Published in final edited form as:

Brain Lang. 1974 October ; 1(4): 351–362. doi:10.1016/0093-934X(74)90012-1.

Dichotic Interactions of Speech Sounds and Phonetic Feature Processing^{1,2}

David B. Pisoni and Sandra D. McNabb

Department of Psychology, Indiana University

Abstract

Identification of CV syllables was studied in a backward masking paradigm in order to examine two types of interactions observed between dichotically presented speech sounds: the feature sharing effect and the lag effect. Pairs of syllables differed in the consonant, the vowel, and their relative times of onset. Interference between the two dichotic inputs was observed primarily for pairs which contrasted on voicing. Performance on pairs that shared voicing remained excellent under all three conditions. The results suggest that the interference underlying the lag effect and the feature sharing effect for voicing occur before phonetic analysis where both auditory inputs interact.

A large body of research suggests that the perception of speech stimuli may involve specialized neural mechanisms that may not be employed in the perception of other auditory signals (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967; Stevens & House, 1972). Most investigators have assumed that the right ear advantage (REA) for dichotically presented linguistic stimuli reflects the general asymmetry of cerebral language function (Kimura, 1961, 1967; Studdert-Kennedy & Shankweiler, 1970). Studdert-Kennedy and Shankweiler have further argued that the right ear advantage for nonsense syllables reflects the operation of a specialized speech processor in the language dominant hemisphere and is not due to specialized auditory processing capacities. In other words, 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 linguistic–phonetic features in the stimulus.

Support for the notion of a unilateral phonetic processor in the language dominant hemisphere rests on a number of general findings about the relations between speech and language (see Mattingly & Liberman, 1969; Mattingly, 1972). However, most of the experimental evidence to date deals primarily with several types of interactions that have

¹This research was supported by NIMH Research Grant No. MH-24027 to the first author and PHS Biomedical Sciences Grant No. S05-RR-7031 to Indiana University. We are grateful to Dr. Franklin S. Cooper who kindly made the unique facilities of Haskins Laboratories available to us for preparation of the stimulus materials. We also thank Professor Michael Studdert-Kennedy for his interest and advice on this project and Jeff Tash for his programming assistance.

²A shorter report of some of these findings was presented at the 87th meeting of the Acoustical Society of America, New York City, April 1974.

been found between left- and right-ear dichotic speech inputs. We consider below two of these interactions—the “feature sharing effect” and the “lag effect.”

The “feature sharing effect” refers to an advantage in identification for dichotic pairs of competing consonants that share phonetic features (e.g., place or voicing). For example, Shankweiler and Studdert-Kennedy (1967) and Studdert-Kennedy and Shankweiler (1970) found that initial stop consonants in CV syllables were identified more often if the two consonant segments shared place (e.g., /ba/—/pa/) or voicing (e.g., /ba/—/da/) than if neither feature was shared (e.g., /ba/—/ta/). This interaction was interpreted as evidence that 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 was used in each syllable, auditory and phonetic information were redundant. However, Studdert-Kennedy, Shankweiler, and Pisoni (1972) examined this finding in more detail and showed that the feature sharing advantage for place was not due to shared auditory information but to shared phonetic information in the two inputs. By varying vowel context, it was possible to compare pairs of consonants that shared both auditory and phonetic information (e.g., /bi/—/pi/) to pairs that shared only phonetic information (e.g., /bi/—/pu/). Since the advantage for shared features (i.e., place) was the same under both vowel conditions in that study, the authors concluded that the feature sharing advantage must have a phonetic rather than an auditory basis. The present experiments which deal with shared voicing do not address this question directly, since auditory and phonetic dimensions were not varied independently. However, by examining the interaction between the feature sharing effect and backward masking, they do suggest a relatively straight forward account in terms of auditory interaction, at least for the feature of voicing.

The second type of dichotic interaction we consider in this paper is the so-called “lag effect.” 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. When the same syllables were mixed and the signal was presented monotonically the lag effect was reversed: The leading syllable was reported more accurately than the lagging. Studdert-Kennedy *et al.* (1970) interpreted the lag effect as a form of interruption of speech processing 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 *et al.*, 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).

However, a number of recent experiments have shown that the lag effect may not be peculiar to speech 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 differing only in fundamental frequency. With ± 25 msec offsets between stimuli, listeners reported the second stimulus more often than the first. Since the lag effect

has been obtained with speech as well as nonspeech stimuli, it seems reasonable to assume that this type of dichotic interaction may have an auditory rather than a phonetic basis. In this context, the lag effect appears to be a variation of a more general result obtained in backward recognition masking experiments: a second stimulus can impede the processing of a preceding stimulus (Kahneman, 1968; Massaro, 1972; Turvey, 1973). Specifically, the lag effect refers to the difference between dichotic forward and backward masking.

However, the locus of the interference underlying the lag effect still remains to be specified. At what stage in the information processing system does the interference arise? The available evidence dealing with this problem to date is ambiguous. For example, Darwin (1971) obtained more backward masking for CV syllables when the masking stimulus was another CV syllable than when the mask was a steady-state vowel or a nonspeech timbre. Unfortunately, Darwin used only one asynchrony value and only one CV syllable mask, thus limiting the generality of his study. In this paper we sought to determine the locus and the nature of the competition needed to produce the interference underlying the lag effect. Except for the brief study by Darwin (1971) little detailed information exists on types of competition that produce backward masking for speech stimuli (however, see Massaro, 1974).

The present experiments show that interference between dichotically presented stop consonant–vowel syllables is not found equally for all stimulus contrasts but appears to be restricted to certain pairs. These are pairs containing CV syllables that do not share the phonetic feature of voicing. We also show that the amount of interference may be manipulated by varying acoustic factors such as target–mask spectral composition and mask intensity.

The method we used to study the feature sharing and lag effects was a dichotic backward recognition masking paradigm. Two CV syllables were presented on every trial. The syllables differed in the consonant, the vowel and their relative times of onset. The S's task was to identify only the leading or target stimulus in an ear monitoring paradigm and to ignore the lagging or masking stimulus. Figure 1A shows the general arrangement of the target and masking stimuli employed in these experiments. With this technique the processing of the target stimulus may be interrupted at various stimulus onset asynchronies by the masking stimulus and thereby provide information about the temporal course of perceptual processing of the target sound. The targets and masks used in the present studies were always drawn from different stimulus ensembles as shown in Figure 1B. There were two voiced targets, /ba/ and /da/, and two voiceless targets, /pa/ and /ta/. The six masks were selected so that they either shared or contrasted with the auditory and phonetic feature composition of the targets. Half of all trials contained pairs of stimuli that shared voicing; half contained pairs that contrasted in voicing. By also varying the vowel context of the mask it was possible to hold the phonetic features constant but vary the auditory features in the target and mask. For example, the target–mask pair /ba/–/ga/ shares the voicing feature and the vowel. However, the pair /ba/–/ge/ still shares voicing but differs in the vowel and hence the auditory information is not redundant.

We are most interested in the comparison of pairs sharing voicing with those that do not share voicing. In addition, we are also concerned with the effect of the vowel in the mask and its influence on the identification of the target, since these comparisons permit us to specify the locus of interaction between the two types of dichotic inputs. For example, if the vowel of the mask had no effect on the identification of the target we would conclude that the interaction has a phonetic basis presumably occurring after the consonant features have been abstracted from the syllables. On the other hand, if the vowel of the mask affects target identification, this would indicate that the interaction occurs at an earlier stage of processing before phonetic analysis. Both of these interpretations were examined for target-mask pairs that either shared voicing (e.g., /b/–/g/) or differed on voicing (e.g., /b/–/k/).

We should note here that our concern in this paper is not with the right ear advantage in dichotic listening experiments nor with its magnitude. Rather, our interest lies in the auditory and phonetic interactions of dichotic inputs and the light these interactions may shed on the more general processes of speech perception.

EXPERIMENT 1

Method

Subjects—Twenty-four undergraduate students at Indiana University served as Ss. All Ss were right-handed native speakers of English with no history of a hearing or speech disorder.

Materials—Ten three-formant CV syllables were synthesized on the Haskins Laboratories parallel resonance synthesizer. Each syllable had a duration of 300 msec. The formant transitions lasted 40 msec whereas the steady-state vowel portion lasted 260 msec. The four target stimuli consisted of the syllables: /ba/, /da/, /pa/, and /ta/; the six masking stimuli were the syllables /ga/, /ka/, /gae/, /kae/, /ge/ and /ke/. The first formant was cut back by 70 msec, and the upper formants were aspirated over this period for the voiceless consonants. Each syllable had a pitch contour which fell linearly from 130 to 90 Hz.

Two dichotic tapes were prepared by a computer controlled procedure which permitted precise control over the temporal onsets of the stimuli. Each of the four targets was paired with each of the six masks to produce 24 paired combinations. These combinations are given in Table 1 for shared or nonshared trials in each of the three vowel contexts. The 24 pairs of syllables were presented at 10 different stimulus onset asynchrony (SOA) values: 0, 10, 20, 30, 40, 50, 60, 75, 100, 150 msec. The targets always appeared on one channel of the audio tape while the masks appeared at a given SOA value on the other channel. There was a 3 sec interval between trials and a 10 sec pause after every 30 trials.

Procedure—The experimental tapes were reproduced on a high quality tape recorder (Ampex AG 500) and were presented dichotically through Telephonics (TDH-39) matched and calibrated headphones at 75 dB SPL. Subjects were run in four counterbalanced groups of six Ss each on three separate days.

On each day *Ss* first received an 80 item diotic identification tape which consisted of 20 replications of each of the four target stimuli in random order. The *Ss* were then given each of the two dichotic masking tapes. Tape presentation and ears were balanced across groups. For the identification test, *Ss* were told to identify the consonant in each stimulus as b, d, p, or t. For the masking test, *Ss* were told to monitor a particular ear (e.g., right or left) for a block of 240 trials and to report only one of the four target stimuli on each trial.

Results and Discussion

Figure 2 shows the main results for shared and nonshared trials as a function of stimulus onset asynchrony. These functions are based on the percentage of target stimuli correctly identified in all vowel contexts. Voiced and voiceless targets are combined in this figure. Performance was consistently higher for pairs that shared voicing than for pairs that contrasted on voicing. The effect is highly significant by analysis of variance, $F(1,23) = 23.77, p < .001$. Performance is relatively stable for shared pairs at all SOA values, whereas performance improves steadily for the nonshared pairs with increases in SOA.

The effect of the mask's vowel on shared and nonshared pairs is shown separately for each of the three vowels in Fig. 3. Backward masking of the target consonant decreases systematically as spectral differences between the vowels increase. The vowel effect is not only significant, $F(2,46) = 81.32, p < .001$, but also interacts with both the feature sharing effect, $F(2,46) = 39.28, p < .001$ and the stimulus onset asynchrony effect, $F(18,414) = 9.06, p < .001$. The latter interaction suggests that nonshared pairs reach asymptotes at different rates in the three vowel contexts. Again, for each vowel context, identification in the shared condition is consistently higher than in the nonshared condition. When the data are scored separately for the voicing feature, performance on shared trials is unaffected by vowel context.

Separate analysis of voiced and voiceless targets revealed that the voiceless stimuli were correctly identified more often than voiced stimuli, $F(1,23) = 31.98, p < .001$. However, the feature sharing effect is still present under both voicing conditions although less strongly with the voiceless targets.

The main results of this experiment indicate that the feature sharing effect and the interference (i.e., backward masking) obtained in the lag effect are distinct types of interactions between dichotic speech inputs. Correct performance on shared trials appears to be affected only minimally by either SOA or vowel context of the mask. In contrast, both of these manipulations influence performance on the nonshared trials.

Together these results present somewhat of a paradox: similarity in consonant voice onset time (i.e., voicing) 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 two syllables, the more likely they are to “fuse” or integrate into one perceptual unit so that the listener has difficulty assigning the correct auditory features to the appropriate syllables. In other words, we can account for variations in the degree of backward masking for nonshared pairs by variations in “acoustic confusability” due to spectral composition of target and mask.

Before attempting an account of the apparent absence of backward masking for the shared pairs, we consider a second experiment in which another acoustic variable, intensity, is manipulated. If acoustic factors are indeed the principal determinants of variations in the degree of backward masking, we would expect intensity variations to have an effect on both shared and nonshared trials as well as the variations in spectral composition.

EXPERIMENT II

Method

Subjects—Thirty undergraduate students at Indiana University served as Ss. None of them had participated in the previous study. They met all the requirements of those used in Expt I.

Materials—The same two dichotic backward masking tapes were employed here.

Procedure—The procedure was identical to that used in the previous experiment except for the following: The channel of the tape containing the target stimuli was set at 65 dB SPL, the intensity of the channel containing the masking stimuli was then adjusted by means of separate decade attenuators (Daven; Models 2511, 2513) to be 0, 10, or 20 dB greater than the target channel. Subjects were run in six counterbalanced groups of five Ss each on three separate days. Subjects heard each dichotic tape once on a given day at one of the three intensity levels. The different intensity levels were balanced over days and groups. On a particular day, the target stimuli were presented to the right ear half the time and to the left ear the other half. The instructions were the same as those used in Expt I.

Results and Discussion

Figure 4 displays the results for shared and nonshared trials as a function of stimulus onset asynchrony for each of the three mask intensity levels averaged over all vowel contexts. Analysis of variance on correct responses revealed highly significant main effects ($p < .001$) for feature sharing, stimulus onset asynchrony, vowel context, and voicing of targets. Though not as strong for the shared as for the non-shared trials, the intensity effect is clearly present for both and is highly significant, $F(2,58) = 6.77$, $p < .005$. The interaction between intensity and feature sharing was also significant, $F(2,58) = 13.33$, $p < .001$. However, all of the remaining higher-order interactions with mask intensity failed to reach significance in this analysis.

The data were also scored separately for voicing, treating a response as correct if voicing was correct. Figure 5 shows the results for shared and nonshared trials at each intensity level. This figure is based on the data for /a/ vowel masks and is representative of the other vowel conditions. Note that most of the intensity effect for shared trials disappears. Evidently, increased mask intensity has its main effect on place, a feature cued by relatively rapid spectral changes during the first 40 msec of the syllable. In contrast, correct identification of voicing for nonshared trials decreases systematically as mask intensity is increased.

Figure 6 separates voiced from voiceless targets, scoring for voicing only, and for the condition where backward masking should be at its maximum, namely, a 20 dB mask

intensity with target and mask vowels identical. The most important aspect of this figure is the interaction between the feature-sharing effect and target voicing. Both voiced and voiceless targets are more or less immune to interference, if target and mask share voicing. If they do not, voiced targets are more severely masked by voiceless masks than the reverse. This result demonstrates a clear effect of an acoustic variable (i.e., voice onset time) on the feature-sharing effect and suggests that, at least under the conditions of this experiment, the effect is susceptible to auditory rather than phonetic manipulations.

But how are we to explain the absence of interference for shared pairs? Recall that such pairs are unaffected by variations in spectral composition of target and mask (Expt I) and, at least for the shared feature of voicing, by variations in mask intensity (Expt II). The clue to understanding is provided by an examination of feature errors. We will not present an exhaustive analysis here since this will be dealt with in another paper (McNabb & Pisoni, 1974). However, Table 2 displays the proportions of voicing and place errors under the two feature-sharing conditions for the condition yielding maximum masking. The main point to note is that while place errors are roughly the same when voicing is not shared as when it is, voicing errors are sharply increased in the nonshared condition. In other words, the feature-sharing effect is confined to the particular feature shared. Previous studies (Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy, Shankweiler & Pisoni, 1972) failed to observe this because they did not score by feature, but by total response. Obviously, if, in the nonshared condition, *S* makes a voicing error, 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 second syllable interferes with processing the voicing feature of the target.

The simplest account of the feature sharing effect is that precisely the same process occurs whether or not voicing is shared: features in the second syllable interfere with processing features in the first. But when voicing is shared the *S*'s "error" is correct: The error has the same feature value as a correct response. Since shared pairs by definition have redundant features, the probability of *S* making an error on the shared feature would be relatively small compared with the probability on nonshared features.

In short, it is not inconceivable that the sharing of phonetic features facilitates phonetic processing as suggested by the previous study of place-sharing (Studdert-Kennedy *et al.*, 1972). However, the present experiments indicate that, at least for voicing, the feature sharing effect (and the lag effect) appear to have their locus at an auditory level prior to phonetic processing. Thus, both auditory and phonetic features determine the degree of interference between dichotic inputs.

REFERENCES

- Darwin CJ. Dichotic backward masking of complex sounds. *Quarterly Journal of Experimental Psychology*. 1971; 23:386–392. [PubMed: 5148924]
- Kahneman D. Method, findings, and theory in studies of visual masking. *Psychological Bulletin*. 1968; 70:404–425. [PubMed: 4889566]
- Kimura D. Some effects of temporal lobe damage on auditory perception. *Canadian Journal of Psychology*. 1961; 15:156–165. [PubMed: 13756014]
- Kimura D. Functional asymmetry of the brain in dichotic listening. *Cortex*. 1967; 3:163–178.

- Liberman AM, Cooper FS, Shankweiler DP, Studdert-Kennedy M. Perception of the speech code. *Psychological Review*. 1967; 74:431–461. [PubMed: 4170865]
- Liberman, AM.; Mattingly, IG.; Turvey, MT. Language codes and memory codes. In: Melton, AW.; Martin, E., editors. *Coding Processes in Human Memory*. Washington, DC: V. H. Winston & Sons; 1972. p. 307-334.
- McNabb SD, Pisoni DB. Dichotic and monotic forward and backward masking of CV syllables. In preparation. 1974
- Massaro DW. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*. 1972; 79:124–145. [PubMed: 5024158]
- Massaro DW. Perceptual units in speech recognition. *Journal of Experimental Psychology*. 1974; 102:199–208. [PubMed: 4811941]
- Mattingly IG. Speech cues and sign stimuli. *American Scientist*. 1972; 60:327–337. [PubMed: 5033272]
- Mattingly, IG.; Liberman, AM. The speech code and the physiology of language. In: Leibovic, KN., editor. *Information Processing in the Nervous System*. New York: Springer-Verlag; 1969. p. 97-117.
- 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. [PubMed: 6041684]
- Stevens, KN.; House, AS. Speech perception. In: Tobias, J., editor. *Foundations of Modern Auditory Theory*. Vol. Volume II. New York: Academic Press; 1972. p. 1-62.
- Studdert-Kennedy M, Shankweiler DP. Hemispheric specialization for speech perception. *Journal of the Acoustical Society of America*. 1970; 48:579–594. [PubMed: 5470503]
- Studdert-Kennedy M, Shankweiler D, Pisoni DB. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*. 1972; 3:455–466. [PubMed: 23255833]
- Studdert-Kennedy M, Shankweiler DP, 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 MT. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*. 1973; 80:1–52. [PubMed: 4689202]

FIG. 1.

Typical arrangement of target and masking stimuli in the dichotic backward masking paradigm (Panel A) and the actual targets and masks employed (Panel B) in both experiments. Targets and masks drawn from the same row share voicing whereas targets and masks drawn from different rows contrast on voicing.

FIG. 2.

Average percent correct identification of target stimuli for shared and nonshared trials as a function of stimulus onset asynchrony in Expt I. These data are averaged over the three vowel masks.

FIG. 3.

Average percent correct identification of target stimuli for shared and nonshared trials under each vowel mask in Expt I.

FIG. 4.

Average percent correct identification of target stimuli for shared and nonshared trials at each of three mask intensities in Expt II. These functions are averaged over the three vowel masks.

FIG. 5.

Average percent correct for the voicing feature on shared and nonshared trials as a function of mask intensity in Expt II. These data are taken from the /a/ vowel mask condition.

FIG. 6.

Average percent correct for the voicing feature on shared and nonshared trials for voiced and voiceless targets in Expt II. These data are taken from the + 20 dB mask intensity condition for the /a/ vowel masks.

TABLE 1
Combinations of Four Target Consonants with Masks that Share or Contrast on Voicing and Vowel Context

Target	Shared pairs		Nonshared pairs	
	Vowel mask		Vowel mask	
	a	æ	ε	æ
Voiced	ba-ga	ba-gæ	ba-gε	ba-kæ
	da-ga	da-gæ	da-gε	da-kæ
Voiceless	pa-ka	pa-kæ	pa-kε	pa-gæ
	ta-ka	ta-kæ	ta-kε	ta-gε

TABLE 2

Proportions of Voicing and Place Errors under Voicing Shared and Nonshared Conditions for the +20 dB /a/ Vowel Masks

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