Abstract. When an [s] or [ɣ] fricative noise is combined with vocalic formant transitions appropriate to a different fricative, the resulting consonantal percept is usually that of the noise. To see if the mismatch affects processing time, five experiments were run. Three experiments examined reaction time for identification of [s] and [ɣ], as well as the whole syllable (in one experiment) or only the vowel (in the others). The stimuli contained either appropriate or inappropriate formant transitions, and the vowel information in the noise was either appropriate or not. Subjects were significantly slower in all tasks in identifying stimuli with inappropriate transitions or inappropriate vowel information. Similar results were obtained with stop-vowel syllables in which the release bursts of syllable initial [p] and [k] were transposed in syllables containing the vowels [a] and [u]. In the fifth experiment, enough silence was introduced between the initial fricatives and vocalic segment for the vocalic formant transitions to be perceived as a stop (e.g., [stu] from [su]). Mismatched transitions then had a much reduced effect on reaction time, while mismatches of vowel quality slowed identification even more. The results indicate that listeners take into account all available cues, even when the phonetic judgment seems to be based on only some of the cues.

INTRODUCTION

It is well known that information about a phone is temporally spread in the speech signal. It is usually impossible to isolate one piece of the signal and identify it as one single phone. Even when such a segmentation results in a stretch of sound that is identifiable as a single phone, information about neighboring phones usually remains. The vowels of consonant vowel syllables, for example, can be identified at better than chance levels from the excised stop consonant release bursts (Blumstein & Stevens, 1980; Kewley-Port, 1960; LaRiviere, Winitz, & Herriman, 1975b) or from the excised fricative noises (LaRiviere, Winitz, & Herriman, 1975a; Yeni-Komshian & Soli, 1981).

The vowel information in stop bursts and frications is quite weak. This is evident in our saying that these vowels can be identified at a "better than chance" level. If the percept were strong, the vowel would be as easy to

Acknowledgment. I would like to thank Alvin M. Liberman, Bruno Repp, Michael Studdert-Kennedy, Harriet Magen, and Sharon Manuel for comments on this paper. The research and writing were supported by NICHD Grant HD-01994.

identify from the part as from the whole syllable. There is not that much information available. Rather than constructing a vowel percept, the subject can infer what vowel must have been present.

The vowel information in a stop release burst is also not a strong enough vowel cue to override the information in the vocalic segment. If a release burst from [pa], for example, is replaced with one from [pu], our perception of the vowel does not change, although there is vowel information in the burst. An artificial mismatch of that sort, in which a cue is put in a new environment in which its cue value is not sufficient to change the phonetic percept, will be called a subcategorical phonetic mismatch. The cue that gets overridden in that way will be called a mismatched cue. There are three ways a listener can treat a mismatched cue: 1) she can reject it, so that a nonspeech click, pop, whistle, etc., is perceived in addition to the speech; 2) she can integrate it with the overriding cue in such a way that within-category variation is perceived (as could be determined with a discrimination test); 3) or she can ignore it. The experiments described in this paper will show that mismatched cues impose a processing load. Thus the "act of ignoring" a cue (or possibly within-category variation) takes time. This supports the notion that listeners are sensitive to all the information they gather and attempt to incorporate it into the percept.

Note that in order to know whether to accept or reject a mismatched cue, the listener must know what a possible speech sound is. If she treats the cue as non-linguistic noise, it must be because she could not make linguistic sense of the auditory pattern. In extreme cases, there may be gross auditory discontinuities. Mismatched cues, in similar but appropriate contexts, can be integrated. Thus it is not sufficient to say that mismatched cues are not speech-like; given the proper environment, they are quite natural and provide phonetic information appropriate to the speech sounds they were originally produced with. It requires a complete knowledge of phonetic possibilities to know whether a cue is in its appropriate environment or not.

Two kinds of mismatched cues were studied in the present experiments: 1) vowel information in fricative noises and stop consonant release bursts, and 2) the place of articulation information in stop bursts and in vocalic formant transitions of vocalic segments occurring with fricatives. The information about a fricative's place of articulation in formant transitions has been shown to influence phonetic identification when the friction cue is ambiguous (Harris, 1958; Mann & Papp, 1980; Whalen, 1981). Unambiguous fricative noises, on the other hand, seem to override mismatched transitions completely in following vocalic segments. The perception of vowels following frications that were originally produced with other vowels is similarly unaffected by that mismatched information.

A similar situation sometimes occurs with syllable-initial stops. If we exchange release bursts from stops produced at different places of articulation, the bursts often determine the place of the resulting stop percept. Other times, however, the transitions will be the deciding cue. Sometimes the perceived place will be different from both that cued by the burst and that cued by the transitions (Fischer-Jörgensen, 1972). (Unlike the fricative noises, no stop burst, it seems, provides an unambiguous cue to place in all vocalic contexts; cf. Blumstein & Stevens, 1980; Dorman, Studdert-Kennedy, &
Raphael, 1977). Yet another parallel occurs with medial stops. If the transitions into and out of medial stops conflict, the second (opening) set usually determines the percept with no audible contribution of the closure transitions (Dorman, Raphael, Liberman, & Repp, 1975; Fujimura, Macchi, & Streeter, 1976). Stimuli with such conflicting transitions are difficult to discriminate from stimuli with matched transitions (Repp, 1977).

In many stimuli with mismatched cues, then, no overt ambiguity results, and the mismatch escapes conscious detection. However, it could be that the assignment to a phonetic category was in fact slower when some cue or another was inappropriate. Delays in identification have been shown in stimuli with overt ambiguities (Fisoni & Tash, 1974; Repp, 1981b). An alternative view hypothesizes that the listener's perceptual system would treat the overriding cue for a phone as sufficient and ignore the "subcategorical" mismatches completely. In this case, a listener would be able to identify, say, an alveolar fricative equally fast whether the transitions of the vocalic segment it occurred with were appropriate or not.

The first view presumes that the perceptual mechanism tries to include the phonetic value of each cue in the percept, whether that cue is strictly necessary to the identification or not. The latter view presumes that the perceptual system attempts to make a justifiable phonetic assignment as soon as possible (as in Blumstein & Stevens, 1980; Cole & Scott, 1974; Klatt, 1979; Stevens, 1975). The former proposal will be called the "integrating" account, since the proposed mechanism attempts to integrate (over time and frequency) all information reaching it into a unified percept (see Liberman, 1979; Liberman & Studdert-Kennedy, 1978; and Repp, 1982, for recent reviews of the relevant literature). The latter will be called the "disposing" account, since its mechanism attempts to dispose of each portion of the speech signal (by passing a phonetic judgment on to another part of the system) as it is received.

Consider first the case of mismatched cues that precede the overriding cue in the speech signal. Several studies have shown that such mismatches slow decision time. Subcategorical mismatches of transitions into medial stop resulted in slower decision times in a speeded lexical decision task (Streeter & Nigro, 1979). (The effect only appeared for words, not for nonwords.) Martin and Bunnell (1981) have shown that identification of final [i] and [u] are slowed when a preceding fricative or fricative-stop cluster was originally produced before the other vowel. Later studies (Martin & Bunnell, 1982) examined vowel to vowel coarticulation with similar results.

The integrating account does not need any additions to explain these results. A listener need only notice that conflicting cues are present, and she will attempt to integrate them into the phonetic percept. That these cues can provide information is shown by their determining the percept when the (normally) overriding cue is ambiguous. The disposing account can, with some additions, also explain the stop data by assuming that a phonetic decision is made on the basis of the closure transitions, but that the decision is not firm enough to allow it to generate the phonetic percept. When the opening transitions conflict with the decision based on the closure transitions, it would presumably take some extra time to set up another phone as the percept. The mechanism of the disposing account must also generate a (preliminary)
vowel percept based on the friction (to account for Martin & Bunnell's, 1982, data).

The situation that distinguishes these theories occurs when the conflicting cues follow the overriding cue. The integrating account predicts that such cues will be as slowing as those that precede the overriding cue. An initial fricative followed by inappropriate transitions should give longer identification times. The disposing account, on the other hand, predicts no delay due to following misinformation, since the correct decision would already have been made.

Figure 1 is a comparison of the predictions of the disposing and integrating accounts. When the mismatched cues precede the overriding cue, both theories predict that mismatches will slow response time. The disposing account assumes that the identification will take longer to reach criterion level, while the integrating account assumes that the integration of conflicting information takes longer than integrating compatible information. (The figure is oversimplified by assuming that integration does not begin until all the cues have been received; this is done for convenience of display only.) When the mismatched cue follows the overriding cue, the disposing theory predicts identical times for both matched and mismatched versions of the stimuli, while the integrating account predicts a delay for mismatches.

The present paper reports five experiments examining speeded identification of fricatives, vowels, stops, and whole syllables with and without mismatched cues. In the first experiment, the overriding cue came after the conflicting cue. This will confirm the other results mentioned above. For three of them, however, the overriding cue came before the conflicting cue. The integrating account predicts a delay, while none is predicted by the disposing account. In the last experiment, the transitions of the fricative-vowel syllables were allowed to affiliate with a different phone (i.e., a stop) by inserting silence between the noise and the vocalic segment. The integrating account predicts a reduction in the effect of mismatches here, while the disposing account still predicts no effect.

**EXPERIMENT 1**

**Experimental Procedure**

Materials. A male native speaker of English recorded ten tokens of each of the syllables [æs], [aθ], [is], [iθ], [oʊ], [oʊ], [uθ] and [uθ] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that the vocalic portion of all eight were of equal duration, the friction of all eight were of equal duration, and, of course, the original syllables and all combined syllables were also of equal duration. All judgments were thus given to stimuli of equal duration. A vocalic segment duration of 200 msec was found naturally in eight syllables. Seven were shortened by cutting off between 10 and 50 msec from the onset of the vowel; the resulting abruptness did not sound unnatural. The eighth vocalic portion was lengthened 20 msec by repeating its first pitch pulse three times. The frictions were 250 msec in duration; nine were shortened by removing between 10 and 50 msec from near the end of the signal.
R ✓ Time for initiation of response to stimulus with matched cue
Rx  Time for initiation of response to stimulus with mismatched cue

Mismatched Cue Precedes Overriding Cue
Mismatched Cue Follows Overriding Cue

Disposing Account

\( t \) \hspace{1cm} R ✓ \hspace{1cm} Rx

\( (s) \) criterion

\( \circ \) information added by one time slice, matched cue stimulus
\( \bullet \) information added by one time slice, mismatched cue stimulus

 Integrating Account

\( t \) \hspace{1cm} R ✓ \hspace{1cm} Rx

\( (s) \)

\( \circ \circ \) accumulation of information, matched cue stimulus
\( \bullet \bullet \) accumulation of information, mismatched cue stimulus
\( --- \) integration time

Figure 1. Comparison of the predictions of the disposing and integrating accounts for preceding and following mismatched cues.
Once the tokens had been selected and the durations equalized, each friction was combined with each vocalic segment, including the one it was originally produced with. The resulting 256 stimuli fell into four categories of interest: 1) The vocalic formant transitions had been produced with the same fricate as the percept generated by the noise ("appropriate transitions") and the vowel was the same as the vowel the fricate had originally been produced with ("appropriate vowel"); 2) The transitions were appropriate but the vowel was inappropriate; 3) The vowel was inappropriate but the transitions were appropriate; and 4) Both the transitions and the vowel were inappropriate.

Some mismatches of vowel and the vowel information in the friction gave rise to perceived [i] or [u] offglides on the vowel (as detailed in Whalen, 1982). Thus there is a mixture of cue status here; some are mismatched, and some are reinterpreted as an added phone. Whalen (1982) showed that the transitions did not contribute to the diphthong percepts. Thus the mismatched transitions are clearly subcategorical mismatches. The effect of mismatched vowel quality was not as readily attributable to subcategorical mismatches, since not all of the vowel quality cues were ignored.

Each session consisted of four blocks of stimuli. Each block contained 128 trials, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). One token of each stimulus occurred once within the first two blocks, and once within the second two; the order was otherwise random. The stimuli were recorded on one channel of an audiotape while, on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The inter-stimulus interval was 2500 msec.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were phonetically trained. Two were left-handed. The naive subjects were 10 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. One was left-handed.

Apparatus. Subjects were seated in a quiet room and heard the stimuli over Telephonics TDH-39 headphones. Their responses were made by pressing one of two buttons on a panel in front of them. The "s" response was on the left and the "sh" response on the right. During the test, if the answer was correct and within a predefined time limit (longer than 100 msec and shorter than one second), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify the fricative as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. The feedback light was explained to them. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-second pause between, followed by a short break. The next two blocks, separated by a thirty-second pause, finished the session.
Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long (over one second) or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 3.4%.

As can be seen from Figure 2, inappropriateness of transition slowed the subjects' identifications, \( F(1,18)=93.225, p < .001 \). The four bars of the graph show mean identification time, respectively, from left to right, for 1) the syllables in which both transition and vowel were matched, 2) those where the transition was mismatched but the vowel was matched, 3) those where the transition was matched but the vowel was mismatched, and 4) those syllables where both transition and vowel were mismatched. On average, subjects were 24 msec faster in their decision when the transition was appropriate (means of 516 msec vs. 540). The inappropriateness of the vowel also slowed the identification times, \( F(3,54)=5.494, p < .01 \), by an average of 9 msec. The effect of appropriateness of transition is seen in the difference between the first two bars as well as the difference between the second two. The effect of appropriateness of vowel is seen in the comparison of the first and third bars and of the second and fourth bars. Further, these two effects were independent, \( F(3,54)=0.918, n.s. \), for the interaction.

The experts were significantly faster than the naive subjects, \( F(1,18)=5.446, p < .05 \). The means were 528 and 588 msec, respectively (measured from the onset of the vowel). The interactions with the two appropriateness factors were not significant, though, indicating that the effects are independent of linguistic sophistication.

The vowels were chosen to contrast in rounding (/o,u/ vs. /a,i/) and (relative) height (/i,u/ vs. /a,o/). Therefore a second analysis was performed in which the appropriate vowel factor was split into appropriate height (where the height of the vowel matched the height of the vowel that the fricative was originally produced with) and appropriate rounding. Appropriate rounding was significant as a main effect, \( F(1,18)=4.625, p < .05 \), but appropriate height was not, \( F(1,18)=2.076, n.s. \). Appropriateness of the transition did not interact with the appropriateness of the vowel for either rounding or height, \( F(1,18)=1.696, 1.129 \). The two types of vowel appropriateness did interact with each other, \( F(1,18)=17.846, p < .001 \). The syllables in which both vowel features were appropriate were identified faster than those where one or both were mismatched. Further work is needed to determine the limits of vowel information in fricative noise; the current results simply show that it is there.

Discussion

The strongest effect from the first experiment is that inappropriate vocaical formant transitions slow identification of a following fricative. While this result makes sense, it is perhaps a bit unexpected. One might assume, as did Cole and Scott (1973, p. 448), that the transitions serve only to keep the fricative noise from "streaming" off and sounding like nonspeech. If the transitions are only an auditory event that leads the hearer to expect a fricative, then any transitions should do. Thus the listener could ignore

173
TIME REQUIRED TO IDENTIFY FINAL FRICATIVE

- Transition and Vowel matched
- Transition mismatched, Vowel matched
- Transition matched, Vowel mismatched
- Transition and Vowel mismatched

Figure 2. Times to identify the fricative as [s] or [ß], Experiment 1.
the place information in the transitions. If this "auditory" integration were sensitive to the place of the fricative, then the transitions would in fact be giving place information and thus be a cue. The present results indicate that, indeed, place information in the transitions is taken into account even when it is overridden by the more salient friction cue.

The vowel effect is less surprising and can be interpreted in terms of coarticulation. We would expect, on articulatory grounds, that rounded vowels would have a large effect on the spectrum of the friction. Studies of vowel information in frications have shown this consistently (LaRiviere et al., 1975a; Yeni-Komshian & Soli, 1981; cf. Whalen, 1982). In the present results, mismatches in rounding did indeed slow identification, while mismatches in height did not. This result must be qualified, however, since the differences in height were not as systematic as those of rounding.

In general, sub-categorical phonetic mismatches can slow identification. The next experiment was designed to see if subjects could avoid such delays when the fricative occurred first in the utterance, that is, when the overriding cue for the fricative preceded the mismatched cue.

**EXPERIMENT 2**

**Experimental Procedure**

**Materials.** A male native speaker of English recorded ten tokens of each of the syllables [sa], [ša], [su] and [šu] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that the friction would be equally long in all eight. A duration of 180 msec was found naturally in seven syllables; the eighth was produced by removing 50 msec from a token with a longer friction duration. The vocalic segments varied in duration, ranging from 255 to 221 msec for [a] and 225 to 188 msec for [u].

One other manipulation was carried out on the stimuli in an attempt to see if the subjects were categorizing the fricative on the basis of the fricative noise alone. Since the noise is the overriding cue, a fricative judgment could be made on it alone. If subjects make their decision rapidly enough, then shortening the friction should have no effect on the reaction time. Since the initial portion of the noise unambiguously specifies the fricative, the response can be initiated without waiting for the vocalic segment. Alternatively, if reaction times vary with the duration of the friction, this would indicate that subjects wait at least until the start of the vocalic segment before initiating their response. A shortened version of each friction was made by excising 50 msec from the middle of the noise. This left the onset and offset amplitudes intact. This procedure caused no audible discontinuity and generated no affricate percepts.

To make sure that there would be occasions on which the subjects would be forced to wait for the vocalic segment before responding, two conditions were run. In the first, only the fricative was identified; in the second, the whole syllable. When identifying the whole syllable, the subjects must wait
for the vocalic segment to occur before they can make their judgment. We can then tell whether inappropriate cues have an effect in all cases, only when the conflicting cues must be waited for, or never.

Once the tokens had been selected and the shortened frications made, each friction was combined with each vocalic segment. This gave 2 (short vs. long friction) x 2 ([s] vs. [ɛ]) x 2 ([a] vs. [u]) x 2 (vowel that the friction was produced with is appropriate to the vowel in the combined syllable vs. inappropriate vowel) x 2 (vocalic formant transitions are appropriate to the friction vs. inappropriate transitions) x 2 (tokens of each vocalic segment) x 2 (tokens of each friction) = 128 stimuli.

Each session consisted of four blocks of stimuli. Each block contained one repetition of each of the 128 stimuli, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). The stimuli were randomized within blocks. Test stimuli were recorded on one channel of an audiotape, while a timing tone was recorded simultaneously on the other channel. The inter-stimulus interval was 2500 msec.

Subjects. The subjects were 20 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. Ten were the naive subjects from Experiment 1. Three were left-handed.

Apparatus. Subjects were seated in a sound-attenuated booth and heard the stimuli over TDH-39 headphones. Their responses were made by pressing one of 2 (condition 1) or 4 (condition 2) buttons on a panel in front of them. In condition 1, the "s" response was on the left and the "sh" response on the right. In condition 2, the "sa" and "sha" responses were on the left, with "sa" being directly above "sha." The "su" and "shu" buttons were arranged similarly on the right. During the test, if the answer was correct and within the stated time limit (longer than 100 msec), and shorter than one second (for condition 1) or one and a half seconds (for condition 2), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify either the fricative (condition 1) or the whole syllable (condition 2) as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-second pause between, followed by a short break. The next two blocks, separated by a thirty-second pause, finished the session.

To see if familiarity with the task made it easier to judge the friction alone, half the subjects were given the four-choice condition (condition 2) first, and half had the two-choice condition first. In each group, half the subjects had participated in Experiment 1 and half had not.
Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long (over one or one and a half seconds) or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 4.7%.

Figure 3 shows the results. The left half shows the results for the condition in which only the fricative was identified; the right half shows the results for the identification of the whole syllable. The four bars of each half show mean identification time (collapsed across original and shortened frictions), respectively, from left to right, for the syllables 1) in which both transition and vowel were matched, 2) those where the transition was mismatched but the vowel was matched, 3) those where the transition was matched but the vowel was mismatched, and 4) those syllables where both transition and vowel were mismatched. The effect of appropriateness of transition, then, is seen in the difference between the first two bars as well as the difference between the second two. The effect of appropriateness of vowel is seen in the comparison of the first and third bars and of the second and fourth bars.

Across conditions, inappropriate transitions significantly slowed identification by 11 msec, $F(1,16)=12.97, p < .01$. The appropriateness of the vowel to the friction was even more significant, $F(1,16)=52.24, p < .001$, with a delay of 20 msec for inappropriateness. The inappropriateness of the vowel slowed responses more (by 27 msec to 14) when the transitions were inappropriate, $F(1,16)=8.01, p < .05$. The difference between the two conditions was highly significant, $F(1,16)=105.05, p < .001$. Since this compared a two-choice test with a four-choice one, the difference is no surprise.

The results for shortened versus original frictions, collapsed over appropriateness of vowel, are shown in Figure 4. (The results with the vowel mismatched were in accordance with the predictions.) The first two columns of each half represent the times for the syllables with the original frictions; the next two, those with the shortened frictions. The first columns of each of those pairs represent the syllables with appropriate transitions, the second, those with inappropriate transitions. Syllables with shortened frictions were identified faster than the originals overall by an average of 33 msec, $F(1,16)=204.05, p < .001$. Still, the speed advantage of the shortened stimuli was significantly larger in the whole syllable condition than in the fricative condition, $F(1,16)=60.04, p < .001$: The shortened frictions resulted in a 46 msec gain in reaction time when the whole syllable was identified, but only 19 msec when the fricative was identified.

These main results conform to the predictions. In the results for the identification of the whole syllable, however, there was one anomaly. The syllables with inappropriate transitions but appropriate vowels were identified faster than the syllables with both transition and vowel appropriate (see Figure 3). This did not result in a significant interaction between condition and appropriateness of transition, $F(1,16)=1.26$, n.s. However, the triple interaction of condition and appropriateness of vowel and of transition was significant, $F(1,16)=8.75, p < .01$. In the whole syllable condition, inappropriateness of the transition slowed identification only if the vowel was inappropriate as well. This unexpected behavior also contributed to the
TIMES FOR INITIAL FRICATIVES AND WHOLE SYLLABLE

- Transition and Vowel matched
- Transition mismatched, Vowel matched
- Transition matched, Vowel mismatched
- Transition and Vowel mismatched

Figure 3. Times to identify the fricative or the whole syllable, Experiment 2.
Figure 4. Times to identify the fricative or the whole syllable, shortened vs. original stimuli, Experiment 2.
interaction of appropriateness of vowels and condition, $F(1,16)=22.92$, $p < .01$. The delay for syllables with inappropriate vowels was 30 msec when the whole syllable was identified, compared with only 11 when just the fricative was judged.

A further set of interactions reveals that the anomaly is limited to the syllables containing shortened frications (see Figure 4). In the fricative condition, inappropriate transitions slowed responses both for the original and the shortened frications. In the syllable condition, however, making the transitions inappropriate actually speeded the decision 3 msec with the shortened friction; the syllables with the original friction showed the expected pattern, $F(1,16)=11.55$, $p < .01$. Even across conditions, appropriateness of transition and shortened friction interacted. When the transitions were appropriate, there was less of an advantage for having the short friction (26 msec compared with 39), $F(1,16)=15.46$, $p < .01$. The same held for appropriateness of the vowel (26 msec vs. 41), $F(1,16)=9.35$, $p < .01$. There was a further interaction of condition and appropriateness of vowel and of transition with length, $F(1,16)=5.71$, $p < .05$. In sum, there was one group of stimuli, the syllables with shortened frications and inappropriate transitions, that behaved unexpectedly when the whole syllable was identified.

Neither prior experience nor order of conditions had a significant effect on reaction time, $F(1,16)=0.29$, 0.075, respectively, n.s. The interaction was not significant either, $F(1,16)=0.65$, n.s. These two variables interacted with the conditions variable, $F(1,16)=7.00$, $p < .05$. No natural explanation for the interaction is obvious. More important is the lack of any interaction with the two appropriateness factors.

**Discussion**

Once again, mismatching the transitions, while it did not change the phonetic identity of the fricative, did slow identification—in this case, of both the fricative and the syllable the fricative was in.Mismatch of the vowel and the vowel that the fricative was originally produced with was a more significant factor in this experiment than in the previous one. In the four-choice condition, this is natural, since the information in the noise could be a partial cue to the identity of the vowel. Yet even in the two-choice condition, where the subject could, in principle, make her decision before she even hears the vowel, there is an effect. Further, the mismatched cues still slow the identification even though the overriding cue is heard first. Therefore the results support an "integrating" account and cast doubt on any "disposing" account. (See the General Discussion for a treatment of a disposing account with a large time window.)

If, in the two-choice condition, subjects were basing their decision about the fricative on the noise alone, we might expect the following three patterns to emerge: 1) Inappropriateness of transition would have an effect only in the four-choice condition, where the subject is required to listen to the whole syllable. 2) Similarly, inappropriateness of vowel would have an effect only in the four-choice condition. 3) In the two-choice condition, there would be no difference in response times for original and shortened frications. None of these expectations is fulfilled. However, there is a tendency in the direction of fulfilling the last two, so the following revision is worth considering: In the two-choice condition, subjects can
occasionally succeed in making their decision before the vocalic segment reaches them. In those cases, the judgment would be "unaffected" by the vocalic segment and the above mentioned expectations would hold. When the subject is not able to ignore the vocalic segment (is "affected" by it), the expectations do not hold; the result would be a mixture of responses in which the effects of conflicting cues are weakened in the two-choice condition. However, two major pieces of evidence conflict with this interpretation.

First, the transition effect is equally strong in the two-choice and four-choice conditions. That is, identification is slowed equally by mismatches in transition whether the whole syllable is identified or only the fricative. If subjects were basing their decision only on the noise, we would expect no effect of mismatched transitions when only the fricative was identified. For the transitions to have an effect, they must be heard. To be heard, at least the beginning of the vocalic segment must be heard. Thus even if the vowel itself was ignored, the 50 msec difference in time should have shown up, as it did in the whole-syllable condition. The difference, however, was only 19 msec.

Second, the higher level interactions show that the division of fricative identifications into "affected" and "unaffected" responses is not straightforward. The time advantage brought about by shortening the friction is quite suggestive: In the four-choice condition, the gained speed (46 msec) is almost equal to the cut in duration (50 msec). For the two-choice condition, the gain is only two-fifths of that (19 msec). This would lead us to expect that subjects could make their decision on the noise alone approximately three-fifths of the time. The discussion of the last paragraph casts doubt on this proportion; other interactions involving inappropriateness of vowel do the same. If decisions were either "affected" or "unaffected," then mismatched vowel and transition cues would either slow decisions equally (in the affected identifications) or be ignored together (in the unaffected cases). Thus there should be an interaction between inappropriateness of transition with condition and interaction between inappropriateness of vowel with condition, but no interaction of the three. In fact, the transition effect is unaffected by condition, the vowel effect is weaker in the identification of just the fricative, and the interaction of all three is significant. The interaction of inappropriateness of vowel and transition itself goes against any simple explanation of the effects of the mismatch.

It thus appears that, whatever the explanation of the effect of shortening the friction, subjects are not ignoring the vocalic segment in any of their judgments. This is not always the case, as is shown in Repp (1981a). In an experiment that tested only identifications of the fricatives [s] and [ʃ], Repp showed that inappropriate transitions did not affect reaction time. Shortening the noise by 50 msec resulted in a significant reduction in reaction time, but the difference was only 8 msec. The subjects in the present experiment may have been more inclined to pay attention to the vocalic segment since half of them participated in the four-choice (identification of whole syllable) condition before the two-choice (identification of fricative only) condition. In addition, some of Repp's subjects had recently participated in fricative discrimination studies, in which they had to concentrate on the spectrum of the noise. However, the lack of an effect of vocalic context
Subcategorial Phonetic Mismatches Slow Phonetic Judgments

does not fit well with the shortened reaction times for shortened frications, even if the difference is smaller.

Some of the interactions might lead to the following proposal: The most typical noise will give the fastest time in all environments. Repp (1981a) also had some evidence that this might be the case for [a]. The noise of [s] is high in frequency, and unrounded vowels result in higher noises for coarticulated fricatives. The converse holds for [z]. With the current stimuli, the [s] noise from [sa] is the most decidedly [s], and the [z] noise from [zu] is the most decidedly [z]. We might expect responses to those noises to be the shortest. For the present data, this is not the case, even when the identification of the fricatives alone is considered. Instead, the identification seems to be sensitive more to appropriateness than absolute typicality.

Many complicated factors seem to be involved in the perception of these modified stimuli. While the exact nature of these factors would require a series of tests manipulating the acoustic structure in a more detailed fashion, the main point is clear: Mismatch of cues results in a delay in identification. The next experiment will demonstrate this result with stops.

**EXPERIMENT 3**

Stop release bursts are in many ways equivalent to fricative noises. They are noises within limited frequencies, and they provide substantial consonant information and some vowel information. The third experiment of this series explores the behavior of mismatched burst cues. In this case the two mismatched cues were combined in one element, the burst, so that both the inappropriate vowel and consonantal information preceded the overriding cues in the transitions and the steady-state vocalic section.

The four-choice condition of the previous experiment, in which the whole syllable was identified, was replaced with one in which only the vowel was identified. Differences between the identification of the consonant and of the vowel would have a better chance of emerging if the different tasks were more similar. Also, the subject must still wait for the mismatched cues to occur before identifying the vowel, yet the task of choosing between two vowel categories is much easier than that of choosing among four syllable categories.

**Experimental Procedure**

Materials. A male native speaker of English recorded ten tokens of each of the syllables [pe], [pu], [ka], and [ku] on magnetic tape. These were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen, with the requirement that the release burst of the stop be 5 msec in duration. The burst was defined as a segment of noise with an amplitude rise and fall occurring before the aspirated formant transitions. The syllables were either 500 msec in duration (with [a]) or 350 msec (with [u]). All the [u]'s were of a much shorter duration, and there was no pressing need to have the stimuli of exactly the same duration, so the syllables were not modified.
Once the tokens had been selected, the bursts were isolated and then recombined with each vocalic segment. The vocalic formant transitions were the overriding cue in all cases for the experimenter. Some subjects complained of disagreement, especially in the [u] syllables. A non-speeded identification of the consonants was added to the experiment to assess the magnitude of the disagreement.

The mismatched cue, the burst, again came before the deciding cue, that is, the vocalic formant transitions. The resulting 64 stimuli fell into four categories similar to those that were of interest before: 1) The information in burst matched both the transitions and the vowel; 2) The vowel information matched but the stop information conflicted; 3) The stop information matched but the vowel information conflicted; 4) Both the vowel and the stop information in the burst conflicted with the transitions and vowel of the syllable.

Each session consisted of two conditions: judging the consonant and judging the vowel. Two blocks of stimuli occurred in each condition. Each block contained 128 trials, plus four "warm-up" stimuli at the beginning (which were not tallied in the results). Two tokens of each stimulus occurred within each block in random order. The stimuli were recorded on one channel of an audiotape while, on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The inter-stimulus interval was 2500 msec.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were phonetically trained. Eight had participated in Experiment 1. Two were left-handed. The naive subjects were 10 young adults, all native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation. Nine had participated in Experiments 1 and 2. One was left-handed.

Apparatus. Subjects were seated in a sound-attenuated booth and heard the stimuli over TDH-39 headphones. Their responses were made by pressing one of two buttons on a panel in front of them. In the consonant condition, the "p" response was on the left and the "k" response on the right. In the vowel condition, the "a" response was on the left and the "u" response on the right. During the test, if the answer was correct and within the stated time limit (longer than 100 msec and shorter than one and one half seconds for the consonant condition, shorter than one second for the vowel condition), a small light on the control box in front of them lit up. Their response time, answer, and the correctness of that answer went into a computer file after each trial.

Procedure

The subjects were instructed to identify the consonant or vowel as quickly as possible. They were told to expect a few mistakes, but to slow down if they made too many. Since subjects were not unanimous in their judgment of the stop identity, they were told to expect to disagree with the feedback in some instances. The feedback light was explained to them. Thirty stimuli were run but not scored to give them practice. After it had been determined that there were no questions, two blocks were run with a thirty-
second pause between, followed by a short break. The next condition, consisting of another two blocks separated by a thirty-second pause, finished the session. Order of the conditions was counterbalanced across subjects.

After the reaction time experiments were over, the first block was presented again for non-speeded identification of the consonants. These results were tallied separately from the speeded identifications.

Results

Only correct responses within the specified time limits were included in the analysis of the results. Thus responses that were too long or too short (under 100 msec) were counted as mistakes. This gave an overall error rate of 4.6%

Figure 5 shows the results in a way that is parallel to the previous results. The effect of the appropriateness of the transition was significant, $F(1,18)=7.679$, $p < .05$. On average, subjects were 4 msec faster in their decision when the transition was appropriate. The effect was only present when the consonant was identified. This is shown by the interaction of condition with appropriateness of transition, $F(1,18)=14.308$, $p < .01$. Inappropriate transitions slowed identification of the consonants (Condition 1) by 13 msec, but sped identification of the vowel by 3 msec.

Inappropriate vowels did not slow identification significantly, $F(1,18)=1.080$, n.s., despite a trend of 2 msec in that direction. Misidentifications of the stop may have obscured this result, so an analysis was done of the data for syllables containing the vowel [a]. The identification of the stops in these syllables was correct 99.4% of the time for all subjects. These results were analyzed in the same manner as the full test results. Inappropriateness of transition did not have any effect, $F(1,18)=0.40$, n.s., but inappropriateness of vowel did, $F(1,18)=6.99$, $p < .05$, for a delay of 7 msec.

The experts were significantly faster than the naive subjects, $F(1,18)=9.067$, $p < .01$. The means were 378 and 500 msec, respectively. This factor was involved in no significant interactions.

Results for the non-speeded identification of the consonants appear in Table 1. They are summarized as percentage of misidentifications of the consonants. Results are collapsed across consonant and vowel category, and are divided in the same manner as the results displayed in Figure 5. The rate of misidentification corresponds to increase in reaction time, but it is not certain that ambiguity in the stimuli is sufficient to account for the results. Four of the subjects accounted for 48.7% of the misidentifications. The other sixteen subjects were correct at least 94.5% of the time. A second analysis was done on the 10 subjects with the highest accuracy. There were no changes in the variables and interactions that were significant. However, the misidentifications still parallel the reaction times (see Table 1).
Figure 5. Times to identify the stop or the vowel, Experiment 3.
Table 1

Results of Consonant Identification Task

<table>
<thead>
<tr>
<th>Transition was matched</th>
<th>Vowel was matched</th>
<th>matched</th>
<th>mismatched</th>
<th>matched</th>
<th>mismatched</th>
</tr>
</thead>
<tbody>
<tr>
<td>% misidentification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>for all twenty subjects</td>
<td>1.6</td>
<td>7.0</td>
<td>4.1</td>
<td>6.4</td>
<td></td>
</tr>
<tr>
<td>for ten best subjects</td>
<td>0.3</td>
<td>3.4</td>
<td>1.9</td>
<td>2.5</td>
<td></td>
</tr>
</tbody>
</table>

Discussion

Overall, inappropriate consonantal information in the burst slowed reaction time. This effect, however, did not appear in the results for the syllables with [a]. Overall, making the vowel information in the stop burst inappropriate to the vowel does not slow identification of that stop. When the results for syllables with [a] are considered alone, however, inappropriateness of vowel does slow reaction time. While these results confirm the previous results for the fricatives to some extent, they must be treated with caution.

Since the bursts were necessarily chosen for their minimal place information, their lack of a slowing effect is not too surprising. The acoustic effects of the articulation on the burst are less clear than the effect on fricative noise, and the stop and transitions interact in complex ways. The stop can be identified to some extent from the burst alone (Kewley-Port, 1980; Tekiel & Cullinan, 1979; Winitz, Scheib, & Reed, 1971), but these bursts must contain less place information for this experimental design.

Vowels can be identified much better than chance from the friction of a coarticulated fricative by itself (Yeni-Komshian & Soli, 1981). The vowel information in release bursts is generally poor, even for bursts of longer duration than the ones used here (Cullinan & Tekiel, 1979; Kewley-Port, 1980). Thus any delay caused by inappropriate vowel information may actually be due to the burst's being taken as appropriate to a stop not among the choices in the task.

Although the vowel effect in the stop syllables is promising, the results of this experiment do not provide strong support for the notion that subcategorical mismatches slow phonetic judgments. For this phenomenon to be studied with stops, it is apparent that more control over the stimuli is needed, which is probably available only in synthesis.
EXPERIMENTS 4 AND 5

In Experiments 1 and 2, formant transitions have been shown to provide information about the fricative that cannot be completely ignored even when that information does not determine the category judgment. If the transitions were taken to give information about a segment other than the fricative, however, we would expect them not to affect the speed with which the fricative is identified. One way to make the transitions "affiliate" with another phone is to insert silence artificially between the friction and the vocalic segment (cf. Best, Morrongiello, & Robson, 1981; Mann & Repp, 1980). With a sufficient amount of silence preceding, transitions can be perceived as stops in fricative-stop clusters.

When 60 msec of silence was introduced between the friction and the first pitch pulse of the fricative-vowel syllables from Experiment 2, stop percepts resulted in about half the cases. Generally, the [s] transitions yielded stops, while the [ʔ] transitions were usually perceived as an interdental fricative [θ]. The unexpectedness of this result led to a reexamination of the particular stimuli used. As seen in Figure 6, there is a portion of the noise just before the onset of voicing that is much lower in amplitude than the rest of the friction (as seen in the waveform), and that has recognizable traces of formant transitions (as seen in the spectrogram). This token of [ʔa] is typical of the eight syllables used in Experiment 2. Although the first pitch pulse has been used as a demarcation between fricative and vowel (including transition) in previous experiments, the transitions need not begin with voicing. When the fricative gesture ends and the vowel gesture begins, there can be a brief period when the tongue is not close enough to the roof of the mouth to produce real friction but voicing has not started. What results then is essentially aspiration. This aspiration can be seen as part of the transitions, just as it is with voiceless stops.

When these observations are taken into account, it is clear that there is just as much justification for treating the "aspiration" as part of the transitions as for excluding it. If the onset of voicing defines a point that excludes some of the transition, it is not as surprising that introducing silence at that point will not always result in the perception of a stop. The "aspiration" deserves to go with the vocalic segment as well. In fact, when an appropriate amount of silence is introduced 10 msec before the onset of voicing (thus including a portion of aspirated transitions with the vocalic portion), stop percepts result with all the syllables of Experiment 2. Stimuli with 60 msec of silence inserted 10 msec before the first pitch pulse were then chosen for an experiment to determine whether the differing transitions slowed identification even when they affiliated with another phone, in this case, a stop. To justify the original result, however, the new location had to be tested in the original paradigm. Experiment 4, therefore, is a replication of Experiment 2, and Experiment 5 tests the theory that the transition effect will disappear when the transitions affiliate with a different phone.

EXPERIMENT 4

The four-choice condition of Experiment 2, in which the whole syllable was identified, was again replaced with one in which only the vowel was identified. In addition to the reasons for the revised procedure given above
Figure 6. Illustration of the low-amplitude, voiceless transitions, from the syllable [§a].
for Experiment 3, there was the added necessity of comparing Experiments 4 and 5. Since the syllables of Experiment 5 consist of three phones, it would be difficult for the subjects to identify only the first and third.

Experimental Procedure

Materials. The syllables [sa], [Sa], [su] and [Su] from Experiment 2 were used. The shortened versions of the frications were not used. Thus there were eight fricative and eight vocalic segments (since two tokens of each type were used), with the difference being that the vocalic segments now contained ten msec of voiceless transitions, and the fricatives were correspondingly shorter. Again, each fricative was combined with each vocalic segment, including the one it was originally produced with. This resulted in 64 unique stimuli, comprising the same groups of interest: 1) both transitions and vowel quality were appropriate; 2) transitions were appropriate but vowel quality was mismatched; 3) vowel quality was matched but transitions were not; and 4) both transitions and vowel quality were inappropriate.

Procedure. Each session consisted of two conditions. In one, subjects identified the fricative as quickly as possible; in the other, they identified the vowel. An unscored practice block of thirty stimuli was given before each condition. Each condition consisted of two blocks separated by a thirty-second pause. The order of the conditions was counterbalanced across subjects. The general procedure was the same as Experiment 2. In the fricative condition, the "s' response button was on the left and the "sh" on the right. In the vowel condition, the "a' button was on the left and the "u" on the right.

Subjects. Two groups of subjects were tested, expert and naive. The expert listeners were 10 researchers at Haskins Laboratories, all of whom were either phonetically trained and/or had experience in phonetic research. One was left-handed. The naive subjects were volunteers who were paid for their participation. None was left-handed.

Results

The error rate was 4.3% overall. Answers longer than one second (in both conditions) were counted as errors.

Figure 7 shows the results in the same manner as before. Inappropriate transitions resulted in a significant 6 msec delay, F(1, 18)=23.35, p < .01. Inappropriate vowels caused a 12 msec delay, F(1, 18)=28.43, p < .01. These two factors were again independent, F(1, 18)=1.85, n.s.

Identification of the fricative was faster than that of the vowel by an average of 68 msec, F(1, 18)=19.82, p < .01. The slowing effect of inappropriate transitions was the same whether the vowel or the fricative was identified, F(1, 18)=0.05, n.s. The vowel effect, on the other hand, was smaller when the fricative had to be identified, F(1, 18)=6.66, p < .05.

The expert subjects were 47 msec faster than the naive subjects (435 vs. 482 overall mean), but this difference was not significant, F(1, 18)=2.382, n.s. None of the interactions with the expert/naive factor was significant.
TIMES FOR IDENTIFICATION OF INITIAL FRICATIVES AND OF VOWELS

- Transition and Vowel matched
- Transition mismatched, Vowel matched
- Transition matched, Vowel mismatched
- Transition and Vowel mismatched

Figure 7. Times to identify the fricative or the vowel, Experiment 4.
TIMES FOR IDENTIFICATION OF INITIAL FRICATIVES AND OF VOWELS

- Transition and Vowel matched
- Transition mismatched, Vowel matched
- Transition matched, Vowel mismatched
- Transition and Vowel mismatched

Figure 8. Times to identify the fricative or the vowel, Experiment 5.
Subcategorical Phonetic Mismatches Slow Phonetic Judgments

Discussion

As before, mismatching the transitions, while it did not change the phonetic identity of the fricative, did slow identification. In this case, the identification was either of the fricative or just of the vowel. The delay caused by the inappropriateness of the vowel was again larger than that caused by inappropriate transitions (12 msec vs. 6 msec). However, the transition effect was more reliable. Also, the transition effect does not weaken in the two-choice condition. That is, identification is slowed equally by mismatches in transition whether the vowel is identified or only the fricative, as we would expect from Experiment 2.

Some of the finer details of this experiment and Experiment 2 do not match, but the overall picture is clear. Inappropriateness of transition leads to a delay in phonetic identification of both the fricative and the vowel; inappropriateness of vowel gives a similar, somewhat larger delay. These two effects are independent. The next experiment explores the effect of the transitions when they affiliate with a phone other than the fricatives they were originally produced with.

EXPERIMENT 5

Experimental Procedure

Materials. The syllables [sa], [əa], [su] and [əu] from Experiment 4 were used, but with 60 msec of silence inserted between the friction and the vocalic segment. This gave rise to a stop percept in all combinations. The procedure was otherwise the same as for Experiment 4.

Subjects. The subjects of Experiment 4 participated.

Procedure. The procedure of Experiment 4 was used.

Results

The error rate was 4.1% overall. Answers longer than one second (in both conditions) were counted as errors.

Figure 8 shows the results in the same fashion as the previous experiments. Inappropriate transitions resulted in a significant 4 msec delay, $F(1,18)=5.41$, $p < .05$. Inappropriate vowels caused a 17 msec delay, $F(1,18)=81.99$, $p < .01$. As before, the slowing effect of inappropriate transitions was the same whether the vowel or the fricative was identified, $F(1,18)=0.56$, n.s. This time, however, the vowel effect was also the same in both conditions, $F(1,18)=2.25$, n.s.

Subjects were significantly slower (by 124 msec) in identifying vowels than fricatives, $F(1,18)=39.69$, $p < .01$. Note that this is almost exactly 60 msec more than the difference in Experiment 4 (without the 60 msec of silence).

The expert subjects were again faster (this time by 47 msec) than the naive subjects (484 vs. 531 overall mean), but this difference was not significant, $F(1,18)=2.37$, n.s. There were no interactions with this factor.
An analysis that compared Experiments 4 and 5 was run. This revealed three interactions of interest. First, responses were slower to the syllables with inserted silence (459 vs. 507 msec), $F(1,18)=23.26$, $p<.01$. This was due largely to the vowel identification, $F(1,18)=23.26$, $p<.01$ (see Table 2). Since the syllables in Experiment 5 were 60 msec longer than those of Experiment 4, it is natural that the vowel judgments should be slower by approximately that much. The consonant judgments were also slower in Experiment 5. A separate analysis of variance of just the fricative identifications, however, shows that this difference is not significant, $F(1,18)=1.81$, n.s. This indicates that, while the listener is waiting long enough to integrate the information of the vocalic segment into the fricative percept, she does not need to wait for the syllable to finish before she makes her judgment.

| Table 2 |

Mean Reaction Times (in msec) for Identification of Fricative vs. Vowel Occurring in Experiments 4 and 5.

<table>
<thead>
<tr>
<th></th>
<th>fricative</th>
<th>vowel</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp 6</td>
<td>425</td>
<td>493</td>
</tr>
<tr>
<td>Exp 7</td>
<td>445</td>
<td>569</td>
</tr>
</tbody>
</table>

The prediction that the effect of inappropriate transitions would be greatly reduced is fulfilled. While the absolute duration is not much shorter (4 msec vs. 6 msec), the transition effect is much less reliable in Experiment 5, $F(1,18)$ of 5.41 for Experiment 5 vs. 23.35 for Experiment 4. It might seem that this is the result merely of physical separation of the two cues (friction and transition). With the same separation, however, the vowel effect strengthened, both in duration of the delay and its reliability: 12 msec, $F(1,18)$ of 28.43 for Experiment 4 vs. 17 msec, $F(1,18)$ of 82.00, for Experiment 5.

Discussion

Inserting silence between the friction and the vocalic segment so that a stop was perceived did not change the perceived phonetic category of the fricative. Nonetheless, the mismatch of transitions did slow the subjects somewhat. The delay caused by the inappropriateness of the vowel was again larger than that caused by inappropriate transitions (17 msec vs. 4 msec). In this instance, the vowel effect was much more reliable.

The transitions of [u] vocalic segments did not significantly affect reaction time, while the effect of inappropriate vowel quality was much greater for [u] than for [a]. Neither pattern showed up in Experiment 4.
Subcategorical Phonetic Mismatches Slow Phonetic Judgments

Since the transitions for high vowels are shorter than those for low vowels, it could be that only the transitions of low vowels give information about a preceding fricative. The previously noted effect of stop-affiliated transitions on fricative percepts (Repp & Mann, 1981) used only the vowel [a].

Even with added silence and a new (stop) percept, the general pattern established in the previous experiments remains: Inappropriateness of transition (in the one case where such an effect had been shown in perceptual studies previously) leads to a delay in phonetic identification of both the fricative and the vowel; inappropriateness of vowel gives a similar, somewhat larger delay. These two effects are independent.

The prediction that the effect of inappropriate transitions would be greatly reduced by the insertion of silence (Experiment 5) is fulfilled. While the absolute duration of the delay caused by inappropriate transitions is not much shorter (4 msec vs. 6 msec), the transition effect is much less reliable in Experiment 5, $F(1,18)=5.41$, than in Experiment 4, $F(1,18)=23.35$. It might seem that this is the result merely of physical separation of the two cues (friction and transition). With the same separation, however, the vowel effect strengthened, both in duration of the delay caused by inappropriateness and the reliability of the effect: 12 msec, $F(1,18)$ of 28.43, for Experiment 4 vs. 17 msec, $F(1,18)$ of 82.00, for Experiment 5.

GENERAL DISCUSSION AND CONCLUSION

The five experiments described in this paper provide convincing evidence that listeners take cues into account even when those cues seem both superfluous and ineffective. The vowel information in fricative noises and stop bursts and the consonant information in vocalic formant transitions both are generally too weak to do more than cause subcategorical variation, yet reliably slow down identifications if they are inappropriate. This slowing occurs whether the information pertains to the particular phone being identified, or to the phone that just happens to be presented at the same time. And finally, the mismatches cause just as much delay whether they precede or follow the overriding cue.

This last result is further evidence that listeners do not interpret the speech stream in a strictly left to right fashion. Other evidence to that effect has been found. For example, Repp, Liberman, Eberardt, and Pesetsky (1978) found that a stretch of silence was or was not treated as a cue to stop manner depending on the phonetic judgment made on the next segment. Miller and Liberman (1979) and Miller (1981) found that speaking rate, as determined by length of a following vowel, influenced the [b]-[w] boundary. Both these and other instances of later information affecting an earlier boundary involve timing. Various "disposing" theories (e.g., Klatt, 1979) have incorporated methods of withholding certain phonetic judgments until length information has been gathered. However, the present judgments do not depend on duration. In the fricative-vowel syllables, the place of the fricative is completely determined by the noise. Length could, in some cases, determine voicing. But there is no apparent reason for waiting until after the transitions have been processed to make the place decision. Thus the speech mechanism seems to integrate all cues available not only across the frequency range, but also across the time and frequency ranges together.

194
It might appear that the difference between the integrating and disposing accounts is the size of the time frame for analysis. This is not the case. The primary distinction is that disposing accounts wish to treat each time slice as a single (auditory) event and to extract all information from just its "gross spectral shape" (Blumstein, Isaacs, & Wertus, 1962; Stevens, 1980). A disposing theory with a large time window would thus need to extract, for example, both the stop consonant and the vowel from one spectral shape. If, on the other hand, the temporal window is increased but more than one spectral analysis is done, then the two theories would be indistinguishable.

Listeners do accumulate information about phones during the reception of the speech signal. It is possible for them, in the proper paradigm, to make decisions of fricative identity based on the noise alone (Repp, 1981a). The accumulation of cues, then, is continuous, even if adjustments to their values are made in response to later cues. When the whole signal must be processed, as in the identification of the vowel in the present experiments, the integration of cues seems to take place consistently.

The present results do not tell us very specifically just how long a listener waits before she reaches a decision. Recent work by Martin and Bunnell (1982) shows that vowel-to-vowel coarticulation, manipulated in much the same way as the present stimuli, holds across intervening consonants. Thus the syllable is not the absolute limit to the subcategorical matching process. A transient cue like a set of formant transitions, though, may be more tightly bound to the syllable in which it occurs. Only further experimentation will decide the issue.

The delays in identification due to phonetic mismatches are small but highly reliable. This suggests that subjects are not overly concerned that one or two minor variations are introduced, but must still take the time to integrate the cues processed. But consider the problem with synthetic speech. Unlike natural speech, which has almost everything right, synthetic speech has just barely enough right to be understood. Even "fully" intelligible synthesis may impose an unacceptable processing load for general usefulness. Those features that make a synthesized syllable just a bit harder to process (for example, getting the transitions slightly wrong after fricatives) may not be apparent even to the most critical listener. Yet the small delays may be adding up, requiring more time to be spent on phonetic processing, and leaving less time for semantic processing. If synthetic speech is to be listened to for long periods with the intention of getting the content straight, the synthesis must be more than interpretable. It must be accurate in ways that the person doing the synthesis cannot hear directly (cf. Nye & Gaitenby, 1973; Pisoni, 1982).

Finally, it should be noted that the proposed attempt by the listener to make sense of all she hears does not contradict the evidence that she can restore parts of the signal that are missing (Samuel, 1981; Warren, 1970). There is a difference between a lack of information and the presence of conflicting information. A demonstration of just that distinction in the present paradigm is being planned. But for now, we still have further evidence that the listener knows what a possible articulation is and attempts to integrate all cues in the construction of her phonetic percept.
REFERENCES


Fischer-Jørgensen, E. Tape cutting experiments with Danish stop consonants in initial position. *Annual Report, Institute of Phonetics, University of Copenhagen*, 1972, 6, 104-168.


