Journal of Experimental Psychology:
Human Perception and Performance
1977, Vol. 3, No. 1, 37–50

Dichotic Competition of Speech Sounds: The Role of Acoustic Stimulus Structure

Bruno H. Repp
Haskins Laboratories, New Haven, Connecticut

Dichotic consonant–vowel syllables contrasting in two features of the initial stop consonant (voicing and place) were presented for identification in a single-response paradigm without selective attention instructions. The acoustic structure of the syllables was varied within categories on both dimensions (voice onset time and formant transitions). These variations (especially those in voice onset time) had a clear influence on the pattern of responses (including blends), which ruled out a simple phonetic feature-recombination model. Rather, the auditory properties of the stimuli seemed to be preserved at the stage of dichotic interaction. A prototype model, which assumed that dichotic integration of information takes place at a stage intermediate between auditory and phonetic processing, was only moderately supported by the data. Nevertheless, some arguments are presented for maintaining this model as a working hypothesis. A new procedure for estimating the dichotic ear advantage is applied for the first time in conjunction with the single-response requirement. Most subjects showed unusually large right-ear advantages, making the present methodology interesting for the study of hemispheric asymmetry.

Many recent studies of dichotic listening have employed synthetic syllables as stimuli, most often the set /ba/, /da/, /ga/, /pa/, /ta/, and /ka/. These syllables offer a number of advantages over other materials. Being synthetic, their acoustic properties can be precisely controlled. Phonetically, they are a homogeneous stimulus set that represents all possible combinations of two values of the voicing feature (voiced and voiceless) and three values of the place feature (labial, alveolar, and velar). They also yield a reliable right-ear advantage that often tends to be larger than the right-ear advantage for other classes of competing speech sounds (Blumstein, 1974; Cutting, 1974; Haggard 1971).

Feature-Recombination Hypothesis

Detailed studies of the dichotic competition between the six stop consonants have revealed several interesting phenomena, one of which is of special interest here: When the two competing stimuli differ on both dimensions (voicing and place, for example, /ba/–/ta/), many errors are obtained that combine correct feature values from the two ears, for example, /pa/ or /da/ as responses to /ba/–/ta/. These responses have been termed *blend* errors (Halwes, 1969; Studdert-Kennedy & Shankweiler, 1970). Blend errors are responsible for another finding often called the *feature-sharing advantage,* which is more properly named *feature-contrast disadvantage.* Dichotic syllables that differ in both features receive fewer correct responses than syllables that contrast only in a single feature (Halwes, 1969; Pisoni, 1975; Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy, Shankweiler, &

Portions of the data were presented at the ninety-first meeting of the Acoustical Society of America, Washington, D.C., April 1976. This research would not have been possible without the generous hospitality of Haskins Laboratories, Alvin Liberman, President. The author was supported by National Institutes of Health Grant T22 DE00202 to the University of Connecticut Health Center. I would like to thank Terry Halwes for comments on a draft of this article.

Requests for reprints should be sent to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

Pisoni, 1972). These two phenomena (which are two sides of the same coin, since blend errors can occur only with double-feature contrasts and, therefore, lead to higher error rates for these dichotic pairs) have provided the primary support for a *feature-recombination model* of dichotic interaction. In its simplest form, this model assumes that phonetic features are (a) independently extracted from the auditory information arriving from each ear, (b) stored in a common feature buffer where information about the origin of the feature values is lost, and (c) finally recombined into percepts or responses. In other words, it is assumed that the interaction between dichotic stimuli takes place after the extraction of phonetic features and that the competing values of a particular feature have roughly equal probabilities of being selected from the feature buffer, independently of other particular features. Although this model has not always been clearly stated in the past, it was implicit in most previous research on dichotic competition (Blumstein, 1974; Cutting, 1976; Halwes, 1969; Pisoni, 1975; Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy et al., 1972).

This simple model makes several strong and easily testable predictions, some of which have been examined by Halwes (1969). If all information about the local origin of the feature values is lost, double-feature contrasts should receive an equal number of correct responses and blend errors; the two possible blend (and correct) responses should also be equally frequent. However, Halwes found correct responses to be twice as frequent as blend errors. This result could be accommodated by assuming that some of the local information is retained, so that feature values that come from the same hemisphere (ear) have a better than even chance of being selected together to form a response. However, Halwes also found wide variation in the frequencies of blend errors for different individual stimulus combinations as well as strong asymmetries in the frequencies of the two possible blend (and correct) responses for individual stimulus pairs. He suggested that unequal salience of different acoustic cues may have played a role, but he did not indicate how this idea could be incorporated in the feature-recombination model (which he did not explicitly reject).

In fact, it is possible to maintain the basic structure of the model, if the additional assumption is made that individual phonetic feature values have different strengths or saliencies, which are reflected in unequal probabilities of being selected from the phonetic feature buffer. The question remains: What determines these strengths? One possibility is that they are inherent, that is, that they have a phonetic basis. The other possibility, suggested by Halwes (1969), is that they reflect the acoustic structure of the stimuli. If the latter hypothesis were true, the simple phonetic feature-recombination model would have to be rejected, since it rests on the basic assumption that dichotic competition is exclusively phonetic in nature.

To test these hypotheses, let us consider another prediction of the model. This prediction is that acoustic stimulus variations *within* phonetic categories should not affect the frequency of blend errors and, indeed, should leave the whole response pattern unchanged. Since the phonetic features are assumed to be extracted independently before the combination of information from the two hemispheres, acoustic within-category variations can affect only the feature extraction process but not the subsequent recombination of the features. However, by definition, within-category variations do not affect the accuracy of phonetic feature extraction (if they do, they are not true within-category variations), so that their effect in dichotic competition should be nil. This null hypothesis, whose maintenance is essential to the survival of the feature-recombination model, is the focus of the present study. A rejection of the hypothesis is expected, since an alternative model is available that predicts specific effects of within-category acoustic variations.

Prototype Model

This alternative model has been proposed by Repp (1976b; Note 1). It differs from the feature-recombination model in that it considers syllables not as bundles of separately extracted phonetic features but as integral multidimensional entities whose dimensions are inseparable aspects of the whole pattern (cf. Garner, 1974; Lockhead, 1970, 1972; see also the present General Discussion section). The dimensions are assumed to reflect the auditory properties of the stimulus and thus are continuous, not binary. Instead of representing speech sounds as matrices of discrete feature values, they are conceptualized as points in a continuous multidimensional perceptual space. In the same auditory space, a limited number of fixed *prototypes* are located, which represent the listener's ideal concepts (his tacit knowledge) of the relevant phoneme or syllable categories. According to this prototype model, a stimulus is identified in the following three stages: (a) First, auditory processing leads to a mapping of the acoustic information into the multidimensional space. (b) Second, in this perceptual space, the stimulus leads to *activation* of the prototypes in its vicinity, the degree of activation being an inverse and probably nonlinear function of the (Euclidean) distance between stimulus and prototype. This results in a *multicategorical vector* whose elements are the activation values of the prototypes. (c) Finally, a probabilistic decision process selects the prototype with the largest activation value as the response (or percept).

In this model, dichotic interaction is assumed to take place at the level of multicategorical representation, in the form of a weighted averaging of the multicategorical vectors for the two stimuli. A single categorical decision then is made on the basis of this average vector. Thus, the model assumes that the competing information is combined and results in a single percept, an assumption that is justified with synthetic syllables that have the same fundamental frequency and the same vocalic context because these stimuli strongly tend to fuse in dichotic competition (Halwes, 1969; Repp, 1976b; Repp & Halwes, Note 2).

The nature of the single categorical percept is determined by two factors: (a) *ear dominance,* represented by the weights in the averaging process and (b) *stimulus dominance,* which is determined by the relative distances of the two competing stimuli from the prototypes in the perceptual space. The model predicts that stimuli that are close to a prototype will tend to dominate stimuli that are far from prototypes; this may be called the *category goodness hypothesis* of dichotic competition. Category goodness, that is, the distance from the correct prototype, is a function of auditory stimulus characteristics, so that the model predicts that stimulus dominance will vary if acoustic within-category variations of the stimuli are introduced. This was confirmed by Repp (1976b) within a restricted stimulus set, that of the voiced stop consonants. By varying the initial formant transitions, the dominance relations between the stimuli from this place continuum could be reliably influenced; the pattern of the data conformed at least qualitatively to the prototype model.

The present experiment investigates the generality of these earlier findings. To be useful, the prototype model should explain the response pattern for all dichotic combinations of the six stop consonants, as well as the effects of variations in cues other than the initial formant transitions. Consider first how the model explains blend responses: Two stimuli, for example, /ba/ and /ta/, will not only activate their correct prototypes (B and T, respectively) but also, to a lesser degree, the blend prototypes (D and P, respectively), that are neighbors in perceptual space. Because of the presumed additivity of prototype activation levels, the blend protoypes may reach activation levels comparable to those of the correct prototypes, to which only one of the two stimuli makes a substantial contribution.

In principle, this model allows for variations in the frequencies of blends between individual stimulus pairs, since they will depend in a complex way on the arrangement of prototypes and stimuli in the perceptual space. A mathematical formulation of the model should be able to predict their pattern. In the present context, however, we will be content with qualitative predictions concerning *changes* in the response pattern, leaving quantitative tests to a future study.

Contrary to the feature-recombination

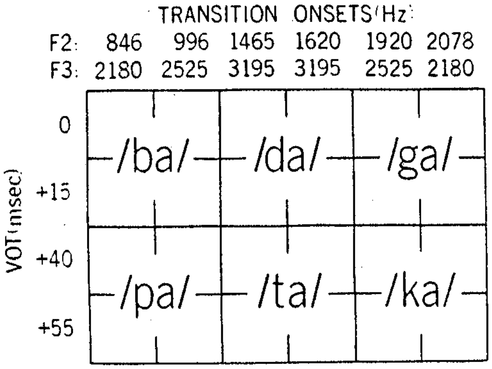

Figure 1. Acoustic stimulus parameters. (The steady-state frequencies for /a/ were 1,232 Hz [F2] and 2,525 Hz [F3]. VOT = voice onset time; F2 = second formant; F3 = third formant.)

model, the prototype model predicts variations in the response pattern with changes in the acoustic structure of the stimuli. Consider again the previous example, the stimulus pair /ba/–/ta/. Assume that we delay the voice onset time (VOT, the important acoustic cue for the voicing feature) of /ba/, so that the stimulus is still identified as B in isolation; however, in the perceptual space, it is farther removed from the B prototype and closer to the P prototype. It will now be closer to the boundary between voiced and voiceless sounds, and it will contribute less activation to B and D and more to P and T than the original /ba/. As a result, the frequencies of P and T responses to the dichotic pair /ba/–/ta/ should increase; those of B and D responses should decrease. Similar predictions may be made for changes of the VOT of /ba/ in the other direction, in the other stimulus (/ta/), or in the formant transitions (the acoustic cue for place of articulation) of either stimulus. A number of other, more detailed, predictions may be derived from the model, some of which will be considered in the Results and Discussion section.

The phonetic feature-recombination model and the prototype model are not the only possible conceptions of the process of dichotic interaction; however, most other plausible models are compromises between these two extremes (see the General Discussion section). The detailed formulation of such models seems less important than the empirical demonstration of within-category effects in dichotic competition, which would rule out a whole class of such models.

In addition to the primary focus on stimulus dominance in dichotic competition, the factor of ear dominance is given attention in the present study. A new method of calculating ear advantage indices, especially designed for the single-response paradigm (Repp, 1976a, 1976b; Repp & Halwes, Note 2), is applied here for the first time. (The present experiment constitutes part of an ongoing series of studies aimed at developing optimal procedures for assessing lateral asymmetries in dichotic listening.)

Method

Subjects

The subjects were eight paid volunteers, four women and four men, mostly Yale University students. All had normal hearing, except one man who claimed to have a slight (5 dB) hearing loss in the right ear. Two subjects were left-handed, one of them only in writing. All were relatively inexperienced listeners.

Stimuli

The stimulus set comprised 24 utterances synthesized on the Haskins Laboratories parallel resonance synthesizer. They sounded like /ba/, /da/, /ga/, /pa/, /ta/, and /ka/. There were four acoustically different versions of each syllable, resulting from all combinations of four different VOTs with six second- and third-formant transitions (see Figure 1). All syllables were 300 msec long, had no initial bursts, the same transition durations (50 msec),[1] and the same constant fundamental frequency (90 Hz).

The experimental tape was recorded using the Pulse Code Modulation system at Haskins Laboratories. The tape contained a series of 120 single syllables consisting of 5 different random sequences of the 24 stimuli. It was followed by 2 blocks of dichotic pairs. Each block contained 192 pairs, representing all possible double-feature contrast combinations of the 24 stimuli: 6 phoneme combinations (/ba/–/ta/, /ba/–/ka/, /da/–/pa/, /da/–/ka/, /ga/–/pa/, /ga/–/ta/), with 2 channel/ear assignments for each, and 16 different acoustic

---

[1] It was discovered after the experiment that the first-formant transitions of the labial consonants were only 40 msec long. However, this was almost certainly of no consequence.

combinations within each phonemic contrast. Their sequence was completely random, with interstimulus intervals of 3 sec. The onsets of the syllables in a dichotic pair were exactly simultaneous (.125 msec maximal error).

Procedure

The subjects were tested in small groups in a single session lasting about 2 hr. The single-channel series was presented monaurally for identification, followed by 2 dichotic blocks. After a break, the tape recorder channels were reversed electronically; the 2 dichotic blocks were presented again, followed by the monaural syllables. All in all, each subject listened to 10 replications of each monaural stimulus and to 4 replications of each dichotic pair (8 replications, if channel/ear assignment is ignored). The tape was played back from an Ampex AG-500 tape recorder through an amplifier/attenuator to Telephonics TDH-39 headphones. The intensities of the two channels were carefully equalized at about 65 dB SPL (peak deflections on a voltmeter).

Table 1
Confusion Matrix of the 24 Stimuli (Monaural Identification)

Stimulus	Response					
VOT and F2	B	D	G	P	T	K
/ba/						
0–low	80	—	—	—	—	—
0–high	80	—	—	—	—	—
+15–low	76	—	—	4	—	—
+15–high	79	—	—	1	—	—
/da/						
0–low	—	28	52	—	—	—
0–high	—	34	46	—	—	—
+15–low	—	27	53	—	—	—
+15–high	—	39	40	—	—	1
/ga/						
0–low	—	5	75	—	—	—
0–high	—	—	80	—	—	—
+15–low	—	16	64	—	—	—
+15–high	—	—	79	—	—	1
/pa/						
+40–low	—	—	—	80	—	—
+40–high	1	—	—	78	—	1
+55–low	—	—	—	80	—	—
+55–high	—	—	—	80	—	—
/ta/						
+40–low	—	—	—	1	22	57
+40–high	—	1	—	1	34	44
+55–low	—	—	—	3	30	47
+55–high	—	—	—	—	61	19
/ka/						
+40–low	—	—	4	—	12	64
+40–high	—	1	2	—	4	73
+55–low	—	—	—	2	6	72
+55–high	—	—	—	2	1	77

Note. VOT = voice onset time; F2 = second formant (starting frequency).

As part of the instructions, the subjects were told about the two features, voicing and place, and the precise stimulus combinations to expect; a diagram on the answer sheet elucidated the explanation. However, they were not informed about the within-category variations until after the experiment. The subjects were asked to write down what the fused stimuli sounded most like; they were instructed to give only a *single* response for each dichotic pair. Naturally, the responses were restricted to the six stop consonants, with the additional admonition to try to give both voiced and voiceless responses.[2]

Results and Discussion

Monaural Intelligibility

As is often the case with synthetic syllables, their intelligibility in the experiment turned out to be somewhat poorer than anticipated. The confusion matrix for all eight subjects is shown in Table 1. It can be seen that the problem lay almost exclusively with /da/ and /ta/, which were more often heard as /ga/ and /ka/, respectively. The absence of a burst, which is especially important in

---

[2] It was thought that some subjects might give predominantly voiceless responses, which would have reduced the information in the data. This suspicion, derived from pilot observations, was apparently unfounded. For the same reason, four subjects (two old and two new) were (re)tested with the same tape with detection instructions. These instructions restricted the response set to either the voiced consonants (B, D, and G) or the voiceless consonants (P, T, and K) only, counterbalanced across blocks within subjects. Since the subjects knew that each dichotic pair contained one voiced and one voiceless consonant, this amounted to a detection task. The main purpose of the detection instructions was to force the subjects to give an equal number of voiced and voiceless responses to each pair; consequently, only the effects of variations in formant transitions could be assessed. These effects agreed with those under standard instructions.

Table 2
Dichotic Stimulus–Response Matrix

	Percentage of responses							
Stimulus	B	D	G	P	T	K	Correct total	Blends total
/ba/–/ta/	11.4	**3.6** +	**3.4**	**56.7**	14.2 +	10.6	36.2	63.8
/ba/–/ka/	13.5	**3.5** +	**6.8**	**56.2**	3.2 +	16.6	33.5	66.5
/da/–/pa/	**4.8**	24.6 +	20.0	23.6	**12.2** +	**14.7**	68.2	31.8
/da/–/ka/	1.2 +	16.5	**37.9**	1.7 +	**8.2**	34.6	(52.3)	(47.7)
/ga/–/pa/	**7.1**	8.1 +	38.4	24.6	**4.3** +	**17.5**	71.1	28.9
/ga/–/ta/	**.8** +	**10.9**	37.5	1.3 +	17.3	**32.2**	(56.1)	(43.9)

Note. Bold type represents blends; plus signs connect responses that were pooled in the analysis; parentheses indicate that the exact proportions of correct responses and blends were uncertain.

alveolar consonants, may have been a factor here. The confusability of these stimuli was not detrimental to the purpose of the experiment, although it had to be taken into account in the analysis of the dichotic data.

Confusions along the voicing dimension were extremely rare and occurred exclusively at the VOTS closer to the boundary. A similar pattern may be seen for /ga/ and /ka/ with respect to place confusions: Alveolar responses were more frequent when the velar transitions were closer to the boundary (low). However, for /da/ and /ta/, the opposite was the case: Velar responses were more frequent when the transitions were farther away from the alveolar–velar boundary (low). This curious reversal has been confirmed in other studies using similar stimuli (Repp, Note 3), but its explanation is far from clear.

Dichotic Response Pattern

The dichotic response pattern for the six phonemic contrasts, disregarding within-category variations, is shown in Table 2. The percentages in bold type represent blends; their total frequencies are given in the last column. It can be seen that blend responses were extremely common but varied in frequency as a function of the stimuli involved: In the two pairs containing /ba/, they comprised almost two thirds of all responses; in the two pairs containing /pa/, they comprised only about one third of all responses; in the remaining two pairs, they comprised somewhat less than half of all responses. In these last two pairs (alveolar–velar contrasts), the exact proportion of blends was uncertain, as indicated by the parentheses in Table 2. Because of the listeners' uncertainties about the place of articulation of the component stimuli, blend responses could have arisen from either blending or from confusion and, likewise, "correct" responses may have included some true blends.

The poor discrimination between alveolar and velar place is also reflected in the responses to the other pairs containing one labial consonant. Since the labials were highly intelligible (see Table 1), alveolar and velar responses were simply grouped together in these dichotic pairs, as indicated by the plus signs in Table 2. (For example, G responses to /ba/–/ta/ were considered blends, whereas K responses were considered correct.) In alveolar–velar pairs, the few labial responses that occurred (probably random errors) were combined with the alveolar responses. These groupings were maintained in all further data analyses.

Table 2 shows enormous variation in the pattern of blend responses. In the two pairs containing /ba/, P responses predominated and were more than twice as frequent as P responses to pairs actually containing /pa/. In terms of the prototype model, this indicates that /ba/ was far from the B prototype on the voicing dimension but close to it on the place dimension; that is, it was weak on the former but strong on the latter; hence, the joint predominance of labial and voiceless responses. This suggests that the response pattern perhaps could be explained in terms of separate and independent competition on the two features, voicing and place, although this would contradict the prototype model. However, in the two pairs containing /pa/, for example, correct responses were much more frequent than predicted by this hypothesis; in pairs containing /ba/, they were less frequent than predicted. Note that the hypothesis of feature independence predicts that responses in the different place categories should be proportional within voicing categories. However, the stimulus pair /ga/–/pa/, for example, received five times as many G responses as B responses but actually fewer K than P responses, which contradicts the hypothesis of feature independence in dichotic competition.

In principle, these results are compatible with the prototype model, although it is not yet clear whether a more rigorous quantitative formulation of the model would be able to explain the detailed response pattern. The feature-recombination model, on the other hand, cannot explain the variations in the proportions of blend responses for different stimulus pairs or the asymmetries in blend responses to individual pairs observed earlier by Halwes (1969).

Effect of Within-Category Variations in Voice Onset Time

The effects of within-category variations in VOT are shown in Table 3. The data are shown as the percentages of voiced and voiceless responses and of correct responses and blends to the four VOT combinations, averaged over the different phonemic contrasts and the variations in formant transitions.

Table 3
Percentages of Voiced and Voiceless Correct Responses and Blends as a Function of Voice Onset Time (VOT) Combinations

	Correct		Blends		Total	
VOT	+40	+55	+40	+55	+40	+55
			Voiced responses			
0	42.8	15.1	19.1	5.7	61.9	20.8
+15	35.1	20.6	18.5	10.4	53.6	31.0
			Voiceless responses			
0	16.3	33.1	21.8	46.1	38.1	79.2
+15	19.5	28.5	26.9	40.5	46.4	69.0
			Total			
0	59.1	48.2	40.9	51.8	100.0	100.0
+15	54.6	49.1	45.4	50.9	100.0	100.0

Obviously, the variations in VOT had a strong effect on the response pattern. The most striking effect was produced by a change in the VOT of the voiceless stimulus. Voiceless stimuli with the shorter VOT (+40) led to a slight predominance of voiced responses, whereas those with the longer VOT (+55) brought about a predominance of voiceless responses. This is in agreement with the prototype model, since there is good reason to assume that a voiceless stimulus with a VOT of +55 will be closer to its prototype than a stimulus with a VOT of +40. On the other hand, the effect of a change in the VOT of voiced stimuli was less striking and showed an interaction with the VOT of the voiceless competitor. When the VOT of the latter was +40, the effect of a VOT change from 0 to +15 in the voiced stimulus was as predicted, that is, it led to a relative decrease in the percentage of voiced responses. However, when the VOT of the voiceless stimulus was +55, the effect of the same change in the VOT of the voiced stimulus had just the opposite effect. This interaction was unexpected and is difficult to explain.

This pattern of results was highly consistent between individual phoneme combinations and individual subjects. Analysis of variance of the percentages of voiced (voiceless) responses yielded a highly significant effect of the VOT of the voiceless stimulus, $F(1,\ 7) = 59.14$, $p < .0002$, and a significant interaction between the VOT of the voiced stimulus and the VOT of the voiceless stimulus, $F(1,\ 7) = 24.63$, $p < .002$. The main effect of the VOT of the voiced stimulus was not significant.

Table 3 also shows that the proportion of correct responses and blends varied as a function of VOT. Correct responses were more frequent where voiced responses were more frequent, whereas blends tended to ac-

Table 4
Percentages of Correct Responses and Blends Sharing Place Feature Values with Voiced (V+) or Voiceless (V−) Stimuli as a Function of Transition Combinations

F2 transitions[a]	Correct		Blends		Total	
	V+ close	V+ far	V+ close	V+ far	V+ close	V+ far
Responses sharing place with V+ stimulus						
V− close	30.9	25.1	29.1	33.0	61.0	58.1
V− far	32.0	26.2	36.0	36.8	68.0	63.0
Responses sharing place with V− stimulus						
V− close	23.4	28.3	15.6	13.6	39.0	41.9
V− far	20.1	25.7	11.9	11.3	32.0	37.0
Total						
V− close	54.3	52.6	45.7	47.4	100.0	100.0
V− far	52.1	51.9	47.9	48.1	100.0	100.0

[a] F2 = second formant.

company voiceless responses. Note that the majority of all voiced responses were correct; whereas, among the voiceless responses, blends were more frequent than correct responses. This indicates that the place feature of voiceless stimuli was weak in competition against the place feature of voiced stimuli. In terms of the prototype model, it suggests that noise-excited formant transitions are a less effective cue to place of articulation than voiced transitions. This is plausible in view of the fact that the present stimuli did not contain any bursts, a second important cue to place of articulation that certainly is more important in voiceless plosives.

Effect of Within-Category Variations in Formant Transitions

The effects of within-category variations in formant transitions are shown in Table 4. The percentages of correct responses and blends are subdivided into those that share the place feature value with the voiced stimulus and those that share it with the voiceless stimulus. The dimensions of each 2 × 2 subtable are the transitions of the voiced stimulus and the voiceless stimulus. The transitions were classified according to whether they were close to or far from the category boundary separating the place feature values of the two competing stimuli. Thus, *close* refers to the higher F2 (second formant) transitions for labials and for alveolars paired with velars but to the lower F2 transitions for velars and for alveolars paired with labials.

It is evident that the effect of variations in the formant transitions was much smaller than that of VOT, but it was in the direction predicted by the prototype model. Responses sharing place with the voiced stimulus were most frequent when the transitions of the voiced stimulus were far and those of the voiceless stimulus were close; they were least frequent when the opposite was the case.[3] This pattern was shown primarily by the correct responses; the blends followed a somewhat different pattern, tending to be least frequent when both stimuli were close and most frequent when both were far.

Analysis of variance of the responses sharing place with the voiced (voiceless) stimulus yielded a highly significant effect of the transitions of the voiced stimulus, $F(1, 7) = 27.17$, $p < .002$, but only a marginally significant effect of the transitions of the voiceless stimulus, $F(1, 7) = 4.79$, $p < .07$, with no significant interaction between the two. Thus, the former was more reliable than the latter, which again indicates that the transitions of voiceless stimuli were weak in their perceptual effect.

There were some consistent deviations from the pattern in Table 4, which are responsible in part for the relatively small average effect. Labial–velar pairs, especially /pa/–/ga/, received more labial responses when the velar transitions were far than when they were close. Pairs containing alveolar consonants, on the other hand, conformed to the predictions, despite the inverted pattern of place confusions in monaural presentation (see Table 1).

---

[3] It may be argued that the within-category effect of the transitions reflected merely changes in the confusion probabilities of alveolar and velar stimuli (cf. Table 1). However, the dichotic effects were only slightly reduced after a correction was applied that took changes in confusion structure into account. Moreover, the transitions of labial consonants (which were rarely confused, as can be seen from Table 1) had a very pronounced effect.

Within-Category Feature Interactions

It has been pointed out above that the response pattern in Table 2 cannot be explained by independent competition on the two phonetic dimensions (phonetic feature independence). Questions of feature independence may also be asked *within* phonemic combinations (auditory feature independence): Did within-category variations in VOT affect competion on the place dimension? Did within-category variations in the formant transitions influence competition on the voicing dimension?

The answer to the first question is negative: Responses sharing place with the voiced (voiceless) stimulus did not vary significantly as a function of VOT. However, a more detailed analysis showed that the VOT of the voiceless stimulus did have a significant influence in some individual stimulus combinations. The largest of these effects was in /ba/–/ka/ and consisted in a decrease in labial responses and an increase in velar responses as the VOT of /ka/ changed from +40 to +55. This effect is in agreement with the prototype model which predicts a certain amount of positive correlation between features: As a stimulus moves closer to its prototype along one dimension, its overall Euclidean distance from the prototype is reduced, and other dimensions will indirectly benefit from this increase in category goodness.

On the other hand, voiced (voiceless) responses showed a significant effect of the transitions of the voiceless stimulus, $F(1, 7) = 22.61$, $p < .003$. Voiced responses were more frequent when the voiceless transitions were closer to the place boundary, which is again in agreement with the prototype model. The (nonsignificant) effect of the transitions of the voiced stimulus, however, was not in the predicted direction. It was also surprising that the voiceless transitions affected compettiton on the voicing feature more than competition on the place feature.

The prototype model also predicted variations in the proportion of blend errors (and correct responses) as a function of joint variation in both stimulus dimensions. Correct responses were expected to be most frequent (and blend responses least frequent) when the two competing stimuli were farthest apart in perceptual space, that is, when they were closest to their respective correct prototypes. The opposite result was predicted when the two stimuli were closest in perceptual space, and thus almost as close to the blend prototypes as to the correct prototypes. This hypothesis was most easily tested by considering only the acoustically most similar and the acoustically most dissimilar pair within each phonemic contrast. (For example, in /ba/–/ta/, the most similar pair would be /ba/ with high F2 transitions and a VOT of +15, paired with /ta/ with low F2 transitions and a VOT of +40; the most dissimilar pair would be /ba/ with low F2 transitions and a VOT of 0, paired with /ta/with high F2 transitions and a VOT of +55.) Of the six phonemic contrasts, only one supported the prediction, whereas four showed differences in the opposite direction. Overall, blends were more frequent when the competing stimuli were acoustically *dissimilar*. This is in contradiction to the prototype model. However, the result is in agreement with and, indeed, a consequence of the earlier observations that (a) variations in the formant transitions had a relatively small effect and that (b) blends tended to accompany voiceless responses that increased greatly in frequency as VOT changed from +40 to +55.

Ear Dominance[4]

The present experiment offered a first opportunity to apply an improved method for calculating an unbiased index of ear dominance recently proposed by Repp (1976a, 1976b). This new index takes into account variations in stimulus dominance by applying the methods of signal detection theory and fitting a linear receiver operating characteristic function to the data points for individual stimulus pairs. The index is a linear transformation of the area under the re-

---

[4] The terms *ear dominance* and *ear advantage* are used interchangeably here, although the former is more appropriate within the single-response paradigm.

Table 5
Individual Ear Advantages

Subject	Voicing	Place	Place without alveolar–velar contrasts
1	+.17	+.09[a]	+.10[a]
2	+.73	+.52	+.64
3	+.89	+.57	+.76
4	+.57	+.82	+.89
5	−.09	+.35	+.35
6	+.90	+.76	+.78
7	+.47	+.14	+.26
8	+.75	+.81	+.98
Average	+.55	+.51	+.60
Author	+.96	+.55	+.64

Note. Subject 1 claimed a 5-dB hearing loss in the right ear. Subject 8 was left-handed; Subject 6 was left-handed in writing only. Data for the author are based on three sessions.

[a] Coefficient is not significant. All other coefficients are significant at $p < .05$ or better. Unbiased coefficients and significance estimates are based on the methods described in Repp and Halwes, Note 2.

ceiver operating characteristic function (cf. Green & Swets, 1966), and it ranges from +1 for a perfect right-ear advantage to −1 for a perfect left-ear advantage. Its derivation and its advantages over other indices are discussed in Repp and Halwes, Note 2.

The calculation of the unbiased ear advantage index presupposes that the responses can be grouped into two exhaustive categories. Double-feature contrasts present a problem here, because of the large proportion of blend errors that are ambiguous with respect to ear dominance. At present, it is not clear how a valid index could be derived from the responses at the phonemic level. However, the problem can be circumvented by separately considering the two features, voicing and place. Ear dominance indices for voicing only are easily calculated by classifying the responses as voiced or voiceless, ignoring the place feature. These indices (and the corresponding receiver operating characteristic function) were based on 24 data points, representing the four VOT combinations for each of the six phonemic contrasts, ignoring variation in the transitions. The results are shown in the second column of Table 5.

Similar indices were calculated for the place dimension by dichotomizing the responses, using the same grouping of place categories as in the earlier data analysis (see Table 2). Each index was based on 24 data points, representing the four transition combinations for each of the six phonemic contrasts, ignoring variations in VOT. These indices are shown in the third column of Table 5. The fourth column of Table 5 shows place feature indices that omit the eight data points representing alveolar–velar contrasts.

Table 5 shows that there was a highly significant average right-ear advantage. Except for one subject on the voicing dimension, all subjects showed right-ear advantages. The most striking result is the magnitude of these effects. The average right-ear advantages as well as most of the individual coefficients are several magnitudes larger than the ear advantages reported in earlier studies of normal subjects. (In fact, several subjects showed right-ear advantages close to the possible maximum.) There are two possible reasons why these indices are so large. One is that conventional indices, such as the phi coefficient (Kuhn, 1973; Repp, 1976b), may underestimate the "true" size of the ear advantage. For example, the average phi coefficient on the voicing dimension was +.30, which is only about half the size of the unbiased index of +.55. Still, however, this phi coefficient is very large compared to those in earlier studies, which required the subjects to give two responses (e.g., Shankweiler & Studdert-Kennedy, 1975, who reported an average phi of +.06). The reason for this difference may be that the single-response paradigm adopted here eliminates much of the noise that is present in two-response data and, therefore, reveals the true magnitude of the ear advantage. There is much to be said in favor of this argument (see Repp & Halwes, Note 2). However, Repp (1976b) reported an average phi coefficient of only +.06 in a single-response experiment with completely fused syllables that contrasted in place only. Clearly, there must be an additional factor beyond the response requirements and the kind of index used. Although previous studies have not indicated a substantial difference in the right-ear advantage for completely fused syllables (as in Repp, 1976b) and partially fused syllables (as in the present study), the present results suggest strongly that such a difference exists; it perhaps was obscured by guessing in earlier studies requiring two responses.[5]

A comparison between the third and fourth columns in Table 5 shows that, for all subjects but one, exclusion of alveolar–velar pairs led to an increase in the ear dominance coefficient on the place dimension. This finding illustrates an important methodological point: Pairs of stimuli that are highly confusable will tend to show a reduced ear advantage. It follows that high intelligibility of the stimuli in a dichotic test is an important requirement and that pairs of confusable stimuli should be omitted from consideration when the ear advantage is determined.

Finally, the indices for voicing and place (Columns 2 and 4 in Table 5) may be compared. While the average indices are similar, there are substantial individual differences. Some of these may be due to chance, but the larger differences (and especially that for the author whose results are based on 2,304 responses) are certainly real. It must be concluded that, for a given individual, the right-ear advantage on the voicing dimension is not necessarily the same as on the place dimension. Underlying these differences may be individual differences in the perceptual representation of the speech sounds and their dimensions (i.e., in the structure of the subjective perceptual space). This points to a substantial problem in measuring the "true" or physiological ear advantage, which we are only now beginning to understand. Future research will have to deal with the possibility of interactions between hemispheric dominance and perceptual organization in individuals.

General Discussion

The present study demonstrates clear effects of within-category acoustic variations on dichotic stimulus dominance relations. This finding constitutes conclusive evidence against a simple phonetic feature-recombination model, as outlined in the introduction. It also renders insufficient a more elaborate version of this model incorporating the concept of inherent phonetic feature strength. Rather, the competitive strengths of phonetic feature values are probably a direct function of the acoustic stimulus structure, and changes in the latter lead to changes in the former. Thus, dichotic interaction does not take place at a strictly phonetic level but at an earlier stage where auditory information is still preserved in some form.

The prototype model provides one possible conception of this auditory representation. According to this model, the dichotic inputs converge in the form of multicategorical vectors, a stage intermediate between continuous auditory and discrete phonetic representation. The multicategorical stage embodies the relation between the variable auditory input and the more or less fixed phonetic categories. This stage has proven useful in conceptualizing the process of dichotic interaction and fusion (Repp, 1976b, Note 1), which so far has been considered only in terms of the auditory–phonetic dichotomy (Cutting, 1976; Pisoni, 1975; Studdert-Kennedy, 1976; Studdert-Kennedy et al., 1972). However, the prototype model was only moderately supported by the present data. Below, I will briefly summarize some of its shortcomings, consider some alternative models, and present some theoretical arguments in favor of maintaining the prototype model as a working hypothesis.

On the whole, the main prediction of the prototype model was confirmed. A dichotic stimulus tends to gain in competitive strength if its acoustic structure is changed, so that it moves closer to its presumed correct prototype and away from category boundaries. However, there were two major exceptions: (a) the inverted effect of a change in VOT from 0 to +15 when the com-

---

[5] It may be noted that none of the four subjects who received detection instructions (see Footnote 2) showed a large right-ear advantage on the place dimension, and one subject showed a marked reduction in right-ear advantage compared to the standard condition. The coefficients for these subjects were +.05, +.20, +.08, and +.12 (alveolar-velar pairs included).

peting stimulus had a VOT of +55 (see Table 3) and (b) the inverted effect of a change in the transitions of velars when paired with labials (mentioned in connection with Table 4). Both effects are very difficult to rationalize, but there is no doubt about their reality.

A follow-up study of dichotic competition along the VOT dimension (Repp, Note 4) has revealed even more bizarre interactions. Note that they cannot be explained by atypical stimulus characteristics (such as synthesis artifacts) or by different assumptions about the location of the prototypes in perceptual space. For example, its has been implicity assumed that a VOT of 0 is closer to the voiced prototype than a VOT of +15 and that a VOT of +55 is closer to the voiceless prototype than a VOT of +40. However, if the obvious hypothesis is introduced that the prototypes represent the modal *production* values of the corresponding articulatory dimensions, the first part of the assumption is probably false: A VOT of +15 is closer to the modal production value than a VOT of 0, at least for alveolars and velars (Klatt, 1973; Lisker & Abramson, 1964; Zlatin, 1974). However, even if this were true—and the data permit this interpretation as well as the opposite—it could not explain the interaction obtained; all that would change is which part of the interaction is considered anomalous. (Note also that the VOT interaction was exhibited by all six phonemic combinations and thus was apparently independent of place of articulation.)

There is little value in discussing the several other respects in which the prototype model has failed. Instead, it seems useful to consider alternative models that perhaps could account for the anomalous findings. Unfortunately, however, the most obvious candidates make rather similar predictions and do not fare better than the prototype model.

It is possible, for example, to consider a pure *auditory averaging model*. This model would assume that the dichotic stimuli are integrated at a strictly auditory level of processing, so that a single stimulus (a kind of auditory average of the two components) is phonetically interpreted. In the present context, this model makes predictions that are quite similar to those of the prototype model; however, in other contexts, differential predictions can be generated and the auditory averaging model has been found insufficient (Cutting, 1976; Repp, 1976b, Note 1). It is quite possible, however, that some auditory interaction is involved in addition to integration at a higher, multicategorical (and, perhaps, even phonetic) level. Such a multilevel model of dichotic interaction would be of considerable complexity, but it is not clear whether it could explain the anomalies in the present data.

Another alternative model that deserves some discussion is the *feature-detector model* which currently enjoys some popularity (Cooper, 1974; Cooper & Nager, 1975; Eimas & Corbit, 1973; Miller, 1975, 1976; Studdert-Kennedy, 1976). This model assumes a separate set of detectors for each feature, with one detector corresponding to each value of a feature (Cooper, 1974; Eimas & Corbit, 1973; Miller, 1975). Effectively, this places the prototypes at the level of auditory analysis. Dichotic interaction may be conceptualized as follows: Each stimulus passes through separate banks of feature detectors and emerges as an array of multicategorical feature codes (i.e., as a multicategorical *matrix*). These matrices then converge upon a single processor where they are averaged. Subsequently, separate feature decision mechanisms select the largest detector response for each feature, and finally, these categorical feature values are combined into a percept or response. Thus, each feature or dimension has its own little perceptual space and its own set of prototypes.

The predictions of the feature-detector model are again rather similar to those of the prototype model, except that, in its simplest form, the former assumes mutual independence of individual features. There are several instances in the present data where this assumption must be rejected, so that rather complex ad hoc assumptions about the interrelations among feature detectors and among feature decisions would have to be introduced. The prototype model, on the other hand, predicts specific interdependencies between different features; some of them were supported by the data but others were not. The data, therefore, do not permit a choice between these alternative models. However, given that they are equally well (or equally poorly) supported, there are some theoretical reasons why the prototype model might be preferred as a working hypothesis.

The voicing and place features of stop consonants are among the best examples of integral dimensions (Lockhead, 1972; Garner, 1974). One cannot exist without the other, and selective attention to one feature is impossible without taking the other feature into account. In fact, there is strong evidence that the whole consonant–vowel syllable is an integral unit of processing (Wood & Day, 1975; Pisoni & Tash, Note 5). Integral units are multidimensional, and their dimensions interact during processing. The feature-detector model can deal with such interactions only by some rather strenuous assumptions that, typically, are made post hoc and often are based on assumptions of serial processing, which are inappropriate with integral dimensions (Garner, 1974). The prototype model, by virtue of its multidimensional Euclidean structure, naturally incorporates such interactions, and it makes predictions that can be quantified and falsified. Moreover, it is somewhat counterintuitive and uneconomical to assume a separate categorical decision for each feature, subconscious as these decisions may be. A single phonetic decision is more in line with subjective experience and certainly more parsimonious.

Lockhead has discussed similar problems with respect to visual stimuli. Lockhead's (1972) views are worth quoting here, since they apply to speech stimuli as well:

> A distinctive feature must be a set of attributes considered in relation to all stimuli; one cannot have distinctive features in a vacuum. . . . We must determine the space, the set of relations, and not just the features, if we are to understand pattern recognition. The basic hypothesis is that observers first locate an object in some complex psychological space and then analyze that locus according to the needs of the task. . . . Perhaps a distinctive feature can be defined as an attribute(s), or the value of an attribute(s), of a stimulus which causes that integral object to be distant from other potential stimuli in the psychological space. . . . [This] directs attention to the possibility that the relations between attributes (which is another way of saying locus in space) may be processed *before* the values of the attributes themselves are processed. (pp. 417–418)

The prototype model is very much in line with Lockhead's views. It adds the assumption of category prototypes, a concept that has been useful in other research on perception and pattern recognition (e.g., Hyman & Frost, 1975; Posner, 1969; Reed, 1972; Rosch, 1973; Smith, Shoben, & Rips, 1974) but has been largely neglected in models of speech perception (however, see Galunov & Chistovich, 1966). Thus, the prototype model has considerable heuristic value and much more evidence will have to be collected before it can be confidently rejected. The achievement of the present study lies primarily in the rejection of the overly simple phonetic feature-recombination model; its contribution to the evaluation of the prototype model remains modest.

An additional result of the present study is the magnitude of the ear advantages obtained. It suggests that the single-response paradigm, together with the unbiased ear dominance index (Repp, 1976a, 1976b; Repp & Halwes, Note 2), is a powerful method for assessing laterality effects, and it is probably one step closer to an optimal dichotic test for diagnostic purposes.

Reference Notes

1. Repp, B. H. Discrimination of dichotic fusions. *Haskins Laboratories Status Report on Speech Research*, 1976, *SR-45/46*, 123–139.
2. Repp, B. H., & Halwes, T. G. *Measuring laterality effects in dichotic listening.* Unpublished manuscript, 1976. (Available from the authors at Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.)
3. Repp, B. H. *The voicing boundary as a function of variations along the place continuum.* Manuscript in preparation, 1977.
4. Repp, B. H. *Dichotic competition of voice onset times.* Manuscript in preparaton, 1977.
5. Pisoni, D. B., & Tash, J. "Same-different" reaction times to consonants, vowels and syllables. In *Research on speech perception* (Progress Rep. No. 1). Bloomington: Indiana University, Department of Psychology, 1974, 129–139.

References

Blumstein, S. E. The use and theoretical implications of the dichotic technique for investigating distinctive features. *Brain and Language*, 1974, *1*, 337–350.

Cooper, W. E. Adaptation of phonetic feature analyzers for place of articulation. *Journal of the Acoustical Society of America*, 1974, *56*, 617–627.

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

Cutting, J. E. Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, 1974, *16*, 601–612.

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

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

Galunov, V. I., & Chistovich, L. A. Relationship of motor theory to the general problem of speech recognition (review). *Soviet Physics-Acoustics*, 1966, *11*, 357–365.

Garner, W. R. *The processing of information and structure.* Potomac, Md.: Erlbaum, 1974.

Green, D. M., & Swets, J. A. *Signal detection theory and psychophysics.* New York: Wiley, 1966.

Haggard, M. P. Encoding and the REA for speech signals. *Quarterly Journal of Experimental Psychology*, 1971, *23*, 34–45.

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

Hyman, R., & Frost, N. H. Gradients and schema in pattern recognition. In P. M. A. Rabbitt and S. Dornic (Eds.), *Attention and performance V.* London: Academic Press, 1975.

Klatt, D. H. Voice-onset time, frication and aspiration in word-initial consonant clusters. *Quarterly Progress Report of the Massachusetts Institute of Technology Research Laboratory of Electronics*, 1973, No. 109, 124–136.

Kuhn, G. M. The phi coefficient as an index of ear differences in dichotic listening. *Cortex*, 1973, *9*, 447–457.

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

Lockhead, G. R. Identification and the form of multidimensional discrimination space. *Journal of Experimental Psychology*, 1970, *85*, 1–10.

Lockhead, G. R. Processing dimensional stimuli: A note. *Psychological Review*, 1972, *79*, 410–419.

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

Miller, J. L. Properties of feature detectors for VOT. *Journal of the Acoustical Society of America*, 1976, *59*, S41. (Supplement)

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* (Vol. I). Hillsdale, N.J.: Erlbaum, 1975.

Posner, M. I. Abstraction and the process of recognition. In G. H. Bower & J. T. Spence (Eds.), *Psychology of learning and motivation* (Vol. 3). New York: Academic Press, 1969.

Reed, S. K. Pattern recognition and categorization. *Cognitive Psychology*, 1972, *3*, 382–407.

Repp, B. H. Acoustic determinants of "stimulus dominance" in dichotic competition. *Journal of the Acoustical Society of America*, 1976, *59*, S6. (Supplement) (a)

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

Rosch, E. H. On the internal structure of perceptual and semantic categories. In T. M. Moore (Ed.), *Cognitive development and the acquisition of language.* New York: Academic Press, 1973.

Shankweiler, D., & Studdert-Kennedy, M. A continuum of lateralization for speech perception? *Brain and Language*, 1975, *2*, 212–225.

Smith, E. E., Shoben, E. J., & Rips, L. J. Structure and process in semantic memory: A featural model for semantic decisions. *Psychological Review*, 1974, *81*, 214–241.

Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics.* New York: Academic Press, 1976.

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

Wood, C. C., & Day, R. S. Failure of selective attention to phonetic segments in consonant-vowel syllables. *Perception & Psychophysics*, 1975, *17*, 346–350.

Zlatin, M. A. Voicing contrasts: Perceptual and productive voice onset time characteristics of adults. *Journal of the Acoustical Society of America*, 1974, *56*, 981–994.

Received April 28, 1976 ■