Subcategorical phonetic mismatches
slow phonetic judgments

D. H. WHALEN
Haskins Laboratories, New Haven, Connecticut

When an [s] or an [š] 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 [š], 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 initials [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 still had an effect on reaction time, as did mismatches of vowel quality. 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.

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 excised stop-consonant release bursts (Blumstein & Stevens, 1980; Kewley-Port, 1980; LaRiviere, Winitz, & Herriman, 1975b) or from excised fricative noises (LaRiviere, Winitz, & Herriman, 1975a; Yeni-Komshian & Soli, 1981).

Nevertheless, the vowel information in stop bursts and frictions 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 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) he or she can *reject* it, so that a nonspeech click, pop, whistle, etc., is perceived in addition to the speech; (2) he or 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) he 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 is a possible speech sound. If he or she treats the cue as nonlinguistic noise, it must be because he or 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 (cf. Mann & Repp, 1980; Whalen, 1981). 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 in formant transitions about a fricative's place of articulation has been shown to influence phonetic identification when the friction cue is ambiguous (Harris, 1958; Mann & Repp, 1980; Whalen, 1981). Unambiguous fricative noises, on the other hand, seem to completely override mismatched transitions in following vocalic segments. The perception of vowels following frictions that were originally produced with other vowels is similarly unaffected by that mismatched information.

Mismatched cues are rather more difficult to find 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-Jørgensen, 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, 1978). 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 (Pisoni & 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; Stevens, 1975). The former proposal will be called the "integrating" acount, 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 stops resulted in slower decision times in a speeded lexical decision task (Streeter & Nigro, 1979). (The effect appeared only 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 he or 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 noise followed by inappropriate transitions should give longer identification times. The disposing account, on the other hand, pre-
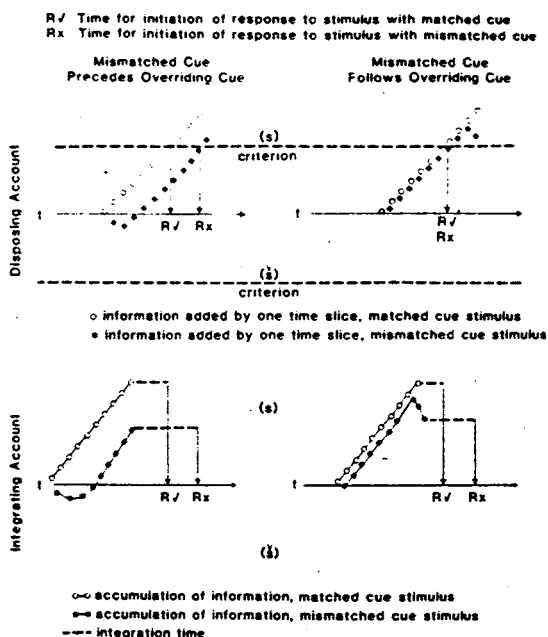
Figure 1. Comparison of the predictions of the disposing and integrating accounts for preceding and following mismatched cues.

dicts 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, whereas 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, whereas 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 fricative cue came after the conflicting cue. This will confirm the other results mentioned above. For three of the experiments, however, the overriding fricative 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, whereas the disposing account still predicts no effect.

EXPERIMENT 1

The first experiment examined two mismatched cues in vowel-fricative syllables. One was fricative information in the vocalic segment (the formant transitions), and the other was vowel information in the friction. For example, a mismatched transition would be contained in the vocalic segment from [uš] when combined with the fricative noise from [us]. Similarly, a mismatch of vowel quality would occur when the vocalic segment from [is] was combined with the fricative noise from [us]. The overriding cue for the fricative was the noise, and the overriding cue for the vowel was the vocalic segment. With the mismatched fricative cue, then, the overriding cue came later in the signal, while the overriding vowel cue came before the mismatched cue.

Method

Materials. A male native speaker of English recorded 10 tokens of each of the syllables [as], [aš], [is], [iš], [os], [oš], [us] and [uš] 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 durations of the vocalic portions of all eight were equal, the friction durations of all eight were equal, and, of course, the durations of the original syllables and all combined syllables were also equal. 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.

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 fricative as the percept generated by the noise ("appropriate transitions"), and the vowel was the same as the vowel the fricative had originally been produced with ("appropriate vowel"); (2) the transitions were appropriate, but the vowel was inappropriate; (3) the vowel was appropriate, but the transitions were inappropriate; 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, 1983). Thus, there is a mixture of cue status here; some are mismatched, and some are reinterpreted as an added phone. Whalen (1983) 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 re-

corded on one channel of an audiotape; on the other channel, a timing tone was recorded simultaneously with the onset of the stimulus. The interstimulus interval was 2,500 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. The 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 1 sec), a small light on the control box in front of them lit up. (The categories were always clear, so the feedback was not strictly necessary; many subjects reported ignoring it altogether.) 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 a 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 30-sec pause between, followed by a short break. The next two blocks, separated by a 30-sec 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 1 sec) 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) = 93.23, p < .001]. The four bars of the graph show mean identification time, respectively from left to right, (1) for the syllables in which both transition and vowel were matched, (2) for those in which the transition was mismatched but the vowel was matched, (3) for those in which the transition was matched but the vowel was mismatched, and (4) for those in which both transition and vowel were mismatched. On average, the subjects were 24 msec faster in their decisions when the transition was appropriate (means of 516 msec vs. 540 msec). The inappropriateness of the vowel also slowed the identification times [F(3,54) = 5.49, 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. Furthermore, 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) = 5.45, 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 were 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 (with the height of the vowel matching 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) = 4.63, p < .05], but appropriate height was not [F(1,18) = 2.08, n.s.]. Appropriateness of the transition did not interact with the appropriateness of the vowel for either rounding or height [Fs(1,18) = 1.696, 1.129]. The two types of vowel appropriateness did interact with each other [F(1,18) = 17.85, p < .001]. The syllables in which both vowel features were appropriate were identified faster than those in which 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 vocalic formant transitions slow identification of a following fricative. While this

TIME REQUIRED TO IDENTIFY FINAL FRICATIVE

☐ Transition and Vowel matched
▨ Transition mismatched, Vowel matched
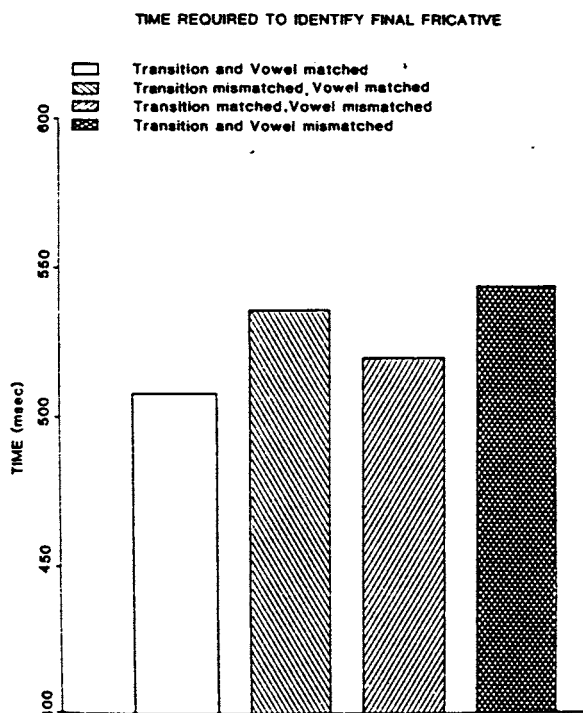▧ Transition matched, Vowel mismatched
▨ Transition and Vowel mismatched

Figure 2. Times to identify the fricative as [s] or [š], Experiment 1.

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, as they claim, are only an auditory event that leads the hearer to expect a fricative, then any transitions should do. Thus, the listener could ignore 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 frictions have shown this consistently (LaRiviere, Winitz, & Herriman, 1975a; Yeni-Komshian & Soli, 1981; cf. Whalen, 1983). 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. That is, the agreement in height between [i] and [u] is closer than that between [a] and [o].

In general, subcategorical phonetic mismatches can slow identification. This was true of the mismatched fricative cue that preceded the overriding cue, but also of the vowel cue that followed the overriding cue.

EXPERIMENT 2

The next experiment was designed to see if subjects could avoid processing delays when the fricative occurred first in the utterance, that is, when the overriding cue for the fricative preceded the mismatched cue. The overriding vowel cue (i.e., the vocalic segment) came first in Experiment 1, but the cue it conflicts with has not been as well established as the transition cues to the fricative. Although identifiable, the vowel information in the fricative noise has so far not been shown to modify the vowel percept, as would be seen in its effect on a vowel continuum. The transitions associated with the fricative, however, have shown just such an effect (Mann & Repp, 1980; Whalen, 1981) on fricative noise continua. A slowing effect of following mismatched fricative cues, then, is even stronger evidence that all cues are integrated.

Method

Materials. A male native speaker of English recorded 10 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 occurred 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 [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 (i.e., during the friction), 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 was. When identifying the whole syllable, subjects must wait for the vocalic segment to occur before they can make a 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 frictions made, each friction was combined with each vocalic segment. This gave 2 (short vs. long friction) × 2 ([s] vs. [š]) × 2 ([a] vs. [u]) × 2 (vowel that the friction was produced with vs. inappropriate vowel) × 2 (vocalic formant transitions are appropriate to the friction vs. inappropriate transitions) × 2 (tokens of each vocalic segment) × 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 interstimulus interval was 2,500 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. The 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 (Condition 1) or four (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 1 sec (for Condition 1) or 1½ sec (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 30-sec pause between, followed by a short break. The next two blocks, separated by a 30-sec 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 were given the two-choice condition (Condition 1) 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 1 or 1½ sec) 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, (1) for the syllables in which both transition and vowel were matched, (2) for those in which the transition was mismatched but the vowel was matched, (3) for those in which the transition was matched but the vowel was mismatched, and (4) for those in which 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 in the difference between the second two. The effect of appropriateness of vowel is
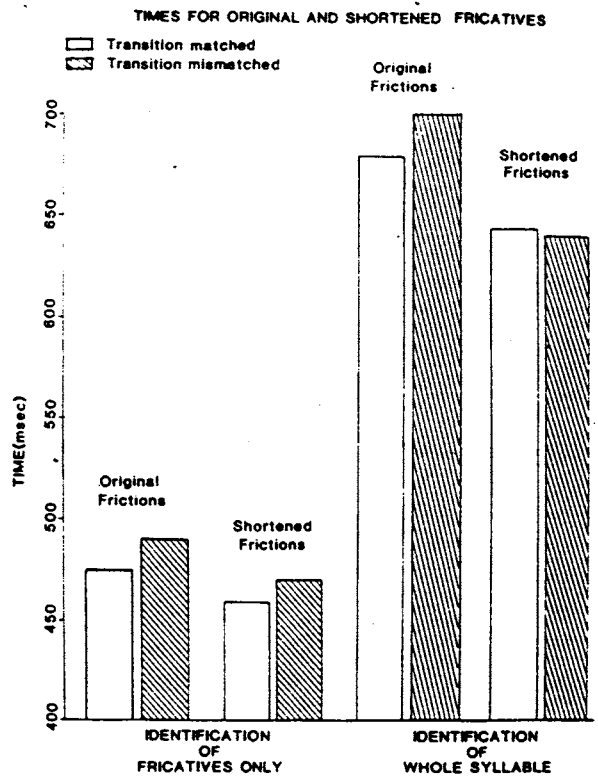

Figure 4. Times to identify the fricative of the whole syllable, shortened versus original stimuli, Experiment 2.
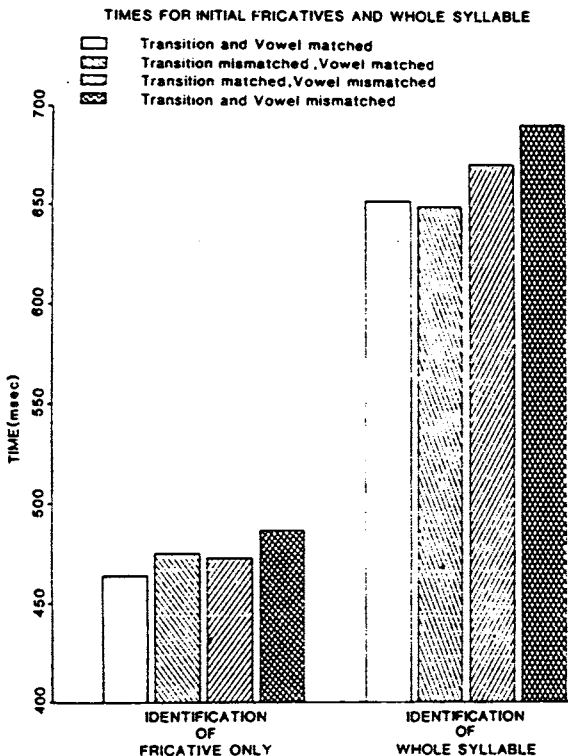
seen in the comparison of the first and third bars and in that of the second and fourth bars.

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 also 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 to 14 msec) 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 represent 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,

Figure 3. Times to identify the fricative or the whole syllable, Experiment 2.

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 in a 19-msec gain 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 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 msec when just the fricative was judged.

A further set of interactions reveals that the anomaly is limited to the syllables containing shortened frictions (see Figure 4). In the fricative condition, inappropriate transitions slowed responses both for the original and the shortened frictions. In the syllable condition, however, making the transitions inappropriate actually speeded the decision by 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 msec) [F(1,16) = 15.46, p < .01]. The same held for appropriateness of the vowel (26 msec vs. 41 msec) [F(1,16) = 9.35, p < .01]. There was a further interacton 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 frictions 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, 0.08 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, although mismatching the transitions did not change the phonetic identity of the fricative, it 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, in which the subject could, in principle, make his or her decision before he or she even hears the vowel, there is an effect. Furthermore, the mismatched fricative cue still slows 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 decisions 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, in which 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 frictions. (Note that, if prediction 2 is incorrect, there would be a difference between the two versions, assuming that the response was based on the offset of the noise.) 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 decisions 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, some of the statistical interactions conflict with this interpretation.

The higher level interactions in these data 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 decisions 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 to their full extent (in the affected identifications) or be ignored together (in the unaffected cases). The size of each effect is not important, and the two effects can be of different sizes (as they in fact are). Thus, there should be an interaction between appropriateness of transition and condition and an interaction between appropriateness of vowel and 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 appropriateness 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 [š], 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 does not fit well with the shortened reaction times for shortened frictions, even if the difference is smaller. If the subjects were using only the beginning of the noise for their decisions, shortening the friction should have had no effect at all. Both this experiment and Repp's (1981a) did show an effect. Further work is needed to discover what is responsible for these differences.

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 [š]. With the current stimuli, the [s] noise from [sa] is the most decidedly [s], and the [š] noise from [šu] is the most decidedly [š]. 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.

Method

Materials. A male native speaker of English recorded 10 tokens of each of the syllables [pa], [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 nonspeeded identification of the consonants was added to the experiment to assess the magnitude of the disagreement.

The mismatched cues to vowel and consonant were in the burst and came before the deciding cues, which were in the vocalic formant transitions and the vocalic segment. 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 interstimulus interval was 2,500 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. The subjects were seated in a sound-attenuated booth, and the stimuli were presented 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 1½ sec for the consonant condition, shorter than 1 sec 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 the subjects were not unanimous in their judgments 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 30-sec pause between, followed by a short break. The next condition, consisting of another two blocks separated by a 30 sec 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 nonspeeded 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 that of the previous results. The effect of the appropriateness of the transition was significant [$F(1,18) = 7.68$, $p < .05$]. On average, the subjects were 4 msec faster in their decision when the transition was appropriate. The effect was present only when the consonant was identified. This is shown by the interaction of condition with appropriateness of transition [$F(1,18) = 14.31$, $p < .01$]. Inappropriate

TIMES FOR IDENTIFICATION OF INITIAL STOPS AND FOR VOWELS

☐ Transition and Vowel matched
▨ Transition mismatched, Vowel matched
▦ Transition matched, Vowel mismatched
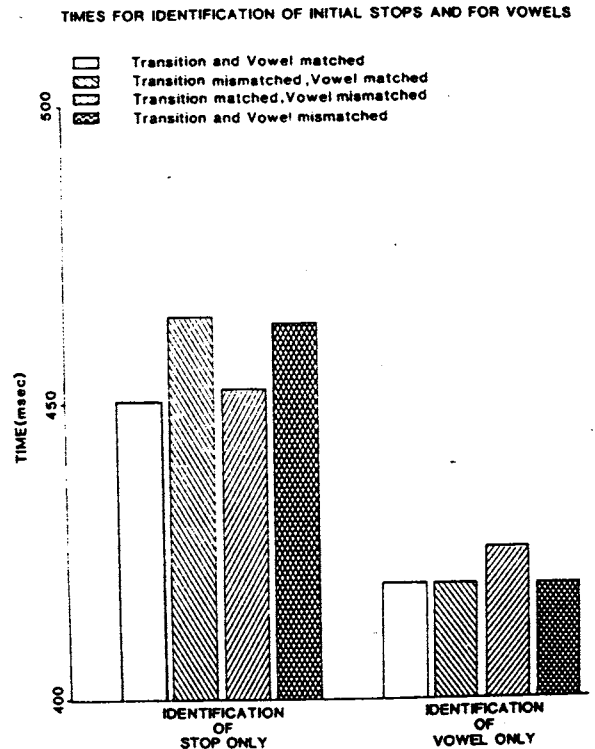▨ Transition and Vowel mismatched

Figure 5. Times to identify the stop or the vowel, Experiment 3.

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 [$F(1,18) = 1.08$, n.s.], despite a trend of 2 msec in that direction. As mentioned, some subjects did not correctly identify the stops that occurred before [u]. Misidentifications of the stop may have obscured the results, so an analysis of the data was done 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 [$F(1,18) = 9.07$, $p < .01$]. The means were 378 and 500 msec, respectively. This factor was involved in no significant interactions.

Results for the nonspeeded 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 misi-

Table 1
Results of Consonant Identification Task (Percent Misidentification)

| | Transition/Vowel | | | |
	Matched/ Matched	Mismatched/ Matched	Matched/ Mismatched	Mismatched/ Mismatched
For All 20 Subjects	1.6	7.0	4.1	6.4
For 10 Best Subjects	0.3	3.4	1.9	2.5

dentification 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 16 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 or interactions that were significant. However, their misidentifications still parallel the reaction times (see Table 1).

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 indicate that the stop bursts are not completely parallel to fricative noises.

Since the bursts were necessarily chosen for their minimal place information (so that the transitions would override the burst), their lack of a slowing effect is not too surprising. The stop can be identified to some extent from the burst alone (Kewley-Port, 1980; Tekieli & Cullinan, 1979; Winitz, Scheib, & Reed, 1971), but that does not tell us whether that information would be overriding in different contexts.

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 & Tekieli, 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 artificially insert silence between the friction and the vocalic segment (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 [š] transitions yielded stops, while the [s] 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 [ša] 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 aspira-
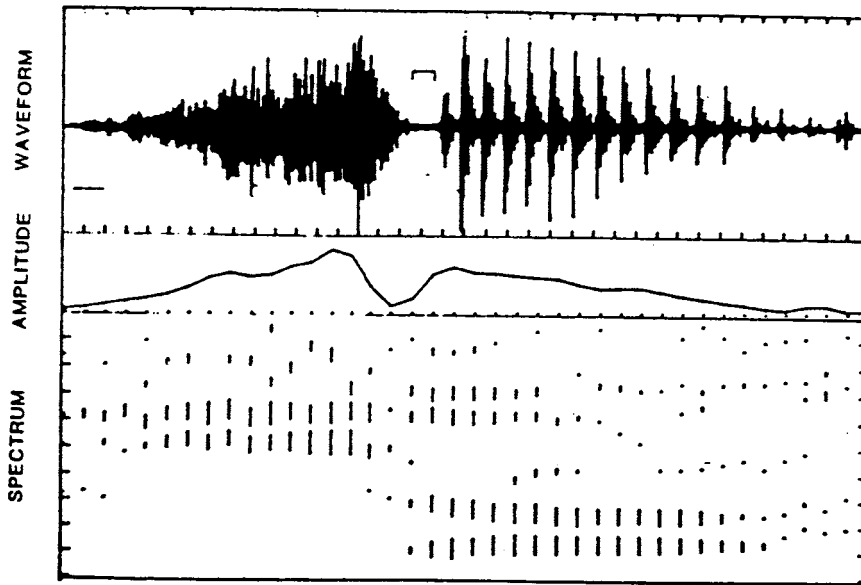
Figure 6. Illustration of the low-amplitude, voiceless transitions, from the syllable [ša].

tion 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 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. The shortened-friction manipulation of Experiment 2 was not carried out.

Method

Materials. The syllables [sa], [ša], [su], and [šu] from Experiment 2 were 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 10 msec of voiceless transitions and the frictions were corres-

pondingly shorter. Again, each friction 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 30 stimuli was given before each condition. Each condition consisted of two blocks separated by a 30-sec pause. The order of the conditions was counterbalanced across subjects. The general procedure was the same as in 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 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 1 sec (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 made for 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 [$F(1,18) = 0.03$, n.s.]. The

TIMES FOR IDENTIFICATION OF INITIAL FRICATIVES AND OF VOWELS

☐ Transition and Vowel matched
▨ Transition mismatched ,Vowel matched
☐ Transition matched.Vowel mismatched
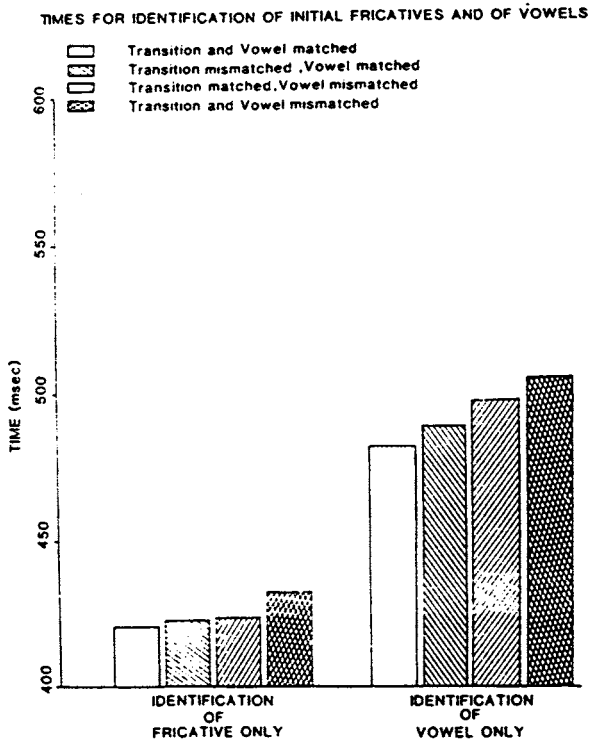▨ Transition and Vowel mismatched

Figure 7. Times to identify the fricative or the vowel, Experiment 4.

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 msec vs. 482 msec 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.

Overall responses were faster in this experiment than they were for the comparable stimuli in Experiment 2. This is probably due to the fact that many of the subjects were, by this point, exceedingly used to the task. Far from allowing the subjects to ignore the mismatched cues, however, this practice merely shortened the overall time and did not affect the factors of interest.

Discussion

As before, although mismatching the transitions did not change the phonetic identity of the fricative, it 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). The transition effect was more reliable than it was in Experiment 2. Also, identification is slowed equally by mismatches in transition whether it is the vowel that 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, but somewhat larger, delay. These two effects are independent. The next experiment explored the effect of the transitions when they affiliated with a phone other than the fricatives they were originally produced with.

EXPERIMENT 5

Experiment 4 verified the results of Experiment 2. Experiment 5 extends the results by seeing the effects of separating the cues. This is predicted to have different effects on the two cues. The fricative cues in the transitions should become, for the most part, cues to the new phonetic percept, the stop. The effect of mismatched transitions should thus be weakened. The vowel effect, on the other hand, should be unaffected, since coarticulation extends over consonant clusters (Martin & Bunnell, 1981). Integration should still take the cue into account.

Method

Materials. The syllables [sa], [ša], [su], and [šu] 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 1 sec (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 made for 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.].

The subjects were significantly slower (by 124 msec) in identifying vowels than in identifying fricatives [F(1,18) = 39.69, p < .01]. Note that this is almost exactly 60 msec more than the 68-msec 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 msec vs. 531 msec overall mean), but this difference was not significant [F(1,18) = 2.37, n.s.]. There were no interactions with this factor.

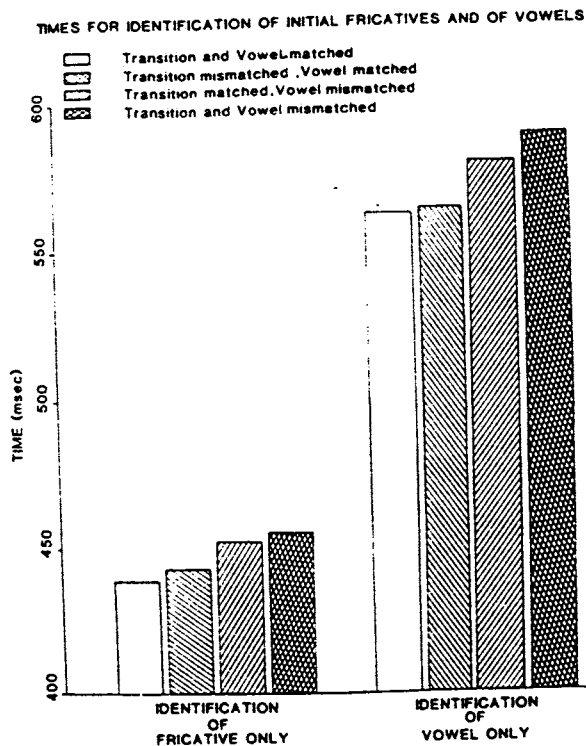An analysis that compared Experiments 4 and 5 was run. This revealed three interactions of interest.

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.

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 as before. The delay caused by the inappropriateness of the vowel was again larger than that caused by inappropriate transitions (17 msec vs. 4 msec). Thus, a large (in auditory terms) separation between the cues did not erase the effects of subcategorical mismatches.

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 not fulfilled. This could mean that the transitions were just as good a cue to fricative identity as before. Repp and Mann (1981), however, found that the stop transitions affected fricative category boundaries much less than transitions adjacent to the fricative noise. Thus, the current results could indicate that the mismatch in the transitions was a mismatch not to the fricative but to the stop. We would expect (from Repp & Mann, 1981) that stops in these contexts would have the same place of articulation as the fricative. The [s] transitions, however, probably cued an alveolar [t], whereas [š] transitions probably cued a velar [k]. (Subjects did not have to identify the stop.) If the fricative noise indicated one stop and the transitions indicated another, there would still be a mismatch, even though it was not on the phone to be identified. As has been shown in every experiment reported here, a mismatch on a phone slows the response, even if an unaffected phone is being identified.

First, responses were slower to the syllables with inserted silence (459 msec vs. 507 msec) [F(1,18) = 8.93, 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, he or she does not need to wait for the syllable to finish before he or she makes his or her judgment.

The prediction that the effect of inappropriate transitions would be greatly reduced is not fulfilled. While there is a trend in the absolute duration (4 msec vs. 6 msec), the transition effect does not give rise to an interaction.

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 they 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. Thus, a mismatch in vowel quality resulted in slower identifications not

Table 2
Mean Reaction Times (in Milliseconds) for Identification of
Fricative vs. Vowel Occurring in Experiments 4 and 5

	Fricative	Vowel
Experiment 4	425	493
Experiment 5	445	569

only of the vowels, but even of the fricatives. 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, Eccardt, 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 (1981) and Miller and Liberman (1979) 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. A "disposing" theory could withhold phonetic judgments until length information had been gathered (in a way similar to Klatt, 1979). If the place judgment was made solely on one time slice, but other judgments had to wait for duration cues, then the final phonetic percept would in fact have been integrated.

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, & Mertus, 1982; Stevens, 1980). (Even this proposal is suspect; see Kewley-Port, Pisoni, & Studdert-Kennedy, 1983.) If the time window were quite large, then both the stop consonant and the vowel information, for example, would have to be extracted from one spectral shape. This would not be possible. If, on the other hand, the temporal window is increased but spectral changes over time are considered, then the two theories would be indistinguishable. This is, in essence, the status of Klatt's (1979) position, in which the spectral shapes in stop release bursts have different values depending on which diphone (i.e., vowel context) they are in.

An interesting type of integrating theory is that of Kewley-Port (1983). The theory is integrating because some cues (such as midfrequency peaks) are taken one way in one context (e.g., the peak does not persist, thus could be an alveolar) and another in another (e.g., the peak does persist, so is a velar). The range of integration proposed, however, is much smaller than needed to account for the data presented here. Presumably the extension of the theory to cover more phone classes would also increase the range of integration. Determining that range, in fact, is a major area for future research.

Disposing theories depend on abrupt spectral changes to trigger phonetic analysis. Thus, it could be argued that the mismatches of Experiments 1-4 resulted in an additional phonetic analysis that would not be present if the signal was untampered with.

(That this must be an interesting auditory event rather than an artifact of digital editing is shown by the fact that half the wholly appropriate syllables had a digital splice in them; they came from different tokens of the same phoneme sequence.) However, Experiment 5 makes such explanations untenable. A 60-msec separation is quite large in auditory terms, yet exactly the same results were obtained in that experiment. The transitions may have been inappropriate for the stop, thus an auditory account might only have to specify that a phonetic expectation should be generated by the (offset of) the noise. The vowel effect, however, shows conclusively that an expectation is made for the vowel, rather than there being some auditory mismatch between, say, the second formant of the vowel and similar information in the friction (cf. Soli, 1981).

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 he or 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, such as 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 (the exception being the occasional misarticulation), 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 he or she hears does not contradict the evidence that he or 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 his or her phonetic percept.

REFERENCES

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 29, 191-211.

Blumstein, S. E., Isaacs, E., & Mertus, J. (1982). The role of the gross spectral shape as a perceptual cue to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 72, 43-50.

Blumstein, S. E., & Stevens, K. N. (1980). Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 67, 648-662.

Cole, R. A., & Scott, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 27, 441-449.

Cole, R. A., & Scott, B. (1974). Toward a theory of speech perception. *Psychological Review*, 81, 348-374.

Cullinan, W. L., & Tekieli, M. E. (1979). Perception of vowel features in temporally-segmented noise portions of stop-consonant CV syllables. *Journal of Speech and Hearing Research*, 22, 122-131.

Dorman, M. F., Raphael, L. J., Liberman, A. M., & Repp, B. H. (1975). *Some maskinglike phenomena in speech perception* (Haskins Laboratories Status Report on Speech Research, SR-42/43, pp. 265-276). New Haven, CT: Haskins Laboratories.

Dorman, M. F., Studdert-Kennedy, M., & Raphael, L. J. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, 22, 109-122.

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

Fujimura, O., Macchi, M. L., & Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, 21, 337-346.

Harris, K. S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1, 1-7.

Kewley-Port, D. (1980). *Representations of spectral change as cues to place of articulation in stop consonants* (Research in Speech Perception, Tech. Rep. 3). Bloomington, IN: Indiana University, Department of Psychology.

Kewley-Port, D. (1983). Time-varying features as correlates of place of articulation in stop consonants. *Journal of the Acoustical Society of America*, 73, 322-335.

Kewley-Port, D., Pisoni, D., & Studdert-Kennedy, M. (1983). Perception of static and dynamic acoustic cues to place of articulation in initial stop consonants. *Journal of the Acoustical Society of America*, 73, 1779-1793.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics*, 7, 279-312.

LaRiviere, C. J., Winitz, H., & Herriman, E. (1975a). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, 18, 613-622.

LaRiviere, C. J., Winitz, H., & Herriman, E. (1975b). Vocalic transitions in the perception of voiceless initial stops. *Journal of the Acoustical Society of America*, 57, 470-475.

Liberman, A. M. (1979). An ethological approach to language through the study of speech perception. In M. von Cranach, K. Poppa, W. Lepenies, & D. Ploog (Eds.), *Human ethology* (pp. 682-704). Cambridge: Cambridge University Press.

Liberman, A. M., & Studdert-Kennedy, M. (1978). Phonetic perception. In R. Held, H. W. Leibowitz & H. L. Teuber (Eds.), *Handbook of sensory physiology. Vol. VIII: Perception* (pp.143-178). New York: Springer-Verlag.

Mann, V. A., & Repp, B. H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 28, 213-228.

Martin, J. G., & Bunnell, H. T. (1981). Perception of anticipatory coarticulation effects in /stri, stru/ sequences. *Journal of the Acoustical Society of America*, 69, S92.

Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, 8, 473-488.

Miller, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, 38, 159-180.

Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, 25, 457-465.

Nye, P. W., & Gaitenby, J. H. (1973). *Consonant intelligibility in synthetic speech and in a natural speech control (Modified Rhyme Test results)* (Haskins Laboratories Status Report on Speech Research, SR33, pp. 77-91). New Haven, CT: Haskins Laboratories.

Pisoni, D. B. (1982). Perception of speech: The human listener as a cognitive interface. *Speech Technology*, 1(2), 10-23.

Pisoni, D. B., & Tash, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, 15, 285-290.

Repp, B. H. (1977). *Perceptual integration and selective attention in speech perception: Further experiments on intervocalic stop consonants* (Haskins Laboratories Status Report on Speech Research, SR-49, pp. 37-69). New Haven, CT: Haskins Laboratories.

Repp, B. H. (1981a). *Influence of vocalic context on perception of the [ʃ]-[s] distinction: V. Two ways of avoiding it* (Haskins Laboratories Status Report on Speech Research, SR-66, pp. 251-262). New Haven, CT: Haskins Laboratories.

Repp, B. H. (1981b). Perceptual equivalence of two kinds of ambiguous speech stimuli. *Bulletin of the Psychonomic Society*, 18, 12-14.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92, 81-110.

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 4, 621-637.

Repp, B. H., & Mann, V. A. (1981). Perceptual assessment of fricative-stop coarticulation. *Journal of the Acoustical Society of America*, 69, 1154-1163.

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, 110, 474-494.

Soli, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, 70, 976-984.

Stevens, K. N. (1975). On the potential role of property detectors in the perception of consonants. In G. Fant & M. A. A.

Tatham (Eds.), *Auditory analysis and the perception of speech.* New York: Academic Press.

STEVENS, K. N. (1980). Acoustic correlates of some phonetic categories. *Journal of the Acoustical Society of America,* 68, 836-842.

STREETER, L. A., & NIGRO, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America,* 65, 1533-1541.

TEKIELI, M. E., & CULLINAN, W. L. (1979). The perception of temporally segmented vowels and consonant-vowel syllables. *Journal of Speech and Hearing Research,* 22, 103-121.

WARREN, R. M. (1970). Perceptual restoration of missing speech sounds. *Science,* 167, 392-393.

WHALEN, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary. *Journal of the Acoustical Society of America,* 69, 275-282.

WHALEN, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech,* 26, 91-100.

WINITZ, H., SCHEIB, M. E., & REED, J. A. (1971). Identification of stops and vowels from the burst portion of /p,t,k/ isolated from conversational speech. *Journal of the Acoustical Society of America,* 51, 1309-1317.

YENI-KOMSHIAN, G. H., & SOLI, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America,* 70, 966-975.