

*Psychological Review*  
1976, Vol. 83, No. 2, 114-140

Auditory and Linguistic Processes in Speech Perception:  
Inferences from Six Fusions in Dichotic Listening

James E. Cutting

# Auditory and Linguistic Processes in Speech Perception: Inferences from Six Fusions in Dichotic Listening

James E. Cutting  
Wesleyan University and  
Haskins Laboratories, New Haven, Connecticut

A number of phenomena in speech perception have been called fusion, but little effort has been made to compare these phenomena in a systematic fashion. The present paper examines six of them. All can be exemplified using the syllable /da/ as in *dot*, and all occur during dichotic listening. In each type of fusion, the robustness of the fused percept is observed against variation in three parameters: the relative onset time of the two opposite-ear stimuli, their relative intensity, and their relative fundamental frequency. Patterns of results are used to confirm the arrangement of the six fusions in a hierarchy, and supporting data are summoned in an analysis of the mechanisms that underlie each with reference to speech. The six fusions are sound localization, psychoacoustic fusion, spectral fusion, spectral/temporal fusion, phonetic feature fusion, and phonological fusion. They occur at three, perhaps four, different levels of perceptual analysis. The first two levels are characterized by perceptual integration, the other(s) by perceptual disruption and recombination.

Many accounts of speech and language emphasize a hierarchical structure (see, e.g., Fry, 1956; Pisoni, in press; Studdert-Kennedy, 1974, in press). Recently, the interface between two particular levels in this hierarchy has aroused much attention: the general level logically just prior to linguistic analysis, typically called the *auditory* level, and the first tier of the language hierarchy logically just subsequent to that auditory analysis, the *phonetic* level (Cutting, 1974; Pisoni, 1973; Studdert-Kennedy, Shankweiler, & Pisoni, 1972; Wood, 1975; Wood, Goff, & Day, 1971). These and other levels of processing appear to operate in parallel, but the outcome at one level appears to be contingent

on the output at a prior level (Marslen-Wilson, 1975; Wood, 1974, 1975).<sup>1</sup> The present paper looks at these and other levels from a new vantage point.

Information-processing analyses assume that perception takes time and that systematic disruption or complication of the process can reveal its underlying properties. This epistemological position often leads the researcher to paradigms of masking in both visual (Turvey, 1973) and auditory (Darwin, 1971; Massaro, 1972, 1974) modalities. Masking occurs through the *rivalry* of two stimuli competing for the limited processing capacities of a single processor: Information is lost at a bottleneck in the perceptual system. The reciprocal process to rivalry, one equally suited to information-processing analysis, is *fusion*. In this process, information from the two stimuli is not strictly lost, but rather

---

This research was supported in part by Grant HD-01994 from the National Institute of Child Health and Human Development to the Haskins Laboratories, and by a seed grant from Wesleyan University to the author. Initial portions of this research were reported in Cutting (Note 1) and in Cutting (1973), a doctoral dissertation submitted to Yale University. I thank Ruth S. Day for her helpfulness in all stages of this enterprise, and Michael Studdert-Kennedy, Michael Turvey, Terry Halwes, Bruno Repp, and the reviewers for many suggestions to improve the paper.

Requests for reprints should be sent to J. Cutting, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510.

---

<sup>1</sup> More recently, however, many of the phenomena thought to characterize phonetic perception have been found to occur for music and musiclike sounds (Bever & Chiarello, 1974; Blechner, Day, & Cutting, in press; Cutting, in press; Cutting & Rosner, 1974; Cutting, Rosner, & Foard, in press; Locke & Kellar, 1973). Among other implications, these results suggest that "phoneticlike" perception is characteristic of general higher level processing in the auditory system encompassing both speech and music.

transformed into something new. With the outstanding exception of Julesz (1971), fusion has received little systematic attention in vision, and the phenomenon has received essentially no systematic attention in audition. The present investigation takes a small step in this direction.

One reason that little attention has been paid to auditory fusions may be that, as Julesz (1971, p. 52) suggests, the "auditory system is basically an archaic, diffuse structure that is hard to probe." A second reason may be that seemingly too large a number of auditory phenomena have been called fusion. When one reads a given paper in this field (e.g., Broadbent & Ladefoged, 1957; Day, 1968; Halwes, 1969; Perrott & Barry, 1969; Sayers & Cherry, 1957), it is clear with what phenomenon its authors are concerned; moreover, one feels confident that each author or group of authors has properly labeled each phenomenon as a fusion. However, when one inspects the papers as a group, it is not clear that they share any common ground except for two superficial facts: They all use the word *fusion* in their titles and they all present their stimuli dichotically—that is, one stimulus to one ear and one to the other. Fusion is clearly not one phenomenon but many phenomena, yet how are they related? At best, these findings appear to be just "curious binaural phenomena" (Tobias, 1972); at worst, they may lead the careful reader to *confusion*. The purpose of this paper, then, is (a) to enumerate the different kinds of auditory fusion, (b) to arrange six of the dichotic phenomena relevant to speech processing in a hierarchy according to the processing characteristics implied by each, and (c) to confirm that arrangement by subjecting each fusion to a common set of presentational and stimulus variables that have proved important to auditory processing in general.

The list of fusions considered here is not intended to be exhaustive, but is merely organized with regard to three themes. First, all fusions discussed here are *dichotic* fusions, combinations of stimuli presented to opposite ears. This stipulation eliminates temporal fusions of repeating noise patterns (Guttman &

Julesz, 1963), tone patterns (van Noorden, 1975), and briefly interrupted segments of speech (Huggins, 1964, 1975). Second, all fusions in the present study reflect processes underlying the perception of *speech*. This constraint eliminates consideration of centrally generated perceptions of simple pitches (Bilsen & Goldstein, 1974; Cramer & Huggins, 1958; Fourcin, 1962), patterns of pitches (Kubovy, Cutting, & McGuire, 1974), musical intervals and chords (Houtsma & Goldstein, 1972), musical illusions (Deutsch, 1975a, 1975b; Deutsch & Roll, 1976), or integrated pulse trains (Huggins, 1974). Third, all of the selected fusions are exemplified by a single rule: The fused percept is *different* from the two dichotic inputs. This rule eliminates the dichotic switching-time experiments of Cherry and Taylor (1954) and Sayers and Cherry (1957) using speech stimuli, and the phenomenon of masking-level difference (see, e.g., Jeffress, 1972). Masking-level difference is typically eliminated by the second stipulation—most often, tones are imbedded in noise, and alteration of interstimulus phase relations yields percepts of either tone-plus-noise or noise alone. However, this is not necessarily always the case. Speech stimuli can easily be imbedded in noise and their intelligibility increased through the manipulation of phase relations.

It may seem that these constraints eliminate all the possible fusions that might occur in audition. However, there are at least five others that are relevant to speech, and they will be discussed with regard to a sixth and most basic type of fusion, sound localization. Since previous authors have simply called their phenomena *fusion*, I have taken the liberty in most cases of adding a descriptive adjective. The fusions considered in this paper are (a) sound localization, (b) psychoacoustic fusion, (c) spectral fusion, (d) spectral/temporal fusion, (e) phonetic feature fusion, and (f) phonological fusion. For purposes of comparability, each is discussed primarily with regard to the syllable /da/, as in *dot*, as shown in Figure 1. Schematic spectrograms of the dichotic stimuli are presented to suggest the manner in which items can be perceptually combined. The present paper

TABLE 1

PREVIOUS UPPER LIMITS OF INTERSTIMULUS DISCREPANCIES PERMITTING THE FUSION OF SOUNDS PRESENTED TO OPPOSITE EARS FOR SIX TYPES OF FUSION

Fusion type	Onset time (msec)	Intensity (db.)	Frequency (Hz)
Sound localization	2.5 <sup>a</sup> 10 <sup>b</sup>	60 <sup>c</sup> 65 <sup>d</sup>	25 <sup>e</sup> 80 <sup>f</sup> < 2 <sup>g</sup>
Psychoacoustic fusion <sup>h</sup> Spectral fusion <sup>i,j,k</sup>	< 250 <sup>l</sup> < 5 <sup>l</sup>	— 40 <sup>m,n</sup>	< 25 <sup>l</sup>
Spectral/temporal fusion <sup>m</sup> Phonetic feature fusion <sup>n,o</sup>	— 60 <sup>p</sup> < 120 <sup>q</sup> > 150 <sup>r</sup> < 200 <sup>s</sup>	— 40 <sup>m</sup> 20 <sup>r</sup> > 15 <sup>u</sup>	— > 14 <sup>z</sup> > 20 <sup>u</sup>

<sup>a</sup> Woodworth (1938, p. 528) for clicks.

<sup>b</sup> Cherry & Taylor (1954) for natural speech stimuli. Also see Tobias (1972) and Babkoff (1975).

<sup>c</sup> Groen (1964) for binaural beats of sine waves.

<sup>d</sup> Application of "cyclotean" stimuli of Kubovy, Cutting, & McGuire (1974).

<sup>e</sup> Licklider, Webster, & Hedlund (1950) and Perrott & Nelson (1969) for binaural beats.

<sup>f</sup> Perrott & Barry (1969) for sine waves near 2,000 Hz; considerably greater differences possible for higher frequencies. Also see Thurlow & Eliner (1959) and Tobias (1972).

<sup>g</sup> Halwes (1969) for synthetic speech stimuli.

<sup>h</sup> Existence of this fusion gleaned from Halwes (1969). Also see Repp (1975d).

<sup>i</sup> Broadbent (1955) for natural speech patterns.

<sup>j</sup> Broadbent & Ladeloged (1957) for synthetic speechlike patterns.

<sup>k</sup> Leakey, Sayers, & Cherry (1958) for nonspeech; Matzker (1959), Linden (1964), and Smith & Resnick (1972) for natural speech; Halwes (1969), Ades (1974), and Haggard (Note 2) for synthetic speech. Several examples cited by Tobias (1972) may also fit into this category.

<sup>l</sup> Pilot research by author using metronome-like ticks.

<sup>m</sup> Rand (1974) for synthetic speech.

<sup>n</sup> Nye, Nearey, & Rand (Note 3) and Nearey & Levitt (Note 4) for synthetic speech.

<sup>o</sup> Shankweiler & Studdert-Kennedy (1967, subsequent analysis) for synthetic speech stimuli; Studdert-Kennedy, & Shankweiler (1970) for natural speech.

<sup>p</sup> Estimated from Studdert-Kennedy, Shankweiler, & Schulman (1970) for synthetic speech.

<sup>q</sup> Repp (1975a, 1975b, 1975c) for synthetic speech.

<sup>r</sup> Estimated from Cullen, Thompson, Hughes, Berlin, & Samson (1974) and Speaks & Bissonette (1975) for natural speech stimuli.

<sup>s</sup> Day (1968) for natural speech stimuli.

<sup>t</sup> Day (1970) for synthetic speech; Day & Cutting (Note 5) for natural speech.

<sup>u</sup> Cutting (1975) for synthetic speech.

is concerned with the pressures that can be placed on the perceptual system to inhibit fusion. All fusions are more or less subject to these pressures.

In general, three variables have proven informative in previous investigations of these fusions: relative onset time of the two stimuli, their relative intensity, and their relative fundamental frequency. Table 1 summarizes the results of this previous research, which used many different kinds of stimuli. Only in rare cases were the same stimuli employed to investigate more than one type of fusion or even to investigate more than one parameter within a given fusion type. The following

overview of the six types of fusion incorporates the material from both Figure 1 and Table 1.

### Sound Localization: Fusion of Two "Identical" Events

Sound localization is included here as a reference point to be used when considering the other forms of fusion. All audible sounds, simple or complex, can be localized—and usually are. Sound localization is the most basic form of fusion and occurs for both speech sounds and nonspeech sounds alike. Provided that microsecond accuracy is not crucial, a convenient way to study sound localization in the laboratory is to use the same apparatus needed for studying other types of fusion: a good set of earphones, a dual-track tape recorder, and a two-channel tape with appropriate stimuli recorded on it. Although approximate sound localization can be obtained using just one ear (Angell & Fite, 1901; Mills, 1972; Perrott & Elfiner, 1968), discussion in the present paper is limited to the phenomenon obtained with both ears.

The three primary parameters that affect sound localization were mentioned previously: the relative timing of the events at each ear, the relative intensity of those events, and also their relative frequency. First, consider relative timing. If one presents a brief click simultaneously to each ear, the listener reports hearing one click localized at his or her midline. Delaying one click by as little as .1 msec causes the apparent source of the percept to move away from the midline toward the ear receiving the leading stimulus. Delaying that click by 1 msec causes the apparent source to move to the extreme side of the auditory field away from the delayed click. With delays (onset time differences) of 2.5 msec, the fused percept disintegrates and two clicks can be heard (Woodworth, 1938) shooting across the auditory field, one after the other. Apparently, the effect of disintegration is postponed for longer and more complex stimuli such as speech syllables until relative phase differences (or onset time differences) are as great as 10 msec (Cherry & Taylor, 1954) or more (Tobias, 1972). Thus, when two /da/s are presented to op-

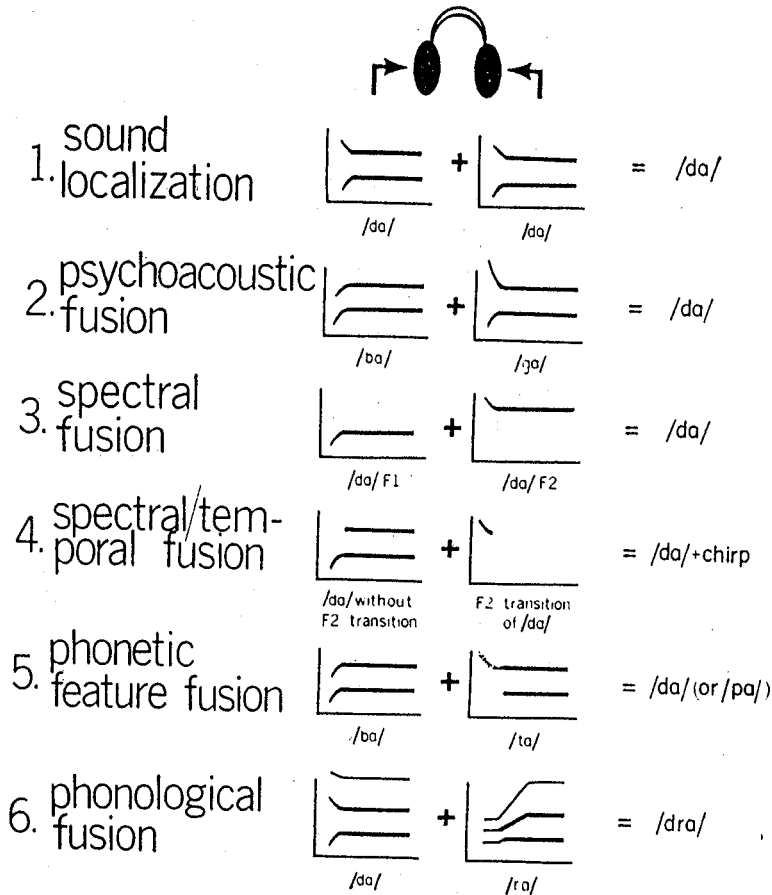


FIGURE 1. Schematic spectrograms of stimuli used in the six fusions. (F1 = Formant 1; F2 = Formant 2.)

posite ears with as much as 10 msec separating their onsets, a single /da/ may still be heard. Experiment 2 is designed, in part, to confirm this finding.

Intensity is a second parameter that affects sound localization. Interaural differences as small as a few decibels or less easily affect the perceived locus of a sound. The problem here, however, is that unlike the potential fused percept in other fusions, the fused percept in sound localization does not "disintegrate." By making one stimulus less and less intense compared to a stimulus of constant intensity in the other ear, the locus of the percept migrates from the midline toward the ear receiving the most intensive stimulus. Most importantly, the difference between percepts in some binaural and monaural condi-

tions is negligible, if detectable at all. I have found that when Stimulus A is presented to one ear at a comfortable listening level and to the other ear at 20 db. lower, and also when Stimulus A is presented to one ear at a comfortable listening level and *no* stimulus is presented to the other ear, one cannot consistently hear the difference between the two trials.

To get around this problem, at least with regard to sine waves, one can look to the phenomenon of binaural beats: a "whooshing" or "roughness" percept, depending on the frequency difference between sine waves (to be discussed below). This phenomenon is perceived through the cross-correlation of opposite-ear signals. Groen (1964) presented data suggesting that with the intensity of the

signals differing by as much as 60 db., a "pulling" is still heard. Although Groen's procedure is not clear, I have replicated that result using the stimuli of Kubovy, Cutting, and McGuire (1974). In the case of their stimuli, a melody is heard through the cross-correlation and subsequent localization of the melodic elements as spatially distinct from a background of complex noise. No melody can be heard in a single stimulus because physically it is not present (which distinguishes the phenomenon from masking-level difference). The percept is robust enough so that if one stimulus is presented to one ear at 100 db. (SPL) and the other stimulus to the other ear at 35 db., a faint melody can still be heard and identified. The fact that it can be heard suggests that the sound localization process is still functional at interaural intensity differences as great as 65 db. Nevertheless, intensity is not considered relevant to the fused percept when /da/ is presented to both ears, because the percept never ceases to be /da/.

The third parameter is frequency. Sine waves may differ in frequency by as much as 25 Hz in certain frequency ranges, and a fused percept, the "roughness" of binaural beats, can be maintained. The principle here is that stimuli differing slightly in frequency can be thought of as stimuli with identical frequencies but with constantly changing phase relations. Differences of 3 Hz or less can be heard as an oscillating stimulus whirling around the head in an elliptical path (Oster, 1973). Greater frequency differences are heard as roughness until the two tones break apart. Outside the realm of binaural beats, Perrott and Barry (1969) found that dichotic tones of considerably greater frequency differences could be heard as a single stimulus, especially above about 2000 Hz. However, when the signals are complex and periodic, the pitch is typically much lower—for speech sounds, in particular, a fundamental frequency of 100 Hz is not uncommon in male speakers. It can be easily demonstrated that a /da/ at 100 Hz presented to one ear and a /da/ of 102 Hz presented to the other ear are heard as two events. Experiment 4 is designed, in part, to confirm this finding.

### *Psychoacoustic Fusion: Fusion of Proximal Acoustic Features by Perceptual Averaging*

Unlike sound localization, about which many volumes have been written, little has been written about psychoacoustic fusion. In this type of fusion, acoustic features from opposite-ear stimuli appear to be averaged to yield an intermediate result. The phenomenon could logically occur for many different kinds of sounds, both speech and nonspeech. Its existence is gleaned from my own experimentation with dichotic synthetic speech stimuli and from a large table of data presented by Halwes (1969, p. 61). Perhaps the best way to demonstrate the phenomenon is to consider the stimuli in Figure 1. If this particular /ba/ is presented to one ear and this particular /ga/ to the other, the listener often reports hearing a single item, /da/. Notice that these stimuli differ only in the direction and extent of the transition in the second formant, the resonance of highest frequency for these items (Delattre, Liberman, & Cooper, 1955). Note further that the second-formant transitions are arrayed such that /da/ lies between /ba/ and /ga/, and that an "average" of the two extreme items would fall very close to /da/.

To date, little is known about this type of fusion, in part because it is less likely to occur with the more complex and more widely differing natural speech syllables. Therefore, Experiment 1 is designed to demonstrate the psychoacoustic nature of the phenomenon. Also in Experiment 1, the relation between this fusion and a similar phenomenon known as the "feature sharing effect" is considered (see Pisoni & McNabb, 1974; Repp, 1975a; Studdert-Kennedy & Shankweiler, 1970; Studdert-Kennedy et al., 1972). Experiments 2, 3, and 4 are designed so that the effect of the three crucial variables on psychoacoustic fusion can be observed.

### *Spectral Fusion: Fusion of Different Spectral Parts of the Same Signal*

Broadbent (1955) and Broadbent and Ladefoged (1957) reported this third phenomenon, which I call spectral fusion. It occurs when different spectral ranges of the same signal are presented to opposite ears. A given stimulus, for example, is filtered into

two parts: one containing only the low frequencies and the other containing only the high frequencies. Each is presented separately but simultaneously to opposite ears. The listener invariably reports hearing the original stimulus, as if it has undergone no special treatment. In his initial study, Broadbent found that this fusion readily occurred for complex stimuli of many types—for non-speech sounds such as metronome ticks as well as for speech sounds. Moreover, when listeners were informed about the nature of the stimuli and asked to report which ear received the low frequency sounds and which ear the high frequency sounds, they performed at chance level.

Relative timing is an important parameter in spectral fusion. Broadbent (1955) found that onset time differences of 250 msec were sufficient to disrupt the fused percept. My own pilot research suggests that this interval may be at least an order of magnitude too large. For example, when the different spectral portions of metronome-like ticks are offset by as little as 5 msec, the listener hears two sets of ticks, not one. As in sound localization, the temporal differences tolerable in spectral fusion may be greater for more complex sounds such as speech items. Therefore, Experiment 2 is directed at finding the relative onset times limiting fusion when the speech syllable /da/ is spectrally split and its first formant presented to one ear and its second formant to the other. For generality, the syllables /ba/ and /ga/ are also used.

The effect of relative intensity on spectral fusion has been explored systematically by Rand (1974; see also Nye, Nearey, & Rand, Note 3; Nearey & Levitt, Note 4). The most interesting case occurs when the second and higher formants, presented to one ear, are decreased in amplitude with respect to the first formant, presented to the other ear. Rand found that decreases of as much as 40 db. have little effect on identification of /ba, da, ga/. This large effect is particularly surprising since attenuations in the upper formants of only 30 db. are sufficient to render the syllables unrecognizable when all formants are presented to both ears. Rand termed the phenomenon "dichotic release from masking," and the release is clearly substan-

tial. However, the emphasis in the present paper is not on masking but on fusion, and since Rand (1974) used stimuli very similar to those used in the present study and Nye, Nearey, and Rand (Note 3) have already replicated those results, intensity effects on spectral fusion are not explored further.

Fundamental frequency is also an important parameter in spectral fusion. Broadbent and Ladefoged (1957) found that the fundamental frequencies of the to-be-fused stimuli must be identical for fusion to occur (that is, for one item to be heard). Differences of 25 Hz inhibited fusion, and Halwes (1969) suggested that differences of 2 Hz may inhibit fusion as well, but these investigators appear to have been dealing with two types of fusion: One concerns the number of items heard (one or two), and the other concerns the identity of the stimulus. While the effect of differences in pitch between the two component stimuli on the *identity* of the fused percept is considered in Experiment 4, the effect of fundamental frequency on the *number* of items heard is considered in Experiment 5.

#### *Spectral/Temporal Fusion: Perceptual Construction of Phonemes from Speech and Nonspeech Stimuli*

Rand (1974) discovered a fourth type of fusion. In addition to dividing the stimulus spectrally and presenting those portions to either ear, as noted previously, he divided the speech syllable both spectrally and temporally. Two-formant renditions of his stimuli are shown schematically in Figure 1. One stimulus is simply the second-formant transition excised from the syllable and presented in isolation. Mattingly, Liberman, Syrdal, and Halwes (1971) noted that these brief glissandi sounded like the discrete elements of birdsong, and they dubbed them "chirps." The other stimulus is the remainder of the syllable without the second-formant transition. It should be noted that the transitionless /da/—that is, the speech sound without a second-formant transition—is *not* identifiable as /da/ when presented in isolation; instead, it is almost 85% identifiable as /ba/. This appears to result from the upward spread of harmonics of the first-formant transition, which seems to mimic the now-absent second-

formant transition in such a manner as to cue /b/.

Perhaps the most interesting aspect of spectral/temporal fusion, and the aspect that distinguishes it from the logically similar spectral fusion, is that the listener hears more than one auditory event. He does not hear two *speech* sounds. Instead, he hears one speech sound, /da/, and one nonspeech sound, a chirp. Note that the perceptual whole is greater than the sum of the parts: The listener "hears" the second-formant transition in two different forms at the same time. One form is in the complete syllable /da/, which would sound more like /ba/ without it. The second form is similar to the transition heard in isolation—a nonspeech chirp. Thus, spectral/temporal fusion is more complex phenomenologically than the three fusions previously considered. It may be possible for spectral/temporal fusion to occur for nonspeech sounds (perhaps a complex nonspeech sound could be segmented spectrally and temporally in the same manner, and analogous percepts obtained). Nevertheless, it is discussed in the present paper exclusively with respect to speech.

Of the three relevant parameters—onset time, intensity, and frequency—only intensity has been explored thus far for spectral/temporal fusion. As in spectral fusion, Rand (1974) attenuated the isolated second-formant transitions of /ba, da, ga/ by as much as 40 db., and identification was largely unimpaired. The result was in marked contrast to that produced by the condition in which the syllable remained as an integral whole but with the second-formant transition attenuated as before: 30 db. was then sufficient to impair identification. As in spectral fusion, the intensity data are not replicated here, but the effects of differences in relative onset time and frequency are explored in Experiments 2 and 4, respectively. Again, /da/ is used as a reference syllable, but /ba/ and /ga/ are also used for the sake of generality.

*Phonetic Feature Fusion: Recombination of Phonetic Feature Values by Perceptual Misassignment*

With this fifth type of fusion we move to a domain that belongs exclusively to speech.

Studdert-Kennedy and Shankweiler (1970), Halwes (1969), and Repp (1975a) have reported that misassignment of phonetic feature values often occurs in the dichotic competition of certain stop-vowel syllables. This "blending" can be thought of as phonetic feature fusion. Consulting Figure 1, one finds that when /ba/ is presented to one ear and /ta/ to the other, the listener often reports hearing a syllable not presented. The most frequent errors are the blends /da/ and /pa/, in which the listener combines the voicing feature value of one stimulus with the place feature value of the other. For example, the voicing value of /b/ is combined with the place value of /t/ and the result is the *fusion* /d/.

Consider a stimulus repertory of six items: /ba, da, ga, pa, ta, ka/. On a particular trial when /ba/ and /ta/ are presented to opposite ears and the subject is asked to report what he or she hears, three types of responses can occur: correct responses /ba/ or /ta/, blend responses /da/ or /pa/, and anomalous responses /ga/ or /ka/. The last two items are anomalous because, although they share the voicing value with one item in the stimulus pair, neither shares place value. Using natural speech stimuli, Studdert-Kennedy and Shankweiler (1970) found that the ratio of blends (phonetic feature fusions) to anomalous responses was about 2:1, a rate significantly greater than chance. For synthetic speech items, Halwes (1969, p. 65) found them to occur at a rate of 10:1 or better (p. 64). Studdert-Kennedy et al. (1972) found these fusions to occur even when the vowels of the two stimuli differed markedly.

Evidence for the effect of relative onset time on phonetic feature fusion is only indirect. Studdert-Kennedy, Shankweiler, and Schulman (1970) found that errors in identification occur *more* often when the relative onset times of competing pairs of stimuli are slightly staggered than when they are simultaneous. The effect decreases substantially for relative onset times of greater than 70 msec or so. The maximum error rate occurs for asynchronies of about 40 msec. If we assume that the ratio of blend responses to anomalous responses is constant for different lead times, maximum phonetic feature fusions should



occur at about 40 msec lead time but should fall off rather rapidly thereafter. Experiment 2 is designed in part to confirm these predictions.

Evidence for the effect of relative intensity on phonetic feature fusion is equally indirect. The data of Cullen, Thompson, Hughes, Berlin, and Samson (1974) demonstrate that dichotic items can compete with one another, that is, yield substantial error rates, when the two items differ by as much as 20 db. If we assume that the ratio of blend responses to anomalous responses is constant for different intensities, phonetic feature fusions should continue to occur rather readily until intensity differences between the two stimuli are greater than 20 db., at which point errors largely cease. Errors (and fusions) should be greatest when the two stimuli have the same intensity. Experiment 3 is designed in part to confirm these predictions.

The effect of fundamental frequency differences between the stimuli is better known for phonetic feature fusion. Halwes (1969) found these fusions to occur almost as frequently when the two competing stimuli had different fundamental frequencies as when they had the same fundamental frequency. Experiment 4 extends the frequency differences well beyond the 14 Hz of Halwes to observe the effect on reported fusions.

#### *Phonological Fusion: Perceptual Construction of Phoneme Clusters*

Phonological fusion occurs when two inputs, each of  $n$  phonemes, yield a response of  $n + 1$  phonemes. Day (1968) found that compatible phoneme strings, one beginning with a stop consonant and the other with a liquid, could be fused into one unit: Given PAHDUCT and RAHDUCT presented to opposite ears, the subject often hears PRODUCT. One of the unique aspects of phonological fusion not found in either psychoacoustic or phonetic feature fusion is that two stimuli, which are presented at the same time and contain different phonetic segments, fuse to form a new percept that is longer and linguistically more complex than either of them. Another unique aspect of this fusion is that the order in which the phonemes fuse is phonologically ruled: BARKET/LANKET yields BLANKET, not LBANKET. Note that in English, initial

stop + liquid clusters occur frequently, but initial liquid + stop clusters never occur: /b, d, g, p, t, k/ can typically precede /l, r/, but the reverse is never true at the beginning of a syllable. When these phonological constraints are lifted, fusion can occur in both directions: Thus, TASS/TACK can yield both TACKS and TASK responses (Day, 1970). Other linguistic influences on phonological fusion are discussed by Cutting and Day (1975) and Cutting (1975).

The effects of relative onset time have been explored by Day (1970), Day and Cutting (Note 5), and Cutting (1975). Their results show that phonological fusion is remarkably tolerant of differences in onset time. When no lead times are greater than 150 msec, fusion occurs readily at all leads. Even when much longer lead times are used, fusion remains frequent at all relative onsets of 150 msec and less. Factors such as whether the to-be-fused stimuli are natural or synthetic speech, whether the inputs are words or nonwords, and whether the stimuli are monosyllabic or disyllabic appear to play a role. Experiment 2 explores the fusion of the synthetic speech items /da/-/ra/ and /ba/-/la/ when relative onset times are varied. Cutting (1975) found that phonological fusion did not decrease with intensity and frequency differences between fusible stimuli of as much as 15 db. and 20 Hz, respectively. Experiments 3 and 4 extend the ranges of these differences to explore possible effects on the fused percept.

#### *Purpose of the Present Experiments*

The purpose of the experiments that follow is fivefold. First, Experiment 1 is designed to demonstrate that psychoacoustic fusion is a separate phenomenon resulting from the perceptual averaging of acoustic features. Second, Experiments 2 through 4 are designed to replicate the results found by previous studies (Table 1), using, as much as possible, the same stimulus or percept for each (/da/) and the same group of listeners. Third, those same experiments are designed to fill in the data gaps for psychoacoustic and spectral/temporal fusion and to confirm the estimated data for phonetic feature fusion. Fourth, Experiment 5 and the discussion that follows it are directed at the interactions of different

fusions. And fifth, from the results of all the studies, the different fusions are considered with respect to the types of mechanisms that must exist at different processing levels and to their relevance in speech perception.

#### EXPERIMENT 1: DEMONSTRATION OF PSYCHOACOUSTIC FUSION

Of the six fusions, least is known about psychoacoustic fusion. Its existence is gleaned from a single table presented by Halwes (1969), and he does not discuss this particular phenomenon. Although several types of different synthetic syllable pairs can yield a single percept *ambiguous* between the two dichotic items—/ba/-/ma/, /ra/-/la/, /pa/-/ba/, to name a few—it may be only the pair /ba/-/ga/ that will frequently yield a percept *different* from either of the two inputs, /da/. What causes such fusion responses? Two hypotheses appear tenable. First, and supporting the notion that this fusion is psychoacoustic, the listener may hear /da/ simply because the acoustic average of the second-formant transitions for /ba/ and /ga/ happens to fall in the middle of the /da/ range. A second, alternative view is that the perceptual averaging may be more abstract. Perhaps linguistic information is extracted from the dichotic syllables with respect to place of articulation (see Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967): /b/ is labial, /g/ is velar, and the articulatory mean between the two gestures is close to the alveolar /d/. These two hypotheses have different predictions about what happens to the fused percept when acoustic variation takes place within the /ba/ and /ga/ inputs. The first hypothesis predicts that the percentage of /da/ responses will vary according to the acoustic structure of the inputs; the second hypothesis, on the other hand, predicts no change in the number of /da/ responses, since all inputs are good exemplars of /ba/ and /ga/, and /da/ is always an articulatory mean between the two.

#### Method

Four stimuli were generated on the Haskins Laboratories parallel-resonance synthesizer. Two were /ba/ and two were /ga/. All four were two-formant, 300-msec stimuli with a constant fundamental fre-

quency of 100 Hz. First formants were centered at 740 Hz, and second formants at 1,620 Hz. First-formant transitions were 50 msec in duration, rising in frequency, and identical for all four items. Second-formant transitions were 70 msec in duration and varied in slope and direction as shown in Figure 2.<sup>2</sup> The two stimuli nearest the /da/ boundaries are called  $b^1$  and  $g^1$  (for /ba/ and /ga/, respectively), whereas the two stimuli farthest from the boundaries are called  $b^2$  and  $g^2$ . Start frequencies for the second-formant transitions were 1,232 and 1,386 Hz for the two /ba/ stimuli, and 1,996 and 2,156 Hz for the two /ga/ stimuli. Boundaries and start frequencies are based on the findings of Mattingly et al. (1971), who used very similar stimuli. Pretesting determined that all items were at least 90% identifiable as the appropriate /b/ or /g/. Items were digitized and stored on disk file for the preparation of dichotic tapes (Cooper & Mattingly, 1969).

Four dichotic pairs were assembled:  $b^1$ - $g^1$ ,  $b^2$ - $g^2$ ,  $b^1$ - $g^2$ , and  $b^2$ - $g^1$ . Each of these pairs was repeated 10 times in a random sequence, with 3 sec between pairs and with channel assignments properly counterbalanced. Stimuli were reconverted into analog form at the time of recording the test tape. Ten Wesleyan University undergraduates participated in the task as part of a course project; four others had been eliminated because they did not identify the stimuli as desired. Each was a native American English speaker with no history of hearing difficulty, no experience at dichotic listening and limited experience with synthetic speech. The tape was played on a Crown CX-822 tape recorder, and signals were sent through a set of attenuators to matched Telephonics headphones (Model TDH39). Stimuli were presented at approximately 80 db. (SPL). Earphone-to-ear assignments were counterbalanced across listeners. They wrote down B, D, or G to identify the item that they heard, and they remained uninformed that only two items were actually presented (note that no /da/ items were presented). Except when noted otherwise, the procedure and apparatus of Experiment 1 were used for all the other experiments.

#### Results and Discussion

The percentages of /b/, /d/, and /g/ responses for the four types of dichotic pairs are shown in Table 2. The largest number of /da/ fusions occurred for the  $b^1$ - $g^1$  pair, and the fewest for the  $b^2$ - $g^2$  pair; intermediate fusion scores occurred for the other

<sup>2</sup> These transitions are longer than those typically found in synthetic syllables, but preliminary testing suggested that longer transitions facilitate psychoacoustic fusion. For other effects with longer transitions see Tallal and Piercy (1974, 1975), whose data support the notion that transition duration has auditory consequences independent of phonemic consequences.

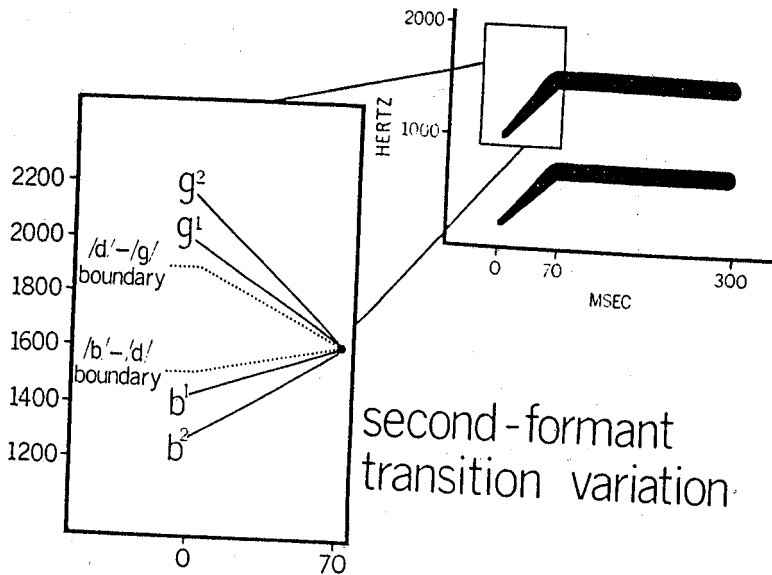


FIGURE 2. Schematic spectrogram and representations of four stimuli, two /ba/s and two /da/s, used in Experiment 1 for psychoacoustic fusion.

two pairs. Of paramount interest is the fact that not only did /da/ fusions decrease for the dichotic pair whose second-formant transitions were farthest from the /d/ boundaries, but *both* /ba/ and /ga/ responses increased as well. The difference between the number of /da/ responses for  $b^1-g^1$  and  $b^2-g^2$  pairs proved significant by a Wilcoxon matched-pair signed-ranks test,  $T(8) = 1.5$ ,  $p < .02$  (two-tailed): This demonstrates that the proximity of the second-formant transitions in the to-be-fused pair to the /da/ boundaries is important to the phenomenon. Thus, simple averaging of formant transitions is insufficient to explain consistent psychoacoustic fusions; it is the averaging of optimally similar transitions that is crucial. This result suggests a

close relationship between rivalry and fusion, as noted earlier.<sup>3</sup>

The responses for the other two dichotic pairs were predictable from the first results. The  $b^1-g^2$  pair had an intermediate number of /da/ fusions and a smaller number of /ba/ responses. The  $b^2-g^1$  pair also had an intermediate number of /da/ fusions but had an increase in the /ba/ responses, largely at the cost of /ga/. Subtracting the number of /ga/ responses from the number of /ba/ responses for each subject, this shift pattern is statistically robust,  $T(7) = 0$ ,  $p < .02$ . It appears that this fusion is psychoacoustic rather than psycholinguistic, since it is quite sensitive to phonemically irrelevant acoustic variation in the stimuli.<sup>4</sup>

TABLE 2  
PERCENTAGES OF /b/, /d/, AND /g/ RESPONSES FOR  
FOUR PAIRINGS OF DICHOTIC STIMULI IN EXPERI-  
MENT 1 (PSYCHOACOUSTIC FUSION)

Pair	Response		
	/b/	/d/	/g/
$b^1-g^1$	9	56	35
$b^2-g^2$	15	24	61
$b^1-g^2$	5	32	63
$b^2-g^1$	21	47	32

<sup>3</sup>Fusion and rivalry are clearly not exclusive alternatives but interact in a probabilistic fashion. In Experiment 1, as in all experiments presented here, I have tried to maximize the probability of fusion in most conditions. The terms fusion and rivalry as used here are intended to parallel their use by Julesz (1971, p. 23). Whether there is a suppression of rivalry by fusion mechanisms in audition like that proposed in vision (Julesz, 1971, p. 218-220; Kaufman, 1963, 1974) is not known, and is beyond the scope of the present investigation.

<sup>4</sup>In a study made independently at the time the present investigation was being prepared for publication, Repp (1975d) found essentially the same re-

A phenomenon similar to psychoacoustic fusion in certain respects is the "feature sharing effect" as formulated by Studdert-Kennedy et al. (1972) and explored parametrically by Pisoni and McNabb (1974). One emphasis of Pisoni and McNabb, like that of Rand (1974) for another type of fusion, is on masking; here, of course, the emphasis is on fusion. Pisoni and McNabb found that, given a target syllable such as /ba/ presented to one ear and a mask such as /ga/ presented to the other ear, listeners made few errors identifying the target regardless of the interval between the onsets of the target and mask. They found that for /b-/g/ pairs it made no difference whether the target and mask shared the same vowel or had different vowels, but they did note that "the more similar the vowels of the two syllables, the more likely they are to 'fuse' or integrate into one perceptual unit so that the listener had difficulty assigning the correct auditory features to the appropriate stimulus" (Pisoni & McNabb, 1974, p. 357). The fusion they allude to is most likely the psychoacoustic fusion reported here. Given a /ba-/ga/ target-mask pair, however, they found few /da/ responses to occur even when the items had simultaneous onsets. The reason for this result most likely stems from subject expectations and the difference between the two procedures. For example, their listeners knew that targets would be presented to one ear and masks to the other, that two items would be presented on every trial, and that they were to report the identity of the first item; here, on the other hand, there were no targets or masks, listeners did not know that two items were presented on each trial, and they were simply to report what they heard. In Experiment 1, relative onset time was not varied as in the Pisoni and McNabb (1974) study; however, Experiment 2 varies onset time in a similar fashion. Just as I have concluded that psychoacoustic fusion occurs prior to phonetic processing, Pisoni and McNabb

(1974) concluded that the feature-sharing effect is also prephonetic.

The present experiment confirms the existence of psychoacoustic fusion and also strongly suggests that its nature is not linguistic but rather a perceptual integration of acoustic features. The major thrust of this paper, however, stems from those experiments that follow—replications and explorations of the effects of varying relative onset time, relative intensity, and relative frequency on the six fusions outlined previously.

### *Experiments 2-5: General Methodology*

#### *Overview*

Sixteen different brief experiments were conducted to study the six fusions. All dealt with the syllable /da/, either as a stimulus or as a potential percept, as shown schematically in Figure 1. In general, three experiments were directed at each type of fusion: In one, the relative onset time of the dichotic stimuli was varied; in a second, the relative intensity was varied; and in a third, the relative fundamental frequency was varied. For simplicity's sake, rather than numbering each separate demonstration as an experiment, all those dealing with relative onset time are considered as part of Experiment 2, those dealing with relative intensity as Experiment 3, and those dealing with relative frequency as Experiment 4. Experiment 5 deals with the interaction of different fusions.

The same 10 listeners that participated in Experiment 1 participated in these experiments. Because of the great number of separate studies, counterbalancing of test order was not attempted; instead, all subjects listened first to the three tests pertaining to phonological fusion, and then to those pertaining to sound localization, psychoacoustic fusion, spectral fusion, spectral/temporal fusion, and phonetic feature fusion, respectively.

#### *Stimuli*

Six speech syllables were generated in several renditions using the Haskins Laboratories' parallel-resonance synthesizer: /ba, da, ga, ta, la, ra/. The /ba/ and /ga/ stimuli were the b<sup>1</sup> and g<sup>1</sup> items used in Experiment 1. All stimuli were 300 msec in dura-

---

results as those reported here, using /bae-/gae/ dichotic pairs. From the results of several experiments on psychoacoustic fusion, he reaches many of the same (but some different) conclusions that I reach.

ion and shared the same /a/ vowel used previously. The stop-consonant stimuli (those beginning with /b/, /d/, /g/, and /t/) consisted of two formants, and the liquid stimuli (those beginning with /l/ and /r/) consisted of three formants. For the stop stimuli, first- and second-formant transitions were 50 and 70 msec in duration, respectively. Start frequency of the second-formant transitions for /da/ and /ta/ was 1,695 Hz. All voiced stops (/b/, /d/, and /g/) had a voice-onset time value of 0 msec, while the voiceless stop (/t/) was aspirated with a voice-onset time of +70 msec (see Lisker & Abramson, 1964). Liquid items began with 50 msec of steady-state resonance in all formants, followed by 100 msec of transitions in the second and third formants (only 20 msec in the first formant), followed by the vowel resonances. An open-response pretest showed that each item was identified correctly on at least 86% of the trials.

The *standard* forms of all these items had a pitch of 100 Hz (like that of an adult male) and an intensity of 80 db. (SPL). *Nonstandard* forms were generated with frequencies of 102, 120, and 180 Hz and with intensities ranging downward to 40 db. in 5 db. steps. For spectral fusion and for spectral/temporal fusion, /ba/, /da/, and /ga/ were also parsed into separate parts as shown in Figure 1: For spectral fusion, items were generated as separate formants, and for spectral/temporal fusion, the second-formant transition was isolated from the remainder of the syllable. All stimuli were digitized and stored on computer disk file for the preparation of dichotic tapes. Except as noted below, procedures were identical to those of Experiment 1.

## EXPERIMENT 2: RELATIVE ONSET TIME IN SIX FUSIONS

### Method

Six different randomly ordered sequences of dichotic pairs were recorded, one for each type of dichotic fusion. Relative onset times were chosen with regard to the temporal range for which pretesting had determined each fusion most sensitive.

*Sound localization.* Tokens of the standard form of the stimulus /da/ (100 Hz and 80 db.) were recorded on both channels of audio tape. The items could have synchronous onsets (0 msec lead time) or asynchronous onsets. Ten asynchronous onsets were selected: 1, 2, 3, 4, 5, 10, 20, 40, 80, and 160

msec lead times. A sequence of 24 items was recorded: (10 asynchronous leads)  $\times$  (2 lead-time configurations, Channel A leading Channel B and vice versa) + (4 simultaneous-onset pairs). Listeners were told to write down the number of items they heard, one or two, paying no special regard to the identity of the stimuli.

*Psychoacoustic fusion.* Tokens of the standard /ba/ were recorded on one channel and tokens of the standard /ga/ on the other. Nine lead times were selected: 0, 1, 2, 5, 10, 20, 40, 80, and 160 msec. A sequence of 36 dichotic pairs was recorded: (9 leads)  $\times$  (2 lead-time configurations)  $\times$  (2 channel assignments, /ba/ to Channel A and /ga/ to Channel B and vice versa). Listeners were instructed to write down the initial consonant of the syllable that they heard most clearly, choosing from among the voiced stops B, D, or G. Note that no /da/ stimuli were actually presented.

*Spectral fusion.* The first formants of the standard items /ba/, /da/, and /ga/ were recorded on one channel and the second formants on the other. Six lead times were selected: 0, 10, 20, 40, 80, and 160 msec. A sequence of 72 items were recorded: (3 stimulus pairs, first and second formants for /ba/, /da/, and /ga/)  $\times$  (6 lead times)  $\times$  (2 lead-time configurations)  $\times$  (2 channel assignments). Listeners wrote down the initial consonant that they heard most clearly: B, D, or G.

*Spectral/temporal fusion.* Each of the 70 msec second-formant transitions was excised from the three standard syllables /ba/, /da/, and /ga/ and recorded on one channel, and the remainder of the syllables were recorded on the other. A sequence of 72 items was recorded following the same format as the spectral fusion sequence. Again, listeners wrote down B, D, or G.

*Phonetic feature fusion.* The standard form of /ba/ was recorded on one channel, and the standard form of /ta/ on the other. Seven lead times were selected: 0, 5, 10, 20, 40, 80, and 160 msec. A sequence of 84 items was recorded: (7 leads)  $\times$  (2 lead-time configurations)  $\times$  (2 channel assignments)  $\times$  (3 observations per pair). Listeners wrote down the initial consonant of the syllable that they heard most clearly, choosing from among B, D, P, and T. Note that no /da/ or /pa/ stimuli were actually presented.

*Phonological fusion.* Two types of dichotic pairs were recorded on opposite channels: /ba/ and /la/, and /da/ and /ra/, all of standard form. Five leads were selected: 0, 20, 40, 80, and 160 msec. A sequence of 40 items were recorded: (2 fusible dichotic pairs)  $\times$  (5 leads)  $\times$  (2 lead-time configurations)  $\times$  (2 channel assignments). Listeners were instructed to write down whatever they heard, following Day (1968).

### Results and Discussion

Relative onset time is a crucial variable for all six fusions, as shown in Figure 3. In general, the greater the interval between onsets

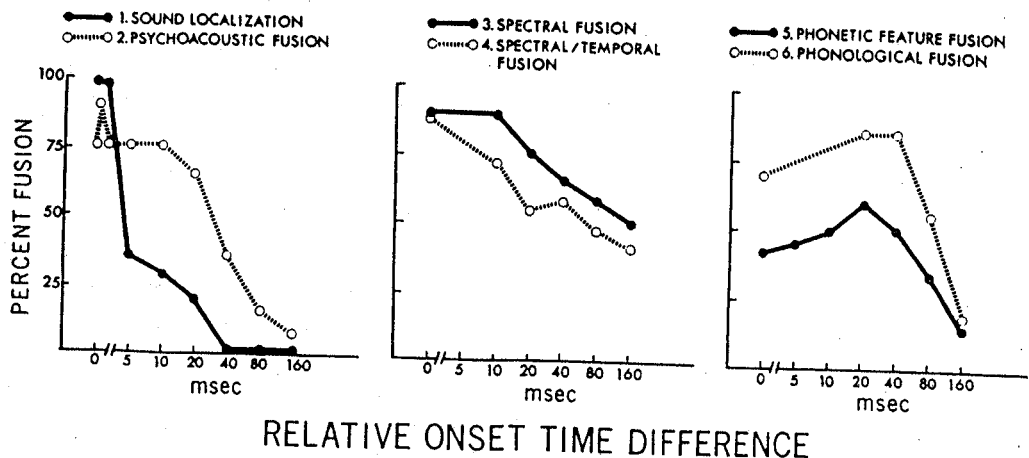


FIGURE 3. The effect of relative onset time in six fusions (data from Experiment 2).

of members of the dichotic pair, the less frequently fusion occurs—or, inversely, the greater the probability of the perceptual disintegration of the fused percept. All results are discussed with regard to that relative onset time at which fusions first occur significantly less frequently than for simultaneous onset pairs, as measured by a Wilcoxon matched-pair signed-ranks test ( $p < .05$ ).

For sound localization, the fusions, or one-item responses, decreased precipitously with relative onset times as small as 4 and 5 msec. (In the left-hand panel of Figure 3, onset differences of 2 and 3 msec are combined, as well as those of 4 and 5 msec, because there was little difference between them.) This estimate is slightly smaller than that given by Cherry and Taylor (1954), but it should be noted that some subjects continued to report hearing only one item for onset asynchronies as great as 20 msec. No one-item responses were given for relative onset times of 40 msec or greater. In the same panel, the data for psychoacoustic fusion present a different pattern. Fusions—that is, D responses, given a /ba/-/ga/ dichotic pair—occurred readily for all trials with relative onset times of up to 20 msec but dropped considerably on trials with greater asynchronies. There was no significant effect of the order in which the two stimuli arrived: /ba/-leading-/ga/ and /ga/-leading-/ba/ trials yielded equally frequent /da/ responses.

The pattern of fused responses for spectral fusion and spectral/temporal fusion (B, D,

or G for the components of /ba, da, ga/, respectively) were markedly parallel. They are shown in the central panel of Figure 3. Significant decreases in the probabilities of fusion responses occurred by 40 msec for spectral fusion and by 20 msec for spectral/temporal fusion. In both, the results presented are those summed over /ba/, /da/, and /ga/ items. Combined, these functions asymptote at slightly greater than 33% performance. Listeners chose from among three responses, one of which had to be correct (unlike psychoacoustic fusion where there were no /da/ items presented). Moreover, the first formant alone sounded somewhat like /ba/ itself: Fully 85% of all responses for both fusions at 160-msec asynchronies were /ba/. This fact contributes to the relatively high asymptote of the combined functions. There was no significant effect of the order of arrival of the dichotic items for either fusion, but B and D fusions were more frequent than G fusions. This result is considered in the general discussion.

Phonetic feature fusion and phonological fusion presented patterns of fused responses not seen for the other types of fusion. In both, fusions were slightly more frequent at brief relative onset times than at simultaneous onsets or at longer relative onsets. For phonetic feature fusion this effect was significant: The 20-msec asynchrony yielded more fused responses than asynchronies of 0, 80, and 160 msec. For phonological fusion, however, the increase was not significant. In phonetic

feature fusion, fused responses were equally frequent regardless of whether /ba/ began before /ta/, or vice versa. For all asynchronies, /pa/ fusions were more frequent than /da/ fusions: The overall ratio was 4:1. In phonological fusion, the nonlinear pattern of fusion responses is slightly complicated by the fact that when, for example, /da/ begins before /ra/, fusions are more frequent than when /ra/ begins before /da/. Whereas the effect was not significant here, the trend is similar to that found by Cutting (1975, Experiment 3; but see Day, 1970; Day & Cutting, Note 5). Fused responses were more frequent for /ba/-/la/ pairs than for /da/-/ra/ pairs, a finding which replicates the stop + /l/ and stop + /r/ findings of Day (1968) and Cutting (1975). In addition, /da/-/ra/ pairs yielded many anomalous fusions; that is, in addition to the cases where listeners wrote down DRA, they also wrote occasional DLA, BLA, and GLA responses. See Day (1968), Cutting (1975), and Cutting and Day (1975) for more specific accounts of this effect.

In general, the results for the six types of fusion fall into three groups when relative onset time is varied. Group 1 consists of sound localization alone, in which the fused percept disintegrates after asynchronies of only 5 msec. Group 2 consists of psychoacoustic fusion, spectral fusion, and spectral/temporal fusion. When corrections are made for differences in lower asymptote, these three functions are much the same: After asynchronies of 20 and 40 msec, the frequency of fusions tapers off rapidly. Phonetic feature fusion and phonological fusion form Group 3, in which fusion increases with lead times of 20 to 40 msec before decreasing with longer lead times.

Consider now the stimulus "units" that appear to be fused in each of these three groups. Sound localization, the only member of Group 1, occurs through cross-correlation of opposite-ear waveforms. In the present study the two signals were speech sounds carried on a glottal source of 100 Hz. Each glottal pulse was thus 10 msec in duration and served as a convenient acoustic "unit" to anchor the cross-correlation process. Microstructure differences between contiguous glottal pulses were slight but may have served as an aid in the localization process (Leakey, Sayers, &

Cherry, 1958). Onset asynchronies less than 5 msec are apparently surmountable in sound localization, whereas larger differences generally are not. It may be that the binaural hearing mechanisms can integrate glottal pulses conveying identical information if they arrive within 5 msec of one another, in part, because each pulse overlaps by at least half a glottal wavelength. With onset times greater than 5 msec, the opposite-ear pulses that arrive most nearly at the same time are not identical, and microstructure differences may impede localization. From this account one would predict that a long, continuous steady-state vowel /a/ with identical microstructures within each glottal pulse would always be localizable as a single item regardless of timing differences. Indeed, Broadbent (1955) and Sayers and Cherry (1957) performed demonstrations similar to this.

Group 2 has three members: psychoacoustic fusion, spectral fusion, and spectral/temporal fusion. In psychoacoustic fusion, the second-formant transitions of the opposite-ear stimuli appear to be the "unit" of fusion. Each is 70 msec in duration, and onset time differences of 20 to 40 msec are needed before the fused percept disintegrates. Again, the tolerable onset time asynchrony appears to be about "half the fused unit," although for this second-level type of fusion that unit is several times larger. In psychoacoustic fusion there is competition of opposite-ear information. If this information is not meshed temporally in the right fashion, rivalry will occur and backward masking is the typical result (see Massaro, 1972, 1974). In spectral fusion and in spectral/temporal fusion, on the other hand, there is no competing information. In both, the "units" are the second-formant transition properly aligned with the first-formant transition to yield the percepts /b/, /d/, or /g/. Again, transition durations are the same and, allowing for differences in asymptote, disintegration of the percept appears to occur at about the same point. Here, information about place of articulation from one ear (either as part of the second-formant resonance or as an isolated chirp) is combined with the information about manner of production in the first-formant transition. In all three fusions, however, the actual fusion

appears to be only incidentally linguistic. Transitions of formants are compared and merged, and subsequent analysis of the fused information reveals it to be linguistically labelable as /b/, /d/, or /g/.

Group 3 consists of phonetic feature fusion and phonological fusion. In both, the "units" are linguistic, but in the first, the units are phonetic features and in the second, they are the phonemes themselves. The increase in fusion at short lead times followed by a decrease at longer lead times separates these fusions from the others. Phonemes and their composite phonetic features can have acoustic manifestations of 20 to 150 msec, depending on the phoneme and on the feature. Thus, it seems to make less sense to talk in terms of "half a unit's duration" as the threshold for the disintegration of the fused percept. Instead, it makes more sense to speak in terms of the time course of processing those features. In an inverted form, the functions in the right-hand panel of Figure 3 look like J-shaped backward masking functions found in vision (Turvey, Note 6). As noted previously, disruption by masking can be thought of as a reciprocal process to fusion. Processing of linguistic features appears to be disrupted most readily after initial processing of the first-arriving item has taken place; earlier or later arrival of the second item decreases the chance of such interference. The disruption process allows for the possible misassignment of the feature values in the case of phonetic feature fusion; in phonological fusion, the disruption allows for the possible combination—and, in the case of /ra/-leading-/da/ items, the misassignment of temporal order—of the phonemes themselves.

These three patterns of results, one for each group, suggest at least three types of analysis relevant to speech perception. Each has its own temporal limit within which fusion occurs, and this limit can be thought of as analogous to different types of perceptual "moments" (see Allport, 1968; Efron, 1970). The smallest type of moment lasts up to about 2 to 5 msec, within which time the waveforms of stimuli presented to opposite ears can be meshed (localized). To exceed this limit is to exceed the resolving capacity of the mechanisms involved. An intermediate-

sized moment lasts up to perhaps 40 msec and allows the acoustic features of opposing stimuli to merge. Again, to go beyond this range is, generally, to go beyond the system's ability to process (fuse) the discrepancy in stimulus information. A third moment lasts 20 to 80 msec, a time limit which provides maximal opportunity for misassigning certain linguistic feature values of competing inputs.

It seems likely that these three types of moments reflect processes that occur concurrently and become relevant to the percept according to variations in the stimuli and in the demands placed on the listener in a particular task. It also seems likely that the different sizes of the moments reflect the level at which fusion occurs: The smaller the interval, the lower in the system that the fusion occurs and, conversely, the larger the interval, the higher in the system the fusion occurs. This general scheme is an extension of that suggested by Turvey (1973), who found that peripheral processes in visual masking occurred over smaller time domains than central processes. Translating the terms *peripheral* and *central* to auditory studies of fusion cannot be straightforward, since there is considerably more pathway interaction between the two ears than between the two eyes before events reach a cortical level. Nevertheless, after substituting the more conservative terms *lower level* and *higher level* for peripheral and central, the extended analogy is worth pursuing. For example, Turvey (1973) found that in visual masking, peripheral processing was characterized by stimulus integration, and central processing, by stimulus disruption. Auditory fusions of Group 1 (sound localization) and Group 2 (psychoacoustic, spectral, and spectral/temporal fusion) are integrations of opposite-ear stimuli, whereas those of Group 3 (phonetic feature fusion and phonological fusion) appear to occur because of an interruption of speech perception in midprocess and a subsequent misassignment of features.<sup>5</sup>

<sup>5</sup> Turvey (1973) noted that, in vision, there appear to be two types of integrative processes, one integrating energies and the other integrating features. Level 1 auditory fusions would appear to be of the first variety, and Level 2 fusions, of the second.



TABLE 3

PERCENTAGES OF "FUSION" RESPONSES AS A FUNCTION OF PRESENTATION MODE WHEN ITEMS HAVE SIMULTANEOUS ONSETS AND SHARE THE SAME FREQUENCY AND INTENSITY

Fusion type and dichotic pair	Dichotic <sup>a</sup>	Binaural <sup>b</sup>
Sound localization /da/ + /da/		
Psychoacoustic fusion /ba/ + /ga/	100 /da/°	100 /da/°
Spectral fusion /da/ Formant 1 + /da/ Formant 2	68 /da/	81 /da/ <sup>d</sup>
Spectral/temporal fusion /da/ without second-formant transition + second-formant transition of /da/	85 /da/	100 /da/°
Phonetic feature fusion /ba/ + /ta/	81 /da/	100 /da/°
Phonological fusion /da/ + /ra/	43 /da/	20 /da/°
	73 /dra/ <sup>f</sup>	15 /dra/ <sup>f,g</sup>

- <sup>a</sup> Dichotic presentation occurs when Stimulus A is presented to one ear and Stimulus B to the other (regardless of whether or not the two items are identical). Data in this column are means found in Experiments 2 through 4.  
<sup>b</sup> Binaural presentation occurs, for the purposes of the present paper, when stimuli are mixed and both are presented to both ears.  
<sup>c</sup> These items are physically identical.  
<sup>d</sup> Determined by pretesting with listeners not participating in present study.  
<sup>e</sup> Computed from Halves (1969, p. 75), who used synthetic speech stimuli very similar to those used in the present studies.  
<sup>f</sup> The response /dra/ here represents all fusion responses for the stimulus pair /da/ + /ra/, including those in which a different stop consonant or a different liquid was reported. See Day (1968), Cutting & Day (1975), and Cutting (1975) for further details.  
<sup>g</sup> Cutting (1975, Experiment 2), using synthetic speech stimuli similar to those used here.

#### An Additional Consideration: Presentation Mode

If lower level fusions are those characterized by perceptual integration, and higher level fusions, by perceptual disruption, presentation mode ought to have a crucial effect on the probability of fused responses and should provide a test of the distinction. Two modes of presentation are of interest: dichotic presentation, the mode used in all experiments in this paper, and a type of binaural presentation. I use the term *dichotic* and *binaural* slightly differently than do Woodworth (1938, p. 526) and Licklider (1951, p. 1026). They view dichotic listening as a special form of binaural listening. Dichotic presentation, as the term is used here, occurs when Stimulus A is presented to one ear and Stimulus B to the other ear, *regardless* of whether or not the two items are identical. Binaural presentation, on the other hand, is defined here as occurring when Stimuli A and B are combined and both items are presented to both ears. Table 3 summarizes the percentages of fusion responses for the six fusions as a function of presentation mode. The dichotic scores are means from the present set of experiments, whereas the binaural scores stem from logical considerations and from data collected elsewhere.

Fusions of Group 1 (sound localization) and Group 2 (psychoacoustic, spectral, and spectral/temporal fusion) reveal a pattern distinctively different from those of Group 3 (phonetic feature fusion and phonological fusion.) For the first four fusions, the number of /da/ responses under binaural conditions is slightly greater than or equal to the number of dichotic fusions. For the other two, however, binaural fusions are considerably less frequent than dichotic fusions. The first four could logically occur at a neural level prior to the cortex, or at least prior to linguistic analysis within the cortex. To a great degree, presentation mode is irrelevant here, and acoustic combination may be similar to neural combination (integration)—a straightforward, primarily additive process. The other two fusions, on the other hand, must occur subsequent to linguistic (and cortical) analysis, since mere mixing of the signals inhibits rather than aids in obtaining the desired percept. This occurs, presumably, because mixing the signals degrades their separate integrities, and the stimuli mask each other effectively before they ever arrive at some central locus for linguistic analysis. Disruption, then, never has a chance to occur because integration has already occurred. In summary, the first four fusions appear to be only incidentally lin-

guistic, whereas the last two are necessarily linguistic.

In addition to providing a framework for the data discussed above, the upper level/lower level scheme, as adapted from Turvey (1973) would predict effects of stimulus energy on fusions at these different levels. Turvey found that stimulus energy affected visual masking at a peripheral level but that it had essentially no effect centrally. Analogously, in auditory fusion, one would predict that stimulus energy (intensity) would have a relatively large effect on lower level fusions and a smaller effect on higher level fusions. Experiment 3 was designed to test these predictions.

### EXPERIMENT 3: RELATIVE INTENSITY IN SIX FUSIONS

#### Method

Three different randomly ordered sequences of dichotic pairs were recorded in a fashion similar to that of Experiment 2, one each for psychoacoustic fusion, phonetic feature fusion, and phonological fusion. No sequences were prepared for the others, since data on these fusions are either irrelevant (sound localization) or readily obtainable elsewhere (spectral fusion and spectral/temporal fusion). Nine relative intensities were used in all three sequences: One stimulus was always the standard 80-db. (SPL) item, and the other decreased in intensity by 0, 5, 10, 15, 20, 25, 30, 35, or 40 db. Sequences for psychoacoustic and phonetic feature fusions consisted of 36 dichotic pairs: (9 relative intensities)  $\times$  (2 intensity configurations, an 80-db. stimulus on Channel A or on Channel B)  $\times$  (2 channel assign-

ments). Again, /ba/ and /ga/ were used for psychoacoustic fusion and /ba/ and /ta/ for phonetic feature fusion. The sequence for phonological fusion was exactly twice as long, allowing for the two fusible pairs /ba/-/la/ and /da/-/ra/. Listeners followed the same instructions for each fusion as in Experiment 2.

#### Results and Discussion

Patterns of results for the three types of fusion in question are shown in Figure 4, along with results for the other fusions adapted from other sources. Again, results are discussed in terms of the relative intensity levels at which significant decreases in fusion first occur, using the same criterion as in Experiment 2.

For psychoacoustic fusion, a significant decrease occurred with a drop of only 10 db. in either stimulus, /ba/ or /ga/ (but see Repp, 1975d). The number of /da/ fusions tapered off at a decreased rate thereafter. For both phonetic feature fusion and phonological fusion, on the other hand, significant decreases first occurred at 30 db. In all three types of fusion, there was no significant effect of *which* stimulus in the fusible pair was the most intense. The data plotted for sound localization are hypothetical, but since the fused percept never disintegrates and since binaural interactions can occur over intensity differences greater than 40 db., a straight line at 100% is drawn. The data for spectral fusion and for "spectral/temporal" fusion are

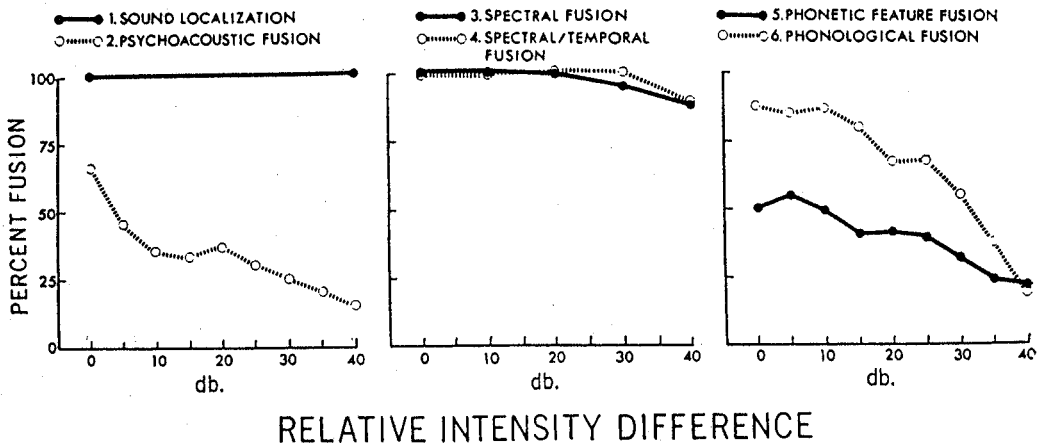


FIGURE 4. The effect of relative intensity in six fusions (data from Experiment 3 for Fusions 2, 5, and 6; from Rand (1974) for Fusions 3 and 4; and from logical considerations for Fusion 1).

adapted from Rand (1974), who used stimuli virtually identical to those in the present studies (except his consisted of three formants rather than two).

For a minor replication of the results of Experiment 2, compare certain aspects of Figures 3 and 4. Notice that the first data point of each function in Figure 3 (0-msec lead) represents exactly the same pairs as those in Figure 4 (0 db.). Note further that the probability of fused responses for each fusion at these points is comparable in each study.

Before pursuing the scheme of higher and lower level fusions, it is necessary first to reconsider the distinction between those fusions whose stimuli "compete" with one another for the same processor and those whose stimuli do not. Competition is here defined as a situation conducive to substantive information loss or alteration. Consider the fusions without competition first: sound localization (Group 1), spectral fusion, and spectral/temporal fusion (both from Group 2). There is no information loss or alteration in sound localization because the two inputs are identical, and the percept can change only in its perceived locus, not in its identity. There is no information loss in either spectral fusion or spectral/temporal fusion because the acoustic information of opposite-ear stimuli is simply restructured back into the original form from component parts in a straightforward manner.

In the other three fusions, however, there is competition. In psychoacoustic fusion (the only other member of Group 2), the second-formant transitions are in the same general space-time region, and as demonstrated in Experiment 1, they appear to contribute best to the fused percept when they are closest together, which apparently enables them to be perceived as a single-formant transition. Information is lost, in that /ba/ and /ga/ are no longer heard, and information is altered, because both items contribute to the percept /da/. In phonetic feature fusion (Group 3), the stimuli compete for the same limited-channel-capacity linguistic processor in the left hemisphere (see Studdert-Kennedy et al., 1970, 1972). Of two values of the voicing

feature and two values of a place-of-articulation feature, only one value of each can typically be processed, while the other is often lost. If they did not originally belong to the same stimulus, a fusion has occurred. Finally, in phonological fusion (Group 3), the stimuli may be processed in opposite hemispheres (Cutting, 1973) with no information loss, but there is considerable information alteration, since the two inputs are combined in the most phonologically reasonable fashion to form a single phoneme string. A more detailed comparison of mechanisms thought to underlie the six fusions is given in the concluding discussion.

Setting aside those fusions in which there is no competition, and hence, essentially no effect of intensity with attenuations of as much as 40 db., one finds that the results of the other three fusions support the higher level/lower level distinction discussed earlier. Psychoacoustic fusion, a lower level process characterized by perceptual integration, was quite sensitive to relatively smaller attenuations of intensity. On the other hand, phonetic feature fusion and phonological fusion, higher level processes characterized by perceptual disruption, were relatively insensitive to intensity variation.

The results of Experiments 2 and 3 and the additional consideration of presentation mode provide clear evidence for the distinction between auditory processes of Group 2 (psychoacoustic, spectral, and spectral/temporal fusion) and linguistic processes of Group 3 (phonetic feature fusion and phonological fusion). This distinction is similar to that made by Studdert-Kennedy et al. (1971), Pisoni (1973), and Wood (1975), among many others, using very different paradigms. However, evidence thus far for the distinction between Group 1 (sound localization) and Group 2 is less impressive (seen only in the left and center panels of Figure 3). Experiments 4 and 5 are directed at supporting this distinction.

#### EXPERIMENT 4: RELATIVE FUNDAMENTAL FREQUENCY IN SIX FUSIONS

##### *Method*

Six different randomly ordered sequences of dichotic pairs were recorded, one for each type of

fusion. Four fundamentals were selected: the standard frequency of 100 Hz, and three others—102, 120, and 180 Hz. Pairs always consisted of one stimulus at 100 Hz and another stimulus at any of the four possible fundamental frequencies, yielding relative frequency differences of 0, 2, 20, and 80 Hz. For the sound localization sequence, /da/ was recorded on both channels in a 16-pair sequence: (4 relative fundamentals)  $\times$  (2 frequency configurations, the 100-Hz item on Channel A or on Channel B)  $\times$  (2 observations per pair). For the psychoacoustic fusion sequence, /ba-/ga/ pairs were recorded in a 32-pair sequence: (4 relative frequencies)  $\times$  (2 frequency configurations)  $\times$  (2 channel assignments)  $\times$  (2 observations per pair).

For both spectral fusion and spectral/temporal fusion, 24-item sequences were recorded. In spectral fusion, the first formant was always held constant at 100 Hz, and in spectral/temporal fusion, the first formant and the steady-state segment of the second formant were also always at 100 Hz: Frequency variation always took place in the second formant or second-formant transition. The dichotic pairs could yield /ba/, /da/, or /ga/ responses. The 24-item sequence consisted of (3 stimulus pairs, those for /ba/, /da/, and /ga/)  $\times$  (4 relative frequencies)  $\times$  (2 frequency configurations). Channel assignments were randomized across pairs.

For phonetic feature fusion, /ba-/ta/ pairs were recorded in a 32-pair sequence following the same format as the psychoacoustic fusion sequence; for phonological fusion, a similar 32-pair sequence was recorded for /ba-/la/ and /da-/ra/ pairs. Again, listeners followed the same instructions for each fusion as in Experiment 2.

### Results and Preliminary Discussion

As shown in Figure 5, frequency differences affected only one type of fusion: sound localization. Fusions, the number of one-item re-

sponses, plummeted from nearly 100% for identically pitched pairs to nearly 0% for pairs with only 2-Hz differences between members. Frequency differences had no significant effect on any of the other five types of fusion. Note again that the fusion probabilities for pairs with 0-Hz differences are similar to the standard pairs in Experiments 2 and 3.

The results suggest a clear distinction between the fusions of Groups 2 and 3 and the sound localizations of Group 1. A problem arises, however, when one considers that these groups correlate perfectly with the type of response required of the listener. In sound localization, the listener reports whether he or she heard one item or two; in all other fusions the listener identifies the item heard. It may be that when frequency varies in these other fusions, the listener can easily report whether one or two items were actually presented but, since a linguistic response is required, may ignore the cues of numerosity. Experiment 5 investigates this possibility.

### EXPERIMENT 5: HOW MANY ITEMS ARE HEARD?

It is clear that sound localization is a very different kind of fusion than the other five, both phenomenologically and in terms of the results of Experiments 2 and 4. In sound localization, the items presented to opposite ears are either integrated into a single percept or they are not, and the identity of the inputs

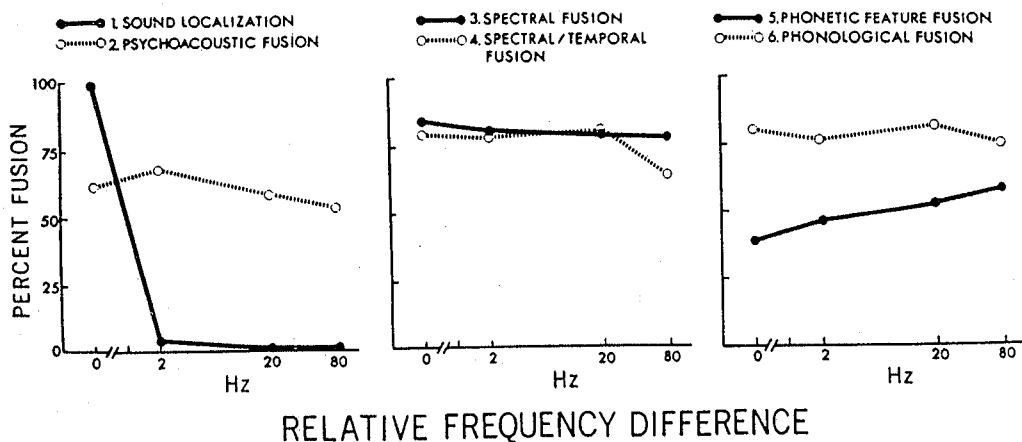


FIGURE 5. The effect of relative fundamental frequency in six fusions (data from Experiment 4).

matters not at all. In all the other fusions, by contrast, it may be possible for the listener to be aware that more than one item is presented on a given trial but find that the fusion response is the best label for what he hears. Differences of only 2 Hz convince the listener of the presence of two different items in sound localization. Is this true of other fusions as well? In other words, is it necessary that the items fuse into a single acoustic percept for them to be labeled and judged as a single linguistic percept?

### Method

One pair of stimuli was chosen to represent each of the six fusions. These pairs are shown schematically in Figure 1. Pairs were either of the standard form (both items at 100 Hz), or they differed by 2 Hz (one item at 100 Hz and the other at 102 Hz). The standard (80 db.) intensities were used. A sequence of 48 simultaneous-onset pairs was recorded: (6 pairs, one for each fusion)  $\times$  (2 relative frequencies, 0- or 2-Hz difference)  $\times$  (2 channel arrangements)  $\times$  (2 observations per pair). Twenty listeners, 10 from the previous experiments and 10 others selected according to the same criteria, wrote down 1 or 2, indicating the number of items that they heard on each trial. No practice was given.

### Results and Discussion

The results for all six types of fusion are shown in Table 4. For four types of fusions

TABLE 4  
PERCENTAGES OF ONE-ITEM RESPONSES GIVEN TO  
DICHOTIC PAIRS IN EXPERIMENT 5

Fusion type and dichotic pair	Stimulus condition	
	Both items at 100 Hz	One item at 100 Hz, other at 102 Hz
Sound localization /da/ + /da/	99	9 <sup>a</sup>
Psychoacoustic fusion /ba/ + /ga/	78	10 <sup>a</sup>
Spectral fusion /da/ Formant 1 + /da/ Formant 2	60	2 <sup>a</sup>
Spectral/temporal fusion /da/ without second-formant transition + second-formant transition of /da/	15	12
Phonetic feature fusion /ba/ + /ta/	22 <sup>b</sup>	8
Phonological fusion /da/ + /ra/	46	5 <sup>a</sup>

<sup>a</sup> Across the two stimulus conditions, differences are significant,  $p < .01$ , by Wilcoxon matched-pairs signed-ranks test.

<sup>b</sup> Halves (1969), using similar stimuli, found that listeners reported hearing only one sound when both items shared the same pitch. The difference between his results and those of the present study may be attributable to a contrast effect here induced by mixed presentation of fusible pairs. Halves blocked pairs of same and different pitches.

the number of one-item responses dropped significantly when fundamental frequency varied: sound localization (90% decrease), psychoacoustic fusion (68%), spectral fusion (58%), and phonological fusion (41%). In the other two types of fusion the decreases were considerably smaller: only 3% for spectral/temporal fusion and 14% for phonetic feature fusion.

Figure 5 and Table 4 demonstrate conclusively that sound localization is different from the other fusions. In the other five fusions the number of items perceived plays no role in the linguistic identity of the fused percepts: In the present experiment, standard pairs may be perceived (a) as a single item most of the time (as in psychoacoustic fusion); (b) as a single item about half of the time (as in spectral fusion and phonological fusion); or (c) nearly always as two items (as in spectral/temporal fusion or phonetic feature fusion). The 2-Hz difference (nonstandard) pairs fuse as readily as the standard pairs, yet all nonstandard pairs are perceived as two-item presentations.

In view of the initial formulation of six different fusions in dichotic listening, the most important result here is that spectral fusion and spectral/temporal fusion differ significantly in the number of items perceived when the to-be-fused stimuli have the same pitch,  $T(15) = 0$ ,  $p < .01$ . Respective one-item response frequencies were 60% versus 15%. A review of Figures 3, 4, and 5 shows no impressive difference between the two fusions: Both are moderately insensitive to relative onset time differences and very insensitive to relative intensity and relative frequency differences. In Table 3, however, one finds support for their separation in the evidence that the fusions are phenomenologically different for the listener.

### GENERAL DISCUSSION: SIX FUSIONS IN SPEECH PERCEPTION

#### Overview

There are four primary results to be emphasized from the five experiments presented here. First, Experiment 1 was successful in demonstrating that psychoacoustic fusion is the result of perceptual averaging of optimally

TABLE 5

UPPER LIMITS OF INTERSTIMULUS DISCREPANCIES  
PERMITTING CONSISTENT AND FREQUENT FUSIONS  
OF /da/ FOR SIX TYPES OF FUSION  
(UPDATING OF TABLE 1)

Fusion type	Onset time (msec)	Intensity (db.)	Frequency (Hz)
Sound localization	< 5	— <sup>a</sup>	< 2
Psychoacoustic fusion	< 40	< 10	> 80
Spectral fusion	< 40	40 <sup>b</sup>	> 80
Spectral/temporal fusion	< 40	40 <sup>b</sup>	> 80
Phonetic feature fusion	< 80	25	> 80
Phonological fusion	> 80	25	> 80

<sup>a</sup> Intensity differences are not relevant to sound localization as discussed here, since the fused percept never disintegrates with such variation.

<sup>b</sup> Rand (1974), using stimuli very similar to those used in the present study.

similar acoustic features of opposite-ear stimuli. Second, Experiments 2-4 replicated the general findings and estimates reported in Table 1, and filled in the empty cells for psychoacoustic fusion and spectral/temporal fusion. These data, shown in revised form in Table 5, are all based on the syllable /da/. Third, the results of Experiments 2-5 provided patterns of results that were used to differentiate the six fusions and to arrange them into three groups according to a levels-of-processing analysis. Those levels are discussed below. Fourth, the results of Experiments 4 and 5 suggested that the perception of a single event is not necessary for the assignment of a single linguistic response in five of the fusions (excluding sound localization).

The results of these experiments and the supporting evidence cited throughout the paper preclude the possibility that the six fusions can be accounted for in terms of a single general mechanism: The large variation in sensitivities to relative onset time differences and to intensity differences, and the effects of pitch on number versus identity of the fused percepts prevent any suggestions of a simple fusion system. Instead, at least three, perhaps even four, perceptual levels are needed, one for each different kind of perceptual combination.

#### *Level 1: Fusion by Integration of Waveforms*

*Sound localization* is the only fusion to occur at this first and lowest level, and the mechanism involved is one that cross-relates the waveforms of opposite-ear stimuli.

Three kinds of evidence separate this fusion from the other five. First, as shown in Experiment 2, no other fusion is as sensitive to differences in relative onset time. Onset differences of 4 and 5 msec are sufficient to inhibit fusion, nearly an order of magnitude less than those intervals necessary for other fusions to disintegrate. Second, extreme sensitivity to relative frequency differences in two speech sounds is very clear from Experiment 4, whereas frequency differences do not affect other fusions. Third, on logical grounds alone, sound localization is unique because it is the only fusion based on number of percepts rather than in their identity. See Deatherage (1966) for an account of the physiology of the binaural system and Sayers and Cherry (1957) for an indication of interactions involved in sound localization.

#### *Level 2: Fusion by Integration of Acoustic Features*

Three fusions appear to occur at this second level: *psychoacoustic fusion*, *spectral fusion*, and *spectral/temporal fusion*. Evidence for their common allocation comes from five sources. First, each is moderately sensitive to differences in relative onset time, yielding markedly similar patterns, especially when corrections are made for differential floor effects. In Experiment 2, each of these fusions withstood temporal differences of between 20 and 40 msec without marked disintegration of the fused percept. Second, and also stemming from Experiment 2, the shape of the functions of these three fusions as onset interval increases is quite different from that of the functions of phonetic feature fusion and phonological fusion, a fact that suggests entirely different processes are occurring. Third, none of these three fusions is dependent on prior perception of the dichotic inputs as a single event. Experiments 4 and 5, taken together, demonstrate that frequency differences have no effect on the probability of fusion responses, but that for psychoacoustic fusion and spectral fusion, frequency differences do significantly affect the numbers of items perceived to occur on any given trial. Fourth, although this third stipulation is also true for phonetic feature fusion and phono-

logical fusion, these three fusions differ from the two remaining fusions with respect to the importance of presentation mode. As shown in Table 3, there are as many or more "fusions" for these three phenomena when the items are presented binaurally than when they are presented dichotically. (Phonetic feature fusion and phonological fusion, on the other hand, show the reverse trend.) These data and those for similar phenomena reported elsewhere (Pisoni & McNabb, 1974; Repp, 1975a) suggest that this second level must be prephonetic. Fifth, relative stimulus intensity plays an important role in the only one of these fusions that occurs through dichotic competition—psychoacoustic fusion. Sensitivity to relative energy levels is indicative of lower level processing in audition.

Although they occur at the same perceptual level, these three fusions are separate phenomena. Psychoacoustic fusion, a general phenomenon that is most dramatically represented by the fusion of /ba/ and /ga/ into /da/, occurs through the perceptual averaging of similar but slightly discrepant information presented to opposite ears. The same averaging could be accomplished using synthetic steady-state vowels with slightly different vowel color: A vowel of midcolor between the two inputs is easily perceived. No such averaging occurs in spectral or spectral/temporal fusion, since there is no discrepant information competing between the two ears.<sup>6</sup>

Spectral fusion and spectral/temporal fusion differ strikingly in perceptual "appearance" to the listener. When fundamental frequencies of the to-be-fused stimuli are the same, listeners generally report hearing only one item in spectral fusion but two items in spectral/temporal fusion. This remarkable fact is sufficient evidence to consider them separate phenomena and necessitates further consideration of spectral/temporal fusion.

In spectral/temporal fusion, the second-formant transition is "heard" in two forms: as part of a speech syllable giving it its identity as /ba/, /da/, or /ga/, and as a brief glissando, or chirp. How does this dual perception come about? In part, one must appeal to dual perceptual systems of speech and nonspeech (Day & Bartlett, 1972; Day,

Bartlett, & Cutting, 1973; Mattingly et al., 1971; Day & Cutting, Note 7). The brief chirp appears to be fused (integrated) with the transitionless stimuli at Level 2 and then identified by a "speech processor." The information in this chirp also appears to remain in a relatively raw acoustic form—a brief acoustic blip against a background of a continuous periodic speech sound. In spectral fusion, by contrast, there is no brief signal to establish this figure-ground relationship. Logically, however, there are three possible percepts in the spectral/temporal fusion of /da/, as shown in Figure 1: the speech sound /da/ and the chirp, the two sounds that are actually heard, and the transitionless /da/, which is not heard. Why not? The following account is somewhat complex, but appears to explain the phenomenon and an additional anomaly.

The transitionless-/da/ stimulus is about 85% identifiable as /ba/. It is identified as /ba/, presumably, as noted earlier, because the harmonics of the first-formant transition mimic the absent second-formant transition and mimic it in such a fashion as to be appropriate for /b/. Thus at some level, this /ba/ competes with a redintegrated /da/. The redintegrated /da/, like the original stimulus, is a highly identifiable, hypernormal item (Mattingly, 1972). It competes against the considerably weaker /ba/, readily identifiable but without a prominent transitional cue typical of synthetic speech syllables. The /da/ easily "wins" in this mismatch and is perceived instead of /ba/. An anomaly that supports this account stems from the fact that fewer /ga/ spectral/temporal fusions occur than /da/ fusions, given the appropriate stimuli (Nye, Nearey, & Rand, Note 3; see also Experiment 2, present paper). Here the transitionless /ga/ (again perceived as /ba/) competes with the redintegrated /ga/. The

<sup>6</sup> The results of Perrott and Barry (1969), which demonstrate fusion of sine waves whose frequencies differ beyond the range of binaural beats, might be thought of as a form of psychoacoustic fusion, but since their task required detection of one versus two signals, and not the "identity" of the signals, their phenomenon appears to be one more closely related to sound localization.

result is that fewer /ga/ percepts arise and a number of /da/ responses are reported instead. This /da/-for-/ga/ substitution may arise from the psychoacoustic fusion of the transitionless /ga/ (perceived as /ba/) and the redintegrated /ga/.

If psychoacoustic, spectral, and spectral/temporal fusion occur at the same level of processing in the auditory system, it should be possible, as suggested above, for them to interact directly to create composite fusions. Indeed, this appears to be possible. Consider two composite fusions, the first a combination of psychoacoustic fusion and spectral fusion and the second a combination of psychoacoustic fusion and spectral/temporal fusion. In the first case, if the syllable /ba/ is presented to one ear and the second-formant of /ga/ to the other, the listener can easily hear /da/, probably with approximately the same probability as he or she heard it when the stimuli were /ba/ and /ga/. Berlin et al. (1973) and Porter (1975) have performed experiments similar to this, and from their results the fusion in this situation seems likely. In the second case, the syllable /ba/ might be presented to one ear and the /ga/ second-formant transition to the other. In this situation the listener may report hearing /da/ plus chirp. Pilot research supports the likelihood of these two composite fusions.

### *Level 3: Fusion by Disruption and Recombination of Linguistic Features*

*Phonetic feature fusion and phonological fusion* can be separated from the other fusions by three findings. First, the functions revealed when relative onset time is varied are unique: For both phenomena, fusions increase slightly with small onset-time asynchronies, only to decrease after onset asynchronies of greater than about 40 to 80 msec. This nonlinearity suggests a multistage process similar, perhaps, to that proposed by Turvey (1973). Fusion occurs best, it would seem, only after the first-arriving stimulus has been partially processed and features extracted from it, but it is then immediately disrupted by the arrival of the second item. Thus, these fusions are perceptual confusions resulting from the misassignment of linguistic features.

Second, these fusions are only moderately sensitive to intensity differences between the two stimuli and provide sharp contrast to psychoacoustic fusion, the only other of these six phenomena that occurs through perceptual competition of similar information presented to both ears. Insensitivity to relative energies is characteristic of higher level processes (see Turvey, 1973, for the visual parallel).

Third, as cited previously in this discussion, phonetic feature fusion and phonological fusion are the only fusions to suffer when presentation mode is changed from dichotic to binaural. This occurs, presumably, because the mixing of the signals degrades their intelligibility, and the stimuli mask one another through integration before disruption can ever take place.

These are the two fusions in which the actual combinatory process is necessarily linguistic. With the possible exception of spectral/temporal fusion, the responses of the listeners for other fusion phenomena are only incidentally linguistic. Although the results presented here do not distinguish phonetic feature fusion from phonological fusion, several other results and logical considerations may warrant separate linguistic levels for the two.

*Levels 3 and 4?* Empirical findings and several logical considerations distinguish the two language-based fusions. In Figure 3, it appears that phonological fusion is only slightly more tolerant of lead-time differences than is phonetic feature fusion. This apparent similarity may be misleading. Day (1970), Day and Cutting (Note 5), and Cutting (1975) have shown that when longer and more complex speech stimuli are used, such as one- and two-syllable words, tolerance to onset-time asynchronies can increase from about 80 msec to at least 150 msec, considerably beyond any consistent effects of phonetic feature fusion. Since higher level fusions typically allow for greater tolerance to relative onset differences, phonetic feature fusion might be thought to occur at Level 3 and phonological fusion at a new level, Level 4. A second finding that might support the separation of the two linguistic fusions is



that several previous studies in phonological fusion (Cutting, 1975; Cutting & Day, 1975) have found that pairs analogous to /da/-leading-/ra/ fuse more readily than pairs such as /ra/-leading-/da/. Such asymmetry does not occur for phonetic feature fusion.

In addition, logical considerations support the separation of the two phenomena. Phonological constraints, those which dictate the logic of contiguity for phonemes in a given language, are higher level language constraints than are phonetic feature analyses (see, among others, Studdert-Kennedy, 1974, in press; Kirstein, Note 8). For example, whereas phonetic features are largely universal across all languages, phonologies are language specific. In English, liquids (/l/) and /r/) cannot precede stop consonants (/b/, /g/, /p/, and /k/, for example) in initial position, but stops readily come before liquids. This fact appears to account for the fact that given a dichotic pair such as BANKET/LANKET the listener rarely reports hearing LBANKET. Instead, he or she often hears BLANKET.

Another consideration is also important and may separate phonological fusion from all other fusion phenomena. Day (Note 9, Note 10), Cutting and Day (1975), and to a lesser extent Cutting (1975) found that there are marked individual differences in the frequency of phonological fusions across different subjects. Some individuals fuse very frequently, others fuse relatively less often, and few individuals fuse at rates in between these two modes. Moreover, these differences correlate with those found on other tasks with the same stimuli and on tasks involving very different stimuli (Day, 1974; Day, Note 9). Preliminary results suggest that such radical and systematic differences may not occur elsewhere in the six fusions.

At least one additional consideration, however, supports the notion that the two linguistic fusions do indeed occur at the same level. Hofmann (Note 11) and Menyuk (1972) have suggested that clusters of phonemes, including stop-liquid clusters, may be parsimoniously described as single, underlying phonemes with their own unique phonetic features (see also Devine, 1971). The results

of Cutting and Day (1975, Experiment 4) appear to support this conclusion, in that certain aspects of the dichotic presentation may mimic certain phonetic feature values and contribute substantially to the perception of a stop-/l/ cluster. Thus, "blending" of phonetic features might account for both phonetic feature fusion and phonological fusion. In summary, then, data and logical considerations may suggest a separation of the two fusions, but the separation cannot yet be affirmed.

#### SUMMARY AND CONCLUSION

Fusion is not a single phenomenon in speech perception, but many phenomena. Six dichotic fusions were considered, and five of them are distinctive in that the fused percept differs from either of the two inputs. The robustness of these phenomena was measured against variation in three parameters: relative onset time of the two stimuli, relative intensity of the stimuli, and their relative frequency. Results, gathered here using the same subjects and essentially the same stimulus repertory for each fusion, agree with previously published accounts or, where there are no prior data, fit nicely into the scheme of upper and lower level fusions that is developed in this paper. The various fusions cannot occur at a single perceptual level: At least three, perhaps four, levels are needed.

Fusion on the one hand and rivalry and masking on the other allow reciprocal glances at the same phenomena. The levels of perceptual processing developed in the present study for audition are quite similar to those developed elsewhere in audition (Studdert-Kennedy, 1974; Wood, 1974, 1975) and also to those developed in vision (Turvey, 1973). With the exception of Julesz (1971), in both modalities, most research concerned with stages of processing has used rivalry-masking paradigms. The findings in the present paper suggest that fusion paradigms can also be used to probe the speech-processing system.

#### REFERENCE NOTES

1. Cutting, J. E. *A preliminary report on six fusions in auditory research* (Status Report on Speech Research, SR-31/32). New Haven, Conn.: Haskins Laboratories, 1972.

2. Haggard, M. *Asymmetrical analysis of stimuli with dichotically split formant information* (Speech Perception, 2). Belfast, Northern Ireland: Queen's University, Department of Psychology, 1975.
  3. Nye, P. W., Nearey, T. M., & Rand, T. C. *Dichotic release from masking: Further results from studies with synthetic speech stimuli* (Status Report on Speech Research, SR-37/38). New Haven, Conn.: Haskins Laboratories, 1974.
  4. Nearey, T. M., & Levitt, A. C. *Evidence for spectral fusion in dichotic release from upward spread of masking* (Status Report on Speech Research, SR 39/40). New Haven, Conn.: Haskins Laboratories, 1974.
  5. Day, R. S., & Cutting, J. E. *Levels of processing in speech perception*. Paper presented at the 10th meeting of the Psychonomic Society, San Antonio, Texas, November 1970.
  6. Turvey, M. T. Personal communication, January 1974.
  7. Day, R. S., & Cutting, J. E. *What constitutes perceptual competition in dichotic learning?* Paper presented at the meeting of the Eastern Psychological Association, New York, April 1971.
  8. Kirstein, E. F. *The lag effect in dichotic speech perception* (Status Report on Speech Research, SR-35/36). New Haven, Conn.: Haskins Laboratories, 1973.
  9. Day, R. S. *Individual differences in cognition*. Paper presented at the 13th meeting of the Psychonomic Society, St. Louis, November 1973.
  10. Day, R. S. *Temporal-order judgments in speech: Are individuals language-bound or stimulus bound?* (Status Report on Speech Research, SR-21/22). New Haven, Conn.: Haskins Laboratories, 1970.
  11. Hofmann, T. R. *Initial clusters in English* (Quarterly Progress Report No. 84). Cambridge: Massachusetts Institute of Technology, Research Laboratory of Electronics, 1967.
- REFERENCES
- Ades, A. E. Bilateral component in speech perception? *Journal of the Acoustical Society of America*, 1974, *56*, 610-616.
- Allport, D. A. Phenomenal spontaneity and the perceptual moment hypothesis. *British Journal of Experimental Psychology*, 1968, *59*, 395-406.
- Angell, J. R., & Fite, W. The monaural localization of sound. *Psychological Review*, 1901, *8*, 225-246.
- Babkoff, H. Dichotic temporal interaction: Fusion and temporal order. *Perception & Psychophysics*, 1975, *18*, 267-272.
- Berlin, C. E. et al. Dichotic signs of the recognition of speech elements in normals, temporal lobectomies, and hemispherectomies. *IEEE Transactions on Audio and Electroacoustics*, 1973, *AU-21*, 189-195.
- Bever, T. G., & Chiarello, R. J. Cerebral dominance in musicians and nonmusicians. *Science*, 1974, *195*, 537-539.
- Bilsen, F. A., & Goldstein, J. L. Pitch of dichotically delayed noise and its possible spectral basis. *Journal of the Acoustical Society of America*, 1974, *55*, 292-296.
- Blechner, M. J., Day, R. S., & Cutting, P. E. Processing two dimensions of nonspeech stimuli: The auditory phonetic distinction reconsidered. *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- Broadbent, D. E. A note on binaural fusion. *Quarterly Journal of Experimental Psychology*, 1955, *7*, 46-47.
- Broadbent, D. E., & Ladefoged, P. On the fusion of sounds reaching different sense organs. *Journal of the Acoustical Society of America*, 1957, *29*, 708-710.
- Cherry, E. C., & Taylor, W. K. Some further experiments upon the recognition of speech with one ear and with two ears. *Journal of the Acoustical Society of America*, 1954, *26*, 554-559.
- Cooper, F. S., & Mattingly, I. G. Computer-controlled PCM system for investigation of dichotic speech perception. *Journal of the Acoustical Society of America*, 1969, *46*, p. 115. (Abstract)
- Cramer, E. M., & Huggins, W. H. Creation of pitch through binaural interaction. *Journal of the Acoustical Society of America*, 1958, *30*, 413-417.
- Cullen, J. K., Thompson, C. L., Hughes, L. F., Berlin, C. I., & Samson, D. S. The effects of varied acoustic parameters on performance in dichotic speech perception tasks. *Brain and Language*, 1974, *1*, 307-322.
- Cutting, J. E. Levels of processing in phonological fusion (Doctoral dissertation, Yale University, 1973). *Dissertation Abstracts International*, 1973, *34*, p. 2332B. (University Microfilms No. 73-25, 191)
- Cutting, J. E. Two left-hemisphere mechanisms in speech perception. *Perception & Psychophysics*, 1974, *16*, 601-612.
- Cutting, J. E. Aspects of phonological fusion. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, *1*, 105-120.
- Cutting, J. E. The magical number two and the natural categories of speech and music. In N. S. Sutherland (Ed.), *Tutorial essays in psychology*. Hillsdale, N.J.: Erlbaum, in press.
- Cutting, J. E., & Day, R. S. The perception of stop-liquid clusters in phonological fusion. *Journal of Phonetics*, 1975, *3*, 99-113.
- Cutting, J. E., & Rosner, P. S. Categories and boundaries in speech and music. *Perception & Psychophysics*, 1974, *16*, 564-570.
- Cutting, J. E., Rosner, B. S., & Foard, C. F. Perceptual categories for musiclike sounds: Implications for theories of speech perception. *Quarterly Journal of Experimental Psychology*, in press.
- Darwin, C. J. Dichotic backward masking of complex sounds. *Quarterly Journal of Experimental Psychology*, 1971, *23*, 386-392.
- Day, R. S. Fusion in dichotic listening (Doctoral dissertation, Stanford University, 1968). *Disserta-*

- tion *Abstracts International*, 1969, 29, p. 2649B. (University Microfilms No. 69-211)
- Day, R. S. Temporal-order perception of a reversible phoneme cluster. *Journal of the Acoustical Society of America*, 1970, 48, p. 95. (Abstract)
- Day, R. S. Differences in language-bound and stimulus-bound subjects in solving word-search puzzles. *Journal of the Acoustical Society of America*, 1974, 55, p. 412. (Abstract)
- Day, R. S., & Bartlett, J. C. Separate speech and nonspeech processing in dichotic listening? *Journal of the Acoustical Society of America*, 1972, 51, p. 79. (Abstract)
- Day, R. S., Bartlett, J. C., & Cutting, J. E. Memory for dichotic pairs: Disruption of ear performance by the speech/nonspeech distinction. *Journal of the Acoustical Society of America*, 1973, 53, p. 358. (Abstract)
- Deatherage, B. An examination of binaural interaction. *Journal of the Acoustical Society of America*, 1966, 39, 232-249.
- Delattre, P. C., Liberman, A. M., & Cooper, F. S. Acoustic loci and transitional cues for consonants. *Journal of the Acoustical Society of America*, 1955, 27, 769-773.
- Deutsch, D. Two-channel listening to musical scales. *Journal of the Acoustical Society of America*, 1975, 57, 1156-1160. (a)
- Deutsch, D. Musical illusions. *Scientific American*, 1975, 233(4), 92-105. (b)
- Deutsch, D., & Röll, P. L. Separate "what" and "where" decision mechanisms in processing a dichotic tonal sequence. *Journal of Experimental Psychology: Human Perception and Performance*, 1976, 2, 23-29.
- Devine, A. M. Phoneme or cluster: A critical review. *Phonetica*, 1971, 24, 65-85.
- Efron, R. The relationship between the duration of a stimulus and the duration of a perception. *Neuropsychologia*, 1970, 8, 37-55.
- Fourcin, A. J. An aspect of the perception of pitch. In A. Sovijarvi & P. Aalto (Eds.), *Proceedings of the Fourth International Congress of Phonetic Sciences*. The Hague, Netherlands: Mouton, 1962.
- Fry, D. P. Perception and recognition in speech. In M. Halle, H. G. Lunt, and C. H. Schoonveld (Eds.), *For Roman Jakobson*. The Hague: Netherlands: Mouton, 1956.
- Goen, J. J. Super- and subliminal binaural beats. *Acta Oto-laryngologica*, 1964, 57, 224-230.
- Guttman, N., & Julesz, B. Lower limits of auditory periodicity analysis. *Journal of the Acoustical Society of America*, 1963, 35, p. 610.
- Halwes, T. G. Effects of dichotic fusion on the perception of speech (Doctoral dissertation, University of Minnesota, 1969). *Dissertation Abstracts International*, 1970, 31, p. 1565B. (University Microfilms No. 70-15, 736)
- Houtsma, A. T. M., & Goldstein, J. L. The central origin of the pitch of complex tones: Excerpts from musical interval recognition. *Journal of the Acoustical Society of America*, 1972, 51, 520-529.
- Huggins, A. W. F. Distortion of the temporal pattern of speech: Interruption and alteration. *Journal of the Acoustical Society of America*, 1964, 36, 1055-1064.
- Huggins, A. W. F. On the perceptual integration of dichotically alternating pulse trains. *Journal of the Acoustical Society of America*, 1974, 56, 939-943.
- Huggins, A. W. F. Temporally segmented speech. *Perception & Psychophysics*, 1975, 18, 149-157.
- Jeffress, L. A. Binaural signal detection: Vector theory. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.
- Julesz, B. *Foundations of cyclopean perception*. Chicago: University of Chicago Press, 1971.
- Kaufman, L. On the spread of suppression and binocular rivalry. *Vision Research*, 1963, 3, 401-415.
- Kaufman, L. *Sight and mind*. New York: Oxford University Press, 1974.
- Kubovy, M., Cutting, J. E., & McGuire, R. M. Hearing with the third ear: Dichotic perception of a melody without monaural familiarity cues. *Science*, 1974, 186, 272-274.
- Leakey, D. M., Sayers, B. M., & Cherry, E. C. Binaural fusion of low- and high-frequency sounds. *Journal of the Acoustical Society of America*, 1958, 30, 22.
- Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. Perception of the speech code. *Psychological Review*, 1967, 74, 631-661.
- Licklider, J. C. R. Basic correlates of the auditory stimulus. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951.
- Licklider, J. C. R., Webster, J. C., Hedlund, J. M. On the frequency limits of binaural beats. *Journal of the Acoustical Society of America*, 1950, 22, 468-473.
- Linden, A. Distorted speech and binaural speech resynthesis tests. *Acta Oto-laryngologica*, 1964, 58, 32-48.
- Lisker, L., & Arbamson, A. S. A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, 20, 384-422.
- Locke, S., & Kellar, L. Categorical perception in a nonlinguistic mode. *Cortex*, 1973, 9, 355-369.
- Marslen-Wilson, W. D. Sentence perception as an interactive parallel process. *Science*, 1975, 189, 226-228.
- Massaro, D. W. Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, 1972, 79, 124-145.
- Massaro, D. W. Perceptual units in speech recognition. *Journal of Experimental Psychology*, 1974, 102, 199-208.
- Mattingly, I. G. Speech cues and sign stimuli. *American Scientist*, 1972, 60, 327-337.
- Mattingly, I. G., Liberman, A. M., Syrdal, A., & Halwes, T. Discrimination in speech and non-speech modes. *Cognitive Psychology*, 1971, 2, 131-157.
- Matzker, J. Two new methods for the assessment of central auditory function in cases of brain dis-

- ease. *Annals of Otolaryngology, Rhinology, and Laryngology*, 1959, 68, 1185-1197.
- Menyuk, P. Clusters as single underlying consonants: Evidence from children's productions. In A. Rigault & R. Charbonneau (Eds.), *Proceedings of the Seventh International Congress of Phonetic Sciences*. The Hague, Netherlands: Mouton, 1972.
- Mills, A. W. Auditory localization. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 2). New York: Academic Press, 1972.
- Noorden, L. P. A. S. van. *Temporal coherence in the perception of tone sequences*. Eindhoven, Netherlands: Instituut voor Perceptie Onderzoek, 1975.
- Oster, G. Auditory beats and the brain. *Scientific American*, 1973, 229(4), 94-103.
- Perrott, D. R., & Barry, S. H. Binaural fusion. *Journal of Auditory Research*, 1969, 3, 263-269.
- Perrott, D. R., & Elfiner, L. F. Monaural localization. *Journal of Auditory Research*, 1968, 8, 185-193.
- Perrott, D. R., & Nelson, M. A. Limits for the detection of binaural beats. *Journal of the Acoustical Society of America*, 1969, 46, 1477-1481.
- Pisoni, D. B. Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 1973, 13, 253-260.
- Pisoni, D. B. Speech perception. In W. K. Estes (Ed.), *Handbook of learning and cognitive processes* (Vol. 5). Hillsdale, N.J.: Erlbaum, in press.
- Pisoni, D. B., & McNabb, S. D. Dichotic interaction of speech sounds and phonetic feature processing. *Brain and Language*, 1974, 1, 351-362.
- Porter, R. J. Effect of delayed channel on the perception of dichotically presented speech and non-speech sound. *Journal of the Acoustical Society of America*, 1975, 58, 884-892.
- Rand, T. C. Dichotic release from masking for speech. *Journal of the Acoustical Society of America*, 1974, 55, 678-680.
- Repp, B. H. Dichotic forward and backward "masking" between CV syllables. *Journal of the Acoustical Society of America*, 1975, 57, 483-496.
- (a)
- Repp, B. H. Dichotic masking of consonants by vowels. *Journal of the Acoustical Society of America*, 1975, 57, 724-735. (b)
- Repp, B. H. Distinctive features, dichotic competition, and the encoding of stop consonants. *Perception & Psychophysics*, 1975, 17, 231-242. (c)
- Repp, B. H. Perception of dichotic place contrasts. *Journal of the Acoustical Society of America*, 1975, 58 (supplement), p. 76. (Abstract) (d)
- Sayers, B. M., & Cherry, E. C. Mechanism of binaural fusion in the hearing of speech. *Journal of the Acoustical Society of America*, 1957, 29, 973-987.
- Shankweiler, D., & Studdert-Kennedy, M. Identification of consonants and vowels presented to left and right ears. *Quarterly Journal of Experimental Psychology*, 1967, 19, 59-63.
- Smith, B. A., & Resnick, D. M. An auditory test for assessing brain stem integrity: Preliminary report. *Laryngoscope*, 1972, 82, 414-424.
- Speaks, C., & Bissonette, L. J. Interaural-intensive differences and dichotic listening. *Journal of the Acoustical Society of America*, 1975, 58, 893-898.
- Studdert-Kennedy, M. The perception of speech. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12). The Hague, Netherlands: Mouton, 1974.
- Studdert-Kennedy, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, in press.
- Studdert-Kennedy, M., & Shankweiler, D. P. Hemispheric specialization for speech. *Journal of the Acoustical Society of America*, 1970, 48, 579-594.
- Studdert-Kennedy, M., Shankweiler, D. P., & Pisoni, D. B. Auditory and phonetic processes in speech perception: Evidence from a dichotic study. *Cognitive Psychology*, 1972, 3, 455-466.
- Studdert-Kennedy, M., Shankweiler, D. P., & Schulman, S. Opposed effects of a delayed channel on perception of dichotically and monotonically presented CV syllables. *Journal of the Acoustical Society of America*, 1970, 48, 599-602.
- Tallal, P., & Piercy, M. Developmental aphasia: Rate of auditory processing and selective impairment of consonant perception. *Neuropsychologia*, 1974, 12, 83-93.
- Tallal, P., & Piercy, M. Developmental aphasia: The perception of brief vowels and extended stop consonants. *Neuropsychologia*, 1975, 13, 69-73.
- Thurlow, W. R., & Elfiner, L. F. Pure-tone cross-ear localization effects. *Journal of the Acoustical Society of America*, 1959, 31, 1606-1608.
- Tobias, J. V. Curious binaural phenomena. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol 2). New York: Academic Press, 1972.
- Turvey, M. T. On peripheral and central processes in vision: Inferences from an information-processing analysis of masking with patterned stimuli. *Psychological Review*, 1973, 80, 1-52.
- Wood, C. C. Parallel processing of auditory and phonetic information in speech discrimination. *Perception & Psychophysics*, 1974, 15, 501-508.
- Wood, C. C. Auditory and phonetic levels of processing in speech perception: Neurophysiological and information-processing analyses. *Journal of Experimental Psychology: Human Perception and Performance*, 1975, 1, 3-20.
- Wood, C. C., Goff, W. R., & Day, R. S. Auditory evoked potentials during speech perception. *Science*, 1971, 173, 1248-1251.
- Woodworth, R. S. *Experimental psychology*. New York: Holt, 1938.