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 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.

Insert Figure 1 about here

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

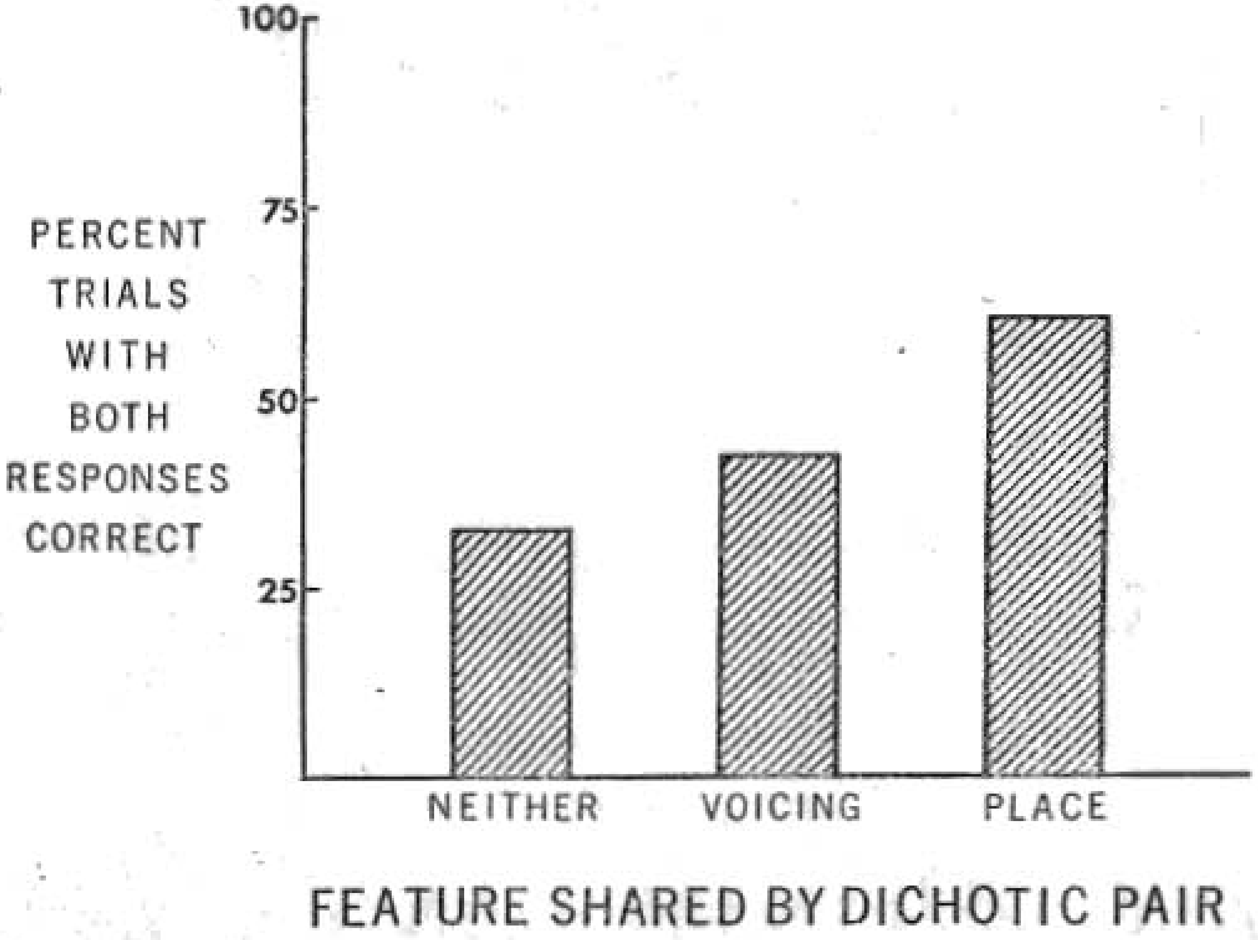
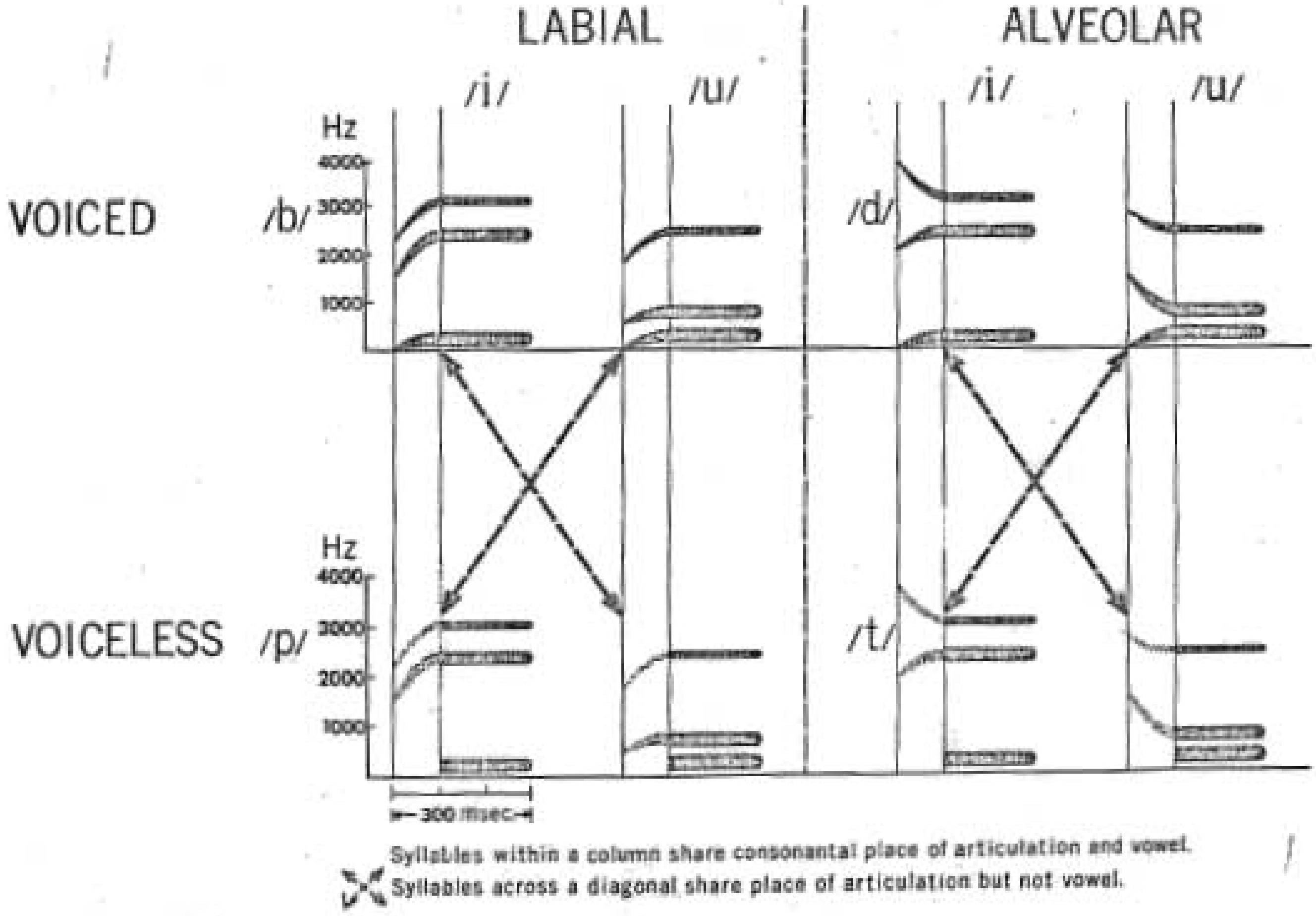


Figure 1. The percentage of trials on which both responses were correct as a function of the consonant feature shared by the dichotic CV pairs (after Studdert-Kennedy, Shankweiler & Pisoni, 1972).

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 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)."

Insert Figure 2 about here

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.



Syllables within a column share consonantal place of articulation and vowel.

Syllables across a diagonal share place of articulation but not vowel.

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 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).

Insert Figure 3 about here

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 ± 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

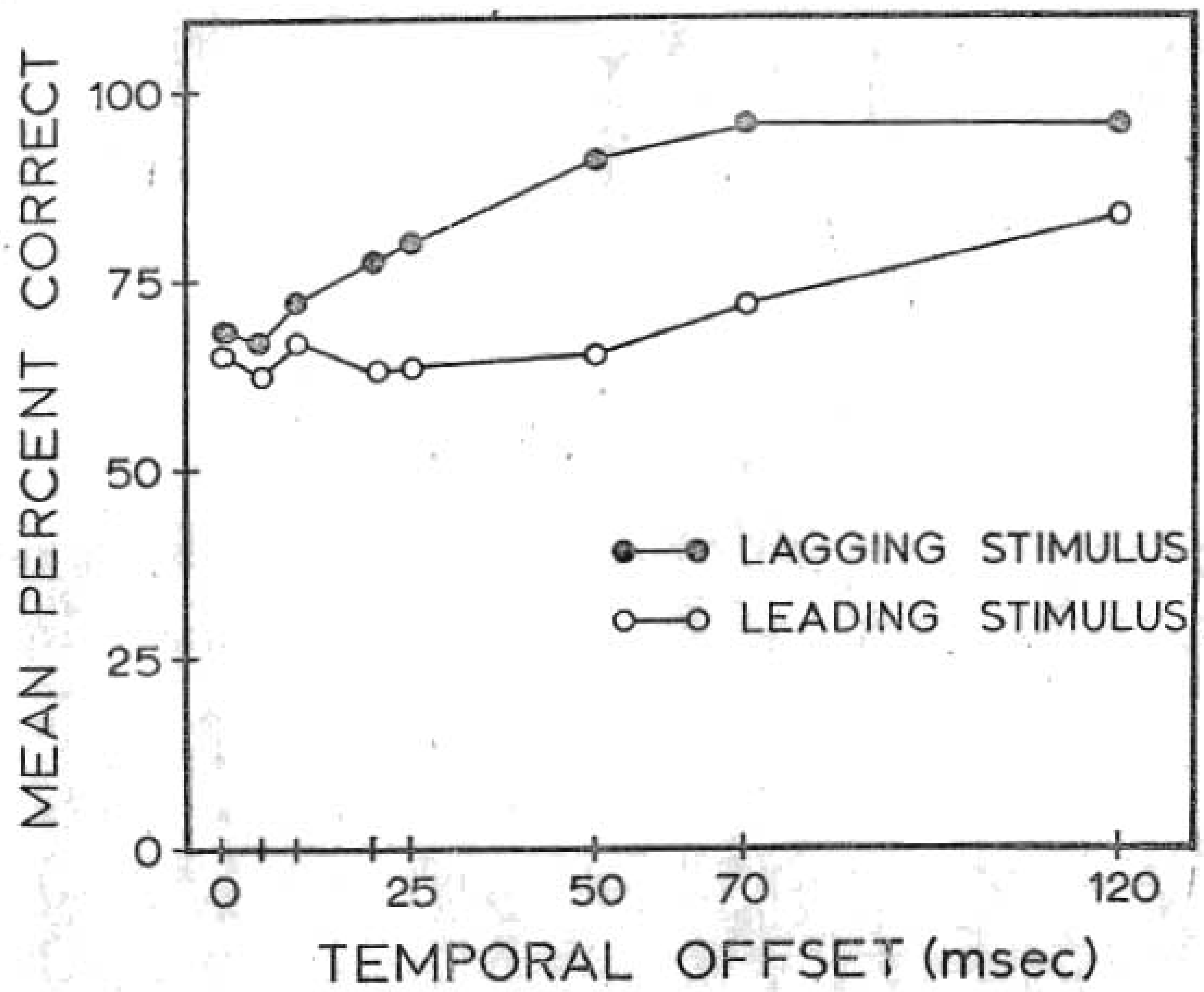


Figure 3. Mean percent correct for leading and lagging dichotically presented CV syllables based on the data from Studdert-Kennedy, Shankweiler & Schulman, 1970.

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]

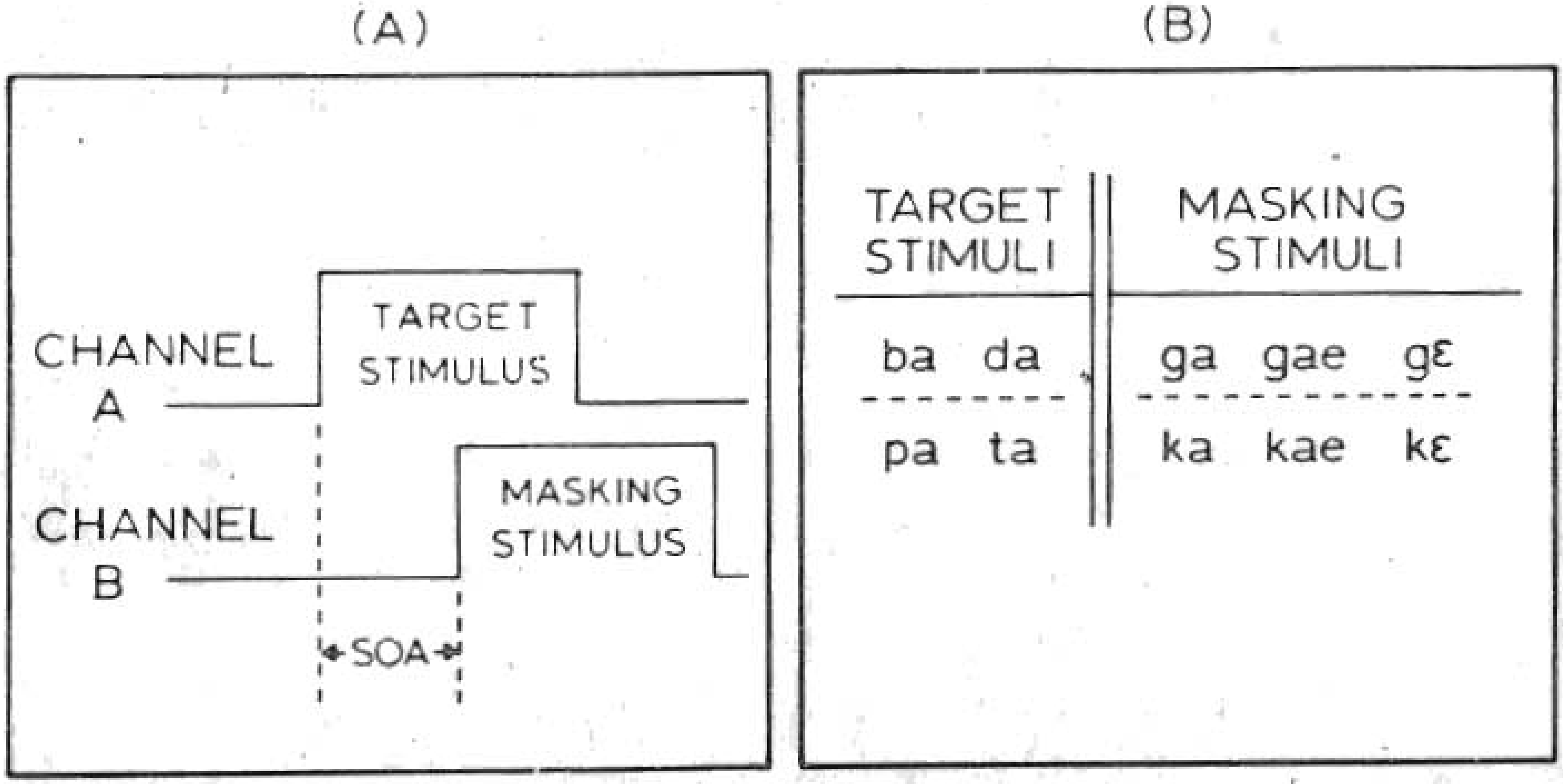


Figure 4. Arrangement of target and masking stimuli in the dichotic backward masking paradigm (Panel A) and the targets and masks (Panel B) used in the experiments. Targets and masks drawn from the same row share the feature of voicing whereas targets and masks drawn from different rows contrast on voicing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

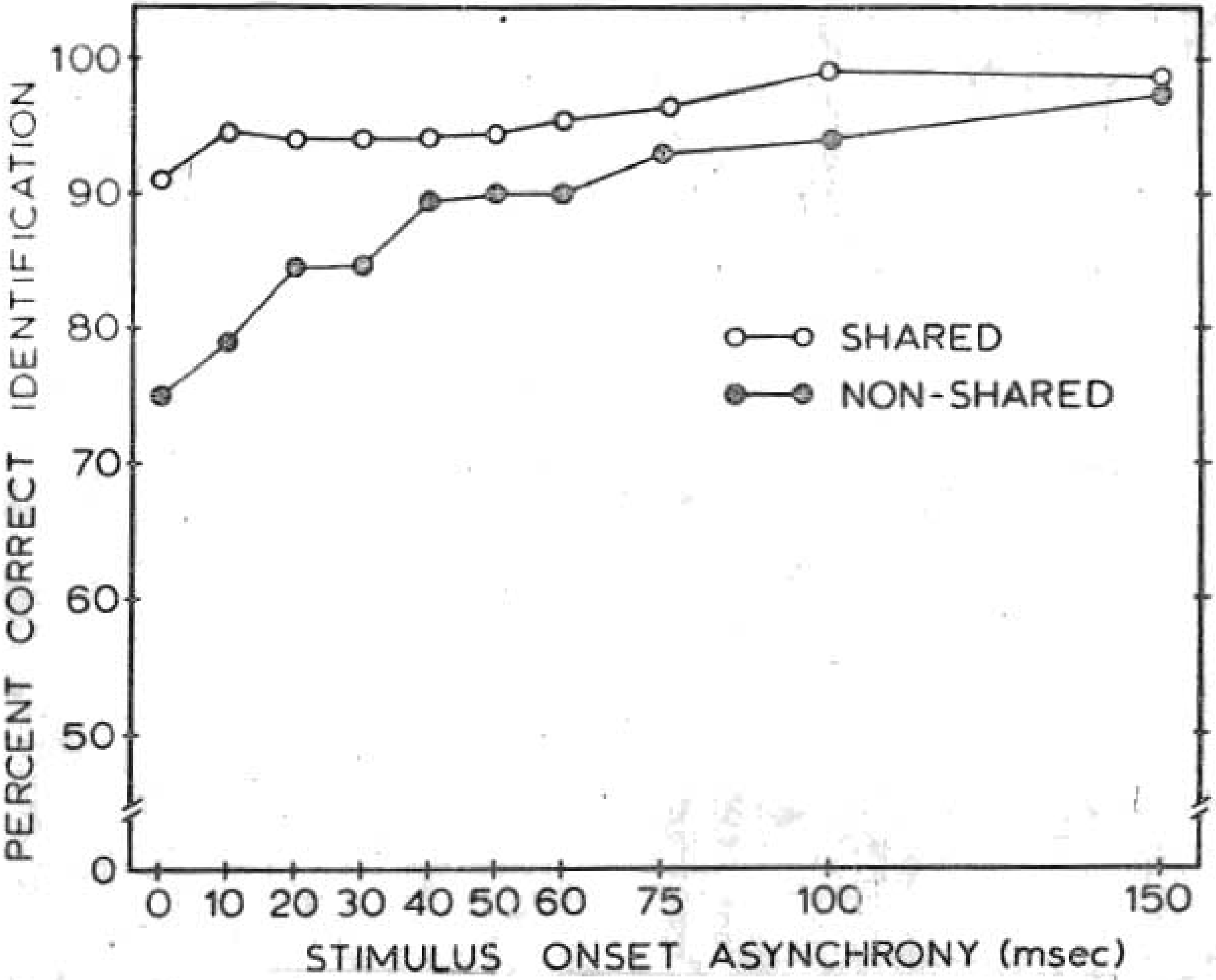


Figure 5. Percent correct identification of target stimuli for shared and non-shared trials as a function of stimulus onset asynchrony. The data are averaged over the three vowel masks (after Pisoni & McNabb, 1974).

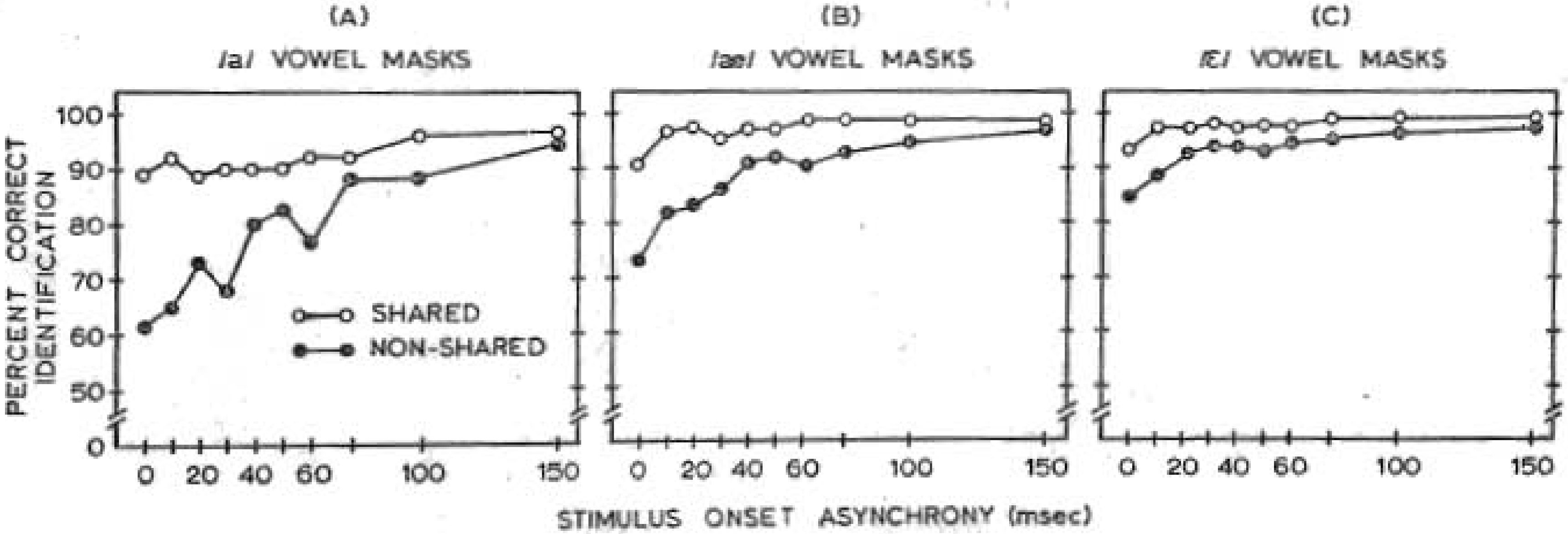


Figure 6. Percent correct identification of target stimuli for shared and non-shared trials under each vowel mask (after Pisoni & Mcnabb, 1974).

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

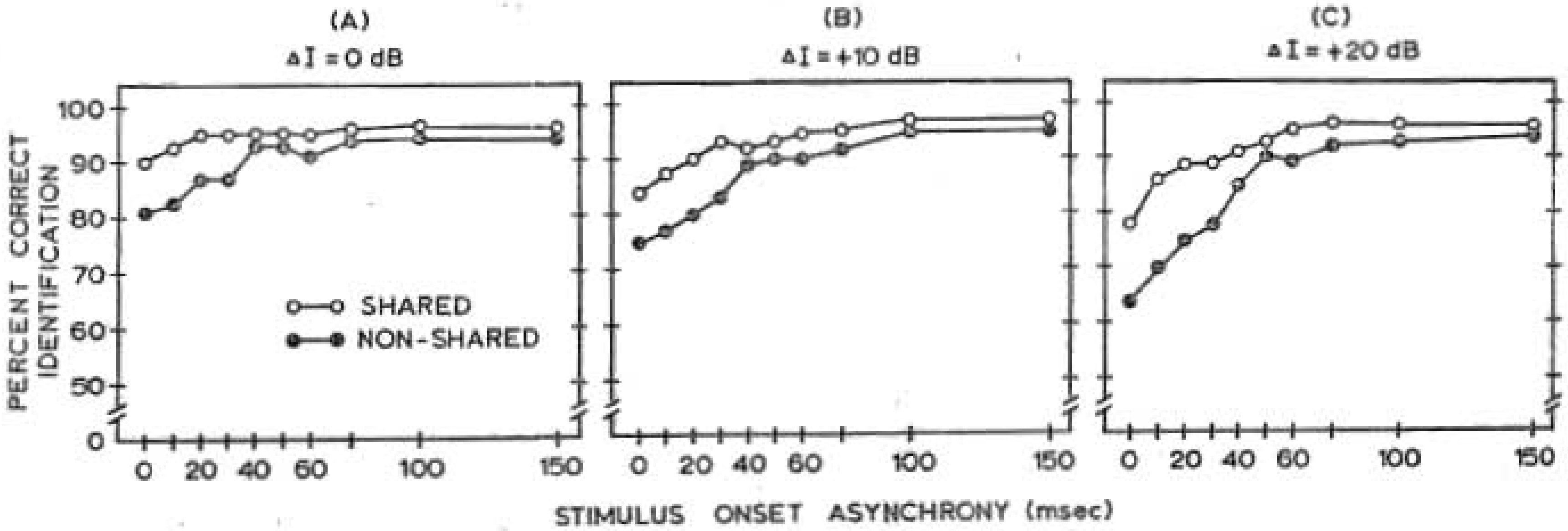


Figure 7. Percent correct identification of target stimuli for shared and non-shared trials at each of three mask intensity levels (after Pisoni & McNabb, 1974).

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 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.

Insert Figure 8 about here

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

Insert Table 1 about here

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

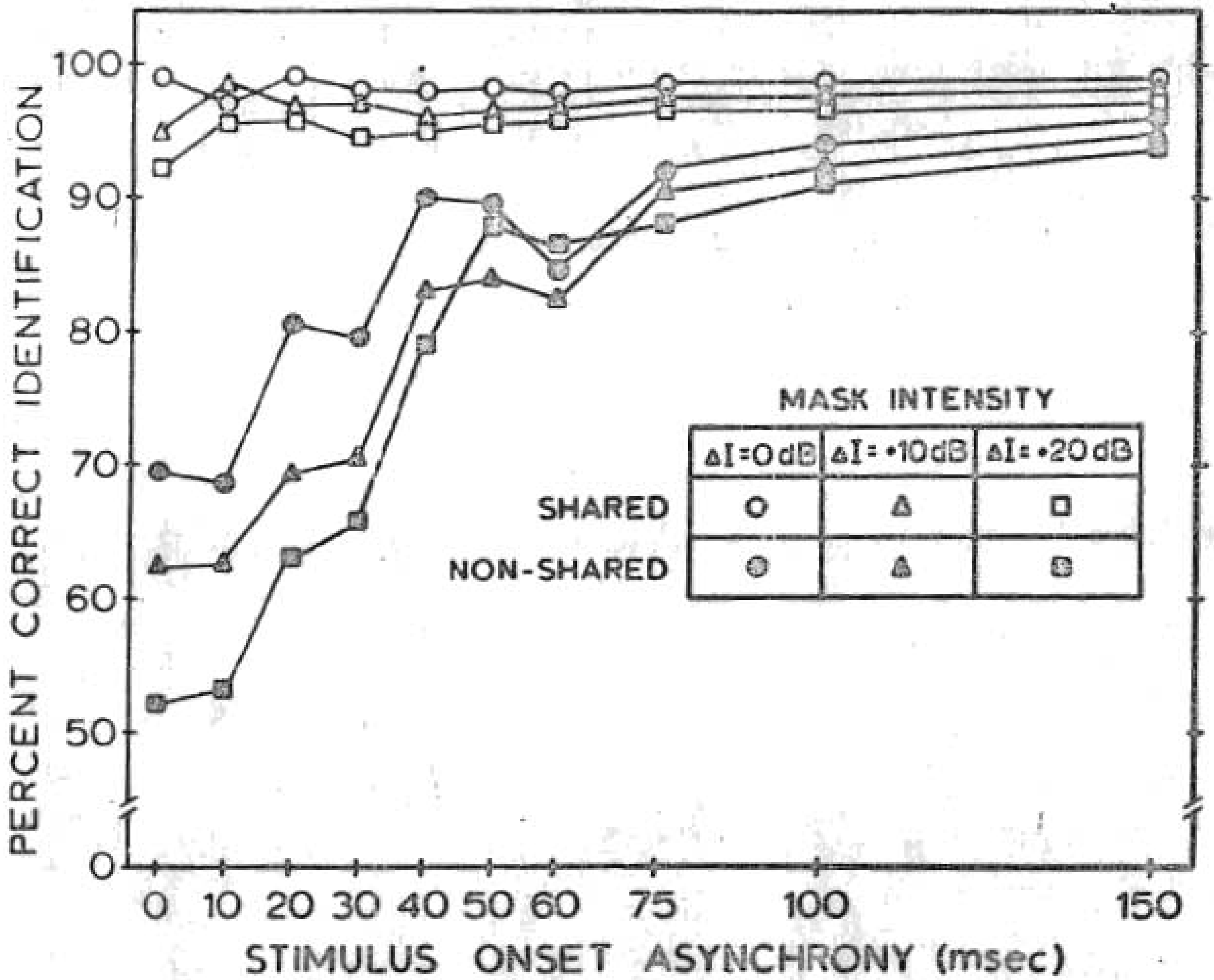


Figure 8. Percent correct identification of the voicing feature on shared and non-shared trials as a function of mask intensity from the /a/ vowel mask condition (after Pisoni & McNabb, 1974).

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

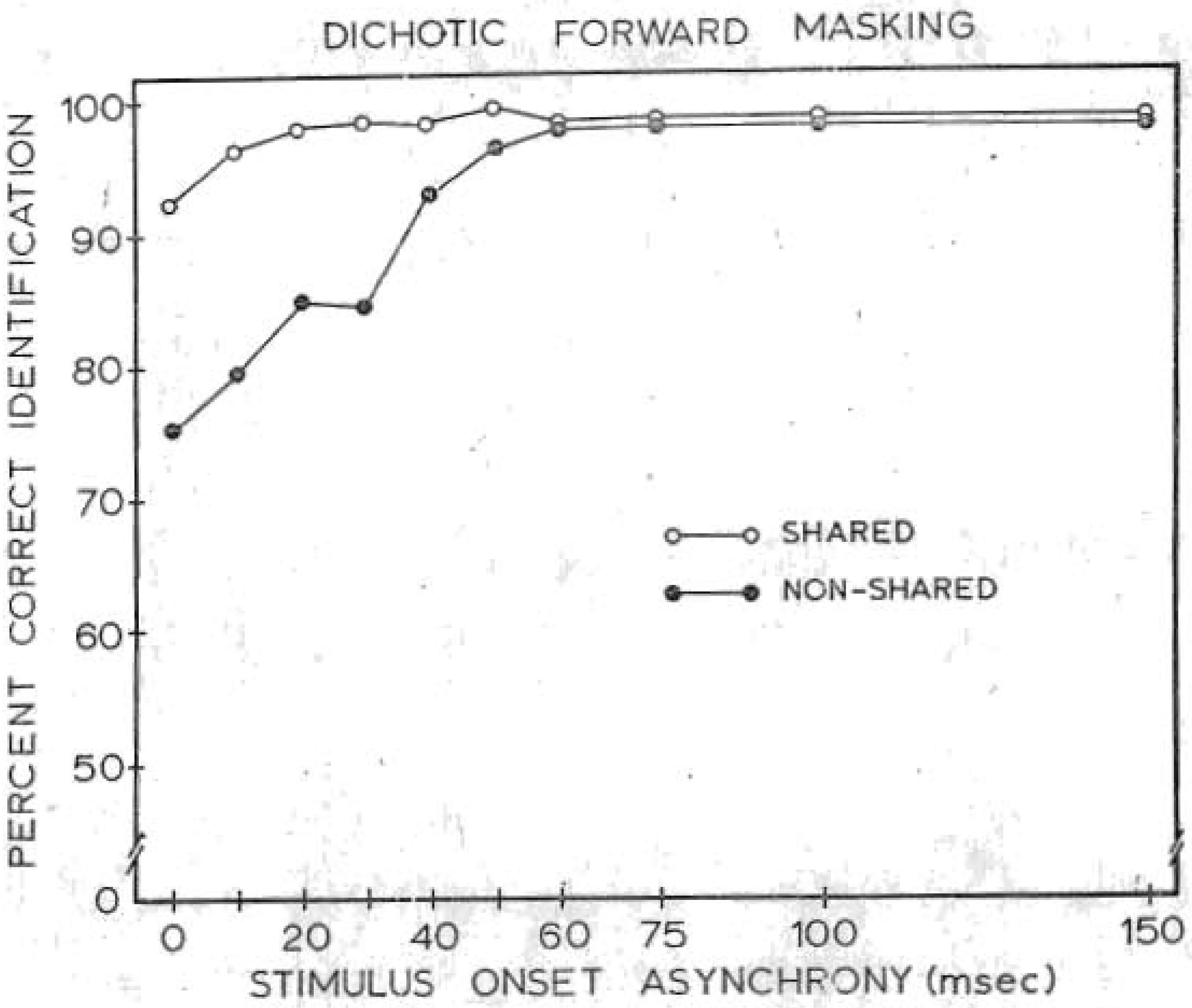


Figure 9. Percent correct identification of target stimuli in forward masking condition for shared and non-shared trials as a function of stimulus onset asynchrony.

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 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.

Insert Figure 10 about here

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 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).

Insert Figure 11 about here

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

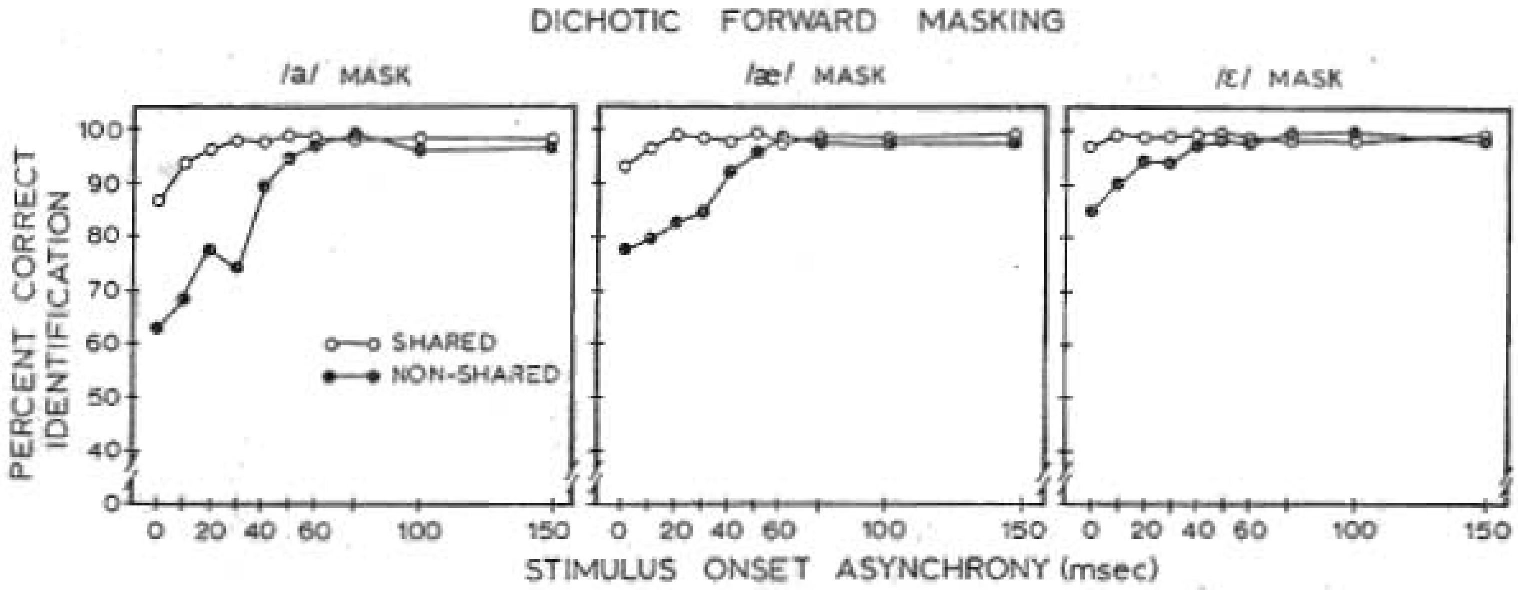


Figure 10. Percent correct identification of target stimuli in forward masking condition under each vowel condition.

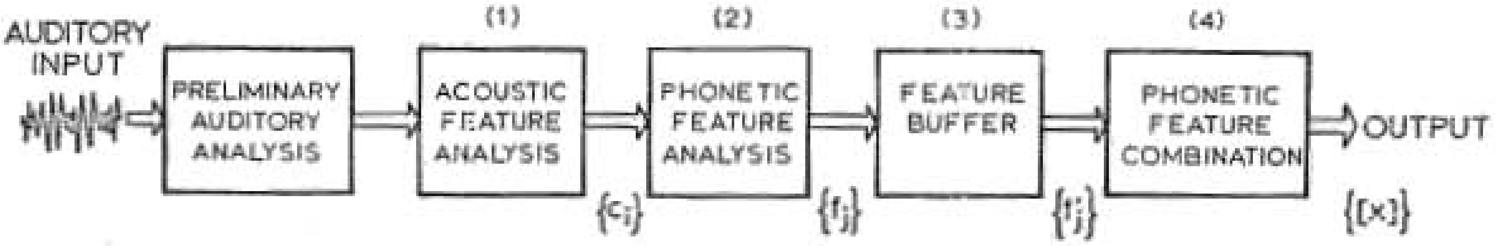


Figure 11. Stage model of levels of processing in phonetic recognition. Auditory input is processed progressively through several levels of analysis.

"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.