A Preliminary Report on Six Fusions in Auditory Research

James E. Cutting
Haskins Laboratories, New Haven

Since the 1950's a number of auditory phenomena have been called "fusion" by various researchers. Broadbent (1955), Day (1968), and Halwes (1969), among others, have described experimental situations in which two auditory signals are perceived as one. From the titles of their papers one would assume that they are concerned with the same process: "On the fusion of sounds reaching different sense organs" (Broadbent and Ladefoged, 1957); "Fusion in dichotic listening" (Day, 1968); and "Effects of dichotic fusion on the perception of speech" (Halwes, 1969).

However, fusion is not just one phenomenon, but many phenomena which are, at best, only tenuously related. Subsuming them all under the single label "fusion" with no descriptive adjective easily leads to confusion. The purpose of this paper is to act as a preliminary report, delimiting the various types of auditory fusion, investigating their similarities and dissimilarities, and arranging them on a cognitive hierarchy according to the processing characteristics in each. We will consider six different types of fusion, beginning with the more primitive fusion phenomena and moving towards more complex phenomena.

Cognitive Levels and Criteria for Fusion Classifications

Before considering the various types of fusion, it is necessary to define the levels which we will discuss and to establish criteria for judging the placement of fusions at these levels. We will consider two general levels: designated "higher" and "lower" levels.

Lower-level fusions are energy-dependent. Pitch and intensity are examples of energy parameters. When lower-level fusions are involved small differences in pitch (2 Hz) or small differences in intensity (2 db) between the two to-be-fused stimuli may inhibit fusion or change the fused percept. Timing, in terms of the relative onset time of the two stimuli, is another important parameter. If one stimulus precedes the other by a sufficient interval, fusion no longer occurs and two stimuli are heard. This interval is very small for lower-level fusions, often a matter of microseconds (microsec).

Higher-level fusions are energy-independent. Pitch and intensity may vary between the two stimuli within a much greater range in the higher-level fusions. Differences in the stimuli of 20 Hz or 20 db may not inhibit fusion at all.

†Also Yale University, New Haven.

[HASKINS LABORATORIES: Status Report on Speech Research SR-31/32 (1972)]
Relative onset time also plays a lesser role; the two stimuli may vary in relative onsets within a range well beyond that of the lower-level fusions. Higher-level fusions are often insensitive to differences of from 25 to 150 msec. Information, not energy or timing, seems to be important in these fusions, and, as we shall see later, information in the to-be-fused stimuli can often override energy differences.

Other, more psychological characteristics also distinguish the higher fusions from lower fusions. Lower-level fusions are generally passive phenomena, whereas higher-level fusions are more constructive in nature. In the lower-level fusions the subject listens to one clear auditory image and is usually unaware that two stimuli are being presented. In the higher-level fusions, however, the subject listens to a more diffuse auditory image; he may even report hearing two sounds or one sound that sounds a bit strange.

In all six types of fusion, the to-be-fused stimuli are presented dichotically, one stimulus to the right ear and the other stimulus to the left ear. In each case the subject is asked to report what he heard. A diagram of each type is shown in schematic form in Figure 1; they will be discussed in turn below. There is a temptation to think of all six fusions primarily as central processes; stimuli transmitted by different channels and integrated into a single percept. This judgment may be misleading. In vision research, Turvey (in press) has noted that peripheral processes tend to be those which are affected by changes in the energy of the stimuli, while central processes tend to be those which are independent of stimulus energy and are more concerned with the information in the stimuli. This peripheral/central dichotomy parallels the lower-level/higher-level distinction outlined above for auditory fusions. If lower-level fusions are energy-dependent, perhaps they are primarily concerned with peripheral mechanisms. If, on the other hand, higher-level fusions are energy-independent, perhaps they are primarily concerned with central mechanisms. For the purposes of this paper peripheral and central processes will be synonymous with lower and higher cognitive levels.

1. SOUND LOCALIZATION: Fusion of two identical events.

Sound localization has been included as a form of fusion to give a reference point in considering other types of fusion. All audible sounds, simple or complex, can be localized—and usually are. It is the most basic form of auditory fusion and occurs for both speech and nonspeech sounds. The best 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.

Three parameters affect sound localization: pitch, intensity, and timing. First, consider pitch. If two tones are presented, one to each ear, the subject may fuse (localize) them. If the tones have the same pitch, fusion occurs and one tone is heard. If the tones differ by 2 Hz, fusion begins to disintegrate and a more wavering tone is heard. Differences of more than 2 Hz often inhibit fusion altogether, and two tones are heard.¹

¹This range is particularly relevant for tones below 1000 Hz. Above 1000 Hz the effect is produced by slightly larger pitch differences.
Figure 1: Six fusions of /da/. Schematic spectrograms of speech and speech-like stimuli in six types of auditory fusion.
Intensity is a second parameter which affects localization. If we present two pure tones, one to each ear, at the same frequency and the same intensity the subject usually reports that the apparent source of the tone is "in the middle of the head," or at the midline. If one tone is increased by 2 db without changing the intensity of the other stimulus, he normally reports that the sound has moved in the direction of the ear that received the more intense stimulus. The more we increase the intensity of the louder stimulus, the more the apparent source moves away from the midline and towards the ear with the more intense sound.

The third parameter which affects sound localization is timing. If we present a brief click simultaneously to each ear, the subject reports hearing one click localized at the midline. Delaying one click by 500 microsec causes the apparent source of the click to move away from the midline toward the ear with the leading click. If we delay one click by only 1 msec, the apparent source moves far enough from the midline so that the subjective percept is that one click was presented to one ear with nothing presented to the other. With delays of longer than about 2 msec fusion disintegrates and two clicks are heard. Another form of timing differences also affects localization. If two tones are presented, one to each ear, at the same pitch and intensity, but one tone is slightly delayed, the two tones will be out of phase. In other words, the two ears receive different aspects of the same waveform at a given point in time. If the delay is about 500 microsec the apparent source of the sound moves away from the midline towards the ear that received the leading sound. This perception is produced by phase differences in the stimuli, not by relative onset differences. The fact that localization is highly sensitive to energy parameters of the stimuli and is intolerant of small differences in stimulus timing suggests that it is a lower-level process.

Both speech and nonspeech sounds are fused in sound localization. The first display in Figure 1 shows an example of the localization of speech sounds. If /da/ is presented to both ears at the same time, at the same intensity, and with the same fundamental pitch, a single /da/ will be perceived by the subject at the midline. The same result occurs when a nonspeech sound replaces the /da/ stimuli.

2. SPECTRAL FUSION: Fusion of different spectral parts of the same signal.

Broadbent (1955) and Broadbent and Ladefoged (1957) reported a second type of fusion. Spectral fusion occurs when different spectral ranges of the same signal are presented to opposite ears. A given stimulus is filtered into two parts: one containing the low frequencies and one containing the high frequencies. Each is then presented separately but simultaneously to a single ear. The subject reports hearing the original stimulus, as if it had undergone no special treatment. In his initial study Broadbent found that fusion readily occurred for many complex sounds. Subjects fused metronome ticks and fused certain speech sounds when these stimuli were filtered and presented to opposite ears. When the subjects were informed about the nature of the stimuli and asked to report which ear had the low frequency sounds and which ear had the high frequencies they performed at chance level.

Pitch is an important parameter in spectral fusion. Just as there is no sound localization for dichotic tones of different pitches, there is no spectral fusion of complex dichotic stimuli with different pitches (fundamental frequencies). Broadbent and Ladefoged (1957) found that the fundamental frequencies of
the to-be-fused stimuli must be identical for fusion to occur. They presented the first formant of a steady-state vowel to one ear, and the second formant to the other. Subjects fused the stimuli when they had the same pitch, but did not fuse when they had different pitches. Instead, subjects heard two nonspeech sounds. Halwes (1969) found that fundamental frequency differences of 2 Hz are sufficient to inhibit this type of lower-level fusion.

The fundamental frequency of a speech stimulus is analogous to the carrier-band frequency of a radio station signal. If the to-be-fused spectral stimuli have different pitches, their entire waveforms are as different as if they had been broadcast by two different radio stations; the sound waves of the stimuli are not of the same duration, or even multiples of one another. Like a radio set which cannot integrate (receive) two different radio stations at once, the subject cannot integrate (fuse) two spectral stimuli with different pitches. The information that the two stimuli are the formants appropriate for the perception of a vowel cannot overcome pitch disparities in the signals, and no fusion occurs.

The effect of intensity differences between the stimuli has not been explored systematically. Nevertheless, preliminary investigations indicate that spectral fusion is quite sensitive to small differences in intensity between the to-be-fused stimuli.

When the relative onset time of the high and low frequency components of the same signal is altered beyond a few msec fusion no longer occurs and the subject hears two separate signals. For example, when the different spectral parts of metronome ticks are offset by as little as 5 msec, the subject hears two sets of ticks, not one. If filtered speech passages begin at slightly different times, the subject reports hearing two speech-like passages, not one. Thus, spectral fusion is very sensitive to small changes in timing.2

Speech sounds more complex than steady-state vowels are also subject to spectral fusion. The second display of Figure 1 shows the first formant of /da/ presented to one ear, and the second formant to the other ear. Provided these stimuli have the same pitch the subject will report hearing one stimulus, the syllable /da/.

---

2 There is, however, an exception to this timing sensitivity. Using the filters explained previously, Broadbent (1955) presented the first formant of the steady-state vowel /i/ as in BEET to one ear and the second and third formants to the other. The vowel sound was continuous over a duration of many seconds, with a constant pitch, and was recorded on a tape loop. One part of the loop was presented to one ear and a different part of the loop to the other, with the relative timing of the filtered segments off by as much as two or three seconds. Yet fusion occurred: the subjects heard /i/. The explanation for this result is quite simple. Timing is not important in a steady-state vowel with a constant pitch; unlike most other speech sounds, one section is exactly like any other section. Therefore, one section of the filtered vowel fuses with any other appropriately filtered section. This may be the exception which proves the rule that spectral fusion is highly sensitive to timing differences in the stimuli.
3. **PSYCHOACOUSTIC FUSION**: Fusion of one feature by acoustic averaging.

Psychoacoustic fusion is a third type of fusion which occurs at lower levels of cognitive processing. Although it probably occurs for both speech and nonspeech stimuli, the examples considered here involve speech stimuli. Unlike the stimuli in previously discussed fusions, the stimuli in psychoacoustic fusion have different information (in terms of phonetic features) in the same spectral range.

To demonstrate psychoacoustic fusion, let us choose two consonant-vowel (CV) stimuli, /ba/ and /ga/. Both are restricted to two formants, and parameters such as pitch, formant frequencies, and duration must be identical. The only difference between the stimuli are the direction and extent of the second formant (F2) transition. As shown in the third display of Figure 1, if /ba/ is presented to one ear and /ga/ to the other, the most frequent error is /da/ (Halwes, 1969:61). Halwes found that in such a situation the subject often reports hearing one of the given stimuli, /ba/ or /ga/; however, when he does make an error it usually is the fusion /da/.

Given the stimuli /ba/-/ga/, how does /da/ result? The stimuli differ in terms of the F2 transitions. However, the remainder of both stimuli are identical and are localized at the midline. The two F2 transitions, one rising in /ba/ and one falling in /ga/, appear to be algebraically averaged in such a manner that the subject perceives a stimulus with an intermediate F2 transition, /da/. (Note that an analogous situation arises when the phonemes /p/ and /k/ are in competition: /t/ is often perceived.)

Psychoacoustic fusion appears to be a lower-level, peripheral process because, as in the previously discussed phenomena, fusion occurs only when the pitch of the two stimuli is identical. Preliminary studies show that if the pitch of the two stimuli, /ba/ and /ga/, differs by as little as 2 Hz, the subject rarely reports hearing /da/, thus indicating that no acoustic averaging takes place.

The effects of intensity and timing differences between the stimuli have not been systematically studied. Nevertheless, pilot work suggests that psychoacoustic fusion is sensitive to small stimulus differences in both parameters.

We will reconsider psychoacoustic fusion after we have considered phonetic feature fusion. The two processes are similar yet different, in many ways. Appropriate comparisons will be made between the two.

4. **PHONETIC FEATURE FUSION**: Fusion of two features by phonetic blending.

With this fourth type of fusion we move to a more central, higher-level process. This is not to say that peripheral mechanisms are inactive, but these fusions cannot be explained wholly by such mechanisms. Halwes (1969) and Studdert-Kennedy and Shankweiler (1970) have reported that phonetic "blending" occurs in the dichotic competition of stop-vowel syllables. This "blending" is phonetic feature fusion. In the fourth display of Figure 1, we note that when the syllable /ba/ is presented to one ear and the syllable /ta/ to the other ear, the subject often reports hearing a syllable which was not presented. The most frequent errors are the "blends" /da/ and /pa/. Here, the subject appears to combine the voicing feature of one of the stimuli with the place feature of
the other. For example, the voicing feature of /b/ is combined with the place feature of /t/ and the result is the fusion /d/.

Let us assume a stimulus repertory of six items: /ba, da, ga, pa, ta, ka/. On a particular trial, three types of responses may occur: correct responses, "blend" errors, and anomalous errors. Given a stimulus pair such as /ba/-/ta/, "blend" errors are much more common than anomalous errors. The anomalous errors for this example are /ga/ and /ka/; both share a voicing feature with one of the pair but they do not share the place feature with either. Using natural speech stimuli Studdert-Kennedy and Shankweiler (1970) found that the ratio of "blend" errors to anomalous errors was 2:1, a rate significantly greater than chance.

Phonetic feature fusion is more central than the fusions discussed above because, for the first time in this discussion, the subject fuses even when the stimuli do not have the same pitch. Using synthetic stimuli Halwes (1969) reports that "blend" errors occur almost as frequently when the two competing stimuli have different pitches as when they have the same pitch.

The effect of intensity and timing differences between the stimuli have not been systematically studied in phonetic feature fusion. However, we may draw on experimental evidence from other sources. Thompson, Stafford, Cullen, Hughes, Lowe-Bell, and Berlin (1972) have noted that dichotic competition occurs for stop-vowel syllables even when one stimulus is 30 dB louder than the other. When stimuli such as /ba/ and /ta/ compete phonetic feature fusion may occur, even with such large intensity differences. We will reconsider the nature of dichotic competition and perceptual fusion later in this discussion.

Evidence concerning the effect of timing differences in phonetic feature fusion is also indirect. Studdert-Kennedy, Shankweiler, and Schulman (1970) have shown that when two CV syllables are presented at various relative onsets, the subject tends to report the syllable which began second better than the syllable which began first. This tendency is very pronounced between relative onset differences of 25 to 70 msec, and has been called the "lag effect." In this region the first stimulus appears to be masked by the second stimulus. If such masking occurs, fusion cannot occur because the phonetic information in the first stimulus is lost. Thus, from these results, we may assume that phonetic feature fusion occurs for temporal onset differences of up to about 25 msec, but that beyond 25 msec the "lag effect" inhibits any fusion that might occur.

Recent evidence indicates that parameters other than pitch, intensity, and timing may differ between the stimuli, and fusion rate still remains the same. For example, the vowels of the stimuli may be different, and fusion still occurs. The data of Studdert-Kennedy, Shankweiler, and Pisoni (1972) show that phonetic feature fusion occurs almost as readily when the stimuli are /bi/-/tu/ as when they are /bi/-/ti/. In both cases the subject is likely to respond with syllables beginning with the "blend" phonemes /d/ or /p/.

A comparison of psychoacoustic fusion and phonetic feature fusion. Because psychoacoustic fusion (fusion type 3) and phonetic feature fusion (fusion type 4) are highly confusable, it is important to make direct comparisons between them. Both processes involve the simultaneous presentation of phonetic

---

3 D. B. Pisoni, personal communication.
information, and both result in general information loss. In the case of psychoacoustic fusion the /b/ and /g/ features merge into a /d/ feature, and the /b/ and /g/ are lost. In the case of phonetic feature fusion the subject hears two stimuli, and when he does not perceive both of them correctly, he tends to "blend" them. In the process of listening to the dichotic pair, the features associated with a particular stimulus appear to lose their source. The /b/ in /ba/ ceases to be a /b/ and becomes a series of features which are appropriate to /b/: stop consonant, voiced, labial. When another stop consonant competes with /b/, the organization of these features appears to be disrupted, and they frequently get inappropriately reassigned to those of the competing stimulus. Thus, /ba/ plus /ta/ may become /da/, and the /b/ and /t/ are lost.

Dimensions which distinguish the two types of fusion are: 1) the pitch of the stimuli, 2) the vowels of the stimuli, and 3) the phonetic features of the stop consonants. Let us consider pitch and vowel requisites together. In the case of psychoacoustic fusion the two stimuli must have the same pitch and the same vowel. In the case of phonetic feature fusion, the two stimuli must have either different pitches or different vowels; these differences ensure that the phonetic features of the stimuli cannot be acoustically averaged. The other important dimension concerns the features of the to-be-fused stimuli. In psychoacoustic fusion the stop consonants must differ in only the place feature, and the voicing feature must be shared by the two stimuli (for example, /ba/ and /ga/). In phonetic feature fusion, on the other hand, the two stimuli must differ along both dimensions, place and voicing (for example, /ba/ and /ta/).

It is possible that psychoacoustic fusion and phonetic feature fusion may occur at the same time for the same competing stop-vowel stimuli. If /ba/ and /ta/ are presented at the same pitch, an ambiguous experimental situation results. The identical pitches and the shared vowel of the two stimuli set up a situation in which psychoacoustic fusion may occur. The unshared place and voicing features of the two stops, on the other hand, set up a situation in which phonetic feature fusion may occur. If the subject reports hearing /da/, we cannot be sure if the fusion is purely phonetic in nature, or if an element of psychoacoustic averaging contributed to the percept. Perhaps both processes are involved. In such cases, the fusion of /ba/ and /ta/ at the same pitch may be a hybrid of psychoacoustic fusion and phonetic feature fusion.

A note on dichotic competition, perceptual rivalry, and perceptual fusion. Dichotic stop-vowel stimuli are normally thought to "compete" with one another. This competition is perhaps more clearly defined as perceptual rivalry. When two CV syllables are presented dichotically, the subject typically reports hearing one or both of them. The stimuli are rivals, and we have thought that they are not usually combined into a single percept. However, as we have seen in psychoacoustic fusion (fusion type 3) and in phonetic feature fusion (fusion type 4) the subject often fuses stop-vowel stimuli. Thus, perceptual rivalry and perceptual fusion appear to converge, since both processes can occur for the same stimuli.

Although rivalry and fusion may occur simultaneously within the same stimuli, they do not appear to occur at the same level. Consider phonetic feature fusion. Given /ba-/ /ta/ the subject never reports hearing /tba/ or /bta/; the phonological constraints of English do not permit two stop consonants
to cluster within a syllable. Thus, we have a case of clear phonological rivalry; at the phoneme level the two stop consonants compete for the same processor. At another level, however, there is fusion. The /b/ and /t/ do not share the same values of place and voicing features. Any combination of labial and alveolar features with voiced and voiceless features yields a permissible stop consonant. Since there are no shared place and voice features in /b/ and /t/, there is no phonetic rivalry between them. The result of this pairing is often fusion at the phonetic level and rivalry at the phonological level.

A similar pattern occurs in psychoacoustic fusion. Given /ba/-/ga/ the subject never reports hearing /bga/ or /gba/; again, there is a clear phonological rivalry. At another level, however, there is fusion, this time at a psychoacoustic level. The place feature of /b/ merges with the place feature of /g/, and the intermediate phoneme /d/ is perceived.

5. CHIRP FUSION: Perceptual construction of phonemes from speech and nonspeech stimuli.

Rand (in preparation) discovered a fifth type of fusion. In chirp fusion there are no competing phonemes, nor is there information loss; instead, different parts of the same speech signal are presented to either ear. The fifth display of Figure 1 shows the syllable /da/ divided into two stimuli. One stimulus contains all the acoustic information in /da/ except the F2 transition, namely the entire first formant and most of the second formant. It is important to note that /da/ without the F2 transition is difficult to identify and is more readily perceived as /ba/ than /da/. The F2 transition alone is the second stimulus. It is very brief—a rapidly falling pitch sweep similar to a bird's twitter, hence the name "chirp." When the /da/ chirp is presented to one ear and the remaining portions of the syllable to the other, the subject reports hearing the full syllable /da/ plus the nonspeech chirp.

Perhaps the most interesting aspect of chirp fusion is that the subject hears more than one auditory image. As in phonetic feature fusion (fusion type 4) he hears two sounds, but he does not hear two speech sounds; instead, he hears one speech sound, /da/, and one nonspeech sound, the chirp. Note that the perceptual whole is greater than the sum of the parts: the subject "hears" the chirp in two different forms at the same time. One form is in the complete syllable /da/, which would sound more like /ba/ without the chirp. The second form is similar to the F2 transition heard in isolation—a nonspeech chirp.

Chirp fusion is more complex than previous fusions we have considered. Pilot studies indicate that many of the energy characteristics of the /da/ chirp may be different from those of the remainder of the /da/ syllable. For example, the two stimuli may have different pitches and fusion rates appear to be unattenuated. Relative differences of 20 Hz appear to have no effect on fusion response levels. Relative intensity differences also do not affect chirp fusion; chirp fusion occurs even if the chirp stimulus is decreased by as much as 30 db relative to the "chirpless" /da/. The chirp in this case is only about

---

4 The material gathered for this section has come from numerous discussions with Tim Rand.
1/1000 as loud as the "chirpless" /da/. This intensity difference is more impressive when one considers that the same chirp (at ~30 db), when electrically mixed with the chirpless /da/ and presented monaurally (instead of dichotically), still sounds more like /ba/ than /da/. Thus, for these stimuli, the chirp is a more potent speech cue when presented to the opposite ear than when presented to the same ear concatenated onto the chirpless stimulus.

Large differences in intensity or pitch between the two stimuli have no apparent effect on fusion rate in chirp fusion. This fact, coupled with "hearing" the chirp in both a speech and nonspeech form suggests that chirp fusion is a more central, higher-level process than the four types of fusion previously discussed. There is yet another dimension which distinguishes it from the lower-level fusions—timing. In chirp fusion the chirp and the chirpless stimulus need not begin at the same time. Rand has done pilot work which indicates that the relative onsets of the two stimuli may differ by as much as ± 25 msec. In other words, the chirp stimulus can begin 25 msec before the onset of the chirpless /da/, the two stimuli may be simultaneous with respect to their relative onsets, or the chirp can begin 25 msec after the onset of the chirpless /da/. The result for all three cases appears to be the same: the subject hears /da/ plus a chirp. When relative onsets of greater than 25 msec are used, fusion breaks down and the subject begins to hear /ba/ plus a chirp. Nevertheless, relative onset differences tolerated in chirp fusion are much greater than those permissible in the lower-level fusions.

In both phonetic feature fusion (fusion type 4) and chirp fusion (fusion type 5) the features of the two stimuli are combined to form a new speech unit. In phonetic feature fusion, place and voicing information is extracted from separate sources. In chirp fusion, manner and voicing cues in the chirpless stimulus are combined with the place feature extracted from the chirp. Thus, both fusions operate on the level of a single phoneme. By contrast, the next type of fusion to be considered deals with the combination of phonemes into a cluster.

6. PHONOLOGICAL FUSION: Perceptual construction of phoneme clusters.

Day (1968) was first to discover that compatible stop + liquid strings could fuse into one unit: given BANKET and LANKET presented to opposite ears, the subject often hears BLANKET (BANKET/LANKET→ BLANKET). One of the unique aspects of phonological fusion is that, unlike psychoacoustic fusion (fusion type 3) and phonetic feature fusion (fusion type 4), two stimuli which contain entirely different segments are presented at the same time, and yet there is no information loss. The segments of both stimuli are combined to form a new percept which is longer and more linguistically complex than either of the two inputs. The sixth display of Figure 1 shows a sample phonological fusion: the inputs /da/ and /ra/ often yield the fusion /dra/. This fusion contains all the linguistic information available in both stimuli. Thus, unlike other fusions, there is no phonological rivalry between the two stimuli.

5 The arrow (→) should be read as "yields."
Note that there is a preferred order of phonemes in the fusion response for these stimuli. Given BANKET/LANKET the subject almost never reports LBANKET. Instead, all of his fused responses are BLANKET. This finding appears to be based on the phonological constraints of English: initial liquid + stop clusters are not allowed. Day (1970) has shown that when such constraints are removed, fusion occurs in both directions. Given the stimuli TASS/TACK, subjects give both TASK and TACKS responses.

Day (1968) has shown that phonological fusion is not a function of response biases for acceptable English words. Both a stop stimulus and a liquid stimulus must be present for a stop + liquid fusion to occur. If one presents different productions of BANKET to either ear, the subject reports hearing BANKET (BANKET/BANKET→ BANKET). That is, the subject does not report the acceptable English word that corresponds most closely to the nonword inputs. Likewise, LANKET/LANKET→ LANKET. However, if the stimuli are BANKET/LANKET, the subject often reports hearing BLANKET regardless of which stimulus is presented to which ear. Furthermore Day (in preparation-a) has shown that even if the subject is informed as to the nature of the stimuli, he still fuses: he perceives BLACK both when the stimuli are BACK/LACK and when they are BLACK/BLACK.

Although there are undoubtedly peripheral elements involved in phonological fusion, various studies suggest that it is primarily a central, higher-level cognitive process. Phonological fusion is insensitive to large energy differences between the dichotic stimuli. The stimuli PAY and LAY fuse to PLAY regardless of their pitch. Relative differences of 20 Hz in fundamental frequency have no effect on the characteristics of fusion responses (Cutting, in preparation-a). Intensity differences also have no effect: one stimulus may be 15 db louder than the other and fusion rates do not change (Cutting, in preparation-a).

Phonological fusion is also insensitive to gross differences in the relative onsets of the two stimuli. Day (in preparation-b) has shown that the stimulus onsets may be staggered by as much as 150 msec and fusion rates do not change substantially. Note that these relative onset times are considerably longer than those permissible in any other fusion. Fusion occurs if the stimuli are simultaneous, if the stop stimulus (BANKET) begins 150 msec before the liquid stimulus (LANKET), or even if the liquid begins 150 msec before the stop. Thus, within a wide range of relative onsets, the actual ordering of the phonemes in real time appears to have little effect on fusion rate.

Phonological fusion also appears to be insensitive to changes in dimensions other than pitch, intensity, and timing. Subjects fuse whether the stimuli were uttered by the same vocal tract or by vocal tracts of different sizes (Cutting, in preparation-a). For example, PAY/LAY→ PLAY even if PAY has been synthesized to resemble the utterance of a normal adult male, and LAY has been synthesized to resemble the utterance of a midget or small child.

Phonological fusion occurs most readily when the same vowel follows both the initial stop and the initial liquid. Nevertheless, while PAY/LAY→ PLAY and GO/LOW→ GLOW, the pairing of PAY and LOW can yield PLO or PLAY. Fusion rates are reduced here, but are still at a fairly substantial level (Cutting, in preparation-b). In fact, fusion even occurs with stimuli that have almost no phonemes in common. Day (1968) has shown that two stimuli as different as
BUILDING and LETTER can "fuse" into BILTER, LILTER, or even BLITTERING; such cases involve phonemic exchanges between the two stimuli.

One distinction which appears to be unique to phonological fusion is that not all subjects fuse. In the five types of fusion previously discussed it appears that all subjects fuse equally readily. In phonological fusion, on the other hand, using natural speech stimuli, Day (1969) has shown that some subjects fuse on nearly all trials, while others fuse only occasionally, if at all. Moreover, few subjects score between these extremes. Cutting (in preparation-a), using synthetic speech stimuli, has also found a bimodal distribution of subjects with respect to their fusion rates. Large individual differences are found in many higher-level processes. For example, Turvey (in press) has shown in vision research that individual differences are larger for central masking than for peripheral masking. It appears that in the case of phonological fusion we have a task which is complex enough so that alternative modes of processing are possible. Studies are now underway to explore the possibility that groups of people who perform differently on the fusion task may also retain their group identity on other auditory and visual tasks (Day, in preparation-c).

Finally, phonological fusion appears to have certain linguistic constraints that no other fusion has. For example, some consonant + liquid stimuli fuse more readily than others. Day (1968) has shown that stop + /l/ stimuli fuse more readily than stop + /r/ stimuli. Day (1968) and Cutting and Day (1972) have shown that stop + /r/ stimuli often elicit a stop + /l/ response, whereas the reverse situation rarely occurs. Thus, PAY/RAY may yield PLAY. Cutting (in preparation-a) has found that stop + liquid stimuli fuse more readily than fricative + liquid stimuli: BED/LED→ BLED more readily than FED/LED→ FLED. These findings cannot be accounted for by the relative frequency of occurrence of these clusters in English. In fact, frequency data show the reverse trends: stop + /r/ clusters outnumber stop + /l/ clusters (Day, 1968) and /f/ + liquid clusters outnumber most other consonant + liquid clusters (Cutting, in preparation-a).

Phonological fusion appears to be the highest level process considered in this paper. Like other higher-level fusions it is insensitive to large stimulus differences in pitch, intensity, and timing. Other dimensions may also be varied with little effect, such as vocal tract size and vowel context. Certain variables, however, do affect fusion rate. Day (1968) has shown that semantics at the word level is one such variable. Fusion occurs most readily when the fused percept is a meaningful word, although nonword fusions do occur (e.g., GORIGIN/LORIGIN→ GORIGIN). Cutting (in preparation-a) has shown that semantics at the sentence level can also influence fusion. Fusion rates are higher when the fusible pair appears in a sentence context. For example, PAY/LAY→ PLAY more readily when the stimuli are THE TRUMPETER PAYS FOR US and THE TRUMPETER LAYS FOR US than when PAY and LAY are presented as an isolated pair.

**REVIEW OF FUSIONS**

We have looked at six phenomena in which stimuli are sent separately to each ear and the subject is asked to report what he heard. The effect of changing various parameters between the two inputs is summarized in Table 1.
TABLE 1: Dimensions which are relevant for the separation of lower- and higher-level fusions. Tolerances are listed within each cell. Specific numbers reflect current knowledge.

<table>
<thead>
<tr>
<th>LOWER-LEVEL FUSIONS</th>
<th>PITCH</th>
<th>INTENSITY</th>
<th>TIMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Sound Localization</td>
<td>&lt;2 Hz</td>
<td>&lt;2 db</td>
<td>&lt;2 msec</td>
</tr>
<tr>
<td>2. Spectral Fusion</td>
<td>&lt;2 Hz</td>
<td>*</td>
<td>&lt;5 msec</td>
</tr>
<tr>
<td>3. Psychoacoustic Fusion</td>
<td>&lt;2 Hz</td>
<td>*</td>
<td>*</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>HIGHER-LEVEL FUSIONS</th>
<th>PITCH</th>
<th>INTENSITY</th>
<th>TIMING</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Phonetic Feature Fusion</td>
<td>20 Hz</td>
<td>30 db+</td>
<td>25 msec+</td>
</tr>
<tr>
<td>5. Chirp Fusion</td>
<td>20 Hz</td>
<td>30 db</td>
<td>25 msec</td>
</tr>
<tr>
<td>6. Phonological Fusion</td>
<td>20 Hz</td>
<td>15 db</td>
<td>150 msec</td>
</tr>
</tbody>
</table>

* systematic data not available
+ indirect evidence

1. **Sound localization** occurs for all audible sounds, speech and nonspeech. The first display of Figure 1 shows that when /da/ is presented to both ears at the same time, pitch, and intensity, the subject perceives one /da/ localized "in the center of the head," or at the midline. Pitch variations of 2 Hz can inhibit fusion, such that two stimuli will be heard. Intensity variations of 2 db are sufficient to change the locus of the fusion. Timing differences of 2 msec are sufficient to cause the fused percept to disintegrate into two elements.

2. **Spectral fusion** occurs for speech sounds and for complex nonspeech sounds. The second display in Figure 1 shows that when F1 of /da/ is presented to one ear and F2 to the other, the subject perceives the fused /da/. Pitch variations of 2 Hz can inhibit fusion. Timing differences of about 5 msec can inhibit the spectral fusion of metronome ticks.

3. **Psychoacoustic fusion** probably occurs for both speech sounds and non-speech sounds. We have considered only speech sounds. For example, in the third display of Figure 1, /ba/ is presented to one ear and /ga/ to the other at the same pitch. The resulting perception is /da/. Pitch differences of 2 Hz can inhibit fusion.

4. **Phonetic feature fusion** occurs for competing speech segments. In the fourth display we note that when /ba/ and /ta/ are presented at different pitches, the subject often reports hearing the "blend" /da/. Pitch does not appear to be an important parameter in this fusion; preliminary work suggests that differences of 20 Hz are easily tolerated and fusion rates are unattenuated. Intensity has not been systematically explored, but data from other sources suggest that differences of as much as 30 db will not inhibit fusion. Our knowledge concerning the effect of timing differences is also indirect. Relative onsets greater than 25 msec inhibit fusion.
5. Chirp fusion is demonstrated in the fifth display of Figure 1. The "chirpless" /da/ is presented to one ear and the /da/ chirp to the other. Subjects perceive the entire syllable /da/ plus a nonspeech chirp. Pilot work has shown that pitch differences of 20 Hz do not alter fusion rates. Intensity differences of 30 db also do not affect chirp fusion. Relative onset differences of 25 msec are tolerable in chirp fusion, but larger relative onset differences inhibit fusion.

6. Phonological fusion occurs for pairs of phonemes which can form clusters, for example BANKET/LANKET → BLANKET. Display six shows that when /da/ is presented to one ear and /ra/ to the other, the subject often perceives /dra/. Large differences in pitch (20 Hz) and in intensity (15 db) do not appear to affect fusion rate. Gross differences in timing also appear to have no effect; differences of as much as 150 msec in the to-be-fused stimuli do not appear to alter the rate of fusion for stimuli such as BANKET/LANKET. Variations in vocal tract size are also tolerated. Certain linguistic variables, however, do influence fusion rate: two such variables are the types of consonant and liquid phonemes involved in the fusion, and the semantic context in which the stimuli appear.

CONCLUSION

There are at least six different types of fusion in auditory perception. They have been compared and contrasted with respect to three primary parameters: pitch, intensity, and timing. The first three fusions discussed (sound localization, spectral fusion, and psychoacoustic fusion) are sensitive to small changes in any of these parameters. Sensitivity to small differences in stimulus energy and stimulus timing has been noted to be a property of peripheral mechanisms (Turvey, in press). Thus, for the purposes of this paper, these three fusions are considered lower-level, peripheral processes.

In contrast, the other three fusions (phonetic feature fusion, chirp fusion, and phonological fusion) appear to be higher-level processes. In general, these fusions are insensitive to large stimulus differences in pitch, intensity, and timing. Since relative insensitivity to stimulus energy and stimulus timing has been noted as a property of central mechanisms (Turvey, in press) these three fusions may be considered higher-level, central processes.

The six fusions run the gamut from primitive to highly complex levels of processing. Both man and animals can localize sound, a very low level of fusion. Phonological fusion, at the other end of the fusion continuum considered here, appears to be a situation complex enough to allow alternative modes of processing: some subjects fuse the stimuli presented to opposite ears and give a single linguistic response, while others do not. All six processes that we have considered are fusions: the subject combines two signals into a single percept. Yet they are clearly different in many ways. Therefore, we have described, compared, and contrasted the various kinds of fusion so that more precise experimental questions can be posed in order to unravel the processes involved.

REFERENCES


Cutting, J. E. (in preparation-b) Phonological fusion of "incompatible" stimuli.


