Identification of Dichotic Fusions*

Bruno H. Repp

ABSTRACT

Seven synthetic syllables from a "place continuum" (/bæ - dæ - gæ/) were presented in all dichotic combinations for identification. These syllables fused completely, so that dichotic pairs were perceived as single stimuli. The response pattern could not be easily explained by an "auditory averaging" hypothesis. Rather, stimuli that were good instances of a category seemed to "dominate" stimuli that were closer to a category boundary. To account for this finding, a three-stage pattern recognition ("prototype") model is proposed according to which the information from the two ears is integrated after auditory but before phonetic-categorical processing, at a "multicategorical" stage. Electronically mixed stimuli led to a similar response pattern, suggesting that competing transitional cues remain intact up to the multicategorical stage. It is demonstrated that these fusions cannot be reliably discriminated from binaural stimuli, and that selective attention to one ear has little effect. For the purpose of assessing ear advantages, dichotic fusions offer methodological advantages over other dichotic stimuli. The problem of determining the "true" ear advantage is discussed.

INTRODUCTION

In recent years, dichotic listening has received much attention, both as a research tool for the investigation of the processes involved in speech perception and as a diagnostic technique for assessing hemispheric dominance for

*A substantially revised version of this paper is to be published in the Journal of the Acoustical Society of America. Authors who wish to refer to this research are urged to consult the revised version.

†Also University of Connecticut Health Center, Farmington.

Acknowledgment: This research was conducted at Haskins Laboratories and would not have been possible without the extraordinary hospitality of this institution and its director, Alvin Liberman. I thank him, Michael Studdert-Kennedy, James Cutting, Terry Halwes, Gary Kuhn, and David Paul for comments and discussions related to this paper. The author was supported by NIH Grant T22 DE00202 to the University of Connecticut Health Center.

[HASKINS LABORATORIES: Status Report on Speech Research SR-45/46 (1976)]
speech.\(^1\) Both aspects are addressed by this paper, which, on the basis of a
detailed analysis of the dichotic interaction between the voiced stop conso-
nants, makes recommendations for a possible methodological refinement of dichotic
testing.

Dichotic tests composed of synthetic stop-consonant-vowel syllables have
become widely accepted as the most precise instruments currently available for
assessing ear advantages in speech perception (Shankweiler and Studdert-Kennedy,
1967a, 1975). The control of stimulus characteristics and channel synchroniza-
tion made possible by modern speech synthesizers and specialized computer sys-
tems, together with the balanced stimulus set of the six stop consonants, gives
these tests a distinct advantage over other materials and procedures. Neverthe-
less, some problems remain. One is the kind and number of responses to be re-
quired from the listeners: two responses (with or without restrictions on their
order) or one response (with or without selective-attention instructions)?
Variants of both response modes have been used at one time or another, but two-
response paradigms have dominated the scene. However, because of the occurrence
of confusions, intrusions, and guessing, and the lack of a good theory taking
these phenomena into account, the two responses cannot be unequivocally assigned
to the stimuli that evoked them, so that errors and correct responses are not
clearly separated in scoring the results (cf. Repp, 1975a, 1976). Selective-
attention instructions offer no remedy, since selective attention is very diffi-
cult with precisely aligned dichotic syllables, and intrusions from the unattend-
ed channel are common (Halwes, 1969; Haggard, 1975; Repp, 1975a).

Another problem has been the derivation of an index for the ear advantage.
Simple percentage differences have the disadvantage that they depend on the
overall performance level and therefore do not adequately represent the degree
of an ear advantage but merely measure its direction. The proposal of Kuhn
(1973) to use the \(\phi\) coefficient as a measure of the ear advantage has been an
important step forward. However, Kuhn's index is designed for two-response
paradigms (or single-response paradigms with selective-attention instructions)
and therefore does not solve the problem of unraveling correct responses and
errors.

Halwes (1969) and Studdert-Kennedy and Shankweiler (1970) have pointed out
the low information content of the second of two responses. This observation
suggests that it may be more appropriate to ask for a single response only. In
fact, it seems that listeners often perceive only a single syllable when a
dichotic pair is presented. This tendency is more pronounced with syllables
contrasting in only a single distinctive feature\(^2\) (voicing, for example,

---

\(^1\)See, for example, *Brain and Language*, 1974, Vol. 1, No. 4 and 1975, Vol. 2,
No. 2.

\(^2\)A comment on terminology is in order here. Many authors refer to "shared fea-
tures" rather than "feature contrasts," for example, /ba/ and /pa/ "share
place" (Studdert-Kennedy and Shankweiler, 1970; Pisoni and McNabb, 1974). This
terminology is awkward, for several reasons: (1) Any characterization in terms
of shared features is indeterminate unless all shared features are enumerated
(which includes many irrelevant features), whereas mentioning the contrasting
features is informative even without precise knowledge of the complete stimulus
set. (2) Features are dimensions and therefore are always shared, precisely
than with syllables contrasting in both features (for example, /ba+ta/): in a "same-different" judgment task, the former receive more incorrect "same" responses than the latter. Moreover, within the single-feature contrasts, place contrasts are much harder to discriminate from identical (binaural) syllables than voicing contrasts (Halwes, 1969; Blumstein and Cooper, 1972; Repp, 1976). In other words, precisely aligned simultaneous dichotic syllables that differ only in the direction of their initial formant transitions strongly tend to fuse and sound like a single syllable originating in the middle of the head (if their intensities are equal).

Cutting (1972, 1976) has proposed a classification of dichotic fusions that includes "psychoacoustic fusions": when /ba+ga/ is presented, /da/ is often heard. We will follow Cutting and use the term "psychoacoustic fusion" only for this specific phenomenon. However, it should be clear that fusion in the more general sense—hearing only a single stimulus when two are presented—occurs independently of the nature of the phonetic percept. Thus, /ba+ga/ sounds just as fused when /ba/ or /ga/ is heard as when /da/ is heard, and /ba+da/ fuses just as well, although it will never give rise to a "new"response.

These considerations suggest that it is useless to require a listener to give two responses when a dichotic place contrast is presented. A single response will contain virtually all the information available to the listener. (However, it may be usefully supplemented by a measure of response uncertainty, such as confidence ratings, reaction times, or response distributions.) The principal question is then: How is the information from the two ears combined into a single percept? Cutting (1972, 1976) has suggested that psychoacoustic fusion is a relatively low-level auditory averaging phenomenon. Any such explanation should apply to all dichotic place contrasts. The present experiments attempt to investigate this question further by examining the identification of dichotic fusions in some detail.

From a methodological standpoint, it is important to determine whether dichotic fusions lead to the right-ear advantage (REA) commonly found in dichotic listening. Several studies have indicated that place contrasts show a somewhat speaking. It is their values that may differ, and this seems to be somewhat better captured in the term "feature contrast" (that is, a contrast with respect to a feature) than in "shared feature." (3) Most importantly, feature sharing has often been interpreted as a factor facilitating dichotic perception. However, there is no known factor in dichotic listening that facilitates perception relative to monaural or binaural presentation; rather, performance is impaired by competition as a consequence of feature contrasts. Therefore, the latter term will be used here exclusively.

3 The notation i+j will be used to indicate a dichotic stimulus pair regardless of channel/ear assignment of the component stimuli, while i−j and j−i will designate the two specific channel assignments (i and j stand for stimulus numbers; see Table 1).

4 Conversely, it may also be argued that, within the set of the six stop consonants at least, there is characteristically only one perceptual result, regardless of whether phenomenological fusion occurs.
smaller REA than other feature contrasts (Shankweiler and Studdert-Kennedy, 1967a; 1967b; Studdert-Kennedy and Shankweiler, 1970). Since the place contrasts in these studies may not have been perfectly fused, the difference may in fact be larger. This is interesting with regard to the question at which level(s) in processing the REA arises. If it were the case that dichotic place contrasts fuse at a very early stage in processing and then are transmitted in this form to each hemisphere, there should be no REA, since the REA is usually attributed to transcallosal transmission loss of left-ear information, assuming functional independence of the dichotic inputs prior to their convergence upon the dominant hemisphere (Studdert-Kennedy, 1975). On the other hand, fusion may either occur at a higher level (after central convergence) or be an entirely autonomous phenomenon mediated by an independent low-level cross-correlation mechanism, so that fused syllables are processed in basically the same way as less completely fused syllables; in this case, there should be no difference in REAs between the two.

EXPERIMENT I

The first experiment examined the identification of fused dichotic stimuli from a "place continuum" (Pisoni, 1971) obtained by systematically varying the starting frequencies of the initial formant transitions. The principal questions were whether identification responses could be predicted by a simple auditory averaging model, whether a significant REA of "normal" magnitude exists, and whether psychoacoustic fusions are as common as suggested by Cutting (1972). The effects of variations in the acoustic properties and relationships of the fused stimuli were of prime concern with respect to all three questions.

Method

Subjects. Thirteen paid volunteers participated, seven males and six females, all right-handed, unaware of any hearing trouble, and relatively inexperienced listeners. The data of two additional subjects were eliminated because they were too noisy.

Stimuli. The stimuli were seven syllables ranging perceptually from /bae/ to /dæ/ to /ga/. They were produced on the Haskins Laboratories parallel resonance synthesizer. All syllables were of 280-msec duration, had a constant fundamental frequency (114 Hz), a voice onset time of -15 msec (that is, prevoicing), 45-msec linear transitions, and no bursts but an abrupt onset of energy following the prevoicing. The syllables differed only in the onset frequencies of the second-formant (F2) and third-formant (F3) transitions, which are shown in Table 1.

---

5The recent paper of Cutting (1976) was not available at the time of the experiment.
TABLE 1: Starting frequencies (in Hz) of second-formant (F$_2$) and third-formant (F$_3$) transitions of the seven stimuli.

<table>
<thead>
<tr>
<th>Stimulus Number</th>
<th>F$_2$</th>
<th>F$_3$</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1312</td>
<td>2348</td>
</tr>
<tr>
<td>2</td>
<td>1456</td>
<td>2694</td>
</tr>
<tr>
<td>3</td>
<td>1620</td>
<td>3026</td>
</tr>
<tr>
<td>4</td>
<td>1772</td>
<td>3026</td>
</tr>
<tr>
<td>5</td>
<td>1920</td>
<td>2694</td>
</tr>
<tr>
<td>6</td>
<td>2078</td>
<td>2348</td>
</tr>
<tr>
<td>7</td>
<td>2234</td>
<td>2018</td>
</tr>
<tr>
<td>Steady-state /æ/</td>
<td>1620</td>
<td>2862</td>
</tr>
</tbody>
</table>

(binaural) pairs and 42 nonidentical (dichotic) pairs. Five different random sequences of the 84 stimuli were recorded. The interstimulus interval was 3 sec.

Procedure. The subjects were tested individually or in small groups in a single session lasting approximately 90 minutes. Playback was from an Ampex AG-500 tape recorder through an amplifier to Grason-Stadler TDH-39 earphones. Playback intensity was adjusted and monitored on a Hewlett-Packard voltmeter, and special care was taken to equalize the intensities of the two channels at about 85 dB SPL (peak deflections).

Each subject listened twice to the five blocks of 84 stimuli. The channels were reversed electronically after the first five blocks. The instructions were to write down one response for each syllable heard: B, D, or G, whatever the syllable sounded most like.

The subjects were generally not informed until after the experiment that different inputs were presented to the two ears in half of the stimuli. (There were some exceptions, because some subjects had previously participated in related experiments with dichotic fusions.) Most subjects agreed when questioned that they heard only single syllables and showed surprise when told about their actual nature. This, together with the experimenter’s impression, was considered sufficient evidence for the adequate fusion of the stimuli. (Formal tests were conducted later in Experiment III with different subjects.)

Results and Discussion

The response pattern. The pooled results of the 13 subjects are shown in Figure 1. The numbers in the graphs represent identical (binaural) pairs, and the dashed lines connecting them trace the categorical identification functions for the seven stimuli. It can be seen that stimuli 1 and 2 were generally identified as B; 3 and 4, as D; and 6 and 7, as G. Stimulus 5 was the only truly ambiguous syllable, with somewhat more D than G responses. (The stimulus numbers refer to Table 1.) Some subjects produced noisy data, which is reflected in the averages; for example, G responses to stimuli 6 and 7 reached only 85-86 percent.
Figure 1: Percentages of B responses (upper left-hand panel), G responses (upper right-hand panel), and D responses (lower panels) to binaural (numbers) and dichotic (filled symbols) pairs. Each function connects the combinations of a constant stimulus (represented by the number at one of the endpoints of the function) with the stimuli along the abscissa. The dashed lines connect the binaural pairs and trace the single-stimulus labeling functions.
Consider now the other symbols in Figure 1 that represent the dichotic combinations of different stimuli. Each function connects the pairs formed by one particular stimulus (denoted by the number at one end of the function) and the stimuli along the abscissa. The pattern may be described as follows:

(1) When a particular stimulus was paired with other stimuli, the percentage of responses in the relevant category tended to decrease as the competing stimuli were further and further removed on the continuum. This was especially clear for D responses, while the functions for B and G responses became flat and even nonmonotonic when /ba/ and /ga/ stimuli were paired with stimuli more than two or three steps removed on the continuum. Note that B responses were at a minimum in pairs with stimulus 4, while G responses tended to be at a minimum in pairs with stimulus 3.

(2) The percentages of responses in the three categories generally remained in proportion to the binaural identification results for the component syllables of a dichotic pair; for example, the B-function for stimulus 2 (upper left-hand panel in Figure 1) lies uniformly lower than that for stimulus 1, and that for stimulus 3 is even lower. More interesting, however, is the fact that a similar difference exists between the G-functions for stimuli 6 and 7 (upper right-hand panel in Figure 1), although these two stimuli showed identical binaural identification scores. In addition, there is one crossover of functions: the D-function for stimulus 3 lies above that for stimulus 4 in pairs with stimuli 5, 6, and 7 (lower left-hand panel in Figure 1).

(3) There was a tendency for /ba/ stimuli (especially 1) to dominate /ga/ stimuli (6 and 7). A change in acoustic structure at the /ba/ end of the continuum had a greater effect than an equivalent change at the /ga/ end, as indicated by the wider spacing of the B functions (cf. upper panels in Figure 1).

(4) Psychoacoustic fusions were clearly present but rather infrequent, especially in pairs containing stimulus 1. The numerical results for the four relevant stimulus pairs are shown in Table 2.

<table>
<thead>
<tr>
<th>Stimulus pair</th>
<th>Responses</th>
</tr>
</thead>
<tbody>
<tr>
<td>B</td>
<td>D</td>
</tr>
<tr>
<td>2+6</td>
<td>38.5</td>
</tr>
<tr>
<td>2+7</td>
<td>45.0</td>
</tr>
<tr>
<td>1+6</td>
<td>67.3</td>
</tr>
<tr>
<td>1+7</td>
<td>60.8</td>
</tr>
</tbody>
</table>

Psychoacoustic fusions: Three averaging hypotheses. The pattern of results just described (particularly under paragraphs 2 and 4) definitely rules out a "phonetic averaging" (attention-switching or rivalry) hypothesis. If, for example, the two stimuli competed for a single phonetic processor, so that one syllable gained access to the processor in a certain percentage of the trials while the other syllable was lost, the distribution of identification responses
for a dichotic pair would be a weighted average of the response distributions for the two component stimuli in isolation. The same would be true if both syllables were categorized independently in separate processors, and an attentional mechanism with limited capacity selected one or the other outcome on a probabilistic basis. Instead, the existence of psychoacoustic fusions and of effects of acoustic within-category differences is evidence that the dichotic information interacts prior to the completion of phonetic processing.

A second hypothesis may be termed "articulatory averaging" (Cutting, 1976). It is similar to the phonetic averaging hypothesis, except that it allows for psychoacoustic fusions by perceptual-articulatory interpolation at the feature level. However, it excludes any interaction between the acoustic properties of the stimuli and therefore is clearly disconfirmed both by the present data and by Cutting's own.

On the other hand, the data are superficially in accord with an "auditory averaging" hypothesis, which assumes that the formant transitions of the two competing stimuli (or rather, their equivalent auditory codes in the brain) fuse to yield new, intermediate transitions, and the resulting new information is then phonetically interpreted. This hypothesis has also been considered by Cutting (1976), who independently investigated the effect of acoustic stimulus variations on the frequency of psychoacoustic fusions. However, one prediction would then be that /bæ + ɡæ/ stimulus pairs, such as 1+7 and 2+6, which have about the same "average," should yield the same percentage of D responses. Instead, the acoustically more similar pair, 2+6, led to more psychoacoustic fusions than the acoustically more dissimilar pair, 1+7 (cf. Tables 2 and 3, upper left-hand quadrant), which parallels the results of Cutting (1976). Therefore, Cutting's conclusion that simple averaging of formant transitions is an insufficient explanation also applies to the present data.⁶

Another problem with the auditory averaging model is its deterministic nature. There is no /bæ + ɡæ/ stimulus pair for which only D responses are obtained. In fact, the frequency of psychoacoustic fusions in the present experiment was surprisingly low. Nine of the thirteen subjects showed negligible frequencies (less than 7 percent, after a correction for expected confusions). One reason for this may have been the presence of F₂ transitions, which were rising for /bæ/ and /ɡæ/ stimuli but falling in /dæ/ stimuli. In /bæ + ɡæ/ pairs, the "average" F₂ transition may have been in conflict with the "average" F₃ transition, so that the responses tended to shift among all three alternatives. The classical studies of Harris, Hoffman, Liberman, Delattre, and Cooper (1958) and Hoffman (1958) have shown (incidentally, also in the context /æ/ ) that F₃ transitions have a strong influence on the tendency to give D responses, with F₂ transitions held constant: rising transitions decrease and falling transitions increase D responses. Cutting (1976) used two-formant syllables and obtained higher percentages of psychoacoustic fusions than the present study;

⁶ Of course, the assumption of a linear (unweighted) auditory averaging process is naive and probably wrong. However, the conclusion that acoustic similarity plays a role seems nevertheless justified. The present results differ from those of Cutting (1976) with respect to the relative weight of low-frequency and high-frequency transitions. Here, low-frequency changes had a greater effect, while Cutting's data (for /bæ + ɡæ/) show precisely the opposite.
however, he also encouraged D responses by presenting only /ba+ga/ pairs to uninformed subjects who were given three response alternatives.

In order to check further on the role of F3 transitions, a new stimulus tape was prepared that contained all dichotic and binaural pairs of seven syllables identical with those of Experiment I, except that they had no third formant. BHR, who had also participated in five sessions of Experiment I, listened to 30 random blocks of 49 stimulus pairs each, in three sessions. The results closely resembled his results with three-formant syllables, except for two of the four /ba+ga/ combinations. These results are shown in Table 3 (upper portion). The pooled response distribution for the four two-formant /ba+ga/ pairs differed significantly from that for the corresponding three-formant pairs ($\chi^2(2) = 7.6, p < .05$) but, clearly, the difference was due only to 2+6 and 2+7, which showed greatly increased frequencies of D responses. (Note that BHR generally gave an unusually high percentage of psychoacoustic fusion responses.)

<table>
<thead>
<tr>
<th>Stimulus pair</th>
<th>Responses</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Three formants</td>
<td></td>
<td>Two formants</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>D</td>
<td>G</td>
<td>B</td>
<td>D</td>
<td>G</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dichotic</td>
<td>2+6</td>
<td>28.0</td>
<td>47.0</td>
<td>25.0</td>
<td>26.7</td>
<td>70.0</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+7</td>
<td>36.0</td>
<td>38.0</td>
<td>26.0</td>
<td>23.3</td>
<td>53.4</td>
<td>23.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+6</td>
<td>71.0</td>
<td>26.0</td>
<td>3.0</td>
<td>73.3</td>
<td>23.4</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+7</td>
<td>63.0</td>
<td>30.0</td>
<td>7.0</td>
<td>71.7</td>
<td>25.0</td>
<td>3.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Mixed</td>
<td>2+6</td>
<td>55.0</td>
<td>25.0</td>
<td>20.0</td>
<td>36.7</td>
<td>58.3</td>
<td>5.0</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>2+7</td>
<td>96.3</td>
<td>2.5</td>
<td>1.2</td>
<td>86.7</td>
<td>11.7</td>
<td>1.7</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+6</td>
<td>46.3</td>
<td>38.8</td>
<td>15.0</td>
<td>30.0</td>
<td>56.7</td>
<td>13.3</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>1+7</td>
<td>77.5</td>
<td>10.0</td>
<td>12.5</td>
<td>80.0</td>
<td>6.7</td>
<td>13.3</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

It may be concluded that, in two pairs at least, the conflict between F2 and F3 transitions probably played a role. However, even in the absence of a third formant, psychoacoustic fusions were far from the 100 percent predicted by a simple auditory averaging hypothesis. If this hypothesis is to be maintained, considerable random variability in the weighting function of the averaging process must be assumed. This assumption will be tested in Experiment II.

**Ear dominance and stimulus dominance.** In order to correct for perceptual confusions between the stimulus categories (especially those provided by an ambiguous stimulus), left-ear and right-ear scores were derived for each stimulus pair. This was done by weighting each response by the relative frequencies of this particular response category for the two component stimuli in isolation and by subsequent summation of these weights for each ear. Expressed formally,
the right-ear score for a given dichotic pair i-j (with i in the right ear and j in the left ear) was computed as

\[
T_{\text{RE}(i)} = \frac{\sum_{k=1}^{3} f(R_k | i-j) f(R_k | i)}{f(R_k | i) + f(R_k | j)}
\]  

(1)

where \( f(R_k | i-j) \) is the frequency of response category \( R_k \) for the dichotic pair, \( f(R_k | i) \) and \( f(R_k | j) \) are the frequencies of response \( R_k \) to \( i \) and \( j \), respectively, when presented in isolation, and the summation is over the three response categories. For the left ear, \( T_{\text{LE}(j)} = N - T_{\text{RE}(i)} \), where \( N \) is the total number of responses to this stimulus pair. The weight (the fraction) in Eq. (1) was set equal to 0.5 whenever the combined responses to \( i \) and \( j \) in a particular category constituted less than 10 percent. The resulting scores are free from overt variations in performance level, since the scores for the two ears always sum up to \( N \), that is, there are no errors by definition. Because of the weighting procedure, individual variations in accuracy (which do exist) play only a negligible role as long as the "noise" does not exceed a certain level.

The two scores for a given dichotic pair, \( T_{\text{RE}(i)} \) and \( T_{\text{LE}(j)} \), have counterparts in the two scores for the other channel assignment of the same stimulus combination, \( T_{\text{RE}(j)} \) and \( T_{\text{LE}(i)} \). These four scores were arranged in two different two-way contingency tables, and two \( \phi \) coefficients were calculated: the stimulus dominance index

\[
\phi_D = \left( T_{\text{RE}(i)} - T_{\text{RE}(j)} \right)^2 / \left( T_{\text{RE}} T_{\text{LE}} \right)^{1/2} \quad \text{with} \quad T_{\text{RE}} = T_{\text{RE}(i)} + T_{\text{RE}(j)}
\]

\[
T_{\text{LE}} = T_{\text{LE}(i)} + T_{\text{LE}(j)}
\]  

(2)

which indicates the degree to which stimulus \( i \) "dominates" stimulus \( j \); and the ear dominance (or ear advantage) index

\[
\phi_E = \left( T_{\text{RE}(i)} - T_{\text{LE}(i)} \right)^2 / \left( T_{(i)} T_{(j)} \right)^{1/2} \quad \text{with} \quad T_{(i)} = T_{\text{RE}(i)} + T_{\text{LE}(i)}
\]

and \( T_{(j)} = T_{\text{RE}(j)} + T_{\text{LE}(j)} \)  

(3)

which describes the relative dominance of the right ear over the left ear. Overall indices were obtained by calculating \( \phi \) coefficients from summed response frequencies, with separate summations for \( i-j \) and \( j-i \) pairs (arbitrarily assuming that \( i<j \) on the stimulus continuum).\(^7\) The significance of these indices was tested by \( \chi^2 \) (cf. Kuhn, 1973).

\(^7\)The denominator in the formula for the \( \phi \) coefficient is the geometric mean of the two unequal marginal sums in the contingency table (the other two marginals being equal to \( N/2 \)). Unless the difference between these marginals is very large, their geometric mean is similar to their arithmetic mean, which equals \( N/2 \). \( \phi_D \) [Eq. (2)] is therefore usually well approximated by \( 2(T_{\text{RE}(i)} - T_{\text{RE}(j)}) / N \), and \( \phi_E \) [Eq. (3)] is usually almost identical to \( 2(T_{\text{RE}(i)} - T_{\text{LR}(i)}) / N \), except in cases of extreme stimulus dominance. If the entries in the contingency table are expressed as percentages (that is, divided by \( N/2 \)), \( \phi_D \) and \( \phi_E \) can be estimated at glance. This relationship also justifies the calculation of an overall index from summed response frequencies, which usually deviates only very slightly from the average of the coefficients for individual stimulus pairs.
The crucial question was whether the REA obtained from Eq. (3) would be comparable to the REA found in a two-response paradigm with a larger stimulus ensemble. The results are shown in the left third of Table 4. The 13 subjects exhibited a significant average REA, with six significant individual REAs but only one significant left-ear advantage. These results were compared with those of a recent study that used the complete set of six stop consonants and reported the distribution of Kuhn's (1973) $\phi$ coefficient for 22 subjects (Shankweiler and Studdert-Kennedy, 1975). The two distributions were virtually identical (Mann-Whitney test: $Z = 0.03$). To the degree that the two-ear-advantage indices are indeed equivalent, and within the limits imposed by the small sample sizes, this comparison indicates that dichotic fusions show just the same degree of an average REA as less completely fused syllables (which make up the majority of the combinations of all six stop consonants), so that phenomenological fusion is probably unrelated to the degree of REA obtained. The smaller REAs reported for place contrasts in the past were most likely artifacts of the two-response requirement and of the ear-advantage indices used.

![Table 4: Dichotic ear dominance indices, and dichotic and mixed stimulus dominance indices for individual subjects.](image)

Table 4 also shows a highly significant REA for BHR. Interestingly, however, his REA with two-formant stimuli was much smaller and did not reach significance. This finding, which suggested that auditory stimulus complexity may influence the REA, was followed up in Experiment IV.

105
Actually, the ear advantages were slightly underestimated because one-step contrasts, which were mostly within categories (e.g., 1+2 and 3+4), were included. Of the 21 individual stimulus pairs, 20 showed a positive average \( \phi \). There was a tendency toward larger REAs with increasing separation of the component stimuli on the continuum: the average \( \phi \) increased from +0.04 (two-step pairs) to +0.08 (three-step pairs) to +0.11 (four-, five- and six-step pairs), despite the occurrence of uninformative psychoacoustic fusions at the largest separations. Hence, acoustic stimulus disparity may play a role in determining the magnitude of the REA, a question of considerable theoretical importance that deserves further study.

Table 4 (center) also shows the average stimulus dominance (\( \phi_D \)) indices for the individual subjects. These indices express the average dominance of \( i \) over \( j \), summed over all \( i<j \); or, in other words, the degree of perceptual dominance of lower-frequency \( F_2 \) transitions over higher-frequency \( F_2 \) transitions (assuming that competition between \( F_3 \) transitions plays only a minor role). This average index is rather crude, but it captures some striking individual differences. The overall \( \phi_D \) was positive and highly significant, indicating strong dominance of lower-frequency transitions. However, 2 of the 13 subjects had highly significant negative coefficients.

The \( \phi_D \) indices for the individual stimulus pairs, which were of primary interest, were by no means homogeneous, as was already evident from Figure 1. Only a few pairs were in perceptual equilibrium (\( \phi_D = 0 \)), and stimulus dominance effects were considerably stronger than ear dominance effects. The stimulus dominance pattern for a subgroup of 7 of the 13 subjects is illustrated in Figure 2 (filled triangles). This subgroup was selected for reasons of comparison with the results of Experiment II; their data are representative of all 13 subjects, except that the average \( \phi_D \) was somewhat reduced. Discussion of the dominance pattern will be reserved for the General Discussion section following the description of Experiment II.

**EXPERIMENT II**

The relatively low percentages of psychoacoustic fusions in Experiment I may have been due to random variability in ear dominance or stimulus dominance from trial to trial. Psychoacoustic fusions may occur only when the two syllables receive very nearly equal weights in the hypothetical auditory averaging process; a slight tip of the balance in favor of one stimulus may lead to perceptual dominance of that stimulus. However, when the two syllables in a pair are acoustically combined before they reach the ear, the potential factor of variability in ear dominance is excluded. In addition, auditory averaging may occur at a more peripheral stage and may reduce any variability arising at more central levels. Therefore, this hypothesis predicted an increase in psychoacoustic fusions for mixed stimuli.

A comparison of dichotic and mixed pairs promised to be interesting with respect to the whole "dominance pattern" of individual stimulus combinations. The peripheral interactions coming into play in the mixed mode (acoustic interference, auditory masking) may well lead to an entirely different response pattern than in the dichotic mode. On the other hand, any significant similarities between the two situations will have to be ascribed to common central processing levels.
Method

Subjects. Nine of the thirteen subjects in Experiment I participated, one of them prior to Experiment I. The data of one additional subject were eliminated because they were too noisy.

Materials. The same stimulus tape as in Experiment I was used.

Procedure. The procedure was identical with that of Experiment I except that the output of the two tape recorder channels was mixed electronically and presented binaurally. The intensity was readjusted to about 85 dB SPL. Special care was exercised in equating the intensities of the two channels before they entered the mixer. There was no reversal of channels here.

Results and Discussion

Controls. A comparison of the response distributions for pairs of identical syllables in the dichotic and mixed conditions revealed significant differences for six of the seven syllables. However, the changes consisted primarily in a reduction of the "noise" and an increase in response consistency, so that familiarity and practice were the most likely cause. In view of these changes in the "baseline" scores, it was especially important to compare the response patterns in the two conditions by means of a measure that takes these changes into account. This was achieved by weighting the data as in Experiment I [cf. Eq. (1)], with "channels" replacing "ears." Subsequently, \( \phi_D \) and \( \phi_C \) ("channel dominance") coefficients were calculated [cf. Eqs. (2) and (3)].

While, at the levels used here, intensity differences of a few decibels have little effect in dichotic listening (Speaks and Bissonette, 1975), the mixing procedure was likely to be sensitive to small channel imbalances. The \( \phi_C \) coefficients served as a check on the proper equalization of the two channels prior to mixing. Two subjects indeed showed highly significant \( \phi_C \) coefficients, both in the same single session. This indicated a calibration error, and the data were excluded from further consideration.

Psychoacoustic fusions. Table 5 compares the responses to the four /bæ + ɡæ/ pairs in the dichotic and mixed conditions for the same seven subjects. Surprisingly, D responses were clearly less frequent in the mixed condition than in the dichotic condition, with B responses making up for most of the difference. This was probably not a practice effect, since BHR—who again participated in five sessions—showed precisely the same decline in psychoacoustic fusions (Table 3, left portion), and a correction for expected D confusions did not eliminate the difference. It may be noted that the data of Halwes (1969) showed a similar reduction in psychoacoustic fusions for mixed syllables.

Since it was conceivable that again the presence of a third formant somehow played a role, BHR once more served as a control subject and listened to mixed two-formant syllables (30 blocks in 3 sessions). The results showed an increase in psychoacoustic fusions with respect to mixed three-formant syllables but a reduction with respect to dichotic two-formant syllables (Table 3). This shows that the reduction was not due to a change in the salience of the third formant.

The stimulus dominance pattern. The overall \( \phi_D \) coefficient was again significant and in favor of the stimuli with the lower numbers on the continuum
TABLE 5: Response percentages for four dichotic /bæ+/gæ/ pairs: within-subject comparison of dichotic and mixed conditions (seven subjects).

<table>
<thead>
<tr>
<th>Stimulus pair</th>
<th>Dichotic</th>
<th></th>
<th></th>
<th>Mixed</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>2+6</td>
<td>32.1</td>
<td>32.1</td>
<td>35.8</td>
<td>42.1</td>
<td>19.3</td>
<td>38.6</td>
</tr>
<tr>
<td>2+7</td>
<td>38.6</td>
<td>32.1</td>
<td>29.3</td>
<td>46.4</td>
<td>17.9</td>
<td>35.7</td>
</tr>
<tr>
<td>1+6</td>
<td>55.7</td>
<td>16.4</td>
<td>27.9</td>
<td>76.4</td>
<td>2.9</td>
<td>20.7</td>
</tr>
<tr>
<td>1+7</td>
<td>50.7</td>
<td>15.0</td>
<td>34.3</td>
<td>59.3</td>
<td>5.7</td>
<td>35.0</td>
</tr>
</tbody>
</table>

(lower-frequency F2 dominance) but slightly reduced in comparison to the dichotic condition (Table 4). Again, there were large individual differences, also from one condition to the other (cf. Table 4).

The stimulus dominance indices for the individual stimulus pairs in the two conditions are compared in Figure 2. The \( \phi_p \) values in Figure 2 represent the dominance of the stimulus held constant in each panel over the stimuli on the abscissa. (Each individual stimulus combination, \( i+j \), may be found twice in Figure 2, once in the panel for \( i \) with \( j \) on the abscissa, and once in the panel for \( j \) with \( i \) on the abscissa, with a \( \phi_p \) coefficient of opposite sign. Of course, \( \phi_p = 0 \) for identical pairs.) It is evident that, with few exceptions, the functions for the mixed condition exhibit the same basic peaks and valleys as those for the dichotic condition. There are some consistent differences as well, primarily in pairs containing stimuli 1 and 2: in the mixed condition, these /bæ/ stimuli showed increased dominance over /gæ/ stimuli (5, 6, 7) but reduced dominance over /dæ/ stimulus (3, 4). The dominance relationship between /dæ/ and /gæ/ stimuli did not change very much.

BHR's data were in excellent agreement with those of the seven subjects. The dominance pattern of BHR's two-formant results was virtually identical to that of his three-formant results, in both the dichotic and mixed conditions, suggesting a negligible role of the third formant apart from its effect on the frequency of psychoacoustic fusions (which were neutral with regard to dominance relationships). Consequently, the differences between the dichotic and mixed conditions were the same for two-formant and three-formant syllables.

GENERAL DISCUSSION: DICHOTIC INTEGRATION

It was noted earlier that a simple "auditory averaging" model—which assumes that a single auditory stimulus, somehow intermediate between the component stimuli, is interpreted phonetically—is somewhat inadequate in explaining the data. It predicts more psychoacoustic fusions than were actually obtained, especially in the mixed condition where auditory averaging should have been perfect, and it cannot account for the effect of stimulus dissimilarity on psychoacoustic fusions (found also by Cutting, 1976). The model may be modified to include random variation in the weights of the averaging process, although the source of the variation is obscure in the mixed condition. Alternatively, one could assume that, in analogy to vision, fusion (auditory averaging) alternates with rivalry (dominance), the probability of rivalry increasing with
Figure 2: Within-subject comparison of dichotic and mixed conditions: dominance indices ($\phi_D$) for each of the seven stimuli (panels 1-7) when paired with the stimuli on the abscissa.
stimulus dissimilarity (Cutting, 1976). While this would account for the pattern of psychoacoustic fusions, the usefulness of a special model for this specific phenomenon is limited. Clearly, psychoacoustic fusions should be explainable by the same principles of interaction as other responses. In other words, an appropriate model should explain the total dominance pattern.

The simple auditory averaging model and Cutting's fusion-rivalry model allow for variable dominance relationships between pairs of stimuli, but only in a form that is related to auditory parameters. For example, consider the dominance function for stimulus 1 in Figure 2: since the starting frequency of the $F_2$ transition increases monotonically with stimulus number, the dominance function for 1 was expected to be a monotonic function (rising if lower frequencies tend to dominate higher frequencies, and falling if the opposite is true). Because of the possibly special status of straight formants, a smooth curvilinear function would also be reasonable. (BHR's data suggested that the third formant played only a negligible role.) However, there is no straightforward auditory explanation for the abrupt and striking dip of the function at stimulus 4 (that is, for the pair 1+4) and the equally abrupt reversal at stimulus 5 (1+5). Similar observations may be made in several other panels of Figure 2 (for example, panels 3 and 6). The data from the mixed condition weigh especially heavy here. Apparently, then, even when two stimuli are acoustically superimposed and/or perceived as a single syllable, the perceptual mechanism does not treat the composite information simply as the auditory average of its two constituents.

Therefore, we must turn to a different model. The model to be suggested assumes that the acoustic cues of the component stimuli remain independent and largely intact beyond the auditory processing stage, even in mixed syllables, where a stimulus with a rising transition plus one with a falling transition results in a fused stimulus with both a rising and a falling transition. We assume that to this composite information a pattern recognition process is applied that consists in comparing it with "ideal" representations ("prototypes" or "schemata"—cf. Posner, 1969; Rosch, 1975) of the relevant speech sounds in long-term memory. From these ideal representations or prototypes, the one is selected that matches the input most closely.

This process of speech recognition can be conceived as active or as passive (Morton and Broadbent, 1967). The active form is usually referred to as analysis-by-synthesis, pattern matching, or hypothesis testing. The passive form, which is preferred here on heuristic grounds, may be formulated in terms of Morton's "logogen model" (Morton, 1969) or in terms of banks of selectively tuned feature detectors (e.g., Cooper and Nager, 1975). An equivalent but more abstract conception is in terms of a multidimensional perceptual space whose dimensions are the derived auditory characteristics of the relevant set of speech sounds. The relevant response alternatives are located as fixed "ideal points" in this $n$-dimensional space, while an incoming stimulus generates a point at some location corresponding to its auditory properties. Because synthetic stimuli are acoustically much simpler than real speech (which the prototypes represent), they will be mapped into a subspace of lower dimensionality, for example, a $F_2$-$F_3$-transition-frequency plane, in the present case. The distances from the stimulus point to all prototypes are assessed in parallel, and a subsequent decision process selects the prototype with the shortest associated distance as response. A more concrete conceptualization of the calculation of distances is in terms of a "spread of excitation" from the stimulus points.
which leads the prototypes to be activated or to "resonate" in proportion to their distance from the stimulus point.

The model thus comprises three states: (1) Auditory processing, which maps an acoustic stimulus into perceptual space; (2) multycategorical processing, which generates a multycategorical vector of prototype activation values; and (3) a (uni)categorical decision, which selects the response category by determining the largest element in the multycategorical vector. (Stages 2 and 3 constitute what has been traditionally called phonetic processing—Pisoni, 1975; Studdert-Kennedy, in press.)

Random variability may arise at any of the three processing levels: in the representation of the stimulus points in perceptual space ("perceptual noise"—cf. Repp, 1975b), or in the baseline activation levels of the prototypes, or perhaps in the final decision process itself. These details will not concern us further here. The point to be made is that a stochastic pattern recognition model of this sort may provide a useful framework for explaining the speech recognition process, even when applied in an informal (that is, nonnumerical) fashion.

This model should apply to dichotic fusions or mixed stimuli as well as to any single input. Since the location of a stimulus point in the hypothetical perceptual space is determined by its derived auditory characteristics (its "acoustic cues"), and since fused or mixed stimuli contain multiple cues (for example, two different transitions of the same formant), they will lead to two stimulus points in perceptual space. The listener is usually not aware of this fact but only of the perceptual outcome that will be determined by the prototype that reaches the highest level of activation from the simultaneous presence of the two stimulus points.

This assumption predicts the most important feature of the data: the pattern of dominance relationships. The model implies that, of two fused stimuli, that stimulus will dominate that is closer to a prototype in perceptual space. In other words, stimuli close to a category boundary and far from the category prototypes will tend to be dominated by stimuli that are far from a category boundary and close to a prototype. This is what Figure 2 seems to show, on the whole. Stimulus 1, for example, dominates 5 precisely because the latter is ambiguous, whereas it does not dominate 4, which is a good /dæ/; and it dominates 7 only slightly, since 7 is a good /ɡæ/. Stimulus 2, which is a less

---

8 This "holistic" model automatically takes into account certain interactions between the processing of different features of a speech sound. An alternative model might postulate that "multicategorical" processing takes place at the auditory level, by means of selectively tuned feature detectors (e.g., Cooper and Nager, 1975) that act as auditory prototypes. This auditory stage would then be followed by a series of feature decisions whose outcomes are finally combined into a response. However, this model would have to explain why the feature detectors are selectively tuned as they are, and it would have to include additional mechanisms for the interaction of different feature decisions. It is worthwhile, therefore, to adopt the holistic model as a working hypothesis, until there is sufficient reason to reject it. We cannot decide between the two models on the basis of the present data because only a single feature is involved.
perfect /bæ/, tends to be dominated by most other stimuli, and so on. The predictions of the model are not confirmed in every detail, but they nevertheless seem to provide the best explanation of the overall pattern.

However, there are other features that the model cannot explain as it stands. Note that stimulus 7 is dominated most strongly by 3, while 1 is dominated most strongly by 4 (Figure 2). In addition, psychoacoustic fusions and the differences between dichotic and mixed pairs need to be accounted for.

Psychoacoustic fusions are explained as follows: if a stimulus in isolation receives 100 percent B responses, this does not mean that only the B prototype has been activated by this stimulus. Because of the hypothetical spread of excitation, all prototypes will be activated to some degree; but if the stimulus is sufficiently close to the B prototype and the noise in the system is not too high, the activation levels of the other prototypes will never exceed the level of the B prototype. However, in dichotic competition the activation resulting from the two stimulus points will be integrated by the prototypes, and since the D prototype is likely to lie somewhere between the B and G prototypes in perceptual space, it will profit most from this integration. If both the /bæ/ and the /gæ/ stimulus in a pair are close to the D boundary, their joint activation of the D prototype may even exceed that of the B and G prototypes. So, for example, 2 + 7 should yield more D responses than 1 + 7; and 2 + 6, more than 1 + 6, which was in fact obtained (cf. also Cutting, 1976). The component stimuli, 6 and 7, on the other hand, had no differential effect of D responses, which seems to imply that their activation of the D prototype was equal in degree. This is not quite in accord with the model, but it is plausible that differences at higher frequencies have a smaller effect than differences at lower frequencies.

The same reasoning explains why 1 was dominated most strongly by 4, but 7 was dominated most strongly by 3. Clearly, 3 is more likely to activate the B prototype than 4, so that, in the pair 1 + 3, the B activations will summate and outweigh the D activation due primarily to 3 alone. In 1 + 4, 4 will contribute less to the activation of the B prototype and D will have a stronger stand against B. The opposite argument applies when 3 and 4 are paired with 7. (These relationships are also predicted by the auditory averaging model.)

The prototype model cannot account for the differences between the dichotic and mixed conditions. Most likely, this difference can be traced back to peripheral auditory masking, which comes into play in the mixed condition. The data suggest that, in mixed syllables, rising transitions (in stimuli 1 and 2) tended to mask (dominate) falling transitions, and relatively flat formants (stimuli 3 and 4) tended to mask rising transitions. The first effect may reflect the "upward spread of masking" familiar from the auditory masking literature, while the second effect may reflect a higher susceptibility to masking of transitions in general, as compared to steady-state formants. The reduction in psychoacoustic fusions in the mixed condition was most likely due to the masking of /gæ/ by /bæ/, so that B responses increased at the expense of D and G responses.

The results pertaining to ear advantages will be discussed after two additional experiments have been reported.
EXPERIMENT III

This brief experiment served to demonstrate what had been based only on introspective evidence in Experiment I, viz. that dichotic fusions are difficult or impossible to discriminate from binaural syllables. In the Introduction, I have referred to the results of several experiments that seemed to show that place contrasts frequently, but not always (in about 60 percent of the cases), sound like a single syllable (Halves, 1969; Blumstein and Cooper, 1972; Repp, 1976). However, these studies did not differentiate between voiced and voiceless place contrasts (the latter may be less completely fused than the former), and they employed only a single, unambiguous token from each category, so that the frequent ambiguity of dichotic fusions may have assumed the role of a distinctive cue. To test the proposition that binaural and dichotic pairs cannot be distinguished, ambiguity must be made irrelevant. This is at least partially achieved by using syllables from a place continuum, so that at least one of the identical pairs will be ambiguous (stimulus 5, in the present case). The false-alarm rates ("different" responses) for this ambiguous pair should reveal whether the ambiguity cue plays any role.

Method

Subjects. Eight subjects (four men and four women) participated who had not taken part in Experiments I and II. All subjects were right-handed and without hearing trouble, with the exception of one subject who claimed to have a 5-dB hearing loss in the right ear.

Materials. The stimulus tape of Experiment I was used.

Procedure. This discrimination task was appended to Experiment IV, taking up the last 20 minutes of a session. Each subject listened first to one block of 84 syllable pairs (half identical, half nonidentical) and wrote down "1" when he thought a pair consisted of two identical syllables and "2," when it consisted of two different syllables. (To avoid confusion with the stimulus numbers, these responses will be referred to as "same" and "different," respectively.) During the next block of 84 syllable pairs, the subject merely followed the correct responses that had been filled in on the answer sheet. After this feedback trial, another block of judgments followed. The subjects were instructed that there was an equal number of identical and nonidentical pairs, and that ambiguity was not an indication that two different syllables had been presented.

Results and Discussion

As predicted, average performance was very poor, although slightly above chance (56 percent correct). The performance of three individual subjects was significantly above chance (67, 62, and 58 percent correct, respectively). BHR, who participated in four sessions, performed at chance level (51 percent correct), and so did another highly experienced listener who listened informally. The feedback did not improve performance.

A more detailed analysis was conducted in order to find out whether ambiguity played a role and whether accuracy increased with the acoustic dissimilarity of the syllables in a pair. The data are shown in Table 6. The most ambiguous identical pair, 5+5, did not show an increased false-alarm rate, suggesting that ambiguity did not serve as a distinctive cue in this task. On the other
hand, the "hit rate" for nonidentical pairs increased monotonically with the number of steps separating the two syllables in a pair. At the first glance, this seemed to suggest that within-pair acoustic dissimilarity played a role. However, a closer look at the data showed that this was probably not true, and that the result was due to the confounding of acoustic separation with the acoustic characteristics of the component syllables. (Pairs with large separations did not contain any stimuli from the middle of the continuum.) Table 6 shows that both the hit rates for nonidentical pairs and the false-alarm rates for identical pairs were greatly increased when a pair contained stimulus 1, indicating a strong bias to respond "different." Hit rates were also increased for most pairs containing stimuli 2 or 7, relative to the remaining pairs. However, within these groups of pairs (holding one stimulus constant), no clear relation to acoustic dissimilarity could be discerned.

<table>
<thead>
<tr>
<th>Stimulus number</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>61</td>
<td>69</td>
<td>75</td>
<td>75</td>
<td>69</td>
<td>75</td>
<td>81</td>
</tr>
<tr>
<td>2</td>
<td></td>
<td>54</td>
<td>50</td>
<td>31</td>
<td>50</td>
<td>50</td>
<td>50</td>
</tr>
<tr>
<td>3</td>
<td></td>
<td></td>
<td>40</td>
<td>28</td>
<td>22</td>
<td>38</td>
<td>47</td>
</tr>
<tr>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td>26</td>
<td>28</td>
<td>34</td>
<td>53</td>
</tr>
<tr>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>30</td>
<td>25</td>
<td>47</td>
</tr>
<tr>
<td>6</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>16</td>
<td>19</td>
</tr>
<tr>
<td>7</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The most likely explanation of this pattern of results is that the stimuli from the ends of the continuum had some peculiar acoustic properties, perhaps owing to the steep slope of their transitions. This artifact, which may have been due to limitations of the synthesizer or may have been psychoacoustic in nature, was apparently interpreted incorrectly as a relevant cue. The only exception to this interpretation is the very low rates of "different" responses to the pairs 6+7 and 7+7 (Table 6).

Apart from this issue, the data do provide some evidence of better-than-chance performance of some subjects, which remains an astonishing and somewhat puzzling feat. For all practical purposes, however, it may be concluded that dichotic voiced place contrasts are perceived as single syllables.

EXPERIMENT IV

The fourth experiment served three purposes. First, it attempted to demonstrate the ineffectiveness of selective-attention instructions with dichotic fusions. Although Halves (1969:Experiment 5) found no effect of selective attention in "fused" syllable pairs, a subsequent experiment of his showed a slight effect (Halves, 1969:Experiment 6). His stimuli actually included all six stop consonants and were called "fused" only because they had the same fundamental frequency. Repp (1973, 1976) has also demonstrated small selective-attention effects for such stimuli. The question here is whether the components of perfectly fused voiced place contrasts can be attended selectively.
The second purpose was a test of the hypothesis suggested by BHR's smaller REA for two-formant stimuli than for three-formant stimuli in Experiment I (Table 4). It may be that stimulus complexity (which in turn may be related to speech-likeness and naturalness) is positively correlated with the REA obtained. For this purpose, two-formant and three-formant pairs were compared in the same design. The role of the third formant in stimulus dominance relationships was also of interest.

The third purpose of Experiment IV was simply to create a more typical test situation, using only one token from each category, in order to find out how serious the problems of stimulus dominance, stimulus heterogeneity, and individual differences actually are in this more "natural" setting. Any such problems encountered should reinforce the methodological suggestions to be made in the final Discussion.

Method

Subjects. The same subjects as in Experiment III participated. However, the data of one subject who did not hear any /gæ/’s at all were excluded and replaced by data for BHR as a subject (from the first of four sessions in which he participated).

Materials. The stimuli were three syllables, with or without third formants, from the same place continuum as in the earlier experiments. The /de/ was stimulus 4 of Table 1, the /be/ had slightly more extreme transitions than stimulus 1 of Table 1 (starting frequency of F₂: 1232 Hz; F₃, if present: 2180 Hz), and the /gæ/ was intermediate between stimuli 6 and 7 (F₂: 2156 Hz; F₃: 2180 Hz).

The experimental tape contained a brief monaural practice list of 30 random syllables (five replications of each of the six stimuli). This was followed by two blocks of 180 dichotic pairs. Each block contained 10 subblocks, each representing a different randomization of 18 dichotic pairs made up from the nine possible combinations of the three syllables with two formants and with three formants, respectively. (Two-formant and three-formant stimuli were never paired with each other.)

Procedure. After trying to identify the practice syllables (and repeating the series, if necessary), the subjects listened twice to the experimental tape, that is, to four blocks of 180 dichotic pairs. For two of these blocks, the subjects were instructed to shift their attention to one side, by whatever means they found suitable. It was explained that the syllables actually consisted of two different inputs, and that only the syllables in the designated ear were to be identified. In the remaining two blocks, no selective attention was required, and the subjects simply wrote down what the fused syllables sounded like. The sequence of attention/no-attention conditions and of left-ear and right-ear selective attention was counterbalanced across subjects.

Results

The data were analyzed as in Experiment I. There was a significant overall REA (φₑ = .07, p < .01). Five of the eight subjects showed significant REAs,
one subject a significant LEA. The hypothesis of a difference in REA for two-formant and three-formant syllables was not confirmed. Although individual subjects showed considerable differences, the average $\phi_p$ indices were identical. BHR even showed a slightly larger REA with two-formant syllables, contrary to the opposite difference in Experiment I, which had given rise to the hypothesis in the first place.

The effect of selective attention was very peculiar: the differences were precisely in the wrong direction. The $\phi_p$ coefficients were $+0.12$ for left-ear attention, $+0.03$ for right-ear attention, and $+0.07$ for no-attention. The effect was very similar for two-formant and three-formant stimuli. However, no individual subject showed any clear evidence of consistent positive or negative selective attention effects, so that the inverted pattern may have been due to chance. Two subjects showed an inversion of the REA as a function of selective attention but regardless of the ear attended to.

The frequency of psychoacoustic fusions was low (12 percent), as expected with acoustically dissimilar stimuli. This percentage excludes the data of BHR who, as in Experiment I, showed a much higher frequency (35 percent). Quite surprisingly, and contrary to BHR's control results in Experiment I, psychoacoustic fusions were more frequent with three-formant than with two-formant stimuli (15 vs. 8 percent for the seven subjects; 41 vs. 28 percent for BHR).

There was a reliable difference in the stimulus dominance pattern between two-formant and three-formant syllables, which is shown in Table 7 and may be characterized as a reduction in the "strength" of /dæ/ when the third formant was removed. This was already evident in the identification of binaural pairs: the two-formant /dæ/ received only 86 percent correct responses, while the three-formant /dæ/ received 94 percent. (The intelligibility of the other stimuli did not change.) Table 7 shows that, with three formants present, /dæ/ dominated /ba/ and /gæ/. With two formants, the pattern was reversed. This indicates that an $F_3$ transition was more important for /dæ/ than for /bæ/ and /gæ/; and it supports the hypothesis, set forth earlier, that a poor representative of a category will be dominated by better examples of other categories. Again, however, there were large individual differences in dominance patterns.

The $\phi_p$ coefficients for the three individual stimulus pairs (which were similar for two- and three-formant stimuli) are also shown in Table 7. Surprisingly, /dæ+gæ/ pairs did not exhibit an average REA. BHR (who participated in four sessions) even showed a LEA with this pair, but a clear REA with the other two. However, apart from BHR's data, this phenomenon was not reliable for individual subjects who showed large variations in their ear advantages for individual pairs. Both the /dæ+gæ/ anomaly and the high variability are somewhat disconcerting. It will be recalled that Experiment I did not show any comparable effect.

---

9 This was the subject who claimed to have a 5-dB hearing loss in the right ear. However, it would be quite surprising if this had been the cause of the dichotic asymmetry, considering that channel differences much larger than 5 dB have only little effect on the dichotic ear advantage at the intensities used here (Speaks and Bissonette, 1975).
TABLE 7: Stimulus dominance indices for individual stimulus pairs, and ear dominance indices (averaged over two- and three-formant stimuli). (Note: A positive $\phi_D$ index indicates dominance of the stimulus named first.)

<table>
<thead>
<tr>
<th></th>
<th>/bæ + dæ/</th>
<th>/dæ + gæ/</th>
<th>/bæ + gæ/</th>
</tr>
</thead>
<tbody>
<tr>
<td>three-formant $\phi_D$</td>
<td>-0.31</td>
<td>+0.40</td>
<td>+0.45</td>
</tr>
<tr>
<td>two-formant $\phi_D$</td>
<td>+0.09</td>
<td>-0.31</td>
<td>+0.32</td>
</tr>
<tr>
<td>average $\phi_E$</td>
<td>+0.14</td>
<td>-0.01</td>
<td>+0.11</td>
</tr>
</tbody>
</table>

GENERAL DISCUSSION: II. MEASURING THE EAR ADVANTAGE

The presence of a significant average REA for dichotic fusions is evidence that, despite the subjective impression of a single syllable, the information from the two ears remains functionally separated until it converges upon the dominant hemisphere. It makes unlikely a low-level auditory mixing mechanism that combines spectrally similar information and routes it to both hemispheres, because such a mechanism would have to be influenced by hemispheric dominance. Rather, it seems that each stimulus first arrives at the contralateral hemisphere, and integration takes place only when the information is recombined after considerable auditory (and perhaps even initial phonetic) processing in each hemisphere, which has been a common assumption in dichotic listening research (Studdert-Kennedy and Shankweiler, 1970). The REA for dichotic fusions challenges an interpretation in terms of spatial location only (Morais and Bertelson, 1973; Morais, 1975). Since only a single stimulus is heard that is localized in the median plane, the hypothesis that stimuli that come from the right are perceived more accurately does not apply.

The subjective phenomenon of fusion (hearing only a single stimulus) probably does arise from a low-level cross-correlational mechanism, but it is apparently separate from, and unrelated to, the subsequent allocation and integration of information. This has two interesting implications: (1) in the limiting case, identical binaural stimuli may also be independently transmitted to their respective contralateral hemispheres and perceptually combined only at a central level; and, more importantly, (2) the identification of less completely fused dichotic stimuli (e.g., voicing contrasts) should be explainable by the same principles as the identification of dichotic fusions, for example, by the prototype model proposed earlier. This view is in basic agreement with the conclusions of Halwes (1969), who found that subjective fusion versus nonfusion was largely irrelevant to the pattern of responses.

It also follows from these conclusions that other types of dichotic contrasts should lend themselves to the one-response, no-attention requirement ("What does it sound like?") whose advantages over the two-response paradigm have already been outlined in the Introduction (cf. Geffner and Dorman, in press, who used this method successfully with four-year-old children). However, what makes voiced place contrasts especially convenient from a methodological standpoint is (1) that the task is "natural" because the listeners are not aware of different inputs to the two ears, (2) that the fused stimuli do not sound strange (as other dichotic contrasts often do) but similar to binaural syllables, (3) that they do not invite selective attention strategies (however ineffective
they may be), and (4) that relatively few responses are given that are ambiguous with respect to ear dominance (psychoacoustic fusions). The last problem can be completely eliminated by simply omitting /bæ+ɡæ/ pairs from dichotic tests. A dichotic test composed only of /bæ+ɡæ/ and /dæ+ɡæ/ pairs, interspersed with binaural controls, should be a useful instrument to try out.

However, such a test still presents some major problems. Foremost among these is the phenomenon of stimulus dominance and the large individual variations connected with it. Extreme dominance of one stimulus in a pair must be prevented; otherwise, this dichotic pair will provide no information about ear dominance. Then, there is the important question of the relationship between stimulus dominance and ear dominance that parallels, but is not identical with, the question of the relationship between performance level and ear dominance in the two-response paradigm (Kuhn, 1973). Finally, there is the question of item homogeneity: Do different dichotic pairs measure the ear advantage to the same degree, even if they have equal stimulus dominance coefficients?

Unlike performance level in the two-response paradigm, which is a global index and cannot be manipulated by the experimenter, stimulus dominance is a characteristic of individual stimulus pairs and can be controlled to a certain degree by manipulating stimulus parameters, as demonstrated in Experiment I. There are two possible ways of making use of this control. One is to try to minimize stimulus dominance and to bring all stimulus pairs as close to equilibrium ($\phi_D = 0$) as possible. Because of individual differences, construction of a single optimal test is out of the question. An appropriate method would be testing under computer control, where, during an initial adaptive phase of testing, the computer keeps track of the responses and adjusts the stimulus parameters to reduce asymmetries. Such a procedure is worth exploring but has some drawbacks: it does not guard against drifts of stimulus dominance during the actual testing phase, and it requires sophisticated equipment and therefore is of little value outside the laboratory. The other alternative is to construct a test containing a variety of stimuli, so that the individual pairs span a wide range of stimulus dominance relationships (as in Experiment I). In order to derive a valid measure of ear dominance, in this case, the nature of the relationship between stimulus dominance and ear dominance must be known. Since it is reasonable to expect that ear dominance will be maximal when stimulus dominance is minimal, a global $\phi_E$ index obtained from summed response frequencies (as in Experiment I) or from averaged ear dominance coefficients for individual pairs will underestimate the "true" ear advantage and will not be comparable from individual to individual, because of different individual stimulus dominance patterns. A method for inferring the true ear advantage is needed.

The situation is formally analogous to that in signal detection. Ear dominance represents "sensitivity" and stimulus dominance represents "bias." When there is extreme bias ($\phi_D = \pm 1$), sensitivity cannot be determined ($\phi_E = 0$). When sensitivity is optimal ($\phi_E = \pm 1$), there cannot be any bias ($\phi_D = 0$).

---

10 The question of performance level also arises in the present paradigm, in the form of confusions. As long as the confusions are not too numerous, however, their impact is negligible because of the weighting procedure employed [Eq. (1)]. There are some individuals, however, who seem to be unable to give consistent identification responses to the synthetic syllables used here.
Between these extremes, the two tendencies mutually constrain each other. For example, when $T_{1}(1)/N = 0.8 (\phi_{P} = +0.75)$, it can easily be shown that $T_{RE}/N$ is restricted to the range between 0.3 and 0.7 ($\phi_{E}$ between $\pm 0.5$); and $\phi_{E}$ constrains $\phi_{P}$ in a similar fashion. In order to apply the methods of signal detection theory, one event (for example, responding i when i-j is presented, with i in the right ear) may be arbitrarily chosen to represent "hits," and another event (responding i when j-i is presented, with i in the left ear), "false alarms." However, the crucial requirement is that sensitivity (namely, the "true" ear advantage) be independent of the bias (stimulus dominance). Since stimulus dominance is varied by changing the characteristics of the stimuli (rather than by manipulating the listeners' criteria), it is an important empirical question whether all items are homogeneous (in the test-theoretical sense) and measure the same kind of ear advantage, so that all stimulus pairs can be represented as points on the same single receiver-operating-characteristic function.

The results of the present experiments create some doubts about whether the homogeneity assumption will be tenable. When plotted as "hits" versus "false alarms," the stimulus pairs of Experiment I exhibited considerable scatter, perhaps owing to the high individual variability in the data. There was also a tendency for $\phi_{E}$ to increase with the acoustic dissimilarity of the component stimuli in a dichotic pair. At the same time, there was no negative correlation between $\phi_{E}$ and $|\phi_{P}|$ ($r = +0.04$), so that an increase in $\phi_{E}$ could not be explained by a simultaneous decrease of dominance asymmetries. In Experiment IV, one of the three stimulus pairs showed no REA. Again, this was not related to stimulus dominance (cf. Table 7). As a result, no monotonic receiver-operating characteristic function will fit these data well. Further research will be required to determine the reliability of the present findings. It may be useful to compare variations in stimulus dominance produced by varying stimulus parameters with similar variations introduced by other means, such as adaptation (Cooper, 1974; Miller, 1975).

A more explicit model of dichotic interaction would also contribute to the solution of this methodological problem. In mathematical terms, stimulus dominance (bias) and ear dominance (sensitivity) mutually constrain each other. However, in the actual processing chain, the constraint may well be unidirectional, since it is highly likely that the two asymmetries arise at different stages in processing. Since stimulus dominance effects were more pronounced than ear dominance effects but did not correlate with the latter, the present data suggest that the cause of ear dominance precedes the cause of stimulus dominance in the processing hierarchy. This is in agreement with the hypothesis that ascribes ear dominance to transcallosal transmission loss but stimulus dominance to subsequent integration of information in the dominant hemisphere.

REFERENCES


