RESEARCH ON SPEECH PERCEPTION

Progress Report No. 1

October 1973 - September 1974

David B. Pisoni
Principal Investigator

Department of Psychology

Indiana University

Bloomington, Indiana 47401

Supported by:
Department of Health, Education and Welfare
U. S. Public Health Service
National Institute of Mental Health
Grant No. MH-24027-01

Contents

INTRODUCTION

This is the first report of research on speech perception at the Department of Psychology, Indiana University. The main purpose of this report is to summarize activities over the past year and make them readily available to interested colleagues. Some of the papers in this report are extended manuscripts that have been prepared for publication. Other papers are reports of research that have been presented at professional meetings or brief progress reports of ongoing research. We would be most grateful for copies of reprints, preprints and progress reports dealing with related research.

Correspondence should be directed to:

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

EXTENDED MANUSCRIPTS

Dichotic Listening and Processing Phonetic Features

David B. Pisoni

Indiana University

Chapter to appear in F. Restle, R.M. Shiffrin, N.J. Castellan, H. Lindman and D.B. Pisoni (eds.), Cognitive Theory: Volume I. Potomac, Maryland: Erlbaum Associates, 1975 (In Press).

Dichotic Listening and Processing Phonetic Features

David B. Pisoni
Indiana University

For over a hundred years it has been known that the left hemisphere of man is specialized for various types of linguistic processes. Evidence supporting this view has come from a variety of sources including both clinical and experimental studies of normal and brain damaged subjects (Geschwind, 1970; Milner, 1971). However, only within the last decade have investigators begun to identify some of the stages and operations that underlie this asymmetric representation of language in the brain (Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy, 1974a,b; Wood, 1973).

Some of the strongest support for specialized neural processes in normal subjects has been obtained in dichotic listening experiments (Kimura, 1961, 1967; Shankweiler & Studdert-Kennedy, 1967; Studdert-Kennedy & Shankweiler, 1970). In this paradigm, pairs of stimuli are presented simultaneously to right and left ears and listeners are asked to identify, discriminate or recall these sounds. Depending on the types of stimuli employed, two main findings have been repeatedly observed. First, if the pairs of stimuli are linguistic such as words, digits or syllables, subjects usually report the stimulus presented to the right ear more accurately than the stimulus presented to the left ear (Bartz, Satz, Fennell & Lally, 1967; Kimura, 1961; Shankweiler & Studdert-Kennedy, 1967). Secondly, if the pairs of stimuli are non-linguistic such as melodies, tones, sonar signals or environmental sounds, the opposite effect is observed, namely, subjects report the left ear stimulus more accurately than the right ear stimulus (Kimura, 1964; Shankweiler, 1966; Curry, 1967).

Most investigators have assumed that the right ear advantage (REA) for linguistic stimuli is a reflection of the general asymmetry of cerebral dominance for language function (Kimura, 1961, 1967; Bryden, 1967; Studdert-Kennedy & Shankweiler, 1970). Explanations of the REA have generally been as follows. First, it is assumed that there is a functional prepotency of the contralateral auditory pathways from right ear to left hemisphere. This is supported by physiological evidence which indicates that the contribution of the contralateral pathways is greater than the ipsilateral pathways (Rosenzweig, 1951; Bocca, Calearo, Cassinari and Migliavacca, 1955). Second, under dichotic stimulation of the left ear signal undergoes a relatively greater "loss" than the right ear signal because it must first travel to the right hemisphere before it is transmitted to the left hemisphere via the corpus callosum. There is also evidence that the ipsilateral pathways are occluded or inhibited during dichotic stimulation (Milner, Taylor & Sperry, 1968). However, at the present time the exact locus of the REA still remains unspecified. It could occur immediately before, during or immediately after the interface between auditory processing and initial phonetic analysis. Studdert-Kennedy and Shankweiler have further argued that the right ear advantage observed under dichotic stimulation reflects the operation of a "specialized" speech processor in the language dominant hemisphere and is not simply due to additional auditory processing capacities. They claim that both cerebral hemispheres are capable of processing the auditory information in the speech signal but only the language dominant hemisphere is involved in the identification and recognition of phonetic features in the stimuli.

Support for the notion of a unilateral phonetic processor in the language dominant hemisphere rests on several general findings about the relations between speech and language function (see Mattingly & Liberman, 1969; Wood, Goff & Day, 1971; Liberman, 1972; Wood, 1973). However, most of the experimental evidence to date deals primarily with the types of interactions that have been observed between left- and right-ear dichotic speech inputs. In the present chapter I consider two of these dichotic interactions in some detail--the "feature sharing advantage" and the "lag effect." Both findings are central to a number of recent theoretical efforts in speech perception and have been the focus of a great deal of recent research (Studdert-Kennedy, Shankweiler & Pisoni, 1972; Blumstein, 1974; Benson, 1974; Speaks, Gray, Miller, Rubens & Waller, 1974).

The plan of this chapter is as follows: First, I consider the distinction between auditory and phonetic stages of processing since this underlies much of the work to be described. Second, I review the feature sharing advantage and lag effect in dichotic listening experiments. Third, I present the results of several recent dichotic recognition masking experiments that have examined these types of interactions in more detail. Fourth, I propose a rough model of some of the stages involved in phonetic processing and show how the model can account for the types of feature interactions observed between dichotic speech inputs. Finally, I briefly consider the relation between the right ear advantage and the lag effect in dichotic listening.

Auditory and Phonetic Stages of Processing

Although the distinction between phonetic structure and higher levels of analysis is commonly accepted in linguistic theory, the distinction between auditory (i.e., acoustic structure) and phonetic levels of analysis has not been widely recognized. The auditory stage may be thought of as the first level of analysis between the acoustic signal and perceived message (cf. Studdert-Kennedy, 1974a). At this level the acoustic waveform is transformed (i.e., recoded) into some "time-varying" neurological pattern of events in the auditory system. Acoustic information such as spectral structure, fundamental frequency, intensity, and duration is extracted by the auditory system. All subsequent stages of analysis beyond the auditory stage of analysis are thought to be abstract and based on an analysis of these initial auditory features. The phonetic level, the second stage of analysis, is assumed to be closely related to the first stage. Here, segments and features necessary for phonetic classification are abstracted or derived from the auditory representations of the acoustic signal. At the output of this stage, the continuously varying auditory stimulus has become transformed into a sequence of discrete phonetic segments. Information about the feature specification of these phonetic segments in the form of an abstract distinctive feature matrix is then passed on to higher levels of processing for phonological and syntactic analysis.

Thus, we may think of the auditory level as that portion of the speech perception process which is "nonlinguistic." It includes processes and mechanisms that operate on speech and nonspeech signals alike. On the other hand, processes and mechanisms at the phonetic level are assumed to perform a linguistic abstraction process whereby a particular phonetic feature is identified or recognized from some configuration of auditory features (i.e., acoustic cues) in the acoustic input. The details of this process are central to all current theories of speech perception (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Stevens & House, 1972; Fant, 1973; Bondarko, et al., 1970; Studdert-Kennedy, 1974a,b; Massaro, 1972).

There is still little agreement among investigators as to exactly how auditory and phonetic features are processed during speech perception. Nevertheless, the general "lack of invariance" between the acoustic signal and segments in the linguistic message establishes that the recognition process cannot be a simple one-to-one matching of phonetic features in the long-term memory with acoustic features in the speech stimulus (Liberman et al., 1967). As a result, a number of investigators have suggested that speech sound perception may involve specialized neural mechanisms that may not be employed in the perception of other auditory signals (Liberman et al., 1967; Stevens & House, 1974).

One broad aim of dichotic listening experiments has been to provide evidence for the existence of some type of specialized speech processing mechanism (Milner, 1962; Sparks & Geschwind, 1968). Recent work employing the selective adaptation paradigm to study feature detectors in speech perception has also been aimed in this direction (see for example, Eimas & Corbit, 1973; Eimas, Cooper & Corbit, 1973; Cooper, 1974, this volume). However, a second related aim of dichotic listening has been to study the more general processes of speech and language function. Specifically, a number of recent dichotic listening experiments have been concerned with defining the stages of processing and describing the types of operations that take place at each of these stages. In this sense, dichotic listening is simply one of a number of experimental techniques that can be used to study the processing of speech sounds.

The concern in this chapter is not primarily with the nature of the right ear advantage in dichotic listening nor with its magnitude under various experimental conditions. The literature is much too extensive to even attempt to review it here coherently. Moreover, some efforts have already been made along these lines in several recent papers (see for example: Studdert-Kennedy & Shankweiler, 1970; Berlin, Lowe-Bell, Cullen, Thompson & Loovis, 1973; Berlin & McNeil, 1974). Rather, auditory and phonetic feature interactions between dichotic inputs are examined in order to begin to describe some of the stages of processing by which phonetic features are identified.

Feature Sharing Advantage

The feature sharing advantage refers to a gain in identification for dichotic pairs of consonant vowel (CV) syllables that share phonetic features (e.g., place or voicing). The effect is shown in Figure 1 which is based

---

Insert Figure 1 about here

---

on data from Studdert-Kennedy & Shankweiler (1970). The probability that both initial stop consonants will be correctly identified is greater if the two consonant segments shared the place feature (e.g., /ba/-/pa/) or the voicing feature (e.g., /ba/-/da/) than if neither feature was shared (e.g., /ba/-/ta/). This interaction was interpreted by Studdert-Kennedy and Shankweiler as providing evidence that both dichotic inputs converge on a single phonetic processing center before the extraction of phonetic features. The authors suggested that "duplication of the auditory information conveying the shared feature value gives rise to the observed advantage (Studdert-Kennedy & Shankweiler, 1970, p. 589)." This conclusion seemed reasonable at the time. Since the same vowel (i.e., /a/) was used in each syllable, auditory and phonetic features were redundant.

The context conditioned dependence of consonant cues on vowel context should be emphasized here. One of the best known facts about phonetic perception is that the acoustic cues for a particular consonant segment, especially

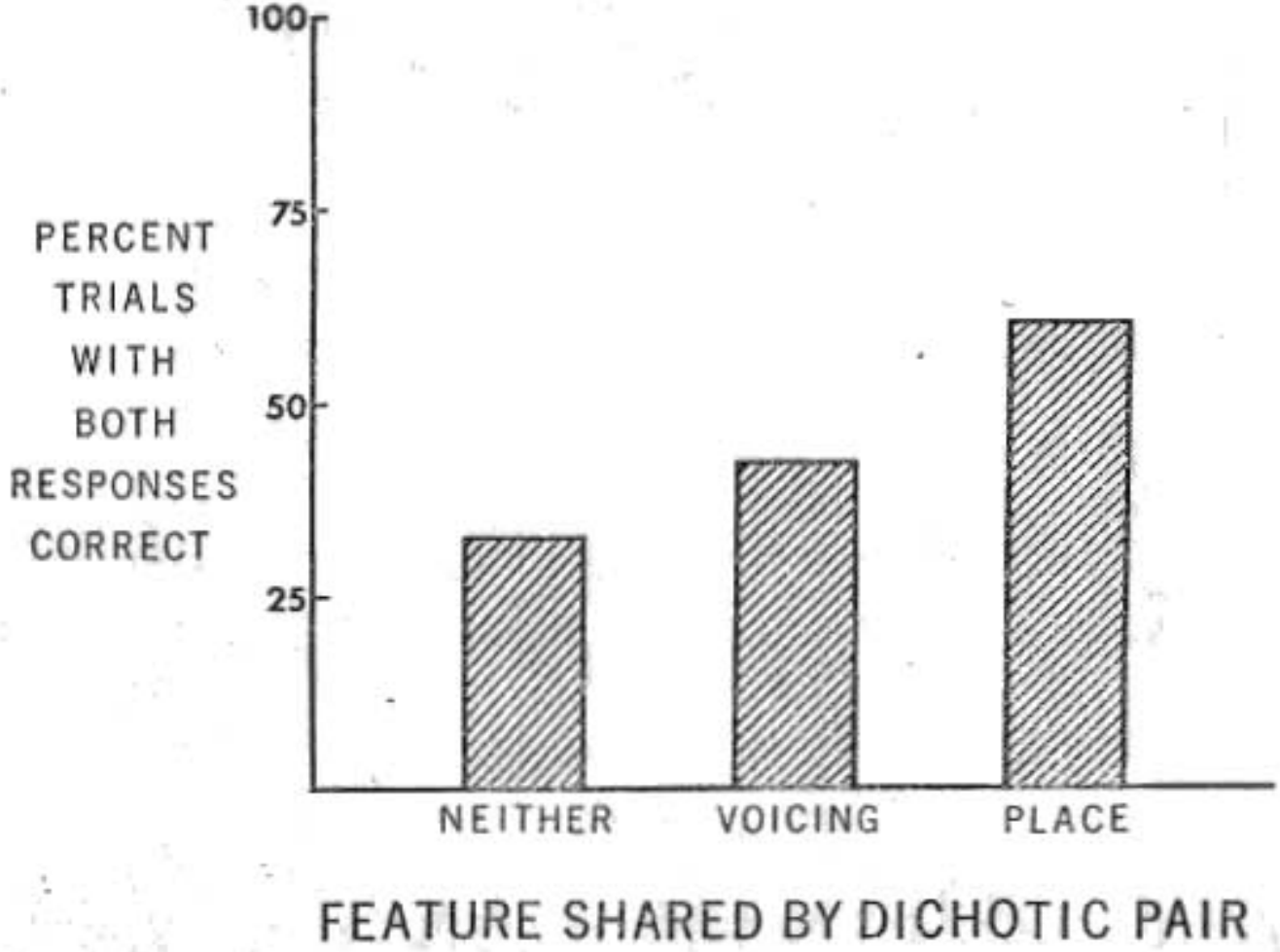


Figure 1. The percentage of trials on which both responses were correct as a function of the consonant feature shared by the dichotic CV pairs (after Studdert-Kennedy, Shankweiler & Pisoni, 1972).

stop consonants, vary as a function of vowel context, position in the syllable, stress, speaking rate and speaker.[1] Thus, when the vowel is the same, particularly with synthetic stimuli, the acoustic cues that underlie a consonant feature are also the same. The acoustic cues for a particular consonant feature vary only when vowel context or some additional parameter is manipulated. Thus, although the feature sharing advantage was originally thought to be due to commonality of the auditory features in the two inputs, the effect could also be due to shared phonetic features. To test this hypothesis we studied the feature sharing advantage under two conditions (Studdert-Kennedy, Shankweiler & Pisoni, 1972). In one condition, vowel context remained the same for both dichotic inputs, in the other condition vowel context was varied. Schematized spectrographic patterns of the stimuli which illustrate this comparison are shown in Figure 2. Eight CV syllables were formed from all possible combinations

---

Insert Figure 2 about here

---

of the four stop consonants (/b,p,d,t/) and the two vowels (/i,u/). As shown in this figure, all within column pairs (e.g., /bi-pi, bu-pu, di-ti, du-tu/) share both place of articulation (i.e., labial, alveolar) and the following vowel. These pairs have identical formant transitions and, therefore, the same auditory features underlie the phonetic feature of place of articulation. The cross-column pairs which are shown by the arrows (/bi-pu, bu-pi, di-tu, du-ti/) also share place of articulation but contrast on the vowel. Thus, these pairs have different formant transitions and, therefore, different auditory features cue the same phonetic feature. As in the earlier experiment, CV syllables that have the same vowel share both phonetic and auditory features. Pairs that contrast on the vowel shared only phonetic features. The results of that experiment replicated the earlier feature sharing results; correct performance for both stimuli is greater for dichotic pairs that share a feature in common. But of most interest was the finding that there was no effect of vowel context on correct recognition. Thus, we concluded that the feature sharing advantage was due to the shared phonetic features in the two inputs and not shared auditory features. The feature sharing advantage is assumed to have a phonetic rather than auditory basis. These results suggested to us at the time that the feature sharing advantage arises after phonetic analysis during output or response organization-- "activation of a feature processor for one response facilitates its activation for another temporally contiguous response (Studdert-Kennedy, Shankweiler & Pisoni, 1972, p. 463)."

The feature sharing advantage in dichotic listening may be considered to be a facilitatory effect at the phonetic feature level. Features in both inputs have been recognized and appear to be present in short-term memory. This idea is supported by the presence of "blend" and feature reversal errors in Ss' responses. Both types of errors occur when the features in a stimulus presented to one ear are incorrectly combined with the features in the other ear. For example, a "blend" error occurs if /ba/ and /ka/ are presented dichotically and the S reports /ga/; the voicing feature from /ba/ is combined with the place feature of /ka/ to produce a response having both component features. A feature reversal error occurs when the S reports /ba/ and /ta/ when the input stimuli were /pa/ and /da/; all the component features of the input stimuli are present in the responses but the features have been recombined incorrectly.

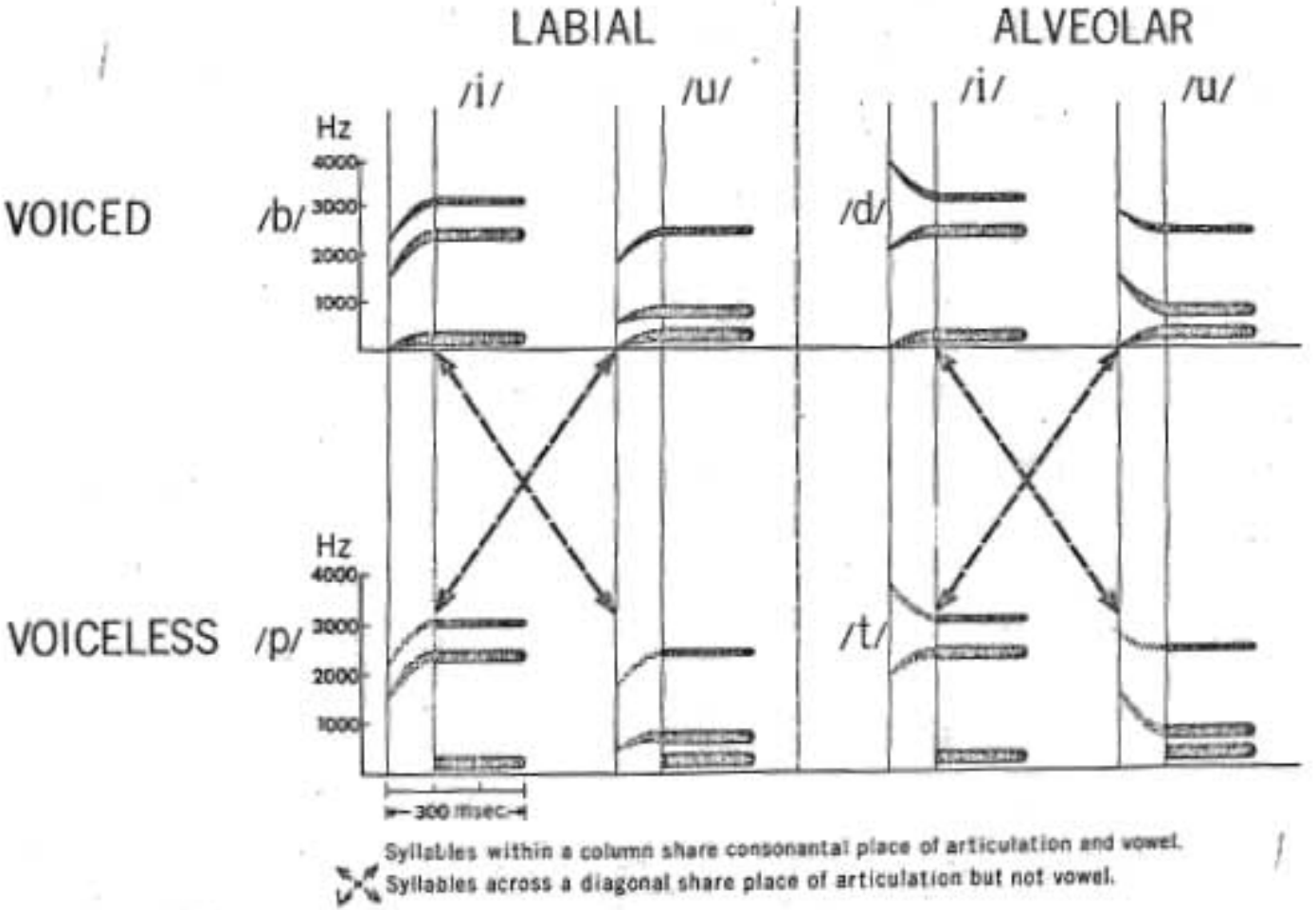


Figure 2. Schematic spectrograms of the eight synthetic CV syllables used in the feature sharing experiment (after Studdert-Kennedy, Shankweiler & Pisoni, 1972).

Theoretical interest in these types of phonetic interactions is twofold. First, they provide additional support for the idea that phonetic features are recognized more or less independently during perceptual processing. This stage of processing, however, should be distinguished from the earlier stage where auditory features are processed. Current evidence suggests that auditory features are not processed independently of each other (Holloway, 1971; Haggard, 1970; Smith, 1973; Sawusch & Pisoni, 1974). A second reason for interest in these feature interactions is that they indicate that recombination of the component features from each stimulus must have a common locus, presumably after phonetic processing in the language dominant hemisphere. Indeed, most of the support for a unilaterally represented phonetic processor rests on these types of phonetic feature interactions (Studdert-Kennedy & Shankweiler, 1970). If recombination of the component features occurred separately for each ear, there would be little possibility for the phonetic features from each ear to recombine in the form of blend and feature reversal errors.

Lag Effect

The second type of interaction to be considered is the so-called "lag effect" in dichotic listening. This effect occurs when the dichotic inputs are presented with varying temporal delays. Studdert-Kennedy, Shankweiler and Schulman (1970) reported that Ss identify the second or lagging syllable of a dichotic pair of temporally overlapping stimuli more accurately than the leading syllable. The effect is shown in Figure 3 which has been replotted from the original report. As shown here performance is better on the lagging

---

Insert Figure 3 about here

---

syllable than the leading syllable. When the same syllables were mixed and the signal presented monotically to one ear, the lag effect was reversed; the leading syllable was now reported more accurately than the lagging syllable. Studdert-Kennedy et al., (1970) originally interpreted the lag effect as a form of "interruption" of speech processing presumably occurring at a central level of perceptual analysis. They suggested that "the lag effect is tied to speech, and, specifically, to those components of the speech stream for which a relatively complex decoding operation is necessary (Studdert-Kennedy, Shankweiler & Schulman, 1970, p. 601)." Indeed, the lag effect has been used recently as evidence to support the general argument that speech perception engages specialized processes that differ from those of nonspeech auditory perception (Liberman, Mattingly, & Turvey, 1972).

The lag effect appears to be a variation of a more general result obtained in backward masking experiments: a second stimulus can impede the processing of a preceeding stimulus (Kahnemen, 1968; Massaro, 1972; Turvey, 1973). As used in the speech perception literature, the lag effect actually deals with the relative difference between forward and backward masking; there appears to be more dichotic backward masking than forward masking for CV syllables.

A number of recent dichotic experiments have shown that the lag effect may not be peculiar to speech sounds since it has been obtained with nonspeech timbres, vowels and other sounds. For example, Darwin (1971), using a directed attention paradigm, has reported a lag effect for stimuli that differ only in fundamental frequency. With $\pm$ 25 msec offsets between stimuli, listeners reported the second stimulus more often than the first. Although Porter, Shankweiler, and Leberman (1969) initially failed to obtain a lag effect for steady-state vowels, Kirstein (1971) obtained the effect with slightly different procedures. Since the lag effect has been found with speech as well as nonspeech stimuli, it seems reasonable to suppose that this type of dichotic

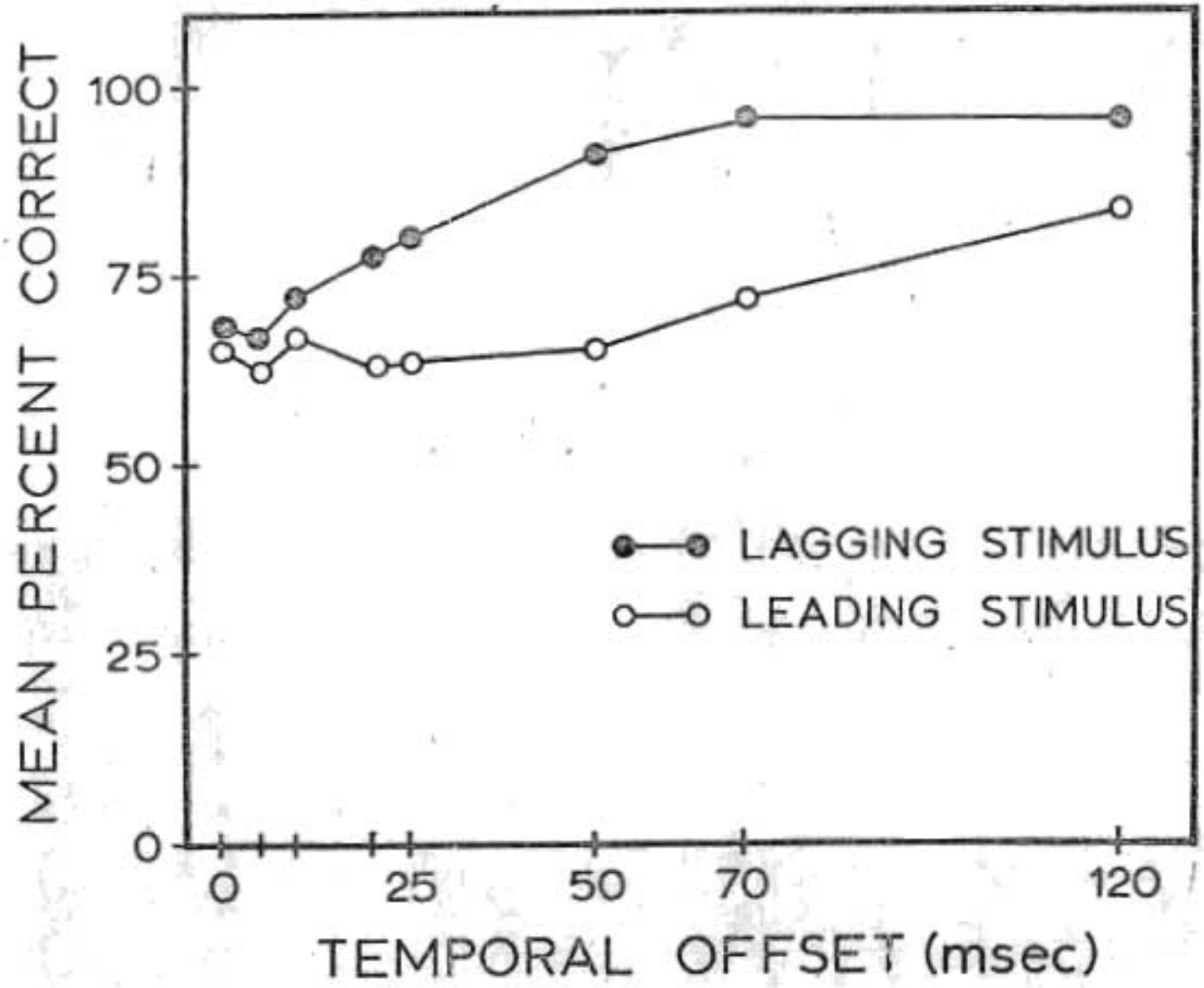


Figure 3. Mean percent correct for leading and lagging dichotically presented CV syllables based on the data from Studdert-Kennedy, Shankweiler & Schulman, 1970.

interaction has an auditory rather than phonetic basis. The interaction between the inputs may occur at the auditory feature level prior to phonetic analysis.

We may think of the lag effect as a form of interference in dichotic listening. But what is the locus and nature of this form of interference? At what stage in the information processing system does the interference arise? The dichotic recognition masking experiments to be described were aimed at these questions. If the masking that underlies the lag effect occurs at an early stage of processing prior to auditory analysis any CV syllable should interfere with the processing of a preceeding syllable. This is essentially a stop processing or interruption hypothesis. On the other hand, if the lag effect occurs after auditory analysis only certain types of stimulus contrasts should produce interference. These masking experiments indicate that only certain types of stimulus contrasts should produce interference. These masking experiments indicate that interference is not found equally for all stimulus contrasts. The greatest interference occurs on trials that contain CV syllables that do not share phonetic features. Thus, the feature sharing advantage and the lag effect provide evidence for distinct auditory and phonetic feature interactions in dichotic listening. Furthermore, these types of interactions provide some basis for formulating a rough model of the stages of processing in phonetic perception.

Dichotic Recognition Masking

The method used to study the feature sharing advantage and the lag effect was a dichotic recognition masking paradigm. Two CV syllables were presented on each trial, a target and a mask. The syllables differed in the consonant, the vowel and their relative times of onset. The subjects' task was always to identify the target stimulus in an ear monitoring paradigm and to ignore the masking stimulus. Figure 4A shows the general arrangement of

---

Insert Figure 4 about here

---

the target and masking stimuli used in the backward masking experiments. For the forward masking experiments, the configuration of target and mask was simply reversed. In backward masking, the S identified the first stimulus, in forward masking he identified the second stimulus.

With this technique the processing of a target stimulus may be probed by a masking stimulus at various stimulus onset asynchronies and thereby provide us with some information about the temporal course of perceptual processing of the target sound (Massaro, 1972, 1974). The targets and masks used in these experiments were always drawn from different stimulus ensembles as shown in Figure 4B. There were two voiced targets, /ba/ and /da/, and two voiceless targets /pa/ and /ta/. The six masks that we used were selected so that they either shared or contrasted with the auditory and phonetic feature composition of the targets. As in the previous dichotic experiment with Studdert-Kennedy and Shankweiler, the vowel context was varied in order to manipulate the auditory features which underlie a particular phonetic feature. However, the phonetic feature studied in these experiments was voicing whereas in the previous experiment the feature was place of articulation. We should note here that the place feature in stop consonants is cued primarily by rapid transitional changes in the spectrum (Liberman, Delattre, Cooper & Gerstman, 1954). On the other hand, the voicing feature is cued primarily by the timing of the onset of first formant relative to the second formant (Liberman, Delattre & Cooper, 1958).[2]

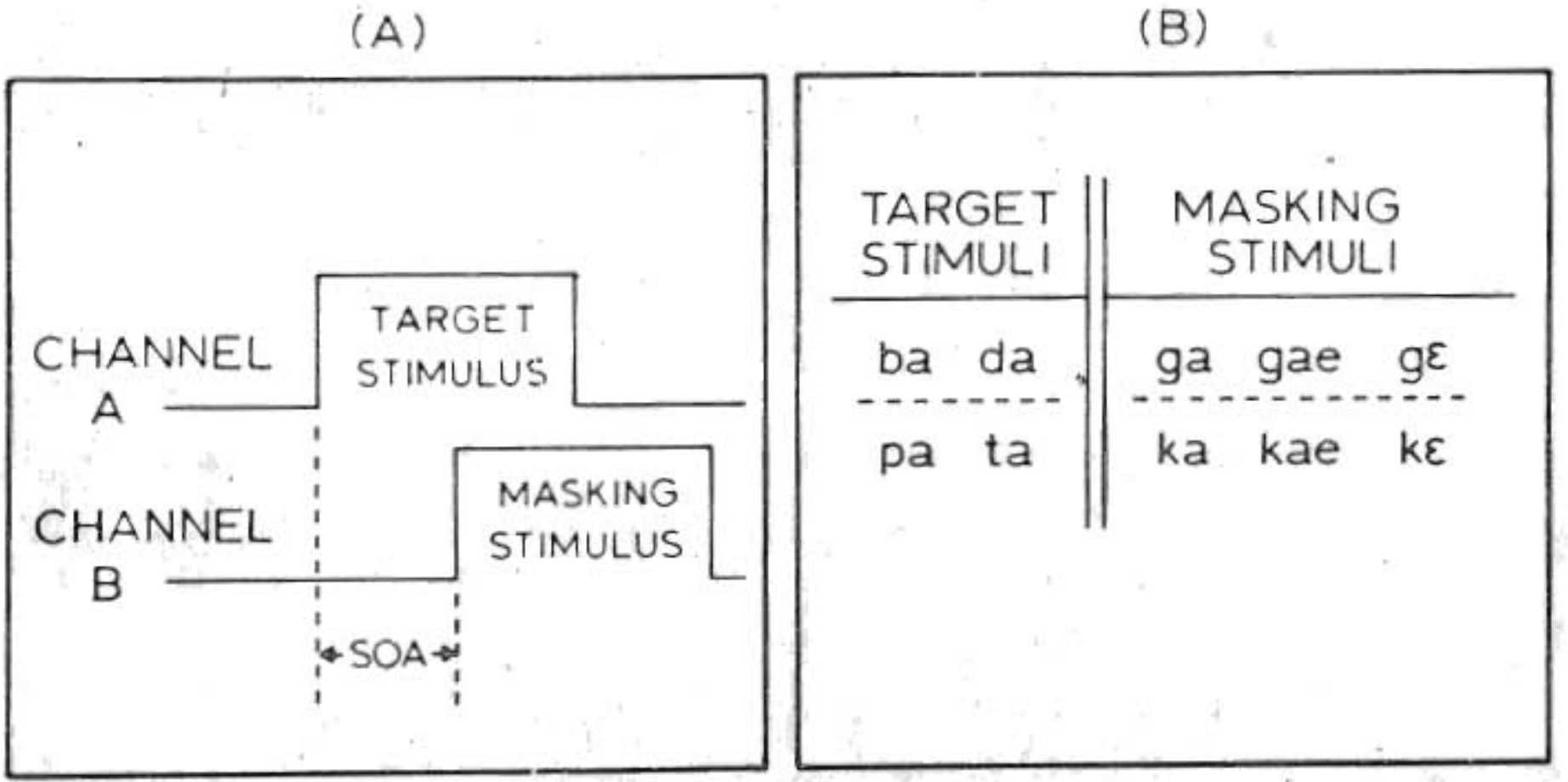


Figure 4. Arrangement of target and masking stimuli in the dichotic backward masking paradigm (Panel A) and the targets and masks (Panel B) used in the experiments. Targets and masks drawn from the same row share the feature of voicing whereas targets and masks drawn from different rows contrast on voicing (after Pisoni & McNabb, 1974).

By varying the vowel context the over-all spectral composition of the target and mask could also be manipulated. For example, the target-mask pair /ba/-/ga/ shares the voicing feature (+ voicing) and the vowel. The pair /ba/-/ga/ still shares the voicing feature but now differs in the vowel. Half of all trials in these experiments contained pairs of stimuli that shared the voicing feature; half contained pairs that contrasted on voicing.

Two comparisons are of interest here as a function of time. First, is there a difference in recognition between pairs of stimuli that share or contrast on the voicing feature? Second, what is the effect of the vowel in the mask on identification of the target? The latter comparison should permit us to specify the locus of the interactions between the dichotic inputs. For example, if the vowel context of the mask has no effect on the identification of the target, we would conclude that the interaction between the inputs occurred at the phonetic level. This would be anticipated if the consonant segments had already been abstracted from the syllables. On the other hand, if the vowel context systematically affects target identification, this would indicate that, at least, some component of the interaction occurs at an earlier stage of analysis either before or during phonetic processing.

Backward Masking. In the first experiment backward masking was examined for shared and non-shared trials as a function of stimulus onset asynchrony (SOA). The main results are shown in Figure 5 averaged over the three vowel contests.

Insert Figure 5 about here

Voiced and voiceless targets have also been combined in this figure. Performance was consistently higher for pairs that shared voicing than pairs that contrasted on voicing. Performance is relatively stable for shared pairs at all SOA values whereas performance improves steadily for non-shared pairs as SOA increased. When we scored the data for correct recognition of the voicing feature alone, performance in the shared condition was virtually perfect. For example, if /ba/ was the target and the S responded with /da/, we scored this as a correct response of the voicing feature; stimulus and response were both (+ voiced). In contrast, performance for the voicing feature on the non-shared trials remained the same as in the previous analysis of correct responses.

The effect of vowel context of the mask on shared and non-shared trials is shown separately for each of the three vowel conditions in Figure 6. The

---

Insert Figure 6 about here

---

influence of the vowel is restricted primarily to the non-shared pairs. Performance on these trials was lowest for /a/ vowel masks, highest for /ɛ/, and midway between the two for /ae/. Identification in the shared condition is consistently higher under each vowel condition than in the non-shared condition.

The main results of this experiment suggest that the feature sharing advantage and the interference obtained in the lag effect are distinct types of interactions between dichotic speech inputs, presumably occurring at different levels of analysis. Overall performance is affected by both SOA and vowel context of the mask. However, the difference between shared and non-shared trials still maintains itself under these conditions.

These results replicate and extend the previous findings on the feature sharing advantage reported by Studdert-Kennedy and Shankweiler (1970) and Studdert-Kennedy et. al., (1972). As noted earlier, these findings were interpreted as evidence that the feature sharing advantage occurred on the output side of phonetic analysis during response organization. However, in the present experiment

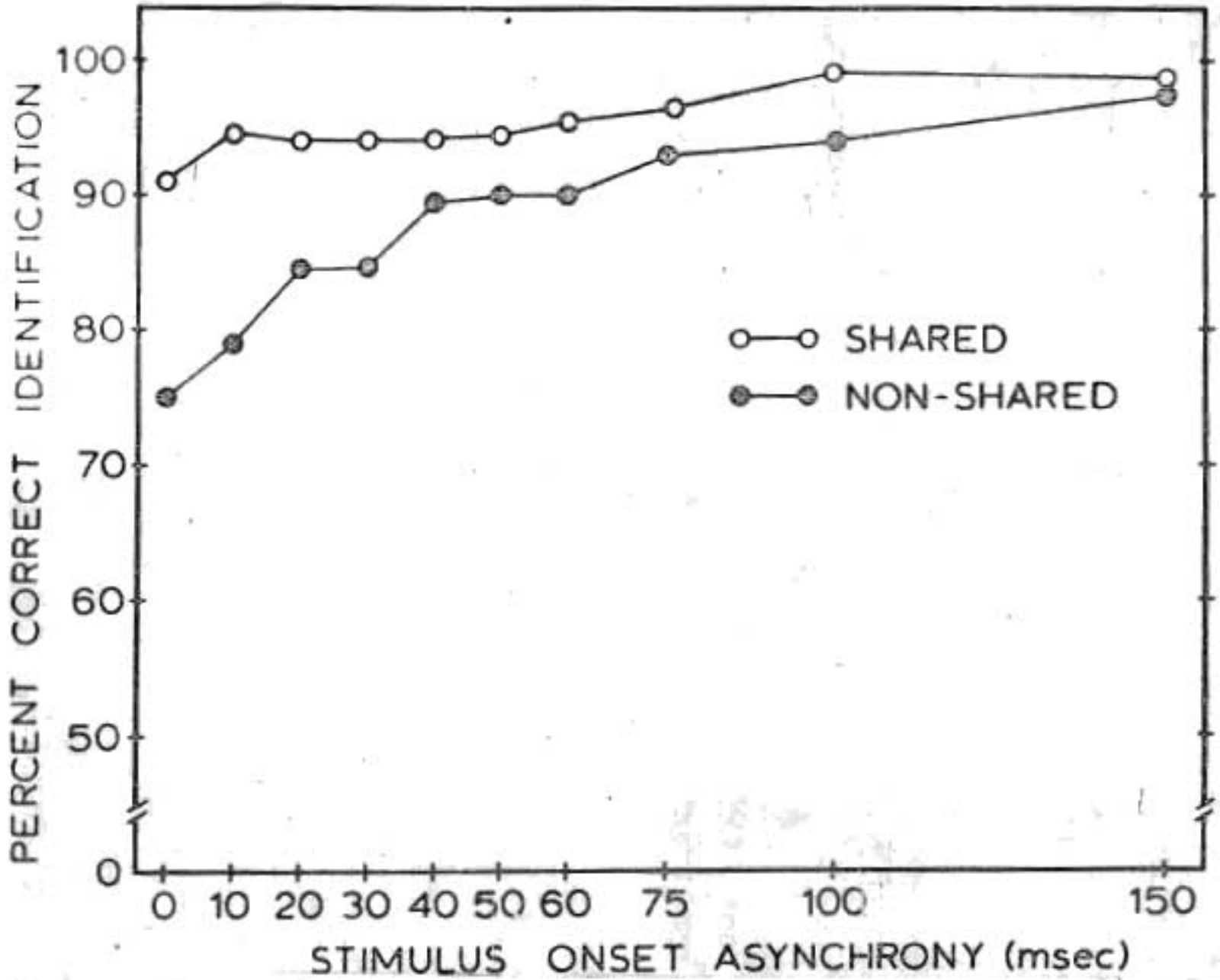


Figure 5. Percent correct identification of target stimuli for shared and non-shared trials as a function of stimulus onset asynchrony. The data are averaged over the three vowel masks (after Pisoni & McNabb, 1974).

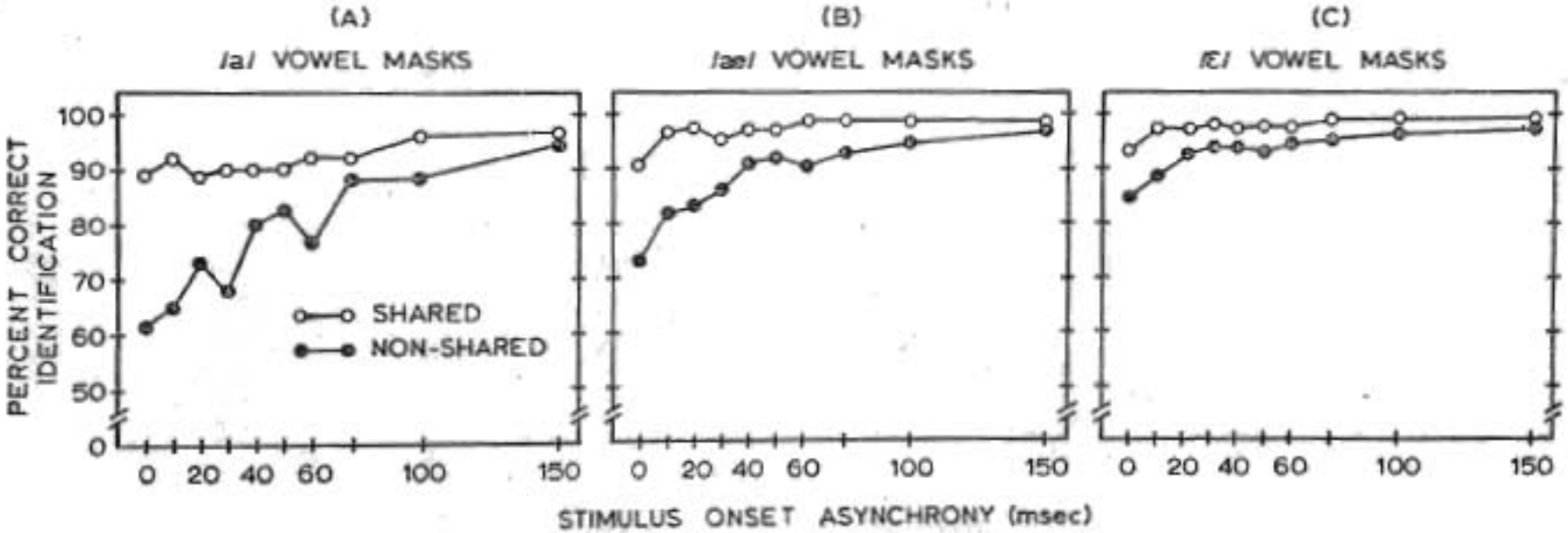


Figure 6. Percent correct identification of target stimuli for shared and non-shared trials under each vowel mask (after Pisoni & Mcnabb, 1974).

the feature sharing advantage still occurs and with considerable magnitude when only one response is required. Thus, we can infer from this result that the feature sharing advantage probably lies somewhere before response organization after the features have been identified. We will return to a more detailed account of the feature sharing advantage later on.

These results also provide some insight into the type of interaction underlying the lag effect. For non-shared trials we observe that performance increases as the interval between the onset of the target and mask is increased. Increases in SOA provide increases in processing time for recognition of the auditory features in the target stimulus. Since recognition of the target stimuli is affected systematically by the vowel context of the masking syllable, one component of the interaction must occur before phonetic analysis while the auditory features in the syllables are still being processed. If the interaction occurred after the consonant features had been abstracted from the target, the vowel context should not have affected the identification of the target. These results suggest that the locus of the interference underlying the lag effect occurs at an auditory feature level.

At first glance, the results of this experiment present somewhat of a paradox: similarity in the consonant voicing feature (i.e., voice onset time) reduces interference, similarity in the vowel increases interference. The latter effect is not difficult to understand. We have only to suppose that the more similar the vowels of the target and mask, the more likely the two syllables are to "fuse" or integrate into one perceptual unit so that the listener has difficulty assigning the correct auditory features to the appropriate stimulus (see also Cutting, 1972). This account of the vowel effect argues against a strict interruption or stop processing explanation. If the second stimulus simply terminated the readout of auditory features from the first stimulus, vowel similarity should not have had any effect on target recognition. Any speech stimulus should have terminated processing. In addition, we would not expect to find an interaction between the phonetic feature composition of the consonant targets and the vowel context of the mask. Both findings suggest an account of masking based on some form of integration at an auditory level. Auditory features from both stimuli merge together to form a composite stimulus which is then made available for subsequent phonetic analysis. Thus, variations in the degree of backward masking can be accounted for by variations in"acoustic confusability" due to overall spectral composition of target and mask. We are going to assume that the vowel effect is due to relatively low level binaural interaction in the auditory system (see Durlach & Colburn, In Press; Colburn & Durlach, In Press).

But how are we to account for the apparent lack of interference for pairs of stimuli that share the voicing feature? Before attempting an account of the absence of masking in this condition, we consider another experiment where mask intensity is manipulated. If auditory factors are the principal determinants of variations in the degree of backward masking, we would expect intensity variations to have an effect on both shared and non-shared trials as well as the variations in spectral composition. Intensity as a gross physical parameter should also have its effect at relatively early stages of processing.

We carried out another backward masking experiment where the intensity of the mask differed from the target by 0, +10, or +20dB. Figure 7 shows

---

Insert Figure 7 about here

---

the results of this experiment for shared and non-shared trials as a function of SOA for each mask intensity level. These functions are averaged over all

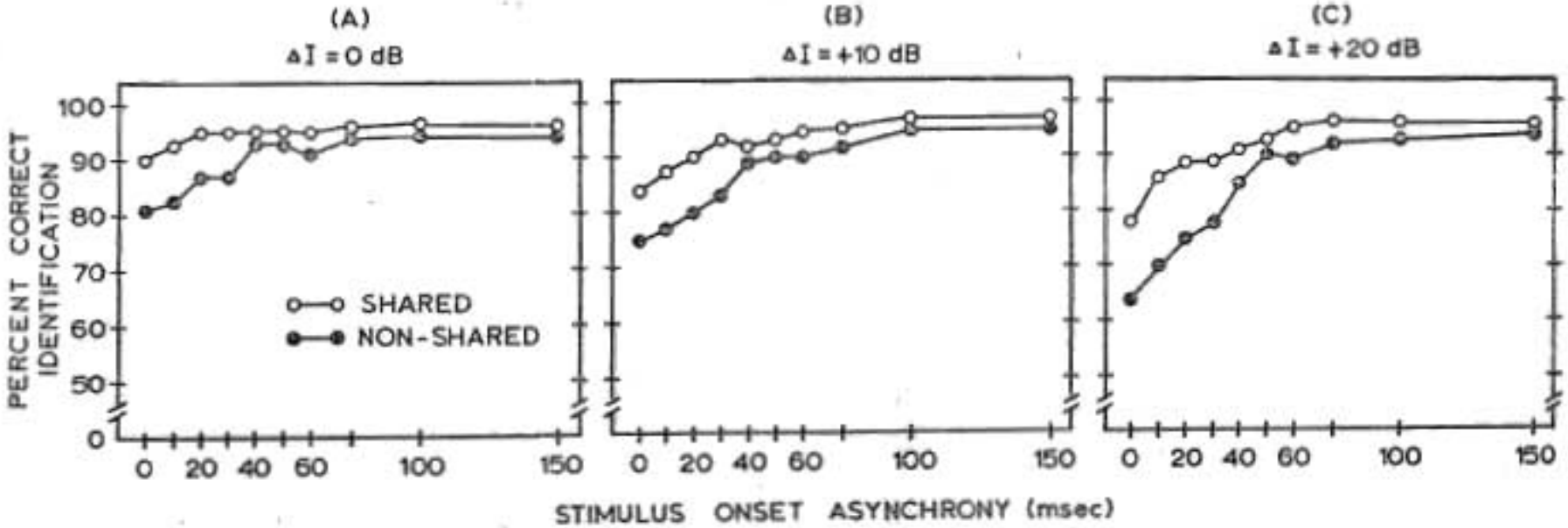


Figure 7. Percent correct identification of target stimuli for shared and non-shared trials at each of three mask intensity levels (after Pisoni & McNabb, 1974).

vowel contexts. Note that the effect of mask intensity is clearly present for both shared and non-shared trials; performance on the target systematically decreased as mask intensity increased. The difference in recognition between shared and non-shared trials is, however, still present under all three intensity conditions.

When the data were scored separately by voicing, treating a response as correct if voicing was correct, the intensity effect for the shared trials disappears. This result is shown in Figure 8 which is based on the data from

---

Insert Figure 8 about here

---

the /a/ vowel mask condition. Thus, increased mask intensity for shared pairs apparently has its main effect on the place feature which is cued by relatively rapid spectral changes during the very early portion of the syllable. In contrast, correct identification of the voicing feature for the non-shared pairs decreases systematically as mask intensity is increased.

A clue to understanding the absence of interference for shared pairs is provided by an examination of the feature errors. Table 1 displays the

---

Insert Table 1 about here

---

proportions of voicing and place feature errors for shared and non-shared trials in the conditions yielding maximum masking, namely, a +20 dB mask intensity with target and mask vowels identical. The main point to note in this table is that while place errors are roughly the same when the voicing feature is shared as when it isn't, voicing errors are sharply increased in the non-shared condition. In other words, the feature-sharing advantage is confined to the particular feature shared. The previous studies by Studdert-Kennedy and Shankweiler (1970) and Studdert-Kennedy et. al., (1972) failed to observe this because they did not score the S's response by feature by only by total response. Thus, if a S makes a voicing error on a non-shared trial, his response must contain the voicing feature of the mask. The high rate of errors on voicing is then due to the fact that the voicing feature of the mask interacts with the voicing feature of the target. This result should be emphasized since it clearly suggests that the feature sharing advantage occurs at the phonetic feature level and not earlier.

Forward Masking. The backward recognition masking results could be explained by a simple masking or interruption hypothesis (Massaro, 1972, 1974). The second stimulus terminates the read-out of auditory features from the preceeding stimulus. However, some complexities arise when we consider the case of forward recognition masking. In this experiment, Ss identify the second stimulus rather than the first. The forward masking experiment is important for several reasons. First, if the interference between target and mask were due strictly to interruption, no forward masking would be anticipated since processing time for the target is unlimited. Second, the presence of forward masking would lend additional support to the integration hypothesis outlined earlier. The target and masking stimuli merge to form a composite stimulus containing auditory features of both stimuli.

In this forward masking experiment, all stimuli and experimental conditions were identical to the first backward masking experiment described earlier except that a new group of Ss was employed. The main results are shown in Figure 9 averaged over the three vowel contexts. The difference in correct identification

---

Insert Figure 9 about here

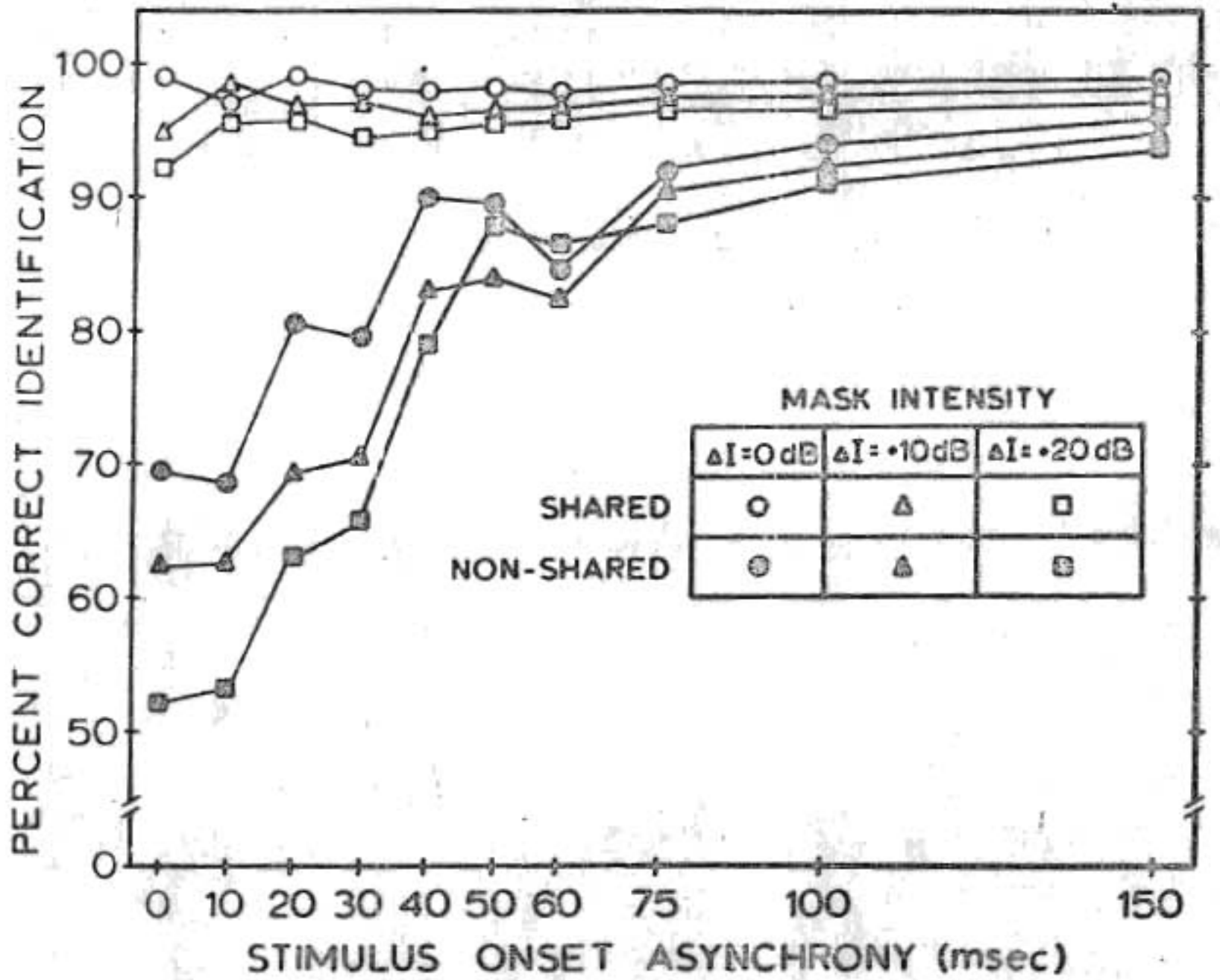


Figure 8. Percent correct identification of the voicing feature on shared and non-shared trials as a function of mask intensity from the /a/ vowel mask condition (after Pisoni & McNabb, 1974).

Table I

Proportions of voicing and place errors under voicing shared and non-shared conditions for the +20 dB /a/ vowel masks from Pisoni & McNabb, 1974.

	Feature	Voicing Shared	Voicing Non-Shared
Voicing	Voiced	.05	.31
	Voiceless	.03	.16
Place	Labial	.16	.12
	Alveolar	.03	.04

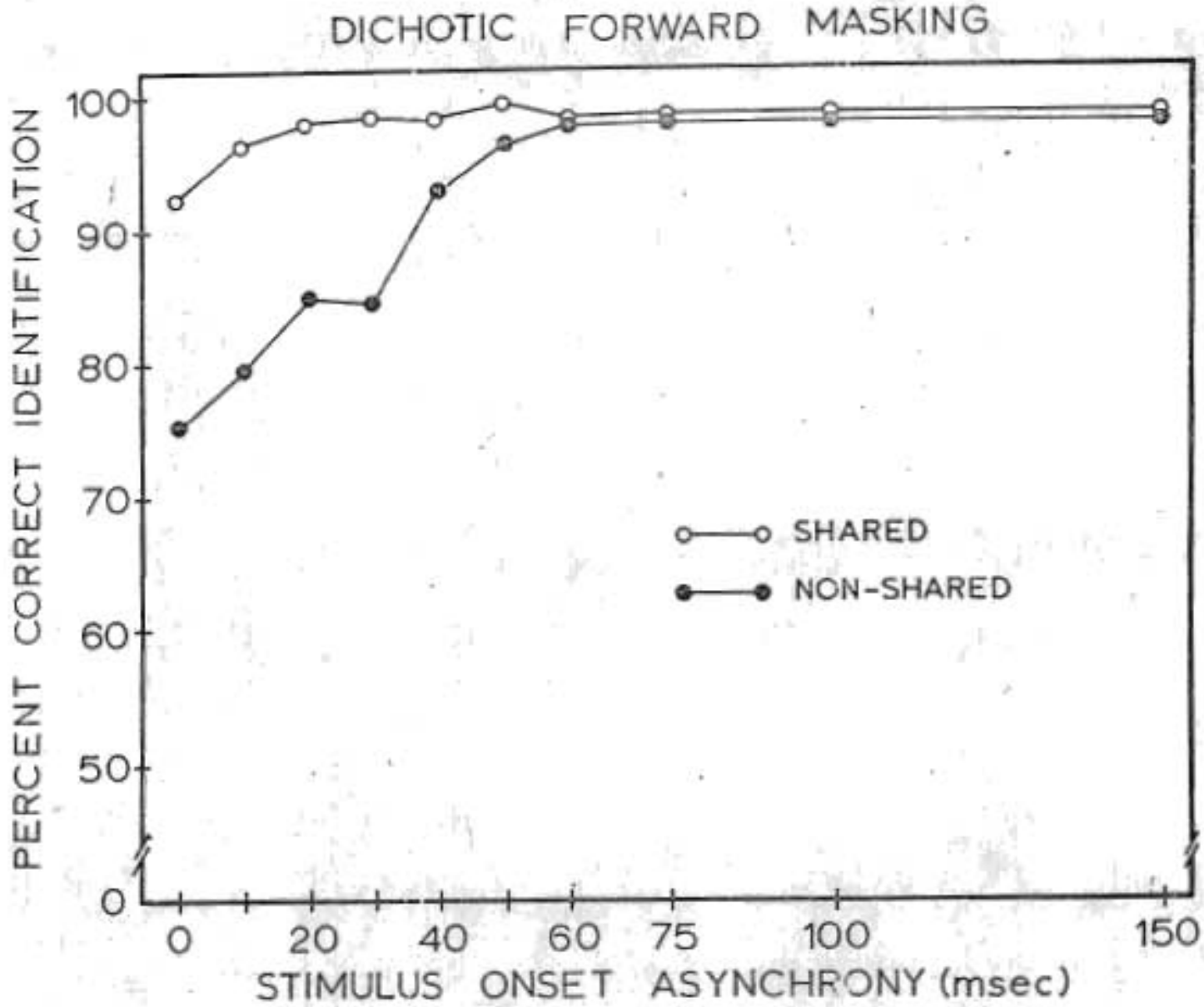


Figure 9. Percent correct identification of target stimuli in forward masking condition for shared and non-shared trials as a function of stimulus onset asynchroay.

of the targets between shared and non-shared trials is quite similar to that found in the earlier backward masking experiments. Performance improves steadily as a function of SOA for both types of trials. The effect of the vowel context is shown separately again for each vowel in Figure 10. The effect of the vowel

---

Insert Figure 10 about here

---

on target identification is remarkably similar to that found in the backward masking case; overall performance is inversely related to the spectral composition of the vowel context of the target and masking syllables.

We can summarize the results of these experiments quite simply. First, forward and backward masking functions appear to be essentially the same. Differences in relative onset of target and mask, spectral similarity, and mask intensity influence the overall level of performance for both shared and non-shared trials. Furthermore, shared and non-shared trials continue to show differences in performance under these experimental manipulations. These results suggest several stages at which dichotic speech inputs can interact. In order to describe these interactions in more detail we consider a rough model of the stages of processing in phonetic perception.

Stages of Processing

Taken together the forward and backward dichotic masking results provide some insight into the recognition process. Earlier we described the distinction between auditory and phonetic stages of processing. However, based on the present findings, this dichotomy appears to be much too gross and additional stages are required. Figure 11 shows a qualitative model of the stages of

---

Insert Figure 11 about here

---

processing involved in phonetic recognition. Auditory input first undergoes preliminary auditory analysis. The output is assumed to be some type of spectral display in terms of frequency, time, and intensity. Sensory input is then processed progressively through several levels of analysis. Processing stages have been arranged here serially only for convenience since we do not have sufficient experimental evidence to argue for parallel or serial processing between these stages at the present time (see Wood, 1974, this volume).

Acoustic Feature Analysis is the first stage of the recognition process. Here, auditory features of the speech signal are identified by a system of individual auditory feature detectors (Stevens, 1973; Cooper, this volume). For example, in the case of a simple CV syllable, we assume that specialized detectors will respond selectively to some of the following types of auditory information: (a) presence or absence of a rapid change in the spectrum, (b) direction, extent and duration of a change in the spectrum, (c) duration and intensity of noise, (d) frequency of noise segment or burst, (e) presence or absence of the fundamental frequency from the beginning of the syllable, (f) abrupt rise in the frequency of the fundamental at the transition from consonant to steady-state vowel. The output of Acoustic Feature Analysis is some set of acoustic cues or auditory features, $\{c_i\}$, which forms the input to the next stage of processing.

In Stage 2, Phonetic Feature Analysis, we assume that a set of decision rules is employed to map multiple auditory features into phonetic features. It is assumed that this is a many-to-one mapping where several different auditory features provide information about a particular phonetic feature (e.g., see Hoffman, 1958; Liberman, Delattre & Cooper, 1958). Rather than assume that a

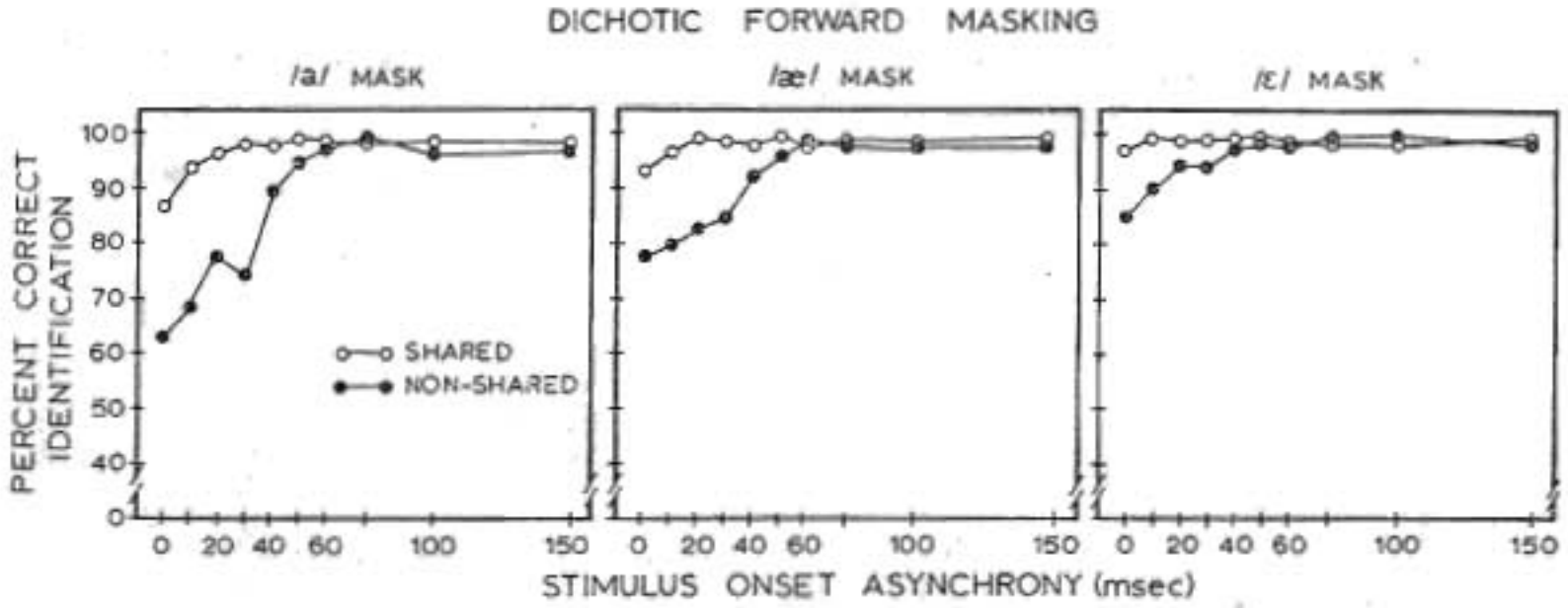


Figure 10. Percent correct identification of target stimuli in forward masking condition under each vowel condition.

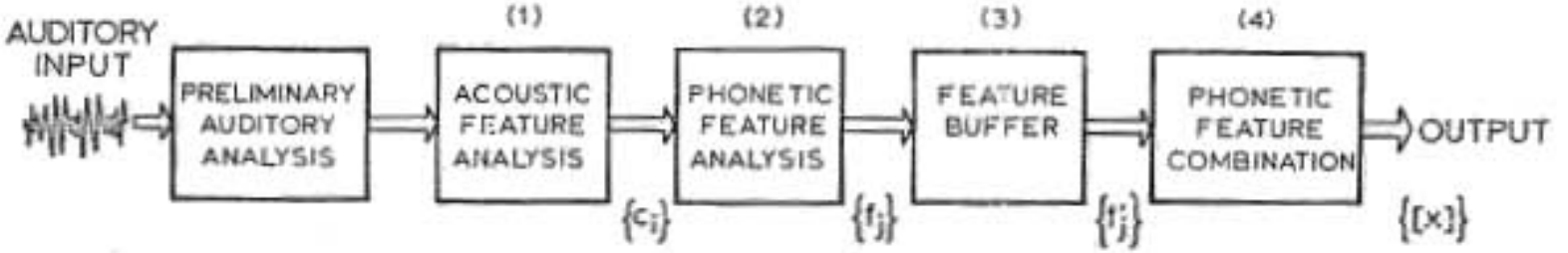


Figure 11. Stage model of levels of processing in phonetic recognition. Auditory input is processed progressively through several levels of analysis.

"phonetic processor" exists as a distinct physiological mechanism, we would prefer, at the present time, to describe its function simply in terms of decision rules. Decisions about a particular feature are based on auditory information distributed across the whole syllable (Liberman, 1970; Massaro, 1972; Studdert-Kennedy, 1974 a, b). It is at this stage that processing becomes lateralized in the language dominant hemisphere. The outputs of acoustic feature analysis, $\{c_i\}$, from each hemisphere converge for phonetic feature analysis. The output of phonetic feature analysis is a set of abstract phonetic features $\{f_j\}$.

Phonetic features are subsequently maintained in Stage 3, the Feature Buffer. This may be thought of as a holding mechanism which maintains decisions about the feature composition of a particular syllable. We distinguish the output of the feature buffer, $\{f'_j\}$, from the input, $\{f_j\}$, since information can be lost by interference or decay and confusion among features can result. There are two reasons for postulating a feature buffer. First, not all phonetic features are assumed to be processed (i.e., recognized) at the same rate. Secondly, some memory process is needed to preserve or maintain phonetic features more-or-less independently for subsequent stages of linguistic processing.

Feature information is then used in Stage 4, Phonetic Feature Combination, where individual features are recombined to form discrete phonetic segments (i.e., phonemes). The output of Stage 4 is a phonetic segment, (X), where the feature specification is, for example, some form of an abstract distinctive feature matrix. This information is passed on to higher levels of linguistic analysis (i.e., phonological).

The model as we have described it is still preliminary and a number of details remain to be worked out. However, the model can account in a qualitative way for a number of the dichotic listening results discussed so far. For example, the feature sharing advantage probably arises after Phonetic Feature Analysis in the Feature Buffer. Redundant features do not have to be maintained separately in the buffer and there is less chance of confusion. The feature reversal and blend errors described earlier probably result from confusions among features in the buffer before recombination into phonetic segments. Since these errors involve only the loss of local sign (i.e., ear of origin) it is clear that the features have been identified and they are being maintained in some form independent of context.

Forward and backward masking appears to arise before acoustic feature analysis. Since relative onset time, spectral similarity and mask intensity all effect overall performance for both shared and non-shared trials it seems safe to assume that these gross physical parameters affect processing at relatively early stages. Thus, the advantage for sharing a phonetic feature must occur relatively late in the processing sequence since the difference between these two pairs is still present regardless of large acoustic differences between the target and masking stimuli.

The Right-Ear Advantage and the Lag Effect

Throughout most of this chapter we have focused on the interactions between dichotic speech inputs and essentially ignored asymmetries between the ears. In this section we briefly deal with the right ear advantage for speech stimuli and its relation to the lag effect.

In a recent paper, Weeks (1973) has called attention to an apparent paradox between the right-ear advantage and the lag effect in dichotic listening experiments. Most investigators have assumed that the right-ear advantage is due to some loss of information from the left-ear input. Loss may result from the additional time necessary for the left-ear input to reach the dominant hemisphere since the signal must transverse a longer distance via the corpus

callosum. Weeks (1973) has called this a queueing or "delay" hypothesis. It is assumed that the ipsilateral input from the left ear arrives at the dominant hemisphere some time later than the contralateral input from the right ear. However, loss of information from the left ear may also be due to some impairment in the ipsilateral signal as a result of interhemispheric transfer. This is the currently favored explanation of the right ear advantage which has been coined the "degradation hypothesis" by Weeks. Feature extractors in the dominant hemisphere receive a poorer or more degraded signal from the ipsilateral ear.

The apparent paradox between the right-ear advantage and the lag effect is as follows. Both the delay and degradation hypotheses of the REA assume that the left-ear stimulus arrives at the dominant hemisphere some time later than the right-ear stimulus. Thus, there is an inherent temporal asymmetry and masking should occur between left- and right-ear stimuli, The left-ear stimulus should interrupt the processing of the right-ear stimulus. However, available evidence indicates that the right-ear advantage and lag effect are more or less independent of each other (Kirstein, 1970, 1971; Berlin, et. al., 1973). Thus, the interpretation of the lag effect as a form of interruption of processing through backward masking is in serious conflict with the interpretation of the right-ear advantage in terms of some inherent delay of the left-ear stimulus. This paradox can be resolved easily, however, by assuming as we have in this chapter that the lag effect results from integration of the two dichotic inputs. Thus, the two dichotic stimuli are not functionally independent of each other and therefore each hemisphere probably receives a different composite of both stimuli. The interference underlying the lag effect arises, therefore, prior to the stage at which the right-ear advantage occurs. Several experiments are currently in progress which deal with this particular problem.

Final Remarks

In this chapter I have tried to show how dichotic listening techniques can be used to study some of the more general processes in speech perception. A good part of the recent dichotic listening literature has focused on the types of auditory and phonetic interactions that occur between dichotic speech inputs. These interactions appear to occur at a number of different processing stages and provide some insight into the general organization of information processing in speech perception. However, we are a long way off from a really well-developed model of the speech perception process. Many details still need to be worked out and many of the conclusions arrived at through dichotic listening experiments will need to be evaluated in other experimental paradigms. In the future, however, we can probably expect to see an increase in the use of various types of brain-damaged subjects in speech perception experiments. These experiments should help to bridge the gap between our knowledge of underlying physiology of speech and language function and the processes we have imparted to little boxes that have appeared in such ever-increasing proclivity over the last few years.

Acknowledgments

The research reported in this paper was supported by NIMH Research Grant MH-24207-01 to Indiana University. Preparation of the manuscript was supported in part by a Faculty Fellowship from the Office of Research and Advanced Studies, Indiana University. I am very grateful to S.D. McNabb for help and assistance in all phases of this work and to M. Studdert-Kennedy for critical comments and suggestions.

Footnotes

[1]Throughout this chapter I use the terms "acoustic cue" and "auditory feature" somewhat interchangeably since there is a one-to-one mapping of acoustic cue to auditory feature.

[2]The voicing feature in the present experiments is cued by voice onset time (VOT), the temporal interval between the release of stop closure and the onset of laryngeal pulsing. Since VOT is a temporal cue, manipulating vowel context does not necessarily entail a strict independence between auditory feature and phonetic feature as was the case with the place cue.

References

Bartz, W.H., Satz, P., Fennell, E. & Lally, J.R. Meaningfulness and laterality in dichotic listening. Journal of Experimental Psychology, 1967, 73, 204-210.

Benson, P. Phonetic and auditory features in dichotic listening. Paper presented at the 87th Meeting of the Acoustical Society of America, New York City, April, 1974.

Berlin, C.I., Lowe-Bell, S.S., Cullen, J.K., Thompson, C.L., and Loovis, C.F. Dichotic Speech Perception: An Interpretation of Right-Ear Advantage and Temporal Offset Effects. Journal of the Acoustical Society of America, 1973, 53, 699-709.

Berlin, C.I. and McNeil, M.R. Dichotic Listening. In N.J. Lass (Eds.) Contemporary Issues in Experimental Phonetics. Springfield, Illinois: C.C. Thomas, 1974.

Blumstein, S. The use and theoretical implications of the dichotic technique for investigating distinctive features. Brain & Language, 1974, 1, 000-000.

Bocca, E., Calearo, C., Cassinari, V., & Migliavacca, F. Testing "Cortical" Hearing in Temporal Lobe Tumors. Acta Oto-laryngologia, 1955, 45, 289-304.

Bondarko, L.V. et al. A Model of Speech Perception in Humans. Working Papers in Linguistics No. 6. Computer & Information Science Research Center, Ohio State University, Columbus, Ohio. Technical Report 70-12, 1970.

Bryden, M.P. An evaluation of some models of laterality effects in dichotic listening. Acta Oto-laryngologica, 1967, 63, 595-604.

Colburn, H.S. & Durlach, N.I. Models of Binaural Interaction. In E.C. Carterette & M.P. Friedman (Eds.) Handbook of Perception. New York: Academic Press (In Press).

Cooper, W.E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974, 55, 000-000.

Cooper, W.E. Selective Adaptation to Speech, In F. Restle, R.M. Shiffrin N.J. Castellan, H. Lindman & D.B. Pisoni (Eds.) Cognitive Theory: Volume I. Potomac, Maryland: Erlbaum Associates, 1975, Pp. 000-000

Curry, F.K.W. A Comparison of Left-handed and Right-handed Subjects on Verbal and Non-verbal Dichotic Listening Tasks. Cortex, 1967, 3, 343-353.

Cutting, J.E. A Preliminary Report on Six Fusions in Auditory Research. Haskins Laboratories: Status Report on Speech Research, SR - 31/32, 1972, 93-107.

Darwin, C.J. Dichotic Backward Masking of Complex Sounds. Quarterly Journal of Experimental Psychology, 1971, 23, 386-392.

Durlach, N.I. & Colburn, H.S. Binaural Phenomena. In E.C. Carterette and M.P. Friedman (Eds.) Handbook of Perception. New York: Academic Press (In Press).

Eimas, P.D., Cooper, W.E. and Corbit, J.D. Some Properties of Linguistic Feature Detectors. Perception & Psychophysics, 1973, 13, No. 2, 274-252.

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

Fant, G. Speech Sounds and Features. Cambridge, The M.I.T. Press, 1973.

Geschwind, N. The organization of language and the brain. Science, 1970, 170, 940.

Haggard, M.P. The use of voicing information. Speech Synthesis & Perception, 1970, 2, 1-15, University of Cambridge, London.

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

Holloway, C.M. A test of the independence of linguistic dimensions. Language & Speech, 1971, 14, 4, 326-340.

Kahneman, D. Method, Findings, and Theory in Studies of Visual Masking. Psychological Bulletin, 1968, 70, 404-425.

Kimura, D. Some Effects of Temporal Lobe Damage on Auditory Perception. Canadian Journal of Psychology, 1961, 15, 156-165. (a)

Kimura, D. Left-right Differences in the Perception of Melodies. Quarterly Journal of Experimental Psychology, 1964, 16, 355-358.

Kimura, D. Functional Asymmetry of the Brain in Dichotic Listening. Cortex, 1967, 3, 163-178.

Kirstein, E.F. Selective Listening for Temporally Staggered Dichotiv CV Syllables. Journal of the Acoustical Society of America, 1970, 48, 95 (A).

Kirstein, E.F. Temporal Factors in Perception of Dichotically Presented Stop Consonants and Vowels. Ph.D. Dissertation, University of Connecticut, 1971.

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

Liberman, A.M. The specialization of the language hemisphere. Invited paper presented at the Intensive Study Program in the Neurosciences at Boulder, Colorado, July, 1972.

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

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

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

Liberman, A.M., Mattingly, I.G. & Turvey, M.T. Language Codes and Memory Codes. In A.W. Melton and E. Martin (Eds.), Coding Processes in Human Memory. Washington, D.C. V.H. Winston & Sons, 1972, Pp. 307-334.

Massaro, D.W. Preperceptual Images, Processing Time, and Perceptual Units in Auditory Perception. Psychological Review, 1972, 79, 2, 124-145.

Massaro, D.W. Perceptual units in speech recognition. Journal of Experimental Psychology, 1974, 102, 2, 199-208.

Mattingly, I.G. & Liberman, A.M. The Speech Code and the Physiology of Language. In K.N. Leibovic (Ed.) Information Processing in the Nervous System. New York: Springer - Verlag, 1969, Pp. 97-117.

Milner, B. Laterality effects in audition. In V.B. Mountcastle (Ed.) Interhemispheric Relations and Cerebral Dominance. Baltimore: Johns Hopkins University Press, 1962, Pp. 177-193.

Milner, B. Interhemispheric differences in the localization of psychological processes in man. British Medical Bulletin, 1971, 27, 3, 272-277.

Milner, B., Taylor, L. & Sperry, R.W. Lateralized Suppression of Dichotically - presented Digits after Commissural Section in Man. Science, 1968, 161, 184-185.

Pisoni, D.B. & McNabb, S.D. Dichotic Interactions and Phonetic Feature Processing. Brain & Language, 1974, 1, 000-000.

Porter, R.J., Shankweiler, D., Liberman, A. Differential Effects of Binaural Time Differences on Perception of Stop Consonants and Vowels. Proceedings of the 77th Annual Meeting of the American Psychological Association, Washington, D.C., 1969, Pp. 15-16.

Rosenzweig, M.R. Representations of the two ears at the auditory cortex. American Journal of Physiology, 1951, 167, 147-158.

Sawusch, J.R. & Pisoni, D.B. On the identification of place and voicing features in synthetic stop consonants. Journal of Phonetics, 1974, 2, 3, 201-214.

Shankweiler, D.P. Effects of temporal-lobe damage on perception of dichotically presented melodies. Journal of Comparative and Physiological Psychology, 1966, 62, 115-119.

Shankweiler, D. & Studdert-Kennedy, M. Identification of Consonants and Vowels Presented to Left and Right Ears. Quarterly Journal of Experimental Psychology, 1967, 19, 59-63.

Smith, P.T. Feature-testing models and their application to perception and memory for speech. Quarterly Journal of Experimental Psychology, 1973, 25, 511-534.

Sparks, R. & Geschwind, N. Dichotic listening in man after section of Neocortical Commissures. Cortex, 1968, 4, 3-16.

Speaks, C., Gray, T., Miller, J., Rubens, A. & Walker, M. Interference with processing dichotic pairs of CV syllables after temporal-lobe lesion. Paper presented at the 87th Meeting of the Acoustical Society of America, New York City, April, 1974.

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

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

Studdert-Kennedy, M. The Perception of Speech. In T.A. Sebeok (Ed.), Current Trends in Linguistics, Vol. XII. The Hague: Mouton, 1974 (a).

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

Studdert-Kennedy, M. & Shankweiler, D.P. Hemispheric Specialization for Speech Perception. Journal of the Acoustical Society of America, 1970, 48, 2, 579-594.

Studdert-Kennedy, M., Shankweiler, D. & Pisoni, D.B. Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study. Cognitive Psychology, 1972, 3, 455-466.

Studdert-Kennedy, M., Shankweiler, D.P. & Schulman, S. Opposed Effects of a Delayed Channel on Perception of Dichotically and Monotically Presented CV Syllables. Journal of the Acoustical Society of America, 1970, 48, 599-602.

Turvey, M.T. On Peripheral and Central Processes in Vision: Inferences from an Information - Processing Analysis of Masking with Patterned Stimuli. Psychological Review, 1973, 80, 1-52.

Weeks, R.A. A Speech Perception Paradox?: The Right-Ear Advantage and the Lag Effect. Haskins Laboratories Status Report on Speech Research, SR-33, 1973, 29-35.

Wood, C.C. Levels of Processing in Speech Perception: Neurophysiological and Information-processing Analyses. Unpublished doctoral dissertation. Yale University, 1973. (Also appears in Status Report on Speech Research, SR-35/36, Haskins Laboratories, New Haven.)

Wood, C.C. Parallel processing of auditory and phonetic information in speech perception. Perception & Psychophysics, 1974, 15, 000-000.

Wood, C.C. (This volume).

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

Selective Adaptation of Auditory Feature Detectors in Speech Perception

by

Jeffrey B. Tash

Submitted to the Faculty of the Graduate School in partial fulfillment of the requirements for the degree of Master of Arts in the Department of Psychology
Indiana University
August, 1974

ACKNOWLEDGEMENTS

I wish to express my deepest appreciation to Dr. Lloyd R. Peterson who served as a member of my committee, and to Dr. David B. Pisoni, chairman of the committee, whose criticisms and suggestions were invaluable in the design of this study and in the preparation of this thesis.

In addition, I would like to thank Dr. James Cutting, Yale University and Haskins Laboratories, for constructing the experimental materials used in this investigation.

I also wish to express my gratitude to Dr. David Pisoni for providing the funds (NIMH Research Grant MH 24027-01) and the facilities without which this study could never have been done.

Above all, I wish to thank my wife Lois for her invaluable contributions. Not only did she assist with many technical and clerical aspects of this thesis, but more importantly, she provided encouragement, understanding and loving toleration throughout many frustrating moments.

JBT

INTRODUCTION

It is generally well established that the conversion of a continuously varying acoustic waveform into a string of discrete phonetic segments involves a number of distinct stages or levels of perceptual analysis (Fant, 1967; Stevens and Halle, 1967; Stevens and House, 1972; Studdert-Kennedy, 1974). At the lowest level the physical signal is analyzed into a set of time-varying auditory dimensions such as timbre, pitch and loudness. This low-level auditory information is then operated upon by the next higher stage of processing for the extraction of abstract phonetic features. This distinction between an auditory and a phonetic level of speech processing has been experimentally supported by a number of different researchers (e.g., Studdert-Kennedy and Shankweiler, 1970; Studdert-Kennedy, Shankweiler and Pisoni, 1972; Wood, 1974; Pisoni and Tash, 1974). The present study intends to examine the role of this dichotomy in terms of a feature detector analysis of speech perception.

The notion of a feature detector originally comes from the electrophysiological investigations of single cell neurons. Lettvin, Maturana, McCulloch, and Pitts (1959), studying the visual system of the frog, discovered that specialized neural receptors in the frog's eye extract relatively restricted patterns of information from the visual signal. In their experiment, four classes of

feature detectors were observed: edge detectors, moving detectors, dimming detectors and convex edge detectors. Edge detectors responded whenever a border between light and dark occurred within a specific receptive field. Moving detectors were activated by the presence of a moving edge. Dimming detectors reacted to an overall decrease in illumination. And convex edge detectors responded whenever a small, dark, moving object appeared in the visual field (i.e., a bug). The nature of these detector mechanisms suggest that they provide exactly the visual information necessary for the frog to survive in its sensory-restricted environment.

In a series of later experiments, Hubel and Wiesel developed a technique which enabled them to record from single cell units in the visual cortex of the cat (1962, 1965) and of the monkey (1968). What they discovered was the existence of highly specialized neural mechanisms in the cortex capable of extracting abstract features or patterns from the visual signal. Furthermore, Hubel and Wiesel noted that this extraction process involves multiple stages of analysis, where depth of processing is directly related to the complexity of feature abstraction. For example, detector mechanisms at the lowest level of perceptual analysis effectively function as abstract pattern recognizers by extracting basic features from small, specific receptive fields. These low-level feature detectors which are located on the retina correspond to

the pattern analyzers discovered by Lettvin et. al. These processes, however, represent only the first levels of perceptual analysis. The neural signals generated by these peripheral feature detectors are then sent to the visual cortex for more advanced stages of processing. At the lowest levels of cortical analysis, feature detectors exist which monitor the output of the retinal detector mechanisms. Since the information being monitored is already abstract in nature, the task of the cortical pattern analyzers is to detect abstract features from an abstract message. Similarly, information extracted by these low-level cortical detectors is then sent to higher-level detector mechanisms for subsequent pattern analysis. In this hierarchical manner, the system is able to continually extract relatively more and more abstract features from the original physical signal. Thus, a perceptual system based on increasingly complex levels of detector mechanisms, provides the power required for the recognition of abstract patterns and features.

Evidence for the existence of feature detectors involved in the perception of speech signals was originally presented in a study by Eimas and Corbit (1973). Their intention was to demonstrate by means of a selective adaptation procedure that the perception of voicing contrasts in speech is mediated by linguistic feature detectors, each sensitive to a restricted range of voice onset times.

Voiced onset time (VOT) is a major acoustic cue under-

lying the perceived phonetic distinction between voiced and voiceless stop consonants. For example, in English it distinguishes /b/ from /p/, /d/ from /t/, and /g/ from /k/. In terms of production, VOT has been defined as the interval between the release of the articulators and the onset of laryngeal pulsing (Lisker and Abramson, 1964). Acoustically, it refers to the delay in the onset of the first-formant relative to the second- and third-formants. Additionally, when the first-formant is absent, the second- and third-formants are noisy rather than voiced (Lisker and Abramson, 1970). The amount of delay in VOT required for a stop to be heard as voiceless rather than voiced is normally about 30 - 40 msec (Abramson and Lisker, 1970).

In order to assess the degree by which selective adaptation alters voicing perception, Eimas and Corbit constructed a continuum of synthetic consonant vowel (CV) syllables, by systematically varying the stimuli in equal steps of VOT. Identification functions were then obtained for listeners in the unadapted state and after adaptation. Adaptation was accomplished by repeatedly presenting a CV syllable selected from either extreme end of the VOT continuum.

Eimas and Corbit reasoned that if a given detector is selectively sensitive to a particular feature in a stimulus pattern, then repeated presentation of that feature should fatigue the detector and reduce its sensitivity. As such, they predicted that adaptation of the

voicing feature should cause a shift in the locus of the phonetic boundary in the direction toward the adapting stimulus' end of the VOT continuum. The results confirmed their predictions. Adaptation with /ba/ caused the phonetic boundary between /ba/ and /pa/ to shift toward the /ba/ end of the continuum. In other words, stimuli near the boundary which were identified as /ba/ when the listener was in an unadapted state, were subsequently labeled /pa/ after adaptation. Similar results were obtained when /pa/ was the adapting stimulus.

The results discussed thus far do not conclusively demonstrate that the observed effects are due to the selective adaptation of "linguistic" feature detectors. Alternatively, it may be that the sound patterns corresponding to the phonetic units are being adapted, and as such, the feature detectors may be auditory rather than phonetic in nature.

Eimas and Corbit rejected this alternative explanation by pointing out that the effects of adaptation are not class specific as indicated by the presence of crossed-consonant shifts. For example, adaptation with the voiceless bilabial stop /p/, produced approximately equivalent effects on the identification functions for a series of alveolar (/d/ and /t/) stop consonants as it did for a series of bilabial (/b/ and /p/) stops. In both cases the locus of the phonetic boundary shifted toward the voiceless end of the continuum, indicating that a greater number of identification responses belonged to the voiced or unadapted

category. From these results taken together with some additional findings based on a discrimination task, Eimas and Corbit concluded that the perception of voicing contrasts involves two distinct classes of feature detectors, each class being specifically tuned to a restricted range of VOT values.

In a subsequent study reported by Eimas, Cooper, and Corbit (1973), an attempt was made to corroborate the earlier interpretation that the effects of adaptation are the result of phonetic rather than auditory perceptual analysis. They reasoned that if the information concerning VOT is extracted by detector mechanisms operating at the auditory level of processing, then repeated presentation of just the essential acoustic information required to specify VOT should produce an equivalent shift in the phonetic boundary as that incurred when voicing information is presented in a speech context. If, however, adaptation of the VOT detectors occurs only with a speech pattern as the adapting stimulus, then it would appear reasonable to infer that the voicing detectors are operating only during phonetic processing.

In order to test these assertions, listeners were repeatedly exposed to a synthetic CV syllable /da/, selected from the extreme end of the VOT continuum. The results from this condition replicated the findings obtained earlier in the Eimas and Corbit experiments. The locus of the phonetic boundary shifted toward the voiced end of the

continuum. In a different experimental condition, identification functions were obtained after adaptation with d-chirps. These adapting stimuli consisted of only the initial 50 msec of the /da/ syllable. Although these initial portions of the speech pattern carry the same acoustic information concerning VOT as the entire /da/ syllable, they do not sound at all like speech, but rather like glissandoes or the chirps of a bird (Liberman, 1970). Adaptation with d-chirps produced no significant changes in the phoneme boundary. These results lend additional support to the notion of linguistic feature detectors specifically sensitive to the distinctive features of speech which are engaged by the language processing system during the phonetic stage of perceptual analysis.

All of the studies discussed thus far have dealt exclusively with the phonetic feature of voicing. More recent experiments have examined the effects of adaptation along the phonetic dimension place of articulation. In English, the feature of place serves to distinguish among the voiced stop consonants /b,d,g/ and the voiceless stops /p,t,k/. The major acoustic cues which carry this information are the transitions (i.e., relatively rapid changes in the frequency of the formants) of the second- and third-formants (Liberman, Cooper, Shankweiler, and Studdert-Kennedy, 1967). However, unlike the feature of voicing, there exists no invariant range of acoustic cues which determine place distinctions. Instead, the major acoustic

cues for place are highly subject to contextual variation. This is so, because the formant transitions, at every instant, simultaneously provide information about both the stop consonant and its vowel environment. The classic example illustrating this point comes from an examination of the syllables /di/ and /du/. In the case of /di/, the acoustic cue for the perception of the consonant /d/ is a rising second-formant transition. It begins at 2200 Hz and climbs up to 2600 Hz. In /du/, the phoneme /d/ is cued by a second-formant transition that falls from about 1200 to 700 Hz. Thus, the same perceived phoneme is cued, in different contexts, by auditory features that are vastly different in acoustic terms.

In a study by Cooper (1974a) an attempt was made to determine whether the selective adaptation procedure can produce alterations in perception along the acoustically non-invariant place feature in a manner analogous to that observed for voicing. It was argued that if adaptation operates only on invariant acoustic information, then perceptual shifts should not occur for the feature of place. If, on the other hand, adaptation is a function of phonetic processing, then perceptual shifts should be observed. To examine these issues, Cooper constructed a set of thirteen synthetic speech syllables ranging perceptually from /bae/ to /dae/ to /gae/, by systematically varying the starting frequency of the second- and third-formant transitions. By then presenting listeners with repeated

presentations of either one of the endpoint stimuli or the midpoint stimulus, it was possible to evaluate the effects of the adaptation procedure on the loci of the two phonetic boundaries. The obtained results indicate that a selective adaptation effect can be induced for the feature place of articulation. Thus, repeated presentation of /bae/ produced a shift in the locus of the /bae/-/dae/ boundary toward the /bae/ end of the place continuum. However, since place of articulation is a tri-valued dimension, it is necessary to also examine the /dae/-/gae/ phonetic boundary.

If adaptation is operating on stable feature modes such as bilabial, alveolar, and velar, then no shift is predicted for the alveolar-velar phonetic boundary when the adapting stimulus is bilabial. On the other hand, if it is found that this boundary does shift toward the bilabial end of the place continuum, such that stimuli that were originally identified as alveolar were, after adaptation, identified as velar, then this would indicate that adaptation was effecting a purely relative feature analyzing system. The results suggest that the former is correct. Adaptation with /bae/ produced no significant change in the /dae/-/gae/ boundary. Similarly, adaptation with /gae/ caused the /dae/-/gae/ boundary to shift toward /gae/ but had no effect on the /bae/-/dae/ phonetic boundary. Finally, adaptation with /dae/ (the midpoint stimulus) produced a shift in the loci of both the /bae/-/dae/ and

the /dae/-/gae/ boundaries. These results, then, suggest that place information is decoded in terms of three separate analyzer modes, and that the locus of a phonetic boundary corresponds to the point along the place continuum where two adjacent analyzer modes respond with equal strength.

After having obtained results consistent with Eimas et. al.'s original "phonetic" interpretation of feature detectors for the feature place of articulation, Cooper set out to examine the extent to which the adapting stimulus could be varied. In one condition, a test was conducted to determine the role of vowel context on the adapting stimulus. It was argued that if adaptation operates primarily at an acoustic-invariant level of perceptual processing, then no boundary shifts should occur in a "crossed-vowel" adaptation test since the frequency values of second- and third-formant transitions are highly vowel dependent. On the other hand, if boundary shifts do occur, then this would suggest that the phonetic feature detecters fatigued during adaptation are specifically sensitive to the distinctive features of speech. To examine this problem, listeners were tested along the /bae/-/dae/-/gae/ identification series after adaptation with the real speech syllable /bi/. The obtained identification functions indicate the presence of a significant shift in the locus of the /bae/-/dae/ phonetic boundary. However, the magnitude of this shift was considerably less than those obtained after adaptation with the real speech syllable /bae/.

In a second experimental condition, the real speech syllable /bae/ was compared with the synthetic speech syllable /bae/ in terms of their effectiveness as an adapting stimulus for the synthetic /bae/-/dae/-/gae/ test series. The results showed that both stimuli were able to produce a shift in the /bae/-/dae/ phonetic boundary, but that the shift incurred by the synthetic speech syllable was significantly larger than that incurred by the real speech syllable.

Finally, a third experimental condition was run in which a "crossed-consonant" adaptation stradegy was employed. Listeners in this condition were adapted with the real speech voiceless bilabial stop CV syllable /p^hae/ and then tested for identification with the synthetic speech series /bae/-/dae/-/gae/. Although the results indicate that /p^hae/ was effective in producing a perceptual shift along the /bae/-/dae/-/gae/ continuum, the magnitude of this shift represented a significant decrement when compared with those obtained when the adapting stimulus was the real speech voiced bilabial stop CV syllable /bae/.

The results from these three different experimental conditions strongly suggest that Eimas' interpretation may be too simplistic. Perhaps instead, it would be more beneficial to view adaptation as a multi-component process which operates during multiple levels of perceptual analysis, such that part of the adaptation effect is attributable to the fatiguing of feature detectors at the auditory level,

while another part is due to the fatiguing of feature detectors at the phonetic level. Thus, /bi/ may have been a less effective adapting stimulus than /bae/ because it was only fatiguing the feature-specific component of adaptation operating at the phonetic level of processing. Similarly, the synthetic speech syllable /bae/ may have produced more auditory adaptation than the real speech syllable /bae/ because the real speech syllable contained more acoustic information irrelevant to the task of perceiving phonetic distinctions based solely on differences in the second- and third-formant transitions. However, this interpretation still does not explain why the real speech syllable /bae/ should have been a more effective adapting stimulus than the real speech syllable /p^hae/. At the auditory level, the information specifying place should have been nearly identical for both of these syllables. Also, at the phonetic level both should have been classified as bilabial. Thus, it appears that some additional explanatory component is required to account for the observed difference.

Cooper's explanation for the significant difference in shift magnitude between the real speech /bae/ and /p^hae/ adaptation conditions was based on the postulation of a high-level "phonetic unit" component of adaptation. This phonetic unit component presumably operates on consonantal sounds as a unit rather than on individual distinctive features. Therefore, a larger shift occurred for the adapting stimulus /bae/ because this syllable's

consonant was represented directly in the /bae/-/dae/-/gae/ test series whereas the syllable /p^hae/'s consonant was not. Thus, it can be seen that Cooper's interpretation of the selective adaptation results assumes not only that the processing of speech sounds takes place at more than one level of perceptual analysis, but also that more than one such level is capable of being adapted (Cooper, 1974b). This approach closely resembles Hubel and Wiesel's account of a perceptual system based on increasingly complex levels of detector mechanisms. However, this is not the only possible explanatory model of the selective adaptation phenomena.

A different explanation can be derived from Helson's Adaptation Level Theory (1964). According to this theory, listeners should partition the /bae/-/dae/-/gae/ test continuum into three equivalent categories when each stimulus occurs with an equal probability. However, when one stimulus occurs more often than any of the other stimuli, the theory predicts that the phonetic boundary should shift toward the more frequently occurring stimulus' end of the continuum. Thus, for the selective adaptation procedure, both a feature detector theory and Helson's adaptation level theory predict that the locus of a phonetic boundary should shift toward the adapting stimulus' end of the test continuum. For the feature detector theory, the effect is due to a generalized decrease in the sensitivity of a feature detector across its entire response range incurred by repeated presentation of its adequate

stimulus. For the adaptation level theory, the effect is attributed to simple response bias. That is, during adaptation the listener is exposed to many more instances of the category from which the adapting stimulus was drawn, and as a result, he has a tendency to identify sounds as belonging to a category other than the one assigned to the adapting stimulus. What Helson's theory is describing is a phenomenon commonly referred to in the psychophysical literature as the contrast effect.

In an attempt to distinguish between these two differing interpretations of the selective adaptation results, Sawusch and Pisoni (1973) and Sawusch, Pisoni, and Cutting (1974) examined the effects of unbalanced probabilities of occurrance of stimuli on the identification functions of a voicing (/ba/-/pa/) and a place (/bae/-/dae/) test continuum. By manipulating the distribution of probabilities such that one of the endpoint stimuli appeared more often than any of the other members of the test series, it was possible to determine whether frequency of occurrence can, by itself, alter phonetic boundaries. According to adaptation level theory, the boundary should shift toward the more frequently occurring stimulus. Feature detector theory, on the other hand, makes no such prediction since relatively no adaptation (fatiguing) is operating within this test procedure. The results indicate that for the class of stop consonants phonetic boundaries do not shift as a function of probability of stimulus occurrence.

It was thus concluded that a response bias explanation cannot account for the boundary shifts found with the selective adaptation paradigm. This conclusion cannot, however, be considered definitive. First of all, there exists a problem related to the degree of probabilistic asymmetry. In the Sawusch et. al. experiments, the more frequently occurring stimulus appeared only twice (voicing) or four times (place) as often as each of the rest of the test stimuli. In contrast, the adapting stimulus in a typical selective adaptation experiment can occur as much as 100 - 200 times more frequently than any other test item. Given the relatively small shifts obtained after adaptation, it is not surprising to find that no shifts resulted from such minor probability manipulations. A second, and perhaps more serious problem involves the assumption that a contrast interpretation and a feature detector interpretation are mutually exclusive. It is possible that a response bias component is confounded with the effects of feature detector fatigue in the selective adaptation procedure's single response measure. If so, then some type of response bias explanation may account for why, in Cooper's experiment, /bae/ was a more effective adapting stimulus than /p^hae/.

Regardless of whether a response bias component is involved in the selective adaptation procedure, the evidence from Cooper's crossed-vowel and crossed-consonant conditions clearly indicate the presence of a feature-specific component. In order to obtain further information concerning

the organizational properties of these feature-specific analyzing mechanisms, Cooper and Blumstein (1974) investigated the adaptation effects when the adapting and test stimuli belonged to different consonant categories. It was reasoned that perceptual shifts should be obtained if the detectors which extract place information operate irrespective of manner information. In contrast, the failure to obtain perceptual shifts would tend to indicate that the adaptation effects operate on analyzers of a more restricted nature which decode place information for each of the major consonant classes individually. To test this, identification functions were obtained for the /bae/-/dae/-/gae/ test series with five different adapting stimuli. The five adapting syllables selected were the voiced stop /bae/, the voiceless stop /p^hae/, the nasal /mae/, the voiced fricative /vae/, and the semiconsonant /wae/. Since they all contained a labial initial segment, a positive finding would be indicated by a shift in the locus of the /bae/-/dae/ phonetic boundary toward the /bae/ end of the place continuum. This prediction was partially confirmed. The results obtained with the adapting syllables /bae/ and /p^hae/ replicated the findings of Cooper (1974a). They showed a significant shift in the /bae/-/dae/ phonetic boundary, the shift being directed toward the /bae/ category. Similarly, the adapting stimuli /mae/ and /vae/ produced the expected perceptual shifts, thus demonstrating that place information is extracted from

consonants independently of their particular manner of articulation. The adapting stimulus /wae/, however, displayed only a slight, non-significant mean shift in the /bae/-/dae/ boundary, although it was in the predicted direction. Since /w/ is a semiconsonant, it was suggested that this single discrepent finding may indicate that the defining limit of the "place" analyzers is that they extract information only for the class of true consonants.

In a more recent experiment, Cooper (1974c) has investigated the effects of vowel environment on the feature-specific component of adaptation. To accomplish this, listeners were adapted with an alternating sequence of two different adapting syllables, /da/ and /t^hi/, and then tested for identification with stimuli selected from two different VOT continua. One VOT series ranged from /ba/ to /p^ha/ and the other from /bi/ to /p^hi/. Since /da/ and /t^hi/ represent both extremes of the VOT dimension, their repetitive alternating presentation should simultaneously adapt both the voiced and voiceless VOT detectors, and as such, produce no perceptual shifts along either test series. This, of course, is assuming that the effects of adaptation are operating solely on the consonant feature of voicing. The results, however, demonstrated a differential effectiveness for the two adapting stimuli as reflected by the occurrence of opposite shifts in the identification functions of the /ba/-/p^ha/ and /bi/-/p^hi/ series. For the /ba/-/p^ha/ series, the phonetic boundary shifted toward

the /b/ category. Conversely, the /bi/-/p^hi/ series exhibited a shift toward the /p/ category. Thus, perception of the /ba/-/p^ha/ stimuli was primarily influenced by the adapting stimulus /da/, while the perception of the /bi/-/p^hi/ stimuli was mainly influenced by the adapting stimulus /t^hi/. These results indicate that adaptation operates on voicing perception in a vowel-contingent manner.

The results from Cooper's contingent-adaptation experiment makes it necessary to reevaluate the claim that the feature-specific component of adaptation is selectively sensitive to a set of "phonetic distinctive features" in the Chomsky - Halle (1968) sense. Alternatively, it is possible that the feature-specific component reflects the operation of high-level auditory detector mechanisms that are specifically sensitive to the acoustic cues which underlie the phonetic distinctive features of speech. Taking this latter interpretation, Stevens and Klatt (1974) have shown how the perception of voicing contrasts can be accounted for strictly in terms of an acoustically-based feature detector model. Similarly, the findings obtained with the selective adaptation procedure for the feature place of articulation have not distinguished phonetic similarities from the acoustic similarities that underlie them. For example, in both Cooper (1974a) and Cooper and Blumstein (1974), it was found that /p^hae/ was a less effective adapting stimulus than /bae/, when the test series

ranged from /bae/ to /dae/ to /gae/. This result can be understood in terms of an acoustic pattern interpretation by noting that /p^hae/ contained second-and third-formant transitions that were relatively weak in energy, whereas the adapting stimulus /bae/ and the /bae/ members of the test series both contained strong second-and third-formant transitions. Recall also that in the Cooper and Blumstein study, relatively strong adaptation effects were obtained for the adapting stimuli /mae/ and /vae/, but not for the adapting stimulus /wae/. This finding can be explained by examining the slopes of the formant transitions. For /mae/ and /vae/, the second- and third-formant transitions were, like the /bae/ members of the identification series, both rising and steeply sloped. The syllable /wae/, on the other hand, contained rising transitions which sloped only gradually. Thus, from the presently reported data, it is impossible to determine whether the fatigued detectors are specific for distinctive features, or rather for the acoustic cues that underlie these feature.

In an attempt to distinguish between an acoustic and a phonetic interpretation, Ades (1974) examined the effects of a CV adapting stimulus on a VC identification series, and vice versa. To accomplish this, two different test continua were constructed. One series ranged from /bae/ to /dae/, and the other (its mirror-image) ranged from /aeb/ to /aed/. The set of adapting syllables consisted of the four endpoint stimuli, /bae/, /dae/, /aeb/, and /aed/.

Listeners were run in eight different experimental sessions; four sessions with the /bae/-/dae/ test series and four with the /aeb/-/aed/ series. In each session a different adapting stimulus was used. It was reasoned that if adaptation operates at a truly phonemic level, then repetition of any sound containing a /b/ should shift the ID boundary toward /b/. Likewise, repeated presentation of a sound containing the phoneme /d/ should shift the ID boundary toward the /d/ category. The results showed that adaptation with either /bae/ or /dae/ produced positive adapting effects on the /bae/-/dae/ test series. Similarly, adaptation with either /aeb/ or /aed/ produced a positive shift in the locus of the /aeb/-/aed/ phonetic boundary. However, when the adapting and test stimuli were drawn from different continua, there were no differential adaptation effects. Thus, /b/ and /d/ in final position were unable to fatigue the detectors responsible for the perception of /b/ and /d/ in initial position, and vice versa. This finding, although not strong enough to refute a phonetically-based feature detector model, clearly specifies the limits of such a model on adaptation.

If the feature detector systems responsible for the perception of speech sounds are sensitive to individual auditory features rather than for entire "linguistic features" (alla Chomsky and Halle), then it should be possible to demonstrate this fact experimentally. Recall, however, that in the Eimas, Cooper and Corbit (1973) study,

it was found that the initial 50 msec of the syllable /da/ (d-chirp) was ineffective in producing alterations in the perception of voicing contrasts. As Ades(1973) has pointed out, though, this finding can be accounted for if it is assumed that the hypothetical "voicing detectors" are only sensitive to the relative onset times of fairly sustained components. As such, the d-chirps would not convey enough information about voice to affect the "voicing detectors." With regard to the feature place of articulation, Ades(1973) has provided results which demonstrate a substantial adaptation effect with chirps (chirps were, however, significantly less effective adapting stimuli than syllables). Furthermore, since listeners found it relatively easy to identify the chirps as speech-like (e.g., b-like or d-like), adaptation was also attempted with an even less speechlike stimulus, the tweet. Tweets consisted of only F2 and F3 transitions (exactly that part of the sound that distinguishes it from the others in the place series). Again, significant adaptation effects were observed, although the effect was relatively small in comparison to those obtained with syllables and chirps.

The question of concern now is how to interpret Ades' results. Clearly, they do not suggest that adaptation is operating solely at the auditory level of perceptual analysis. Had this been the case, then syllables, chirps and tweets should all have produced equivalent adaptation effects since each contained identical acoustic information

relevant to the perception of the feature place of articulation. Especially interesting is a comparison of the effectiveness of chirps and tweets as adapting stimuli. Since the only difference between these two classes of sound patterns is the presence or absence of F1, then this component of the adapting stimulus should account for a relatively large proportion of the auditory adaptation. This conclusion, however, is unreasonable since the first formant can in no way help distinguish individual members of a place series. Perhaps, instead, Ades' findings can best be understood in terms of a strictly phonetic model of feature analyzers. It is possible to view phonetic features as purely relative events. For example, Ades' Ss reported that chirps were b-like or d-like. In this respect syllables can be considered as being more b-like or d-like than chirps, whereas tweets can be thought of as being less b-like or d-like. Thus, it is possible to construct a continuum for the phonetic features bilabial and alveolar in which the syllables /bae/ and /dae/, the b-chirps and d-chirps, and the b-tweets and d-tweets represent strong, moderate, and weak exemplars, respectively. Such an approach can easily account for the observed data. Alternatively, Ades' results are generally consistent with Cooper's (1974b) multi-component model of adaptation which states that part of the adaptation effects are attributable to the fatiguing of feature detectors at the auditory level while another part is due to the fatiguing

of feature detectors at the phonetic level. Within the framework of this approach, the tweet adaptation condition can be viewed as reflecting only the adaptation of auditory detector mechanisms. Adaptation with chirps and syllables, on the other hand, involves both the auditory and the phonetic feature detector systems. The difference in effectiveness between chirps and syllables can be explained in terms of depth of processing. For instance, syllables may also be adapting higher-level "syllable" feature detectors in addition to the auditory and phonetic feature detectors. The finding that Ss were able to identify the chirps as b-like or d-like tends to indicate that they are processed at least during the early stages of phonetic processing.

The purpose of the present experiment was to provide additional information concerning the nature of the detector mechanisms operated upon during selective adaptation. Specifically, this study attempted to resolve the issue of whether the effects of adaptation are the result of feature detector fatigue at both the auditory and the phonetic level of perceptual analysis, or alternatively, whether they only reflect the adaptation of feature detectors operating during the phonetic stage of speech processing. To accomplish this, it was necessary to construct an adapting stimulus which would preserve all of the acoustic properties underlying the perception of place information while dissociating any corresponding

phonetic information. The sound pattern chosen for this task was the "speech-embedded chirp."

The speech-embedded chirp consisted of a CV syllable's initial 50 msec segment (i.e., a chirp) preceded by a steady-state vowel whose formant values were fixed equal to the starting frequencies of the formant transitions. The major advantage of the speech-embedded chirp is that the formant transitions occur in final position. This contrasts with the role of the corresponding acoustic segment in the CV syllable where the same acoustic information occurs in initial position.

By representing the same acoustic information in different serial positions in the adapting and test stimuli, the possibility of phonetic adaptation is greatly reduced. Since the acoustic cues which underlie place information are highly context-dependent, it is very unlikely they would produce identical phonetic transformations in both initial and final position. However, even if they did, it is still unlikely that phonetic adaptation would occur. This assumption is based on Ades' (1974) findings that a phonetic segment in final position was unable to fatigue the same phonetic segment in initial position, and vice versa. Thus, it would appear reasonable to infer that any positive adapting effects obtained with speech-embedded chirps would represent adaptation operating solely during auditory analysis. The failure to find such effects, however, does not rule

out the possibility of auditory adaptation. A negative finding could indicate that auditory adaptation is, like phonetic adaptation, dependent upon serial position.

METHOD

Subjects

The listeners were five paid volunteers, all of whom responded to an advertisement in the Indiana University student newspaper. All Ss were right-handed, native speakers of American English with no known history of a hearing or speech disorder. Ss were paid at the rate of $2.00 per hour. No S had had any prior experience with the selective adaptation procedure, although two Ss had had some previous experience with synthetic speech stimuli.

Stimuli

All of the stimuli used in this experiment were three-formant speech patterns constructed on the parallel resonance speech synthesizer at Haskins Laboratories, and recorded on magnetic tape.

The test stimuli consisted of a series of seven synthetic CV syllables 300 msec in duration. These stimuli ranged perceptually from /ba/ to /da/. Stimuli differed from one another only in the starting frequency and direction of the second- and third-formant transitions. The different starting frequencies of F2 and F3 are displayed in Table 1. F1 always started at 412 Hz. All transitions were 50 msec in duration and linear. The final 250 msec of the CV syllables consisted of steady-state formants appropriate for the English vowel

/a/. These fixed steady-state formants were centered at 769 Hz (F1), 1232 Hz (F2), and 2525 Hz (F3).

Insert Table 1 about here

In addition to the seven test stimuli, two additional stimuli were synthesized, the b-"Speech-Embedded Chirp" (b-SEChirp), and the d-"Speech-Embedded Chirp" (d-SEChirp). These stimuli were constructed in the following manner: First, a 250 msec three-formant steady-state was inserted at the beginning of the two endpoint stimuli in the test series (i.e., Stimulus 1 or /ba/, and Stimulus 7 or /da/). The frequency values of these new steady-state formants were set equal to the starting values of Stimulus 1's and Stimulus 7's formant transitions i.e., 412 Hz (F1), 996 Hz (F2), and 2180 Hz (F3) for the b-SEChirp stimulus, and 412 Hz (F1), 1465 Hz (F2), and 3195 Hz (F3) for the d-SEChirp stimulus. Finally, the original 250 msec steady-states (i.e., the vowel /a/) were deleted. Thus, the b-SEChirp and d-SEChirp stimuli can be characterized as 300 msec sound patterns with 250 msec steady-states in initial position and 50 msec transitions in final position. Stimulus 1, Stimulus 7, b-SEChirp, and d-SEChirp are displayed schematically in Figure 1.

Insert Figure 1 about here

TABLE 1

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

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

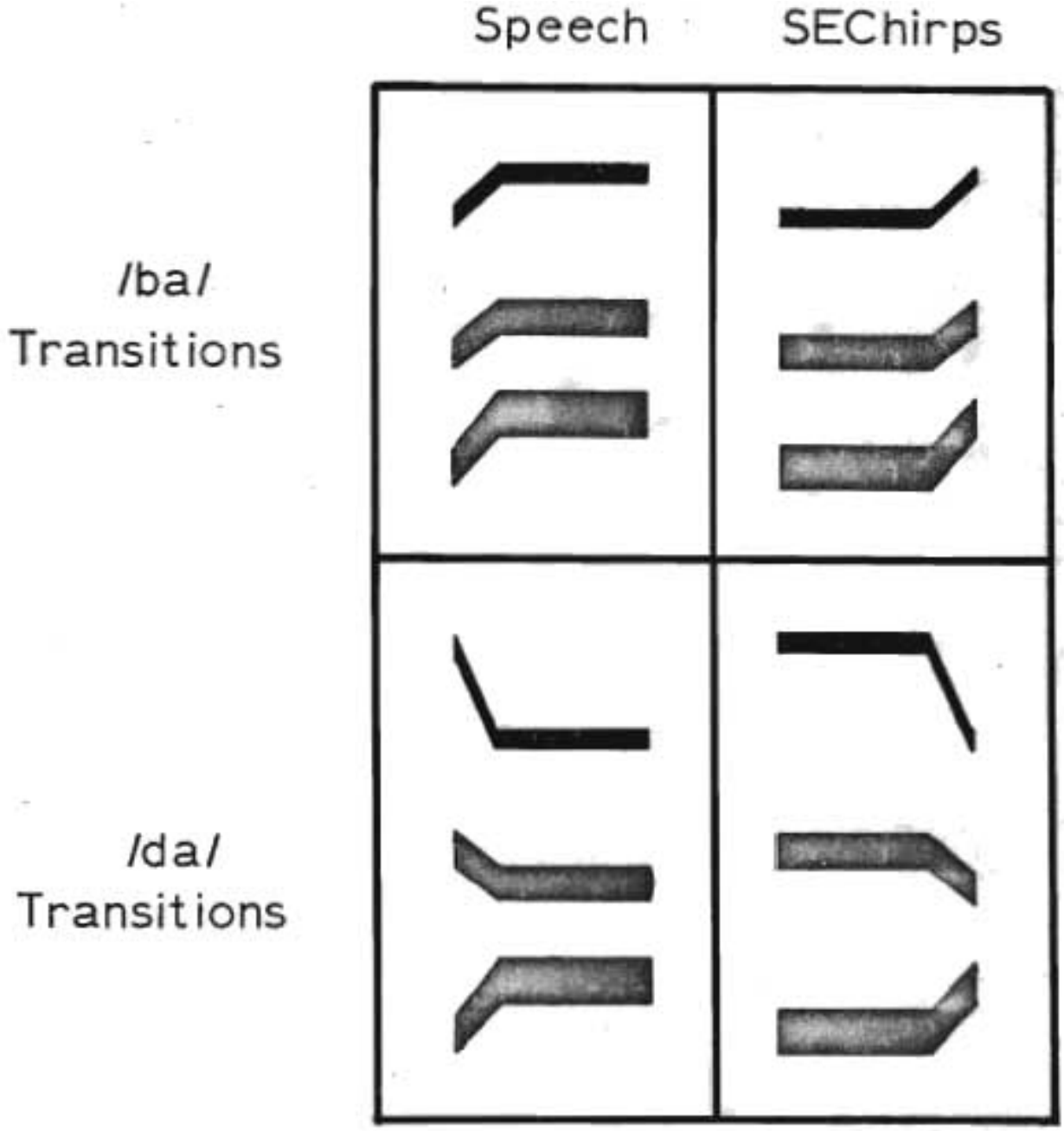


Figure 1. Schematized sound spectrograms of the adapting stimuli /ba/, /da/, b-SEChirp, and d-SEChirp.

Stimuli 1 and 7, and the two speech-embedded chirps were used to construct four different adaptation sequences. Each adaptation sequence consisted of 100 repetitions of the adapting stimulus with a 225 msec interstimulus interval.

Apparatus

All experimental materials were recorded on audio tape and reproduced on an Ampex AG-500 two-track tape recorder and were presented dietically through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the headphones equivalent to 80 dB SPL re 0.0002 dynes/cm^2 for a vowel-like /a/ calibration signal. Measurements were made on a Hewlett-Packard VTVM (model 400) prior to the presentation of each experimental tape. All five Ss were run together in a small experimental room.

Procedure

Baseline identification functions were obtained for all listeners in the unadapted state by presenting twenty random sequences of the seven test stimuli with an interstimulus interval of 3 sec. The Ss were instructed to identify each stimulus as either /ba/ or /da/ by writing the appropriate consonant letter on answer sheets. The listeners were told to respond to every identification stimulus even if they had to guess.

On the day after the initial identification test, a series of four different adaptation tests was conducted, each test lasting roughly 1 hour, and taking place at 24 hour intervals. The order of presentation of the adapting sessions was as follows: Session 1. /b/ - Adapt; Session 2. /d/ - Adapt; Session 3. b-SEChirp Adapt; and Session 4. d-SEChirp Adapt.

Each of the four adaptation tests was conducted in the following manner: Listeners were first presented with two consecutive adaptation sequences (200 presentations) of the selected adapting stimulus (/ba/, /da/, b-SEChirp, or d-SEChirp). Following this "warm-up" period of adaptation (after Cooper, 1974a), ten adaptation trials were administered. Each adaptation trial was composed of 100 presentations of the adapting stimulus with 225 msec between repetitions (i.e., one adaptation sequence). This was followed by 2 sec of silence and then the presentation of the five middle stimuli from the original test series (Stimuli 2-6). The Ss were instructed to identify each of these five test stimuli as either /ba/ or /da/ by writing the appropriate consonant letter on response sheets. The five middle test stimuli occurred in random order with 4 sec between each. After the fifth stimulus was presented for identification, 5 sec intervened before the onset of the next adaptation trial. Each of the ten adaptation trials had a different random order of the five test stimuli. Each stimulus occurred

in each of the five test positions twice. After one presentation of the experimental adaptation tape, Ss were given a short break after which the same tape was rewound and played again. In this manner, each of the five middle stimuli in the test series was presented for identification a total of twenty times within a single adaptation session.

RESULTS

Table 2 shows the individual and mean phonetic boundaries for each of the five experimental sessions: one identification session without adaptation and four adaptation sessions. Each phonetic boundary was computed by finding the point along the stimulus scale which would, by extrapolation, receive 50% /ba/ responses and 50% /da/ responses. In all, there were twenty instances of attempted adaptation, four adapting conditions for each of five subjects. In all but one instance there was a shift in the locus of the /ba/-/da/ phonetic boundary in the predicted direction. The only exception was S 2 who showed virtually no shift at all in the d-SEChirp adaptation test.

Insert Table 2 about here

In Figure 2 the group identification and adaptation functions, averaged over all five Ss, are plotted. It should be noted that the shifts produced by selective adaptation were not accompanied by a decline in the steepness of the response function slopes.

Insert Figure 2 about here

TABLE 2

Individual and Mean Loci of Phonetic Boundaries for Each Test Condition

Syllable Conditions

Subjects	Without Adaptation	Adaptation with /ba/	/da/
1	4.095	3.250	5.250
2	3.594	2.600	4.357
3	3.571	2.625	4.625
4	3.950	3.100	5.643
5	4.095	2.588	5.643
$\bar{X}$	3.797	2.766	5.133

Speech-Embedded Chirp Conditions

Subjects	Without Adaptation	Adaptation with b-SEChirp	d-SEChirp
1	4.095	3.556	4.500
2	3.594	3.412	3.588
3	3.571	3.474	4.167
4	3.950	3.556	4.286
5	4.095	3.714	4.500
$\bar{X}$	3.797	3.534	4.228

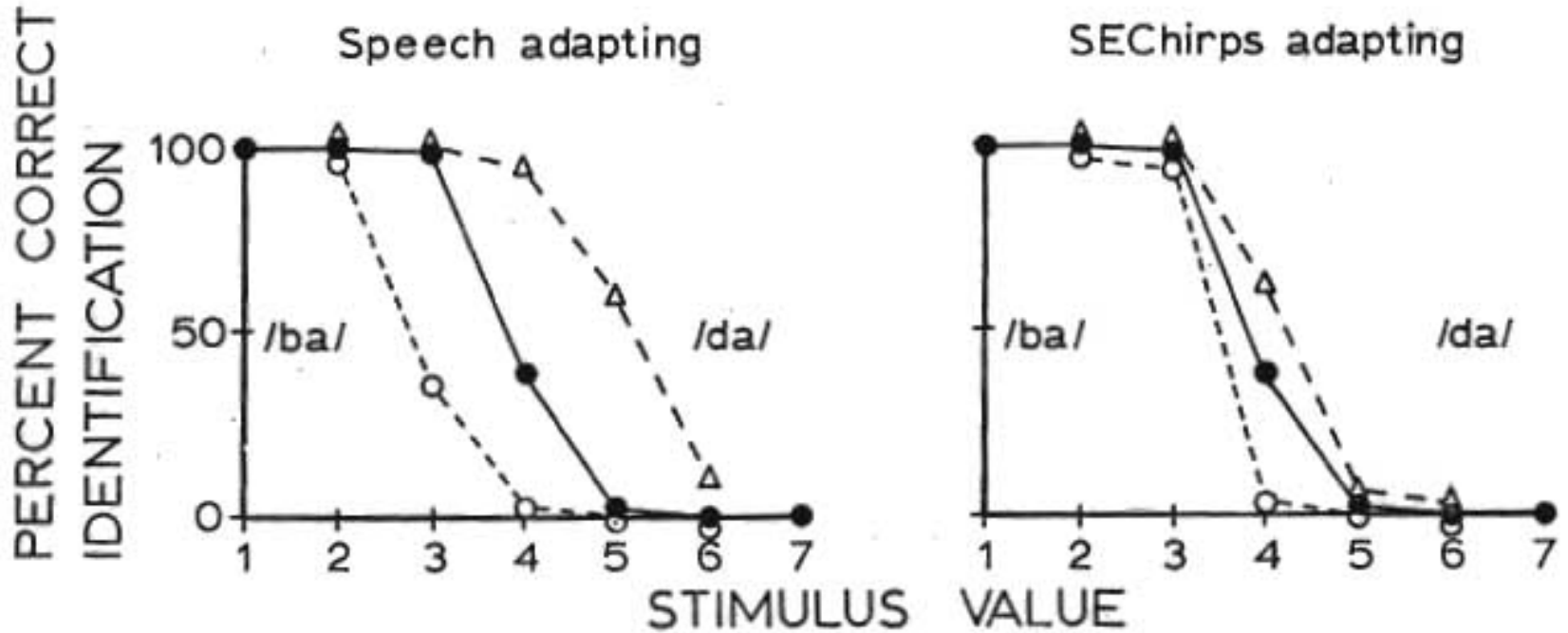


Figure 2. Group identification and adaptation functions.

Shown in Table 3 are the results of tests of significance for the shifts in the locus of the /ba/-/da/ phonetic boundary. Only one-tailed tests of significance were applied, the predicted direction of the boundary shift being toward the category of the adapting stimulus. All shifts in the phonetic boundary incurred by adaptation were found to be significant in comparison to the baseline identification data.

Insert Table 3 about here

For the /ba/ and /da/ syllable adaptation conditions, the respective mean shifts in the locus of the /ba/-/da/ phonetic boundary of 1.031 and 1.336 stimulus value units were found to be highly significant (Correlated t-tests: t=8.36 for /ba/ adaptor, $p < .005$; t=7.36 for /da/ adaptor, $p < .005$). Moreover, the magnitude of these shifts are comparable in direction and consistency to those found in previous studies which investigated adaptation effects on a place series (e.g., Cooper, 1974a; Cooper and Blumstein, 1974; Ades, 1973; Ades, 1974).

For the speech-embedded chirp adaptation conditions, the mean boundary shifts of .263 (b-SEChirp) and .431 (d-SEChirp) stimulus value units were also found to be significant (Correlated t-tests: t=4.02

TABLE 3

Tests of Significance for Shifts in Phonetic Boundary Loci Between the Unadapted Condition and Each Condition of Adaptation

Syllable Conditions

Adaptation with:	/ba/	/da/
p-value:	$p < .005$	$p < .005$

Speech-Embedded Chirp Conditions

Adaptation with:	b-SEChirp	d-SEChirp
p-value:	$p < .01$	$p < .02$

Note: All tests were one-tailed. The direction of each significant shift was toward the category of the adapting stimulus.

for b-SEChirp adaptor, $p<.01$; $t=3.53$ for the d-SEChirp adaptor, $p<.02$). Within these test conditions, the direction of the significant shift was toward the phonetic category from which the adapting stimulus' formant transitions were originally obtained. Thus, after adaptation with the b-SEChirps, Ss made fewer /ba/ identification responses. Similarly, after adaptation with the d-SEChirps, the listeners assigned fewer identification responses to the /da/ category.

DISCUSSION

The fact that speech-embedded chirps were able to produce alterations in the perception of a place series strongly supports the notion that one component of the selective adaptation process reflects the fatiguing of acoustically-oriented feature detectors that operate during the auditory stage of perceptual analysis. The reasoning underlying this conclusion is based on the assumption that any phonetic information conveyed by the speech-embedded chirps was irrelevant in the present task. This assumption was derived from Ades' (1974) results which demonstrated that repeated presentations of a phonetic segment in final position were unable to fatigue the feature detector mechanisms responsible for the perception of the same phonetic segment in initial position.

The validity of this study's conclusion, however, does not depend on the validity of its underlying assumption. In fact, if we reject Ades' results as the product of statistical error, and consider the possibility that phonetic adaptation was operating within the present experimental paradigm, then the results point even more definitively toward an acoustic interpretation. This is so because a phonetic account of this study's design would have predicted shifts in the locus of the /ba/ - /da/ phonetic boundary opposite

those predicted by the acoustic model for the speech-embedded chirp adaptation conditions. Thus, a phonetic interpretation would predict fewer responses assigned to the /d/ category after adaptation with the /b/ - speech-embedded chirps and fewer /b/ responses after adaptation with the /d/ - speech-embedded chirps. The logic behind these somewhat counter-intuitive predictions can best be understood by examining the acoustic structures characteristic of the phonemes /b/ and /d/ in final position. Given either the steady-state formants assigned to the b-SEChirps or the steady-state formants assigned to the d-SEChirps, the phoneme /b/ in final position would have been characterized by falling transitions in all three formants. The phoneme /d/, on the other hand, would have been characterized by rising second- and third-formant transitions, and a falling first-formant transition. In the actual b-SEChirps, all three formants were rising. In the actual d-SEChirps, the first-formant transition was rising while the second- and third-formant transitions were falling. However, since F1 does not convey any information concerning place of articulation, it would be expected that if adaptation was operating during the phonetic stage of perceptual processing then, based solely on the slopes of the second- and third-formant transitions, the listener's /ba/ - /da/ phonetic boundary should have shifted toward the /da/ category after adaptation with b-SEChirps, and

toward the /ba/ category after adaptation with d-SEChirps. Of course, the obtained findings displayed the opposite effects indicating that adaptation was operating during the auditory stages of perceptual processing.

An interesting follow-up to the present study could attempt to replicate the findings obtained here using speech-embedded chirps with falling first-formant transitions. Then, assuming that Ades' results are not the product of statistical error, it should be possible to shift a listener's /ba/ - /da/ phonetic boundary toward the /ba/ category using an adapting stimulus with a d-like quality in final position (b-SEChirp with falling F1 transition). Similarly, it should be possible to shift the /ba/ - /da/ boundary toward the /da/ category using a b-like adapting stimulus (d-SEChirp with falling F1 transition).

Of particular interest in the present investigation was the finding that the magnitude of the shifts incurred during the speech-embedded chirp adaptation conditions closely resemble the shift magnitudes obtained by Ades (1973) after adaptation with tweets (only F2 and F3 transitions). If you recall, it was suggested that tweets represented the adaptation of only the auditory component of adaptation. As such, it would appear that both results reflect the adaptation of the same underlying perceptual mechanisms.

The existence of detector mechanisms operating at the auditory level of speech analysis has two very important implications. First, it provides the missing link in a hierarchical feature detector network such as that proposed by Hubel and Wiesel. And second, it corroborates the findings of several different paradigms which have demonstrated a distinction between auditory and phonetic stages of processing.

REFERENCES

Abramson, A. S. and Lisker, L. Discriminability along the voicing continuum: Cross-language tests. In Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. Prague: Academia, 1970, 569-573.

Ades, A. E. Some effects of adaptation on speech perception. Quarterly Progress Report of the Research Laboratory of Electronics, M. I. T., 111, 121-129.

Ades, A. E. A study of acoustic invariance by selective adaptation. Perception & Psychophysics, 1974, in press.

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

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974a, in press.

Cooper, W. E. Selective adaptation to speech. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, and D. B. Pisoni (Eds.), Cognitive Theory, Volume I. Potomac, Maryland: Erlbaum, 1974b, in press.

Cooper, W. E. Contingent feature analysis in speech perception. Perception & Psychophysics, 1974c, in press.

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

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

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

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

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

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

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

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

Lettvin, J. Y., Maturana, H. R., McCulloch, W. S., and Pitts, W. H. What the frog's eye tells the frog's brain. Proceedings of the Institute of Radio Engineers, New York, 42, 1959, 1910-1951.

Liberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), Perception and its disorders, Proceedings of the A. R. N. M. D. Baltimore: Williams and Wilkins, 1970.

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

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

Lisker, L. and Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. Proceedings of the 6th International Congress of Phonetic Sciences, 1970, 563-573.

Pisoni, D. B. and Tash, J. Reaction times to comparisons within and across phonetic categories. Perception & Psychophysics, 1974, 15, 2, 285-290.

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

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

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

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

Stevens, K. N. and Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction of stops. Journal of the Acoustical Society of America, 1974, 55, 3, 653-659.

Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), Current Trends in Linguistics, Volume XII, The Hague: Mouton, 1974 (Also appeared in Haskins Laboratories Status Report on Speech Research, SR-23, 1970, 15-48).

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

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

Wood, C. C. Parallel processing of auditory and phonetic information in speech perception. Perception & Psychophysics, 1974, in press.

Simultaneous Adaptation of Two Features in a Bidimensional Speech Series

James R. Sawusch

Indiana University

Abstract

Recent studies have explored the context or limits within which feature adaptation occurs in speech perception. Vowel quality and position of the consonant feature in the syllable have been found to be important variables. The present experiment examines the role of simultaneous variation in two consonantal features in determining adaptation. Three stimulus series were used; a place series (/ba/-/da/), a voicing series (/ba/-/pa/), and a bidimensional series (/ba/-/ta/), combining the two single feature series. Ss identified the bidimensional series into three categories of stimuli, /ba/, /pa/, and /ta/. Discrimination functions obtained for this series revealed categorical perception. Ss could discriminate pairs of these stimuli no better than they could absolutely identify them. After adaptation with /ba/, /da/, or /pa/ the shifts in the bidimensional series on the individual features were examined. Ss exhibited shifts in the bidimensional series that were not found in the single feature series. The results suggest that the consonantal features of place and voicing are not processed indepentently and that adaptation on the place feature depends on the value of the voicing feature in both the adaptor and the test stimuli.

Simultaneous Adaptation of Two Features in a Bidimensional Speech Series

James R. Sawusch

Indiana University

A new emphasis in theoretical accounts of speech perception has recently taken hold. The primary impetus for this has been the results of work with the selective adaptation paradigm first used by Eimas and his coworkers (Eimas and Corbit, 1973; Eimas, Cooper and Corbit, 1973). The results of work with this paradigm, which will be briefly reviewed here, have suggested the existence of auditory and/or phonetic feature detectors in speech perception. These feature detectors are assumed to mediate the perception of the acoustic information in speech that is necessary to identify phonemes or syllables.

In the original work with this paradigm, Eimas and Corbit (1973) used a set of three formant synthetic stop consonant-vowel (CV) syllables that ranged perceptually from /ba/ to /pa/. This series varied along the linguistic dimension of voicing. The /b/ phoneme is considered voiced and the /p/ phoneme is considered voiceless. The acoustic cue underlying this phonetic distinction is voice onset time (VOT). This is a compound cue consisting of two components (Liberman, Delattre, and Cooper, 1958; Lisker and Abramson, 1964). One of these is the lag in the onset of the first formant relative to the second formant. This cue is called cutback and is made by physically deleting part of the beginning of the first formant transition. The second acoustic cue in VOT is aspiration (noise) in the second and third formant transitions, in place of the band of harmonics.

Eimas and Corbit (1973) presented Ss with a /ba/-/pa/ series for identification into two categories, /ba/ and /pa/. In subsequent sessions, Ss heard one minute of repeated presentations (75) of the /ba/ stimulus representing the voiced end of the /ba/-/pa/ series. This was followed by one of the /ba/-/pa/ series stimuli for identification. This format was repeated, varing the stimulus to be identified until each of the 14 stimuli had been identified 10 times. Results showed a small but consistent shift in Ss' identification functions as a result of adaptation to the /ba/ syllable. When compared to the unadapted identification function, the /ba/ adapted function had shifted toward the adapting stimulus. This same type of effect was also found when the voiceless syllable /pa/ was the adapting stimulus. The identification function shifted toward the /pa/ end of the series.

Eimas and Corbit also used the alveolar set, /da/ and /ta/, as adapting stimuli. The results were the same as those found with /ba/ and /pa/. When /da/ was the adapting stimulus the identification function shifted toward the voiced (/ba/) end of the series, relative to the unadapted function. The converse was true of using /ta/ as the adapting stimulus. A shift was also found in Ss' discrimination functions for the same stimuli. The peak in discrimination shifted with the identification boundary.

Eimas and Corbit interpreted these findings as reflecting the adaptation of a phonetic feature detector for voicing. At least two other hypotheses are also consistent with these findings. Stevens and Klatt (1974) questioned whether the feature detectors being adapted were indeed phonetic. The acoustic cues for voicing in the stimuli

Eimas used were the same for the two voiced stimuli, /ba/ and /da/. Similarly, the acoustic cues for voicing in the voiceless adaptors were also identical. Since the voicing feature does not change acoustically between different voiced (voiceless) stimuli, the mechanism mediating the observed adaptation may have an acoustic rather than phonetic basis. An acoustic feature analyser for detecting the relative onset between two frequency bands could be adapting in these experiments.

The third candidate for explaining these shifts in identification functions is a response bias mechanism. Response bias theory would predict a shift is an identification function roward a more frequently occuring stimulus (in this case, the adapting stimulus). However, response bias can be ruled out as a primary explanatory factor. Sawusch and Pisoni (1973) and Sawusch, Pisoni and Cutting (1974) have shown that the consonantal feature boundaries for place of production and voicing do not shift when only the probability of occurrence of various stimuli is changed. This is in contrast to the shift found for the identification of the non-linguistic dimension of fundamental frequency in the same stimuli. The shift found for fundamental frequency was toward the more frequently occuring stimulus, relative to as equiprobability control. This is as predicted by a response bias theory.

It seems reasonable to conclude that the recent adaptation effects found with consonants occur during relatively early perceptual processing, instead of during a later response stage. Other researchers have extended these initial findings with features other than voicing.

These results have served to demonstrate part of the acoustic and phonetic context within which feature adaptation can occur.

Results of Ades (in press) indicate that a CV stimulus will not adapt a VC series and visa versa, even though the vowels are identical and the consonant of the adapting stimulus comes from the test series. Cooper and Blumstein (in press) have found adaptation on a /ba/-/da/ (place) series with /ba/, /p^ha/, /ma/, and /va/ as adapting stimuli. All of these adapting stimuli have the same phonetic value on the place feature (i.e., bilabial). All of these stimuli produced a shift in the /ba/-/da/ identification function toward the /ba/ (bilabial) end of the series.

Cooper (in press) has also reported results when the vowel in a CV syllable is varied. Cooper used two different adapting stimuli, /da/ and /t^hi/. Ss heard three presentations of one and then three of the other, over and over again as the adapting series. This sequence of alternations was followed by a test on four syllables. These four were randomly drawn from two test sequences. One series ranged from /ba/ to /p^ha/ and the other from /bi/ to /p^hi/. Results showed differential adaptation, depending on the vowel environment. The /ba/-/p^ha/ series exhibited a shift toward the /ba/ end of the series. The /bi/-/p^hi/ set showed a shift toward the /p^hi/ end of the series. This would seem to indicate that not only is adaptation sensitive to the vowel environment but also that two feature mechanisms exist for the voicing feature. One is for voiced consonants and the other is for voiceless consonants. It also appears that these two mechanisms operate somewhat independently of each other.

One other result of Cooper (in press) should be mentioned here. A three category, place of production series which varied perceptually from /bae/ through /dae/ to /gae/ was used. Upon adaptation with /bae/, only the /bae/ to /dae/ boundary shifted. The /dae/ to /gae/ boundary did not move. Similarly, upon adaptation with /gae/, only the /dae/ to /gae/ boundary shifted (toward /gae/). When /dae/ was used as the adapting stimulus, both boundaries shifted toward the /dae/ stimulus. Cooper concluded that there are three detectors operating for the place feature and that adjacent pairs of the three detectors have an overlap in sensitivity. During adaptation, one of these detectors is fatigued and consequently becomes less sensitive. As a result, the region of overlap in sensitivity between this detector and adjacent detectors shifts. The feature boundry, which is in this overlap region, also shifts.

To summarize these results, there seems to be a feature detection system operating in the perception of stop consonants. These detectors appear to be sensitive to repeated presentations of a particular feature that causes fatiguing of the detection mechanism. This result manifests itself in a shift in Ss' identification functions for stimuli differing on the particular feature. The identification boundary shifts toward the adapting stimulus. However, the context within which adaptation takes place is highly restricted. Adaptation seems to be relatively vowel specific and does not generalize past the feature boundary in question. Adaptation also seems to be very sensitive to the position of the consonant within the syllable. These results also

indicate that adaptation on one consonantal feature is relatively independent of the particular values of other consonantal features.

The purpose of the present experiments was to further investigate the context of feature adaptation. The adaptation paradigm was used to further investigate the way in which voicing and place features are combined in speech processing. Three sets of synthetic speech stimuli were used in the present experiments. The place series varied perceptually from /ba/ to /da/. The voicing series ranged from /ba/ to /pa/. The third series combined the acoustic cues of the place and voicing series in a pairwise fashion. Each value on the place dimension was paired with one value on the voicing dimension. This series will be called the bidimensional or /ba/ to /ta/ series. Ss percieved the end point stimuli as /ba/ and /ta/ respectively. The /ba/ endpoint stimulus is acoustically identical to the /ba/ endpoint stimuli in the place and voicing series. Stylized sound spectrograms of the four end point stimuli (/ba/, /da/, /pa/, and /ta/) are shown in Figure 1.

This type of bidimensional series has been used previously by Sawusch and Pisoni (in press). When Ss were given four response categories to use in labeling these stimuli (B, D, P, and T) they used the P label very consistently for the middle stimuli. The D label was seldom, if ever, used. This series would seem to offer a good opportunity to test the context of feature adaptation. If the adaptation on the bidimensional series (with four response alternatives) is the sum of the adaptation on the component feature dimensions then it would seem that the consonantal features of place and voicing are being processed

and adapted somewhat independently. On the other hand, if adaptation on one feature is found to be dependent on the value of the other feature, then evidence for the non-independence of feature processing will have been found. Whichever result is found, further information as to the context within which feature adaptation results in consonant perception will also have been found.

However, before proceeding with this experiment a second and prior question must be answered. What is the nature of the response categories that Ss employ in the bidimensional identification task? In the component place and voicing series, perception is categorical (Liberman, Harris, Hoffman, and Griffith, 1957). That is, Ss can discriminate stimuli from different categories almost perfectly but they can discriminate stimuli from the same category no better than chance. The test stimuli in previous adaptation experiments all exhibit this categorical form of perception which is a characteristic of speech perception (Liberman, 1970). The purpose of Experiment I was first to investigate whether the stimuli in the bidimensional series are indeed perceived categorically. Experiment II focuses on the role of various consonantal features in adaptation. Specifically, Experiment II will look at whether the place and voicing features are processed seperately or if there is an interaction in processing.

In the first experiment, ABX discrimination functions for the two single feature series and the bidimsional series were examined. In the ABX discrimination paradigm, Ss listen to three stimuli; the first two are always different, the third is identical to either the first ot the second. The Ss task is to determine whether the third stimulus is the

same as the first or the second stimulus. If Ss can discriminate stimuli no better than they can absolutely identify them, it is assumed that they are percieved categorically. This would support the conclusion that Ss identify stimuli in the bidimensional series on a phonetic rather than an acoustic basis (Liberman et al., 1957). The strong categorical assumption, that Ss can discriminate no better than they can absolutely identify, has been formalized as an equation for predicting Ss' discrimination on the basis of their identification of the same stimuli (Liberman, et al.,1957; Pollack and Pisoni, 1971; Fujisaki and Kawashima, 1970). In order to facilitate evaluation of the discrimination data, the predictation equation of Fujisaki and Kawashima (1970) was employed. This formula is summarized in equation 1 below.

$$D_{AB} = 1/2\,[(P_1-P_2)^2 + P_1(1-P_2) + P_2(1-P_1)] + M[P_1P_2 + (1-P_1)(1-P_2)] \quad (1)$$

Here, D_{AB} represents the predicted discriminability of stimuli A and B in an ABX triad. P_1 represents the probability of identifying the first (A) stimulus as a particular phoneme and P_2 the probability of identifying the second (B) stimulus as the same phoneme. M is a constant that allows for estimating the Ss ability to use acoustic information in discrimination within a phonetic category. For our purposes, M is assumed to be .5 (Ss can discriminate within a phonetic category no better than chance).

This equation works well for the /ba/-/da/ and /ba/-/pa/ series which have only two response alternatives. For four response alternatives, this equation is inappropriate. However, if two

assumptions are made about Ss discrimination, then an appropriate prediction equation can be constructed. The first assumption is that Ss are discriminating phonetic features, not phonemes. Hence, the probability of identification of a given feature rather than phoneme is the basis of predicting discrimination.

The second assumption is that the two phonetic features are discriminated seperately and that Ss base their ABX judgment on whichever dimension affords them better discrimination. The result of using these two assumptions is summarized quantitatively in equation 2:

$$D_{AB} = \max \{1/2[(P_{V1}-P_{V2})^2 + P_{V1}(1-P_{V2}) + P_{V2}(1-P_{V1}) + P_{V1}P_{V2} + (1-P_{V1})(1-P_{V2})],\ 1/2[(P_{P1}-P_{P2})^2 + P_{P1}(1-P_{P2}) + P_{P2}(1-P_{P1}) + P_{P1}P_{P2} + (1-P_{P1})(1-P_{P2})]\} \quad (2)$$

P_{V1} represents the probability of identifying the first stimulus in the ABX triad as voiced. P_{V2} represents the probability of identifying the second stimulus in the ABX triad as voiced. Similarly, P_{P1} and P_{P2} represent the probabilities of identifying the first and second ABX triad stimuli as bilabial (the place feature of /b/ and /p/). The parameter M from equation 1 has been assumed equal to one half.

To use equation 2 the identification functions for the features of place and voicing in the bidimensional series are calculated seperately. The probability of identifying a particular feature is found by summing the identification probabilities of all of the stimuli which contain that feature. For example, to get the probability that a stimulus is identified as voiced, the probabilities of identifying

the stimulus as /ba/ and /da/ are added. On the basis of the two feature functions, equation 2 can be used to predict the discrimination function for the bidimensional series. This predicted function can be used for evaluating how closely the obtained data conform to categorical perception.

Experiment I

Method

Subjects. Ss were ten students in introductory psychology participating as a part of the course requirement. All Ss were native American speakers of English, right-handed and reported no history of any speech or hearing disorder.

Stimuli. The three synthetic speech syllable series were /ba/ to /pa/, /ba/ to /da/, and /ba/ to /ta/. Each series contained 7 stimuli. All stimuli were three formant patterns of 300 msec total duration. This included a 50 msec initial transition and a 250 msec steady state vowel (/a/). The /ba/ to /da/ series varied in the initial frequencies of the second and third formant transitions. The second formant varied from an initial value of 996 Hz (/ba/) to an initial value of 1465 Hz (/da/) in six equal steps. Likewise, the third formant varied from an initial value of 2180 Hz to an initial value of 3530 Hz in six equal steps. The /ba/ to /pa/ series varied in VOT from 0 msec VOT (/ba/) to +60 msec VOT (/pa/) in 10 msec steps. Aspiration replaced harmonics in the second and third formant transitions for the duration of the first formant cutback. The /ba/ to /ta/ series combined these two component changes in a one-to-one fashion, resulting in the third seven

step series. These three series of synthetic stimuli were prepared on the speech synthesizer at Haskins Labratories and recorded on magnetic tape to produce three identification test tapes and three ABX discrimination test tapes. All ABX triads were constructed from stimuli that were two steps apart (1-3, 2-4, 3-5, 4-6, 5-7).

Procedure. The experimental tapes were reproduced on a high quality tape recorder (Ampex AG-500) and were presented binaurally through Telephonics (TDH-39) matched and calibrated headphones. The gain of the tape recorder playback was adjusted to give a voltage across the headphones equivalent to 80 dB SPL re 0.0002 dynes/cm^2 for the steady state calibration vowel /a/.

On any one identification tape Ss heard 10 trials of each of the seven stimuli in random order with 4 sec between stimuli. The discrimination tapes contained 4 occurences of each of the 4 orderings (ABB, ABA, BAB, BAA) of each of the 5 ABX triads. There were 4 sec between ABX triads and one-half second between stimuli within an ABX triad.

The order of tape presentation was counter balanced between groups. One group heard the place tapes before the voicing tapes on the first day and visa verse on the second day. The second group recieved the opposite order of presentation. Both groups listened to the bidimensional tapes on the third day.

For each of the identification tapes Ss were told that they would hear synthetic speech syllables and that they were to identify them as /ba/ or /da/, /ba/ or /pa/, or /ba/, /da/, /pa/, or /ta/. Ss were

told to record their identification response to each stimulus by writing down the initial consonant in prepared booklets. For the ABX tapes, Ss were told that they would hear a sequence of three synthetic speech stimuli and that they were supposed to judge whether the third stimulus was most like the first stimulus or the second. Ss recorded their response (a 1 or 2) in prepared response booklets.

Results and Discussion

The group identification functions for the /ba/-/da/ and /ba/-/pa/ series are shown in Figure 2. Results are averaged over all 10 Ss. The averaged discrimination functions are also plotted along with the predicted discrimination functions (using equation 1) for comparison. These results are in accord with previous experiments using single feature CV sreies. Ss show a peak in discrimination for stimulus pairs taken from across the category boundary. For stimuli taken from within a category, Ss show little better than chance discrimination. The match between predicted and actual discrimination functions is quite good.

The group identification function for the bidimensional (/ba/-/ta/) series is shown in Figure 3. Again, results are averaged over all 10 Ss. The partitioning of the bidimensional series into three categories by Ss in this experiment replicates the findings of Sawusch and Pisoni (in press). The almost total absence of /da/ responses to the bidimensional series also conforms to earlier results.

The discrimination results are also shown in Figure 3. As in the single dimension series, Ss show a peak across category boundaries and

a trough within categories. The two discrimination peaks correspond very well to the category boundaries. The fit of the predicted discrimination function to the obtained one is also very good.

From these discrimination results it seems reasonable to conclude that Ss are categorizing the bidimensional stimuli on a phonetic basis. The good match between the obtained and the predicted functions supports this conclusion since the predicted function is based on the notion that only phonetic information is entering into the discrimination. If auditory information was being used, we would expect better than chance discrimination within phonetic categories. Thus, the bidimensional series is a valid set of stimuli for testing the adaptation phenomenon. This series conforms to the classical test of perception in the speech mode. The bidimensional series is perceived categorically. Experiment II used this same bidimensional series in an adaptation paradigm. If the features of place and voicing are not processed seperately then the identification boundary shifts found for the features of place and voicing in the two component series should be different from the shifts found in the bidimensional series.

Experiment II

Method

Subjects. Ss were five Indiana University undergraduates. Ss were paid for their participation. All Ss were righthanded, native American speakers of English and reported no history of either speech or hearing disorders.

Stimuli. The stimuli were the same as those used in the previous experiment. The stimuli were prepared on the speech synthesizer at

Haskins Labratories and recorded on magnetic tape to produce a total of ten experimental test tapes. The three identification tapes were identical to those used in Experiment I. In addition, seven adaptation test tapes were constructed. Each of these tapes consisted of ten adaptation and test sequences. The adapting stimulus and the test stimuli on any one tape were always the same. One adaptation and test sequence was composed on 1 min (100 presentations) of the adapting stimulus with 300 msec between repetitions. Following these repetitions there was 2 sec of silence and then the five middle stimuli from one of the three series. The stimuli occured in random order with 4 sec between stimuli. This short identification sequence was followed by 5 sec of silence and then the cycle repeated. Each of the ten cycles had a different random order of the five test stimuli. Esch stimulus occured in each of the five test positions twice (10 cycles). The /ba/ and /da/ end point stimuli were used as adaptors for the place series, /ba/ and /pa/ end point stimuli were used as adaptors for the voicing series and the /ba/, /da/, and /pa/ end point stimuli were used as adaptors for the bidimensional series. This provided seven adaptation test tapes.

Procedure. The experimental tapes were reproduced as in Experiment I. All Ss were run together in a group. On the first day, all Ss listened to the three identification test tapes. Instructions for classifying the stimuli were the same as in Experiment I.

Beginning on the second day, Ss heard two presentations of one of the adaptation test tapes each day. Ss took a five minute break between tape presentations. On consecutive days, neither the adapting stimulus

nor the testing sequence were the same. On the ninth day Ss heard the /ba/-/da/ test with /ba/ as the adapting stimulus for the third and fourth times. This was done in order to assess the reliability of any shift found. The tenth day consisted of listening to the three identifiaction test tapes for a second time to assess the post adaptation baseline identification functions.

Results and Discussion

All Ss showed a shift in their /ba/-/da/ (place) identification function when adapted with either /ba/ or /da/. Following adaptation to /ba/, the function shifted toward the /ba/ (bilabial) end of the series. The shift found for presentation of this tape on the ninth day was the same as that found for the second day presentation. Thus, the shift seems to be quite reliable, despite intervening exposure to many other, different stimuli. Conversely, following adaptation with /da/, the function shifted toward the /da/ (alveolar) end of the series. The group identification function, averaged over all five Ss, is shown in Figure 4A, along with the adapted functions. The magnitude of these shifts is comparable to that found by Cooper (in press) for a three category place series.

In contrast to the place results, the /ba/-/pa/ (voicing) series showed a shift in only one direction. All five Ss exhibited a shift toward the voiceless /pa/ end of the series after adaptation with the voiceless /pa/ stimulus. However, only two Ss showed any discernable shift toward the voiced end of the series (/ba/) after adaptation with the voiced /ba/ stimulus. No shift was found for two Ss and one S

showed a small shift in the opposite direction (toward the voiceless, /pa/ end). The group results are shown in Figure 48. The /ba/ adapted function shows no shift in the group results. This finding is somewhat contradictory to earlier results. However, the failure to find a shift after adaptation to a voiced stimulus may be due to insensitivity in the measurement. The increment between voicing series stimuli in this experiment was a relatively large 10 msec VOT. In previous studies of adaptation in the voicing feature (Eimas and Corbit, 1973; Eimas et al., 1973) the increment between stimuli was 5 msec VOT. The average shift toward the voiced end of the series was 6.1 msec as opposed to an average shift of 10 msec toward the voiceless end of the series. It is possible that our 10 msec VOT increment between stimuli was just too large to pick up a shift for the two Ss who showed no shift. As in previous studies, a larger shift was found toward the voiceless end of the series (approximately 5 msec VOT) then was found toward the voiced end of the series (no shift).

For the bidimensional series, results were analyzed by feature rather than by response. To obtain the probability for any stimulus of a voiced response the probability of a /ba/ response is added to that of a /da/ response. These two responses represent a decision that the stimulus was voiced. Similarly, the /ba/ and /pa/ responses represent a decision that a stimulus has the bilabial value of the place feature. Consequently, the probabilities of a /ba/ and /pa/ responsa are added together to obtain the probability of a bilabial response for a given stimulus. The results of this feature analysis of the /ba/ - /ta/ series are presented in Figures 5, 6, and 7. These figures show the

results of adaptation on the place and voicing features with /da/, /pa/, and /ba/ adapting stimuli respectively. All results are averaged over all 5 Ss. Individual Ss showed the same trends as the group.

Figure 5 shows the results for identification of the features of place and voicing in the /ba/ - /ta/ series before and after adaptation with /da/. The results are consistent with those found for adaptation on the two single feature series. The voicing feature boundary shows no shift relative to the standard non-adapted function (see Figure 5B). This was expected since the /ba/ adapting stimulus had no effect in the voicing (/ba/ - /pa/) series and /da/ and /ba/ are both voiced.

The place feature boundary does show a large shift in the /ba/ - /ta/ series after adaptation with /da/ (see Figure 5A). Stimulus 6, which was identified as alveolar (/ta/) 60% of the time in the non-adapted condition is identified as alveolar only 14% of the time after adaptation. Unfortunately, Stimulus 7 was not presented for identification after adaptation. This makes precise comparison of the magnitude of the shifts in the place feature between the /ba/ - /ta/ and /ba/ - /da/ series impossible. However, the magnitudes of shift do appear to be roughly comparable. Additivity of feature shifts does seem to hold for this adapting stimulus.

The effects of adaptation with /pa/ on the /ba/ - /ta/ series also seem to support an additive interpretation. The voicing feature boundary shifts toward the voiceless end of the series, as shown in Figure 6B. The magnitude of this shift, about 7 msec VOT, is approximately the same as that found in the voicing feature series (about 5 msec VOT) with the same /pa/ adapting stimulus.

The place feature boundary also shows a shift toward the adapting (bilabial) stimulus (see Figure 6A). Again, the shift is roughly comparable in magnitude to that found on the place feature series (/ba/ - /da/) with /ba/ as the adapting stimulus.

The results of adapting the /ba/ - /ta/ series with the /ba/ stimulus are somewhat different. As expected, the voicing boundary shows no shift (see Figure 7B). The /ba/ adapting stimulus apparently has no effect on this feature boundary. This is the same result as that found for the /ba/ to /pa/ (voicing) series with the same /ba/ adapting stimulus.

The difference lies in the effect of the /ba/ adapting stimulus on the place feature boundary in the /ba/ - /ta/ series as shown in Figure 7A. After adaptation with /ba/ the identification function for the bilabial place feature does not show a simple shift toward the bilabial end of the series. Rather, the function dips to approximately 50% at Stimulus 3 and then rises again at Stimulus 4 before dropping off completely. This second drop in the bilabial identification function gives rise to a shift in the place feature boundary relative to the non-adapted function. The magnitude of this second place shift is approximately the same as the magnitude of the shift caused by adapting with /ba/ in the single place feature /ba/ - /da/ series. The drop at Stimulus 3 is almost entirely due to Ss identifying this stimulus as /da/ (alveolar place) 49% of the time (and /ba/ 48%, /pa/ 3%). At Stimulus 4 the /pa/ response is dominant (61%) which accounts for most of the reinstatement of the bilabial place feature. The shifts found for the different features in all three series are summarized in Table 1. The shifts in the voicing feature are given in msec of VOT. The shifts in

the place feature are given in terms of the proportion of one stimulus increment that the boundary shifts.

The shift found in the place feature as a resultof adaptation with /ba/ cannot be accounted for by an additive model. The adaptation of the bilabial feature appears to be conditional upon the particular value of the voicing feature in the adapting stimulus and in the test stimulus. There appear to be two shifts in the identification of the place feature when /ba/ is the adapting stimulus. The first occurs at a VOT value of approximately 20 msec. The value of the place feature at this point corresponds to the boundary point in the place feature for the single feature /ba/ - /da/ series with /ba/ as the adapting stimulus. However, at slightly larger VOT values (+30 msec) the bilabial response is reinstated. It is as if this value of VOT is too large to support a /da/ response and given that the stimulus is categorized as voiceless, the place value is not enough to support a /ta/ response.

These results offer further support to those found by Sawusch and Pisoni (in press). The features of place and voicing in CV syllables do not appear to be extracted separately. Rather, the extraction of the place feature, as indicated by a shift due to adaptation, appears to be dependent on the particular value of the voicing feature. This inter-dependence of the features of place and voicing in a consonant is in direct contrast to the classical assumption that the phonetic features of a consonant are processed independently. These results support a model of speech processing in which a particular phonetic feature analyzer has access to the acoustic cues for other phonetic features. The information

about other acoustic cues then serves to modify the processing of the directly relevant acoustic cues.

Conclusion

Evidence indicating that features are not processed independently in speech perception has been found using the selective adaptation paradigm. The shift found for the place feature in a bidimensional series is not dependent on only the value of place in the adapting and test stimuli. The value of the voicing feature for both test and adapting stimuli has a marked effect. These results suggest an exchange of information in the processing of features rather than independent and isolated processing for each feature.

References

Ades, A. E. A study of acoustic invariance by selective adaptation. Perception & Psychophysics, 1974, in press.

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. Journal of the Acoustical Society of America, 1974, in press.

Cooper, W. E. Contingent feature analysis in speech perception. Perception & Psychophysics, 1974, in press.

Cooper, W. E. & Blumstein, A "labial" feature analyzer in speech perception. Perception & Psychophysics, 1974, in press.

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

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

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

Liberman, A. M. Some characteristics of perception in the speech mode. In D. A. Hamburg (Ed.), Perception and its disorders, Proceedings of the A. R. N. M. D. Baltimore: Williams and Wilkins, 1970.

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

Liberman, A. M., 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, 5, 358-368.

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

Pollack, I. & Pisoni, D. B. On the comparison between identification and discrimination tests in speech perception. Psychonomic Science, 1971, 24, 6, 299-300.

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

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

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

Stevens, K. N. & Klatt, D. H. Role of formant transitions in the voiced-voiceless distinction for stops. Journal of the Acoustical Society of America, 1974, 55, 3, 653-659.

Table 1

The mean boundary shifts of the features of place and voicing averaged over Ss for each of the three adapting stimuli.[a]

Adapting Stimulus	Series / Feature			
	/ba/-/pa/ Voicing	/ba/-/da/ Place	/ba/-/ta/ Voicing	/ba/-/ta/ Place
/ba/	0.0 msec	1.07	0.0 msec	2.58, 1.05[b]
/pa/	5.0 msec	----	7.0 msec	1.45
/da/	--------	1.05	0.0 msec	1.00[c]

[a]Units for the voicing feature shift are msec of VOT. Shifts on the place feature are stated in terms of stimulus units used in the /ba/-/da/ series.

[b]Two boundaries appear for the place feature with /ba/ as the adapting stimulus.

[c]This is only an approximate value since there were not enough data points to determine the actual feature boundary.

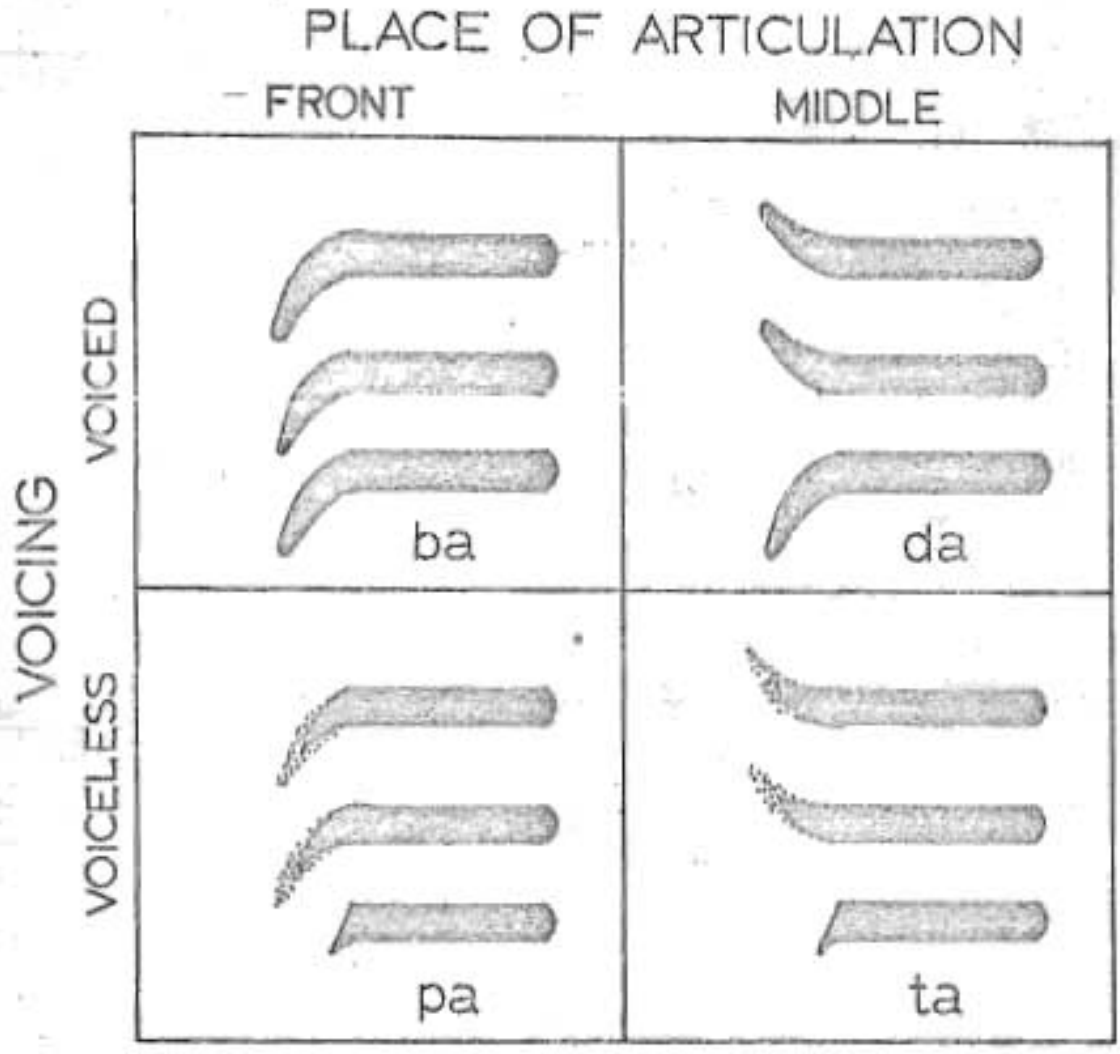


Figure 1. Schematized sound spectrograms of the syllables /ba/, /da/, /pa/, and /ta/ as used in the present experiments.

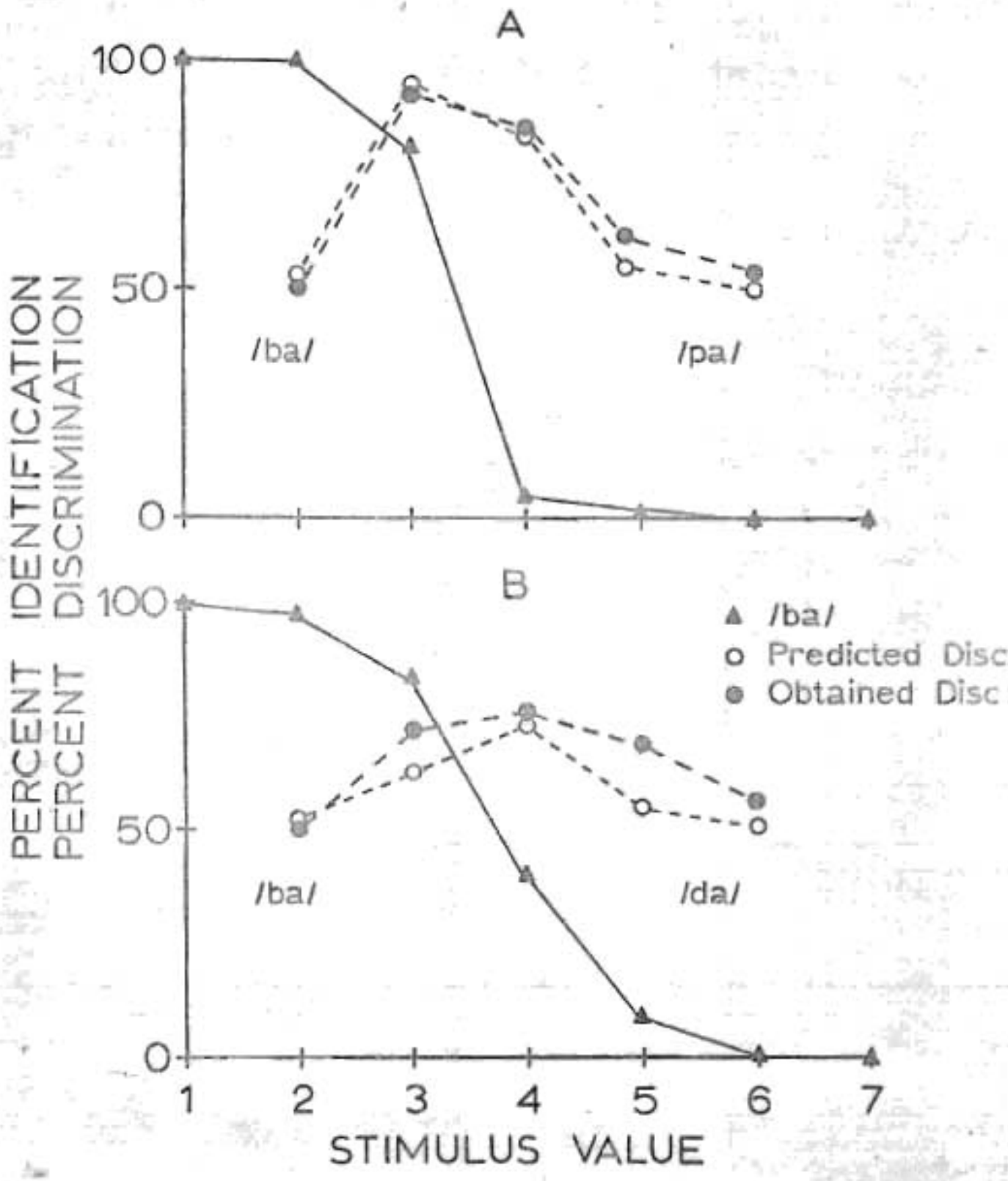


Figure 2. Group identification and ABX discrimination functions (observed and predicted) for the /ba/-/pa/ (A) and /ba/-/da/ (B) series.

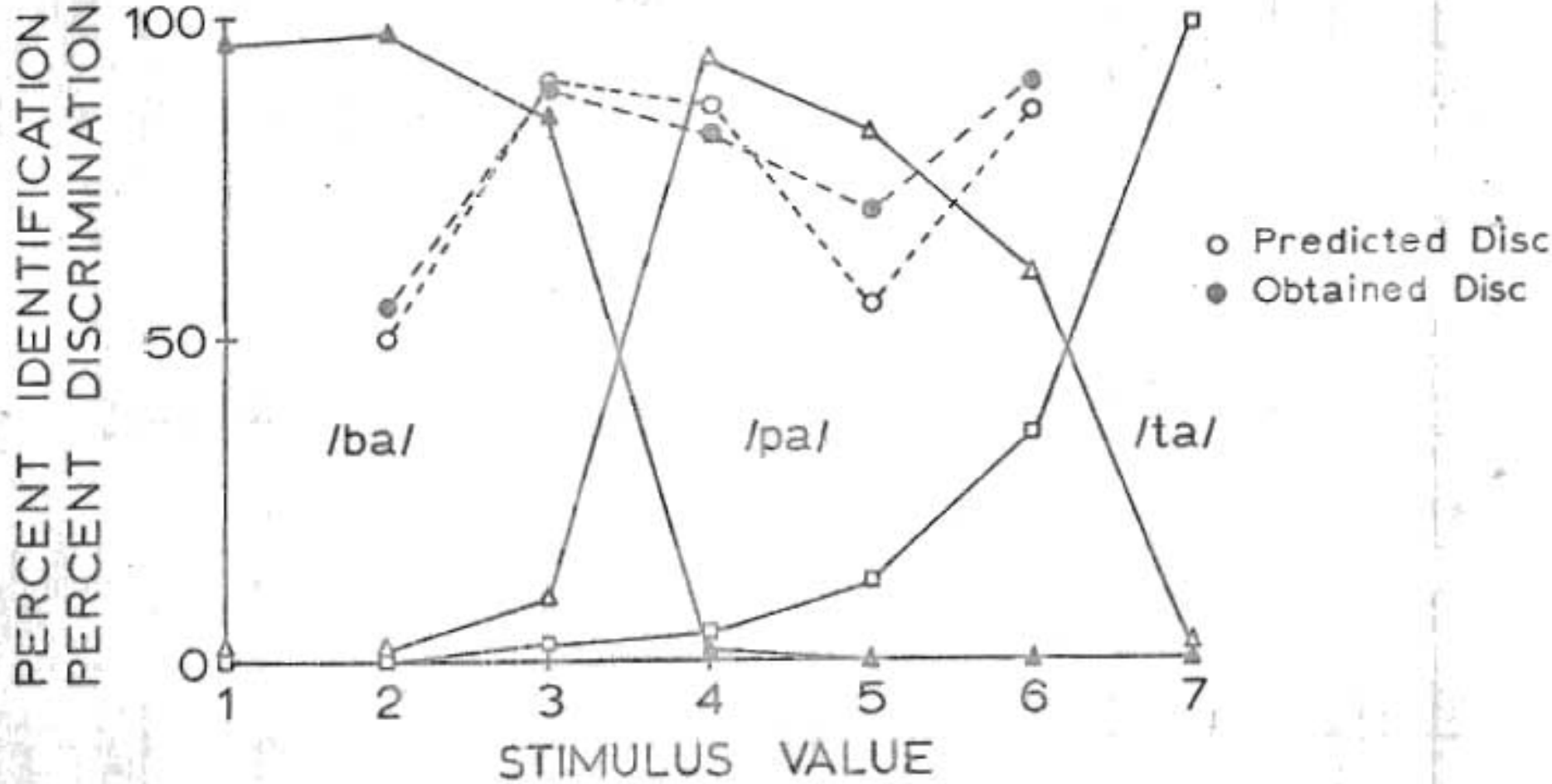


Figure 3. Group identification and ABX discrimination functions (observed and predicted) for the bidimensional /be/-/ta/ series.

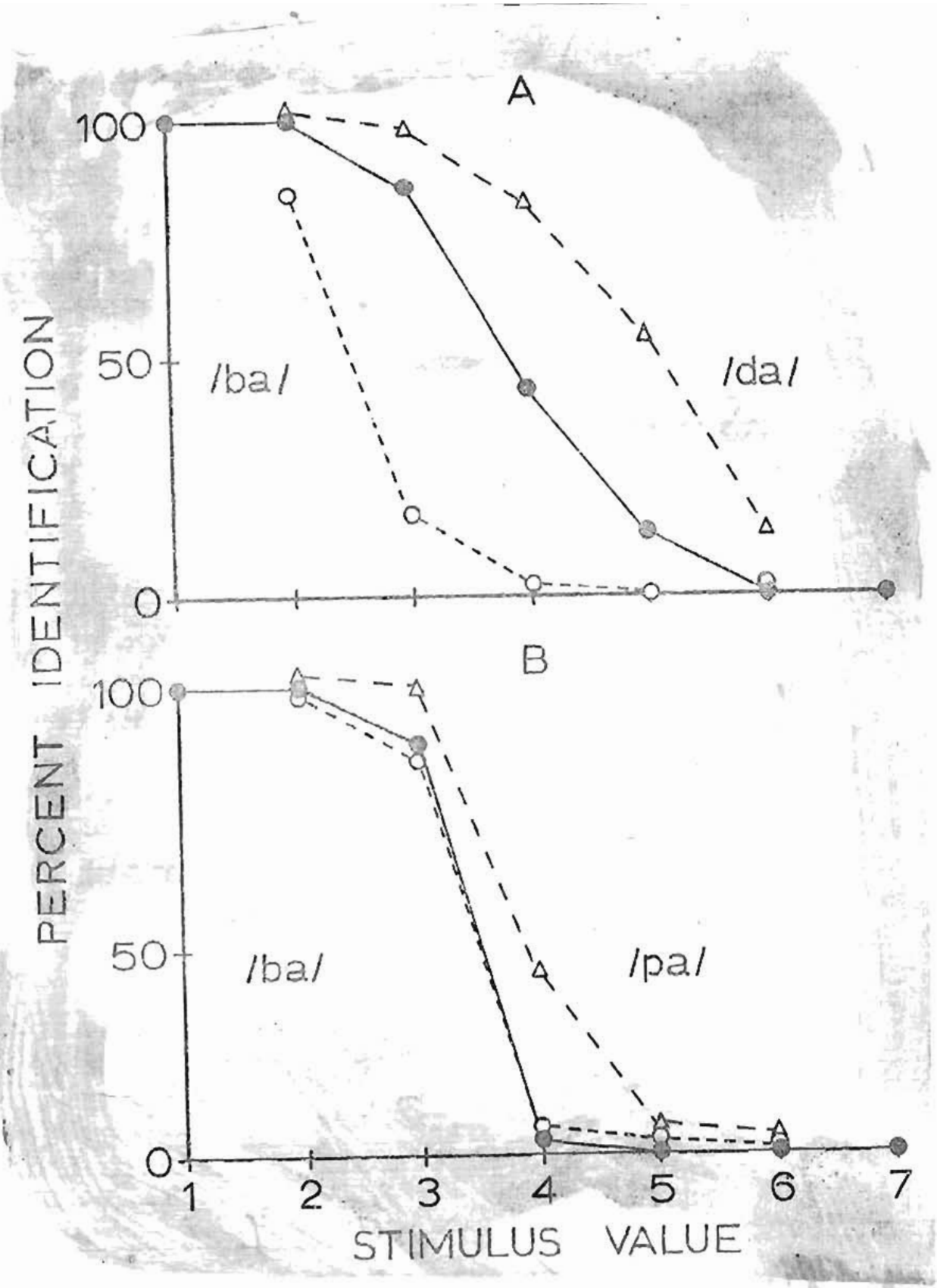


Figure 4. Group identification functions for the non-adapted and adapted series. Part A is for the /ba/-/da/ (place) series with /ba/ and /da/ adaptors. Part B is for the /ba/-/pa/ (voicing) series with /ba/ and /pa/ adaptors.

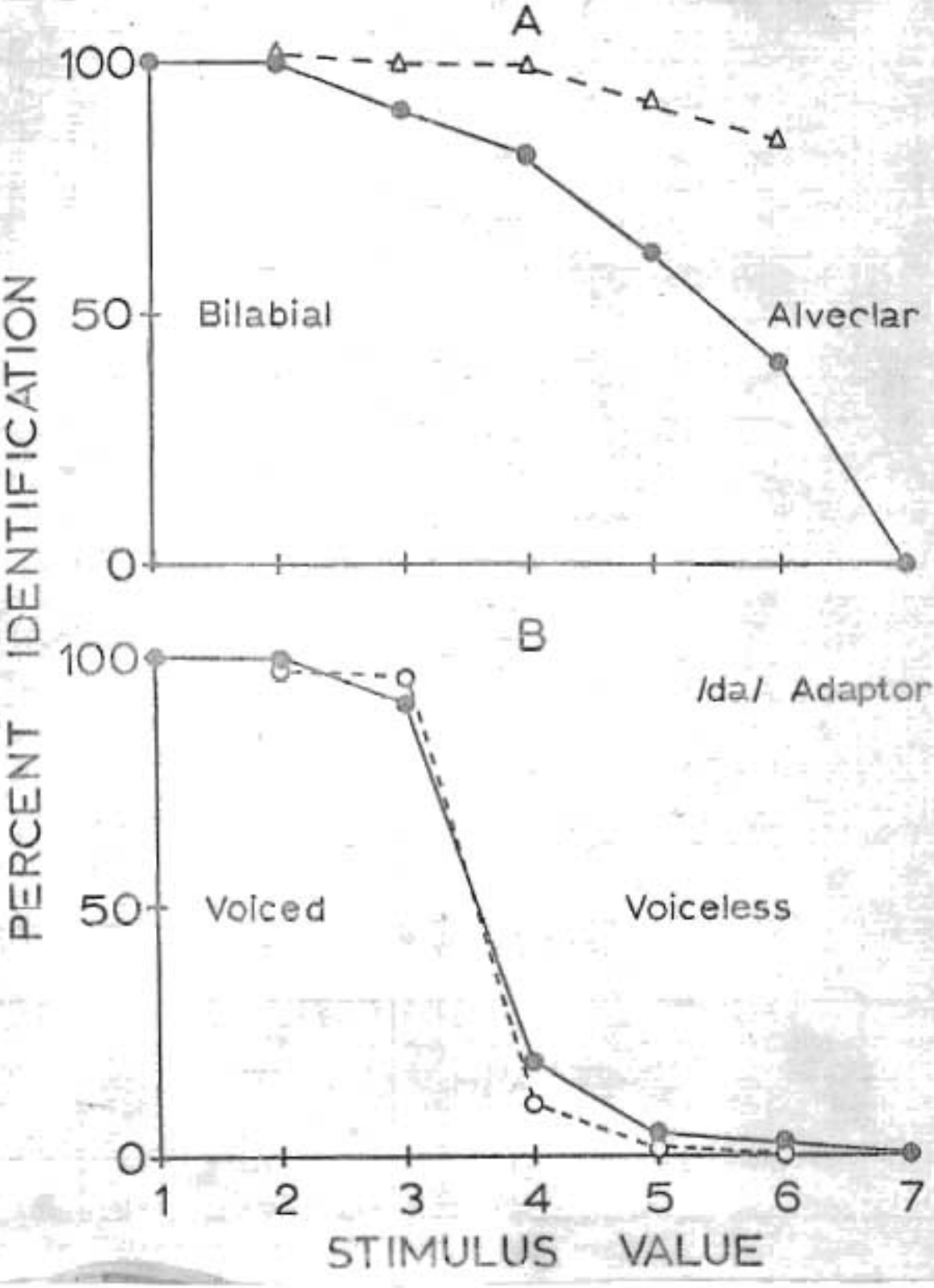


Figure 5. Group identification functions for the /ba/-/ta/ series. The features of place (A) and voicing (B) are shown unadapted and after adaptation with the syllable /da/.

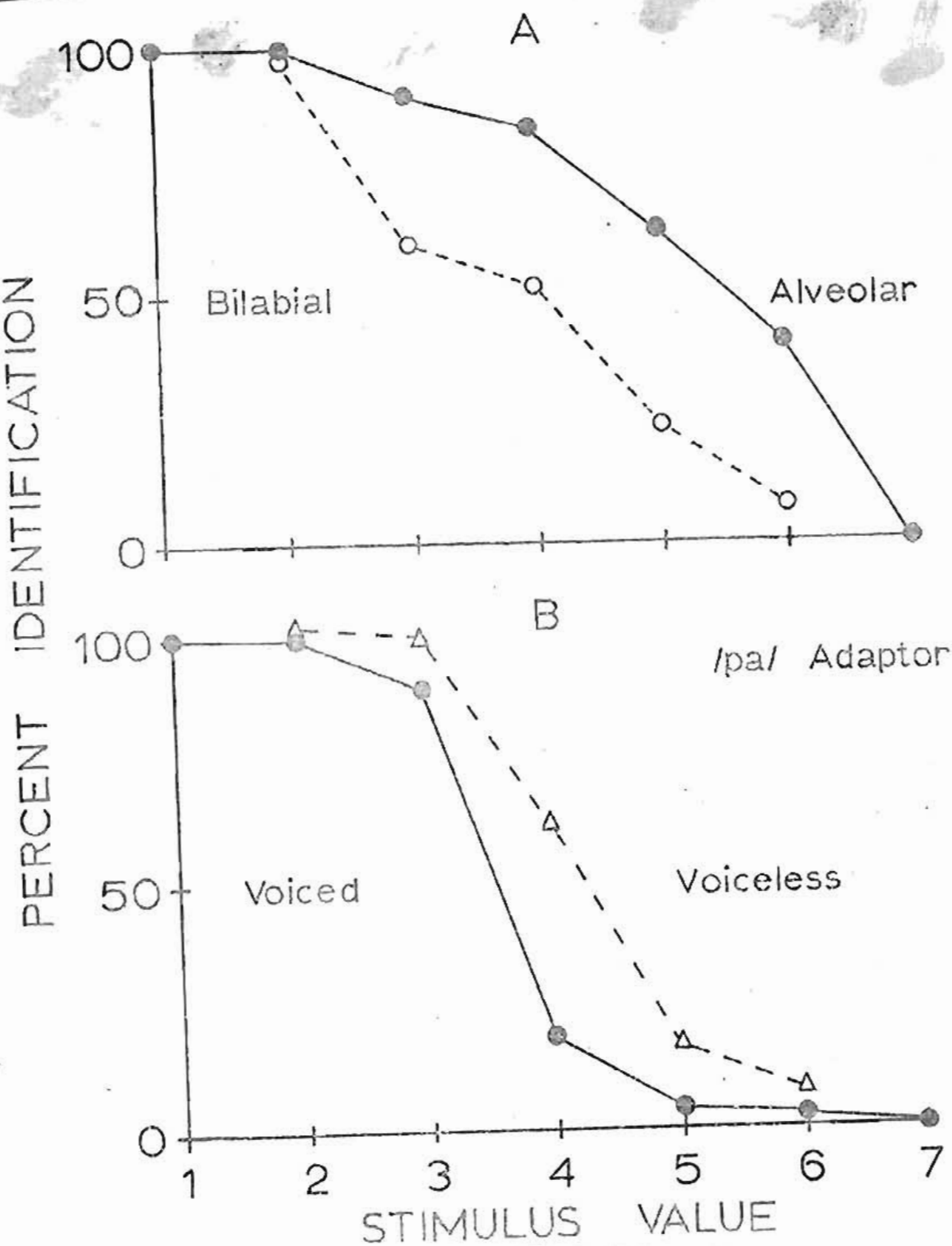


Figure 6. Group identification functions for the /ba/-/ta/ series. The features of place (A) and voicing (B) are shown unadapted and after adaptation with the syllable /pa/.

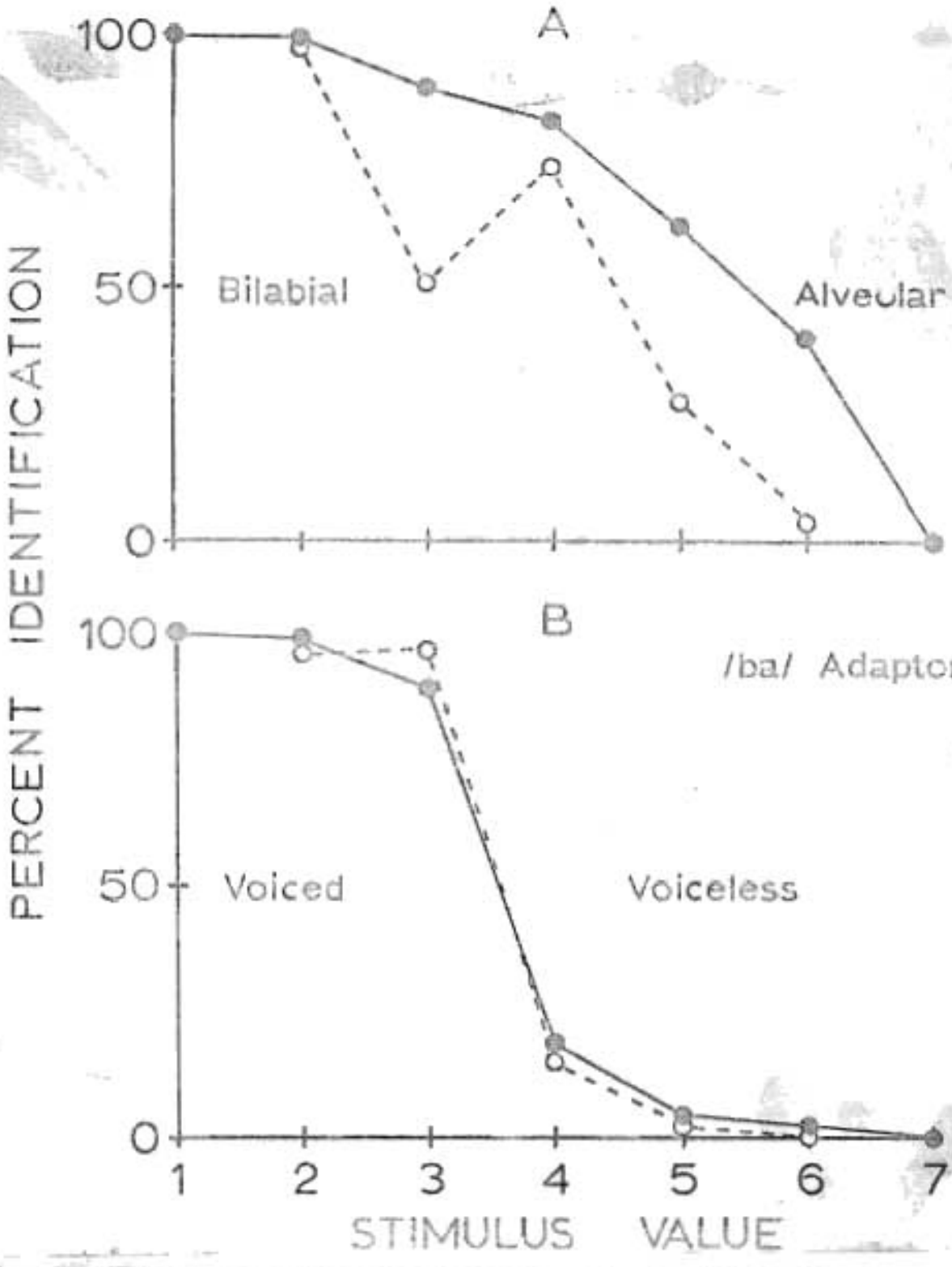


Figure 7. Group identification functions for the /ba/-/ta/ series. The features of place (A) and voicing (B) are shown unadapted and after adaptation with the syllable /ba/.

SHORT REPORTS AND CONVENTION PAPERS

Decision Processes in Speech Discrimination as Revealed by Confidence Ratings

D. B. Pisoni and David L. Glanzman
Department of Psychology
Indiana University
Bloomington, Indiana 47401

Standard speech discrimination tests require the listener to make a decision about a given sequence of stimuli. For example, with the ABX test the listener is required to determine whether the third stimulus is most like the first or most like the second. This test along with several other forced-choice procedures has the excellent property that the E need not specify the dimension along which the stimuli differ, e.g., select the stimulus with the higher pitch. The typical measure of the listener's performance is the percentage of stimuli correctly discriminated. However, in recent attempts to examine the processes underlying the identification and discrimination of speech sounds, it has become apparent that the listener's task, i.e., providing just a single response, may not accurately reflect all the information that the listener may have available to him about the stimuli. For example, sometimes a listener may be quite certain that his response was correct whereas other times the listener may be very uncertain about his response. In this study we examined how listeners assign confidence ratings to discrimination judgments for a set of synthetic stop consonants. The confidence ratings obtained with both ABX and 4IAX discrimination procedures carry additional information aobut the stimulus properties of consonants and provide some insight into the decision processes employed in the discrimination of these speech sounds.

Decision Processes in Speech Discrimination as Revealed by Confidence Ratings

D. B. Pisoni and David L. Glanzman

Indiana University

Bloomington, Indiana 47401

Recent work in speech perception has suggested that the perception of speech sounds may involve processes and mechanisms that are somehow different from those involved in the perception of other auditory stimuli (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Stevens & House, 1972). Support for this conclusion has come from several areas of investigation including: dichotic listening, identification and discrimination experiments, and more recently, from the study of speech perception in infants. In all of these experiments listeners are asked to make a decision about a stimulus or sequence of stimuli. For example, subjects may be asked to identify stimuli into categories defined by the experimenter, or to determine whether two sounds are the "same" or "different."

Several of the current theoretical approaches to speech perception assume that decisions are made relatively late during perceptual processing and that low-level acoustic information is unavailable due to constraints placed on the organism (Eimas & Corbit, 1973; Eimas, Cooper, & Corbit, 1973).

Insert Figure 1 about here

Figure 1 shows two possible hypothetical models of the processes involved in speech und perception. Both models are identical except for where the decision component may operate. In Model 1 decisions occur relatively late in time. This model assumes that early stages of processing are obligatory and automatic, and thus not under the control of the listener. In contrast, Model 2 assumes that decisions are made at all stages of analysis. Low level acoustic information may be accessible to subjects, although this will depend on a variety of factors including: the task demands, state of the organism, and the particular criterion employed by the listener.

In the present study we examined how listeners assign confidence ratings to discrimination judgments for synthetic speech sounds. We were concerned with three basic questions. First, are the discrimination functions obtained with confidence ratings comparable to those functions obtained without confidence ratings? Second, can listeners in a speech discrimination task assign confidence ratings in a non-chance manner which is related to their observed discrimination performance? And third, do the confidence ratings provide some additional insight into the decision processes in speech sound perception?

To answer these questions we looked at discrimination functions under several experimental conditions for a set of synthetic stop conconant vowel syllables.

Method

The stimuli we used were a set of seven three-formant patterns appropriate for the initial bilabial stop series. They were produced

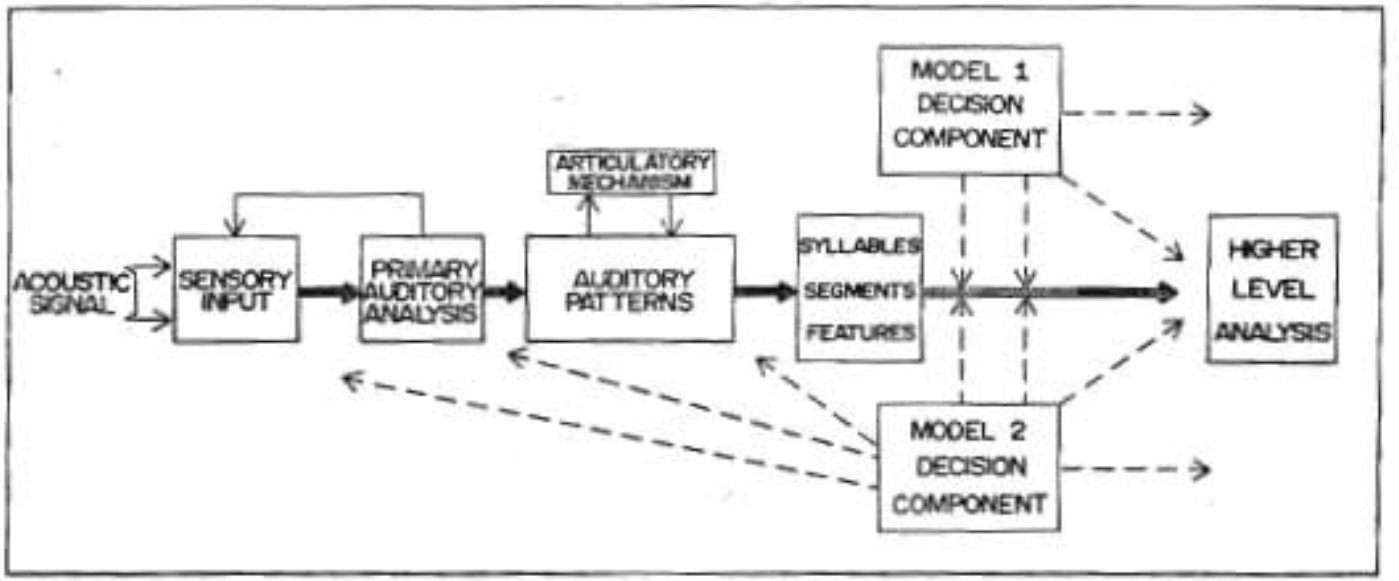


Figure 1. Outline of hypothetical model of speech perception process showing where decision component may operate.

on the parallel resonance synthesizer at Haskins Laboratories and were similar to those employed by Lisker and Abramson (1967). The seven stimuli varied in 10 msec steps along the voice onset time continuum from 0 msec VOT through + 60 msec VOT. The stimuli were recorded on magnetic tape to produce two types of discrimination tests, a standard ABX test and the 4IAX test of paired similarity (Pisoni, 1971; Pisoni & Lazarus, 1974).

Insert Figure 2 about here

The details of these two discrimination tests are shown in Figure 2. All possible pairs of stimuli one and two-steps apart along the continuum were arranged in either an ABX or 4IAX format. In the ABX test, subjects are required to determine whether the third stimulus is most like the first or most like the second. In the 4IAX test, the subject is required to determine which pair of stimuli was the same; the first or the second pair.

Eight subjects who were all undergraduate students at Indiana University were assigned to each of the two conditions of discrimination.

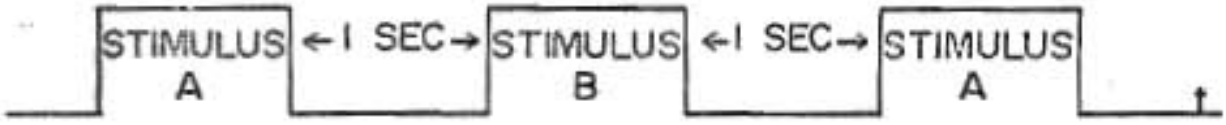
Insert Figure 3 about here

Figure 3 shows the confidence rating scale used in the discrimination tests. After making a discrimination judgment, subjects were required to rate how confident they were in their decision on a four point scale from (+++), positive my response is correct, to (-), my response represents no better than a chance guess.

Subjects were run for an hour a day on three days. Each session

DISCRIMINATION TESTS

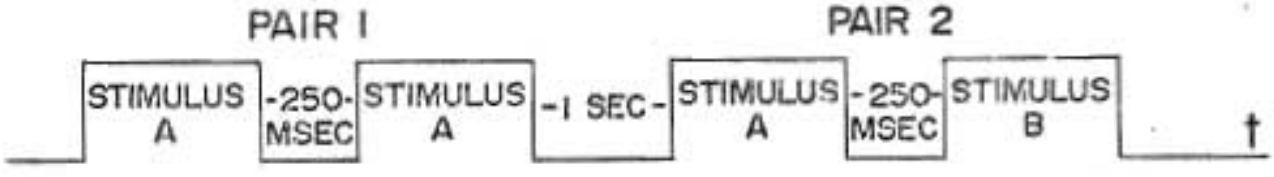
1) ABX TEST - PAIRS OF STIMULI ARRANGED IN TRIADS: ABA, BAB, ABB, BAA



QUESTION: IS THE THIRD STIMULUS MOST LIKE THE FIRST OR SECOND STIMULUS?

RESPONSE: FIRST STIMULUS!

2) 4IAX TEST - TWO PAIRS OF STIMULI ARE PRESENTED ON EACH TRIAL. ONE PAIR IS ALWAYS THE SAME AND ONE PAIR IS ALWAYS DIFFERENT: A-A—A-B, A-B—A-A, A-A—B-A, ETC.



QUESTION: WHICH PAIR WAS MOST SIMILAR - THE FIRST PAIR OR THE SECOND PAIR?

RESPONSE: FIRST PAIR!

Figure 2. Details of the two types of discrimination tests employed; the ABX test and the 4IAX test of paired similarity.

CONFIDENCE RATING SCALE IN DISCRIMINATION

+++	POSITIVE THAT MY RESPONSE IS CORRECT
++	FAIRLY CERTAIN THAT MY RESPONSE IS CORRECT
+	CAN'T DECIDE BUT THINK MY RESPONSE IS CORRECT
−	RESPONSE REPRESENTS NO BETTER THAN CHANCE GUESS

Figure 3. Confidence Rating scale used for discrimination judgments.

began with a standard identification test followed by a discrimination test.

Insert Figure 4 about here

Figure 4 shows the identification functions for the two groups of subjects.

Both groups divided the continuum into two discrete categories. It may be observed that we obtained the usual type of categorical partitioning of the continuum. The phonetic boundary is at about +30 msec VOT.

Insert Figure 5 about here

Figure 5 shows the percent correct discrimination functions for the ABX and the 4IAX tests ignoring for the moment any of the confidence rating data.

These discrimination functions are quite similar to those obtained under the usual conditions of discrimination without confidence ratings. There is a peak between phonetic categories and a trough within categories. Note that discrimination is better in the /ba/ range along the continuum than the /pa/ range, especially for the 4IAX data.

Based on only percent correct discrimination, then, our subjects show "categorical-like" discrimination of the voicing feature. But when we look at how our subjects' confidence ratings are related to observed discrimination performance, the story becomes more interesting.

Insert Figure 6 about here

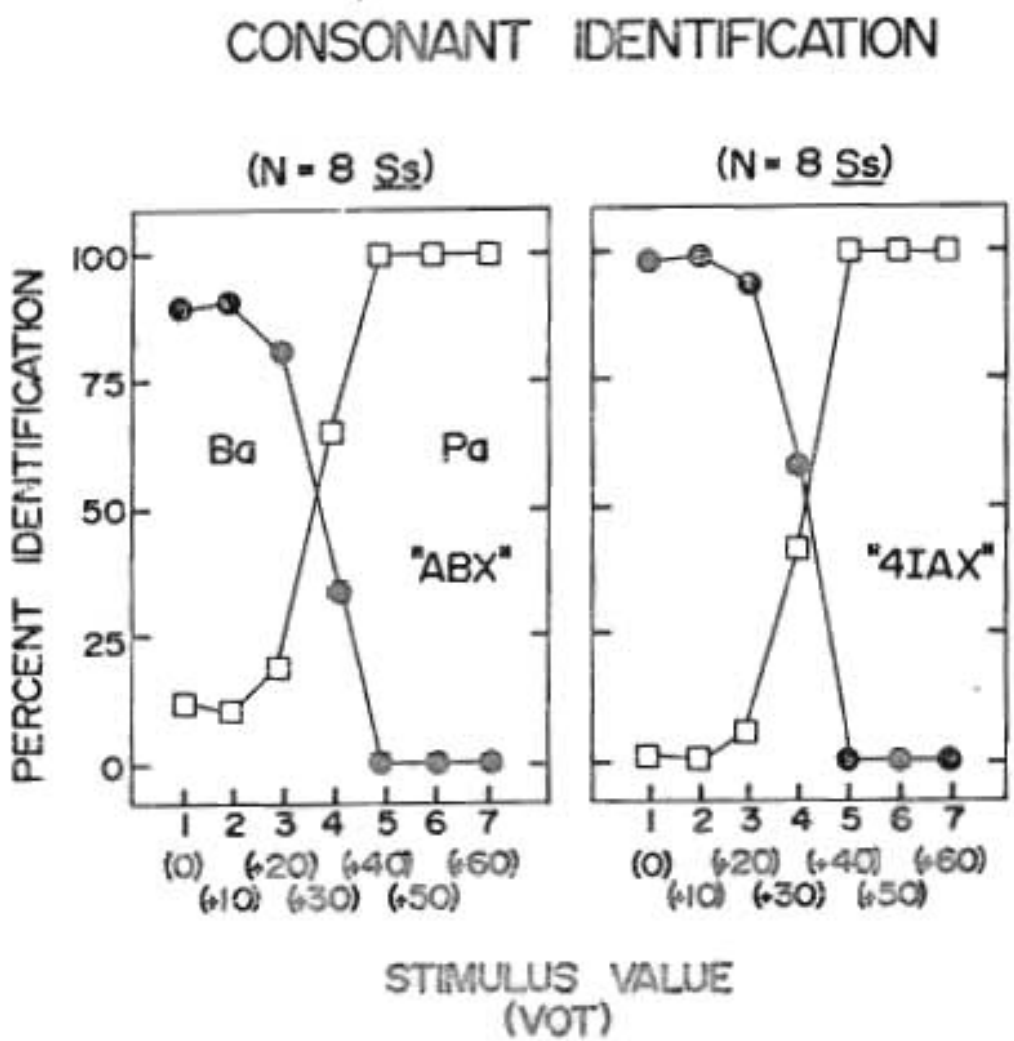


Figure 4. Average consonant identification functions for two groups of listeners.

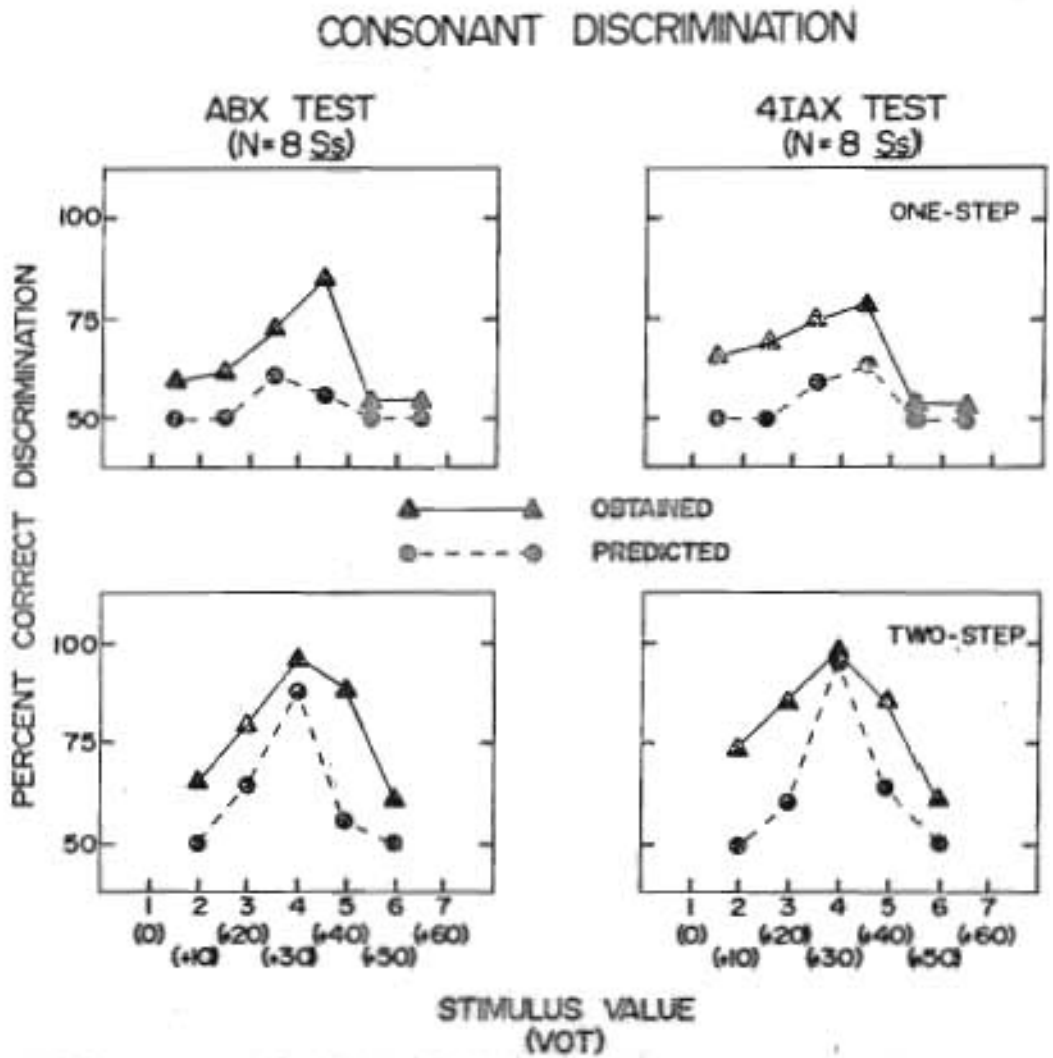


Figure 5. Average consonant discrimination functions for the ABX and 4IAX test conditions.

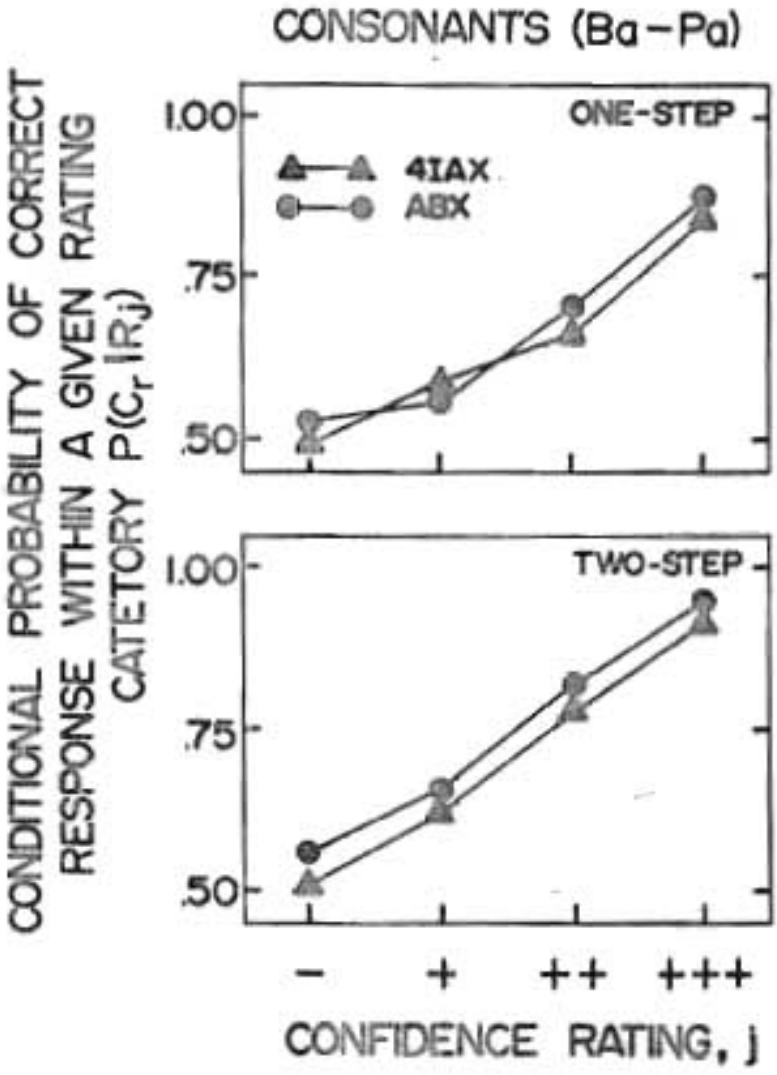


Figure 6. Relationship between a correct response and rating category averaged over the VOT continuum.

This figure shows the probability of a correct response within a given rating category for the one-step and two-step discrimination data. The ordinate is the proportion of correct responses out of the the total number of responses assigned within each rating category.

These data indicate that confidence rating assignment is systematically related to discrimination accuracy. Clearly, listeners can assign confidence ratings in a non-chance manner.

Insert Figure 7 about here

Figure 7 shows the conditional probability of a correct response within a particular rating category across the VOT stimulus continuum. The probability of being correct is much higher within categories when the subjects are very confident (i.e., +++) than might be expected on the basis of the standard percent correct discrimination measure. Furthermore, as subjects become less and less confident in their judgment, within category performance appears to decrease more than between category performance. This suggests that a change in criterion does not affect all stimulus comparisons along this continuum equivalently.

To summarize, we have found answers to each of the three questions we raised at the beginning of this paper. First, when subjects are required to use confidence ratings in speech discrimination, the discrimination functions based on percent correct scores are quite similar to those found without confidence ratings. Second, confidence rating assignment is systematically related to overall accuracy of discrimination. Better performance, therefore, leads to assignment of higher confidence ratings. Moreover, when high confidence ratings are assigned, performance within categories is far superior to that normally expected on the basis

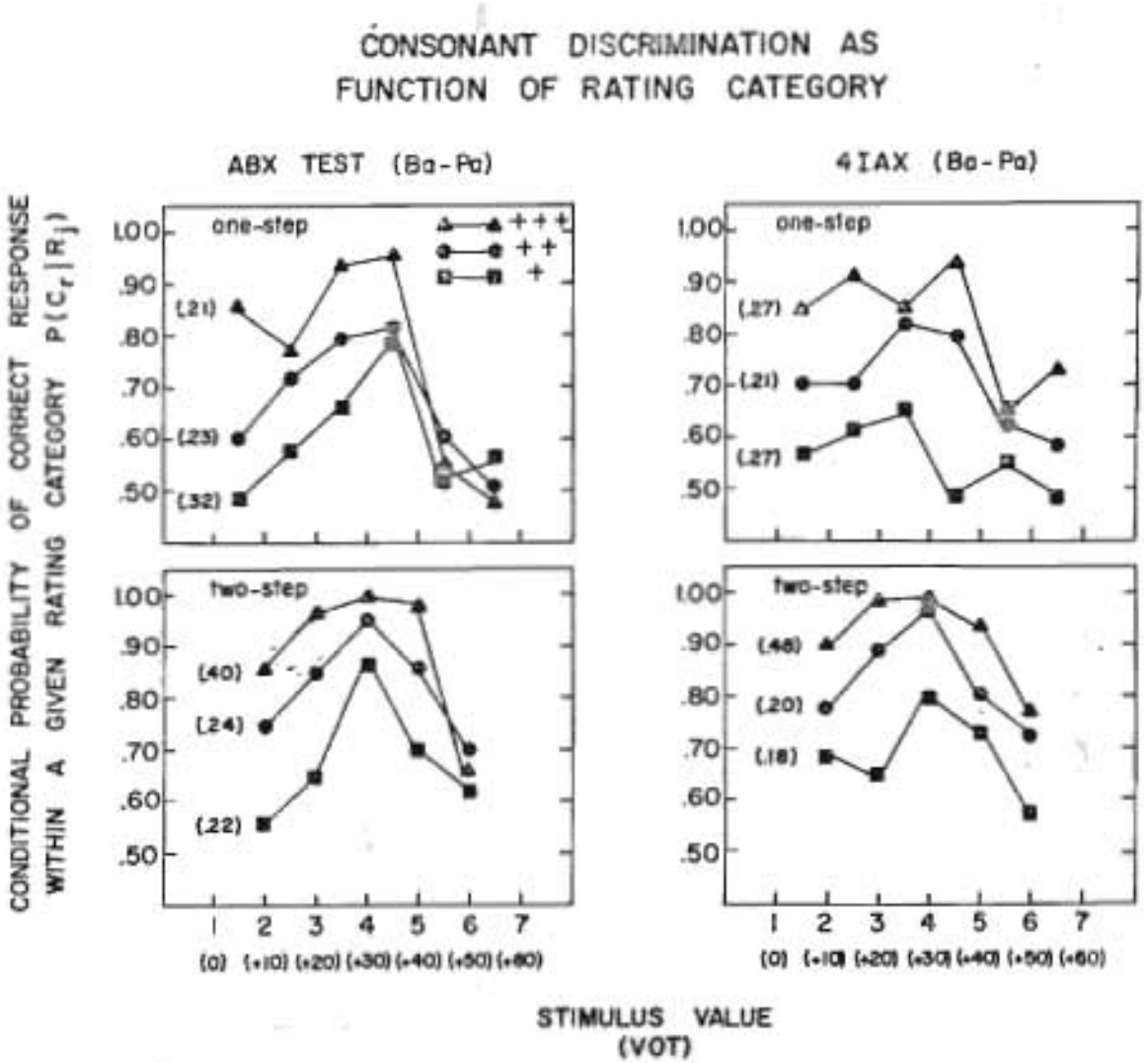


Figure 7. Consonant discrimination at a given rating category as a function of VOT across the continuum.

of percent correct discrimination scores alone.

These findings have several implications for understanding speech perception, especially perception of the voicing feature. Models which assume that decisions are made relatively late in perceptual processing would appear to be inadequate to account for the present results. Rather, models of the type described earlier, which assume that decisions may take place at many levels of processing, would appear to be more consistent with our findings. Thus, categorical perception, at least along the voicing continuum, may therefore be the consequence of the decision rules employed in discrimination rather than a limitation on the sensory (low-level) capacities of the organism. In line with some of Steven's (1972) remarks, it would seem reasonable to suppose that different decision rules are employed in discrimination when the sitmuli are drawn from within the same phonetic category than when they are drawn from across phonetic categories.

REFERENCES

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

Eimas, P. D., Cooper, W. E., and Corbit, J. D. Some properties of linguistic feature detectors. Perception and Psychophysics, 1973, 13, No. 2, 247-252.

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

Lisker, L., and Abramson, A. S. The voicing dimension: Some experiments in comparative phonetics. In Proceedings of the Sixth International Congress of Phonetic Sciences, Prague, 1967. Prague: Academia, 1970, Pp. 563-567.

Pisoni, D. B. On the nature of categorical perception of speech sounds. Status Report on Speech Research (SR-27), Haskins Laboratories. New Haven, 1971, Pp. 101.

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

Stevens, K. N. The quantal nature of speech: Evidence from articulatory-Acoustic data. In E. E. David, Jr. and P. B. Denes (Eds.) Human Communication: A Unified View. New York: McGraw-Hill, 1972, Pp. 51-66.

Stevens, K. N. and House, A. S. Speech perception. In J. Tobias (Ed.) Foundations of Modern Auditory Theory. Volumn II. New York: Academic Press, 1972.

"Same-Different" Reaction Times to Consonants, Vowels and Syllables*

D.B. Pisoni and J. Tash
Department of Psychology
Indiana University
Bloomington, Indiana 47401

It has been proposed that phonetic segments are encoded into perceptual units of syllable size. The aim of this study is to investigate how consonant and vowel information is transmitted (i.e., serial vs. parallel), and to what extent auditory and phonetic levels of analysis may interact in perceptual analysis. Synthetic CV syllables, /ba/, /bae/, /da/, and /dae/, were arranged to produce four types of stimulus pairs: 1) consonant same - vowel same (CS-VS, e.g., /ba/-/ba/), 2) consonant same - vowel different (CS-VD), e.g., /ba/-/bae/, 3) consonant different - vowel same (CD-VS, e.g., /ba/-/da/), and 4) consonant different - vowel different (CD-VD, e.g., /ba/-/dae/). Same-different reaction times (RTs) were obtained separately for consonants, vowels, and syllables. In the consonant and vowel conditions, RTs for "same" responses were faster when the irrelevant feature was the same (i.e., CS-VS); "different" response RTs were faster when the irrelevant feature was different (CD-VD). In the syllable condition, RTs for different responses were faster when the vowel was different and the consonant was the same (i.e., CS-VD) than when the consonant was different and the vowel was the same (i.e., CD-VS). This suggests that vowel information may be more easily accessible for a comparison than consonant information. These results provide evidence for the parallel transmission of information about stop consonants and vowels in syllables, and an interaction of auditory and phonetic levels of processing speech perception.

* This is a draft of a paper to be presented at the 86th meeting of the Acoustical Society of America, November 1, 1973, Los Angeles, California. The research was supported in part by PHS grants S05 RR 7031 and MH 24027 to Indiana University and a grant from NICHD to Haskins Laboratories. We wish to thank Professor Frank Restle for the use of his computer in conducting this work.

"Same-Different" Reaction Times to Consonants, Vowels and Syllables

D.B. Pisoni and J. Tash

Indiana University

What are the basic units of perceptual analysis in speech sound perception? Are they syllables, segments or features? Obviously each of these units is in some sense psychologically "real" since their existence has been verified repeatedly in experiments dealing with speech perception and production. But the important theoretical question now is how are these units represented at various levels of perceptual analysis. It is generally agreed that some units are more abstract than other units although the nature of these arguments have not been detailed very precisely. For example, syllables are usually thought to be less abstract than phonetic segments because they exist as both articulatory and acoustic units. On the other hand, phonetic segments are considered to be more abstract because they are, in general, not directly represented by sound segments in the speech signal.

It has been suggested that the syllable is the carrier of phonetic information and that phonetic segments are encoded into perceptual units of syllable size. Moreover, it has been suggested by a number of investigators that the acoustic information in a syllable is simultaneously providing information about two or more segments at the same time. In other words, there is a form of parallel transmission of information about phonetic

segments encoded in syllables. If this is true then the processing of one phonetic segment may be effected concurrently with the processing of another segment within the same syllable. It follows from this that decisions about specific segments may require the use of information distributed over an entire syllable.

In the present study we were concerned with the way consonant and vowel information is represented within a syllable and the types of decisions that can be made when different levels of perceptual analysis are required. The technique we used to explore this problem was a "same"-"different" reaction time task. On any trial two synthetic CV syllables were presented to a subject and he was required to determine whether the two stimuli were the "same" or "different." Reaction times were obtained under three separate conditions: (1) comparison of syllables, (2) comparison of vowels, and (3) comparison of consonants.

Slide 1 please

Method and Procedure

Slide 1 shows the stimulus conditions used in the present experiment. The cv syllables /ba/, /da/, /bae/ and /dae/ were arranged in all possible pairs to produce four experimental conditions: (1) consonant same - vowel same, (2) consonant same - vowel different, (3) consonant different - vowel same and (4) consonant different - vowel different. The two cells marked with an asterisk, the CS-VS and the CD-VD cells, are completely redundant with the required response. Thus, the CS-VS cell always required a "same" response and CD-VD cell always required a "different" response. Responses

to the other two cells varied depending on whether subjects were comparing syllables, vowels or consonants.

For the syllable condition, Ss were told to respond "same" only if both members of a stimulus pair were identical syllables, they were told to respond "different" if the two syllables differed in any way. For the vowel condition they were told to respond "same" only if the vowels in each syllable were identical and to respond "different" if the vowels were different. They were told to ignore the consonants. Similar instructions were used in the consonant condition. The order of presentation was counterbalanced across six groups of four subjects each. The stimuli were 300 msec. three-formant patterns produced on the synthesizer at Haskins Laboratories. All responses and reaction times were recorded automatically under the control of an IBM 1800 computer. Reaction times were measured from the offset of the last stimulus.

Results and Discussion

A number of very interesting results were obtained in this experiment but because of time limitations we will only describe the major findings and leave some of the finer details for another time.

Slide 2 please

Slide 2 shows the mean reaction times for "Same-Different" responses for each of the three conditions: syllables, vowels and consonants. "Same" responses are faster than "different" responses in each condition. More importantly, however, decisions about syllables are consistently faster than vowels and decisions about vowels are consistently faster than decisions

about consonants. This is true for both "same" and "different" responses. Note in particular that "same" responses to syllables are unusually fast suggesting that this decision may occur even before the second syllable has terminated. Thus, there is a hierarchy in terms of the decision times from syllable to syllable nucleus to consonant. This is not the whole story since in this slide we have collapsed over all "same" and "different" conditions.

Slide 3 please

In this slide we now have the reaction times broken down into the four stimulus conditions. In the syllable condition we can see that "different" responses are ordered systemtically. Reaction time is fastest when both the consonant and vowel differ, somewhat slower when only the vowels differ and slowest when only the consonants differ within a syllable. Thus, it comes as no surprise to find that consonant and vowel information is processed differently even in the syllable condition. But let us move to the data for the vowel and consonant conditions for a moment. Note that the redundant cells, those marked with an asterisk are consistently faster than the non-redundant cells and this is true in both the vowel and consonant conditions. Thus, when a subject is required to compare two vowels in a CV syllable, processing of the consonant apparently interferes with the decision about the vowel. This result is also true when the subject is required to compare only the consonants. Processing of the vowel also interferes with the decision about the consonant. This difference is in the order of about 100 msec. for the "same" responses but only around 30 msec. for the "different"

responses. These findings which are quite similar to Day and Woods' results with an identification paradigm suggest that the acoustic information for the consonant and vowel segments in a CV syllable may be transmitted simultaneous and in parallel. If the information were not transmitted in a parallel form we would not expect differences in consonants to affect the vowel decision and differences in vowels to affect the consonant decision.

How can we summarize these results on reaction time to syllables and segments encoded within syllables? We do not have an exact model worked out yet to handle all these results but we think we have some idea as to what the model might look like.

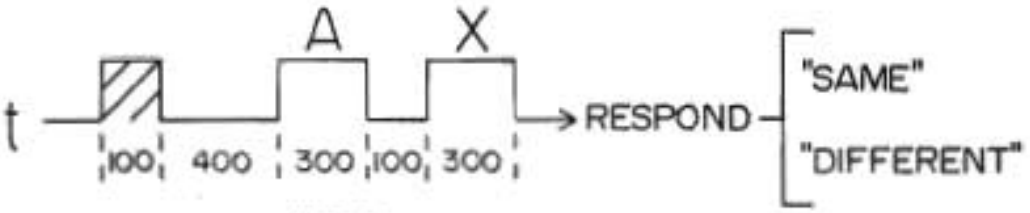
Slide 4 please

In this flow chart we have indicated some of the stages and operations that may be involved in "same"-"different" task. The key to this model is the notion of "depth of processing." We think that the reaction times in the syllable condition can be explained by a relatively simple decision rule at a very early stage of analysis. For example, at Stage 2 the overall gross acoustic similarity of the pair of stimuli is evaluated against a criterion. For the vowel and consonant conditions additional stages and operations must be proposed. We have two explanations for the finding that vowels can be compared more rapidly than consonants. First, the information needed for a decision about the vowels may be more readily available than the information in the consonants. Second, the vowel comparison may take place at an earlier stage of processing than the consonant comparison. Experiments dealing with both possibilities are currently underway at Indiana.

In summary, "same"-"different" reaction times were obtained to pairs of CV syllables. When Ss were required to compare syllables their decisions were faster than when they were required to compare specific segments within syllables. Reaction-time to vowel segments was faster than reaction time to consonants. Moreover, the fact that variations in consonants affect processing of vowels and variations in vowels affect processing of consonants was taken as evidence for the parallel transmission of acoustic information for consonant and vowel segments within syllable sized perceptual units.

STIMULUS CONDITIONS

TIMING:



STIMULI:

/ba/, /da/, /bæ/, /dæ/

EXPERIMENTAL CONDITIONS:

	VOWEL SAME	VOWEL DIFFERENT
CONSONANT SAME	ba-ba da-da bæ-bæ dæ-dæ (CS-VS) *	ba-bæ da-dæ bæ-ba dæ-da (CS-VD)
CONSONANT DIFFERENT	ba-da da-ba bæ-dæ dæ-bæ (CD-VS)	ba-dæ da-bæ bæ-da dæ-ba (CD-VD) *

Figure 1.

MEAN RTs (msec)
"SAME" - "DIFFERENT" RESPONSES
(N = 24 Ss)

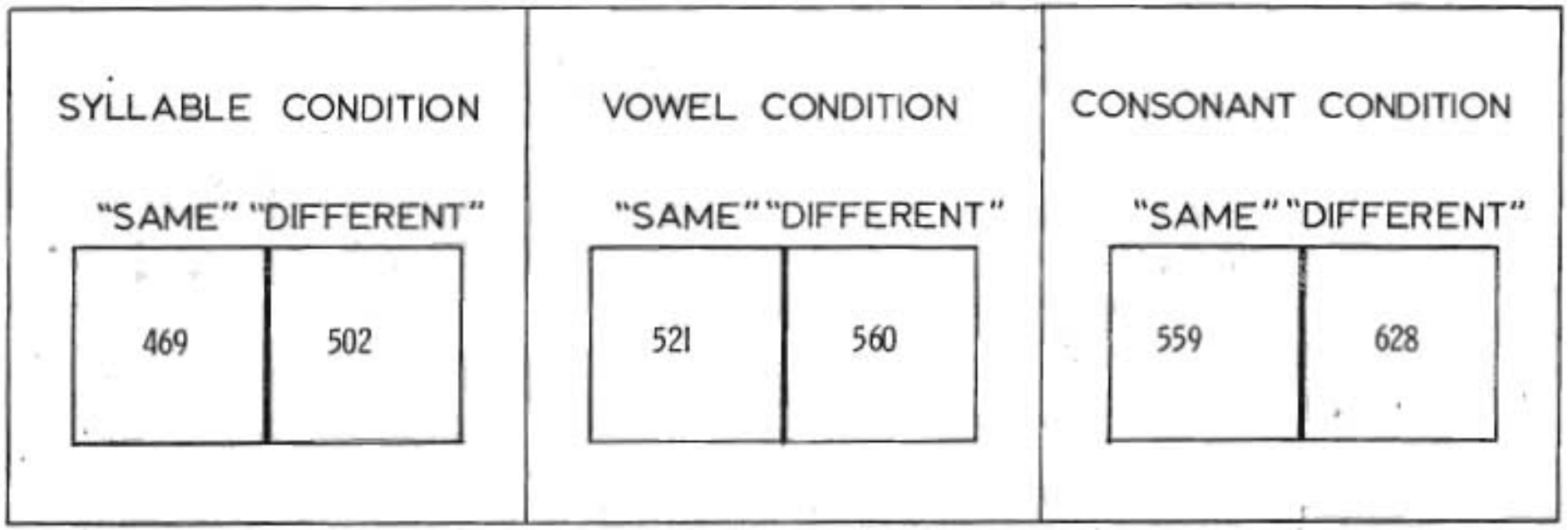


Figure 2.

MEAN RTs (msec)
"SAME" - "DIFFERENT" RESPONSES
(N=24 Ss)

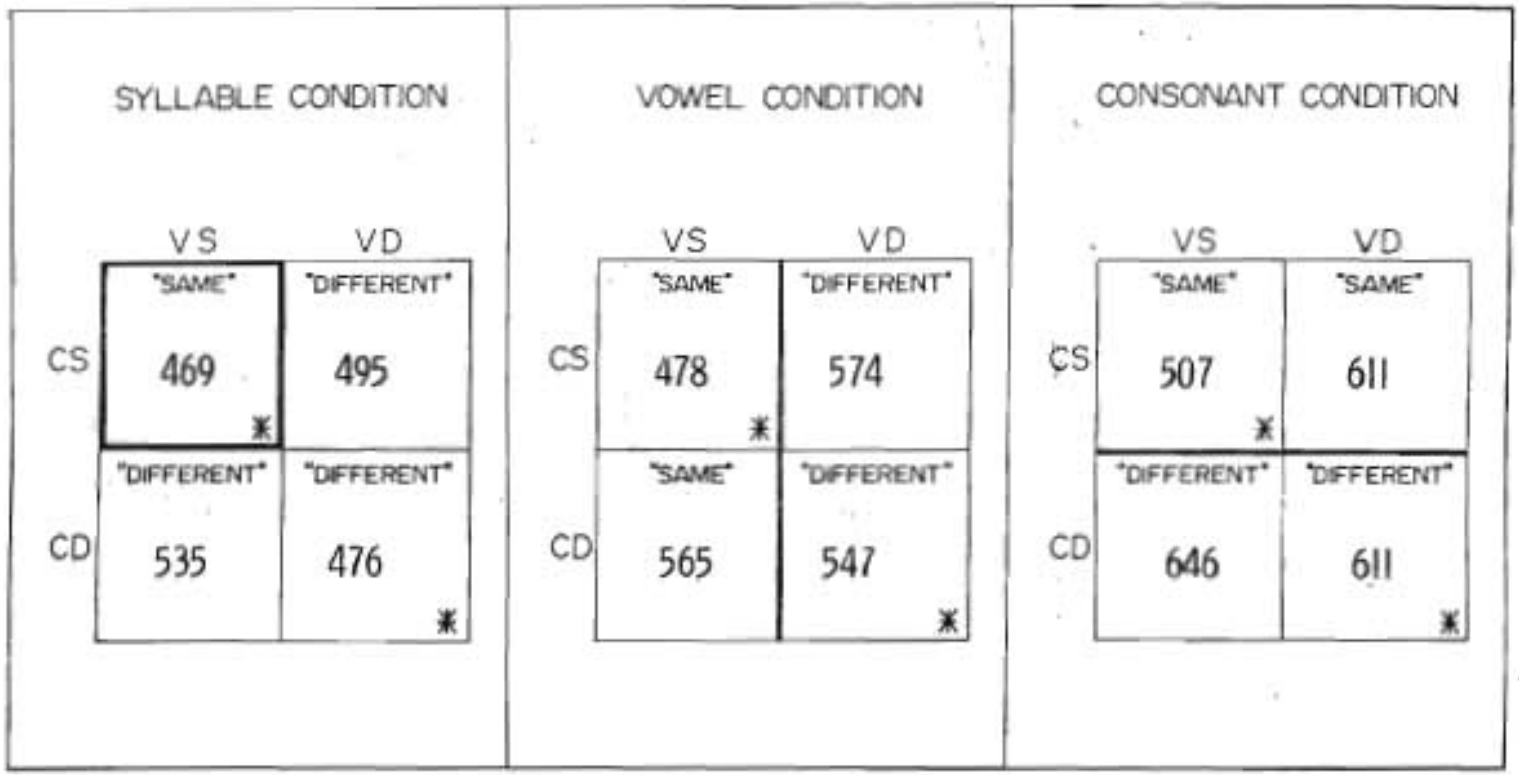


Figure 3.

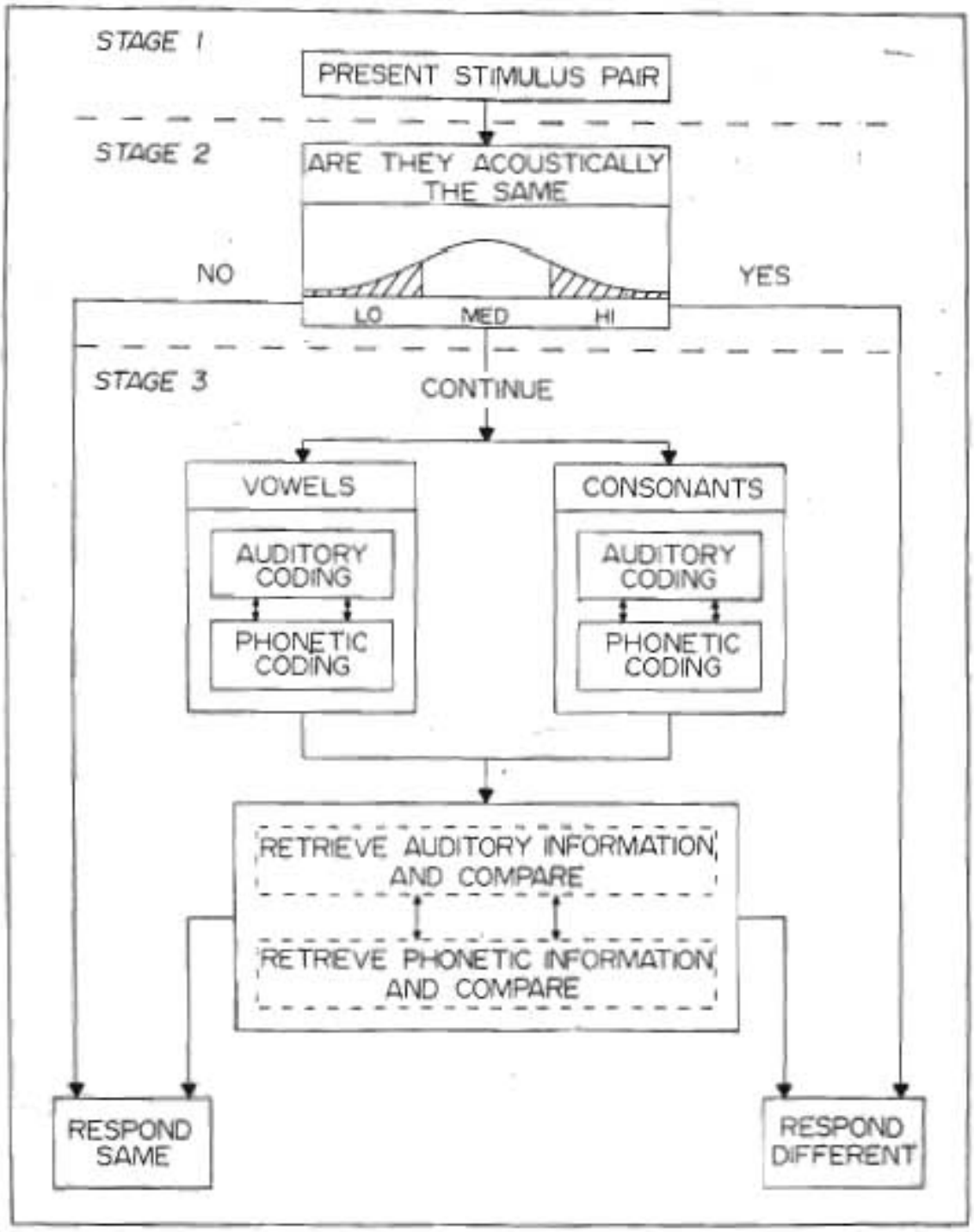

STAGE 1
PRESENT STIMULUS PAIR
STAGE 2
ARE THEY ACOUSTICALLY THE SAME
NO
YES
LO
MED
HI
STAGE 3
CONTINUE
VOWELS
AUDITORY CODING
PHONETIC CODING
CONSONANTS
AUDITORY CODING
PHONETIC CODING
RETRIEVE AUDITORY INFORMATION AND COMPARE
RETRIEVE PHONETIC INFORMATION AND COMPARE
RESPOND SAME
RESPOND DIFFERENT

Figure 4.

Category Boundaries for Speech and Nonspeech Sounds*

J. R. Sawusch and D. B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47401

Recent experiments employing an adaptation paradigm have demonstrated the possibility of phonetic feature detectors in speech perception (Eimas and Corbit, 1973). These results could be explained by Adaptation Level Theory (ALT), assuming the reference for categorizing a speech sound is external. In the present experiment, identification functions were obtained for a series of synthetic speech sounds ranging perceptually from /ba/ to /pa/ and a series of tones varying in intensity from 60 to 84 dB SPL. The distribution of occurrences of each stimulus in a series was varied for both tones and speech stimuli. The category boundaries for the tones shifted as a function of the relative number of occurrences of each tone as predicted by ALT. However, the phonetic boundaries for the speech stimuli failed to show the analogous shift. These results suggest that the response criteria for phonetic boundaries may be mediated by an internally-generated reference. In contrast, the reference for the nonspeech category boundaries appears to be under external stimulus control.

*This is a draft of a paper to be presented at the 86th meeting of the Acoustical Society of America, November 1, 1973, Los Angeles, California. The research was supported in part by PHS grants SO5 RR 7031, MH 24027, and TO1 MH 11219-04 to Indiana University. We wish to thank Dr. Franklin S. Cooper for making the facilities of Haskins Laboratories available to us for the preparation of the stimulus materials and Professor Frank Restle for his interest in this work.

Category Boundaries for Speech and Nonspeech Sounds

J. R. Sawusch and D. B. Pisoni

Indiana University

In recent years a large body of evidence has been accumulated to suggest that the perception of speech sounds may be quite different from the perception of other auditory stimuli. Although several theories of speech perception have been proposed, they are for the most part quite vague and general and it is relatively difficult to derive any specific predictions that are testable. In the present study we wish to consider the nature of category judgments and specifically the nature of the boundaries between categories for speech and nonspeech sounds. We have chosen this particular problem to study primarily because we feel that two of the current theories of speech perception, the Haskin's Motor Theory and Steven's Quantal Theory, would make specific predictions about the nature of category boundaries for certain classes of speech sounds.

We may think of two relatively broad views of the nature of category boundaries for speech sounds. One view is that the category boundaries between phonetic segments are arbitrary in the sense that they are simply the consequence of a psychophysical partitioning of a stimulus continuum into equivalent response categories. In contrast, an alternative view and one which could be predicted from either speech theory is that the boundaries between phonetic segments are not arbitrary. Rather, due to constraints on the articulatory mechanism and the resultant changes in the acoustic signal, the boundaries between segments may be relatively fixed.

If the perceptual boundaries between phonetic segments are arbitrary in the sense of a simple psychophysical partitioning of the stimulus continuum, then it should be relatively easy to produce systematic changes in the location of the boundary by manipulations of the probabilities of occurrence of different stimuli. For example, Adaptation Level Theory, which has been used extensively in psychology to account for changes in the judgment of brightness, hue, loudness, and pitch, could be applicable to the judgment of phonetic segments.

Slide 1 please

Let us consider such a prediction in detail. Slide 1 shows an idealized identification function for a two category absolute identification task. In the control condition, each stimulus occurs with an equal probability and the subject partitions the continuum into two equivalent categories. When the probabilities are unbalanced and, for example, stimulus number one occurs more often than any of the other stimuli, the boundary should shift toward the more frequently occurring stimulus, or anchor. The same effect should be obtained when stimulus seven occurs more often than any of the other stimuli; the category boundary should shift toward that stimulus. Adaptation Level Theory would predict these results for both speech and nonspeech continua since it is assumed that the standard or reference used to categorize a particular stimulus is for the most part under the control of external stimuli.

In the present study, we were concerned with the effect of unbalanced probabilities of occurrence of stimuli on the identification of speech and

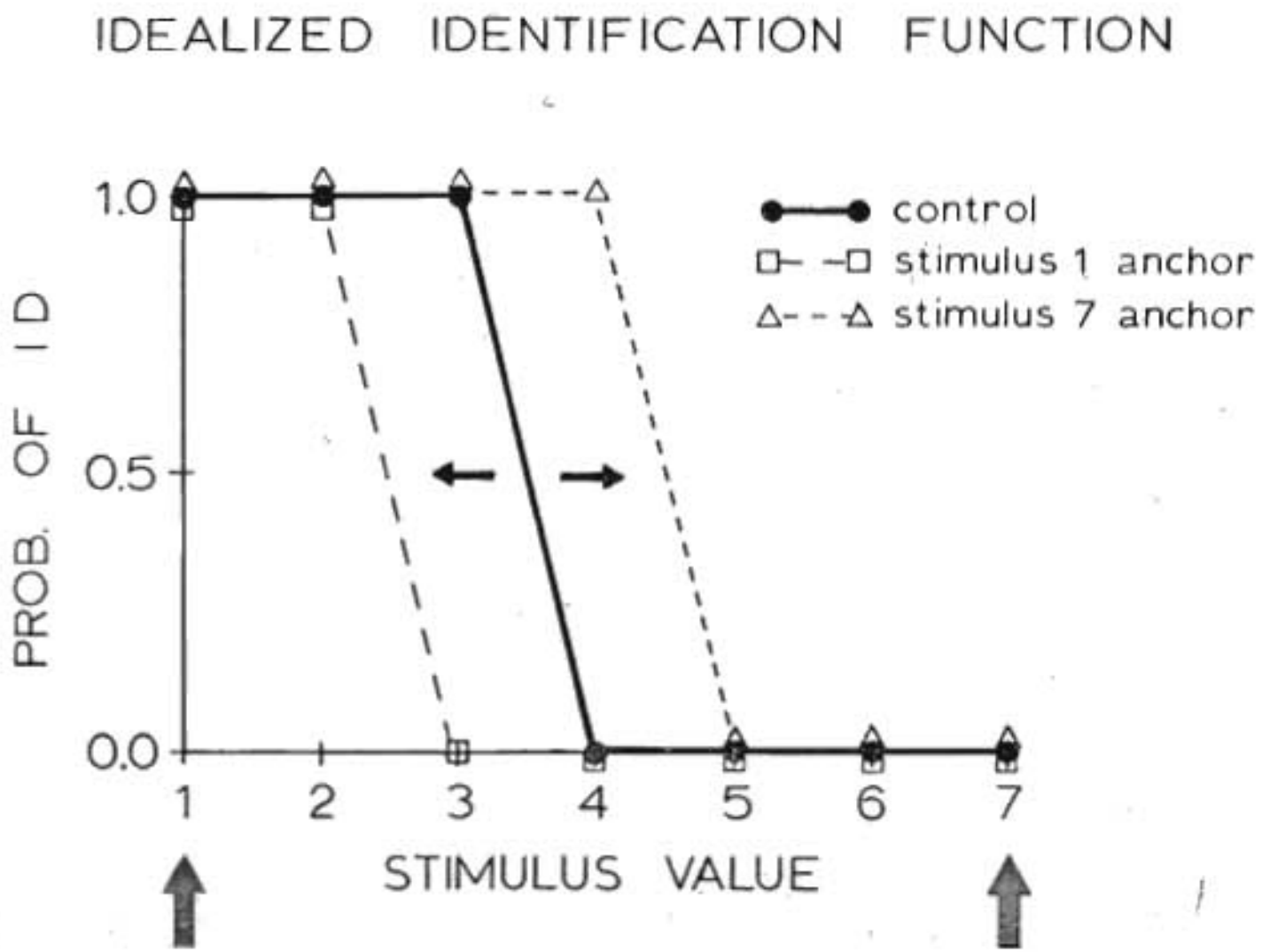

IDEALIZED IDENTIFICATION FUNCTION
1.0
0.5
0.0
PROB. OF I D
control
stimulus 1 anchor
stimulus 7 anchor
1
2
3
4
5
6
7
STIMULUS VALUE

Figure 1.

nonspeech continua. If the boundaries between phonetic segments are arbitrary and under the control of external stimuli we should expect to find a shift in the identification function as the relative frequency of occurrence of the end or anchor stimuli are changed. On the other hand, if we do not find a shift in the identification function for the speech continuum but we do find one for a nonspeech continuum, we may conclude, at the very least, that the boundaries between phonetic segments are not simply due to an arbitrary partitioning of a stimulus continuum.

Method

Two sets of stimuli were used in this experiment, a speech continuum and a nonspeech continuum. The speech stimuli consisted of a set of seven three-formant patterns ranging perceptually from /ba/ through /pa/ and were produced on the speech synthesizer at Haskins Laboratories. The seven stimuli varied in 10 msec steps along the voice onset time continuum from 0 msec VOT to +60 msec VOT. The stimuli were recorded on magnetic tape in random order to produce three identification tests. In the control tape, each stimulus occurred equally often. In the /ba/ anchor tape, the stimulus with 0 msec VOT occurred twice as often as each of the other six stimuli. In the /pa/ anchor tape, the stimulus with +60 msec VOT occurred twice as often as each of the other stimuli.

The nonspeech continuum consisted of a set of seven tones varying in intensity. The tones varied in 4 dB steps from 60 dB to 84 dB. These stimuli were recorded in random order on magnetic tape to produce three analogous identification tests: a control tape, a loud anchor tape, and a soft anchor tape.

Subjects listened to four different tapes: a speech control, a tone

control, a speech anchor, and a tone anchor. In the speech condition subjects were told to identify each stimulus as a /ba/ or a /pa/. In the nonspeech condition subjects were told to identify each tone as loud or soft.

Results and Discussion

Slide 2 please

Slide 2 shows the average identification functions for the nonspeech condition. In Group I shown on the left subjects heard the loud anchor tape. The identification function shows a consistent shift toward the loud stimulus, relative to the control identification function. In Group II shown on the right subjects heard the soft anchor tape. This identification function shows a shift but this time it is toward the soft stimulus, relative to the control function. Both shifts, which would be predicted by Adaptation Level Theory, reveal how the loud-soft judgment is arbitrary in the sense that it is under the control of the stimuli occurring during the test. Now let us turn to the speech data for comparison.

Slide 3 please

Slide 3 shows the average identification functions for the /ba/-/pa/ continuum for the same subjects. Group I heard the /ba/ anchor tape. The identification function shown on the left shows no shift relative to the control condition. The same is true for the /pa/ anchor group. There is no shift relative to the control tape. If the speech stimuli were evaluated

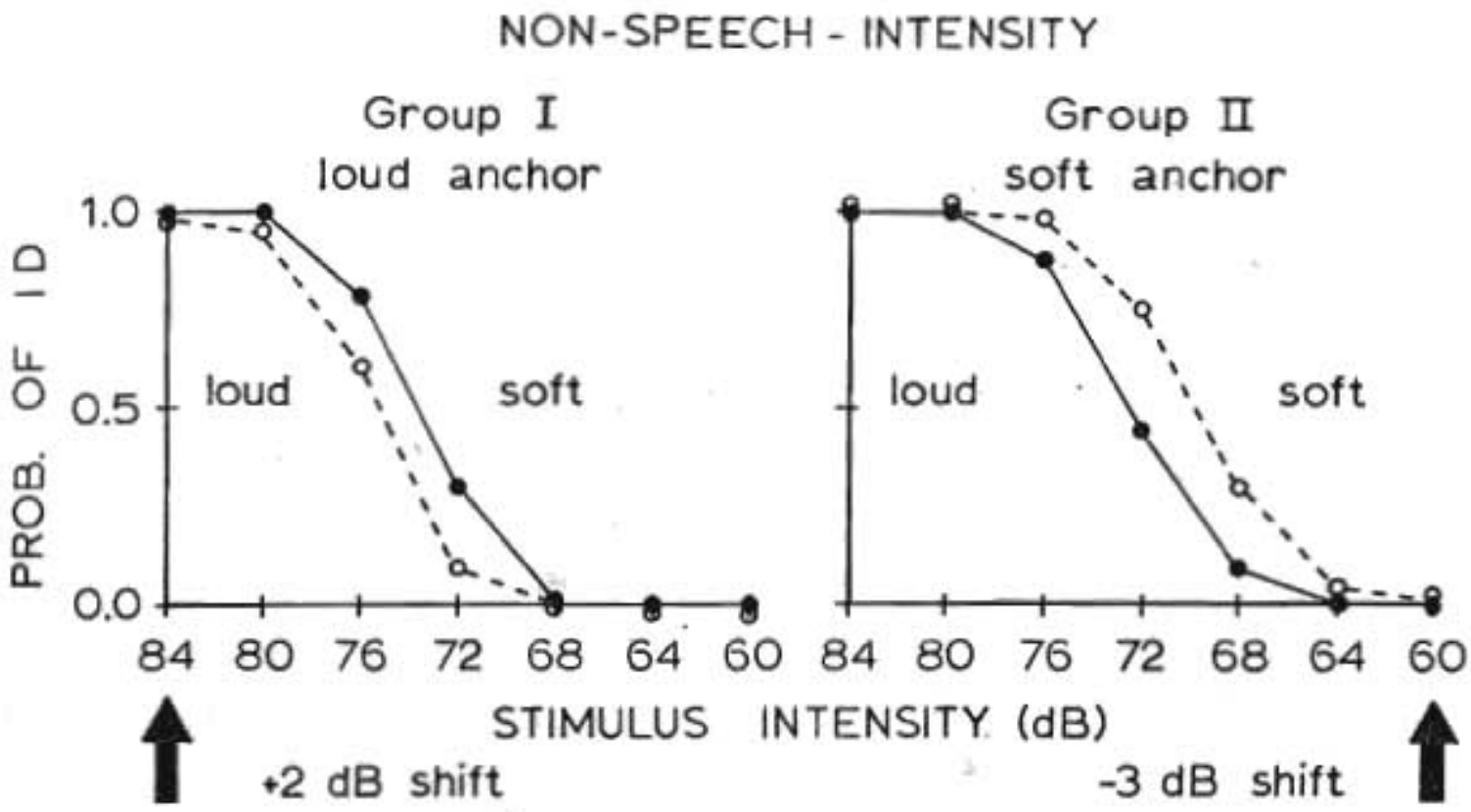

NON-SPEECH - INTENSITY
Group I
loud anchor
Group II
soft anchor
PROB. OF I D
1.0
0.5
0.0
loud
soft
loud
soft
84 80 76 72 68 64 60
84 80 76 72 68 64 60
STIMULUS INTENSITY (dB)
+2 dB shift
-3 dB shift

Figure 2.

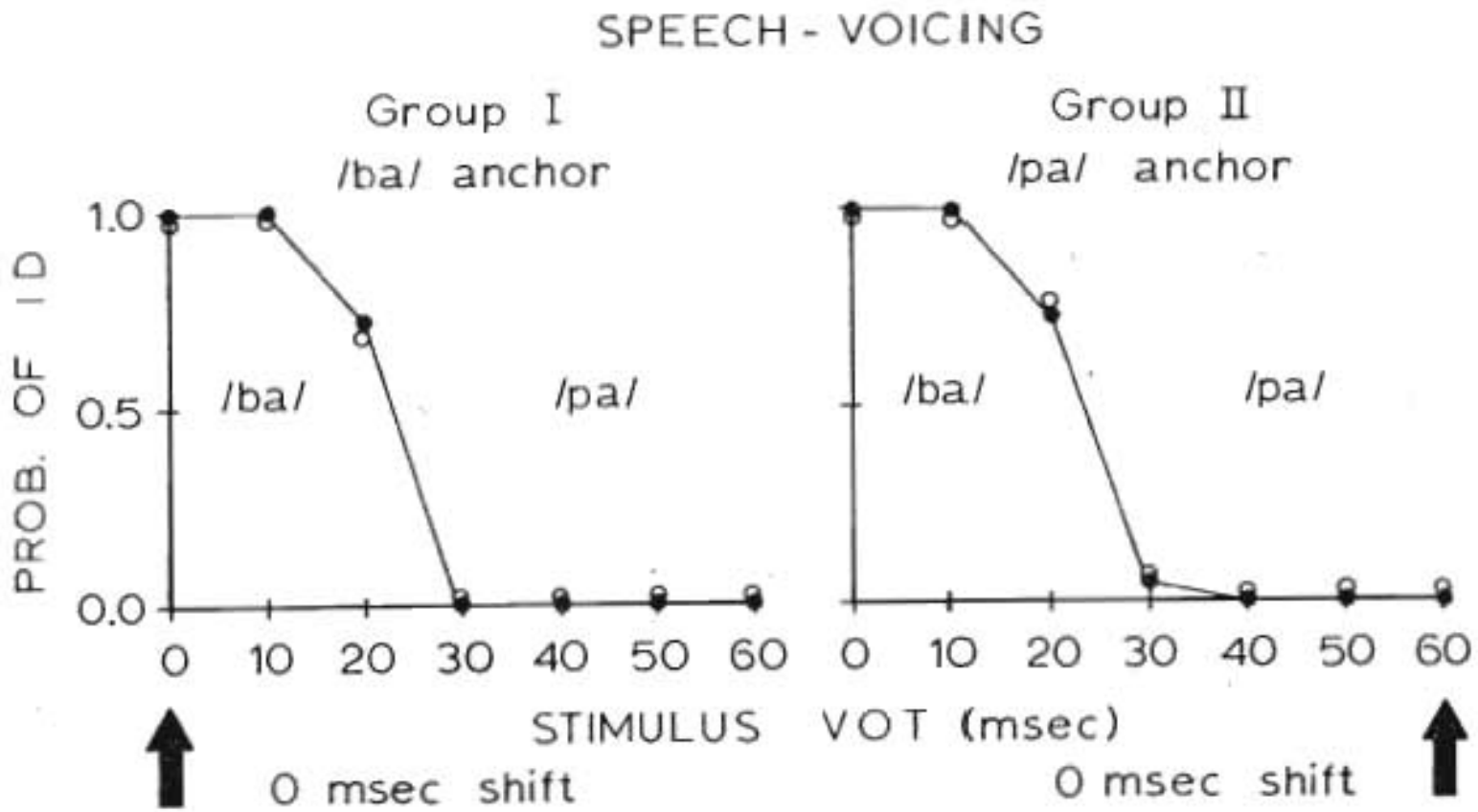

SPEECH - VOICING
Group I
/ba/ anchor
Group II
/pa/ anchor
PROB. OF I D
1.0
0.5
0.0
/ba/
/pa/
/ba/
/pa/
0 10 20 30 40 50 60
0 10 20 30 40 50 60
STIMULUS VOT (msec)
0 msec shift
0 msec shift

Figure 3.

according to the same criterion as the tones we would have expected a shift for speech anchor conditions relative to the control. Surprisingly, no shift occurred for the speech.

Experiments similar in format to these but using synthetic stop consonants varying in place of production and tones varying in frequency have also been run. The results indicate that the category boundary between hi and low frequency tones shifts as predicted by ALT. The category boundary for the place continuum between /bae/ and /dae/ also failed to show any shift.

The view that phonetic categories are arbitrary and simply the result of a psychophysical partitioning of a stimulus continuum seems to be an inadequate explanation of our results. If this were in fact the case we would have expected the speech anchor conditions to show a shift analogous to that found in the tone anchor conditions. It seems that the category boundaries at least for place and voicing in stop consonants result from the application of relatively stable and non-arbitrary criteria. We suggest that these criteria may be mediated more by an internally generated reference than an externally calculated standard.

To summarize, the effect of unbalanced probabilities of occurrence of stimuli produced a shift in the category boundary for a nonspeech continuum but failed to produce a parallel shift in a speech continuum. These results suggest an internal, highly stable and non-arbitrary criteria for the categorization of phonetic segments.

Dichotic Backward Masking, The "Lag Effect" and Processing Phonetic Features*

S. D. McNabb and D. B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47401

The "lag effect" has been used to support the argument that speech perception engages distinctive processes that differ from those of nonspeech auditory perception. However, a number of recent experiments have shown that the effect may not be peculiar to speech since it may be obtained with nonspeech timbres, vowels and other stimuli. In fact, the effect appears to be a variation of a more general result obtained in backward masking experiments: a second stimulus may impede the processing of a preceding stimulus. The present study sought to determine the locus of this effect. Under dichotic presentation, one of four syllable targets (/ba/, /da/, /pa/, /ta/) was followed by one of six possible syllable masks (/ga/, /ka/, /gae/, /kae/, /gɛ/, /kɛ/), each 300 msec in duration. The onset time of the mask relative to that of the target was varied over the range 0 to -150 msec. Subjects identified only the target sound in an ear monitoring task. Two findings were obtained which argue that the lag effect has an auditory basis. First, when the target and mask differed on voicing (i.e., b-k, p-g) performance improved with increases in the onset time of the mask; no interference was obtained when the target and mask shared voicing. Secondly, performance varied inversely with the similarity of the vowels in the mask: performance was lowest with /a/ and highest with /ɛ/. Since the interference occurred only for trials that differ on voicing and these trials were also affected by vowel context, we conclude that the locus of the lag effect lies before phonetic analysis and therefore must have an auditory rather than phonetic basis.

*This paper was presented at the 87th meeting of the Acoustical Society of America, April, 1974, New York City, New York. The research was supported in part by PHS grant MH 24027 to Indiana University.

Over the last few years we have heard a large number of papers at the Society which have dealt with the so-called "lag-effect" in dichotic listening experiments.

Insert Figure 1 about here

The effect is shown here in Figure 1 which we have borrowed from Studdert-Kennedy, Shankweiler & Schulman (1970). The second or "lagging" syllable of a dichotic pair of temporally overlapping stimuli is reported or identified more accurately than the "leading" syllable. I am sure you are all familiar with this data.

Our interest in these findings lies in a number of claims that have been made as to the locus of the effect: That is, where does the interaction between the two inputs occur? Also, is the lag effect peculiar to speech perception or does it result from more general perceptual operations? Studdert-Kennedy, Berlin and others have interpreted the lag effect as a form of "interruption of speech processing" occurring at some "central" level of perceptual analysis. For example, Studdert-Kennedy, Shankweiler & Schulman (1970) state that "the lag effect is tied to speech, and, specifically, to those components of the speech stream for which a relatively complex decoding operation is necessary." Indeed, the lag effect has been used to support the general argument that speech perception engages specialized processes that differ from those of non-speech perception--that is that "speech is special."

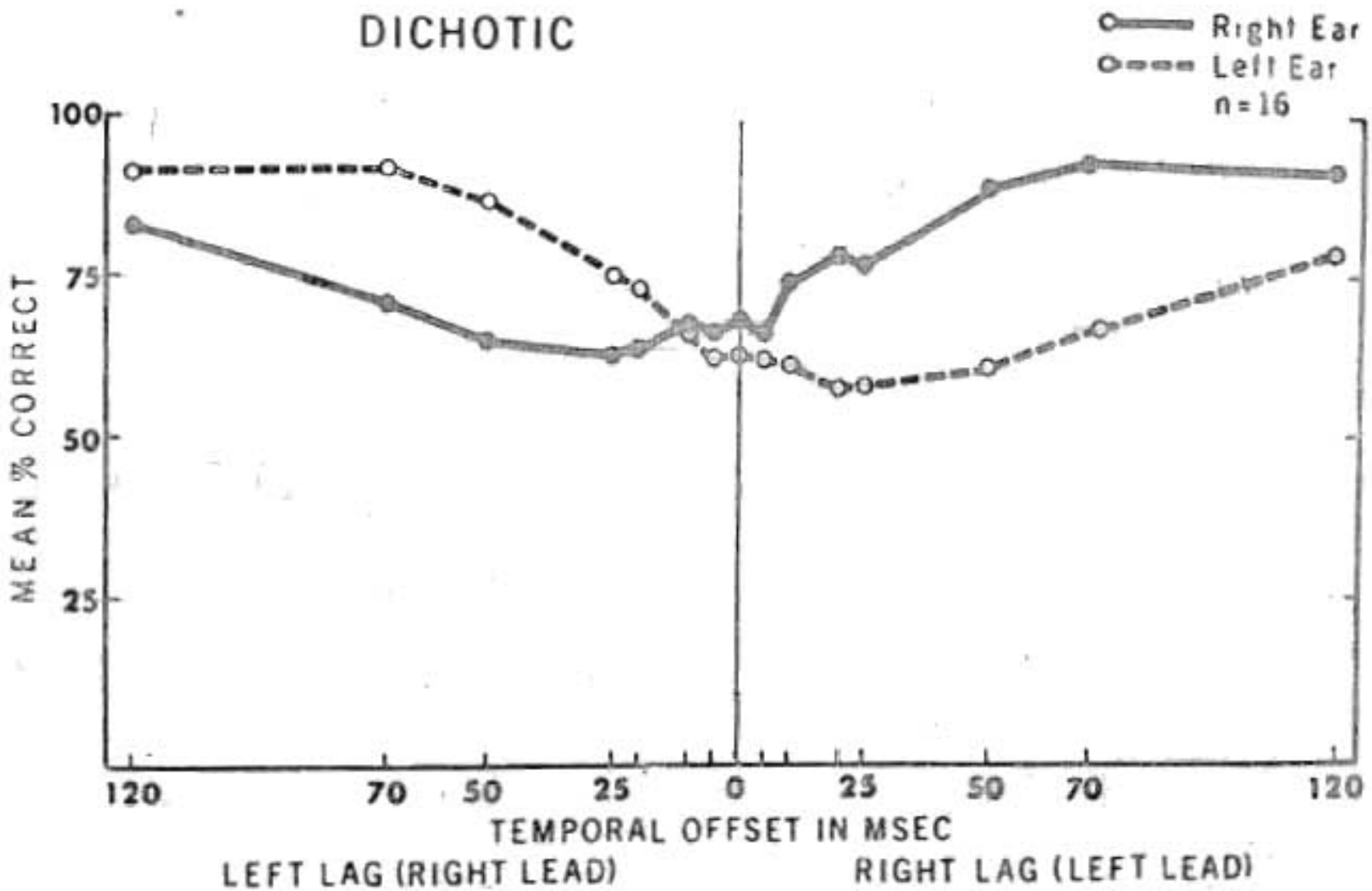

DICHOTIC
Right Ear
Left Ear
n = 16
100
75
50
25
MEAN % CORRECT
120
70
50
25
0
25
50
70
120
TEMPORAL OFFSET IN MSEC
LEFT LAG (RIGHT LEAD)
RIGHT LAG (LEFT LEAD)

Figure 1.

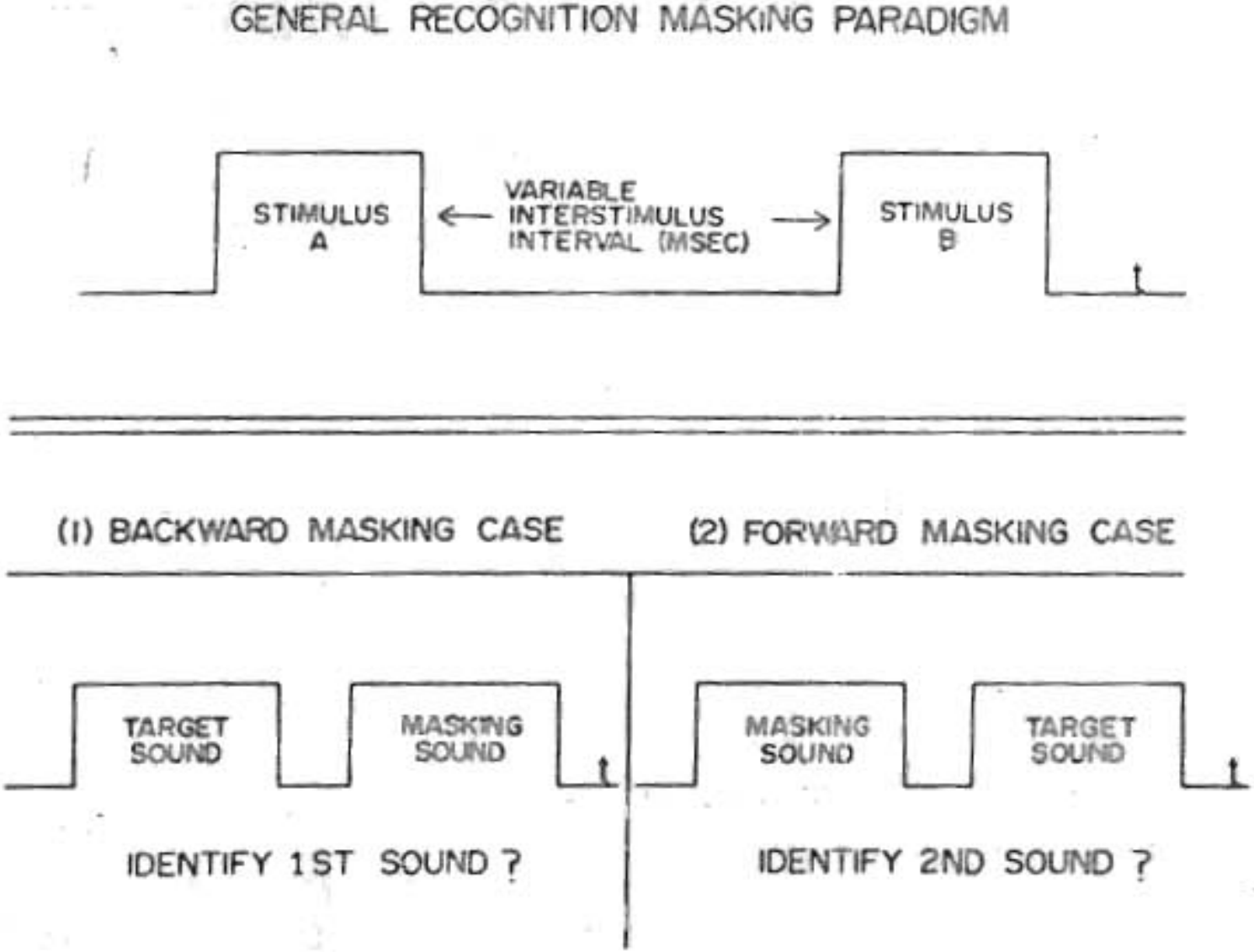

GENERAL RECOGNITION MASKING PARADIGM
STIMULUS A
VARIABLE INTERSTIMULUS INTERVAL (MSEC)
STIMULUS B
(1) BACKWARD MASKING CASE
(2) FORWARD MASKING CASE
TARGET SOUND
MASKING SOUND
MASKING SOUND
TARGET SOUND
IDENTIFY 1ST SOUND ?
IDENTIFY 2ND SOUND ?

Figure 2.

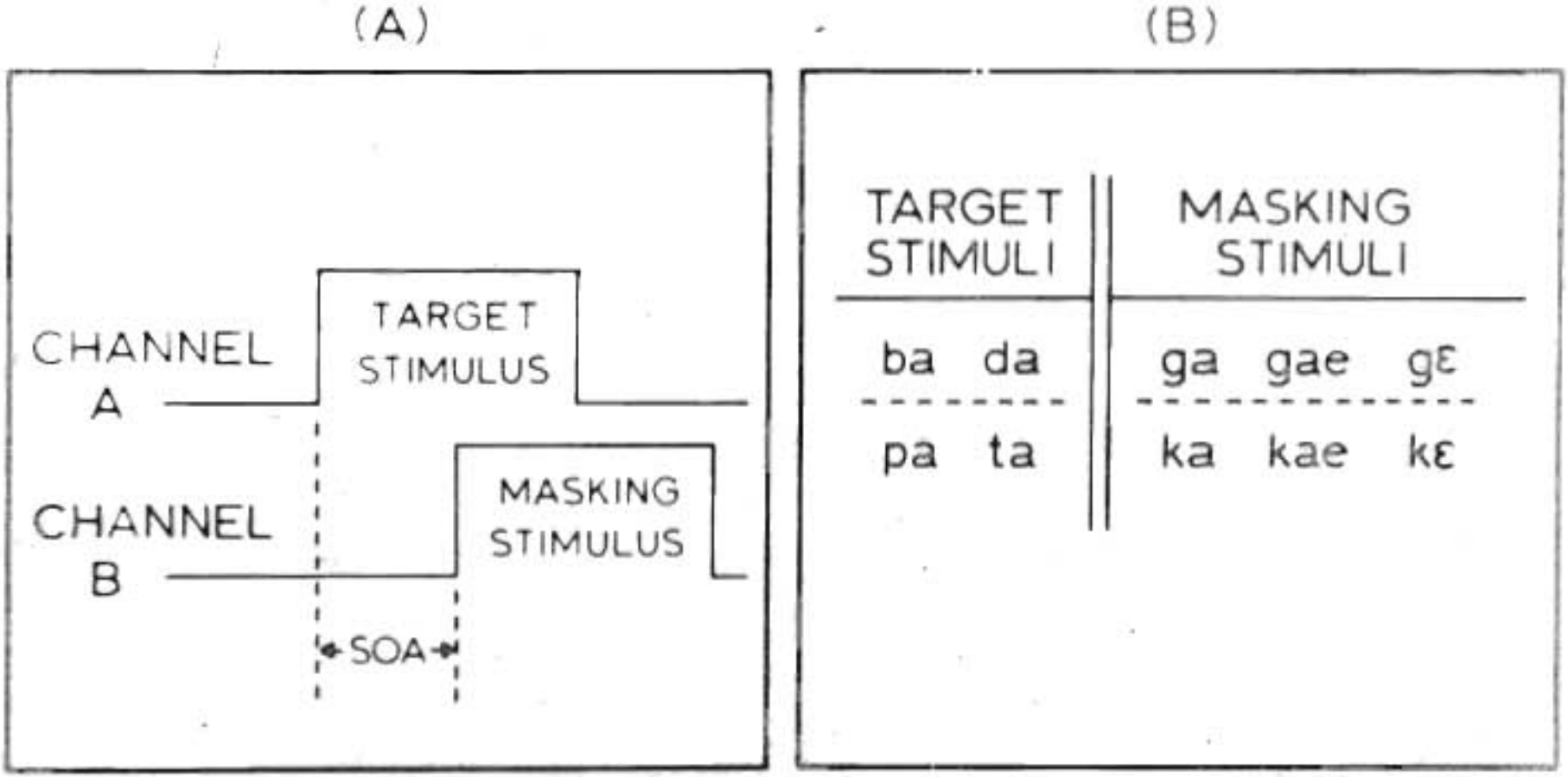

(A)
CHANNEL A
TARGET STIMULUS
CHANNEL B
MASKING STIMULUS
SOA
(B)
TARGET STIMULI
MASKING STIMULI
ba da
ga gae gɛ
pa ta
ka kae kɛ

Figure 3.

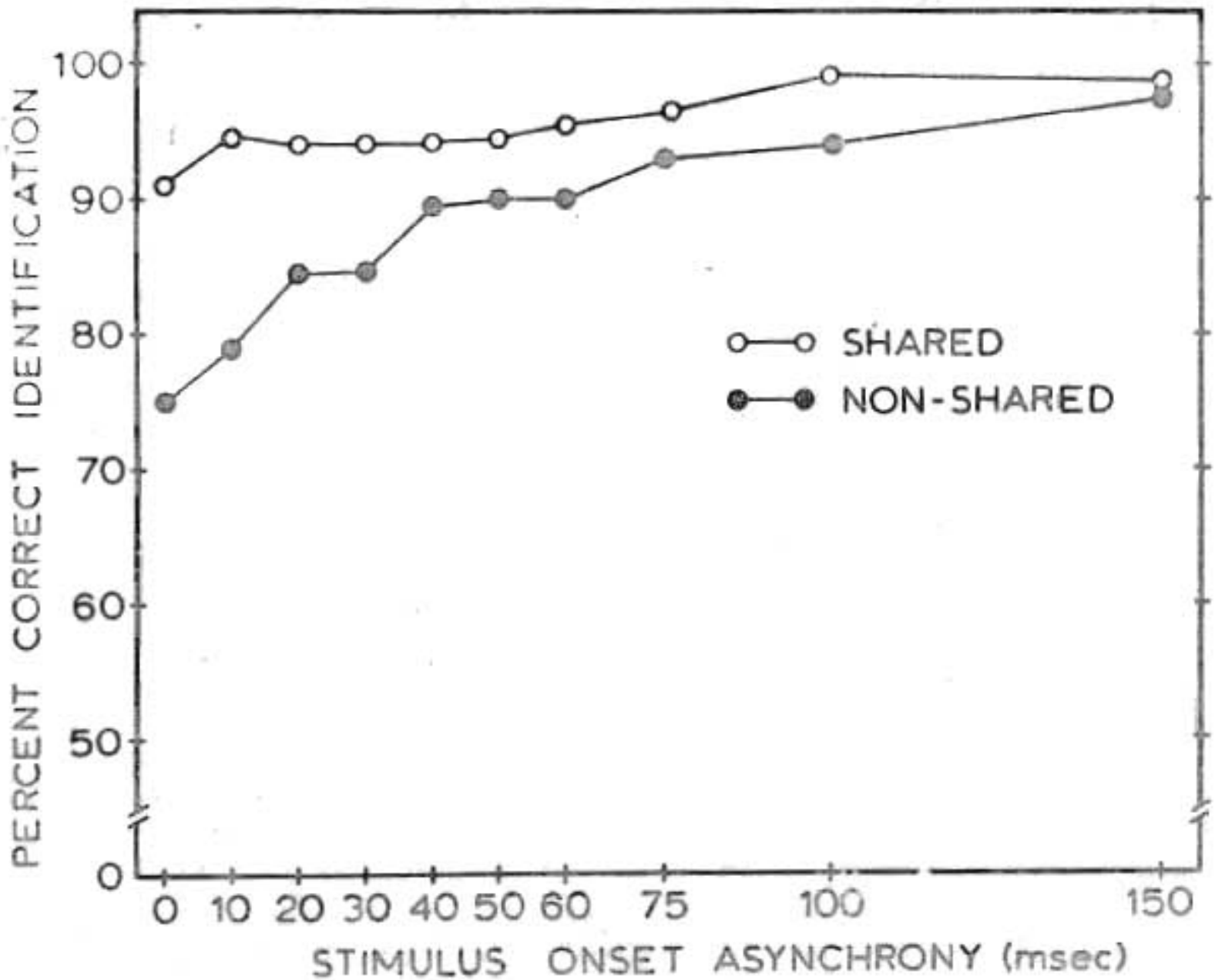

PERCENT CORRECT IDENTIFICATION
100
90
80
70
60
50
0
SHARED
NON-SHARED
0 10 20 30 40 50 60 75 100 150
STIMULUS ONSET ASYNCHRONY (msec)

Figure 4.

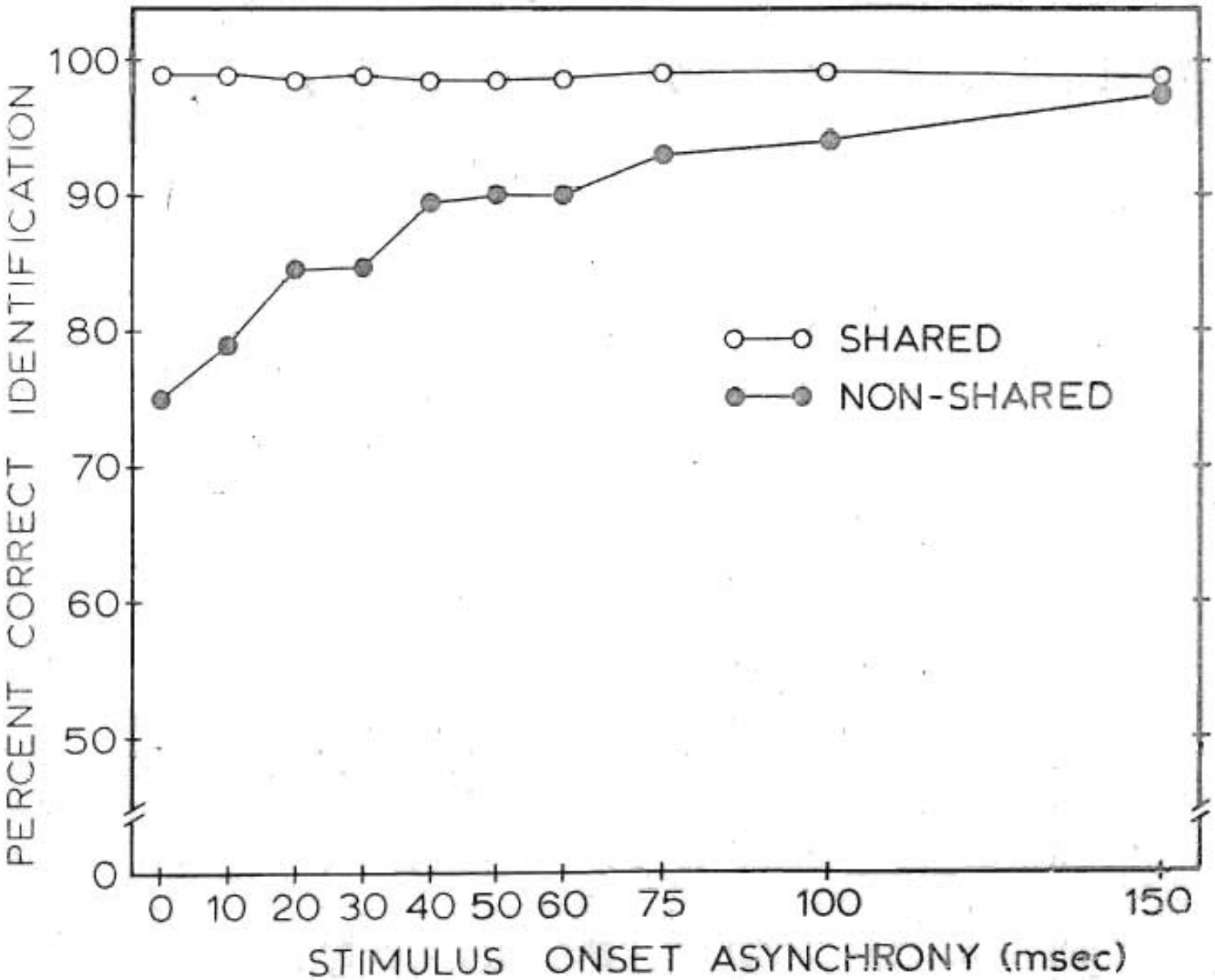

PERCENT CORRECT IDENTIFICATION
100
90
80
70
60
50
0
SHARED
NON-SHARED
0 10 20 30 40 50 60 75 100 150
STIMULUS ONSET ASYNCHRONY (msec)

Figure 5.

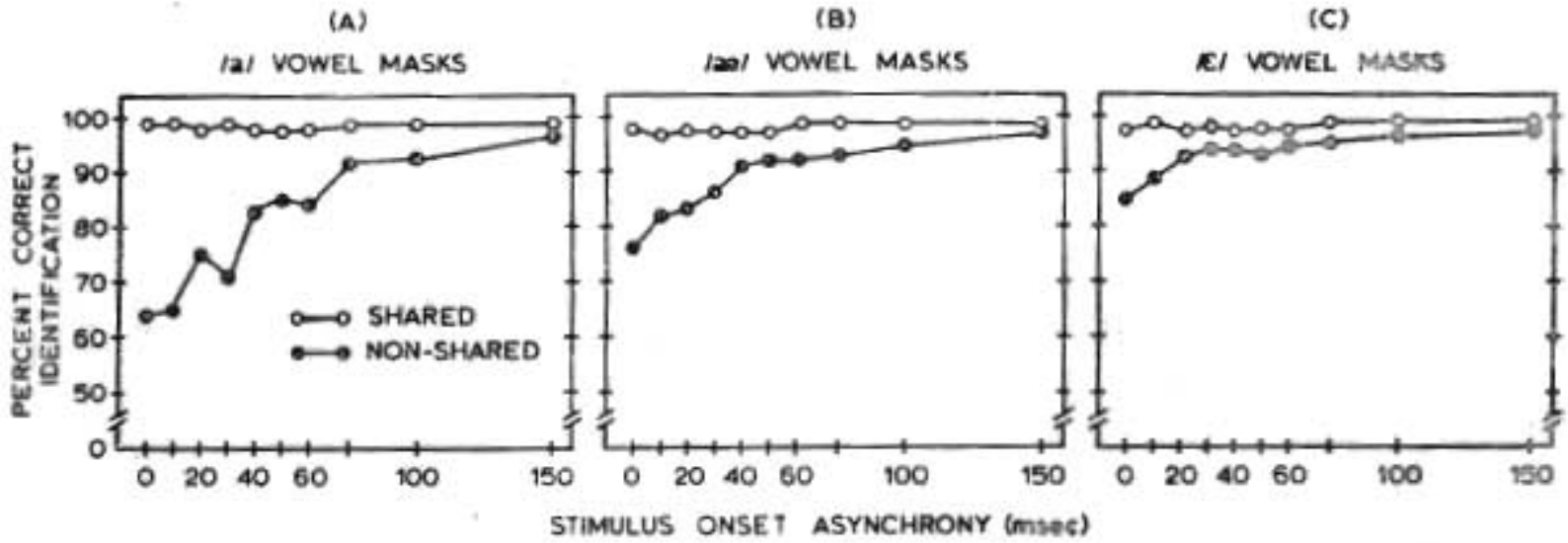

(A)
/a/ VOWEL MASKS
(B)
/aa/ VOWEL MASKS
(C)
Æ/ VOWEL MASKS
PERCENT CORRECT IDENTIFICATION
100
90
80
70
60
50
0
SHARED
NON-SHARED
0 20 40 60 100 150
0 20 40 60 100 150
0 20 40 60 100 150
STIMULUS ONSET ASYNCHRONY (msec)

Figure 6.

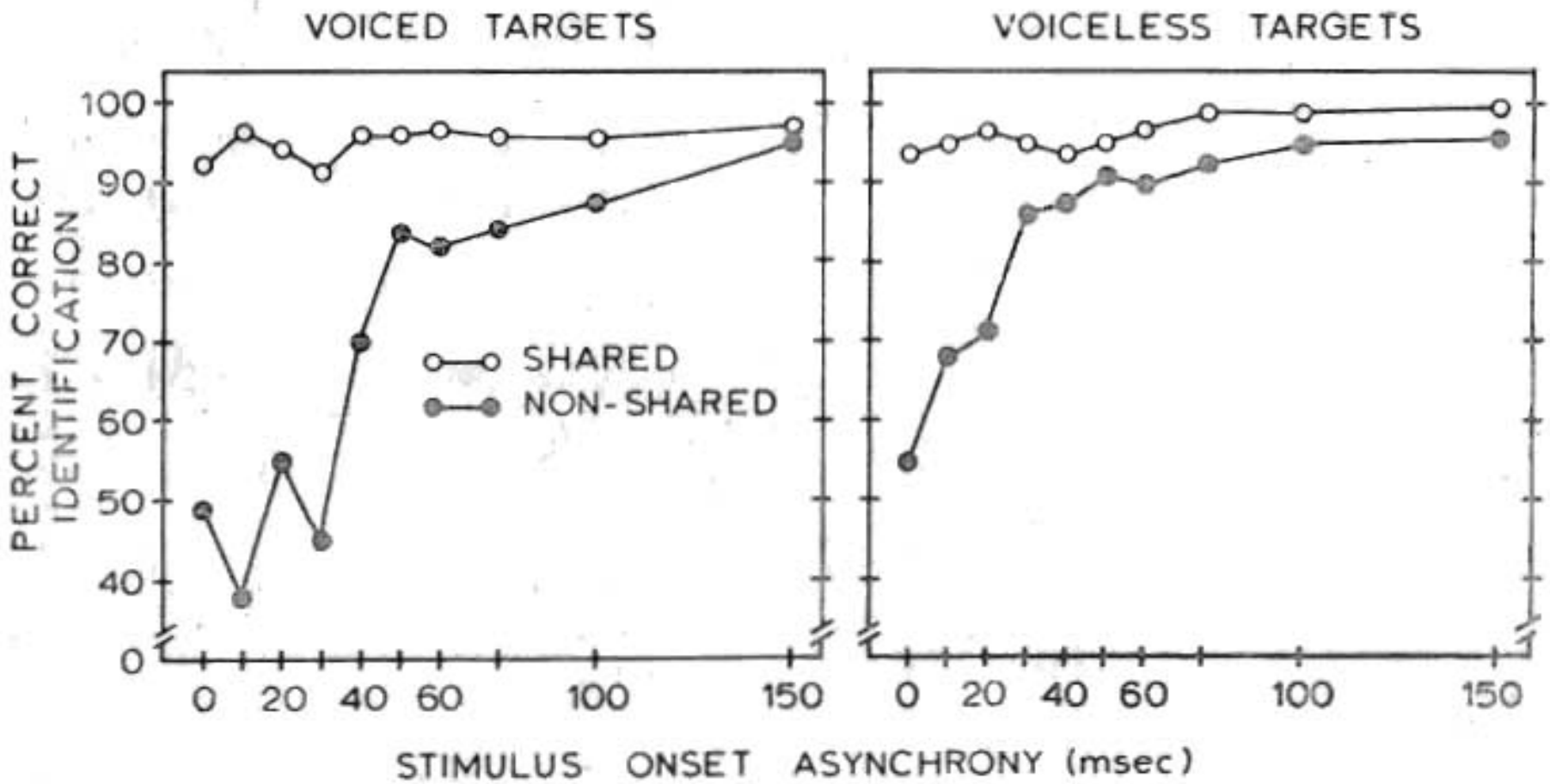

VOICED TARGETS
VOICELESS TARGETS
PERCENT CORRECT IDENTIFICATION
100
90
80
70
60
50
40
0
SHARED
NON-SHARED
0
20
40
60
100
150
STIMULUS ONSET ASYNCHRONY (msec)

Figure 7.

We do not believe that the lag effect is mysterious or peculiar and we hope to show you why. The effect seems to be a variation of a more general result obtained in backward recognition masking experiments: A second stimulus can impede the processing of a preceeding stimulus (Massaro, 1972).

Insert Figure 2 about here

Strictly speaking, the lag effect refers to the relative difference between dichotic forward and backward masking. Under the conditions usually studied there is more dichotic backward masking than forward masking (see also Repp, 1973).

In this study we were concerned with the "locus" of the interaction between the dichotic inputs. Where in the flow of information does the effect arise? Put another way, does the effect have an auditory or a phonetic basis? To study this problem we looked at the effect of three variables on the identification of a known set of "target" CV syllables.

Insert Figure 3 about here

These can be best shown in Figure 3. The first variable is stimulus onset asynchrony (SOA) as shown in panel (A). This was varied over a ten step range from 0 msec. to -150 msec. The second

variable was the feature composition of the target and mask pair. On half the trials the target stimulus and mask shared voicing; on the other half they differed on voicing. For example, /ba/ followed by /ga/ is a shared pair whereas /ba/ followed by /pa/ is a non-shared pair or double contrast trial. The last variable we consider is the vowel in the masking stimulus. As shown in panel (B) of this figure we have 3 vowel contexts for our masking stimuli: /a/, /ae/ and /ɛ/.

If the interference in the lag effect is due to "interruption" of processing, we should expect any mask to be equivalent with regard to its masking effectiveness. On the other hand, if part of the features in the first stimulus have been recognized then whether the target and mask share voicing should affect recognition performance.

If the interference is due to interactions on the phonetic feature level, we should not expect the vowel context of the mask to affect recognition of the target since the consonant would have already been abstracted from the syllable. On the other hand, if the interaction has an auditory basis at an earlier stage of analysis where the syllables interact we would anticipate a systematic vowel effect according to vowel similarity. That is, performance should be lowest with the /a/ vowel masks, highest with /ɛ/ and midway for /ae/.

Twenty-four Ss were run in a dichotic ear monitoring task where on each trial one of the 4 targets was followed by one of the 6 possible masks. These 24 pairs of stimuli were presented at 10 different SOA values in a random order. Ss identified only the targets /ba/, /da/,

/pa/, /ta/ and ignored the other stimuli.

Insert Figure 4 about here

Figure 4 shows the percent correct identification for shared vs. non-shared pairs averaged over all vowel contexts. Performance is relatively good for shared pairs and is not affected very much by SOA. The non-shared condition is lower and is affected by increases in SOA. Performance improves as SOA becomes larger.

Insert Figure 5 about here

This figure shows the same data but scored now for the voicing feature. Performance is perfect for voicing when target and mask share the voicing feature. Performance steadily improves as a function of SOA for non-shared pairs.

Insert Figure 6 about here

This slide shows the effect of the vowel context of the mask. Performance on voicing is lowest for non-shared pairs when the vowel in the mask and target are the same (i.e., /a/). Performance improves as vowel similarity decreases from /a/ to /ae/ to /ɛ/.

Thus, we think we have good grounds for arguing that the interference

obtained in the lag effect has an auditory rather than phonetic basis since the vowel in the mask interacts with recognition of the target stimulus.

Insert Figure 7 about here

This figure shows the trials broken down by voiced and voiceless targets for shared and non-shared trials. The feature effect shows up again. Shared pairs are recognized better than non-shared pairs.

To summarize, we have found that interference in a backward recognition masking experiment does not occur for all stimulus contrasts but only those differing on voicing (i.e., double contrasts). Hence, the interference in the lag effect cannot be due to "interruption" of processing as suggested in earlier reports. We also found that the vowel of the mask systematically effects recognition of the target. We interpret this to mean that the interaction between the inputs (that is, the locus of interaction) occurs before "phonetic" analysis.

We think these data argue strongly that the interference obtained in the lag effect has an auditory rather than phonetic basis and that there is nothing mysterious or peculiar about the lag effect results obtained in previous dichotic listening experiments.

Category Boundaries for Linguistic and Nonlinguistic Dimensions of the Same Stimuli*

J. R. Sawusch and D. B. Pisoni
Department of Psychology
Indiana University
Bloomington, Indiana 47401

and

J. E. Cutting
Haskins Laboratories
New Haven, Connecticut 06510

At the previous meeting of the Society we reported finding a shift in the category boundaries for tonal stimuli as a function of the relative number of occurrences of each stimulus in the series. No such shift was found in the category boundaries for synthetic stop consonant-vowel syllables. However, it could be argued that tonal stimuli are not appropriate control stimuli for speech sounds. To examine this possibility, identification functions were obtained for a series of synthetic CV syllables that varied simultaneously in both place (/ba/ to /da/) and pitch (/lo/ to /hi/). The distribution of occurrences of stimuli for the place and pitch dimensions were varied independently. When Ss judged pitch the category boundary shifted toward the more frequently occurring stimulus. In contrast, when they judged place no shift in the phonetic boundary was observed. These results agree with our previous findings and suggest that unlike the arbitrary categories for nonspeech stimuli, phonetic categories may have a naturally determined basis. These results are also discussed with regard to some recent findings on possible feature detectors in speech perception.

*This paper was presented at the 87th meeting of the Acoustical Society of America, April, 1974, New York City, New York. The research was supported in part by PHS grant MH 24027 and PHS grant MH 11219 to Indiana University.

There has been a great deal of work recently on the phenomenon of adaptation in speech perception. Using the selective adaptation paradigm first employed by Eimas evidence has been found for the existence of both phonetic and acoustic feature detectors in speech perception. Subjects exhibit a shift in the boundary locus of a CV identification function after listening to repeated presentations of an adapting stimulus. However, the selective adaptation results by themselves do not rule out alternative explanations of the shift phenomenon. One explanation of the shift effect is simple response bias. A response bias theory, such as adaptation level theory, would predict movement of an identification boundary toward the more frequently occurring stimulus. In the case of the selective adaptation paradigm, the subjects have been exposed to many more instances of the category from which the adapting stimulus is drawn. As a result the subject is biased to use the other category for responding during the identification test. This type of effect, known as a contrast effect, is well documented in visual perception and in the perception of simple tones.

At the previous meeting of the society in Los Angeles we reported results which examined subjects' identification functions for stop CV syllables and tones under two conditions. All subjects heard a series in which every stimulus occurred equally often. The second series

---

Slide 1 please

---

contained one stimulus that occurred twice as often as each of the rest

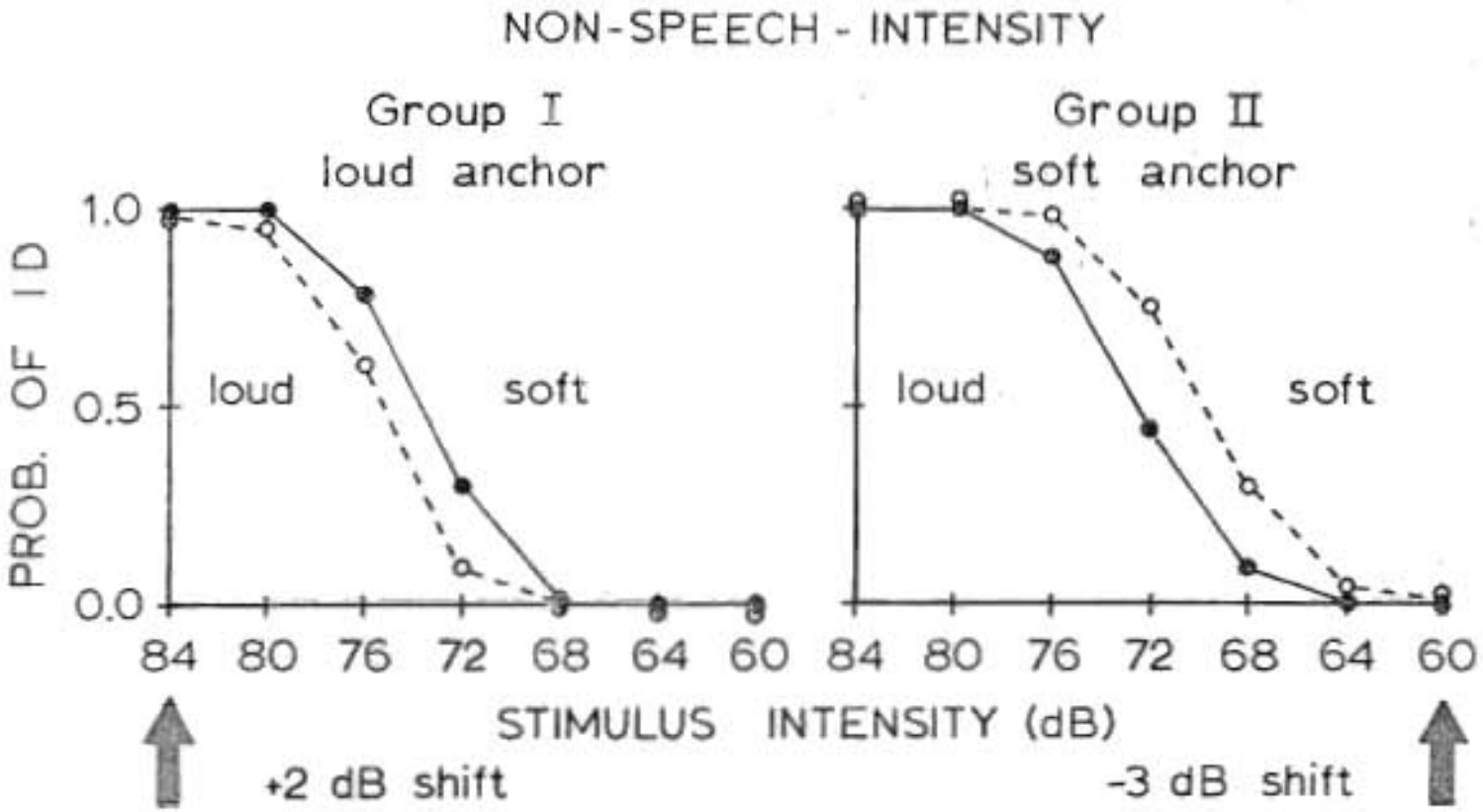

NON-SPEECH - INTENSITY
Group I
loud anchor
Group II
soft anchor
PROB. OF I D
1.0
0.5
0.0
loud
soft
loud
soft
84 80 76 72 68 64 60
84 80 76 72 68 64 60
STIMULUS INTENSITY (dB)
+2 dB shift
-3 dB shift

Figure 1.

of the stimuli. In this type of series, response bias theories predict a contrast effect. The identification boundary in the anchored or unequal series should shift toward the more frequently occurring stimulus, relative to the control condition. This result was found for the tones as shown in slide 1.

Slide 2 please

Under the same conditions, no shift was found for CV syllables varying in either place or voicing. Slide 2 shows this absence of a shift for a voicing series. We concluded that response bias theories which can provide an adequate explanation of the shift in the non-speech stimuli were an inadequate explanation of the category boundaries in CV syllables and that some other mechanism was mediating the category decision.

The data we presented in Los Angeles can be criticized on the grounds that pure tones were an inadequate control for the speech series. Tones are much simpler acoustically and they are also less familiar to the subjects than the speech sounds employed in these studies. In the present experiment, the judgment of place of production in a synthetic CV series is contrasted with the judgment of the fundamental frequency of the same identical stimuli. By using the same speech stimuli as their own control the comparison stimuli are neither simpler nor less familiar. The crucial difference now is whether the dimension being judged is carrying linguistic or non-linguistic information in the speech signal. The predictions of a response bias model such as adaptation level theory are

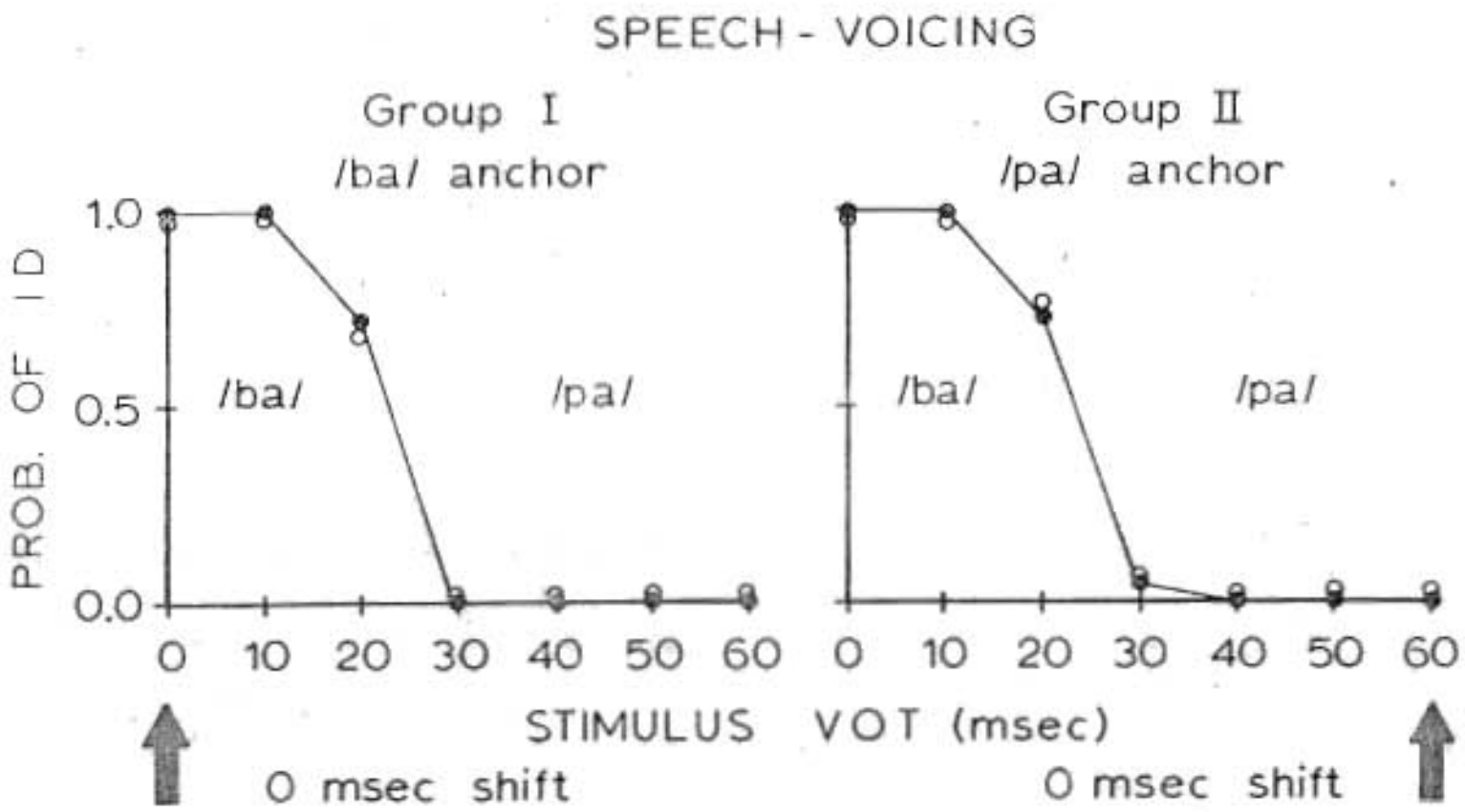

SPEECH - VOICING
Group I
/ba/ anchor
Group II
/pa/ anchor
PROB. OF I D
1.0
0.5
0.0
/ba/
/pa/
/ba/
/pa/
0 10 20 30 40 50 60
0 10 20 30 40 50 60
STIMULUS VOT (msec)
0 msec shift
0 msec shift

Figure 2.

the same as in our previous experiment. In the unequal series, the identification function should shift toward the more frequently occurring stimulus. This should be true whether the subject judges place or pitch.

Method

Slide 3 please

One set of synthetic three formant speech stimuli was used in this experiment. Two dimensions of the stimuli were varied orthogonally. The linguistic dimension was the place of production. The stimuli ranged perceptually from /bae/ to /dae/ in seven equal steps. The second and non-linguistic dimension was fundamental frequency or pitch. This was varied from 114 Hz to 150 Hz in 6 Hz steps. As indicated in slide 3, each value of pitch was paired with each value of place to produce a series of 49 stimuli in which no value on either dimension could be used to predict the value on the other dimension. These 49 stimuli were recorded on magnetic tape at Haskins Laboratories in random orders to produce five identification test tapes. In the control tape, each stimulus occurred twice. In the low anchor tape, each of the seven stimuli with a fundamental frequency of 114 Hz, shown in the dashed box, occurred eight times and the rest of the stimuli occurred twice each. In the bae anchor tape each of the seven stimuli with the bae place of production occurred eight times and the other stimuli occurred twice each. In similar fashion, high anchor, and /dae/ anchor tapes were also constructed.

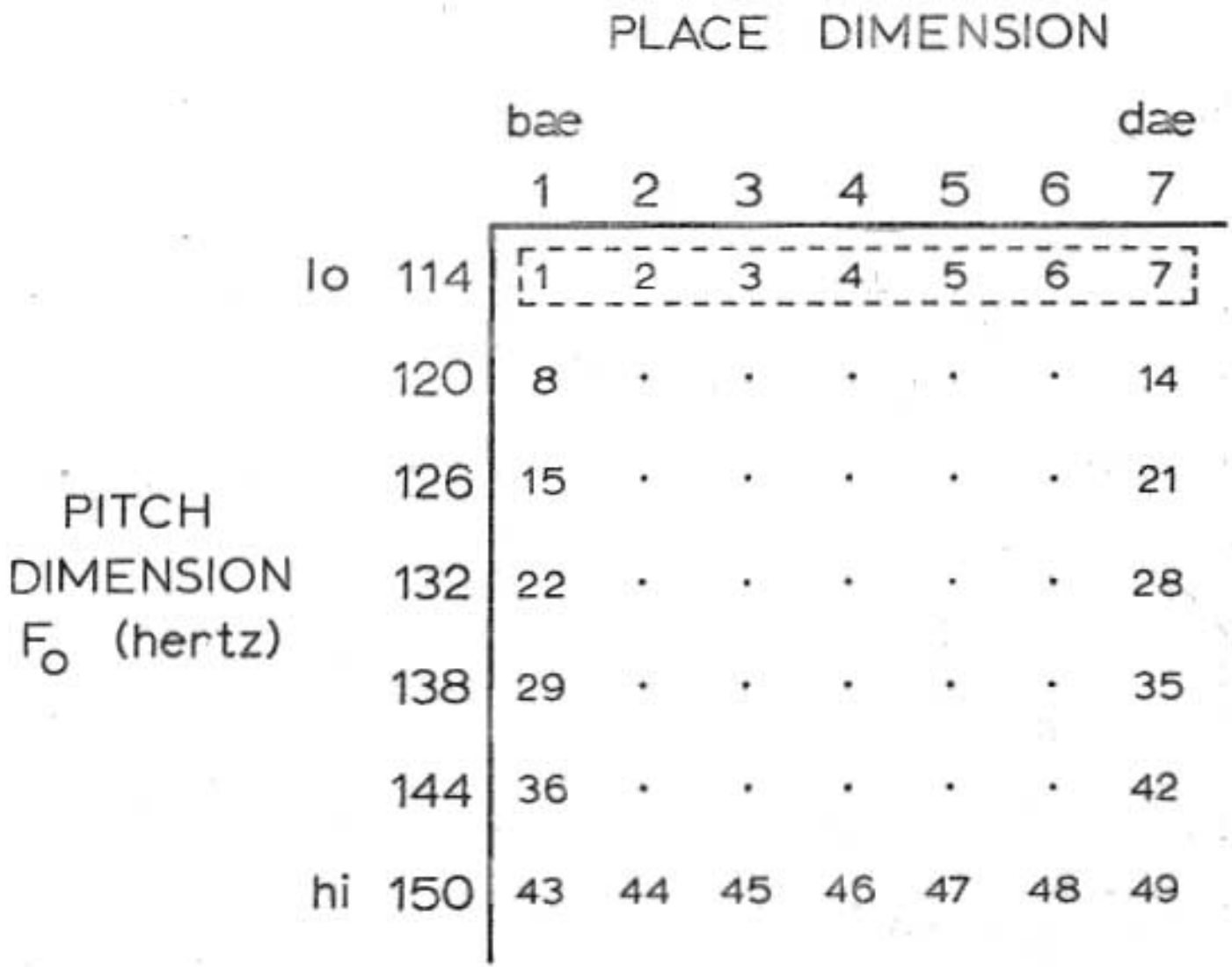

PLACE DIMENSION
bae
dae
1 2 3 4 5 6 7
lo
114
1 2 3 4 5 6 7
120
8 14
126
15 21
132
22 28
138
29 35
144
36 42
hi
150
43 44 45 46 47 48 49
PITCH DIMENSION
F_o (hertz)

Figure 3.

Subjects were divided into four groups of nine subjects each. Each group heard the control tape and one of the anchor tapes. Subjects were told that they would be listening to the syllables /bae/ and /dae/. They listened to examples of the four corner stimuli (a low /bae/, low /dae/, high /bae/ and high /dae/) for practice. Subjects were told the relevant dimension to judge the stimuli on, pitch or place. In the pitch condition they were told to judge the stimuli as high or low pitch. In the place condition they were told to judge the stimuli as bae or dae.

Results and Discussion

Slide 4 please

Slide four shows the average identification functions for the two groups judging pitch. In group I, shown on your left, subjects heard the low anchor tape. The identification function shows a consistent shift toward the low stimulus, relative to the control function. Group II, which heard the high anchor tape, also shows a shift in the identification boundary. The shift is toward the more frequently occurring stimulus, the high pitched one. These results are in accord with the findings obtained earlier using tones.

Slide 5 please

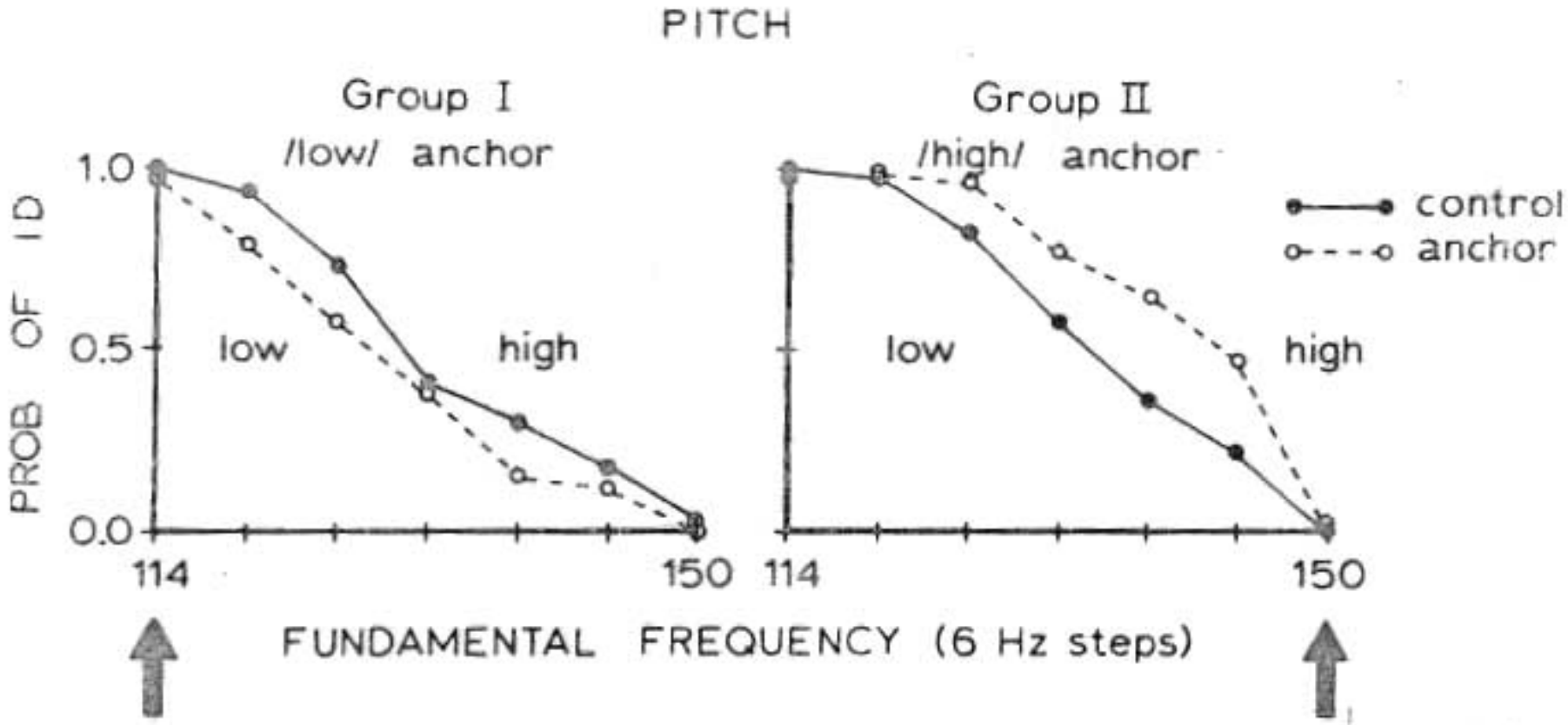

PITCH
Group I
/low/ anchor
Group II
/high/ anchor
control
anchor
PROB OF ID
1.0
0.5
0.0
low
high
low
high
114
150
114
150
FUNDAMENTAL FREQUENCY (6 Hz steps)

Figure 4.

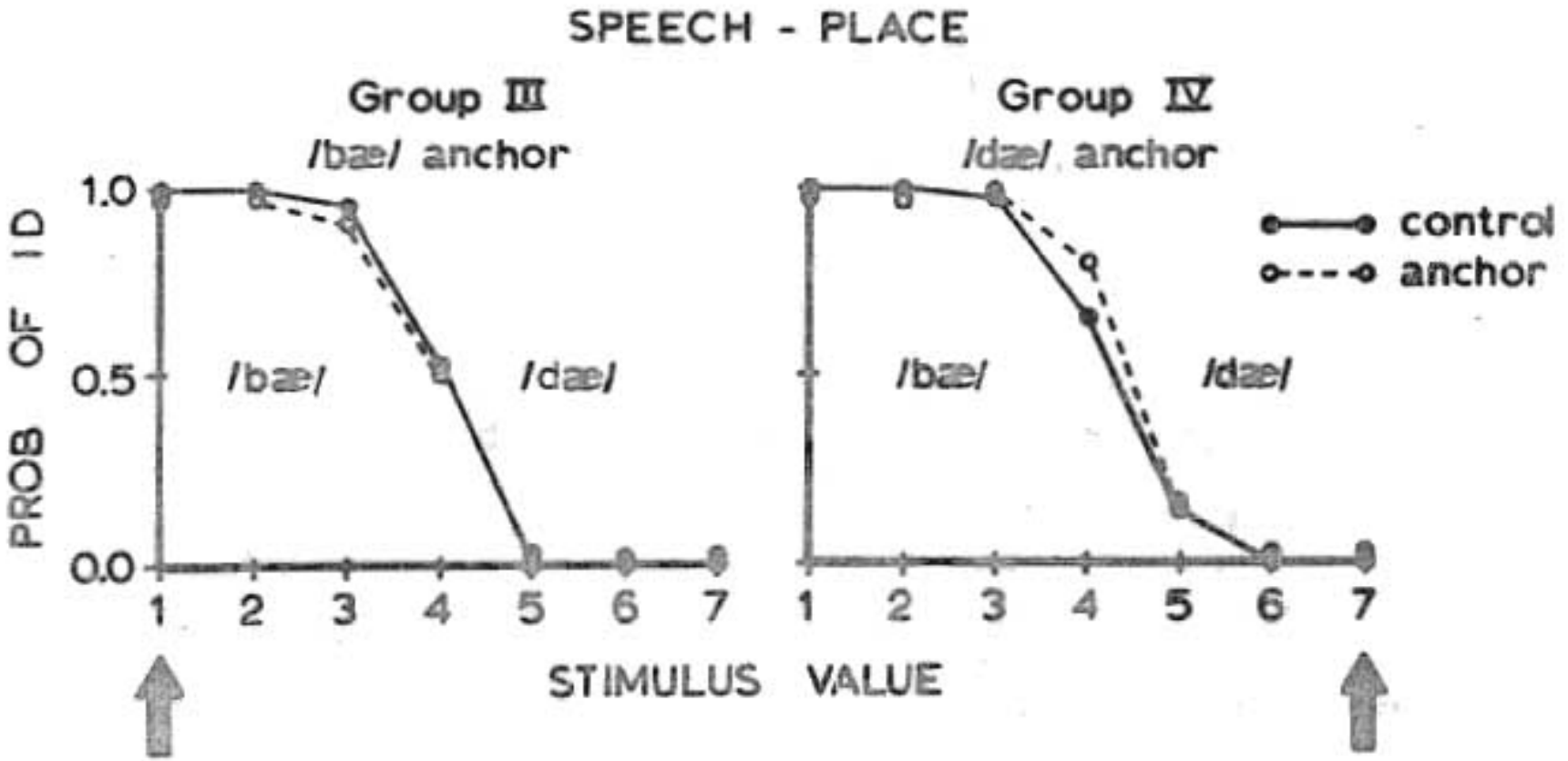

SPEECH - PLACE
Group III
/bae/ anchor
Group IV
/dae/ anchor
PROB OF ID
1.0
0.5
0.0
/bae/
/dae/
/bae/
/dae/
control
anchor
1 2 3 4 5 6 7
1 2 3 4 5 6 7
STIMULUS VALUE

Figure 5.

Slide five shows the averaged identification functions for the two groups juding place. Group III heard the /bae/ anchor tape and Group IV heard the /dae/ anchor tape. These functions with unbalanced probabilities show no shift relative to the control functions. These results are also in accord with those found earlier.

An analysis of variance showed that both the shift and the shift by stimulus dimension interaction were significant beyond .001 level. There is a shift on the non-linguistic pitch dimension and no shift on the linguistic place dimension.

These results support our earlier conclusions that the category boundaries for the phonetic features of place and voicing are non-arbitrary and are not due to a simple partitioning of the stimulus continuum. Even when the same speech stimuli are employed, the non-linguistic dimension of pitch is quite vulnerable to response bias. We would like to suggest that these results rule out a response bias explanation of the category boundary shifts found with the selective adaptation paradigm. The results reported by Eimas, Cooper and others using the selective adaptation paradigm seem to be relatively early perceptual effects rather than changes in the decision mechanism. Whether the shifts are due to the fatiguing of specialized phonetic feature detectors or more generalized auditory processors awaits additional research. However, we think that we have made a good case for ruling out an obvious decision mechanism or response bias explanation of their results.

In summary, when a non-linguistic pitch dimension is being judged, subjects show a shift in their category boundary when the probabilities of occurrence of different stimuli are unequal. The boundary shifts toward the more frequently occurring stimulus. No shift is found when subjects judge the linguistic dimensions of place of production or voicing.

Information Processing and Speech Perception*

David B. Pisoni

Indiana University

Bloomington, Indiana 47401

This paper discusses a number of important concepts within the framework of human information processing and their relevance to speech sound perception. A rough outline of a model of speech perception is described which incorporates the distinctions between sensory, short and long term memory and hierarchical stages of processing. Central to this approach is the continuity of processing and the interrelations between stages of analysis.

* This paper was prepared for the Speech Communication Seminar, Stockholm, August 1 - 3, 1974 and will appear in the proceedings which are to be published by Almqvist & Wiksell and John Wiley & Sons. This work was supported in part by USPHS NIMH Research Grant MH-24027-01 and in part by a Faculty Fellowship from the office of Research & Advanced Studies, Indiana University.

Information Processing and Speech Perception

David B. Pisoni
Indiana University
Bloomington, Indiana 47401 U.S.A.

Introduction

Current theories of speech perception have been quite general and vague, and for the most part, not terribly well developed (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Stevens & House, 1972). Indeed, it is probably fair to say that most of the current theoretical approaches to speech perception are only preliminary "guesses" at what a possible model of the speech perception process might entail. A few quotes should make this point clear. For example, in 1964 Liberman and his colleagues stated at the Symposium on Models for the Perception of Speech and Visual Form that:

> "Since this symposium is concerned with models, we should say at the outset that we do not have a model in the strict sence, though we are in search of one." (Liberman, Cooper, Harris, MacNeilage & Studdert-Kennedy, 1967, p. 68).

At the same meeting, Fant stated that:

> "Any attempt to propose a model for the perception of speech is deemed to become highly speculative in character and the present contribution is no exception." (Fant, 1967, p. 111).

And even more recently, Stevens & House stated in their chapter on Speech Perception that:

> "Since we are still far from an understanding of the neurophysiological processes involved, any model that can be proposed must be a functional model, and one can only speculate on the relation between components of the functional model and the neural events at the periphery of the auditory system and in the central nervous system." (Stevens & House, 1972, p. 47).

Although most investigators agree that the perception of speech sounds may involve processes and mechanisms that are in some way basically different from those employed in the perception of other sounds, very little work has been directed at specifying these differences. A large body of experimental work obtained over the last twenty years suggests that when listeners are presented with speech stimuli their ability to identify and discriminate these sounds on an auditory basis alone is limited to a very substantial degree by their linguistic knowledge. Differences in the perception of speech and non-speech stimuli and differences in perception among various classes of speech sounds have led numerous investigators to propose a special "speech perception mode" (Stevens & Halle, 1967; Liberman, 1970a,b). Other findings employing dichotic listening techniques have suggested that a specialized perceptual mechanism--"a special speech decoder" may exist as a distinct physiological entity for the processing of speech sounds (Studdert-Kennedy & Shankweiler, 1970). Other evidence has been accumulated to suggest that speech perception may involve some sort of active mediation of motor centers associated with speech production (Liberman, Cooper, Harris & MacNeilage, 1963).

In this paper, I consider some of the perceptual processes involved in speech recognition and then describe a rough model for speech sound perception based on recent work in human information processing.

Information Processing Approach

In recent years the study of speech perception has begun to adapt the aims and methods of human information processing models which have been employed quite successfully in the study of visual and auditory perception (Neisser, 1967; Haber, 1969; Massaro, 1972; Reed, 1973). This approach

views perception as a hierarchically organized sequence of events involving stages of storage and transformations of information over time. As Neisser points out, during these stages information is "transformed, reduced, elaborated, stored, recovered and used." A major assumption of this approach to perception is the continuity of different levels of processing. Sensation, perception, memory and thought are considered to be on a continuum of cognitive activity. These stages are thought to be mutually inter-dependent. Furthermore, it is argued that one can only understand perception, especially recognition, identification and perceptual memory by attempting to understand the whole range of these cognitive processes.

Since the information processing approach is fundamental to our approach to speech perception, I will first describe some of the major concepts involved. Then I will describe some of the stages of processing speech perception and then provide some of the details of the current model.

There are three basic assumptions in current information processing models. First, perception is not immediate but is the outcome of distinct operations distributed over time. One goal of information processing models is to attempt to specify the operations which occur from the presentation of a stimulus to the overt response of the observer. The various stages which lie between input and output are typically represented by a flow chart with block design. Much of the recent work on backward masking has been concerned with this question (Pisoni, 1972, 1974; Massaro, 1974).

The second assumption is that there are "capacity limitations" at various stages of processing. Because the nervous system cannot maintain all aspects of sensory stimulations and must integrate energy over time, limits on the

capacity to store and process sensory data occur which require that information be recoded into a different more abstract form. One goal of research in this area has been to identify the locus of these capacity limitations. For example, the recent work by Shiffrin and myself has been specifically directed at this problem (see Shiffrin, Pisoni & Castenada-Mendez, 1974).

The third assumption is that perception necessarily involves various types of memorial processes since recoding and retention of information will occur at all stages of information processing. Hence, the study of speech perception necessarily entails the study of perceptual memory. The work of Fujisaki and myself has shown the importance of short-term memory in speech perception (Fujisaki & Kawashima, 1970; Pisoni, 1971, 1973).

Sensory Memory, Short-term Memory and Long-term Memory

Central to information processing analyses is the motion of an iconic, echoic, or pre-perceptual memory store. This is typically thought of as a very temporary storage medium which preserves all of the stimulus information in a literal or veridical form for several hundred milliseconds. During this time period, the information is converted into a more persistent and abstract form for representation in short-term storage. Short-term (STS) or "working memory" is thought to have a very limited capacity from which information is rapidly lost unless active rehearsal or control processes are operating. Long-term store (LTS) on the other hand, is assumed to be the permanent repository for information. It has an unlimited capacity. Long-term store receives information from short-term store. The process of rehearsal of information in short-term store first regenerates the rapidly decaying memory traces and also causes information in short-term store to be transferred to long-term store.

Recognition is assumed to be a process whereby the sensory input or some derived version of it "makes contact" with a stored representation in long-term memory or some type of representation that has been constructed or generated by rules in long-term memory. Thus, recognition is assumed to take place in short-term memory. The information present in short-term memory is thought to consist of a combination of information from both the sensory input and information from long-term memory.

Sensory information is not simply transferred to short-term store but is "recoded" while still being maintained at the earliest stage of processing. It is generally assumed that the earliest stages of the recognition process occur "automatically" and without conscious control by the subject (see for example, Shiffrin & Geisler, 1973). A good part of the information from the earliest stages of processing is lost by decay and only a relatively abstract representation of the input is maintained in short-term memory.

Stages of Processing in Speech Perception

A number of recent accounts of speech perception have begun to emphasize process and to divide this process into a hierarchy of stages: auditory, phonetic, phonological etc. (see for example, Liberman, 1970; Studdert-Kennedy, 1974a,b; Studdert-Kennedy, Shankweiler & Pisoni, 1972; Wood, 1973). Figure 1 shows some of the processes which are assumed to take place between the initial

---

Insert Figure 1 about here

---

acoustic signal and its final conceptual representation. According to this view, the speech signal undergoes a series of successive transformations whereby information is recoded into more and more abstract forms of representation

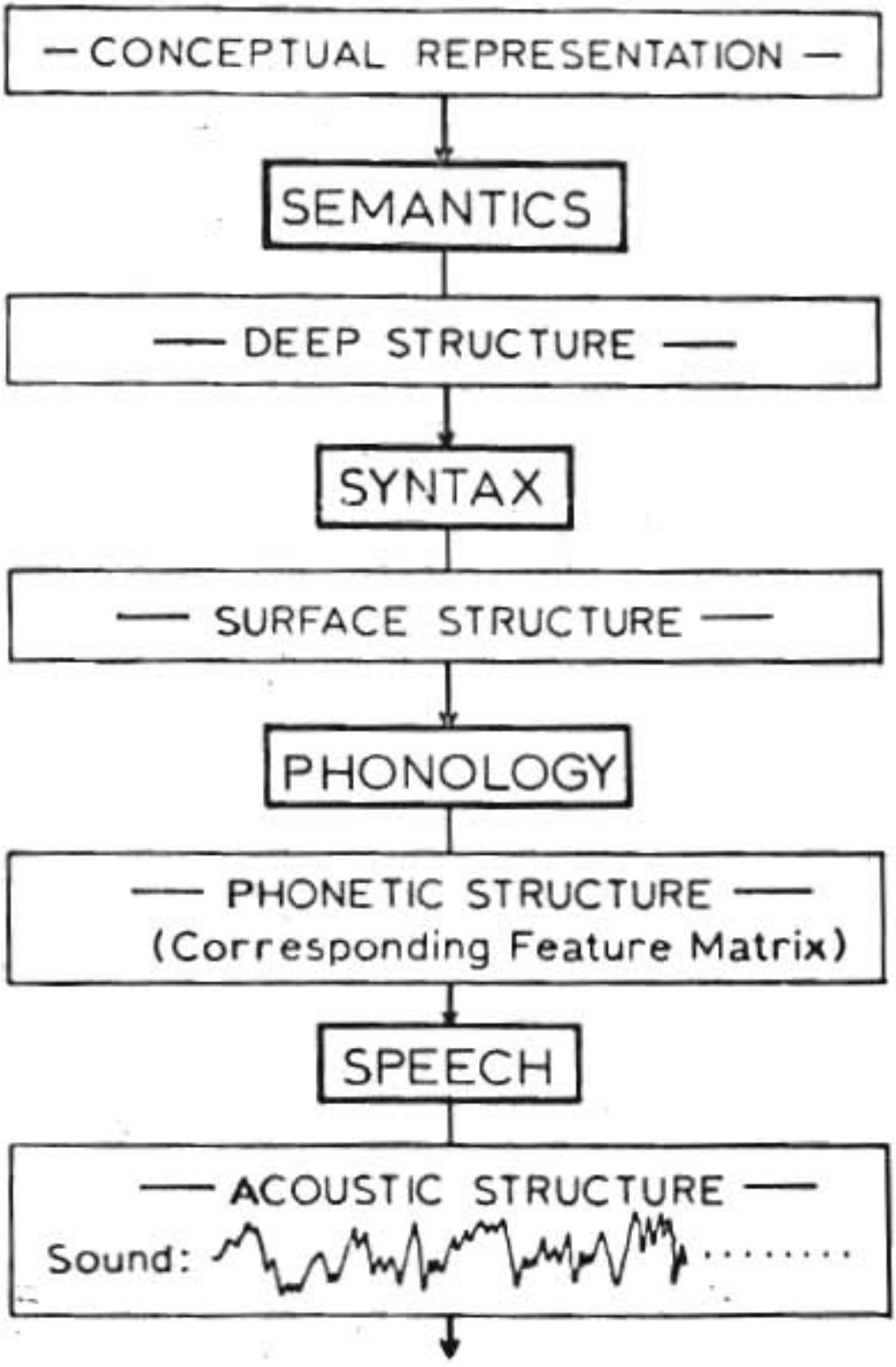

— CONCEPTUAL REPRESENTATION —
SEMANTICS
— DEEP STRUCTURE —
SYNTAX
— SURFACE STRUCTURE —
PHONOLOGY
— PHONETIC STRUCTURE —
(Corresponding Feature Matrix)
SPEECH
— ACOUSTIC STRUCTURE —
Sound:

Figure 1

(see also Liberman, 1970). The stages are thought to be partially successive since spoken language is inherently a temporal phenomena. However, decisions at various stages must also take place in parallel to permit information from higher levels to be employed in processes at lower levels.

Although the distinction between phonetic structure and higher levels of analysis is commonly accepted in linguistic theory, the distinction between auditory (i.e., acoustic structure) and phonetic levels of analysis has not been widely recognized. The auditory stage may be thought of as the first level of analysis. At this stage the acoustic waveform is transformed or recoded into some "time-varying" neurological pattern of events in the auditory system. Acoustic information such as spectral structure, fundamental frequency, intensity and duration is extracted by the auditory system. All subsequent stages of analysis beyond the auditory stage are thought to be abstract and based on these "auditory features." The phonetic stage is closely related to auditory analysis. Here, segments and features necessary for phonetic classification are abstracted or derived from the auditory features of the acoustic signal. At the output of this stage, the continuously varying acoustic stimulus has been transformed into a sequence of discrete phonetic segments. Information about the feature specification of these phonetic segments is then passed on to higher levels of processing for phonological and syntactic analysis.

Thus, the auditory level may be characterized as that portion of the speech perception process which is "non-linguistic," and therefore includes processes and mechanisms that operate on speech and non-speech signals alike. On the other hand, processes and mechanisms at the phonetic level are assumed

to perform a linguistic abstraction process whereby a particular phonetic feature is identified or recognized from some configuration of auditory features.

An Information Processing Model

In this section, I will briefly sketch the structure of the information processing model. Figure 2 shows a block diagram of the components. Auditory

---

Insert Figure 2 about here

---

input enters the system and is processed in progressive stages. The output of Preliminary Auditory Analysis is assumed to be some type of spectral display in terms of frequency, time, and intensity. Sensory input is processed automatically through several levels of analysis without the operation of conscious selective attention. Sensory information is maintained in a relatively gross unanalyzed form in the Sensory Information Store (SIS). Information is further processed by a "recognition device" which is shown as four distinct stages in this figure. Information from any or all of these stages of processing is placed in short-term store where the subject can selectively rehearse, encode or make decisions about it. It is assumed that information in long-term store is employed in the recognition process.

Automatic processing by the recognition device is assumed to take place as follows. In Stage 1, Acoustic Feature Analysis, we assume that auditory features of the speech signal are recognized by a system of individual auditory feature detectors (Stevens, 1973). For example, in the case of a simple CV syllable, we assume that specialized auditory detectors will respond selectively to at least some of the following types of information: (a) presence or absence

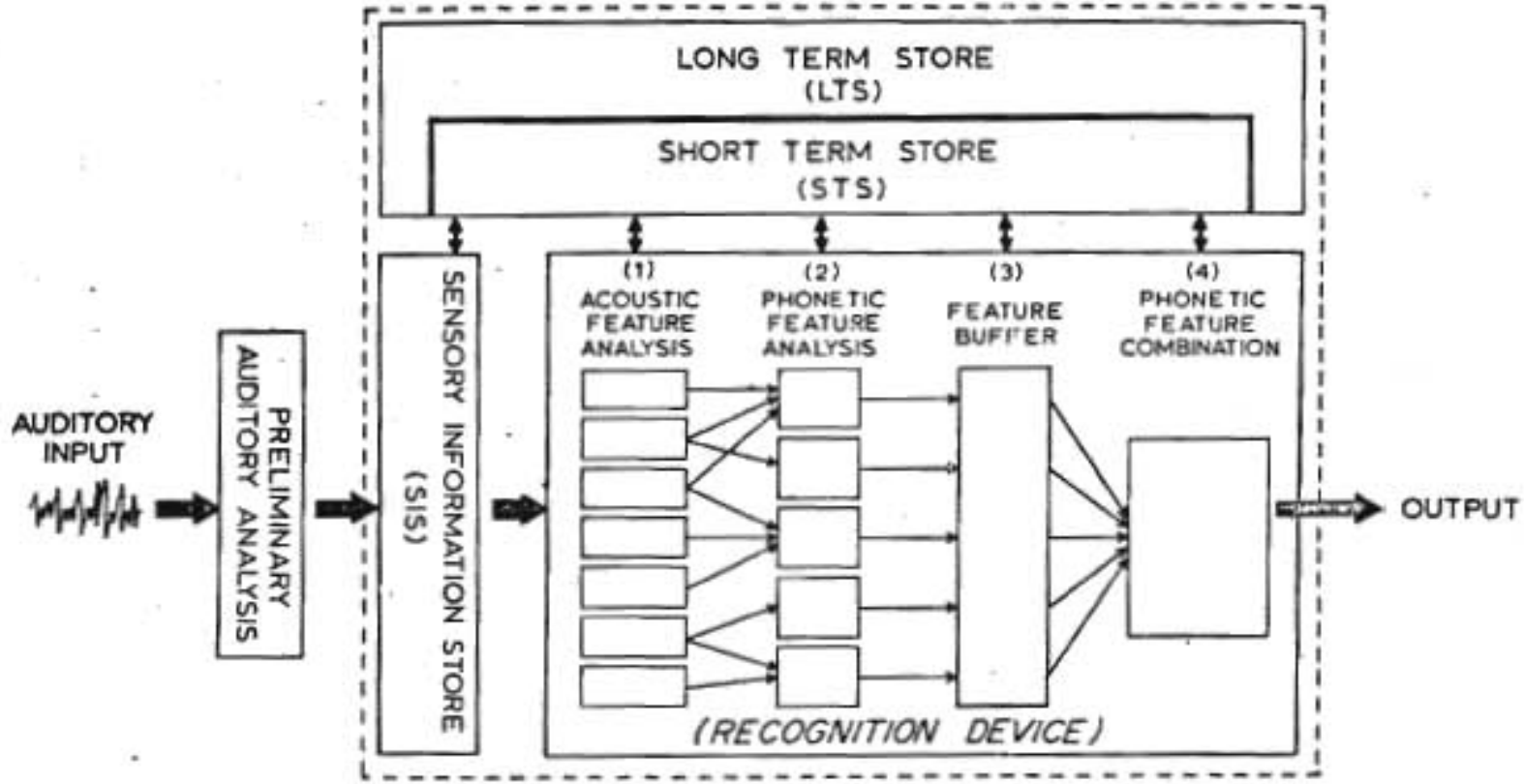

LONG TERM STORE
(LTS)
SHORT TERM STORE
(STS)
AUDITORY
INPUT
PRELIMINARY
AUDITORY ANALYSIS
SENSORY INFORMATION STORE
(SIS)
(1)
ACOUSTIC
FEATURE
ANALYSIS
(2)
PHONETIC
FEATURE
ANALYSIS
(3)
FEATURE
BUFFER
(4)
PHONETIC
FEATURE
COMBINATION
OUTPUT
(RECOGNITION DEVICE)

Figure 2

of a rapid change in the spectrum; (b) direction, extent, and duration of a change in the spectrum; (c) duration and intensity of noise, etc. The output of Acoustic Feature Analysis is some set of acoustic cues or auditory features which forms the input to the next stage of processing.

In Stage 2, Phonetic Feature Analysis, we assume that a set of decision rules is employed to map multiple auditory features into phonetic features. It is assumed that this is a many-to-one mapping where several different auditory features provide information about a particular phonetic feature. The output of phonetic feature analysis is a set of abstract phonetic features. These decision rules can be thought of as having knowledge of articulatory constraints in production although it is not essential for the model in its present form.

These features are subsequently maintained in Stage 3, the Feature Buffer. This may be thought of as simply a holding mechanism which maintains decisions about the feature composition of a particular syllable. There are two reasons for postulating a feature buffer. First, not all phonetic features are assumed to be processed (i.e., recognized) at the same rate. Secondly, some memorial process is needed to preserve and maintain phonetic feature information more-or-less independently for subsequent stages of linguistic processing (e.g., phonological).

Feature information is then used in Stage 4, Phonetic Feature Combination, where individual features are recombined to form discrete phonetic segments. The output of Stage 4 is a phonetic segment, where the feature specification is, for example, some form of an abstract distinctive feature matrix. This information is then passed on to higher levels of processing for phonological and syntactic analysis.

The model as I have described it thus far is still preliminary and a number of changes and revisions will obviously be required. However, I think it has much to offer as a framework for dealing with past research and providing a basis for future work. It combines the virtues of recent information processing models with their emphasis on stages of processing, memory and recoding of information. It also incorporates the recent distinctions between auditory and phonetic stages of processing in speech perception. Finally, I think such a model can be used to generate new and important questions about speech perception that can be tested empirically. For example, what is the general organization of auditory and phonetic stages of processing and the nature of the interaction between them? What is the locus or stage of processing at which processing peculiar to speech is initiated and what is the nature of these perceptual operations?

In a more global sense, the model can be used to specify the ways in which speech sounds may require specialized neural mechanisms for perceptual processing and the ways it may conform to more general principles of human information processing common to other modalities.

References

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

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

Haber, R.N. Information-Processing Approaches to Visual Perception. New York: Holt, Rinehart and Winston, 1969.

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

Liberman, A.M. Some characteristics of perception in the speech mode. In D.A. Hamburg (Ed.) Perception and Its Disorders, Proceedings of A.R.N.M.D. Baltimore: Williams and Wilking Co., 1970. Pp. 238-254.

Liberman, A.M., Cooper, F.S., Harris, K.S., and MacNeilage, P.F. A motor theory of speech perception. In C.G.M. Fant (Ed.), Proceedings of the Speech Communication Seminar, Stockholm, 1962. Speech Transmission Laboratory, Royal Institute of Technology, Stockholm, 1963.

Liberman, A.M., Cooper, F.S., Harris, K.S., MacNeilage, P.F., and Studdert-Kennedy, M. Some observations on a model for speech perception. In W. Wathen-Dunn (Ed.), Models for the Perception of Speech and Visual Form. Cambridge: M.I.T. Press, 1967.

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

Massaro, D.W. Preperceptual Images Processing Time, and Perceptual Units in Auditory Perception. Psychological Review, 1972, 79, 2, 124-145.

Massaro, D.W. Perceptual units in speech recognition. Journal of Experimental Psychology, 1974, 102, 2, 199-208.

Neisser, U. Cognitive Psychology. New York: Appleton, 1967.

Pisoni, D.B. On the Nature of Categorical Perception of Speech Sounds. Status Report on Speech Research (SR-27), Haskins Laboratories, New Haven, 1971, 101.

Pisoni, D.B. Perceptual processing time for consonants and vowels. Paper presented at the 84th meeting of the Acoustical Society of America, Miami Beach, Fla., December, 1972.

Pisoni, D.B. Auditory and Phonetic Memory Codes in the Discrimination of Consonants and Vowels. Perception and Psychophysics, 1973, 13, 2, 253-260.

Pisoni, D.B. Dichotic Listening and Processing Phonetic Features. In F. Restle, R.M. Shiffrin, N.J. Castellan, H. Lindman, and D.B. Pisoni (Eds.), Cognitive Theory: Volume I. Potomac, Maryland: Erlbaum Associates (1974, In Press).

Reed, S.K. Psychological Processes in Pattern Recognition. New York: Academic Press, 1973.

Shiffrin, R.M. & Geisler, W.S. Visual Recognition in a Theory of Information Processing. In R. Solso (Ed.), The Loyola Symposium: Contemporary Viewpoints in Cognitive Psychology. Washington: Winston, 1973.

Shiffrin, R.M., Pisoni, D.B. and Castaneda-Mendez, K. Is attention shared between the ears? Cognitive Psychology, 1974, 6, 2, 190-215.

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

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

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

Studdert-Kennedy, M. The Perception of Speech. In T.A. Sebeok (Ed.), Current trends in linguistics, Volume XII, The Hague: Mouton, 1974(a).

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

Studdert-Kennedy, M. and Shankweiler, D.P. Hemispheric Specialization for Speech Perception. Journal of the Acoustical Society of America, 1970, 48, 2, 579-594.

Studdert-Kennedy, M., Shankweiler, D., and Pisoni, D.B. Auditory and Phonetic Processes in Speech Perception: Evidence from a Dichotic Study. Cognitive Psychology, 1972, 3, 455-466.

Wood, C.C. Levels of Processing in Speech Perception: Neurophysiological and Information-processing Analyses. Unpublished doctoral dissertation. Yale University, 1973. (Also appears in Status Report on Speech Research, SR-35/36, Haskins Laboratories, New Haven.)

Publications:

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

Pisoni, D. B. & Lazarus, J. H. Categorical and noncategorical modes of speech perception along the voicing continuum. Journal of the Acoustical Society of America, 1974, 55, 2, 328-333.

Sawusch, J. R. & Pisoni, D. B. On the identification of place and voicing features in synthetic stop consonants. Journal of Phonetics, 1974, 2, 3, 201-214.

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

Manuscripts to be published:

Pisoni, D. B. Auditory and phonetic memory codes in speech discrimination. In I. G. Mattingly (Ed.) The speech code: Readings in speech perception. (In Press)

Pisoni, D. B. Auditory short-term memory and vowel perception. Memory and Cognition, 1974 (In Press).

Pisoni, D. B. Review of "Acoustic Cues for Constituent Structure" by Robert Scholes. Linguistics (In Press).

Pisoni, D. B. & McNabb, S. D. Dichotic Interactions and Phonetic Feature Processing. Brain & Language, 1974, 1, (In Press).

Pisoni, D. B. Dichotic Listening and Processing Phonetic Features. In F. Restle, R. M. Shiffrin, N. J. Castellan, H. Lindman, & D. B. Pisoni (Eds.) Cognitive Theory: Volume I. Potomac, Maryland: Erlbaum Associates (In Press).