correct per subject per condition. The results are shown in Figures 2.1, 2.2 and 2.3.

The first analysis of the identification task was performed on those trials in which both members of the dichotic pair were reported correctly. The results are shown in Figure 2.1. Serial position was significant ($F(6,66) = 35.89$, $p < .001$) reflecting the systematically bowed shape of the classical serial position curve. Contrary to expectation, the effect of the feature relationships of the dichotic pair did not approach significance ($F < 1.0$), although the feature by position interaction did ($F(12,132) = 1.80$, $p < .055$). In an attempt to further examine this interaction, one-way analyses were performed at each position for the effect of feature. These revealed no significant effects of feature at any position.

A second analysis of variance was done on those trials where at least one of the CVs was correctly reported. Again the effect of position was significant ($F(6,66) = 39.30$, $p < .001$). There was no effect of ear ($F < 1.0$) or feature ($F < 1.0$), and no higher order interactions.

Figure 2.1

Mean percentage correct digit recall contingent upon serial position and feature contrast of the co-occurring dichotic pair in the identification condition of Experiment I for those trials where both CVs were reported correctly.

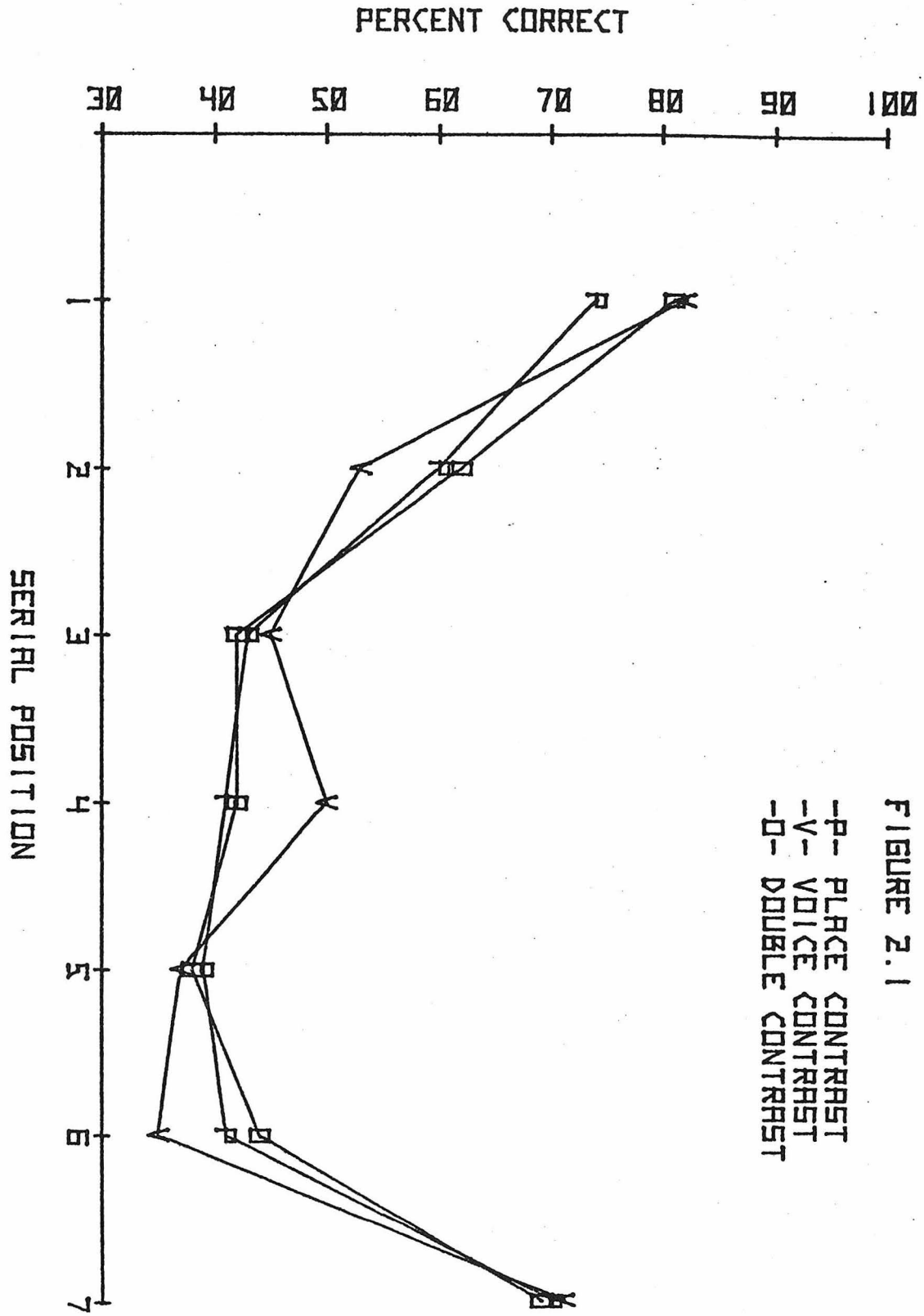


FIGURE 2.1

-P- PLACE CONTRAST
-V- VOICE CONTRAST
-D- DOUBLE CONTRAST

The first analysis of variance on the same-different task was a response (confounded perfectly with correctness) by feature by position analysis. The results of this task are shown in Figures 2.2 and 2.3. Main effects obtained for serial position ($F(6,66) = 41.60$, $p < .001$) and feature ($F(2,22) = 4.23$, $p < .028$), but not for response ($F < 1.0$). The feature by position interaction was significant ($F(12,132) = 2.31$, $p < .011$), as was the response by feature by position interaction ($F(12,132) = 1.95$, $p < .034$). Separate analyses for each level of response yielded a significant main effect of feature for "different" or correct responses ($F(2,22) = 5.45$, $p < .012$) but not for "same" or incorrect responses ($F(2,22) = 1.56$, $p < .23$). For both types of responses the feature by position interaction was significant ("same"; $F(12,132) = 2.40$, $p < .008$; "different"; $F(12,132) = 1.88$, $p < .043$). Analyses for the effect of feature at each position for each response type were performed and are summarized in Table 2.2. It can be seen that "same" responses yield significant memory differences at positions 4 and 6 where voice contrast trials are better than place contrast trials which are better than trials which contrast on both features. In comparison, different responses consistently approach significance at positions 3 through 7 where place contrast trials interfere with memory more severely than trials that contrast on voice or both features.

A separate analysis of variance performed upon mean memory scores on correct dichotic trials revealed no significant difference in interference with STM as a function of type of dichotic task ($F < 1.0$).

Figure 2.2

Mean percentage correct digit recall contingent upon serial position and feature contrast of the co-occurring dichotic pair in the discrimination condition of Experiment I when the dichotic pair was erroneously judged "same."

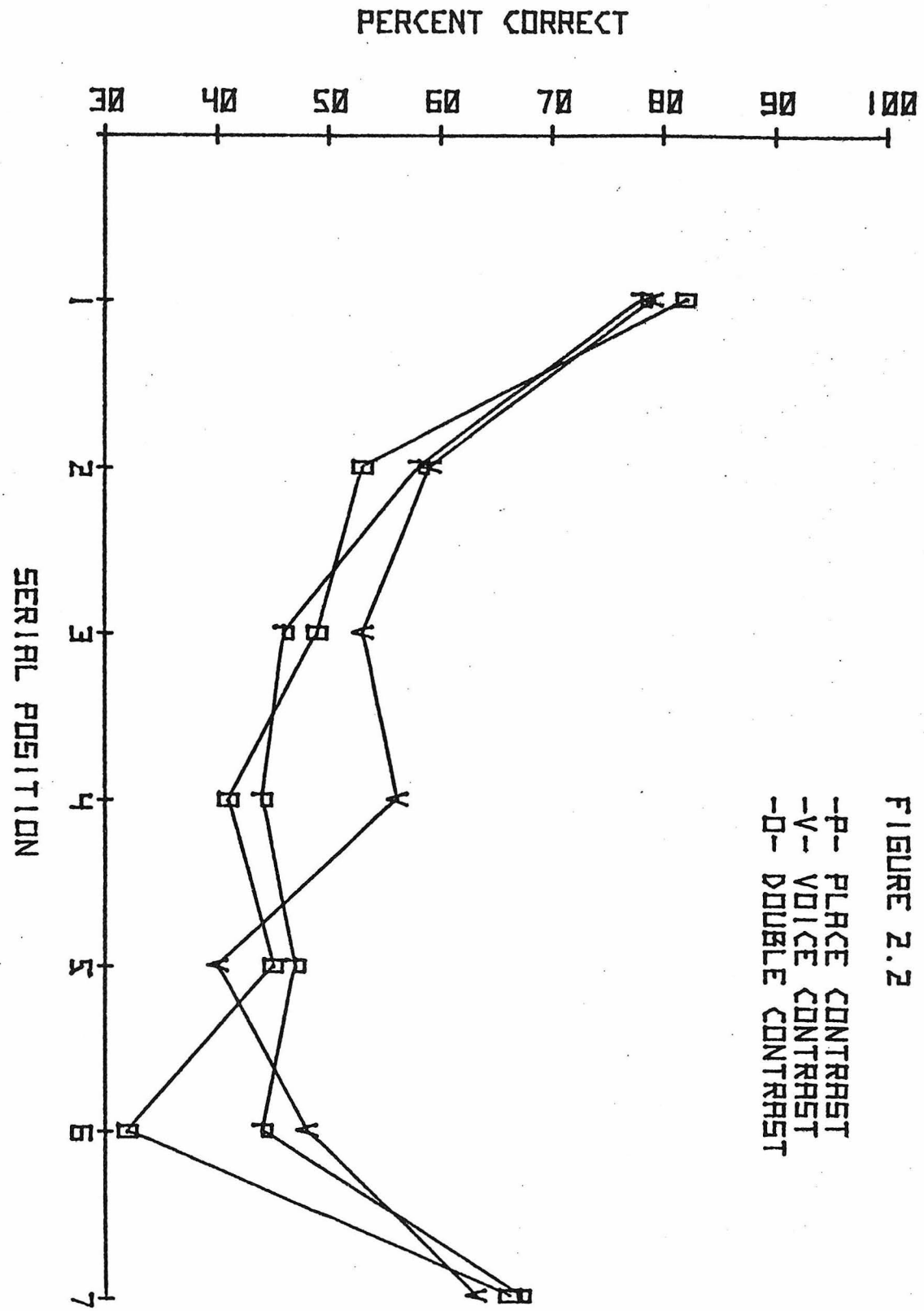

PERCENT CORRECT
30
40
50
60
70
80
90
100
SERIAL POSITION
1
2
3
4
5
6
7
-P- PLACE CONTRAST
-V- VOICE CONTRAST
-D- DOUBLE CONTRAST

FIGURE 2.2

Figure 2.3

Mean percentage correct digit recall contingent upon serial position and feature contrast of the co-occurring dichotic pair in the discrimination condition of Experiment I when the dichotic pair was correctly judged "different."

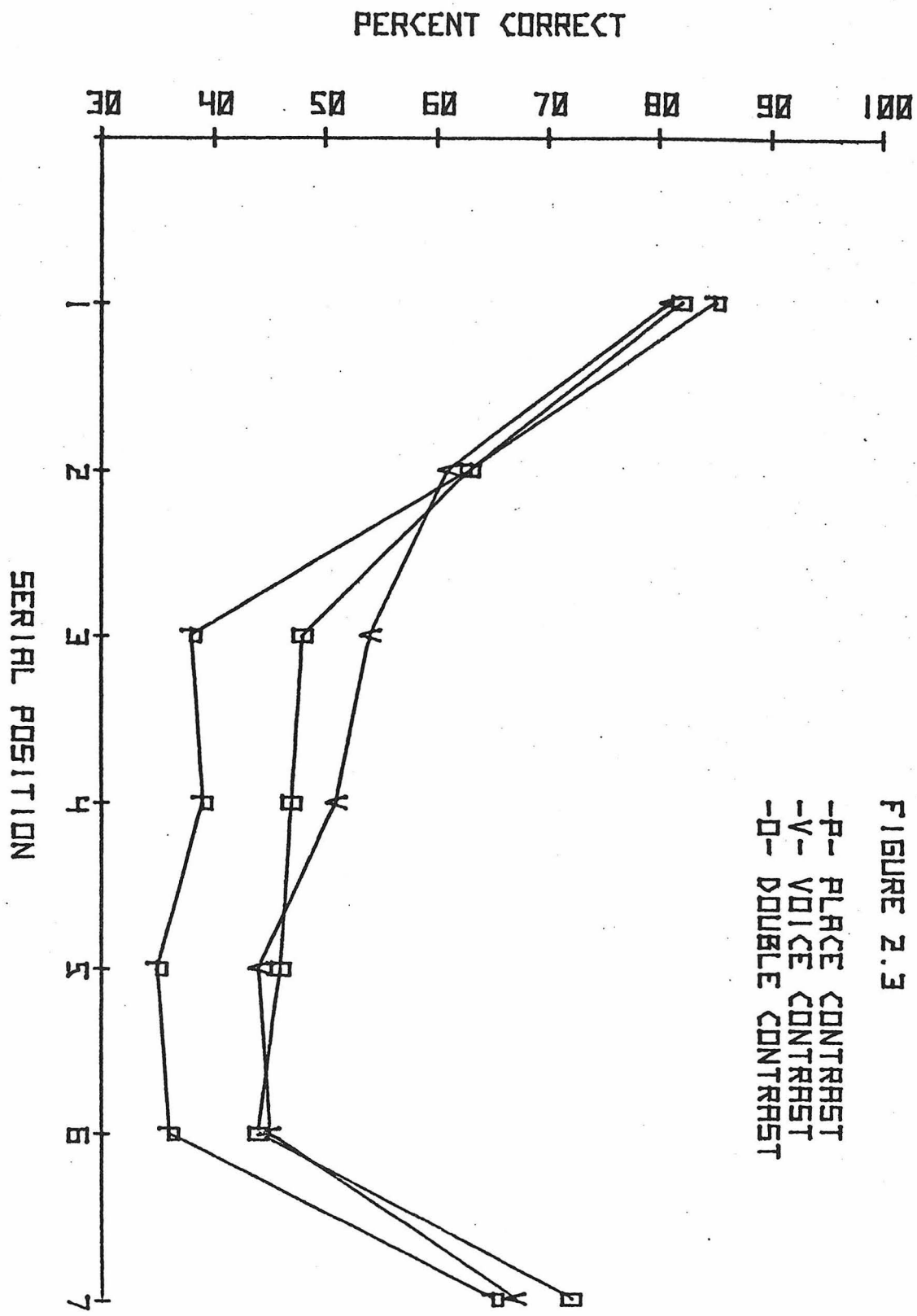

PERCENT CORRECT
30
40
50
60
70
80
90
100
SERIAL POSITION
1
2
3
4
5
6
7
-P- PLACE CONTRAST
-V- VOICE CONTRAST
-D- DOUBLE CONTRAST

FIGURE 2.3

Table 2.2

Summary of one-way analyses of variance for the effect of feature at each position for each response type in Experiment I (df=2, 22 in all cases).

		Serial Position						
		1	2	3	4	5	6	7
Same	F	0.45	0.75	1.43	4.13	1.09	7.09	0.38
	p <	.500	.484	.262	.030	.353	.005	.500
Different	F	0.41	0.14	5.27	3.12	2.81	3.04	3.06
	p <	.500	.500	.014	.065	.082	.069	.068

Discussion

An effect of feature contrasts on recall from STM was obtained only in the discrimination condition. Given that certain perceptual operations must occur both in the identification task and in the discrimination task, and that the expected effect of feature contrasts on memory did not obtain in the identification condition, it is unlikely that the effect of feature contrasts on memory in the discrimination condition is generally related to the perception of speech. Rather, these findings tend to reflect the greater difficulty of discriminating place feature contrasts and will be further discussed in the Appendix. Furthermore, the absence of an effect of feature contrasts on recall from STM in the two-ear identification task does not support the notion that the organization of phonemic responses requires capacity in STM as a function of the number or type of feature values present in the stimuli. If 1) the failure to find an effect of the two-ear identification task on STM scores were taken to indicate a genuine absence of effect, and 2) the likelihood of artifacts from interference of the STM task on dichotic listening were ruled out (cf. Footnote 3), then it would be possible to reject hypotheses in which STM was a necessary mediator of distinctive features in speech perception.

EXPERIMENT II

Another experiment was run to replicate the identification condition of Experiment I in order to rule out the possibility that the absence of an effect was due to the interference of the secondary task with the primary task.

Method

Experiment II was identical to Experiment I with the following exceptions:

1) Twelve different subjects were drawn from the same pool. 2) The memory stimuli were read by the experimenter rather than recorded on tape to attempt to randomize possible irregularities in stimulus presentation. 3) Two-ear identification was the only task associated with the dichotic stimuli. 4) Three conditions of memory load were employed; one, four or seven digits. Strings of one and four digits of first one and four digits of the seven digit strings used in Experiment I. Each subject received all three memory load conditions, one on each of three separate days. The order of these conditions was balanced between subjects.

Results

Dichotic Task

The first analysis of variance on the dichotic listening task was a feature by load analysis performed on those trials where both CVs of the dichotic pair were correctly reported.

The expected feature effect for a two-ear identification task emerged ($F(2,20) = 19.99$, $p < .001$). Those pairs contrasting on voicing alone were not reported with significantly different accuracy than those contrasting on place alone (Newman-Keuls $t = 1.38$, $k = 2$, n.s.) while both voicing contrasts and place contrasts were different from double contrasts (for voicing contrasts $t = 6.28$, $k = 3$, $p < 01$; for place contrasts $t = 4.90$, $k = 2$, $p < .01$).

Table 2.3

Percentage of dichotic pairs reported correctly by memory load and feature contrast in Experiment II.

Memory Load	Feature Contrast		
	Place Contrast	Voice Contrast	Double Contrast
1 Digit	55.47	63.06	38.78
4 Digits	57.95	60.80	36.94
7 Digits	57.25	62.78	38.50

There was no effect of memory load on dichotic performance ($F < 1.0$), nor did there seem to be any systematic trend. The mean for a load of 7 digits was highest (52.46%), while the mean for four digits was lowest (51.89%). In fact, all means were less than .25 standard errors of the mean different from each other. Furthermore, there was no interaction of load with features ($F < 1.0$).

A second analysis of performance on the dichotic task was performed on those trials where at least one CV was correctly reported. There was a strong effect of ear ($F(2,20) = 32.49$, $p < .001$) reflecting the usual right ear advantage for speech ((R-L)/(R+L) = .104). The feature effect was also present ($F(2,20) = 34.16$, $p < .001$) and conformed to that found in the previous analysis. The effect of load was again absent ($F < 1.0$), and as in the previous analysis no trend was evident. There were no higher order interactions which approached significance.

Memory Task

The first analysis of variance on the memory task was performed on the mean percentage of digits reported correctly as a function of load and the feature relations of the co-occurring dichotic pair when both members of that pair were reported correctly. Digit recall performance as a function of memory load, feature relations and serial position is illustrated in Figure 2.4.

The expected effect of load was present ($F(2,22) = 27.61$, $p < .001$) indicating that a 1 digit string was easier to remember than a 4 digit string which was easier, in turn, than a 7 digit string. There was no evidence of a feature effect ($F < 1.0$), and the crucial load by feature interaction also fell somewhat short of significance ($F(4,44) = 2.06$, $p < .103$).

Figure 2.4

Mean percentage correct digit recall in Experiment II contingent upon memory load, serial position and feature contrast of the co-occurring dichotic pair for those trials where both CVs were reported correctly.

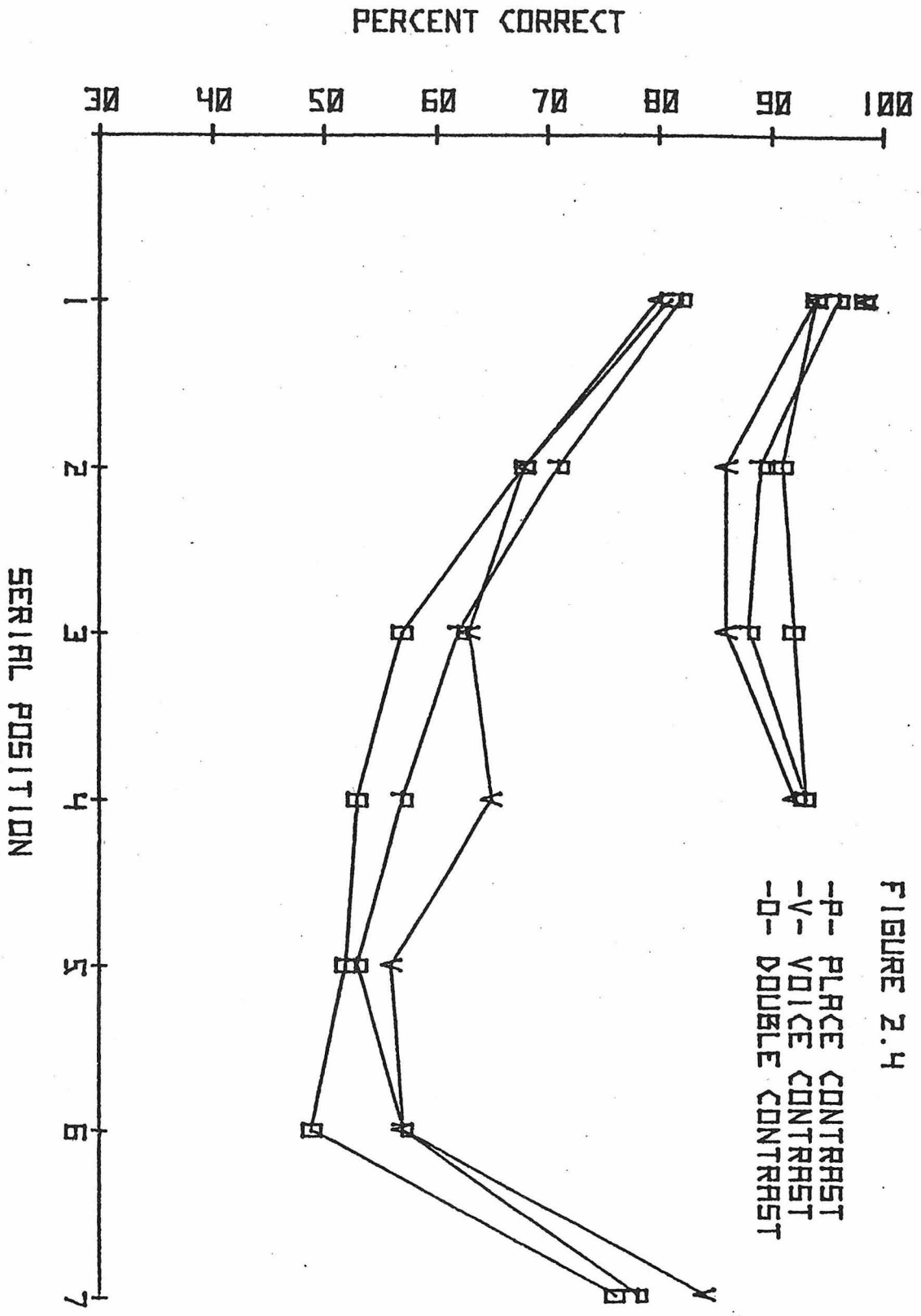

PERCENT CORRECT
30
40
50
60
70
80
90
100
SERIAL POSITION
1
2
3
4
5
6
7
-P- PLACE CONTRAST
-V- VOICE CONTRAST
-D- DOUBLE CONTRAST

FIGURE 2.4

A second analysis was performed on the mean percentage of digits reported correctly as a function of memory load, feature relations of the dichotic pair and ear correct where at least one member of the pair was reported correctly. The effect of ear was not significant ($F(1,11) = 1.64$, $p < .227$) nor were any interactions with ear. Again load was significant ($F(2,22) = 30.87$, $p < .001$), but feature was not ($F < 1.0$), and the load by feature interaction again fell somewhat short of significance ($F(4,44) = 2.10$, $p < .097$).

In order to examine the interaction of load with feature, the critical interaction of this experiment, in more detail, separate feature by serial position analyses were performed for memory loads of 4 digits and 7 digits. However, the effect of feature was not significant for load 4 ($F(2,22) = 1.539$, $p < .237$) or load 7 ($F(2,22) = 1.99$, $p < .161$), nor was the interaction of feature with position in either case. The effect of position was, of course, significant for both load 4 ($F(3,33) = 4.35$, $p < .011$) and load 7 ($F(6,66) = 17.08$, $p < .001$) reflecting the classical bowed shape of the serial position curve.

Discussion

Experiment II provides confirmation of the negative results of Experiment I by replication. Memory for digits as a secondary task does not appear to be differentially sensitive to the feature composition of the dichotic pair in a two-ear identification task. The lack of significant effects of memory load level on dichotic performance in Experiment II tends to rule out the possibility that this negative result might be related in some manner

to interference of the memory task upon the dichotic listening task. Thus, effects of the possible interference between the two tasks were probably not hidden within the primary task in either experiment.

Arguments that the dichotic task did not interact in any way with the digit memory task were countered by presenting ten 7-digit strings from the memory task of Experiment I to five subjects for written recall after a 10 second unfilled retention interval. Mean recall was 90.3%. This figure - a rough estimate of digit recall in the absence of interference from dichotic tasks - was compared with mean recall for digit strings which co-occurred with correct dichotic performance in Experiment I ($\bar{x}$ = 54.6) and the 7-digit condition of Experiment II ($\bar{x}$ = 65.2). These tests indicated that the dichotic task itself caused significant interference (Experiment I, $t = 10.06$, $p < .001$: Experiment II, $t = 3.01$, $p < .008$). Taken with the absence of differential effects on the basis of number or type of feature contrast, these comparisons indicate that other components of the dichotic listening task, exclusive of feature processing, interfere with STM.

We still have no way of knowing whether the failure to find interference of perceptual feature processing requirements of the task within the limited capacity mechanism is due to a genuine lack of effect, or merely to a measure insensitive to feature processing requirements. In common sense terms, for example, it could be that the "chunks" of short-term digit memory are simply too large to be affected by mere sub-phonemic feature values. Another plausible alternative is that features are processed through STM too rapidly to create substantial interference.

On the other hand, it could be that features are perceptually processed through some subset of STM that does not interact with digit memory, such as the feature buffer proposed by Pisoni (1975). This is not to say that distinctive features are not accessed by STM. It is quite likely that STM makes use of features as one of many convenient available codes (cf. Posner, 1969; Paivio, 1969). The finding of feature substitution errors (Wickelgren, 1966; Conrad, 1964) certainly supports this view without necessarily implicating STM as an obligatory stage or process in the perception of speech.

III. EXPERIMENT III-A PROBE REACTION TIME STUDY OF ATTENTION TO DICHOTIC SPEECH

Experiment III was designed to explore the possibilities that processing of distinctive features did not interfere with STM in Experiments I and II because 1) digit-size chunks are insensitive to interference from distinctive feature-size chunks, or 2) that features are processed through STM too rapidly to cause measurable forgetting. In other words, the experiment was designed to explore the possibility that the memory-interference paradigm was insensitive to the postulated capacity demands of distinctive feature processing.

The probe RT paradigm was employed to measure the capacity demands of speech processing because of its sensitivity to small changes in processing capacity and its ability to measure those moment-to-moment changes as a function of the microstructure of the task (Posner & Boies, 1971; Posner & Klein, 1973). The probe RT paradigm involves a simple secondary task arranged so that, at least superficially, it does not conflict with the primary task, either in stimulus or response modality. Thus any interference measured is assumed not to be of peripheral or structural origin. The utilization of discrete stimuli which occur at unpredictable times relative to the primary task provides a measure of momentary capacity demands of the primary task. The use of reaction time as a dependent variable for the secondary task provides a continuous scale potentially sensitive to small and transient changes in available capacity.

Previous probe RT experiments have investigated visual matching tasks (Posner & Boies, 1971; Posner & Klein, 1973; Comstock, 1973), or

kinesthetic (Klein & Posner, 1974) and movement (Ells, 1969) tasks, and have all used auditory stimuli for the probe RT task. The present study was designed to investigate the attention demands of an auditorily based task, therefore a visual stimulus was used in the probe task. Since vision tends to dominate other modalities in perception and is thought to be less alerting and therefore more likely to monopolize voluntary attentive mechanisms than audition or kinesthesis (Posner, Nissen & Klein, 1976), care was exercised in the design and execution of the experiment to emphasize the fact that the speech processing task was primary.

Method

Stimuli

The speech stimuli were the CV syllables /ba, da, pa, ta/, spoken by a trained male phonetician and adjusted to equivalent durations (300 msec) and intensities by computer at Haskins Laboratories. Sixteen dichotic pairs of these stimuli, matched for onset and offset, were prepared on the Haskins computer. Each pair was preceded by 100 msec. of 1000 Hz sine wave and 400 msec. of silence. Four pairs were identical (/ta,ta/, /da,da/ , /pa,pa/ , /ba,ba/), four contrasted on place (/pa,ta/, /ta,pa/, /da,ba/, /ba,da/), four contrasted on voicing (/ba,pa/, /pa,ba/, /da,ta/, /ta,da/), and four contrasted on both features (/ta,ba/, /ba,ta/, pa,da/, /da,pa/). Sixteen non-identical permutations of these sixteen pairs were generated and recorded on tape at equal volume in each channel, with an inter-trial interval of 8 sec.

The visual stimuli were three green light emitting diodes, 5 mm in diameter and 20 mm apart mounted in a black panel. The center light

was directly in front of the subject and the other two lights were 2.5° of visual angle to the left and right of the center light. The rise and fall times of these lights were on the order of 10 nsec, and the lights were clearly visible and approximately equal intensity when lit.

Apparatus

Tapes were played on a Tandberg 1200X tape recorder. The tape recorder was discovered to be approximately 10% slow, and appropriate adjustments were made in the timing of the probe light onsets (See Figure 3.1). The outputs of the tape recorder, amplified by a Pioneer SA-500-A amplifier went to a set of Koss Pro/4AA headphones worn by the subject and to a Lafayette voice-activated relay (model 18010).

The voice operated relay started a Gerbrands digital millisecond timer and created an "on" state on the left-most bit of a standard Hewlett-Packard 11202A TTL I/O interface card. The interface was read by a Hewlett-Packard 9830 calculator which, when the left-most bit went "on", initiated a wait function. The wait function was specified in milliseconds but was controlled not by a real-time clock but by the cycle time of the calculator. Therefore the relationship between the specified time (s) and the actual time (a) was determined using a storage oscilloscope and found to be exceedingly replicable and linear through the range of times used in this experiment ($r = .999$). This relationship is described by the equation: $s = 1.17a - 27.27$.

When the wait function expired, the calculator raised one of the last three bits on the interface which turned on one of the three light-emitting

diodes discussed above. The subject, whose head was in a home-built plexiglass chinrest, registered his response to the light by pressing a key. The key, when pressed, activated a microswitch which stopped the timer. Data were entered into the machine from the calculator keyboard by the experimenter and stored on a Hewlett-Packard 9880 Mass Memory (disk) unit for future analysis.

Subjects

Eight subjects between the ages of 17 and 35 were used in this experiment. All reported that they were right-handed native English speakers with no hearing or neurological deficits and no uncorrected visual deficits. In addition to these eight, three subjects were excused from the experiment due to a prolonged equipment breakdown, one was excused due to scheduling problems, one subject failed to report for the third and fourth days of the experiment, and another subject was excused when a dramatic and previously undetected right ear hearing loss was found in the course of the experiment.

Procedure

Subjects were run for four one hour sessions on four separate days. At the start of day one, written instructions were presented as follows:

> This experiment is complicated, so read these instructions carefully. If you have questions, ask them.
>
> Your main task will be to listen to speech sounds (ba, da, pa or ta) arriving at one ear of your headset. I will tell you which ear to listen to. Report the sound you hear as fast as you can without compromising on accuracy. Ignore any sounds in your other ear. For every correctly identified speech sound you will get 0.4 cents. This adds up to $4.10 if you are correct on every trial.

> Also, during the course of a trial, one of the three lights in front of you may flash briefly. When you see a light, press the key as fast as you can. If you do not respond before the light comes on, and if you are not unreasonably slow to respond,[4] you will be paid 0.1 cent. This can add up to $1.02 by the end of the experiment.
>
> Pay special attention to the center light. If you miss it when it flashes you will lose 1.0 cent. Concentrate your attention on the center light at the start of each trial.
>
> This complicated bonus system is merely to emphasize what you are supposed to do in this experiment. In summary, devote top priority to the speech sounds and secondary priority to the lights, but do not miss a center light.

The payoff matrix described above was quite liberal, and was introduced primarily to reverse a general tendency for a visual stimulus to dominate an auditory one (cf. Posner, Nissen & Klein, 1976), and also to encourage fixation on the center light. The computer automatically calculated the bonus, and the subject was informed of it at the end of each session. Subjects were also paid a $2.00/hour base rate.

On day one, subjects received 64 practice trials, and on days 2-4, 32 warm-up trials were administered prior to the experimental trials. The experimental session itself consisted of 256 trials, with a 5 minute break after trial 128.

A trial always contained a binaural 1000 Hz warning signal followed 400 msec later by a dichotic CV pair. At one of four times during the trial, one of the three lights could be illuminated for 250 msec. Figure 3.1 illustrates the events in a trial in relation to each other in real time.

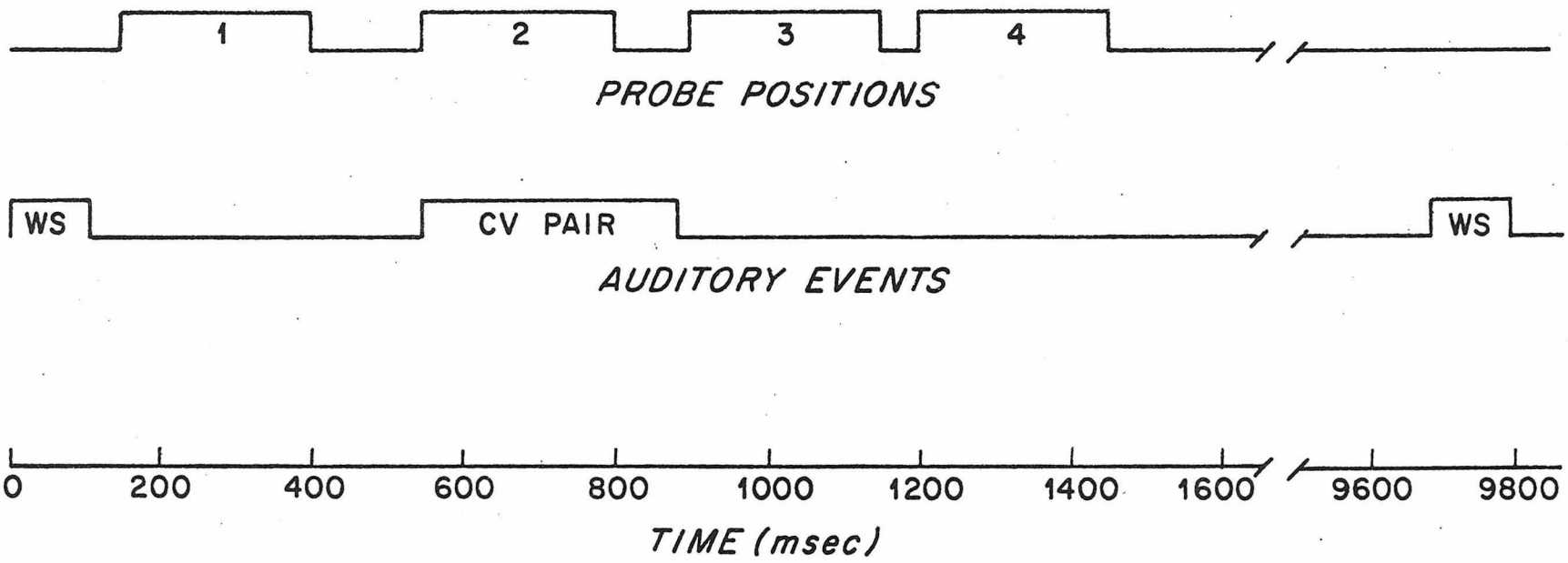
---

[4] Unreasonably slow was defined arbitrarily as 1000 msec. Though such RTs were not paid off, they were included in the analysis of the data.

Figure 3.1

Relative timing of events in Experiment III. Auditory events always occurred in fixed order. A visual probe occurred at a specific probe position with a probability of .1875. The probability of a probe occurring on a given trial was .75.

FIGURE 3.1 RELATIVE TIMING OF EVENTS IN EXPERIMENT III



Six conditions were taken into account in the design of this experiment: response hand, channel-ear assignment, ear monitored, side of light, delay of light, and feature relations of the dichotic pair. Hands were balanced between subjects. Channel-ear assignment and ear monitored were balanced within subjects and between days, with ear monitored balanced within channel-ear assignment. Sides, delays, and features were randomized within days. The randomization of features was fixed across days and subjects - that is, the same tape was used on all four days. Sides and delays were uniquely and randomly assigned to features for each subject on each day with the constraint that there were four replications of each side-delay-feature combination, one for each feature contrast token.

There were four levels of the side factor: center, left, right, and null or no light. Thus 25% of the trials had no light - a control to determine whether the light-speech interference was mutual or one-way. The four delays, in terms of onset asynchrony with the dichotic pair were: -400 msec, 0 msec, 350 msec and 650 msec. The first delay served as a control, the second delay was simultaneous with onset of the speech sound, the third delay occurred 20 msec after offset of the speech stimulus and the fourth delay occurred about at response time for the speech sound.

Results

Dichotic Listening

The results of the dichotic listening task are partially summarized in Table 3.1. A hand by ear by feature by delay by side mixed design analysis of variance was performed on the number of CVs correctly reported from

Table 3.1

Percentage of correctly reported speech sounds as a function of ear and feature contrast in Experiment III.

	Feature Contrast			
	Identical	Place Contrast	Voice Contrast	Double Contrast
Left Ear	99.75	64.38	75.75	73.25
Right Ear	99.75	75.00	85.13	86.00

the monitored ear. Significant main effects of ear, feature and side were obtained.

The main effect of ear ($F(1,6) = 8.01$, $p < .03$) reflected the usual right ear advantage for speech stimuli. The relatively small magnitude of this effect ((R-L)/(R+L) = .019) can probably be attributed to the blocked single ear monitoring requirement of this task.

The main effect of feature ($F(3,18) = 29.67$, $p < .001$) was largely due to identical pairs being reported almost perfectly. They were significantly different on a Newman-Keuls test from voicing contrasts ($t_6 = 22.63$, $k = 4$, $p < .01$), place contrasts ($t_6 = 35.22$, $k = 4$, $p < .01$), and double contrasts ($t_6 = 23.62$, $k = 3$, $p < .01$). The non-identical pairs also showed a feature effect. Place contrasts were significantly worse than voice contrasts ($t_6 = 12.59$, $k = 3$, $p < .01$) and double contrasts ($t_6 = 11.61$, $k = 2$, $p < .01$). Voicing contrasts were not significantly different from double contrasts ($t_6 = .98$, $k = 2$, n.s.). A significant hand by feature interaction ($F(3,18) = 3.44$, $p < .05$) indicated that the feature effect was somewhat stronger for the group that responded with their right hand.

The main effects of ear and feature must necessarily obtain to render the reaction times to the probes interpretable. The significant effect of Side ($F(3,18) = 3.62$, $p < .05$), indicated that there was an effect of the secondary task on the primary task. However, Scheffe tests indicated that the effect was not due to a generalized presence or absence of a secondary task, for the null condition was not significantly different from the center or left conditions. Rather, trials on which the right light flashed showed significant interference with reporting the appropriate CV syllable.

The right light condition was significantly worse than the center, left and null conditions ($p < .05$ in each case by a Scheffe test), as shown in Table 3.3.

There was no main effect of delay ($F < 1$) though there were two higher order interactions with delay, hand by ear by delay ($F(3,18) = 3.41$, $p < .05$) and ear by delay by side ($F(9,54) = 2.57$, $p < .02$). The hand by ear by delay interaction seems to be due to the left light causing greater interference for the left hand response group at delays 2 and 3 ($p < .01$ by a Scheffe test). The ear by delay by side interaction was examined extensively, but appeared to be due to no interpretable pattern and will not be discussed further.

Although the presence of the right light affected performance there was little general effect of secondary task upon primary, as indicated by the absence of a main effect of delay and of a general effect of presence or absence of a light in the main effect of side.

Probe RTs

A hand by ear by feature by delay by side mixed design analysis was performed on the means of those trials where the CV was reported correctly and the RT was longer than 100 msec. Any response shorter than 100 msec including negative RTs was arbitrarily[5] defined as an anticipation. Anticipations accounted for only 0.59% of all responses to the probe that co-occurred

[5]This is a deliberately conservative estimate, for by all accounts the irreducible minimum simple RT to a visual stimulus is around 170 msec (Woodworth and Schlosberg, 1954).

Table 3.2

Effect of side of light upon percent correct identification of CV syllables in Experiment III.

No light	Center	Left	Right
82.88	82.96	83.45	80.13

with a correct response to the dichotic stimuli. Anticipation responses were unrelated to ear monitored ($chi^2(1) = .30$, n.s.), features ($chi^2(3) = 4.4$, n.s.), response hand ($chi^2(1) = .03$, n.s.) or side of light ($chi^2(2) = 3.2$, n.s.), but they were significantly associated with delay ($chi^2(3) = 27.9$, $p < .001$) where longer delays were associated with more anticipations.

The analysis of variance showed significant main effects of feature, delay and side. There was no main effect of ear or hand ($F < 1.0$ in both cases).

The effect of feature ($F(3,18) = 14.73$, $p < .001$) was largely due to trials where identical "pairs" yielded faster probe RT than pairs that contrasted in place, voicing or both features ($p < .01$ in all three cases by a Scheffe test). The significant feature by delay interaction ($F(9,54) = 2.91$, $p < .01$) was due to this pattern at delays 2 and 3 only but not at delays 1 or 4 (see Fig. 3.2). It should be emphasized that post-hoc tests showed no differences approaching significance between non-identical feature contrasts at any delay.

The general form of the delay effect ($F(3,18) = 12.14$, $p < .001$) may also be seen in Fig. 3.2. The improvement in RT from delays 2 to 3 was significant for each level of the feature factor ($p < .01$ by a Scheffe test). The increase in RT from delay 1 to delay 2 was significant ($p < .01$) for the non-identical pairs, whereas the identical pairs show a non-significant improvement in RT over this period. There are no systematic differences for any feature from delays 3 to 4.

The main effect of side of light ($F(2,12) = 5.81$, $p < .025$) is summarized in Table 3.3. Monitoring the left ear seemed to selectively disrupt perception of the right light ($p < .01$ by a Scheffe test).

Figure 3.2

Mean probe reaction times in Experiment III contingent upon probe position and feature contrasts of the co-occurring pair for those trials where the CV was reported correctly.

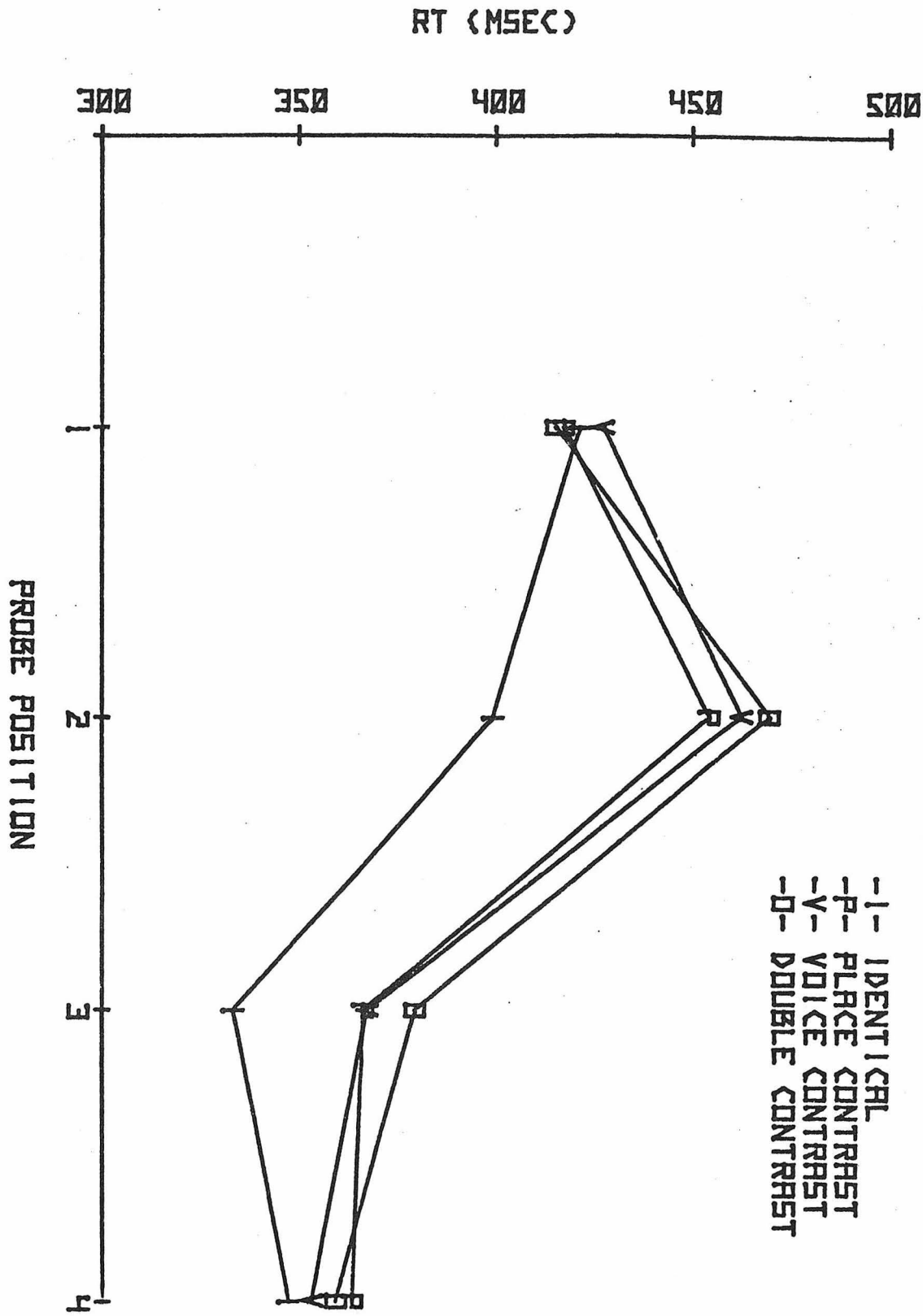

RT (MSEC)
300
350
400
450
500
PROBE POSITION
1
2
3
4
-I- IDENTICAL
-P- PLACE CONTRAST
-V- VOICE CONTRAST
-D- DOUBLE CONTRAST

FIGURE 3.2

Table 3.3

Mean reaction times in milliseconds to probe by ear monitored and side of probe stimulation in Experiment III.

	Side of Probe		
	Center	Left	Right
Left Ear	386.4	389.1	419.3
Right Ear	385.5	398.5	394.7

Discussion

The feature effect obtained in the one-ear identification task of the present experiment more closely resembles the feature effect obtained in the discrimination task of Experiment I than that obtained in the two-ear identification tasks of Experiments I and II. Specifically, in the one-ear identification task and in the discrimination task voicing contrasts and double-contrasts are not different from each other but are both reported more accurately than place contrasts. In the two-ear identification tasks, however, single feature contrasts were reported more accurately than double feature contrasts, but there was no difference between types of single feature contrast. The feature effect from the two-ear identification task was the first reported (Studdert-Kennedy & Shankweiler, 1970), and has been the most studied (Studdert-Kennedy et al, 1972) feature effect in dichotic listening. The configuration of the effect in the two-ear identification task, in which the number of feature contrasts appears to be important, but not the type of contrast, has apparently tended to lead researchers to consider feature effects to be storage or capacity phenomena.

There is clearly an acoustic component of dichotic feature effects, recently elucidated through careful and thorough work on dichotic fusion by Halwes (1969), Cutting (1976) and Repp (1976a,b; also cf. Pisoni, 1975). It seems that dichotic fusion is more likely to occur on pairs that share voicing, making place contrasts more difficult to discriminate or selectively attend to than voicing contrasts.

The fact that the configuration of the feature effect changes as a function of the type of dichotic listening task indicates that there is also

at least one task-specific component to the dichotic feature effect. The requirements of the task seem to modify feature effects in systematic ways, even when the stimuli themselves do not change (also cf. Appendix). This task-specific component does not appear to be related to operations involved in stimulus identification, for major differences in the configuration of the feature contrast effect are clearly in evidence between one-ear (Experiment III) and two-ear (Experiments I and II) identification tasks. By the same logic, it appears that the task-specific component is not related to the use of information from one channel or two. Future research will need to systematically vary requirements of tasks in dichotic listening to understand the way this task-specific component operates.

The lack of systematic effects of feature contrasts of non-identical dichotic pairs on probe RT also argues strongly against feature storage or capacity hypotheses. The notion of selective interference of distinctive features with a limited capacity system has now been tested three times within two different paradigms, all with negative results. Arguments that this negative result was found only because of an insensitive measure now become less convincing because 1) RT is a continuous measure sensitive to small changes unlike digit memory which may be more discrete, and 2) the probe RT measure did, in fact, prove sensitive to one aspect of the dichotic speech signal - i.e. whether or not the dichotic "pair" was identical.

Why did the advantage in probe RT arise for identical "pairs"? First let us consider the general effect of delay depicted in Fig. 3.2. The significant rise in anticipation responses across delays indicates that the absence of an early probe appears to increase the subjective probability of a later

probe. Thus, it appears that more than 25% of trials without any probe are necessary to fully reflect the natural attention demands of the primary task and avoid contamination by subjective expectancy effects. Since evidence of such contamination exists, the interpretation of any general effect of delay becomes difficult. Nevertheless, by comparing the delay effect between identical and non-identical trials, some conclusions may be drawn. From Delays 1 to 2 non-identical trials have a significant rise in probe RT while identical trials actually show a drop. Thus there is something immediate and capacity demanding at, or immediately following onset of a non-identical stimulus pair which is strong enough to override the subjective probability effect. Significant differences between identical and non-identical stimuli exist at Delays 2 and 3. Thus, there is a general demand on capacity for non-identical stimuli from stimulus onset to well after stimulus offset. Processing certain properties of the stimulus, not the physical presence of the stimulus as such, is responsible for this difference.

The difference in probe RT between identical and non-identical stimuli appears to reflect the capacity employed to select one stimulus from two. Since the number of discrete distinctive feature values present in the stimulus pair failed to have an effect in non-identical pairs, it is reasonable to rule out attention to distinctive features as a source of the difference in probe RT between identical and non-identical pairs.

We also know that strong dichotic fusion occurs in pairs of synthetic speech CVs that share voicing (Repp, 1976a,b), yielding one fused percept rather than two distinct ones. Despite the fact that fusion is somewhat

weaker when natural speech is used, as it is in the present experiment, it is likely that fusion contributes to the observed decrease in correct one-ear identification for pairs which contrast only in place (cf. Pisoni & McNabb, 1974). Therefore since probe RT does not differ as a function of type of feature contrast within non-identical pairs, it is reasonable to rule out attention to acoustic features as a source of the difference between identical and non-identical pairs.

One major perceptual difference between identical and non-identical dichotic pairs exists. The identical pairs are heard as unified percepts which are localized in the middle of the head (Cutting, 1976). The non-identical pairs, on the other hand, are heard as percepts which are more diffusely localized and less unified. Shiffrin, Pisoni & Castaneda-Mendez (1974) present evidence which suggests that localization information as such is not useful in attentive processing of speech. They presented one of four stop CVs monaurally at a predictable or a non-predictable ear for identification and found no differences in error rate between the two conditions. An element of difficulty was added to this relatively easy task by the addition of white noise and by the introduction of the non-confusible distractor item /wu/ in the previously empty channel, but still no effect of ear predictability was found.

The fact that non-identical pairs do not yield a unified percept may be important, however. Non-identical pairs contain discrepant phonetic information, thus multiple auditory and phonetic features may be activated. For example, if the dichotic pair were /ba/ - /da/, detectors would be activated for both a rising and a falling second formant. Also both labial and alveolar values of place of articulation would be present. Thus, when the

dichotic pair does not produce a unified and unambiguous percept, more than one phonemic prototype (cf. Repp, 1976a,b) may be activated above its baseline activity rate. Response selection then becomes necessary, and response selection has been shown to demand capacity (Noble et al, 1967). In contrast, the identical "pair" activates features congruent with only one phonemic response, and no selection is necessary. Thus, the response selection stage appears to be the most plausible locus of interference with probe RT for non-identical pairs in the present experiment.

Further evidence of the sensitivity of the present experiment may be found in the existence of laterality effects which are dependent upon aspects of the primary task.

An effect of side of light was present in the dichotic task and the probe RT task. When the right light flashed it was responded to more slowly and caused significant interference with the primary task. In other words, the right light had inhibitory consequences for both tasks. The only apparent explanation for the decrease in right-ear identification performance when the right light flashed is that some aspect of processing the right light was interfering with the processing of the right ear stimulus. Similarly, if one accepts the assumption that if the probe RT task were presented alone there would be no differences in probe RT contingent upon side of visual stimulation (cf. Berlucchi, Heron, Hyman, Rizzolatti and Umilta, 1971), then the logical conclusion would be that the laterality effect in RT was induced by some aspect of the speech perception task. This evidence, which suggests that laterality effects may be induced and modified by aspects of ongoing mental activity is contrary to theories which posit that

lateral specialization arises strictly from the enduring ability of a given cerebral hemisphere to handle certain types of material more efficiently (e.g. Kimura, 1967; White, 1969). Rather, the evidence is more congruent with theories which posit that lateralized processing advantages may be influenced by momentary attentional biases and expectations induced by the nature of a task (e.g. Kinsbourne, 1973). Thus, it is assumed that specialized speech processing mechanisms in the left hemisphere are engaged by the dichotic listening task which a) disrupt or slow processing of a right light which must also be handled by the left hemisphere, and b) are themselves disrupted by the occurrence of a right light. The stage at which this disruption occurs is not isolable within the present experiment.

Superimposed on the main effects of side is an ear by side interaction (see Table 3.2) which can be attributed to orientation or stimulus compatibility effects. Thus, attending to the left ear increases the response time to the right light and vice versa.

The above claims, based on the effects of side of light are strengthened by the finding that the center light was responded to faster than the lateralized lights. This constitutes evidence that subjects generally obeyed instructions and attended to the center light, thus allowing the left light to fall on the right hemiretina and vice versa.

Absence of Evidence for Attentional Processing of Distinctive Features.

The preceding three studies have searched for and failed to find evidence of non-structural interference with the limited capacity system, indicative of attentional processing, in the processing of distinctive features

in speech perception. It is believed that these three experiments, taken together, offer compelling evidence that attentive processing of distinctive feature information does not occur in speech perception.

Given that the processing of distinctive feature information is not attended, it must, then, be automatic. The following experiments were designed to demonstrate such automatic processing.

IV. FATIGUE OF VOICE ONSET TIME DETECTORS WITHOUT ATTENTION

The phenomenon of selective adaptation to speech is now a well established and quite powerful analytic tool in speech perception research (for a review, cf. Cooper, 1975). Eimas & Corbit (1973) conducted the first experimental study of adaptation of voice onset time (VOT).

Perception of the voicing feature in stop consonants is primarily cued by voice onset time (VOT), which is the interval between the release of the stop and the onset of periodic laryngeal pulsing (Lisker & Abramson, 1964). Voiced stops (/b/, /d/ and /g/) generally have short VOTs while voiceless stops (/p/, /t/ and /k/) have longer ones. In English, initial stops with over about 30 msec of VOT are generally heard as voiceless while initial stops with VOT values less than 30 msec are heard as voiced.

Eimas & Corbit (1973) constructed two series of synthetic speech sounds varying in VOT in small steps from /da/ to /ta/ and from /ba/ to /pa/. Subjects were asked to identify randomly selected tokens from these series. Identification functions could be determined from their responses. Category boundaries - the VOT value at which /da/, for example, becomes /ta/ - could then be calculated. After eliciting pre-adaptation identification responses to the series so that a baseline category boundary could be determined, subjects were presented with rapid repetitions of one of the endpoints of the two continua - either /ba/, /pa/, /da/ or /ta/-for at least 1 minute and were then tested for identification of a single token randomly selected from the series. Another identification function was constructed from these responses, and was compared to the pre-adaptation function.

Systematic shifts in identification were found depending on the repeated syllable. When the adaptation syllable contained a voiced consonant, the post-adaptation category boundary had a smaller VOT value, and if the consonant of the syllable was voiceless the VOT at the category boundary increased. These results were interpreted in terms of feature detectors which were fatigued by repetition of the phonemic categories to which they were relevant. This occurred even when the repeated syllable was /ba/ or /pa/ and the identification series was /da - ta/, or vice versa. Thus, Eimas & Corbit (1973) concluded that there exist two detectors optimally sensitive to modal production values for VOT of voiced and voiceless consonants, such that only the detector excited most strongly is capable of reaching higher centers of processing and integration.

Using similar methodology, Cooper (1974a) was able to demonstrate fatigue of phonetic categories along the place dimension. A place identification series was constructed by systematically changing the starting frequencies of the second and third formant transitions while holding everything else constant (Pisoni, 1971). Perceptually, the 13 step series changed from /bae/ to /dae/ to /gae/ in categorical jumps (cf. Liberman, Harris, Hoffman & Griffith, 1957). Category boundaries were expressed in terms of the steps of the identification series. Adaptation with repetitions of /bae/ decreased the size of the /bae/ category (in terms of the number of steps perceived as /bae/), and increased the size of the /dae/ category, leaving the /gae/ category unchanged. Adaptation with /dae/ seemed to increase the size of both the /bae/ and/gae/ categories as it fatigued or decreased the size of the /dae/ category. Adaptation with /gae/ decreased /gae/, increased /dae/ and left /bae/ unchanged. In other words, it appeared

that whatever mechanism was being fatigued was organized to handle place distinctions along a single dimension of analysis corresponding to physical place of articulation in the vocal cavity (cf. Blumstein, 1974). Furthermore, when /bi/ and /pae/ were used as adapters of the test continuum /bae-dae-gae/, they were also effective in decreasing the size of the /bae/ category, though significantly less so than /bae/. Thus, it appeared as if a specifically "bilabial" detector was becoming fatigued.

The results of the Eimas & Corbit (1973) and Cooper (1974a) studies, especially the conditions where the adapting stimuli contained different vowels or consonants than the test series, were interpreted by the authors to be evidence for the existence of speech-specific feature detectors corresponding to phonetic distinctive features (also cf. Eimas, Cooper & Corbit, 1973; and Cooper & Blumstein, 1974).

However, some quite compelling evidence has been obtained indicating that adaptation may be more sensitive to acoustic properties than to phonetic distinctive features. Ades (1974) built two test series /dae-bae/ and /aed-aeb/ and used the four endpoint stimuli from those series as adaptors. He found that the adaptors /dae/ and /bae/ caused significant shifts in the /dae-bae/ series but not in the /aed-aeb/ series. The converse was true when /aed/ and /aeb/ were used as adaptors. Since the phoneme /d/ contains the same linguistic properties or distinctive features whether or not it occurs in /dae/ or /aed/, the evidence from Ades' (1974) study points to habituation of acoustic property detectors rather than phonetic feature detectors.

Pisoni & Tash (1975) reinforced this conclusion by demonstrating a shift in the /ba-da/ boundary after adaptation with stimuli that did not sound like speech. Their "speech-embedded chirps" were constructed by grafting the formant transitions sufficient to produce /b/ and /d/ when they precede the vowel /a/ onto the end of the steady state /a/ formants. The habituation produced by the speech-embedded chirps was significant and in the direction predicted by the acoustic parameters of the adaptor. In other words, the chirp with formants from /ba/ tended to elicit habituation in the same direction as /ba/ itself even though the sound could not be phonetically categorized. Habituation elicited by the speech-embedded chirps was, however, much less than that elicited by the syllables that were perceived as speech.

Along the same lines, Cooper (1974c) used an alternating pair of syllables /da-ti/ as adaptors and tested with two identification series differing on the following vowel, /ba-pa/ and /bi-pi/. He found that adaptation was contingent on the following vowel. That is, in the /bi-pi/ series, identifications shifted in the direction expected from adaptation with /t/, and in the /ba-pa/ series, the shift was in the direction predicted from /d/. Thus, again, when acoustic and phonetic properties were dissociated, fatigue effects followed acoustic properties. Unfortunately, Cooper (1974c) ran no control where /da/ and /ta/ or /di/ and /ti/ were used as adaptors. The category shifts which Cooper obtained were, however, quite small. They averaged less than 2.85 msec of VOT, whereas in the conditions of the Eimas & Corbit (1973) experiment where the consonant of the adapting syllable did not agree in place of articulation with the test series, the average

boundary shift was 7.13 msec. While this comparison is not entirely appropriate, the point is that in the Cooper (1974c) study and in the Pisoni & Tash (1975) study the habituation effects in the conditions that dissociate phonetics and acoustics show weaker adaptation effects than when phonetics and acoustics are not dissociated. Thus, the possibility remains that selective adaptation has a partial phonetic component.

Further evidence supporting this supposition is derived from the phenomenon of perceptuo-motor adaptation. It has been demonstrated that repeated listening to a stop CV can change the average VOT of subsequent production of stop consonants (Cooper, 1974c; Cooper & Nager, 1975) and that repeated production of a stop consonant, even under extreme conditions of white noise masking to eliminate any acoustic feedback, can systematically change identifications in a perceptual test series (Cooper, Blumstein & Nigro, 1975). An acoustic feature detector hypothesis is unable to accommodate these findings. Instead, the findings tend to support the presence of an articulatory and/or phonetic factor in selective adaptation (Pisoni & Tash, 1975; Cutting & Pisoni, 1975).

Selective adaptation was the instrument adopted in the following studies as a measure of speech processing. It was reasoned that the effectiveness of an adaptor in shifting the identification function of a test series would be a measure of the amount of processing of that adaptor. Thus, since processing can be dissociated from any overt responding or other attention demanding operations not directly involved in speech-related auditory processing, the selective adaptation procedure is an excellent tool for studying attention requirements of such processing.

EXPERIMENT IV

Experiment IV was designed to determine whether selective adaptation along the voicing dimension could occur in the absence of selective attention to an adapting stimulus. Given that the three previous studies had not indicated the presence of any attentional effects in the processing of distinctive features, there was every reason to assume that adaptation could occur in the absence of attention. A dichotic tape was prepared with the adaptors /ta/ and /pa/ in random order in one channel and /ba/ and /da/ in the other. It was attempted to limit attention primarily to one channel by the introduction of a target-phoneme monitoring task. The adaptation induced in the various conditions was assessed by means of identification responses of a 14-step VOT series from /da/ to /ta/, administered before and after adaptation. It was expected that the unattended channel would have a significant effect in shifting the test series if non-attended processing were occurring.

Method

Stimuli

The stimuli were constructed on the Haskins Laboratories OVE IIIc synthesizer from two five-formant synthetic stimuli heard as /a.ba/ and /a.da/. The initial steady state portions of the stimuli, prior to the transition of the first formant, were discarded so that the two syllables sounded roughly like /ba/ and /da/. Both stimuli had identical fundamental frequencies and first formants and were 250 msec in duration. They differed only in the direction and extent of their second and third formant transitions.

Formant contours for /ba/ and /da/ are illustrated in Figures 4.1.1 and 4.1.2, respectively.

Two more syllables, /pa/ and /ta/, were created from /ba/ and /da/, respectively, by removing the initial 70 msec of periodic excitation and substituting hiss excitation. Similarly, a 14-step /da - ta/ identification series was created by removing periodic excitation and substituting hiss excitation in 5 msec steps, except for the last step which was a 10 msec step.

These stimuli had no bursts. Instead of bursts, onset amplitudes were increased to create more natural sounding exemplars of the categories employed. Therefore, VOT was defined as the interval from the onset of formant transitions to the onset of periodicity. By this measure the endpoint /da/ and /ba/ had VOT's of 0 msec and the endpoint /ta/ and /pa/ had VOT's of +70 msec. When the above procedures had been executed, all four stimuli sounded like natural speech exemplars of their respective phonetic categories.

A binaural baseline identification tape contained 10 randomized repetitions of each token of the 14-step /da - ta/ continuum. The inter-stimulus interval was 3500 msec. An experimental tape was constructed as follows:

1.) Three "practice" lists of 75 randomized dichotic pairs[6] from the set (/da, ta/, /da, pa/, /ba, ta/, /ba, pa/) were constructed from the

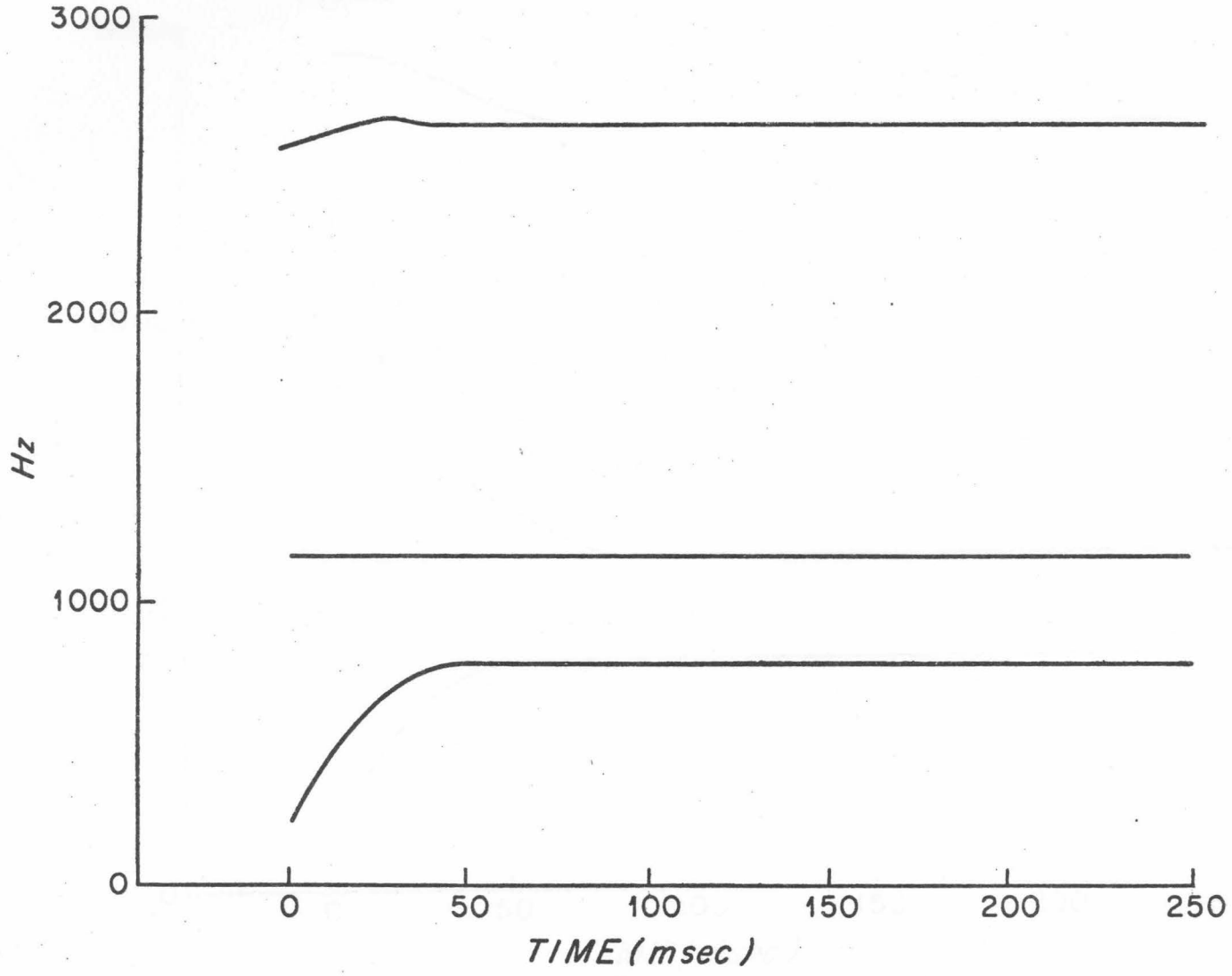
---

[6]The original design called for 76 pairs per list, with equal occurrences of each dichotic pair. Due to a mistake, however, one pair was randomly dropped from each list. The error was not thought to have any material effect on the experiment, and was, of course, taken into account in scoring monitoring performance.

Figures 4.1.1 and 4.1.2

Central formant contours for the stimuli /ba/ and /da/, respectively, used in Experiments IV and V.

FIGURE 4.1.1 FORMANT CONTOURS FOR /ba/



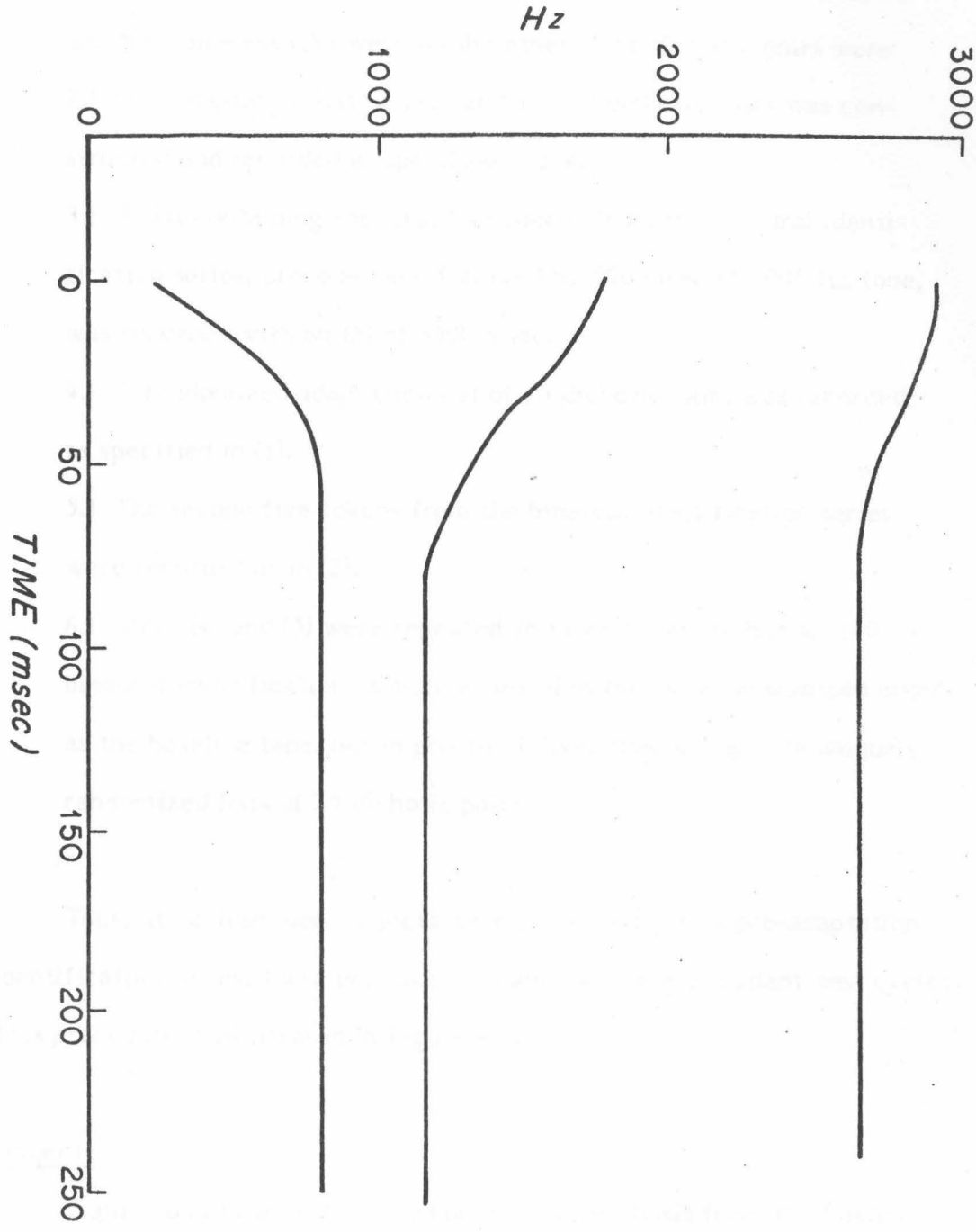


FIGURE 4.1.2 FORMANT CONTOURS FOR /da/

end point examplars described above and recorded with an interpair interval of 500 msec such that the voiced CVs were on one channel and the voiceless CVs were on the other. The dichotic pairs were

2.) One adaptation list of 112 randomized dichotic pairs was constructed and recorded as specified above.

3.) A list containing the first five tokens from the binaural identification series, preceded and followed by 500 msec of 1000 Hz tone, was recorded with an ISI of 3500 msec.

4.) A randomized adaptation list of 75 dichotic pairs was recorded as specified in (1).

5.) The second five tokens from the binaural identification series were recorded as in (2).

6.) Steps (4) and (5) were repeated 26 more times so that all 140 binaural identification stimuli occurred in the same randomized order as the baseline tape, but in groups of five, alternating with uniquely randomized lists of 75 dichotic pairs.

Thus, at each session subjects were presented with a pre-adaptation identification series, three practice lists and twenty-eight adapt-test cycles. This procedure is illustrated in Figure 4.2.

Subjects

Eight subjects were recruited on a volunteer basis from the Boston area and were paid $2.00/hr for their time. All reported that they were

Figure 4.2

Procedure in Experiment IV.

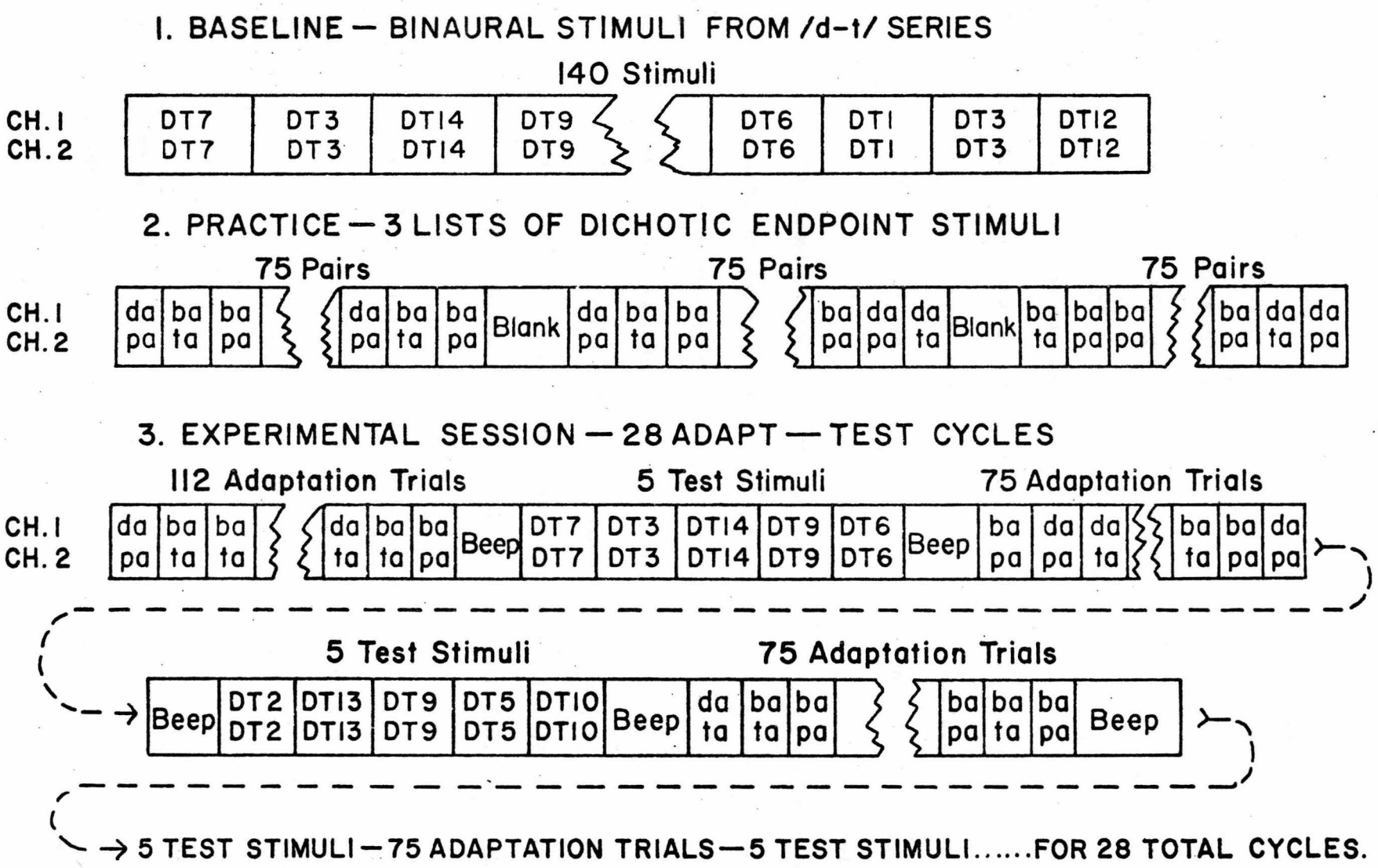


FIGURE 4.2 PROCEDURE IN EXPERIMENT IV

right-handed native English speakers with no hearing deficits. One additional subject was run but was not included in the analysis due to an experimenter error during his last session.

Procedure

Subjects were run individually or in pairs for one hour-long session on each of four consecutive days. Subjects were divided into voiceless and voiced groups on the basis of arrival for their first session. Voiced and voiceless groups monitored voiced or voiceless channels of the dichotic lists, respectively, throughout the experiment. The voiced group always monitored for /da/ and the voiceless group always monitored for /ta/. The monitored channel was in the left ear or the right ear for two days each. On one of those two days the non-monitored channel was on and for the other one it was off. Thus, there were three factors in the experiment - channel monitored, ear monitored and presence or absence of interference.

Stimulus tapes were played on a Sony TC-366 stereophonic tape recorder. The pre-amp outputs of the tape recorder were fed to a Shure Solophone headphone amplifier and then to two pair of Koss K-6LC headphones. The outputs of the two channels were balanced at 80 dB at peak deflection by means of a General Radio sound meter (type 1565Z) at the Solophone. The channels were balanced very carefully due to the fact that physical channels and voicing were perfectly confounded.

Subjects were given a response booklet at the start of each session. The first page corresponded to the baseline identification series and contained 140 consecutively numbered occurrences of the printed letter pair

D-T. Subjects were instructed to circle D when they heard /da/ and T when they heard /ta/ and to always circle one or the other. Pages corresponding to the dichotic lists simply contained columns of consecutive numbers[7] and subjects were instructed to place a check by those numbers corresponding to the trials on which they heard the target phoneme. These pages were alternated with pages containing 5 occurrences of the letter pair D-T. For these pages subjects were instructed as for the baseline identification series. Thus, on every experimental session subjects received a baseline identification series, three "practice" habituation lists, and then 28 habituation-identification cycles.

Results

Category shifts

Category shifts were assessed by the difference between the baseline and experimental identification functions. Two measures were used, the difference in 50% crossover point of the identification function, and the difference in the total number of /d/ responses emitted in the baseline and experimental sessions. The former measure was derived by calculating the baseline and experimental 50% crossover points independently by linear

---

[7] Pilot studies showed that subjects tended to get lost in the monitoring task when going from the bottom of one column of numbers to the top of the next. Therefore, pauses of approximately 1 second were spliced into the adaptation portions of the tape at points corresponding to the ends of columns in the response booklet. Adaptation lists of 75 dichotic pairs had additional 1 second pauses after pairs #25 and #50, while the list with 112 pairs had pauses after pairs #28, #56 and #84.

interpolation and subtracting baseline crossover from experimental. The latter measure was computed by subtracting the number of D's circled in the baseline session from the number circled in the experimental session. The two measures of category shift were in substantial agreement as indicated by the high correlations between them, shown in Table 4.1. Independent mixed design group by interference by ear analyses of variance were also performed on each measure and were in almost absolute agreement on every point. Thus to avoid redundancy, only the analysis of the 50% crossover differences will be reported here.

In this analysis, scores in the direction expected on the basis of the adaptors in the attended ear were expressed as positive numbers and scores in the unexpected direction were expressed as negative numbers. Thus for a subject in the voiced group, a shift toward the voiced end of the series would be a positive number and a shift away would be a negative number. The data on which this analysis is based is presented in Table 4.2.

It can be clearly seen from Table 4.2 that the effect of interference was large and robust ($F(1,6) = 10.95$, $p < .017$). The effect of group approached significance ($F(1,6) = 3.93$, $p < .095$) reflecting a tendency of the voiceless group to show stronger habituation than the voiced group, an oft- reported and ill- explained phenomenon (cf. Eimas & Corbit, 1973). Neither the main effect of ear nor any higher interaction approached significance. The performance of an exemplary subject is illustrated in Figures 4.3.1 - 4.3.4.

Table 4.1

Correlations between two measures of category shift, 50% crossover and number of /d/ responses in Experiment IV.

Ear:	Left		Right	
Interference:	No	Yes	No	Yes
r_{xy} =	.977	.950	.975	.787

Table 4.2

Shifts in 50% crossover point of the /da - ta/ test series in milliseconds of VOT in the expected direction as a function of ear monitored, channel monitored and presence or absence of interference in Experiment IV.

		Left Ear		Right Ear	
		Interference			
S#	Channel Monitored	No	Yes	No	Yes
1	VL	3.50	5.40	19.50	10.65
2	VL	17.50	4.15	7.90	8.95
3	VL	7.50	2.85	1.70	1.45
4	VL	7.00	-3.35	15.50	0.85
5	V	10.85	1.65	6.15	-3.70
6	V	5.85	0.00	6.25	1.15
7	V	5.65	9.90	3.65	1.25
8	V	3.35	5.30	2.10	-5.85
$\bar{x}$		7.65	3.24	7.85	1.85
t_7		4.66	2.30	3.45	.93
p<		.002	.054	.011	n.s.

Figures 4.3.1 - 4.3.4

Baseline and habituation identification functions of an exemplary subject from the voiced group of Experiment IV. Percentage of /da/ responses are contingent upon ear monitored and presence or absence of interference from a channel in the non-monitored ear containing voiceless phonemes.

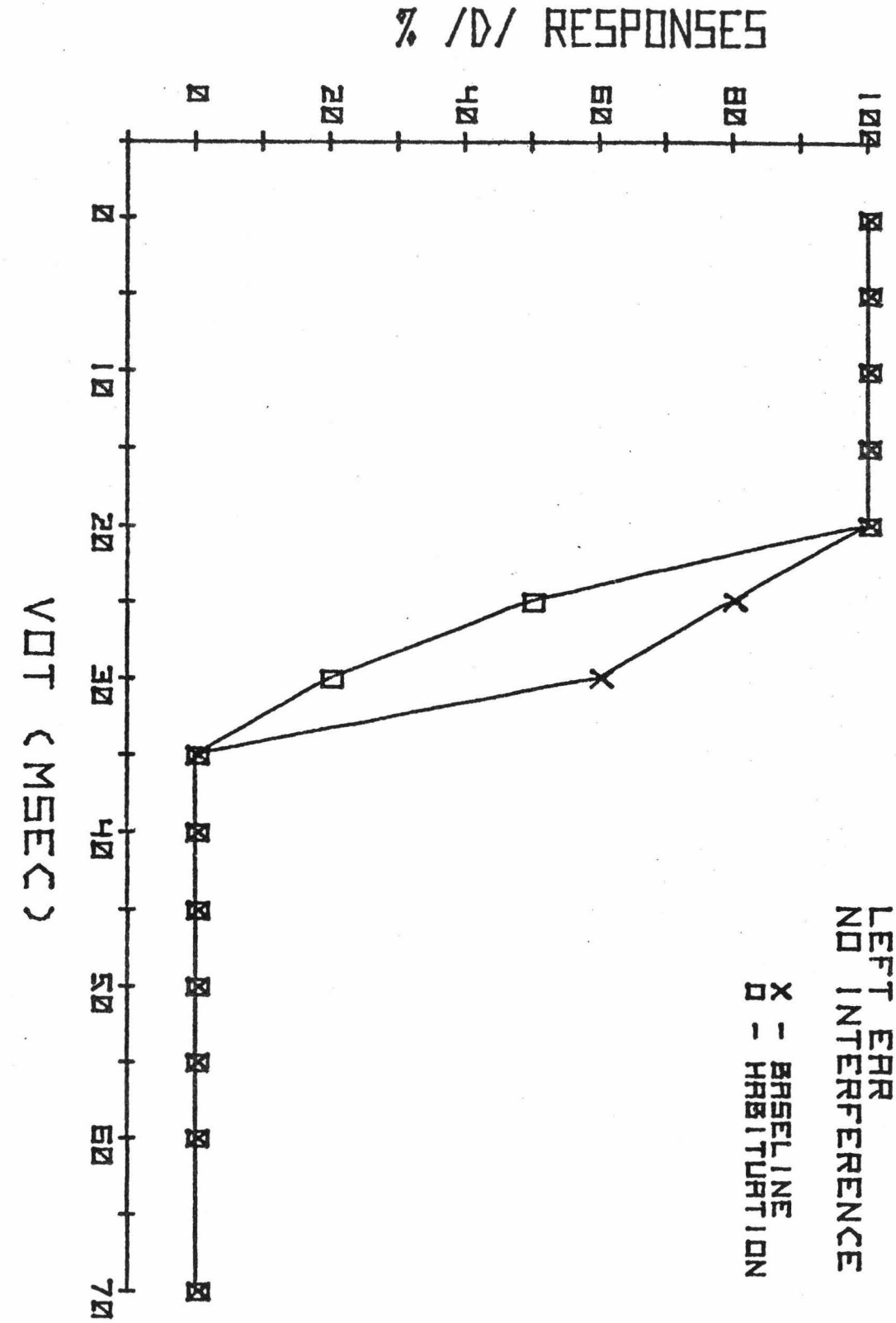


FIGURE 4.3.1
SUBJECT 6
LEFT EAR
NO INTERFERENCE

X - BASELINE
□ - HABITUATION

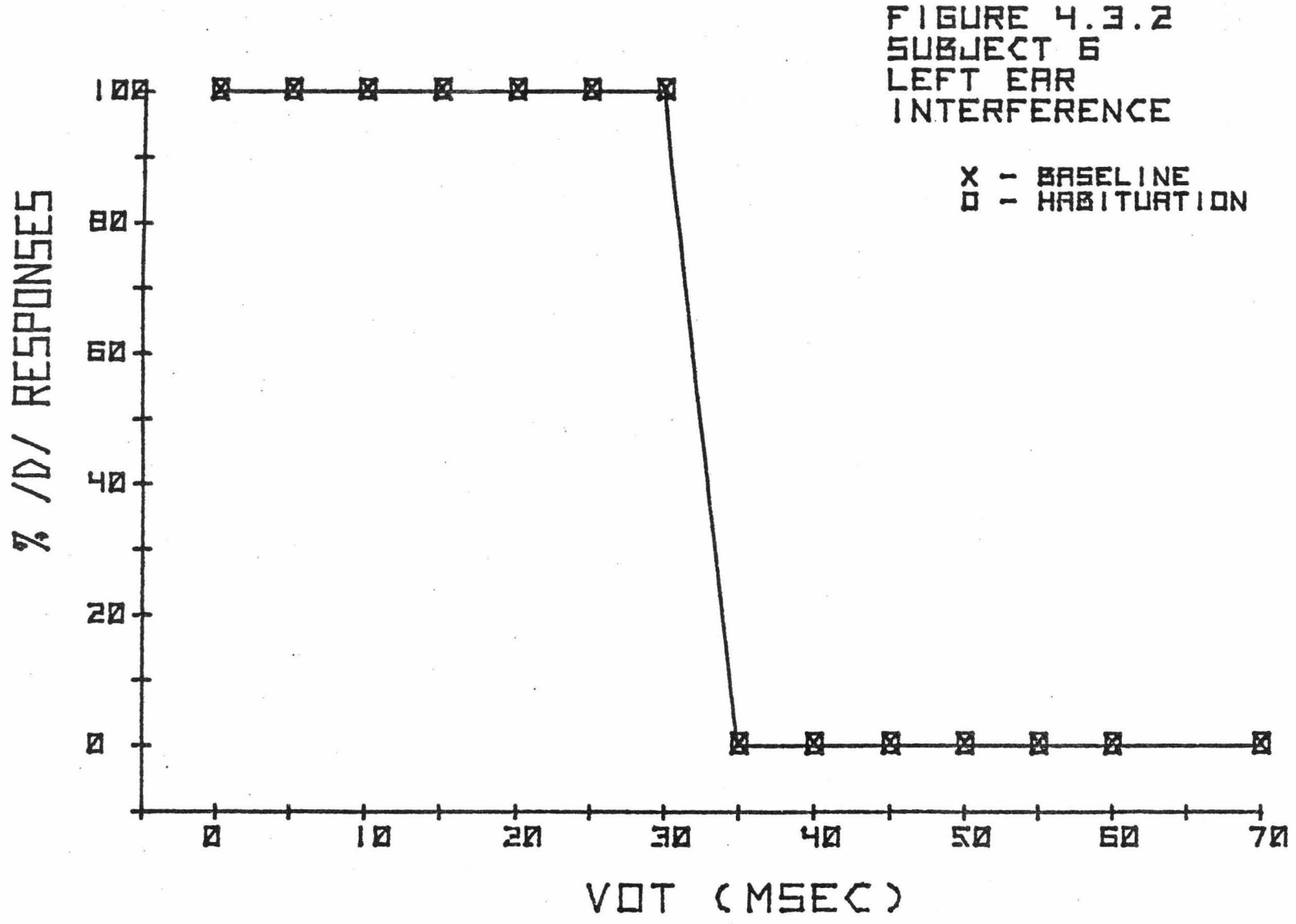


FIGURE 4.3.2
SUBJECT 6
LEFT EAR
INTERFERENCE
X - BASELINE
O - HABITUATION
% /D/ RESPONSES
100
80
60
40
20
0
0
10
20
30
40
50
60
70
VOT (MSEC)

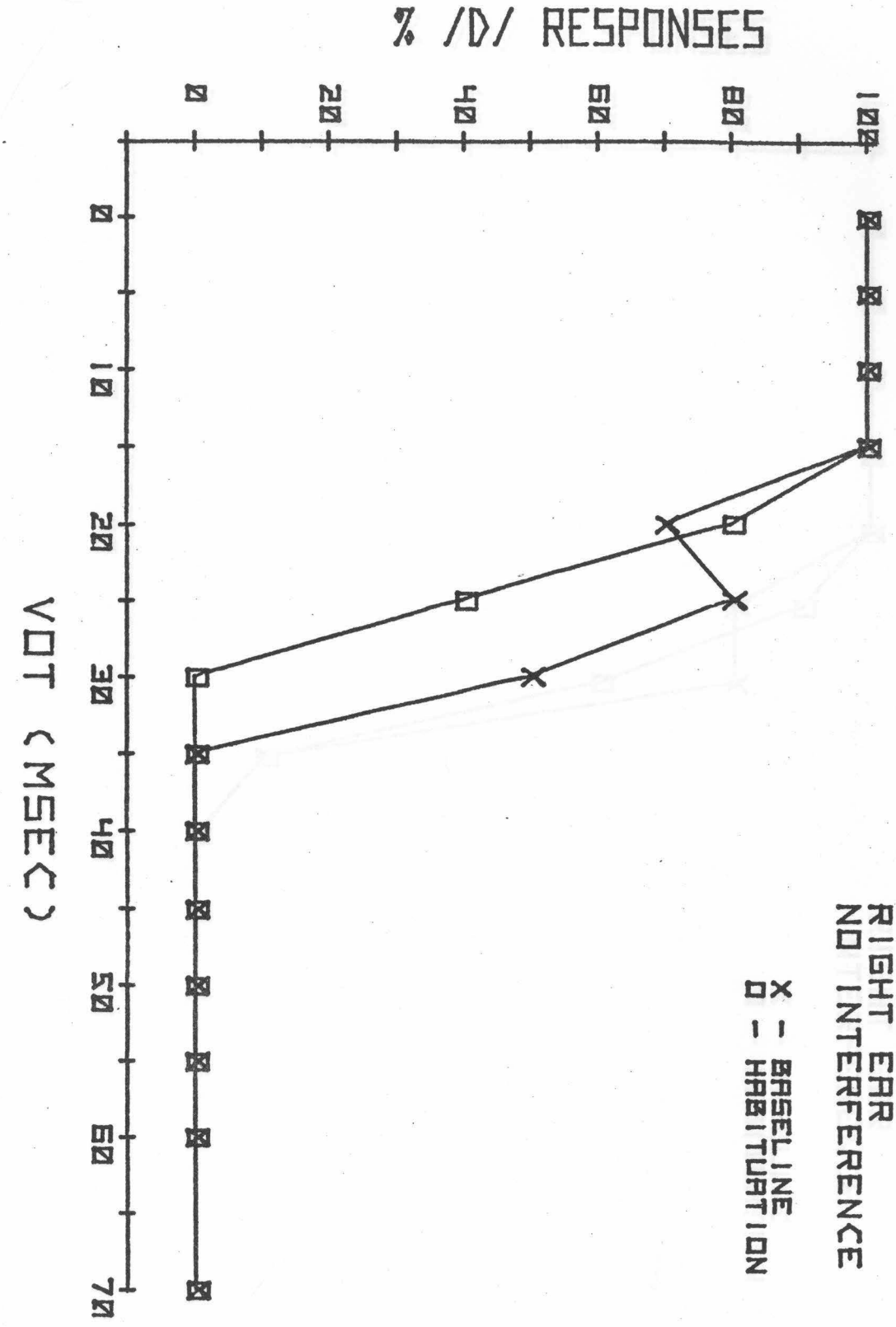


FIGURE 4.3.3
SUBJECT 6
RIGHT EAR
NO INTERFERENCE

X - BASELINE
O - HABITUATION

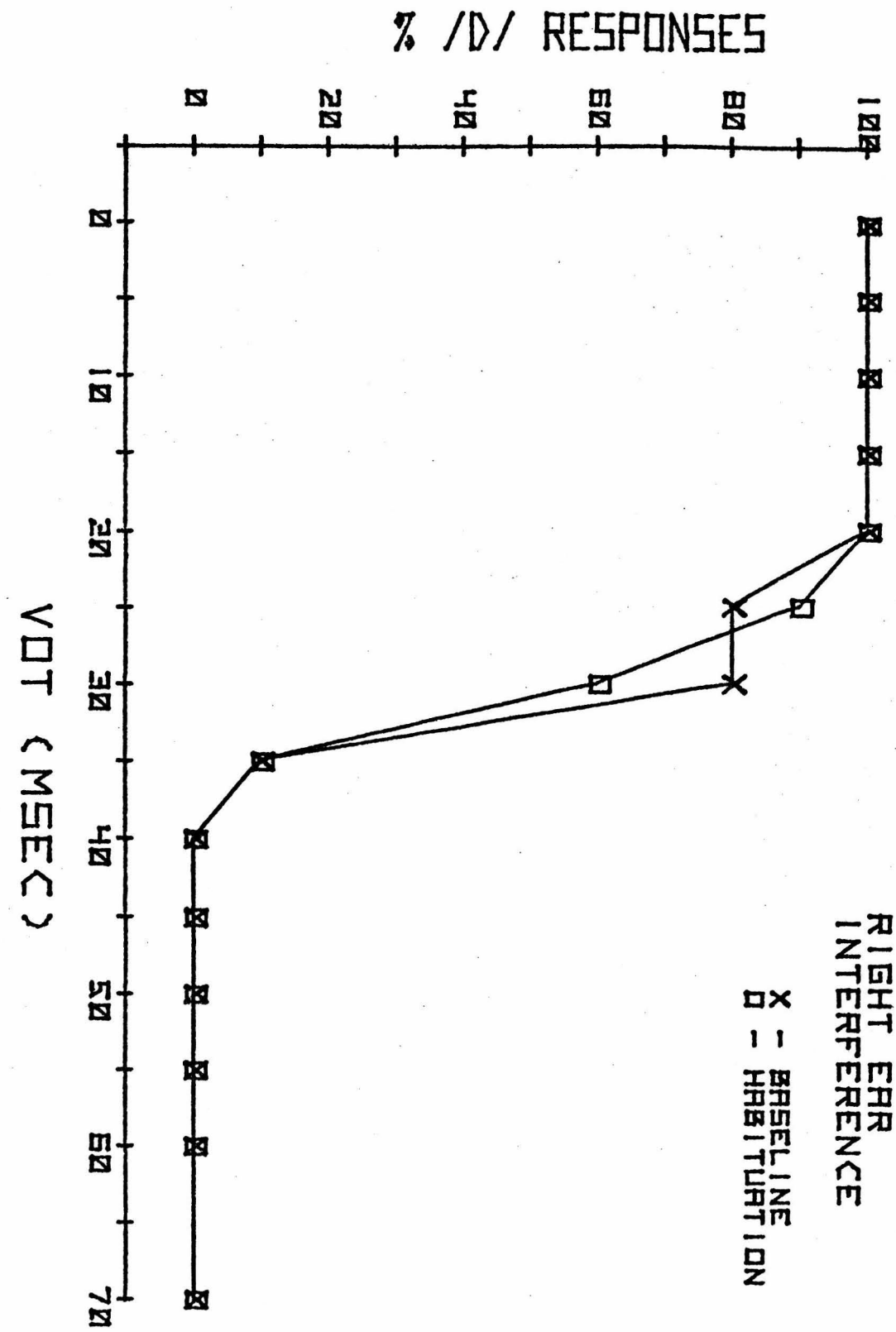


FIGURE 4.3.4
SUBJECT 6
RIGHT EAR
INTERFERENCE

X - BASELINE
□ - HABITUATION

Independent t-tests were performed on each within-subjects condition to assess whether the category shifts were significantly different from zero (see Table 4.2). The no interference conditions clearly showed significant habituation in the expected directions, but in the interference conditions, only the left ear approached significance.

Monitoring performance

Monitoring performance was scored on every fourth list starting with experimental list #4. Percentage correct was subjected to a mixed design group by interference by ear analysis of variance.

The main effect of group was significant ($F(1,6) = 12.39$, $p < .013$) reflecting a tendency for the voiced group to detect /d/ more successfully than the voiceless group detected /t/. The mean percentages correct were 86.20% for voiced and 69.17% for voiceless.

There was also a large effect of interference ($F(1,6) = 46.25$, $p < .001$) reflecting the fact that it was more difficult to detect the target phonemes with interference present in the other channel ($\bar{x} = 71.36\%$) than without ($\bar{x} = 84.01\%$). The interaction of condition by interference approached significance ($F(1,6) = 4.27$, $p < .085$) indicating that this tendency was somewhat stronger for the voiced condition.

The effect of ear, contrary to expectations, did not approach significance ($F < 1.0$) though a right ear advantage was present.

Discussion

The strongest finding of the present study is that the presence of an unattended channel with opposed feature values decreases the strength of habituation induced by an attended channel. It is not likely that this result is due to the distraction of attention to the other channel, for there was little association between shadowing performance and the magnitude of category shift. (see Table 4.3). Furthermore, while it was significantly easier to monitor the voiced channel than the voiceless channel, the tendency was for the monitoring of the voiceless channel to produce stronger habituation.

Thus, it is likely that the decrease in the effectiveness of the attended channel as a habituator was due to the unattended processing of the channel with the opposed feature values.

EXPERIMENT V

Experiment V was designed to 1) replicate the finding that a non-monitored channel could modify habituation induced by an attended channel, 2) rule out distraction as a cause for this modification by holding distraction constant while varying only phonetic structure of the unattended channel and 3) assess any effect of monitoring for a voiced or voiceless target, as distinct from a voiced or voiceless channel.

Method

Stimuli

The four end point stimuli /ba, da, pa, ta/ and the 14-step /d-t/ test series were identical to the stimuli in Experiment IV.

Table 4.3

Correlations of magnitude of 50% crossover shift with percentage correct in monitoring task of Experiment IV.

	Left Ear		Right Ear	
	Interference			
	No	Yes	No	Yes
r_{xy} =	.253	.009	-.042	-.120

Tapes were constructed with the same general format as Experiment IV except for the following differences:

1) The ISI for baseline and experimental binaural test trials was 2500 msec.

2) The ISI for dichotic habituating trials was increased from 500 to 750 msec.

3) All dichotic adaptation lists contained 64 dichotic pairs, including the three practice lists and the initial experimental list. Pauses of 1000 msec occurred after pairs #16, #32 and #48.

4) Two experimental tapes were made. Dichotic lists on the "voiced tape" consisted of the pairs (/da,ba/, /ba,da/, /pa,ba/, and /ta,da/), while on the "voiceless tape" the pairs (/da,ta/, /ba,pa/, /pa,ta/ and /ta,pa/) were used. Channel 1 of both experimental tapes was identical. The terms "voiced tape" and "voiceless tape" refer only to channel 2. Both tapes were recorded in one session to ensure that, aside from the phonemes recorded on channel two, they were identical.

Subjects

Twelve subjects were recruited on a volunteer basis from the Boston Veterans Administration Hospital and were paid $2.00/hr. for their time. Ten were employees of the hospital and two were psychiatric patients who were not on any medication. All reported that they were right handed native English speakers with no hearing deficits.

Two additional subjects were run. One was excused at the end of the first day because of an unresolved confusion about the instructions.

Another subject was excused when she reported hearing the phonemes /k/ and /v/.

Procedure

Subjects were always run individually for one hour-long session on each of three separate days. Subjects were assigned to groups monitoring for /da/ or /ta/ on the basis of their arrival at the first session. Subjects always monitored channel 1, containing the syllables /da/, /ba/, /ta/ and /pa/, which was always in their right ear. On each of three days subjects were presented with one of the following channel 2 stimuli in their non-attended ear: voiced stimuli, voiceless stimuli or, in the control condition, the same stimuli as in the attended ear. The order of presentation of these distractors was counterbalanced within groups and between subjects. Thus, the two factors in this experiment were target and contents of the non-attended channel.

New response booklets were prepared reflecting the fact that adaptation sessions contained 64 stimuli. Five numbered D-T letter pairs to be used in responding to the test series also occurred on the same page.

Stimuli were played on a Teac A-2300 SD tape recorder through Koss Pro-4AA headphones. The channels were carefully balanced at 80 dB peak deflection by means of a General Radio sound meter (type 1565Z). Responses were recorded as in Experiment IV.

Results

Category Shifts

The 50% crossover shift and the shift in the total number of /d/ responses were computed as in Experiment IV. Correlations between the two measures were high (.98, .99 and .95 for control, voiceless and voiced, respectively), and the analyses performed on each measure were in virtually absolute agreement, so only the 50% crossover analysis will be reported to avoid redundancy.

A mixed design unattended ear by target analysis of variance was performed on the absolute direction of the category shift, where a positive shift indicated a shift toward voiceless and a negative shift indicated a shift toward voiced. The data upon which this analysis was performed are shown in Table 4.4.

The contents of the unattended channel had a significant effect ($F(2,20) = 18.03$, $p < .001$). This was due to both the control and voiceless channels being different from the voiced channel ($p < .01$ in each case by a Scheffe test). Neither the target monitored nor the target x unattended channel interaction had any systematic effect ($F < 1.0$ in each case). The performance of an exemplary subject is illustrated in Figures 4.4.1 - 4.4.3.

Monitoring performance

Every fourth list was scored as in Experiment IV. A mixed design unattended channel by target analysis of variance was performed upon the percentages correct. The main effect of condition was significant ($F(2,20) = 11.51$, $p < .001$) indicating that it was easier to monitor for a

Table 4.4

Shifts in 50% crossover of the /da-ta/ test series in milliseconds of VOT as a function of the contents of the non-monitored ear and the target phoneme in Experiment V.

S#	Phoneme Monitored	Control	Voiceless	Voiced
1	T	11.50	15.50	-2.50
2	T	27.85	20.55	0.55
3	T	0.85	-0.75	-4.00
4	T	12.50	-0.50	-23.50
5	T	4.15	7.85	0.15
6	T	15.67	24.00	-2.45
7	D	28.80	20.85	-2.50
8	D	0.70	3.55	-1.65
9	D	10.15	7.00	1.65
10	D	1.00	15.85	-6.65
11	D	-2.50	0.00	-10.85
12	D	2.10	6.25	-2.65
$\bar{x}$		9.40	10.00	-4.50

Figures 4.4.1 - 4.4.3

Baseline and habituation identification functions of an exemplary subject in Experiment V contingent upon type of interference from the non-attended channel.

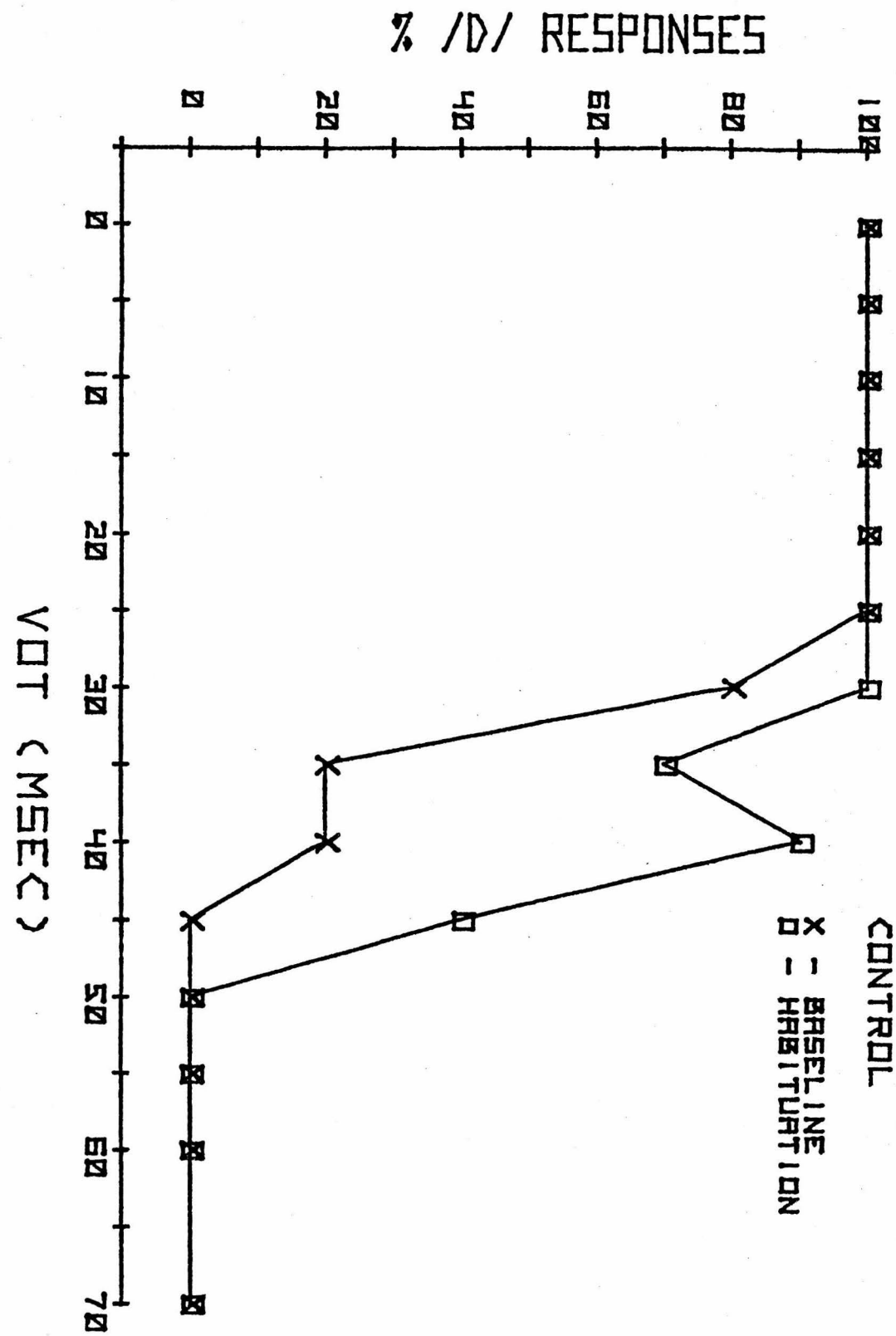


FIGURE 4.4.1
SUBJECT 1
CONTROL

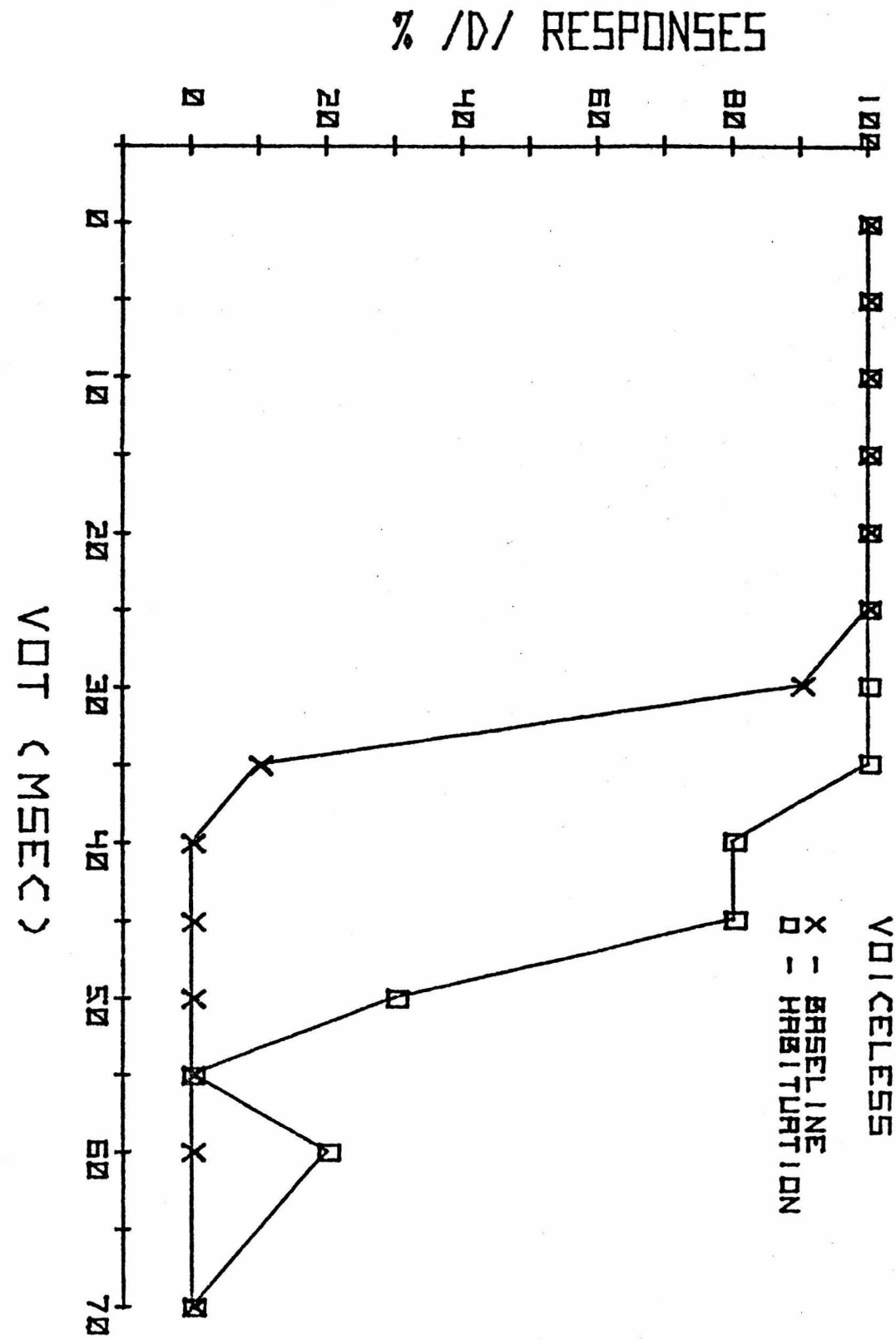


FIGURE 4.4.2
SUBJECT 1
VOICELESS

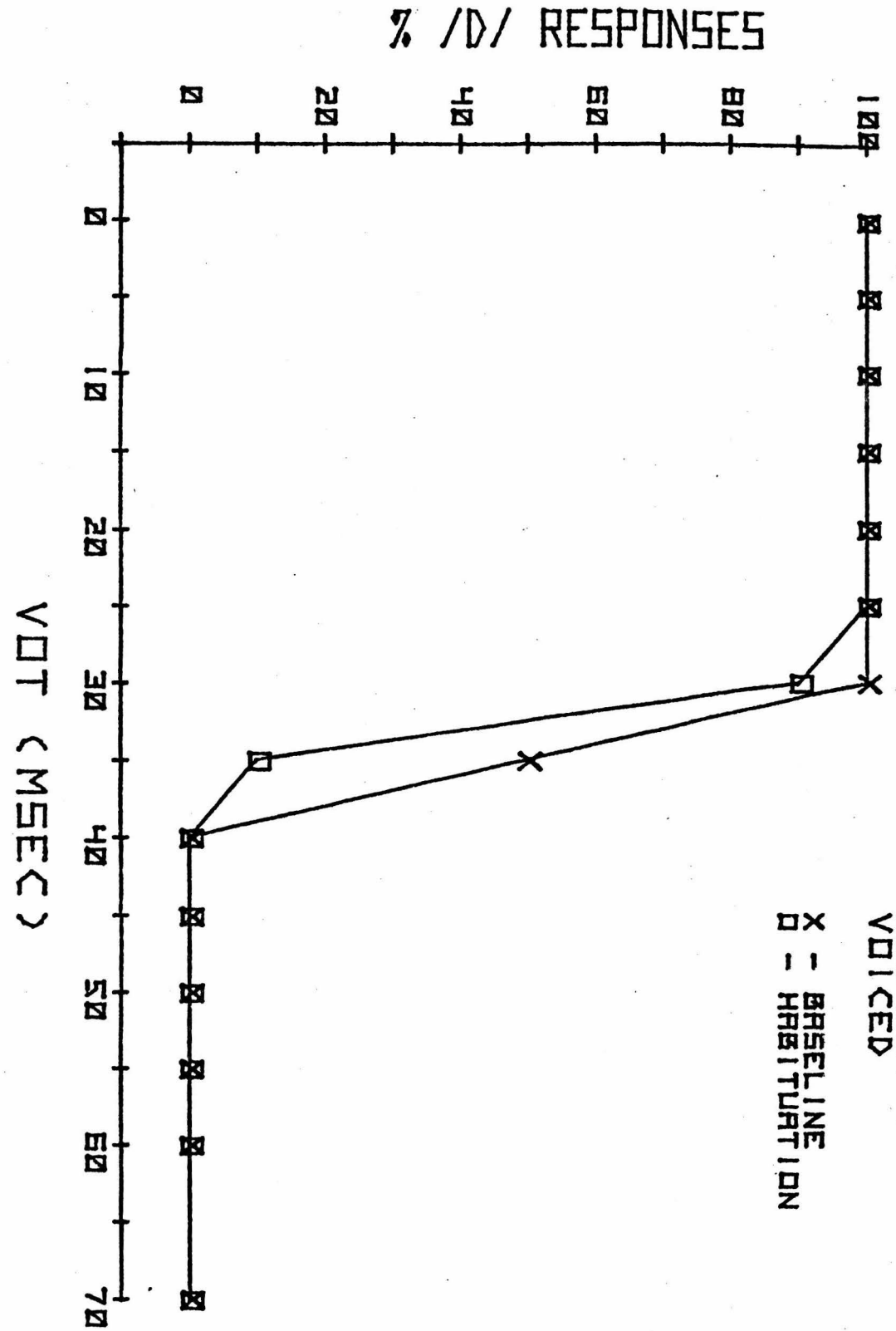


FIGURE 4.4.3
SUBJECT 1
VOICED

X - BASELINE
□ - HABITUATION

target when both channels were the same ($\bar{x}$ = 97.78%) than when the non-monitored channel was different (voiced $\bar{x}$ = 85.58%; voiceless $\bar{x}$ = 84.78%). Both voiced and voiceless conditions were significantly different from the control ($p < .01$ by a Scheffe test in each case), but were not different from each other.

The interaction of unattended channel with target was also significant ($F(2,20) = 30.63$, $p < .001$). This was because monitoring for the target /d/ in the attended channel was more difficult when the unattended channel was voiced ($\bar{x}$ = 76.12%) than when it was voiceless ($\bar{x}$ = 99.07%). Similarly /t/ was more difficult to detect when the unattended ear was voiceless ($\bar{x}$ = 70.49%) than when it was voiced ($\bar{x}$ = 95.01%).

The main effect of target phoneme was not significant ($F(2,20) = 2.11$, $p < .178$).

Discussion

Significant differential effects of the structure of the non-monitored channel upon the direction and extent of the shift of a phonological category were demonstrated. This result is due to the unattended phonetic or speech-related acoustic processing of the non-monitored channel.

It could conceivably be argued, however, that the non-monitored channel was being processed attentively by means of time-sharing or attention-switching (cf. Moray, 1969). Lewis (1970) delineates three factors for use in dichotic shadowing studies to limit the allocation of attention to one channel only. These are 1) The use of unrelated sequences of stimuli, 2) The use of a fast presentation rate and 3) The requirement

of low error rates. The first criterion exists so that subjects are not able to predict stimuli from known sequential dependencies, but must instead listen in real time to achieve correct performance. Experiments IV and V meet this criterion. The second criterion exists so that subjects are not able to completely process a stimulus on the monitored channel and then switch attention to the non-monitored channel. At 750 msec/syllable in Experiment IV subjects made considerable errors even in the non-interference condition. Therefore one channel of the present stimuli could not be perfectly processed at that rate. The presentation rate was slowed to 1000 msec/syllable for Experiment V and near-perfect performance on one channel alone was achieved. However, the fact that error rates increased substantially when a second channel was added is strong evidence that this rate was not too slow. The third criterion is based on the assumption that lapses in shadowing performance could, though must not necessarily, indicate lapses in or misdirections of attention. In Experiment V near-perfect performance was achieved in the conditions where the voicing value of the target was discrepant from that of the non-monitored ear. Habituation elicited in these conditions does not appear to differ from that elicited where the voicing of the target and the non-attended ear matched. In fact, to the extent that attention and correct performance are correlated in the present experiments, a stronger criterion would be that correct performance is unrelated to habituation. In fact, performance and habituation are clearly dissociable in both experiments.

One plausible way to account for the results is based on the phenomenon of dichotic fusion. Dichotic fusion is a phenomenon which tends

to occur when fundamental frequencies of the two members of a well aligned and matched pair of dichotic stop CVs are identical (Halwes, 1969). When such fusion occurs, only or single speech sound is heard. For example, when /ba/ and /ga/ are the two members of the pair, a single fused percept /da/ is often reported (Cutting, 1976). Similarly, when /ba/ and /da/ are presented, either /ba/ or /da/ may be reported depending on stimulus dominance and ear dominance effects (Repp, 1976a).

Given that the fundamental frequencies were identical for all dichotic pairs in the present experiments, it was plausible that fusion was occurring and that the fused percepts produced by it were both being attended and were accounting for the category shift. The present author heard pairs which contrasted only on place as fairly unified percepts, though when a voicing contrast was present the vowel portion sounded fused, but the hiss excitation was clearly isolable from the periodic excitation.

Tartter (personal communication, 1976) reports strong habituation effects when subjects play cards or draw while hearing the adapting stimuli. This is a distracting situation in which fusion cannot occur, for the distracting stimuli are non-auditory. Rather, this evidence tends to confirm that speech-related auditory processing may occur when attention to the acoustic stimuli is greatly reduced. The effects of non-auditory distractors upon the habituation of speech-related features should be investigated more thoroughly and systematically, however, to effectively rule out the alternative explanation based on dichotic fusion. Further studies of the effects of fundamental frequency differences and the presence of fusion in habituation would also be helpful in the attempt to isolate the stage at which habituation occurs.

In the light of the existence of perceptuo-motor adaptation (e.g. Cooper & Nager, 1975; Cooper, Blumstein & Nigro, 1975) it is highly interesting that no effect of target phoneme emerged. Here extreme caution should be exercised in accepting the null hypothesis, for a slight trend in the expected direction does exist in the data of Experiment V. Further exploration of this finding is warranted, for to the extent that monitoring for a phoneme involves the repeated activation of that mental category it provides a test of the limits of perceptuo-motor adaptation. In other words, the phoneme monitoring effect, or lack of it, provides a method to determine whether abstract central categories exist in the same code as peripheral perceptual and motor tokens of that category.

Every study of VOT adaptation where a distinction between voiced and voiceless adaptation could be examined has found voiceless stimuli to be more effective habituators than voiced stimuli (Eimas & Corbit, 1973; Miller, 1975; but cf. Warren & Gregory, 1958). Eimas, Cooper & Corbit (1973) used alternating voiced and voiceless stimuli as adaptors and found a net increase in voiced responses to their test stimuli. This effect has now been replicated in an adjunct study by the present author, using the stimuli of Experiment IV in the absence of selective attention instructions (see Table 4.5). In other words when subjects are asked to simply listen to dichotic stimuli which are voiced in one ear and voiceless in the other ear, more voiced responses to the test stimuli are made ($t = 1.95$, $p < .10$). This finding is possibly best explained by Eimas et al (1973; also cf. Cooper, 1975) that voiced phonemes are more common across languages (Lisker & Abramson, 1964) and occur first in children's language (Port & Preston,

Table 4.5

Shifts in 50% crossover point of the /da-ta/ test series in milliseconds of VOT as a function of channel-ear assignment when the stimuli of Experiment IV were heard without selective attention instructions.

S#	Voiceless-RE Voiced-LE	Voiced-RE Voiceless-LE
1	10.00	-.50
2	0.00	-3.35
3	0.00	2.00
4	4.30	3.65
5	0.85	0.00
6	15.00	7.10
7	0.60	3.00
$\bar{x}$	4.39	1.70

1972). However, a more recent result shows that voiceless stops are more identifiable as a class than voiced stops in unfamiliar languages, whether the identification is performed from a spectrogram (66% versus 35%), the waveform (66% versus 21%) or from the acoustic speech signal itself (77% versus 54%) (Shockey & Reddy, 1974). Thus, the voiced feature is more stable and common, while the voiceless feature is better specified acoustically (also cf. Miller & Nicely, 1955).

The surprisingly strong voiceless adaptation found in the control condition of Experiment V in conjunction with the lack of a significant difference between control and voiceless conditions suggests that fatigue of feature detectors is limited to a certain range close to the category boundary. This suggestion is reinforced by the reversal of direction of category shift in the voiced condition.

The present suggestion is at odds with a hypothesis proposed by Miller (1975). She found that single-ear identifications of good endpoint exemplars of voiced and voiceless stop CVs presented in non-identical dichotic pairs shifted in expected directions after repeated listening to voiced or voiceless adaptors. Her conclusion was that adaptation affected the entire range of the relevant detector's sensitivity. The present data suggest that detectors are only subject to fatigue within a limited range near the category boundary. This conflict is reconcilable in that Miller's data are subject to an auditory averaging explanation (cf. Repp 1976a,b; Cutting, 1976). By this explanation a fused percept from a dichotic pair contrasting on voicing would be ambiguous with respect to voicing and would tend to be identified as voiced after fatigue with voiceless adaptors and voiceless after fatigue with voiced adaptors.

Non-attended processing of a speech-related feature

Experiment IV demonstrated that a non-attended channel could influence identification of a voicing series. Experiment V confirmed this and indicated that the phonemic-acoustic structure of the non-attended channel was responsible for this effect. Taken together with experiments I, II and III, experiments IV and V reinforce the conclusion that speech processing up to the phonetic level is an automatic process in which attention is not required.

V. ON THE ROLE OF ATTENTION IN THE PERCEPTION OF SPEECH

An information processing overview

A view of speech perception as a series of processing stages which are organized hierarchically has emerged and been elaborated in recent years (Liberman, 1970; Studdert-Kennedy, 1974; Cutting & Pisoni, 1975; Cutting, 1976; Sawusch, 1976). Such a view offers a useful framework in which to examine and discuss the results and conclusions of the present experiments.

Serial stage theories all have certain common properties. They all assume at least one auditory analysis stage which is thought to be the first stage of analysis that the acoustic speech signal undergoes. At this stage, representations of time-varying frequency-intensity relations, envelope shape and other auditory properties of the signal are thought to be extracted from the signal. The presence of noise and/or periodicity, the presence, direction and extent of formant transitions and other abstract properties are also thought to be mentally encoded during auditory analysis, though some investigators assign these processes to a second auditory stage (e.g. Sawusch, 1976). Preliminary auditory analysis occurs in all auditory perceptions, and thus is not a speech-specific process. Later auditory analysis processes, however, are thought to be increasingly speech-related (Sawusch, 1976).

A second stage assumed by this approach is a phonetic stage at which phonetic distinctive features present in the signal are specified. This is a stage that is specifically linguistic - the abstract properties of the signal are now assigned to categories which are linguistically relevant in terms

of the articulatory patterns of a language. Since invariant acoustic cues for distinctive features are not always present, the serial stage theories propose that a many-to-one and one-to-many mapping process from auditory properties to phonetic distinctive features occurs here.

A third stage, the phonological level, is commonly postulated. At this level variations which are not relevant to a language are no longer represented. Thus the /t/ in "ten" and /t/ in "latter" are categorized identically.

Were this stage theory to be translated in terms of the analysis by synthesis process (Stevens, 1972), the acoustic level would be the input to the process, the phonetic level would be equivalent to the internal representation system of the process itself, and the phonological level would be the output of the process.

Higher levels of analysis exist - lexical, syntactic, semantic and pragmatic. These are generally treated only in passing by theories of speech perception. Their main function is thought to be "to clean a noisy message" (Studdert-Kennedy, 1974). Stevens (1972), for example, allows that the lexicon may interact with the control process of the analysis by synthesis model. The extent and manner of this interaction are not specified, however, and presumably not crucial to the recognition process. The model is clearly designed with the goal of being an adequate speech recognizer without reference to such higher order processes.

In fact only the first three levels - acoustic, phonetic and phonological - properly belong to the realm of speech perception. Analysis beyond the phonological level is not theoretically necessary for unambiguous categorization of the acoustic speech signal to occur (Studdert-Kennedy, 1974) - the

present studies amply demonstrate this fact. Phonetic and phonological stages, however, are also relevant to higher linguistic levels as well as to speech. As higher linguistic stages they may also be classed with lexical, semantic and syntactic levels.

The experiments, the results and the acoustic-phonetic distinction

The distinction between auditory and phonetic processing is crucial to the interpretation of the present studies. There is no reason to assume that preliminary analysis of the properties of an acoustic signal requires attention. Auditory properties are essentially one-to-one transforms of the acoustic signal - no hypothesis formation, testing, response selection, or other attention demanding processes need be invoked. Pisoni & Sawusch (1975) and others make this assumption explicitly.

The phonetic stage, on the other hand, is where several crucial processes central to speech perception are thought to occur. The specification of a distinctive feature matrix includes features which are variably cued as well as those which are known to be invariant. The specification of a phonetic feature matrix also assumes that segmentation has occurred. Otherwise it would not be possible to know which feature values go with which phonemes. The processes hypothesized to handle variable cues and segmentation at this stage were construed in the present Introduction to require attention.

The feature effect in dichotic listening found by Studdert-Kennedy & Shankweiler (1970) was construed as a phonetic effect because when

acoustic and phonetic levels were experimentally dissociated by manipulating the vowel environment the feature effect was not significantly modified (Studdert-Kennedy et al, 1972). That is, vowels following the consonants in a given dichotic CV pair could be identical or different, but it made no difference to the effect. Furthermore the presence of feature blend errors (Studdert-Kennedy & Shankweiler, 1970; Halwes, 1969) or phonetic feature fusions (Cutting, 1976) seemed to confirm that this effect was truly phonetic - patterns of errors appeared to indicate that at the level of this effect feature values existed in more or less independent form, unbound to distinct phonemes.

However, certain acoustic variables such as differences in relative onset times or intensities of the members of a dichotic pair are known to modify a one-ear identification feature effect in certain systematic ways (Pisoni & McNabb, 1974). These findings imply that the locus of the dichotic feature effect is in part at the auditory level. The presence of effects of auditory manipulations does not, however, necessarily indicate that the locus of feature effects is at the auditory analysis stage. It could be, for example, that the acoustic manipulations of Pisoni & McNabb (1974) were affecting the auditory information available to the phonetic feature analysis stage, rather than suggesting that the auditory stage itself is a more viable locus for the feature effect.

Cutting (1976) demonstrated that phonetic feature blends were subject to modification by different ranges of acoustic parameters than were lower level auditory phenomena such as auditory localization or spectral-temporal

fusion. Feature blends were maximal at greater stimulus onset asynchronies than were "lower" fusions and were not as subject to interaural differences in intensity. Furthermore, binaural mixing of the dichotic stimuli was disruptive, whereas for "lower" auditory fusion phenomena it was actually helpful, indicating that independent peripheral analysis of the members of the dichotic pair was necessary for feature blends to occur.

Feature effects resembling those of Studdert-Kennedy & Shankweiler (1970) and Studdert-Kennedy et al (1972) were found in the two-ear identification conditions of Experiments I and II. The demonstration of Studdert-Kennedy et al (1972) was taken as sufficient evidence that these were phonetic effects.

The extreme variability of dichotic feature effects found in Experiments I and III (as well as Experiment VI in the Appendix) was taken to indicate a strong task-dependent component in dichotic feature processing - reflecting several different ways in which acoustic and phonetic information is used in dichotic listening. This task dependent variability is discussed more fully in the Appendix, where it is concluded that feature effects are in fact multiple component effects reflecting not only acoustic information available to the phonetic stage processing, but also strategic and response organization effects.

No effect of phonetic structure upon the limited capacity system was obtained in any dichotic two-ear identification task (but cf. Appendix for a discussion of the discrimination task results). It was assumed that the short-term digit-memory task provided an adequate assessment of demands upon the limited capacity system.

Shulman & Greenberg (1971) have demonstrated that memory interferes with a perceptual task at central levels as a function of difficulty of the memory task. However, in the present studies the converse finding, that perception interferes with memory as a function of the difficulty of the perceptual task, was not obtained. This was true whether difficulty was measured theoretically by the number of distinctive features to be processed or empirically by correct performance. This absence of an effect was especially surprising since the presence of blend errors in dichotic listening suggests that feature values are preserved in memory in independent form.

A distinction is often made between a sensory memory which preserves relatively raw or precategorical sensory information prior to processing, is limited in capacity only by the characteristics of the specific sensory system and is passively maintained, and a system which preserves categorized information from any cognitive or perceptual operation actively and by rehearsal (cf. Bjork, 1975). Given that Blumstein & Cooper (1972) and others propose that feature effects are due to phonetic features (i.e. categories) interacting within a limited capacity system, the latter active system is implicated. This is the STM system, which is defined in common sense terms as that memory in which one holds a telephone number from the time it is looked up until it is dialed. Thus a digit memory task seemed appropriate to measure capacity requirements in STM. However, no effects of features processed upon this STM task were found.

There is no way of knowing from the present data if feature effects occur in a feature specific memory such as the feature buffer proposed

by Pisoni (1975). Perhaps the auditory processing stage is associated with an auditory-property memory, the phonetic stage with a phonetic feature memory, the phonological stage with phoneme memory and so on.

Interference of processing distinctive features with the limited capacity system was assessed independently of memory by a visuo-motor probe RT task in Experiment III. The associated dichotic listening task involved the identification of only one member of the dichotic pair, but the fact that distinctive feature relationships of the pair affected performance on the dichotic listening task provided evidence that the features of both members of the pair were being processed. Still no effect of processing distinctive features upon the limited capacity system was obtained. An effect of "identical" pairs was obtained and was attributed to the fact that when pairs were non-identical, an attention demanding response selection process was undergone. However, within the non-identical pairs, variations of number and type of feature contrast had no effect. It was concluded that processing phonetic distinctive features takes no space in the central limited capacity mechanism.

It was then attempted to demonstrate that processing of phonetic distinctive features was possible in the absence of attention. Attention was defined here in terms of a phoneme monitoring task - monitoring one channel alone was shown to require a large proportion of available capacity because performance was shown to vary with rate of presentation from Experiment IV to Experiment V. Processing was defined such that if selective adaptation occurred it was assumed that the adapting stimuli were being processed at some level. Evidence regarding whether selective adaptation is phonetic or auditory is reviewed in the introduction to Chapter IV.

It is indisputable that certain components of selective adaptation are strictly auditory, but it also is likely that a phonetic component of the process exists. To the extent that the selective adaptation procedure is phonetic, phonetic processing was demonstrated to proceed in the absence of attention in Experiments IV and V.

A non-monitored channel was introduced and its phonetic structure was varied. The effects of this non-monitored channel upon selective adaptation were large and systematically in the directions predicted by the automatic processing hypothesis. The possibility of attentive processing of this non-monitored channel was ruled out when monitoring performance was found to be unrelated to the obtained habituation effects. It was therefore concluded that phonetic processing, or at least speech-related auditory processing, could occur in the absence of attention.

Some speculations on speech perception as an automatic process

The data from the present studies tend to support the notion that the mental encoding of phonetic distinctive features in speech perception is a direct and automatic process, which may occur without involving the limited capacity system, similar to other types of mental encoding - from letters (Posner & Boies, 1971) and simple visual shapes (Posner & Klein, 1973) to semantic features (Lewis, 1970; Conrad, 1972; MacKay, 1973). The active theories of speech perception (Liberman et al, 1967; Stevens, 1972) contain hypothetical processes assumed to require attention. To the extent this assumption is a valid extention of the theories under consideration, the present studies may be taken as disconfirmation of such theories.

The present studies point to a need to formulate new models of speech perception which incorporate direct and passive encoding as the major perceptual processing mode. The easiest way to do this, of course, would be to assume that the lack of invariance between acoustic and phonetic codes is not a real problem - that invariants in the acoustic speech signal will eventually be found. While it is clear that the speech spectrogram represents the acoustic signal in a way which conveys phonetic information, it is likely that the spectrogram does not convey certain information which may be invariant.

In fact, Stevens & Blumstein (1976) claim to have found acoustic properties for initial stops which appear not to change as a function of following vowel. Such properties are specified in terms of the frequency range and diffuseness of spectral energy at stimulus onset. If this preliminary finding is borne out, it will be able to account for the heretofore problematical perceived identity of /d/ in /di/ and /du/ (see Fig. 1.1). The lack of invariance problem will be solved for a major class of speech sounds.

Furthermore, other types of information not directly represented in spectrograms has been shown to carry significant linguistic information. Mermelstein (1975), for example, has developed an algorithm for the automatic segmentation of speech into syllables which is over 90% accurate. It is based on analysis of the convex hull of the amplitude envelope of the acoustic waveform of the speech signal. Thus, it is likely that human beings are able to use many more acoustic cues than those represented in spectrograms.

One tool which speech scientists will find useful in the search for invariant acoustic-phonetic relations is a recently developed mathematical modelling process called catastrophe theory (Zeeman, 1976). The theory provides ways to mathematically represent the interaction of continuous dimensions to create discontinuous or categorical phenomena. Such a theory will prove useful to the extent that the relevant articulatory-acoustic parameters can be adequately quantified and specified in four or fewer dimensions. If these criteria can be met, catastrophe theory provides a method of direct computation of categorical state, thus obviating computational procedures based on trial and error, as, for example, invention of the calculus obviated more indirect and clumsy methods of finding areas under curves.

The problem of the lack of invariance and segmentation, however, is not specific to speech, but is a general problem for all theories of perception. Consider, for example, the perception of cursive script. Letters are neither invariant forms nor discrete segments. Furthermore, efforts to design automatic handwriting recognition machines, like efforts to build speech recognition machines, have met with only limited success (Eden, 1968). Similarly, the laws of visual form of the Gestalt psychologists (e.g. Koffka, 1935) were formulated to deal with the anomalous relation between unsegmented visual stimuli and segmented perceptions. Thus, the problem of lack of invariance and segmentation is so ubiquitous that for the purposes of this discussion we will assume that it will remain a problem.

An implicit central assumption of those theories of speech perception grounded in acoustic phonetics and built to account for the invariance-segmentation problem in one way or another (Liberman et al, 1967; Stevens,

1960, 1972; Wickelgren, 1967) is, given normal conditions, that phonetic information is perfectly recoverable from the acoustic signal.

Wickelgren (1967) is most explicit regarding this assumption in his discussion of the number of context-sensitive allophones necessary for his theory to account for the recognition of all phonemes in all contexts. Halwes & Jenkins (1971) are also explicit in their attack on Wickelgren's theory as inadequate on the grounds that when coarticulation effects are taken into account there could not possibly be enough context-sensitive allophones for perfect speech recognition.

Liberman et al (1967; also Liberman, 1975; Studdert-Kennedy, 1974) concern themselves with the problems of speech perception in the absence of invariance and segmentation and in the presence of coarticulation effects. They assume implicitly that phoneme recognition is essentially perfect where these phenomena occur. In fact this assumption becomes a corner-stone of their model.

Stevens (1960, 1972), on the other hand, explicitly allows error within his speech recognition devices - he handles the invariance and segmentation problems by generating a series of successive approximations with decreasing error. However, by the time the analysis-by-synthesis process generates phonological outputs, error is at a minimum. Thus, though Stevens allows for the possibility of an occasional error, minimum error obviously means near-perfect perception.

Thus the above theories all contain cumbersome assumptions and hypotheses designed to create invariant segments from the continuously varying acoustic speech signal. The following model, however, assumes

that only those cues which tend to carry invariant or consistent phonetic information are extracted from the preliminarily processed acoustic signal. These cues are then used to automatically activate higher level linguistic structures - phonological, lexical, syntactic and semantic.

When one hears the sentence "Apples grow on _____," there is no articulatory knowledge, or even acoustic signal necessary to entertain a strong expectation that the next phoneme is /t/ and that it will be co-articulated with the consonant /r/ and the vowel /i/ and the consonant /z/ (Miller, Heise & Lichten, 1951; also cf. Morton & Long, 1976).

Of course, in running speech the information from the first three words in the acoustic signal would not be fully present itself. "Apples" would itself be activated from such acoustic cues as; 1) initial steady state formants, 2) short silent interval, 3) rapid formant transitions, 4) slower formant transitions, 5) wide band noise with periodic component, and so forth. Other words might also be activated by these acoustic cues, but it is presumed that when cues for "grow" are processed this would effectively eliminate other alternatives for the first word - in the same sense that 'canary' activates 'yellow' (Collins & Quillian, 1969; also cf. Collins & Loftus, 1975). Such higher level structures may then be used as feed-forward in the phoneme recognition process.

An experiment by Shockey & Reddy (1974), in which phonetically trained listeners were required to provide phonetic transcriptions of recorded spoken sentences in unfamiliar languages, yielded a successful phoneme recognition rate of 56%. This figure may, then, represent a

minimum value for pure acoustic transmission of phonetic information in running speech in the absence of higher level information - not only lexical, syntactic and semantic information, but also information concerning language-specific acoustic-phonetic relations (cf. Lisker & Abramson, 1964) and phonological rules (e.g. Day, 1968). However, the correct feature class was identified 61% of the time.

Klatt & Stevens (1973) similarly attempted to estimate the "raw" phonetic transmission of acoustic information in a spectrogram reading experiment. They read their spectrograms under a 300 msec window so that lexical and other higher level hypotheses would not tend to influence their phonetic decisions. A correct phoneme recognition rate of 33%, and a correct partial feature specification rate of 40% was achieved. These rates were considerably lower than the Shockey & Reddy (1974) rates. This was probably due to the visual recognition component of the task, for visual recognition rates of the Shockey & Reddy (1974) utterances were also low (23% correct phoneme transmission rate; 38% correct feature class). However, Klatt & Stevens (1973) were able to achieve a correct word recognition rate of 96% using a computerized lexical search, their own natural semantic and syntactic intuitions, and the verification of these higher level hypotheses against the original spectrographic data. Thus one may speculate that the 56% rate of Shockey & Reddy (1974) - itself a low estimate - would be more than enough information for unambiguous speech recognition.

The present hypothesis owes much to the word recognition theory of John Morton (1968) which postulates internal recognition units called

logogens. Logogens passively accept information from sensory analysis and context relevant to a word. Properties derived from sensory analysis serve to increase the activity rate of the logogen, while contextual properties act by lowering a threshold. When this threshold is crossed either because activity is increased or because it is lowered, the response which the logogen represents becomes available to consciousness.

Thus, one is normally not aware of formulating and testing hypotheses regarding acoustic-linguistic relations - to the extent that internal linguistic structures are sensitive to the relevant criteria of the stimulus and the context they are automatically activated and brought to awareness. Only where multiple structures are activated above threshold levels is attention deployed in the selection process (cf. Posner & Snyder, 1975). Normally the most probable structure, given the acoustic input and most current working hypothesis, is automatically activated and selected.

This model is close to the theoretical position of Chomsky & Halle (1968) in that, "The hearer makes use of certain cues and certain expectations to determine the syntactic structure and semantic content of an utterance. Given a hypothesis as to its syntactic structure.... he uses phonological principles that he controls to determine a phonetic shape. The hypothesis will then be accepted if it is not too radically at variance with the acoustic material.... What the hearer "hears" is what is internally generated by the rules.... We take it for granted, then, that phonetic representations describe a <u>perceptual</u> reality.... There is nothing to suggest that these phonetic representations also describe a physical or acoustic reality <u>in any detail</u>." (Chomsky & Halle, 1968, p. 24-25: emphasis added).

The present formulation contrasts with the above only in that it proposes that "activated structure" be substituted for "hypothesis" and that prior learning and/or physiologically determined plans, should take the place of subjective control of phonological principles. That is, while both theories advocate a heterarchical processing scheme, the present model wishes to make a strong distinction between attentive and automatic processing, while Chomsky & Halle (1968) use terms which are ambiguous with respect to this distinction.

The central and crucial problems of lack of invariance and segmentation in the acoustic signal no longer become problems - which phonetic segments are present and where their boundaries exist is determined by internal structures activated by acoustic cues and context. A definite acoustic event, such as the presence of fricative noise might act as a segmentation cue not only for itself but for other neighboring phonemes.

One study has demonstrated the interaction of multiple levels of analysis in the perception of running speech. Marslen-Wilson (1975) constructed sentences with semantic, syntactic and phonetic violations and asked subjects to shadow them. His dependent variable was the number of restorations - the number of times subjects normalized these violations while shadowing - as a function of level of violation and syllable of phonetic violation.

Word restorations, for example changing the nonsense word "tomorane" into the real word "tomorrow", increased as a function of syllable of violation, from first to third, when no higher-level violations were present. However such violations decreased in frequency when syntactic and semantic

violations were also present. Context restoration errors, where an appropriate word was supplied when a non-appropriate word was present in the stimulus sentence, increased with the level of violation. Such findings are not predicted by a strictly hierarchical model in which phonological processing, for example, must be completed before lexical recognition may take place. Instead, the findings support a model whereby "the listener analyzes the incoming information at all available levels of analysis such that information at each level can constrain and guide simultaneous processing at other levels" (Marslen-Wilson, 1975, p. 227).

Despite Marslen-Wilson's (1975) assumption of simultaneity of interaction at all levels, the present formulation must assume that speech-related acoustic analysis is logically prior to other stages and the ultimate site of resolution of ambiguity.[8] The fact is that variations in the acoustic signal can cause changes in the phonetic percept. Let us assume, then, that once speech-related acoustic analysis is proceeding, it interacts heterarchically with all other stages (cf. Klatt & Stevens, 1973).

Repp (1976a,b) has made an attempt to model such speech-related acoustic processing. "The basic assumption of the 'prototype model'.... is that auditory information enters a pattern recognition process which consists in comparing the stimulus with internal 'ideal' representations of the relevant speech sounds.... (and) one is selected which matches the input most closely.... Each prototype will be 'activated' to a degree that

---

[8]Marslen-Wilson has now abandoned assumptions of parallel processing and simultaneity (personal communication, 1976).

is an inverse function of the distance separating it from the stimulus, and a subsequent decision process selects the prototype with the highest activation level as the response" (Repp, 1976a, p. 462).

The prototype model was originally designed to quantitatively handle specific data from dichotic fusion experiments. The assumption of acoustic prototypicality is essential if the present model is to account for the fact that variations in the acoustic signal produce contingent variations in the perceived phoneme. However, this assumption alone fails to account for findings that the identification of phonemes excised from context changes from when they are heard in context (Fujimura & Ochiai, 1963; Focht, 1963). The notion of prototypicality fits well within a passive speech recognition system. For the purposes of the present discussion, though, we may assume that prototypes are strongly activated by only those acoustic cues which tend to carry phonetically consistent information, and that the rest of the information of speech is carried upon higher linguistic levels.

In summary, the results of the present experiments indicate that attention is not deployed in the processing of phonetic feature information - contrary to the manner in which "articulatory knowledge" theories are construed in the Introduction. Some speculations are advanced concerning plausible theories of speech perception which are based upon processes generally thought to be automatic. Further productive research suggested by such automatic theories includes continuing the search for acoustic-phonetic invariants and studies of the interactions among acoustic cues, phonetic context and higher level linguistic expectations.

What we attend when we listen to speech

William James (1890, ch. 9) was able to discover by introspection that attention was allocated to clauses or sentences. Specifically, his introspections led him to the conclusion that attention builds up within clauses to a maximum at the end of the clause (cf. Posner, Lewis & Conrad, 1972).

Seventy years later Ladefoged and Broadbent (1960) introduced a paradigm that eventually led to the same conclusion. Their paradigm employed the detection of a short burst of noise, perceived as a click, embedded within other auditory material as a sort of "secondary task" to measure the units of perceptual processing of the "primary" auditory message. The first finding with this paradigm was that estimation of the location of a click within a string of auditory events was more accurate for strings of digits than for sentences (ibid). Therefore, by the logic of the dual task paradigm (Kerr, 1973), one may assume that sentences required more attention to process than strings of digits.

Garrett, Bever and Fodor (1966) investigated attention to syntactic processing with an ingenious technique designed to manipulate only syntactic structure while controlling for many variables such as intonation, word frequency and transitional probability of words. They used an identical tape recorded segment of speech to which two alternative syntactic interpretations could be assigned by splicing it to tape recordings of different initial words to form different sentences. When the sentence was, "As a direct result of the new invention's influence the company was given an award," the greatest number of clicks were perceived as being between "influence" and "the". On the other hand, when the sentence was, "The

retiring chairman whose methods still greatly influence the company was given an award," the most clicks were located between "company" and "was". It may be concluded, therefore, that more attention is allocated to processing clauses than to within-clause breaks.

A further methodological refinement was introduced to the click paradigm by Abrams & Bever (1969), who used clicks embedded in sentences as probes in an RT task. They found that latencies were longest at the end of major clauses and shortest at the beginnings of clauses. This finding was reinforced by Bever's (1968) discovery that detectability (d') for clicks was lowest at the end of major clauses. Thus, William James' (1890) introspection that attention to speech builds up during clauses has been confirmed by modern experimental psychology.

The buildup of attention during a clause appears to reflect a tendency for the sentence to act as a large chunk in STM (Miller, 1956) and for man to process linguistic materials in the largest chunks available (McNeill & Lindig, 1973). In support of this notion, Jarvella (1971) found that rote memory for an interrupted story was maintained for approximately one major clause preceding the interruption, while the ability to paraphrase a clause did not decrease as a function of the distance from the interruption.

Other attention demanding aspects of sentences uncovered by relatives of the click paradigm appear to include prosody (Wingfield & Klein, 1971), ambiguity (Foss, 1970), transitional probabilities of words (Morton & Long, 1976) and some aspects of surface structure syntax (Foss & Lynch, 1969; Hakes & Foss, 1970).

If speech is mentally encoded automatically, as are other stimuli, the attentional effects obtained for spoken sentences should also be obtained by appropriate visual presentation techniques (e.g. Forster & Ryder, 1976).

Despite the general preponderance of results suggesting automatic processing in the present studies, two effects of speech processing were obtained which indicate attentive processing.

The first is the tendency for the correct discrimination of place contrast trials to interfere with digit memory. This tendency is indicated by the feature by position interaction in the digit memory scores of Experiment I. The absence of any corresponding effects for identification tasks (Experiments I and II) tends to indicate that the effect of discrimination of dichotic CVs upon memory is not a general phenomenon associated with phoneme perception. Rather, it is a special case where the demands of the task interact with the structure of the stimuli (cf. Appendix).

The second effect is the tendency for the processing of a non-identical dichotic pair to interfere with probe RT more than a single phoneme, shown by the feature by delay interaction of the probe RT means in Experiment III. Ear information as such does not benefit the attentional system (Shiffrin, Pisoni & Castaneda-Mendez, 1974). Furthermore, the absence of feature effects on probe RT within the non-identical pairs rules out explanations of the interference based on increased acoustic or phonetic processing. Thus selection of one phoneme from more than one activated prototype is implicated as the locus of interference.

This is related to the finding of Noble, Trumbo & Fowler (1967) that response selection is an important source of interference. Further studies (e.g. Trumbo & Noble, 1970) have indicated that this finding is related to the general cognitive operation of response selection, and not necessarily to the selection of a spoken response.

From the viewpoint of cognitive psychology, then, the evidence suggests that attended processes employed for speech processing are similarly employed in non-speech cognitive operations.

APPENDIX: ON THE PERCEPTUAL FRAGILITY OF THE PLACE FEATURE

In contrast to the voicing feature, place of articulation has a fragile and variable relation to the auditory features upon which it is carried.

Voicing in initial stop consonants is cued by a complex of events including explosion energy, degree of aspiration and first formant intensity. The simplest and most direct cue to voicing is VOT, the time interval between the release burst or the start of formant transitions and the onset of periodicity in the signal (Lisker & Abramson, 1964). When VOT is short, initial stops are perceived as voiced and when VOT is long, initial stops are perceived as voiceless. Other auditory properties also may accompany the voicing distinction. For example, whereas the formant transitions of the initial voiced stop are well defined regions of periodic excitation, the initial voiceless stop will have no first formant and the initial portions of the higher formants will also be absent or weakly excited by noise (Stevens & Klatt, 1974). Nevertheless, the auditory properties accompanying the voicing distinction are relatively redundant for initial stops within a given language, though they may be experimentally dissociated (ibid). They are also relatively invariant with respect to phonemic context.

There is no such invariant relation between auditory feature and percept for place of articulation. Place is cued in initial stop consonants by the center frequency of the release burst and/or by the directions of transitions of the second and third formants (Liberman, Delattre & Cooper, 1952; Liberman, Delattre, Cooper & Gerstman, 1954). These cues are not invariant across phonemic context - they vary greatly with the following vowel. For example, Liberman et al (1952) found that a burst at certain frequencies was heard as /p/ before the steady state formants sufficient

to cue the vowels /i/ and /u/ but as /k/ before formants which cue /a/. Thus an identical burst can cue labial or velar closure depending on the vowel context.

The auditory features which cue place, formant transitions and burst frequencies, are highly susceptible to acoustic disruption. Miller & Nicely (1955) found that the place feature was most subject to perceptual errors in the presence of low signal to noise ratios and/or the absence of various ranges of spectral frequencies. Similarly, Pisoni & McNabb (1974) found voicing to be stable but place of articulation subject to increased errors as the intensity of CVs in a non-attended channel was increased.

In Experiment I it was relatively difficult to discriminate whether a dichotic CV pair was different when place of articulation was the only basis for discrimination (see Table 2.1). The argument may be made that differences in the formant transitions cueing place were generally too fragile to survive dichotic competition. Thus, those 32% of responses where correct discrimination did occur may be thought of as "mistakes", where the subject, perceiving one fused stimulus, nevertheless reported two.

The memory task data from the discrimination condition of Experiment I support the idea that a qualitative difference exists between place contrasts and voicing contrasts (see Fig. 2.3). One may speculate that the increased interference with the digit task when the co-occurring dichotic pair contrasted on place alone was due to a reanalysis or deeper analysis of the dichotic stimulus in those instances where the response was "different".

EXPERIMENT VI

Experiment VI was designed to further investigate the differences between the dichotic feature effects found in identification and discrimination tasks. Two major overt differences exist between the dichotic identification task and the same-different task. In the latter, subjects must not only respond in a different way than in the former, but are also under the false impression that some of the dichotic pairs are identical. A hybrid of the two tasks was employed to separate these two components and provide a finer analysis of the cognitive operations in dichotic speech perception. The two-ear identification paradigm was employed, but in one condition subjects were under the false impression that the stimulus tape had some identical pairs, while in the other they were told that there were no identical pairs. If the difference in the feature effects for two-ear identification and discrimination were due to only the expectation of identical pairs rather than to a different response mode, one would expect to see two different feature effects like those found in the same-different task and the identification task, respectively.

Method

Subjects

Eight volunteer subjects between 18 and 30 were recruited from among the employees of the Boston Aphasia Research Center. All were right handed native English speakers with no known auditory or neurological deficits, and all were employees of the Center. Only two subjects had not served in previous dichotic listening experiments, but all were naive to the main manipulation of this experiment.

Stimuli

The stimulus tape of 80 dichotic CVs that was used in Experiment I was again employed, though without the digit strings.

Procedure

Subjects were tested individually in a quiet room. A Teac 2340 tape recorder was used in conjunction with Superex Pro-BIV headphones which were balanced at 80 dB with a 1000 Hz calibration tone measured on a General Radio sound meter (type 1565Z).

Each subject was run for 160 trials in each of two conditions, false identity and nonidentity. These two conditions were presented on separate days. They were distinguished only by the instructions given to the subject. In the false identity condition subjects were told, "In this part of the experiment a small proportion of the dichotic pairs will contain identical initial phonemes," while in the nonidentity condition they were informed, "none of the pairs are identical." In both conditions subjects were told, "Simply tell me what phonemes you hear." Subjects were instructed to always give two responses. When indicating an identical pair, subjects were asked to repeat the phoneme they heard twice.

The order of conditions was counterbalanced between subjects. On each day subjects were run for 80 trials, the earphones were reversed and another 80 trials were run. The order of the channel-to-ear assignment was also counterbalanced between subjects. At least 10 practice trials and five warm-up trials were given on days 1 and 2 respectively. One subject was discarded when, at the start of day 2, he remarked that it sounded just like the tape he heard yesterday.

Results

Two within-subjects analyses of variance were performed on the mean percentage correct per subject per cell.

The first analysis of variance, summarized in Table A.1, was a feature by condition analysis on the proportion of trials where both ears were reported correctly. The main effect of condition was significant ($F(1,7) = 6.77$, $p < .036$) reflecting overall less accuracy on the false identity condition. The main effect of feature was significant ($F(2,14) = 4.62$, $p < .029$) as was the interaction of condition with feature ($F(2,14) = 10.48$, $p < .002$). The effect of feature was examined at each condition by one-way analyses of variance, and was significant for both (Identity; $F(2,14) = 4.99$, $p < .024$: Nonidentity; $F(2,14) = 5.66$, $p < .016$). However, these significant effects actually reflect two different feature effects. A Newman-Keuls test of multiple comparisons performed on the data of the nonidentity condition showed the usual identification feature sharing effect where voice and place matching trials are reported correctly more often than trials matching in neither feature ($t = 3.75$, $k = 3$, $p < .01$). Inspection of the data revealed that this was obviously not the case in the false identity condition, and after confirming that there was homogeneity of variance, subsequent t-tests using a pooled error term were performed comparing each feature across conditions. These t-tests revealed a significant difference between trials which contrasted on place alone in the false identity and nonidentity conditions ($t = 5.57$, $p < .001$) while there was no difference between conditions on trials that contrasted on voice alone ($t < 1.0$) or on both features ($t < 1.0$).

Table A.1

Mean percentage of trials in Experiment VI where both ears were reported correctly.

	Place Contrast	Voice Contrast	Double Contrast
Identity	33.20	53.52	38.87
Non-Identity	49.41	55.86	37.11

The second analysis was an ear by feature by condition analysis on those trials where at least one CV was correctly reported. This showed the same general pattern of results, with significant effects for condition ($F(1,7) = 5.55$, $p < .037$) feature ($F(2,14) = 10.34$, $p < .002$) and the interaction of condition with feature ($F(2,14) = 8.58$, $p < .004$). The effect of ear approached but did not reach significance ($F(1,7) = 3.55$, $p < .10$) while none of the interactions with ear approached significance ($F < 1.0$ in all cases).

A further analysis was done on the proportion of identity responses in the false identity condition by ear correct and feature relationships of the pair. A significant effect of feature ($F(2,14) = 86.46$, $p < .001$) revealed that far more identity responses were made to pairs that contrasted on place alone than to pairs that contrasted on voice alone or on both features. This was confirmed by a Newman-Keuls test ($t = 11.18$, $k = 2$, $p < .001$). Also, more identity responses were correct for the right ear than the left ($F(1,7) = 10.56$, $p < .015$), but there was no ear by feature interaction ($F < 1.0$).

Such large proportion of pairs contrasting on place elicited identity responses that these were broken down by value of voicing (voiced vs. voiceless), by place contrasts of the pair (labial-alveolar, alveolar-velar and labial-velar) and ear correct, and were subjected to further analysis of variance. These data are summarized in Table A.2. Ear again reached significance ($F(1,7) = 6.02$, $p < .044$) but did not interact with any other factor, while the main effects of place and voice, shown in Table A.2 were significant (place; $F(2,14) = 6.45$, $p < .011$: voice, $F(1,7) = 31.99$, $p < .001$) but did not interact with each other.

Table A.2

Percentage of voice sharing trials at each place contrast which elicited identity responses in Experiment VI.

	Labial-Alveolar	Alveolar-Velar	Labial-Velar
Voiced			
Left Ear	21.09	29.69	29.69
Right Ear	21.09	42.19	32.81
Voiceless			
Left Ear	3.13	21.88	4.69
Right Ear	5.47	28.13	17.19

Thus, the main manipulation of this experiment had a dramatic effect on pairs that matched on voicing, markedly decreasing correct performance when identity responses were allowed. This was reflected in the large number of identity responses elicited by trials that contrasted on place and especially on trials in which both CVs were voiced, rather than voiceless.

Discussion

Though the manipulation of allowing identity responses did not result in an identification feature effect that resembled the effect in the same-different task, it did markedly change the configuration of the feature effect. This was due primarily to the fact that trials contrasting on place of articulation alone elicited a great number of identity responses. Apparently the acoustic cues for place of articulation were too fragile to withstand dichotic competition, making dichotic fusion much more likely (Repp, 1976a,b).

Greenberg & Jenkins (1964), in a task where subjects were asked to judge the subjective similarity of two successively presented CVs, found that agreement in place of articulation was approximately equivalent to agreement in voicing as a factor influencing subjective similarity. The probability of eliciting an identity response may be thought of as a measure of auditory similarity in the present experiment, but here agreement in voicing is a much more potent cue than agreement in place of articulation. The difference here is due to simultaneous presentation, which tends to mask differences in the place feature, making it more likely that pairs contrasting on place alone will be called identical.

Within those pairs which share voicing, Greenberg & Jenkins found voiced pairs judged more similar than voiceless pairs, and the present data agree well with this finding. The findings for specific feature values of place of articulation are again at odds, with Greenberg & Jenkins reporting that labial-alveolar and alveolar-velar contrasts are more similar than labial-velar contrasts, reflecting actual distances in physical place of articulation (cf. Cooper, 1974a), while the present data indicate that alveolar-velar contrasts are most similar. Here it must be pointed out that Greenberg & Jenkins' subjects were instructed in an ambiguous way so that their ratings could reflect either perceptual or physical articulatory distance. That is, in contrast to the articulatory cues for voicing, those for place (i.e. the positions of the articulators) are relatively accessible to consciousness and require less sophisticated and automatized temporal control. The present study utilizes a measure of phoneme similarity which is largely a measure of acoustic distance. In contrast to Greenberg & Jenkins (1964), this experiment is one case where perception goes with acoustics, not articulation.

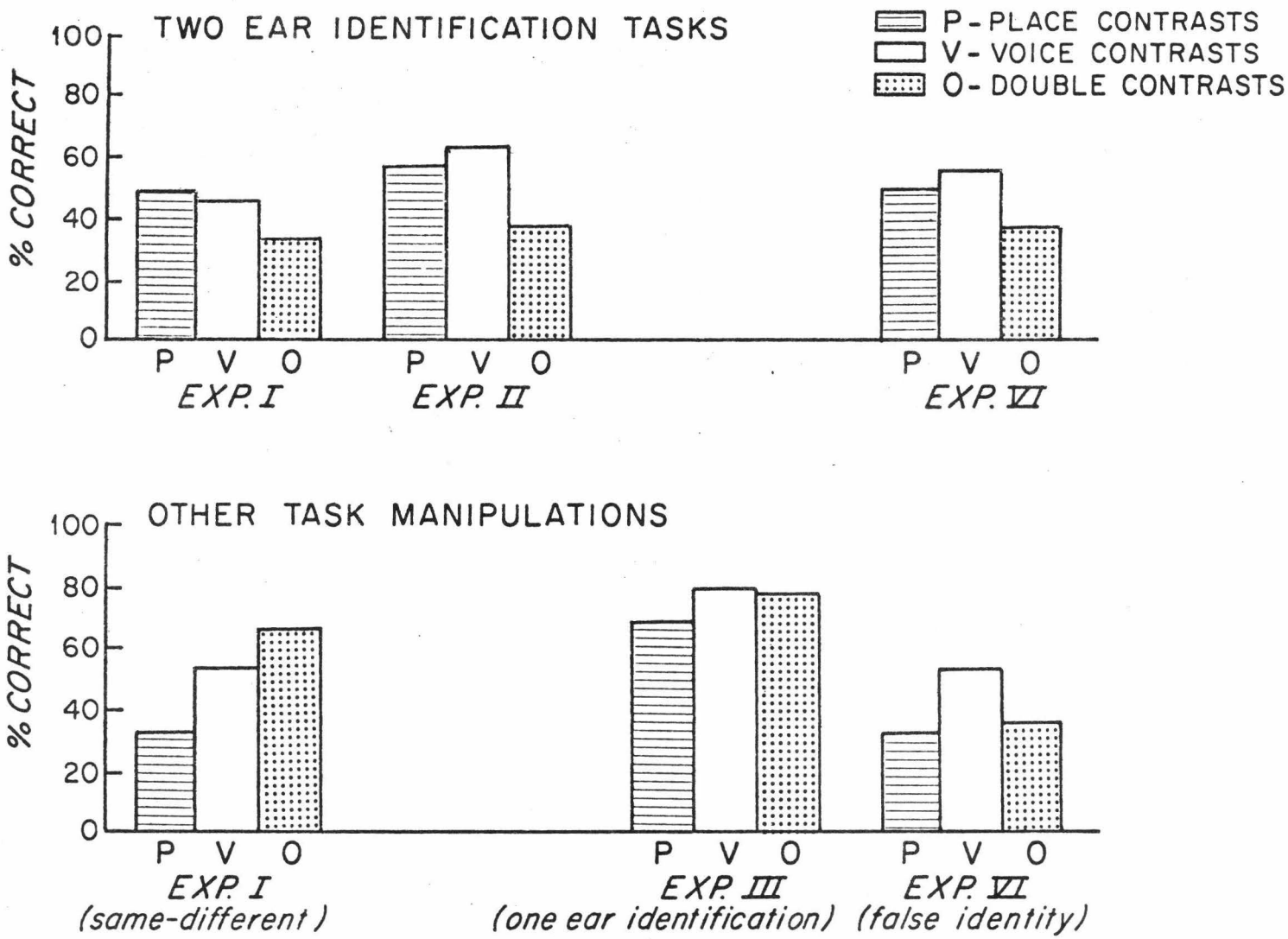
An Overview of Dichotic Feature Sharing Effects

An examination of the dichotic feature effect data collected throughout the present experiments (see Figure A.1) indicates that where voicing alone contrasts, performance is more stable across experiments and task manipulations than when a pair contrasts only on place. The only exception to this generality is Experiment III, where different tokens of the stop consonants and a smaller 'vocabulary' were used. The acoustic cue to a

Figure A.1

Synopsis of dichotic feature effects obtained in Experiments I, II, III and VI.

FIGURE A.1 SYNOPSIS OF FEATURE EFFECTS



voicing contrast in a dichotic pair of stop CVs is quite robust - the onset of periodicity in one ear followed some 30-70 msec later by the onset of periodicity in the other ear. It has long been known that the auditory system is highly sensitive to interaural delays between acoustically similar stimuli of fractions of milliseconds (Woodworth & Schlosberg, 1954, p. 349-361). The much longer delay in onset of periodicity in the present experiments is exceedingly noticeable. Since the transition contours of the formants which cue place of articulation are identical or nearly identical, and the subject knows that both values of voicing are present in two stimuli, it is an easy matter to know which two responses are correct (cf. Pisoni & McNabb, 1974, for a similar argument).

Contrasts on place of articulation, on the other hand, are subject to disruption when the task is not a two-ear identification task. A plausible reason for this is indicated in Experiment VI - subjects are likely to perceive pairs that contrast in place alone as a single speech sound. This speculation is reinforced by observations of several workers in the field (Halwes, 1969; Cutting, 1976; Repp, 1976a,b) that pairs which share the same value of voicing are more apt to give rise to fusion.

The advantage accruing to place contrast pairs in two ear identification tasks may be considered to arise in one of two ways. Place information is abstracted from a fused percept and either 1) a guess between the other two remaining values of place is executed, which is correct 50% of the time, or 2) more information is available than that which is present in the actual percept, and this information, in the form of a partially activated prototype, guides the second response in a probabilistic fashion. The first

alternative would follow from the auditory averaging hypothesis of Cutting (1976) while the later alternative is based on the multicategorical model of Repp (1976a,b). For a comparison of the two models see Repp (1976a).

The present data fit the first alternative well. In those experiments where a direct comparison is possible (Experiments I and VI) place contrasts in the two-ear identification conditions are almost exactly 50% higher than in the other conditions. However, no definitive post-hoc method of eliminating alternative two is available.

The instability of double contrast pairs across type of task is relatively easy to explain. In the discrimination condition of Experiment I, voicing contrasts or place contrasts alone carry enough information to cue correct performance. When both types of contrasts are present in the same pair, performance improves due to redundancy. Conversely, in the false identity condition of Experiment VI, as well as in the two-ear identification conditions of the other experiments, voicing and place information for both members of the pair is necessary for correct identification, and when redundancy is not present for one of the feature values, identification performance suffers. Thus, redundancy for one task is information for the other task.

These studies suggest that dichotic feature sharing effects are not due to one unitary mechanism or process. It appears that the abstraction of differences in place of articulation is more influenced by attentional factors and task demands. Furthermore, high error rates for place contrast trials, both in the same-different task and in the false identity condition of Experiment VI, suggest that strategies of guessing from incomplete information might play a large role.

In contrast, perception of voicing differences is relatively more stable and automatic. Pisoni & McNabb (1974) likewise found voicing to be stable while place was subject to increased errors in a one-ear identification task as the intensity of a CV in the nonattended channel was increased. On the basis of this independent evidence they also conclude that multiple processes operate to yield feature effects in dichotic listening (see also Pisoni, 1975).

In summary, it appears that voicing and place of articulation are processed through different types of mechanisms in dichotic listening. Some of these processes are induced by various aspects of the dichotic listening technique itself. The question of feature processing in the natural perception of running speech remains largely open.

REFERENCES

Abrams, K. & Bever, T. G. Syntactic structure modifies attention during speech perception and recognition. Quarterly Journal of Experimental Psychology, 1969, 21, 280-290.

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

Beller, H. K. Parallel and serial stages in matching. Journal of Experimental Psychology, 1970, 84, 213-219.

Berlucchi, G., Heron, W., Hyman, R., Rizzolatti, G. & Umilta, C. Simple reaction times of ipsilateral and contralateral hands to lateralized visual stimuli. Brain, 1971, 94, 419-430.

Bever, T. G. A survey of some recent work in psycholinguistics. In W. J. Plath (Ed.), Specification and utilization of a transformational grammar: Scientific report number three. Yorktown Heights, N.Y.: Thomas J. Watson Research Center, I.B.M. Corp., 1968.

Bever, T. G., Lackner, J. R. & Stolz, W. Transitional probability is not a general mechanism for the segmentation of speech. Journal of Experimental Psychology, 1969, 79, 387-394.

Bjork, R. A. Short-term storage: the ordered output of a central processor. In F. Restle, R. Shiffrin, N. Castellan, H. Lindman and D. Pisoni (Eds.), Cognitive Theory, Vol. I. Hillsdale, N.J.: Erlbaum, 1975.

Blumstein, S. E. The use and theoretical implications of the dichotic technique for investigating distinctive features. Brain and Language, 1974, 4, 337-350.

Blumstein, S. E., Baker, E. & Goodglass, H. Phonological factors in auditory comprehension in aphasia. Neuropsychologia, 1977, 15, 19-30.

Blumstein, S. E. & Cooper, W. E. Identification versus discrimination of distinctive features in speech perception. Quarterly Journal of Experimental Psychology, 1972, 24, 207-214.

Broadbent, D. E. Perception and Communication. New York: Pergamon, 1958.

Broadbent, D. E. Decision and Stress. New York: Academic Press, 1971.

Cherry, E. C. Some experiments on the recognition of speech with one and two ears. Journal of the Acoustical Society of America, 1953, 25, 975-979.

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

Chomsky, N. & Miller, G. A. 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: Wiley, 1963.

Cohen, G. Hemispheric differences in a letter classification task. Perception and Psychophysics, 1972, 11, 139-142.

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

Collins, A. M. & Loftus, E. F. A spreading-activation theory of semantic processing. Psychological Review, 1975, 82, 407-428.

Collins, A. M. & Quillian, M. R. Retrieval time from semantic memory. Journal of Verbal Learning and Verbal Behavior, 1969, 8, 240-248.

Comstock, E. M. Processing capacity in a letter matching task. Journal of Experimental Psychology, 1973, 100, 63-72.

Conrad, C. Cognitive economy in semantic memory. Journal of Experimental Psychology, 1972, 92, 149-154.

Conrad, R. Acoustic confusions in immediate memory. British Journal of Psychology, 1964, 55, 75-84.

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

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

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

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

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

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

Cooper, W. E., Blumstein, S. E. & Nigro, G. Articulatory effects on speech perception: A preliminary report. Journal of Phonetics, 1975, 3, 87-98.

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.

Corteen, R. S. & Wood, B. Autonomic responses to shock associated words in an unattended channel. Journal of Experimental Psychology, 1972, 97, 303-313.

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

Cutting, J. E. & Pisoni, D. B. An information-processing approach to speech perception. In J. F. Kavanagh and W. Strange (Eds.), Implications of Basic Speech and Language Research for the School and Clinic. Cambridge, Mass.: MIT Press, in press, 1975.

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

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

Eden, M. Handwriting generation and recognition. In P.A. Kolers and M. Eden (Eds.), Recognizing Patterns. MIT Press: Cambridge, Mass., 1968.

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

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

Ells, J. G. Attentional requirements of movement control. Unpublished doctoral dissertation, University of Oregon, 1969.

Fant, C. G. M. A note on vocal tract size factors and nonuniform F-pattern scalings. Quarterly Progress and Status Report (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, Sweden) 1966, QPSR-4.

Festinger, L., Ono, H., Burnham, C. A. & Bamber, D. Efference and the conscious experience of perception. Journal of Experimental Psychological Monograph, 1967, 74 (4), Whole No. 637.

Focht, L. R. Human recognition of sustained phonemes. Journal of the Acoustical Society of America, 1963, 35, 1890 (J10).

Fodor, J. A., Bever, T. G. & Garrett, M. F. The Psychology of Language. New York: McGraw-Hill, 1974 (Chapter 6).

Forster, K. I. & Ryder, L. Perceiving the structure and meaning of sentences. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 285-296.

Foss, D. J. & Swinney, D. A. On the psychological reality of the phoneme: Perception, identification, and consciousness. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 246-257.

Fujimura, O. & Ochiai, K. Vowel identification and phonetic contexts. Journal of the Acoustical Society of America, 1963, 35, 1889 (J4).

Garrett, M. F., Bever, T. G. & Fodor, J. A. The active use of grammar in speech perception. Perception and Psychophysics, 1966, 1, 30-32.

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

Greenberg, J. & Jenkins, J. Studies in the psychological correlates to the sound system of American English. Word, 1964, 20, 157-177.

Hakes, D. T. & Foss, D. J. Decision processes during sentence comprehension: Effects of surface structure reconsidered. Perception and Psychophysics, 1971, 10, 229-232.

Halwes, T. G. Effects of dichotic fusion on the perception of speech. Unpublished doctoral dissertation, University of Minnesota, 1969.

Halwes, T. & Jenkins, J. J. Problem of serial order in behavior is not resolved by context-sensitive associative memory models. Psychological Review, 1971, 78, 122-129.

James, W. Principles of Psychology. Vol. 1. New York: Holt, 1890.

Jakobson, R. Selected writings. I. The Hague: Mouton,1962.

Jakobson,R., Fant, G. & Halle, M. Preliminaries to speech analysis. Technical Report No. 13, Acoustics Laboratory, MIT, May, 1952.

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

Jarvella, R. Syntactic processing of connected speech. Journal of Verbal Learning and Verbal Behavior, 1971, 10, 409-416.

Kahneman, D. Attention and Effort. New York: Prentice-Hall, 1973.

Keele, S. W. Attention demands of memory retrieval. Journal of Experimental Psychology, 1972, 93, 245-248.

Keele, S. W. Attention and Human Performance. Pacific Palisades, California: Goodyear, 1973.

Kerr, B. Processing demands during mental operations. Memory and Cognition, 1973, 1, 401-412.

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

Kinsbourne, M. The control of attention by interaction between the cerebral hemispheres. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.

Klatt, D. H. & Stevens, K. N. On the automatic recognition of continuous speech: Implications from a spectrogram-reading experiment. IEEE Transactions on Audio and Electroacoustics, 1973, AU-21, 210-217.

Klein, R. M. & Posner, M. I. Attention to visual and kinesthetic components of skills. Brain Research, 1974, 71, 401-411.

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

Koffka, K. Principles of Gestalt Psychology. New York: Harcourt, Brace, 1935.

Kozhevnikov, V. A. & Chistovich, L. A. Rech' Artikuliatsia i vospriiatie, (Moscow-Leningrad). Transl. as Speech: Articulation and perception. (Washington, D.C.: Clearing house for federal scientific and technical information) JPRS, 1965, 30, 543.

Ladefoged, P. & Broadbent, D. E. Perception of sequence in auditory events. Quarterly Journal of Experimental Psychology, 1960, 13, 162-170.

Lewis, J. Semantic processing of unattended messages using dichotic listening. Journal of Experimental Psychology, 1970, 85, 225-228.

Liberman, A. M. Some results of research on speech perception. Journal of the Acoustical Society of America, 1957, 29, 117-123.

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

Liberman, A. M. How abstract must a motor theory of speech perception be? Status Report on Speech Research, 1975, SR-44, 1-15.

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

Liberman, A. M., Delattre, P. C., Cooper, F. S. & Gerstmann, L. J. The role of consonant-vowel transitions in the perception of stop and nasal consonants. Psychological Monographs, 1954, 68 (8), Whole No. 379.

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

Lindblom, B. E. F. Spectrographic study of vowel reduction. Journal of the Acoustical Society of America, 1963, 35, 1773-1781.

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

MacKay, D. G. Aspects of the theory of comprehension, memory and attention. Quarterly Journal of Experimental Psychology, 1973, 25, 22-40.

MacKay, D. M. Cerebral organization and the conscious control of action. In Eccles, J.C. (Ed.), Brain and Conscious Experience. New York: Springer-Verlag, 1965.

Marslen-Wilson, W. D. Sentence perception as an interactive parallel process. Science, 1975, 289, 226-228.

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

McNeill, D. & Lindig, L. The perceptual reality of phonemes, syllables, words and sentences. Journal of Verbal Learning and Verbal Behavior, 1973, 12, 419-430.

Mermelstein, P. Automatic segmentation of speech into syllabic units. Journal of the Acoustical Society of America, 1975, 58, 880-883.

Miller, G. A. The magical number seven, plus or minus two, or, some limits on our capacity for processing information. Psychological Review, 1956, 63, 81-96.

Miller, G. A., Galanter, E. H. & Pribram, K. H. Plans and the Structure of Behavior. New York: Holt, 1960.

Miller, G. A., Heise, G. A. & Lichten, W. The intelligibility of speech as a function of the context of the test materials. Journal of the Acoustical Society of America, 1951, 41, 329-335.

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

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

Moray, N. Attention: Selective processing in vision and hearing. London: Hutchinson Educational LTD, 1969.

Morton, J. Interaction of information in word recognition. Psychological Review, 1969, 76, 165-178.

Morton, J. & Long, J. Effect of word transitional probability on phoneme identification. Journal of Verbal Learning and Verbal Behavior, 1976, 15, 43-51.

Neisser, U. Cognitive Psychology. New York: Appleton-Century-Crofts, 1967.

Noble, M., Trumbo, D. & Fowler, F. Further evidence on secondary task interference in tracking. Journal of Experimental Psychology, 1967, 73, 146-149.

Oscar-Berman, M., Zurif, E. B. & Blumstein, S. E. Effects of unilateral brain damage on the processing of speech sounds. Brain and Language, 1975, 2, 345-3 55.

Paivio, A. Mental imagery in associative learning and memory. Psychological Review, 1969, 76, 241-263.

Pisoni, D. B. On the nature of categorical perception of speech sounds. Unpublished doctoral dissertation, University of Michigan, 1971.

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

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

Pisoni, D. B. and Tash, J. B. Reaction times to comparisons within and across phonetic categories. Perception and Psychophysics, 1974, 15, 285-290.

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

Port, D. K. & Preston, M. S. Early apical stop production: A voice onset time analysis. Status Reports on Speech Perception, 1972, SR-29/30, 125-149.

Posner, M. I. Abstraction and the process of recognition. In G. Bower and J. T. Spence (Eds.), Advances in Learning and Motivation, Vol. III. New York: Academic Press, 1969.

Posner, M. I. & Boies, S. J. Components of attention. Psychological Review, 1971, 78, 391-408.

Posner, M. I. & Klein, R. M. On the functions of consciousness. In S. Kornblum (Ed.), Attention and Performance IV. New York: Academic Press, 1973.

Posner, M. I., Lewis, J. & Conrad, C. Component processes in reading: a performance analysis. In J.F. Kavanaugh and I.G. Mattingly (Eds.), Language by Eye and by Ear. Cambridge, Mass.: MIT Press, 1972.

Posner, M. I., Nissen, M. J. & Klein, R. M. Visual dominance: An information-processing account of its origins and significance. Psychological Review, 1976, 83, 157-171.

Posner, M. I. & Snyder, C. R. R. Attention and cognitive control. In R. Solso (Ed.), Information Processing and Cognition: The Loyola Symposium. Potomac, Md.: Erlbaum Associates, 1975.

Repp, B. H. Identification of dichotic fusions. Journal of the Acoustical Society of America, 1976(a), 60, 456-469.

Repp, B. H. Discrimination of dichotic fusions. Status Report on Speech Research, 1976b, SR45/46, 123-139.

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

Sawusch, J. R. The structure and flow of information in speech perception: Evidence from selective adaptation of stop consonants. Unpublished doctoral dissertation, Indiana University, 1976.

Sawusch, J. R. & Pisoni, D. B. On the identification of place and voicing features in synthetic stop consonants. Journal of Phonetics, 1974, 2, 181-194.

Shepard, R. N. Psychological representation of speech sounds. In E.E. David and P.B. Denes, (Eds.), Human Communication: A Unified View. New York, McGraw-Hill, 1972.

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

Shiffrin, R. M., Pisoni, D. B. & Castaneda-Mendez, K. Is attention shared between the ears? Cognitive Psychology, 1974, 6, 190-215.

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

Shulman, H. G. & Greenberg, S. N. Perceptual deficit due to division of attention between memory and perception. Journal of Experimental Psychology, 1971, 88, 171-176.

Singh, S. Cross-language study of perceptual confusions of plosive phonemes in two conditions of distortion. Journal of the Acoustical Society of America, 1966, 40, 635-656.

Sperry, R. W. Neural basis of the spontaneous optokinetic response produced by visual inversion. Journal of Comparative and Physiological Psychology, 1950, 43, 482-489.

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

Stevens, K. N. Segments, features and analysis by synthesis. In J.F. Kavanaugh and I.G. Mattingly (Eds.), Language by Eye and by Ear: The Relationships between Speech and Reading. Cambridge, Mass.: MIT Press, 1972.

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

Studdert-Kennedy, M. Speech perception. In N.J. Lass (Ed.), Contemporary Issues in Experimental Phonetics. Springfield, Ill.: C.C. Thomas, 1975.

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

Studdert-Kennedy, M., Shankweiler, D. P. & Pisoni, D. B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. Cognitive Psychology, 1972, 2, 455-466.

Treon, M. A. Fricative and plosive perception - Identification as a function of phonetic context in CVCVC utterances. Language and Speech, 1970, 13, 54-64.

Treisman, A. M. Effect of irrelevant material on the efficiency of selective listening. American Journal of Psychology, 1964, 77, 533-546.

Trumbo, D. & Noble, M. Secondary task effect on serial verbal learning. Journal of Experimental Psychology, 1970, 85, 418-424.

von Holzt, E. Relations between the central nervous system and the peripheral organs. British Journal of Animal Behaviour, 1954, 2, 89-94.

Wardlaw, K. A. & Kroll, N. E. A. Autonomic responses to shock associated words in a non-attended message: A failure to replicate. Journal of Experimental Psychology: Human Perception and Performance, 1976, 2, 357-360.

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

White, M. J. Laterality differences in perception: A review. Psychological Bulletin, 1969, 72, 387-405.

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.

Wickelgren, W. A. Context-sensitive coding, associative memory, and serial order in speech behavior. Psychological Review, 1969, 76, 1-15.

Wingfield, A. & Klein, J. F. Syntactic structure and acoustic pattern in speech perception. Perception and Psychophysics, 1971, 9, 23-25.

Woodworth, R. S. & Schlosberg, H. Experimental Psychology. New York: Holt, 1954.

Zeeman, E. C. Catastrophe Theory. Scientific American, 1976, 234(4), 65-83.