Perception & Psychophysics
1980, 28 (3), 213-228

Influence of vocalic context on perception of the [ʃ]-[s] distinction

VIRGINIA A. MANN and BRUNO H. REPP
Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510

When synthetic fricative noises from a [ʃ]-[s] continuum are followed by [a] or [u] (with appropriate formant transitions), listeners perceive more instances of [s] in the context of [u] than in the context of [a]. Presumably, this reflects a perceptual adjustment for the coarticulatory effect of rounded vowels on preceding fricatives. In Experiment 1, we found that varying the duration of the fricative noise leaves the perceptual context effect unchanged, whereas insertion of a silent interval following the noise reduces the effect substantially. Experiment 2 suggested that it is temporal separation rather than the perception of an intervening stop consonant that is responsible for this reduction, in agreement with recent, analogous observations on anticipatory coarticulation. In Experiment 3, we showed that the vowel context effect disappears when the periodic stimulus portion is synthesized so as to contain no formant transitions. To dissociate the contribution of formant transitions from contextual effects due to vowel quality per se, Experiment 4 employed synthetic fricative noises followed by periodic portions excerpted from naturally produced [ʃa], [sa], [ʃu], and [su]. The results showed strong and largely independent effects of formant transitions and vowel quality on fricative perception. In addition, we found a strong speaker (male vs. female) normalization effect. All three influences on fricative perception were reduced by temporal separation of noise and periodic stimulus portions. Although no single hypothesis can explain all of our results, they are generally supportive of the view that some knowledge of the dynamics of speech production has a role in speech perception.

Acoustic analyses of speech have revealed that the noise spectrum of fricative consonants varies with the nature of the following vowel (Bondarko, 1969; Fujisaki & Kunisaki, 1978; Heinz & Stevens, 1961; see also our Appendix A). This acoustic context dependency seems to be primarily, although not exclusively, a consequence of anticipatory lip rounding for vowels such as [u] and [o], which results in a lowering of the fricative noise spectrum. (See Zue, Note 1, for a description of analogous effects of a following vowel on the spectrum of stop-consonant bursts.)

This coarticulatory effect has a parallel in perception: Listeners' identifications of fricative consonants are influenced by vocalic context. Although evidence for such a dependency has been scattered through the literature for some time (Delattre, Liberman, & Cooper, 1962; Hughes & Halle, 1956; Hasegawa & Daniloff, Note 2), the clearest demonstration was provided in a recent study by Kunisaki and Fujisaki (Note 3). Using a continuum of synthetic fricative noises varying from [ʃ] to [s], these researchers found that the category boundary shifted in favor of [s] when the following vowel was [u] or [o], relative to the boundary obtained in the context of [a] or [e]. In other words, the phoneme boundary shifted toward lower noise frequencies in the context of rounded vowels, in conformity with the analogous effect of anticipatory lip rounding on fricative noise spectra. Thus, the Japanese listeners seemed to take account in perception of contextual changes characteristic of fricative production, as if their phonetic perception were guided by an intrinsic knowledge of articulatory dynamics.

EXPERIMENT 1

The purpose of our first experiment was to replicate the basic finding of Kunisaki and Fujisaki (Note 3) that the phonetic perception of a fricative noise depends on the nature of the following vowel. This experiment also addressed the question of how the magnitude of that perceptual context effect changes as a function of two variables: the duration of the fricative noise and the presence or absence of a silent interval between the noise and the periodic portion.

It is important to note that changes in noise duration (within the range employed by us) have no gross effect on phonetic perception, whereas insertion of a silent interval induces perception of a stop consonant

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia Mann. Some of the results were reported at the 96th Meeting of the Acoustical Society of America in Honolulu, December 1978, and at the 10th International Congress of Phonetic Sciences in Copenhagen, August 1979. We thank Patti Price and Sarah Peck for their help in conducting and analyzing the experiments, Alvin Liberman for his advice at all stages of this project, Doug Whalen for sharing his observations and insights with us, and David Isenberg for helpful comments on an earlier draft of this paper.

0031-5117/80/080213-16$01.85/0

(cf. Bailey & Summerfield, 1980; Bastian, Eimas, & Liberman, 1961) and thus changes the phonetic structure of the stimulus. In Experiment 1, we were not concerned with distinguishing effects of temporal separation from effects of hearing an additional phonetic segment; this was the purpose of Experiment 2.

Method

Subjects. The 12 subjects included nine paid student volunteers recruited from Yale University, one research assistant, and the two investigators. With the exception of the second author, no subject had had extensive experience in listening to synthetic speech, although some had participated in earlier experiments of a similar nature. All but two of the subjects were native speakers of American English; the remaining two were native speakers of German and Chinese, respectively, but fluent in English. As inspection of individual results suggested that neither experience nor native language affected the pattern of results, the data of all 12 subjects were combined.

Stimuli. A synthetic fricative noise continuum was created on the OVE IIIc serial resonance synthesizer at Haskins Laboratories, following in part the specifications given by Kunisaki and Fujisaki (Note 3). Each noise was characterized by two steady-state poles (formants) produced by the fricative circuit of the synthesizer. No zero (antiformant) was specified. There were nine different stimuli. The center frequencies of both poles increased from Stimulus 1 ([ʃ]-like) to Stimulus 9 ([s]-like) in roughly equal steps; the step size was larger for the second (higher) pole than for the first. These frequencies are listed in Table 1. Each noise reached full amplitude after 40 msec and decreased in amplitude over the last 30 msec. Noise duration was 100 or 250 msec, depending on the condition.

In addition to the fricative noise continuum, we synthesized two periodic stimuli with roughly appropriate initial formant transitions, so as to make the fricative noise and periodic stimulus portions perceptually coherent (see Experiment 3). In isolation, these stimuli sounded like [ta] and [tu] (i.e., /da/ and /du/), respectively. Each was 200 msec in duration, with a 70-msec amplitude ramp at onset, and a fundamental frequency contour that fell linearly from 110 to 80 Hz. The steady-state frequencies of the first three formants were 771, 1,233, and 2,520 Hz for [a], and 250, 800, and 2,295 Hz for [u]. [ta] had 50-msec stepwise-linear transitions in the first and second formants with starting frequencies of 500 and 1,796 Hz, respectively. [tu] had a 70-msec stepwise-linear transition in the second formant only, with a starting frequency of 1,499 Hz.

Table 1
Pole Frequencies of Fricative Noises (in Hertz)

Stimulus	Pole 1	Pole 2
1	1957	3803
2	2197	3915
3	2466	4148
4	2690	4269
5	2933	4394
6	3199	4655
7	3389	4792
8	3591	4932
9	3917	5077

Note—The values given are synthesizer input parameters. Later acoustic analysis suggested that the actual frequencies were about 5% lower. Irregularities in step size were a consequence of using prespecified frequency values in conjunction with the limited frequency resolution of the OVE IIIc synthesizer. Their effect on the results, if any, was to reduce the size of shifts in the [ʃ]-[s] boundary.

The relative amplitudes of the stimulus components are presented in Appendix B, along with a discussion of the influence that amplitude levels might have had on the results.

Design. The experiment had five conditions, distinguished by the composition of the stimuli: (1) isolated 250-msec noises; (2) short (100-msec) noises, immediately followed by either [ta] or [tu]; (3) long (250-msec) noises, immediately followed by either [ta] or [tu]; (4) short (100-msec) noises, followed by a 150-msec silent gap and either [ta] or [tu]; and (5) long (250-msec) noises, followed by a 150-msec silent gap and either [ta] or [tu].

As can be seen, Conditions 2-5 represented the factorial combination of two variables: noise duration (100 or 250 msec) and gap duration (0 or 150 msec). In Conditions 2 and 3, listeners did not perceive any stop consonants because there was no silence indicating closure. Thus, the listeners heard reasonable instances of [ʃa], [sa], [ʃu], and [su]. In Conditions 4 and 5, there was a gap of more than sufficient duration to enable listeners to hear a stop consonant; thus, [ʃta], [sta], [ʃtu], and [stu] were perceived (sometimes, perhaps, [k] instead of [t]—see Experiment 2). Although [ʃt] (or [ʃk]) clusters do not occur in an initial position in English, they appeared to pose no perceptual difficulty for our listeners.

All stimulus sequences were recorded directly from the synthesizer onto magnetic tape. Condition 1 contained three random sequences of 42 stimuli each, with interstimulus intervals (ISIs) of 3 sec, and 6 sec between sequences. The other four conditions each contained five such sequences. In all conditions, the nine stimuli from the fricative noise continuum occurred with unequal frequencies according to a 1-2-3-3-3-3-3-2-1 schedule, which enabled us to collect more observations in the [ʃ]-[s] boundary region than at the ends of the continuum. The resulting basic set of 21 stimuli was replicated once within each sequence in Condition 1, whereas in the other conditions, the two different periodic portions, [ta] and [tu], led to 42 stimuli in each sequence. All in all, each listener gave 15 responses (18 in Condition 1) to each of the more ambiguous fricative noises (Stimuli 3-7 on the continuum).

Procedure. Since informal observations convinced us that practice would play little or no role, the five conditions were presented in the same fixed order (1-5) to all subjects, with brief pauses in between. The subjects were seated in a quiet room and listened over Telephonics TDH-39 earphones. The tapes were played back at a comfortable intensity on an Ampex AG-500 tape recorder. The task was the same in all conditions—to identify in writing the fricative consonant in each stimulus as either "sh" or "s."

Results

The results of this experiment are shown in Figure 1. Consider first the dotted function connecting the triangles in Panel b (duplicated in Panel d). It represents the percentage of "sh" responses to the nine isolated noises (Condition 1). It can be seen that all listeners reliably identified the endpoints of the noise continuum as "sh" and "s," respectively. Stimuli 3-7 showed varying amounts of ambiguity, but there was a reasonably orderly progression from "sh" to "s" responses.

Panels a and b show the effect of immediately following the fricative noises with a periodic portion. It can be seen that the predicted effect of vocalic context was obtained: Listeners were more likely to respond "sh" when [(t)a] followed than when [(t)u] followed. (The parentheses indicate that [t] was not actually perceived.) This effect, which replicates Kunisaki and Fujisaki (Note 3), was obviously very

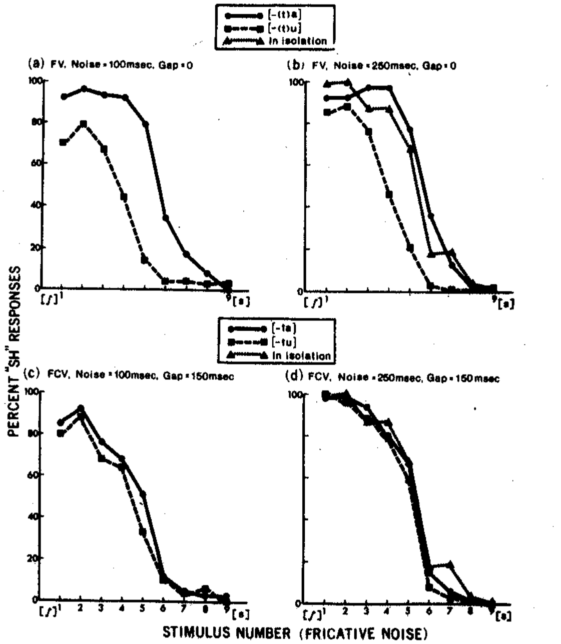

Figure 1. Effect of vocalic context on the [ʃ]-[s] contrast in four conditions (Experiment 1).

large and included even stimuli at the [ʃ]-end of the continuum. Comparison with the baseline results for isolated noises (Panel b) shows that the context effect was primarily due to [(t)u], which pulled the level of "sh" responses down. This is exactly what was to be expected if the perceptual effect of vowel context parallels the coarticulatory effect of anticipatory lip rounding. Since [(t)a] does not involve lip rounding, this context would not be expected to shift responses from the baseline level.

Comparison of Panels a and b indicates that extending the duration of the fricative noise from 100 to 250 msec left the context effect virtually unchanged. On the other hand, a glance at Panels c and d reveals that the introduction of a 150-msec gap between noise and periodic portion practically eliminated the effect. Note that Conditions 3 and 4 (Panels b and c) represent the same interval (250 msec) between noise onset and onset of periodicity; however, in one case, the first 100 msec of the noise were followed by more noise, whereas silence followed in the other case. Clearly, the silent interval in Condition 4 had a different effect on perception than the noise-filled interval in Condition 3. There was also an indication of a slight overall decrease in "sh" responses in Condition 4 (Panel c).

The statistical analysis of Conditions 2-5 confirmed these observations. A three-way analysis of variance was conducted on "sh" response percentages summed over all noise stimuli—a measure roughly equivalent to, and more convenient than, estimates of the category boundary location. The factors were vocalic context, noise duration, and gap. Vocalic context had a highly significant effect [$F(1,11) = 55.7$, $p < .001$], which interacted with gap [$F(1,11) = 62.5$, $p < .001$], but not with noise duration [$F(1,11) = 1.6$]. Interestingly, although the vowel context effect at the 150-msec gap was quite small, it was still highly significant [$F(1,11) = 17.6$, $p < .01$]. Thus, the introduction of the gap substantially reduced the context effect but did not completely eliminate it. In addition, there was a main effect of noise duration [$F(1,11) = 12.0$, $p < .01$] and an interaction of this factor with gap [$F(1,11) = 7.0$, $p < .025$], both effects being due to the decrease in "sh" responses in Condition 4 (short noise plus gap), as confirmed by separate tests. The reason for this decrease is not quite clear.

Discussion

Our first experiment partially replicated Kunisaki and Fujisaki's (Note 3) findings of an effect of vocalic context on perception of the distinction between [ʃ] and [s]. In addition, it extended those findings by examining the influence of two temporal variables on the magnitude of the context effect. The magnitude of the context effect was changed little by increasing the duration of fricative noises from 100 to 250 msec, which suggests that critical perceptual information is located at the point at which the fricative noise adjoins the periodic portion. On the other hand, the context effect was nearly eliminated by the introduction of a silent gap between noise and periodic portion. Apparently, the temporal contiguity of these two stimulus portions is crucial to their perceptual interaction. This is reasonable from a production viewpoint, since anticipatory lip rounding would be expected to affect the later portion of the fricative noise more than the earlier portions (Bondarko, 1969). Moreover, Bell-Berti and Harris (1979) have recently claimed that the onset of lip rounding precedes a rounded vowel by a certain fixed time interval. This implies that fricative noises will not be influenced by anticipatory lip rounding unless they fall within a certain distance from the vowel; thus, if perception parallels production, the contextual effect of rounded vowels on preceding fricatives should be highly dependent on the temporal relationship between noise and periodic portion.

There seems little point in investigating further the variable of noise duration. Given that a 250-msec noise is already beyond the range of durations normally encountered in running speech (Klatt, 1974; Umeda, 1977), extending noise duration further (even though it might eventually lead to a decline of the vowel context effect) would provide data that have little relevance to the perception of speech. However, it is of theoretical interest to pursue the question of why separation between noise and periodic portion reduced the extent of the contextual effect. This is so because an additional factor may have played a role—the perception of an intervening stop consonant, which resulted from the presence of a silent gap. In speech production, Bell-Berti and Harris (1979) have presented electromyographic data showing that anticipatory lip rounding is purely time dependent; the number of phonetic segments preceding the rounded vowel does not seem to matter. If perception parallels production, we should expect temporal separation to be the most critical factor in reducing the perceptual context effect, rather than the perception of an intervening phonetic segment.

EXPERIMENT 2

Assuming that the basic vowel context effect would be replicated when no silence intervened between the noise and the periodic portion, we expected the context effect to exhibit a sharp decline as gaps of increasing duration were inserted. The form of this decline was of special interest: Would it be continuous with increases in gap duration or would it show a discontinuity at the point at which stop consonants began to be heard?

Before conducting Experiment 2, we first collected data for stimuli with a gap duration of 75 msec—halfway between the gap sizes used in Experiment 1 and more than enough for a stop consonant to be heard. The duration of the fricative noise in these stimuli was 150 msec. The stimulus sequence was similar to those of Conditions 2-5 in Experiment 1, and the same 12 subjects listened to it in a separate session. The results showed a highly significant vowel context effect [$F(1,11) = 93.5$, $p < .0001$], which was nevertheless rather small, similar to that obtained with a 150-msec gap duration (Figure 1d).[1] Indeed, the difference between the context effects in the 75- and 150-msec gap conditions fell short of significance in a separate test [$F(1,11) = 4.1$, $p > .05$], and both effects were much smaller than that obtained with no gap at all. These data suggested that a major decrease in the vowel context effect occurs at gap durations shorter than 75 msec.

Method

Subjects. Nine subjects participated in this experiment. They included six new paid volunteers, a new research assistant, and the two investigators.

Stimuli. The stimuli were similar to those used in Experiment 1. The fricative noises were 150 msec long and had 50-msec initial and final amplitude ramps. (See also Appendix B.) There were eight gap durations: 0, 10, 20, 30, 40, 50, 100, and 150 msec.

The experimental tape contained three random sequences of 144 stimuli, separated by 3-sec ISIs. Each sequence contained the 18 combinations of the nine fricative noises with [ta] and [tu] at each of the eight gap sizes. In contrast with Experiment 1, all nine fricative noises occurred with equal frequency. Gap durations were totally randomized in the test sequences.

Procedure. Each subject listened to the experimental tape four times, in two separate sessions. The task was to identify, in each stimulus, both the fricative and any stop consonant perceived. The response choices were "s," "sh," "st," "sk," "sht," and "shk."[2] Half of the data of one subject were rejected since he gave hardly any "s" responses in the first session.

Results

The results are displayed in Figure 2, separately for the nine subjects in order to show the considerable individual differences. Each subject's panel contains four response functions: The two thin lines represent the percentage of stop responses in the [-(t)a] and [-(t)u] contexts as a function of gap duration; the two heavy lines represent the percentage of "sh" responses (averaged over the whole fricative-noise continuum) in the two vocalic contexts. The difference between the latter two functions is a measure of the magnitude of the vowel context effect.

First of all, it is evident that the basic context effect was indeed replicated: All subjects gave more "sh" responses in the [-(t)a] context than in the [-(t)u] context [$F(1,8) = 33.22$, $p < .001$]. There was, however, considerable variability in both the magnitude of the effect and in its relation to gap size. One subject (S.L.) showed a complete disappearance of the context effect at 40 msec of silence; two other subjects (B.H.R. and P.P.) showed a progressive reduction up to that interval. The remaining subjects showed little change in the magnitude of the context effect for gap sizes up to 50 msec. Analysis of variance of the 0-50-msec gaps revealed only a marginally significant and slightly irregular overall decline in the context effect with gap duration [$F(5,40) = 3.31$, $p < .05$]. Evidence for a decline of the context effect at longer gap sizes was more convincing; it was significant in an analysis of variance including the 0-, 50-, 100-, and 150-msec intervals [$F(3,24) = 8.54$, $p < .001$]. Nevertheless, Figure 2 shows that at least three subjects still exhibited sizable context effects at the longest gap duration.

We turn now to stop-consonant perception as a function of gap duration, in order to address the question of whether the perception of an intervening stop limited the occurrence of a context effect between vowel and fricative noises. The stop/no-stop boundaries for four of the nine subjects (B.H.R., V.A.M., P.P., and K.H.) were quite regular: No stop consonants were heard at the shortest gap durations (0, 10 msec), and 30-40 msec of silence were sufficient to hear stops in most cases. The responses of the remaining five subjects were more irregular. One of them (M.L.) heard stops at all gap durations, including stimuli without any true silence at all. Three subjects (S.W., S.L., and G.E.) heard stops in all (or nearly all) [tu] stimuli, regardless of gap size, although they tended to hear no stops in [ta] on at least some trials when gap duration was short. The remaining subject (J.N.) showed no difference between [ta] and [tu] but a moderate tendency to hear stops even at short gap durations.[3]

Despite this striking variability in the onset of stop percepts, the data provide clear evidence against the hypothesis that the perception of a stop consonant blocks the effect of a following vowel on fricative

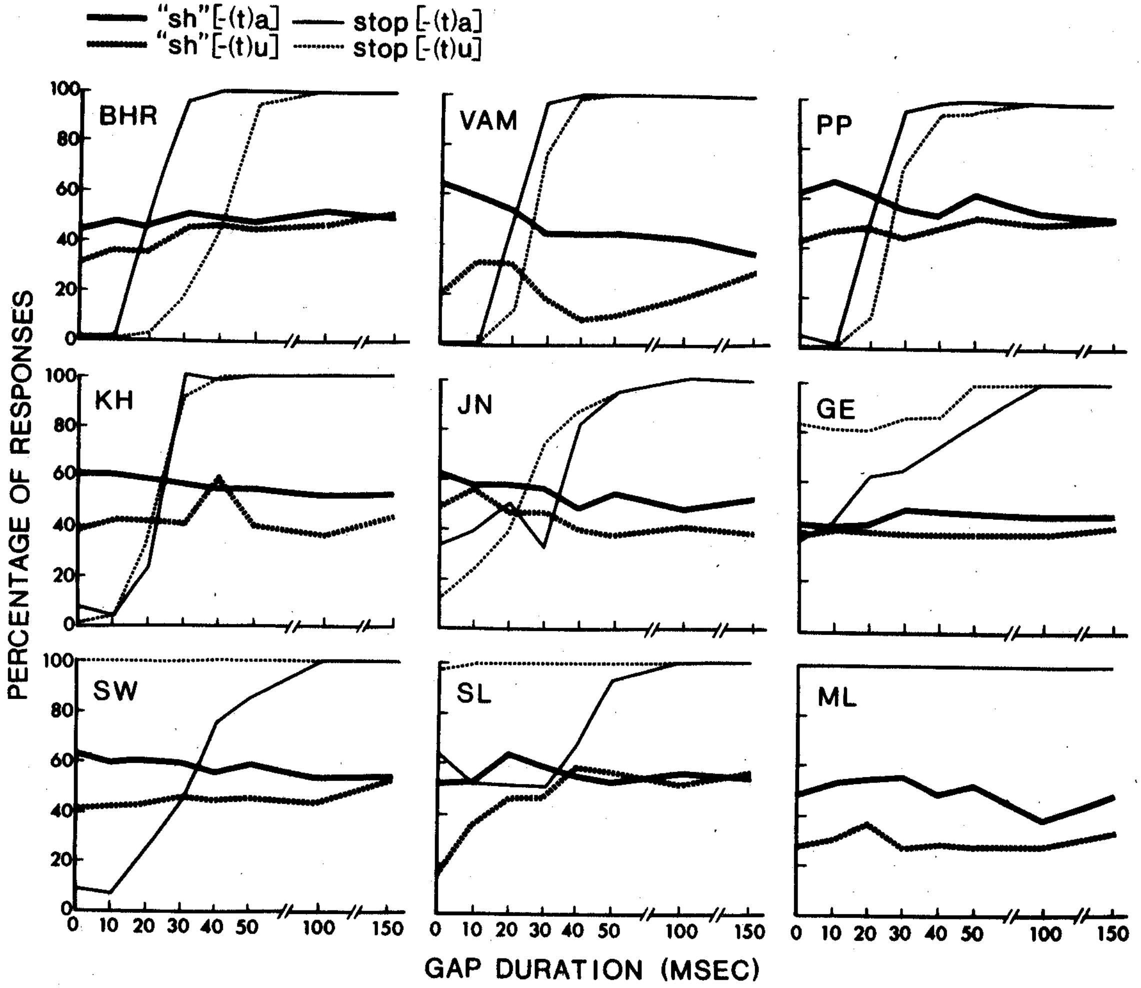

Figure 2. Effect of vocalic context on the average percentage of "sh" responses (heavy lines) and on emergence of stop percepts (thin lines) as a function of silent gap duration. (Individual subject data from Experiment 2.)

perception. Inspection of Figure 2 reveals that, in general, the onset of stop-consonant perception is not accompanied by a marked reduction in the magnitude of the context effect. The only possible exception is Subject S.L., for whom the context effect disappeared as soon as all stimuli were perceived as containing stops. However, this subject (and others as well) did show a large context effect at short gap durations despite a strong tendency to hear stops, which in itself argues against an inhibitory role of stop percepts.[4]

Discussion

The results of Experiment 2 lead us to conclude that the perception of an intervening stop consonant does not prevent effects of vocalic context on fricative labeling. For a few subjects, this context effect may have been slightly reduced by the perception of an intervening segment; however, the majority of subjects remained unaffected and showed only a slow decline of the context effect with increasing temporal separation of fricative noise and periodic portion. Indeed, in some cases, the context effect seemed to extend across more than 150 msec of silence. To the extent that temporal separation was more important than the number of phonetic segments perceived, the present results are in agreement with the speech production data of Bell-Berti and Harris (1979). However, both the large individual differences and the temporal extent of the context effect for some listeners suggest that it may be difficult to directly compare temporal parameters between speech perception and production.

One reason why perception and production are not strictly parallel may be the influence of several other factors on the perceptual context effect. One of them, relative amplitude of stimulus components, is discussed in Appendix B. Another important factor is represented by the formant transitions in the periodic stimulus portion. As we conducted Experiments 1 and 2, we began to wonder whether the perceptual effect of the periodic portion on the fricative was indeed due to vowel quality—as we assumed all along—or whether it was perhaps due, in part or entirely, to the initial formant transitions of the periodic stimulus portions. Although vocalic formant transitions have long been believed to be unimportant for the [ʃ]-[s] contrast (Harris, 1958; LaRiviere, Winitz, & Herriman, 1975), recent experiments by Whalen (Note 4) show that the transitions are a strong cue when the fricative noise is ambiguous (cf. also Delattre et al., 1962, for similar results on voiced fricatives). Thus, if the formant transitions of [tu] happened to be more appropriate for a forward place of fricative articulation than those of [ta], the vowel context effect on fricative perception could have been due to the formant transitions acting as cues to fricative place of articulation. Experiment 2 provided some relevant data on that point. Although both periodic stimuli in isolation sounded to us as beginning with [t], many subjects gave a substantial proportion of "k" responses when the same stimuli were preceded by a fricative noise plus a sufficient amount of silence to permit perception of a stop. Any "k" responses should have been more frequent with [ta] if the transitions of [tu] favored a more forward place of articulation. In fact, the opposite pattern predominated. Of the nine subjects, seven gave "k" responses only or predominantly to our [tu], one subject showed little difference between [ta] and [tu], and only one gave "k" responses to [ta] only.

Thus, it seems that, for the large majority of the subjects, the context effect must have been due to vowel quality, even at short gap durations. Indeed, if the transitions contributed to fricative perception, the transition effect may have partially canceled the vowel quality effect in these subjects, especially at short gap durations. This could have been one reason why there was so little reduction in the overall context effect with increasing gap duration (and with the emergence of stop percepts).

EXPERIMENT 3

The preceding considerations led us to focus on the role played by the vocalic formant transitions. We began by examining whether total elimination of formant transitions reduces the vowel context effect on fricative perception. Since all of our previous stimuli had contained formant transitions, the presumed effect of vowel quality was confounded with whatever effect the transitions themselves might have had on fricative perception. Removal of formant transitions seemed one way of getting rid of this confounding and of assessing the contextual effect due to vowel quality per se.

Method

Subjects. The subjects included all 12 individuals who had previously participated in Experiment 1.

Stimuli and Design. The stimulus materials were highly similar to those employed in Experiment 1, with fricative noises 150 msec in duration. There were three conditions, the first two being replications of the corresponding conditions in Experiment 1: (1) Isolated fricative noises, presented in five randomized blocks of 21 stimuli with ISIs of 2.5 sec. (The number 21 resulted from a 1-2-3-3-3-3-3-2-1 frequency distribution of the nine stimuli on the continuum.) (2) The same noises immediately followed by either [(t)a] or [(t)u]. There were five blocks of 42 stimuli, 21 for each vowel context. (3) The fricative noises followed by either [a] or [u], steady-state vowels obtained by straightening out all formant transitions in [ta] and [tu], leaving all other synthesis parameters unchanged. Otherwise, this condition was identical to Condition 2.

Procedure. All subjects listened to the conditions in the same fixed order (1-3) in a single session. The task was to identify the fricative consonant as "sh" or "s."

Results

The results are depicted in Figure 3 as the percentage of "sh" responses given to each stimulus along the fricative noise continuum. Panel a shows that Condition 2 successfully replicated the basic effect of the following vowel: There were fewer "sh" responses (hence, more "s" responses) in the [-(t)u] context than in the [-(t)a] context [$F(1,11) = 51.7$, $p < .0001$]. The size of the effect was not significantly different from that obtained for the same subjects in Experiment 1, which confirmed our impression that familiarity with the stimuli plays no important role. As in Experiment 1, the effect was almost exclusively due to [-(t)u]; perception of fricative noises in the [(t)a] context was similar to their perception in isolation (Condition 1, dotted function in Panel a).

Panel b shows what happened when the vocalic formant transitions were removed (Condition 3): The context effect practically disappeared and was no longer significant. Curiously, however, the subjects gave fewer "sh" responses to noises followed by either vowel than to the noises in isolation. The reason for this shift in response criteria is not clear.

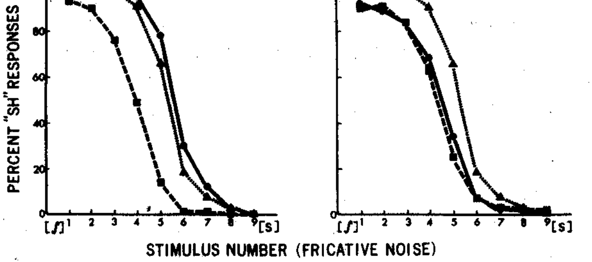

Figure 3. Effects of vocalic context on percentages of "sh" responses to stimuli from a fricative noise continuum in two conditions: with formant transitions (+TRANS.) and without (−TRANS.). The response function for isolated fricative noises (triangles, dotted) is duplicated in the two panels.

Discussion

Why did removal of the formant transitions eliminate the effect of a following vowel on fricative identification? There are two possible answers: (1) The transitions held the fricative noise and the periodic portion together and in this way *mediated* the effect of the vowel on the fricative. (2) Alternatively, the transitions themselves, rather than the steady-state vowel portions, were the source of the context effect.

There is phenomenological evidence in support of the first explanation: Elimination of the formant transitions resulted in a less coherent stimulus percept. In our own perception, the fricative noise seemed to be segregated from the periodic portion and to come from a different source. This is consistent with Cole and Scott's (1973) observation that the acoustic components of iterated transitionless fricative-vowel syllables segregate into separate auditory streams, whereas this tendency is much less strong when transitions are present. Thus, appropriate formant transitions seem to be necessary for auditory (and perceived articulatory) continuity between noise and periodic segments, and our results suggest that such continuity is necessary for a context effect to arise. A similar case has been made to explain the absence of perceptual interactions between stimuli with different apparent sources due to different fundamental frequencies (Ades, 1977; Darwin & Bethell-Fox, 1977; Dorman, Raphael, & Liberman, 1979).

Persuasive though these observations may be, we still need to consider the possibility that the supposed vowel context effect in stimuli containing formant transitions was actually due to the transitions themselves acting as cues to place of articulation of the fricative. Even though we have presented indirect evidence in connection with Experiment 2 that this was highly unlikely, it seemed important to determine directly the relative contributions of formant transitions and vowel quality. This was the purpose of Experiment 4.

EXPERIMENT 4

Experiment 4 had the primary purpose of dissociating the influences of formant transitions and vowel quality on fricative perception. To that end, it was necessary to vary these two factors independently. It would be quite difficult to synthesize transitions for [a] and [u] that are equally appropriate or equally neutral for one or the other fricative place of articulation. Therefore, we decided to combine synthetic fricative noises with periodic portions taken from natural fricative-vowel syllables. In this way, we could be assured that the formant transitions were indeed appropriate for either [ʃ] or [s], depending on the original utterance.

The same technique has recently been used by Whalen (Note 4) for a very similar purpose. Although he found that vocalic formant transitions contribute to the [ʃ]-[s] distinction, especially when the fricative noise is neutralized, he also found an effect of vowel quality that was largely independent of the transition effect. Thus, he anticipated the results of our study. However, his vowels varied between [i] and [u]; they did not include [a]. To enable us to make a more direct comparison with our earlier studies, we attempted to replicate Whalen's experiments using the vowels [a] and [u], at the same time extending the scope of the investigation to include two additional factors of interest.

Experiment 4 extended Whalen's studies by including a condition in which the periodic portion and the fricative noise were separated by a silent gap which led to perception of fricative-stop-vowel syllables. Thus, we examined the question of whether the expected transition and vowel quality effects are differentially affected by temporal separation. We thought that the transitions might contribute to fricative perception only as long as they are interpreted as cues to fricative place of articulation; when they are interpreted as cues to a stop consonant (as is the case when a sufficient amount of silence is inserted), they might lose their effect on fricative perception. The vowel quality effect, on the other hand, might still be present in reduced form when a silent gap is inserted (cf. Experiment 2).

Experiment 4 included yet another interesting variable. In order to assure that our results would not be specific to our selection of natural utterances, we used multiple tokens from two speakers, one male and one female. Consequently, the periodic portions that followed our synthetic fricatives reflected different vocal-tract sizes and source characteristics. We wondered whether these differences (hereafter referred to collectively as the speaker difference) would influence fricative perception. That there are detectable acoustic differences between the fricative noises produced by males and females has been shown by Schwartz (1968): The spectra of female [ʃ] and [s] noises are shifted upward on the frequency scale, relative to those produced by males, presumably because of differences in vocal-tract size. We might expect that speaker-specific information conveyed by the periodic stimulus portion would lead listeners to change their criteria in deciding on the preceding fricative, such that the [ʃ]-[s] boundary on our synthetic noise continuum would shift toward higher frequencies in the context of a female voice. Indeed, precisely such a perceptual normalization effect has been reported by May (Note 5), who followed synthetic fricative noises with synthetic periodic portions whose formants were scaled upward or downward to simulate changes in vocal-tract size. Our Experiment 4 was intended to confirm May's finding with natural-speech periodic portions.

In summary, then, this study examined the effects of orthogonal variations in three parameters—formant transitions, vowel quality, and speaker characteristics—on fricative perception, as well as changes in each of these effects consequent upon introduction of a gap (and a stop-consonant percept) between fricative and vowel.

Method

Subjects. The nine subjects included six paid volunteers, a research assistant, and the two authors. Three additional subjects had to be eliminated because they had difficulties in the gap condition.[5]

Stimuli. Two adults, one male and one female, both native speakers of American English, spoke the utterances [ʃa], [ʃu], [sa], and [su] repeatedly in a random sequence that included several other utterance types. All utterances were recorded on magnetic tape in a soundproof booth and subsequently digitized at 10 kHz using the Haskins Laboratories Pulse Code Modulation (PCM) system. Aided by our ears and by waveform displays, we selected three good tokens of each utterance for use in the experiment. Thus, there were 24 stimuli altogether (2 speakers by 2 fricatives by 2 vowels by 3 tokens). Using the PCM computer programs in conjunction with oscillographic displays, the fricative noises (defined as the signal portion preceding the onset of periodicity) were removed from all stimuli, and the digitized synthetic fricative noises from our nine-member continuum (Table 1) were substituted instead (*no-gap condition*).[6] The noises were 200 msec in duration and had been given a triangular amplitude contour (150-msec rise, 50-msec fall), designed to improve their naturalness. There was a total of $9 \times 24 = 216$ stimuli, which were recorded in two completely random sequences with ISIs of 3 sec and a 6-sec ISI after each group of 24. A second set of stimuli was constructed by inserting an 87-msec period of silence between the synthetic fricative noises and the natural periodic portions (*gap condition*). These stimuli were recorded in identical sequences. (For further details of stimulus structure, see Appendices B and C.)

Procedure. Some subjects listened twice to the no-gap tape before listening twice to the gap tape in a separate session. Others were presented with the no-gap tape followed by the gap tape in each of two sessions. Each subject gave a total of four responses to each individual stimulus (12 responses when ignoring token differences). The task in the no-gap condition was to identify the fricatives as "sh" or "s." In the gap condition, the following stop consonant (if perceived) had to be identified as well; the relevant response choices were "p," "t," "k," or any other label that seemed appropriate.

Results

The fricative identification results are shown in Figures 4 and 5. Figure 4 displays percentages of "sh" responses as a function of stimulus position along the fricative noise continuum. The panels on the left display the no-gap condition, those on the right the gap condition. Top panels are for the male speaker, bottom panels for the female speaker. The four different functions in each panel correspond to the four original utterances, [ʃa], [ʃu], [sa], and [su], from which the periodic stimulus portions derived. The data have been averaged over the three different tokens of each periodic portion. Figure 5 summarizes these data in terms of the overall percentage of "sh" responses (averaged over the nine stimuli on the fricative noise continuum) as a function of original utterance. This figure displays the token variation (data point triplets) and makes the speaker effect (male vs. female) easier to see.

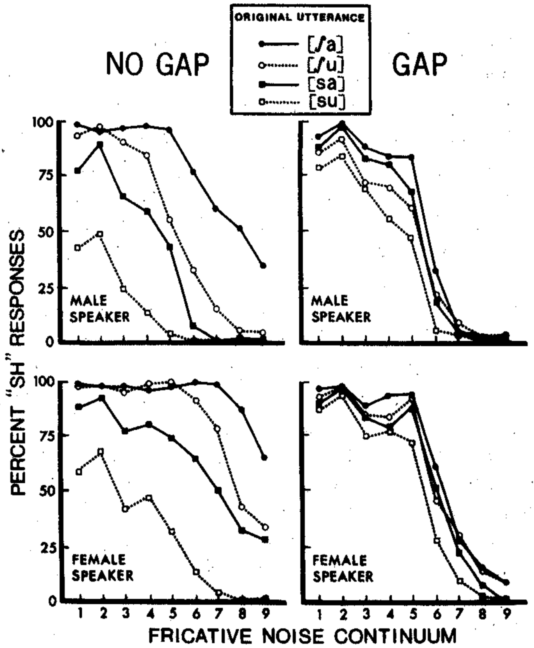

Figure 4. Effects of natural formant transitions and vowel quality on "sh" responses to stimuli from a synthetic fricative noise continuum, for two speakers (male, female) in two conditions (no-gap, gap).

In the no-gap condition, the [ʃ]-[s] distinction was strongly affected by all three factors: vowel quality [F(1,8) = 35.7, $p < .001$]; formant transitions [F(1,8) = 52.6, $p < .001$]; and speaker [F(1,8) = 52.7, $p < .001$]. Listeners gave substantially more "sh" responses to fricative noises followed by [-a], [ʃ] transitions, or a female voice, than to noises followed by [-u], [s] transitions, or a male voice. All three effects were in the predicted direction and so strong that seven out of eight response functions (Figure 4, left panels) did not reach asymptote at both ends. For example, when the periodic portion derived from a female [ʃa], even the most [s]-like noise received 65% "sh" responses; and when the periodic portion derived from a male [su], even the most [ʃ]-like noise received only 50% "sh" responses. The transition effect was larger with [-u] than with [-a] [F(1,8) = 41.7, $p < .001$]; this interaction was more pronounced with the female voice than with the male [F(1,8) = 6.9, $p < .05$].

Consider now the results of the gap condition, shown in the right-hand panels of Figures 4 and 5. As can be seen, all effects are substantially reduced, with response functions close to asymptote at either end and of similar shapes. (The abrupt drop in "sh" responses between Stimuli 5 and 6 is an artifact due to somewhat unequal step sizes on the noise continuum.) A joint analysis of variance of the no-gap

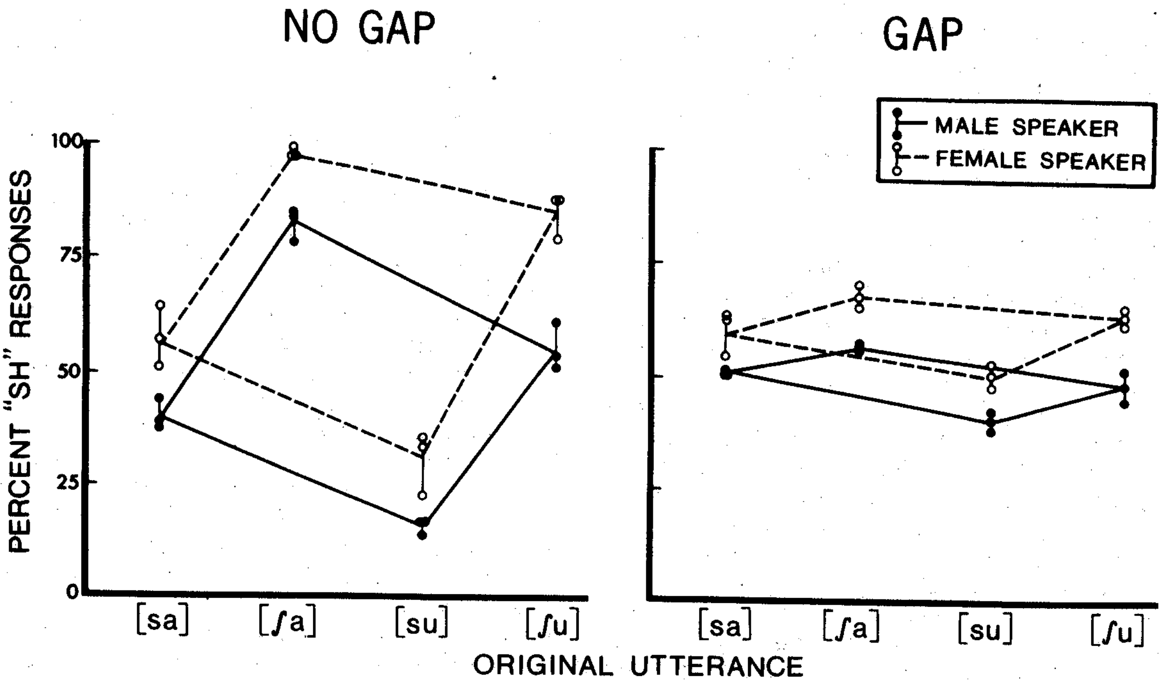

Figure 5. Summary of Figure 4 in terms of overall percentages of "sh" responses, with token variation displayed. Ascending lines, transition effect; descending lines, vowel quality effect; vertical displacement, speaker effect.

and gap conditions showed the decline in magnitude consequent upon introduction of an 87-msec gap to be significant for all three effects: vowel quality [$F(1,8) = 11.8$, $p < .01$]; formant transitions [$F(1,8) = 41.6$, $p < .001$]; and speaker [$F(1,8) = 10.8$, $p < .02$]. Nevertheless, all three effects were still present in the gap condition, as confirmed by a separate analysis: vowel quality [$F(1,8) = 12.9$, $p < .01$]; formant transitions [$F(1,8) = 14.9$, $p < .01$]; and speaker [$F(1,8) = 18.5$, $p < .01$]. There was no longer any interaction between the transition and vowel quality effects.

It is evident from Figure 5 that token variation was small relative to the effects under investigation, even though some token differences appeared to be systematic and reliable. An analysis with token variance as the error term yielded essentially the same results as the earlier analysis (which used Treatment by Subject interactions as error terms). A min F′ analysis (Clark, 1973), which combined subject and token variability, again yielded similar results; in particular, the three main effects remained significant at the $p < .001$ level in the no-gap condition and at the $p < .01$ level in the gap condition.[7]

Discussion

The results of the no-gap condition demonstrate that the [ʃ]-[s] distinction is affected both by the quality of the following vowel and by the vocalic formant transitions. Our findings are in excellent agreement with those of Whalen (Note 4), including the large size of the effects (presumably due to increased perceptual weights of natural acoustic cues, contained in the periodic portion, relative to synthetic cues, contained in the fricative noises). Whalen's, and our, successful experimental isolation of a true vowel quality effect reinforces our earlier conclusion that the "vowel context effect" obtained in our earlier experiments (and by Kunisaki & Fujisaki, Note 3) was indeed primarily due to vowel quality, even though the formant transitions were not strictly controlled.

Despite their paradigmatic similarity, the vowel quality and formant transition effects are very different phenomena from a theoretical perspective. The formant transitions are a consequence of the articulatory movements involved in producing the fricative consonant. Thus, they constitute a perceptual cue to fricative place of articulation; this cue is integrated with others (such as the fricative noise) into a unitary phonetic percept. Vowel quality, on the other hand, is neither a consequence of fricative production nor a direct cue to fricative perception. Rather, it is an independent factor that affects the production of the fricative, and this coarticulatory effect is somehow compensated for in perception. Thus, only the vowel quality effect is a true context effect; the transition effect is best viewed as a manifestation of perceptual cue integration (cf. Repp, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978).

We had this theoretical distinction in mind when we predicted that the vowel quality and transition effects would be differentially affected by insertion of a silent gap between fricative noise and periodic portion. We hypothesized that, when the transitions are interpreted as cues to place of articulation of a stop consonant, they would lose their effect on fricative perception. The vowel quality effect, on the other hand, was expected to persist in reduced form, since this was the result obtained in Experiments 1 and 2. These predictions were only partially confirmed. The vowel quality effect was indeed reduced, and the transition effect even more so. However, in addition to a diminished vowel quality effect, a significant transition effect persisted in the gap condition. This effect cannot be accounted for by the fact that no stop consonants were heard on some trials, despite the gap. For example, the subject with the largest transition effects in the gap condition always heard stops.

This persistence of the transition effect could be explained post hoc in at least two ways. One possibility is that perceptual integration of the fricative noise and transitional cues was not blocked by the perception of an intervening stop consonant. Perhaps blockage did not occur because certain important cues were absent, most notably, the plosive burst following a natural stop closure. After all, the stimuli in the gap condition were derived from fricative-vowel utterances and not from fricative-stop-vowel utterances. Therefore, the nature of the acoustic cues may have promoted integration, regardless of whether or not an intervening stop was perceived. Another possibility is that there is a perceptual dependency between a fricative and a following stop consonant, such that listeners are more likely to hear [s] when [t] follows and [ʃ] when [k] follows. However, Mann and Repp (in press) found that [t] and [k] tend to affect the perception of preceding [ʃ] or [s] in precisely the opposite direction, if at all. Therefore, we opt for the first interpretation—that the acoustic cues, because of their origin in fricative-vowel utterances, promoted perceptual integration despite perception of an intervening phonetic segment. In other words, the formant transitions contributed to the perceived place of articulation of two phonetic segments—the fricative and the stop consonant.

A final comment is necessary concerning the effect of speaker characteristics on fricative perception. We found that, in the no-gap condition, listeners gave substantially more "sh" responses to noises in the context of the female voice. This result confirms May (Note 5) and suggests that listeners compensate, or "normalize," for changes in fricative noise spectrum induced by differences in vocal-tract size. It is interesting to note that the extent of the perceptual compensation seems much larger than actual differences in fricative spectrum between male and female speakers (Schwartz, 1968). Thus, listeners seem to overcompensate (or "hypernormalize") in perception. However, to us, the most interesting finding was that the speaker effect decreased substantially with introduction of a gap. This finding has theoretical implications, since it indicates a divergence of perception and production. Clearly, articulatory effects of vocal-tract size on fricative noise spectrum do not vary with the context in which the fricative occurs; however, perceptual compensation for such effects proves to be context dependent. It may be argued that the speaker effect on fricative perception was reduced in the gap condition because the gap allowed the fricative noise to become perceptually dissociated from the periodic portion, as if they did not belong to the same utterance. However, this does not provide an explanation; it merely describes the possible phenomenological consequence of introducing a gap. The important implication of the result is that perceptual normalization effects are sensitive to local temporal properties of the speech signal. In this way, they seem to be rather similar to perceptual effects of speaking rate (Miller, in press; Summerfield, Note 6). In each case, the perceptual effects seem to operate only over a limited temporal region, suggesting the involvement of a rapidly decaying auditory memory or a sliding perceptual integrator with a time window of a few hundred milliseconds.

SUMMARY AND CONCLUSIONS

The present set of experiments reveals that, when listeners are asked to judge whether a given utterance contains [ʃ] or [s], they do not restrict themselves to acoustic information contained in the fricative noise. Rather, their decision can be influenced by certain attributes of the following signal portion. Here, we have considered the importance of three attributes: the quality of the following vowel, the nature of the formant transitions, and the sex of the speaker. With regard to the first, we have shown that the category boundary on a synthetic [ʃ]-[s] continuum can be shifted by the presence of a following [u], but not by a following [a]. This finding that more "s" responses occur in the context of a rounded vowel (replicating recent work on Japanese listeners by Kunisaki & Fujisaki, Note 3) is a perceptual analogue to the coarticulatory influence of rounded vowels on fricative production, which results in a lowering of the fricative noise spectrum (see Appendix A). Moreover, we have found that temporal separation between fricative noise and periodic portion is the primary determinant of the extent of vowel effects on fricative perception; and Bell-Berti and Harris (1979) report that temporal separation proves to be the primary determinant of coarticulatory influence of following vowels on fricative production. Thus, it seems appropriate to conclude that some tacit knowledge of the dynamics of speech production is involved in speech perception.

This line of reasoning further accords with our finding that formant transitions also influence fricative perception. As we have shown, formant transitions not only preserve the perceptual coherence of the utterance and thus allow vowel quality to have its effect, they also have a significant effect on the perceived place of articulation of the fricative. The reason that listeners should place perceptual weight on both fricative noise and formant transitions is to be found in the fact that the transitions are a consequence of fricative production. As Repp et al. (1978) hypothesized, spectrally quite disparate cues are integrated into a single phonetic percept whenever they reflect a single articulatory act.

Our final experiment reveals that listeners are sensitive not only to the diverse acoustic consequences of the gestures involved in fricative production, but also to the characteristics of the vocal tract in which those gestures seem to have occurred. This point is established by the finding that more [ʃ] percepts occur in the context of a female voice than in that of a male voice—a finding that, once again, has its parallel in speech production. Here, as in the case of vowel quality and formant transition effects, the relation between perception and production is not perfect. The temporal limitations on perceptual context effects do not always reflect the dynamics of speech production, and their extent frequently seems to exceed that of the corresponding coarticulatory shifts. Presumably, this is the price we have to pay for assessing perception-production relationships in the laboratory, where various stimulus and task characteristics influence listeners' response criteria. We should expect the fit between speech production and perception to be much closer in natural communication.

REFERENCE NOTES

1. Zue, V. W. *Acoustic characteristics of stop consonants: A controlled study* (Tech. Rep. 523). Lexington, Mass: Lincoln Laboratory, M.I.T., 1976.

2. Hasegawa, A., & Daniloff, R. G. *Effects of vowel context upon labelling the /s/-/š/ continuum*. Paper presented at the 91st Meeting of the Acoustical Society of America, Washington, D.C., April 1976.

3. Kunisaki, O., & Fujisaki, H. *On the influence of context upon perception of voiceless fricative consonants* (Annual Bulletin, Vol. 11, pp. 85-91). Tokyo: Research Institute for Logopedics and Phoniatrics, University of Tokyo, 1977.

4. Whalen, D. H. *Effects of vocalic formant transitions and vowel quality on the English [s]-[š] boundary* (Status Report on Speech Research, SR-59/60, pp. 35-48). New Haven, Conn: Haskins Laboratories, 1979.

5. May, J. *Vocal tract normalization for /s/ and /š/* (Status Report on Speech Research, SR-48, pp. 67-74). New Haven, Conn: Haskins Laboratories, 1976.

6. Summerfield, Q. *On articulatory rate and perceptual constancy in phonetic perception*. Unpublished manuscript, 1979.

7. Repp, B. H., & Mann, V. A. *Influence of vocalic context on perception of the [ʃ]-[s] distinction: II. Spectral factors* (Status Report on Speech Research, SR-61, pp. 65-84). New Haven, Conn: Haskins Laboratories, 1980.

REFERENCES

Ades, A. E. Source assignment and feature extraction in speech. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 673-685.

Bailey, P. J., & Summerfield, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, **6**, 536-563.

Bastian, J., Eimas, P. D., & Liberman, A. M. Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America*, 1961, **33**, 842. (Abstract)

Bell-Berti, F., & Harris, K. S. Anticipatory coarticulation: Some implications from a study of lip rounding. *Journal of the Acoustical Society of America*, 1979, **65**, 1268-1270.

Bondarko, L. V. The syllable structure of speech and distinctive features of phonemes. *Phonetica*, 1969, **20**, 1-40.

Clark, H. H. The language-as-fixed-effect fallacy: A critique of language statistics in psychological research. *Journal of Verbal Learning and Verbal Behavior*, 1973, **12**, 335-359.

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

Darwin, C. J., & Bethell-Fox, C. E. Pitch continuity and speech source attribution. *Journal of Experimental Psychology: Human Perception and Performance*, 1977, **3**, 665-672.

Delattre, P. C., Liberman, A. M., & Cooper, F. S. Formant transitions and loci as acoustic correlates of place of articulation in American fricatives. *Studia Linguistica*, 1962, **16**, 104-121.

Dorman, M. F., Raphael, L. J., & Liberman, A. M. Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, 1979, **65**, 1518-1532.

Fairbanks, G., House, A. S., & Stevens, E. L. An experimental study of vowel intensities. *Journal of the Acoustical Society of America*, 1950, **22**, 457-459.

Fujisaki, H., & Kunisaki, O. Analysis, recognition, and perception of voiceless fricative consonants in Japanese. *IEEE Transactions (ASSP)*, 1978, **26**, 21-27.

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

Heinz, J. M., & Stevens, K. N. On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, 1961, **33**, 589-596.

Hughes, G. W., & Halle, M. Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, 1956, **28**, 303-310.

Klatt, D. H. The duration of [s] in English words. *Journal of Speech and Hearing Research*, 1974, **17**, 51-63.

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

Lehiste, I., & Peterson, G. E. Vowel amplitude and phonemic stress in American English. *Journal of the Acoustical Society of America*, 1959, **31**, 428-435.

Mann, V. A., & Repp, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, in press.

Miller, J. L. The effect of speaking rate on segmental distinctions: Acoustic variation and perceptual compensation. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the study of speech*. Hillsdale, N.J: Erlbaum, in press.

Repp, B. H. Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 1978, **24**, 471-485.

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

Schwartz, M. F. Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America*, 1968, **43**, 1178-1179.

Umeda, N. Consonant duration in American English. *Journal of the Acoustical Society of America*, 1977, **61**, 846-858.

NOTES

1. The high level of significance of the 75-msec context effect was due to its remarkable consistency across subjects: All 12 listeners showed a small effect in the expected direction.

2. When serving as subjects in Experiment 1, we had noticed a tendency to hear velar stops on occasion, even though the periodic stimulus portions in isolation were heard by us as beginning with alveolar stops. Our informal observation that the tendency to hear velar stops was much stronger following [s] than following [ʃ] led to a series of separate studies of this phenomenon (Mann & Repp, in press). This effect of a preceding fricative on the perceived place of articulation of a stop consonant was also obtained in Experiment 2; we refer to it in Mann and Repp (in press).

3. Bailey and Summerfield (1980) showed that the amount of silence needed to hear a stop between a fricative and a vowel decreases (1) as the extent of the first-formant (F1) transition increases and (2) as F1 onset frequency decreases. These two factors, which are often correlated, were dissociated in our stimuli: [ta] had a larger F1 transition but also a higher F1 onset than [tu]. Three subjects heard stops more readily in [ta], while three other subjects showed the opposite. These individual differences may reflect differential sensitivity to one or the other factor.

4. In pilot studies to Experiment 2, there were two (out of seven) subjects who showed a reduction in the vowel context effect as stop consonants began to be heard. Both subjects, B.H.R. and G.E., also participated in the present study, but only one (B.H.R., one of the authors) continued to show a slight reduction in the context effect at the stop/no-stop boundary.

5. Of these three subjects, one heard far too many instances of [ʃ] in the gap condition, as well as unusual following consonants, such as [n] and [l]. A second subject responded erratically to the fricative noises and heard many instances of [n]. The third subject heard no stop consonants at all (as in the no-gap condition) and showed a nearly random pattern of fricative responses. All three subjects, however, gave a regular pattern of results in the no-gap condition, similar to that exhibited by the other subjects. Their no-gap results were excluded to make possible a pure within-subject comparison of the no-gap and gap conditions.

6. Due to final "null frames" in synthesis (commonly used to avoid transients), these fricative noises had a 12-msec silent interval at the end, which we forgot to delete from the digitized waveforms. Thus, there was an unintended 12-msec interval of silence between fricative noise and periodic portion in the no-gap stimuli. However, since this interval was not sufficient to lead to the perception of stop consonants, its presence was considered inconsequential. (This conjecture was strongly supported by the results of an earlier run-through of Experiment 4, which had used stimuli without any silence but had to be repeated because of a different flaw.) For expository reasons, we will continue to refer to this condition as the no-gap condition. The intended interval in the gap condition was 75 msec; its actual duration is correctly described in the text as 87 msec.

7. The pattern of stop-consonant identification responses in the gap condition showed that the difference between [ʃ] and [s] transitions was perceptually salient even if these transitions did not serve their original function of cuing place of articulation of a fricative, but instead were interpreted as cues to place of articulation of a stop consonant: Stimuli with [ʃ]-transitions received more "k" responses than stimuli with [s]-transitions. For a detailed discussion of these data, we refer to an earlier report (Repp & Mann, Note 7).

APPENDIX A

Coarticulatory Effect of Following Vowel on Fricative Noise Spectrum

We have been unable to find in the literature systematic spectral measurements of American English fricative noises in [-a] and [-u] context. Therefore, we collected some data of our own. Three male native speakers of American English, all experienced phoneticians, spoke the syllables [sa], [ʃa], [su], and [ʃu] as part of a random list containing a number of other utterances. Speaker L.R. provided 12 tokens of each fricative-vowel syllable; Speakers A.A. and L.L., 10 each. The utterances were recorded in a soundproof booth and subsequently digitized at 10 kHz using the Haskins Laboratories PCM system. Spectral cross-sections of the fricative noises were generated in 12.8-msec steps using a Federal Scientific UA-6A spectrum analyzer with a 25.6-msec window. The frequency of the lowest prominent energy peak (presumably coincident with the first pole) was recorded for each individual section. The measurements were then averaged across all sections, resulting in an estimate of the average first-pole frequency of a given noise token. For each speaker, averages and standard deviations of these estimates are displayed in Table 2.

It is clear from inspection of Table 2 that the frequency of the first pole was consistently lower in [-u] context than in [-a] context, both for [ʃ] and for [s]. T tests showed all individual comparisons to be significant ($p < .01$), although the extent of the context effect varied somewhat across speakers. The average shifts were 229 Hz for [ʃ] and 310 Hz for [s]. By comparison, the average perceptual context effect in Experiment 1 (Condition 2), when expressed in terms of the first-pole dimension of the synthetic fricative noise continuum, was about 500 Hz. Our data confirm Kunisaki and Fujisaki (Note 3) by showing that the perceptual effects are considerably larger than the acoustic ones, which suggests either that listeners overcompensate in perception or—more likely—that intrinsic knowledge of coarticulatory dependencies in production is not the only factor guiding perception.

APPENDIX B

Role of Relative Amplitudes of Stimulus Components

One factor that may have had some influence on our results and to which we, at first, did not give sufficient attention is the amplitude relationship among the various

Table 2
Acoustic Variation of [ʃ] and [s] Noise Contingent on the Following Vowel

	Frequency of Lowest Spectral Peak (in Hertz)					
	Speaker L.R.		Speaker A.A.		Speaker L.L.	
Utterance	Mean	SD	Mean	SD	Mean	SD
[ʃa]	2405	94	2548	56	2256	104
[ʃu]	2115	133	2380	88	2028	100
[sa]	3773	149	3688	156	3408	76
[su]	3563	116	3124	56	3252	76

acoustic components of our stimuli. In this section, we first describe the amplitude characteristics of the stimuli used in Experiments 1-4, together with some relevant observations on natural speech, and then report a control study designed to reveal the role played by relative amplitude levels in perception of the [ʃ]-[s] distinction.

Since, at the time of these studies, we did not have easy access to an automatic amplitude display, we determined relative amplitudes by excerpting a brief waveform segment (20-50 msec) from the peak amplitude region of a given stimulus portion, iterating that segment continuously, and getting a reading from the VU meter of a Crown Series 800 tape recorder. Waveform amplitudes were changed digitally to bring signals into the limited range of the VU-meter scale; an appropriate value was then added to or subtracted from the meter reading.

Relative Amplitudes of Fricative Noises

For Experiments 1 and 3, the fricative noises of the nine-member synthetic continuum had identical amplitude specifications at the synthesis stage. The OVE IIIc synthesizer, however, produced output of unequal amplitude, owing to built-in dependencies between the fricative-pole amplitudes, which are intended to mimic natural speech. For this reason, the peak amplitude of the noises actually increased by about 4 dB from the most [ʃ]-like stimulus (No. 1) to the most [s]-like stimulus (No. 9). In Experiments 2 and 4, we decided to eliminate this amplitude gradient. In Experiment 2, this was done by adjusting the amplitude specifications at the synthesis stage; in Experiment 4, the digitized wave-forms of the noises were adjusted instead, which had the advantage of preserving amplitude contours. In each case, later measurements indicated that we had somewhat overshot our goal, resulting in a reversed amplitude gradient (in favor of [ʃ]) of about 2 dB. In any case, although our manipulations of the relative amplitudes of [ʃ] and [s] might have had a slight effect on the precise location of the [ʃ]-[s] boundary (see below), they—unlike the amplitude relationships discussed in the following paragraphs—could not have affected the magnitude of the context effect due to the periodic stimulus portion.

Relative Amplitudes of Periodic Portions

The relative amplitudes of the two periodic portions, appropriate to [a] and [u] (or [ta] and [tu]) are an important parameter: To the extent that they had any effect on fricative perception, their effect was confounded with the effect of vowel quality. We were aware of this problem from the beginning but, in a well-intentioned attempt to deal with it, we may have made it more acute (Experiments 1-3). When we first compared the amplitudes of our synthetic [ta] and [tu] portions, we found [tu] to be 10 dB lower than [ta], apparently due to built-in characteristics of the OVE IIIc synthesizer. This difference seemed too large to be tolerated; therefore, we specified an amplitude for [tu] that was 10 dB higher than that for [ta]. Later measurements indicated, however, that our original assessment of relative amplitudes had been incorrect. In terms of peak amplitudes, the original [ta] and [tu] differed by only 3 dB (in favor of [ta]); hence, the modified periodic portions actually used in our Experiments 1-3 differed by about 7 dB in favor of [tu].

This difference, while not disturbing to listeners, deviates markedly from the normal amplitude relationship between [a] and [u], which seems to be closely approximated by the OVE IIIc synthesizer. Representative values in the literature are 2.0-5.3 dB (Lehiste & Peterson, 1959) and 1.8 dB (Fairbanks, House, & Stevens, 1950) in favor of [a]. The relative amplitudes of the natural [a] and [u] portions in our Experiment 4 were consistent with these previous observations: The peak amplitude of [a] was, on the average, 1 dB above that of [u] for our male speaker and 3 dB above that of [u] for our female speaker. Since Experiment 4 demonstrated a large effect of vowel quality on fricative perception, the analogous effects observed in Experiments 1-3 should, at least in part, have reflected the same factor. However, they may have been enhanced or reduced by the unnatural amplitude difference between the periodic stimulus portions. Below we report a control study that examined that possibility.

Relative Amplitude of Fricative Noise and Periodic Portion

The amplitudes of our fricative noises were relatively low: In Experiments 1 and 3, their peak amplitudes were 18-22 dB (depending on the particular noise) below the amplitude of the [ta] portion, and 25-29 dB below that of the [tu] portion. In Experiment 2, the more [s]-like noises were further attenuated, resulting in amplitude differentials of 22-24 dB with [ta] and 29-31 dB with [tu]. In Experiment 4, the synthetic fricative noises had a more appropriate amplitude relationship to the following periodic portions; the amplitude difference averaged 10 dB but varied from 5 to 15 dB, depending on the particular stimulus. It is possible that the relatively low noise amplitudes in Experiments 1-3 resulted in an artifactual enhancement of the vowel context effect. This issue was also addressed in the control study.

A Control Study

This study involved the orthogonal variation of two factors: (1) the relative amplitude of the two periodic portions, and (2) the amplitude of the fricative noise relative to the periodic portion. All amplitude modifications were performed on the digitized waveforms of synthetic stimuli used in Experiment 1. There were four conditions, each of which contained a random sequence of 140 stimuli, resulting from the combination of seven fricative noises (Stimuli 2-8 of Table 1) with two periodic portions, presented 10 times each. In Condition 1 (which replicated Condition 2 of Experiment 1), [ta] was 7 dB below [tu], and the fricative noises were 18-22 dB below [ta]. In Condition 2, the [tu] portion was attenuated by 14 dB, so that it was now 7 dB below [ta]. In Condition 3, all fricative noises were amplified by 12 dB. In Condition 4, noise amplification was combined with attenuation of [tu]. Thus, the effect of the relative amplitudes of the periodic portions can be assessed by comparing Conditions 1 and 3 with Conditions 2 and 4, and the effect of the amplitude relationship between fricative noise'and periodic portion can be assessed by comparing Conditions 1 and 2 with Conditions 3 and 4.

Eight subjects (ourselves and six colleagues at Haskins Laboratories) listened to the four tapes in counterbalanced order. Four of the listeners were familiar with the stimulus materials; the other four were relatively inexperienced. Combined across subjects, the results showed the expected overall effect of vowel context [F(1,7) = 26.3, p < .001]: All subjects gave more "sh" responses in the [(t)a] context than in the [(t)u] context. Attenuation of [(t)u] by 14 dB resulted in a smaller vowel context effect [F(1,7) = 10.1, p < .025], primarily due to an increase in "sh" responses in the soft-[(t)u] context. However, no subject showed a complete disappearance or even reversal of the context effect in this condition. Expressed as the difference in percent "sh" responses (averaged across the seven members of the noise continuum) between [(t)a] and [(t)u] contexts, the size of the context effect was 12.2% when [(t)u] was loud and 9.1% when [(t)u] was soft. The other experimental manipulation, a 12-dB amplification of fricative noises relative to the periodic portions, surprisingly resulted in an *increase* in the average context effect from 9.4% to 11.8% [F(1,7) = 7.1, p < .05]. There was no significant interaction between the effects of the two amplitude manipulations.

Thus, the results of this control study indicate that although our choice of amplitudes for [(t)a] and [(t)u] in Experiments 1-3 may have resulted in a slight artifactual enhancement of the context effect, that fact may safely be neglected, since we have now shown that context effects of nearly the same magnitude are obtained when the relative amplitudes of [(t)a] and [(t)u] are reversed. Apparently, the relatively low amplitude of our fricative noises in Experiments 1-3 was not responsible for any portion of the context effect; on the contrary, it may have reduced the effect somewhat by making [ʃ] and [s] more difficult to discriminate. Thus, the two minor artifacts may actually have canceled each other.

APPENDIX C

Acoustic Measurements of [ʃ] and [s] Formant Transitions

Having demonstrated large effects of formant transitions on fricative perception (Experiment 4), we wondered about the nature and extent of the difference between the transitions characteristic of [ʃ] and [s]. As far as we know, this difference has not been described in any detail in the literature. Since [ʃ] has a more posterior place of articulation than [s], one might expect the onsets of F2 and F3 in the periodic portion to be less separate for [ʃ] than for [s], in analogy to the pattern of transitions observed for stop consonants with comparable places of articulation. To examine the extent of this difference, we first analyzed the stimuli actually used in Experiment 4 and then, to establish the generality of our observations, measured formant onset frequencies in similar utterances pronounced by four additional speakers.

Rather than using spectrograms which do not offer the necessary resolution, we relied on visual inspection of spectral cross-sections which were generated by a Federal Scientific UA-6A spectrum analyzer and displayed as point plots on a Hewlett-Packard 1300A scope, together with a computer-generated spectrogram and a waveform display. The spectral cross-sections were computed over 25.6-msec time windows, smoothed and preemphasized, with a distance of 12.8 msec between successive sections. The maximal resolution was approximately 40 Hz. One drawback of this equipment was that the precise location of the time

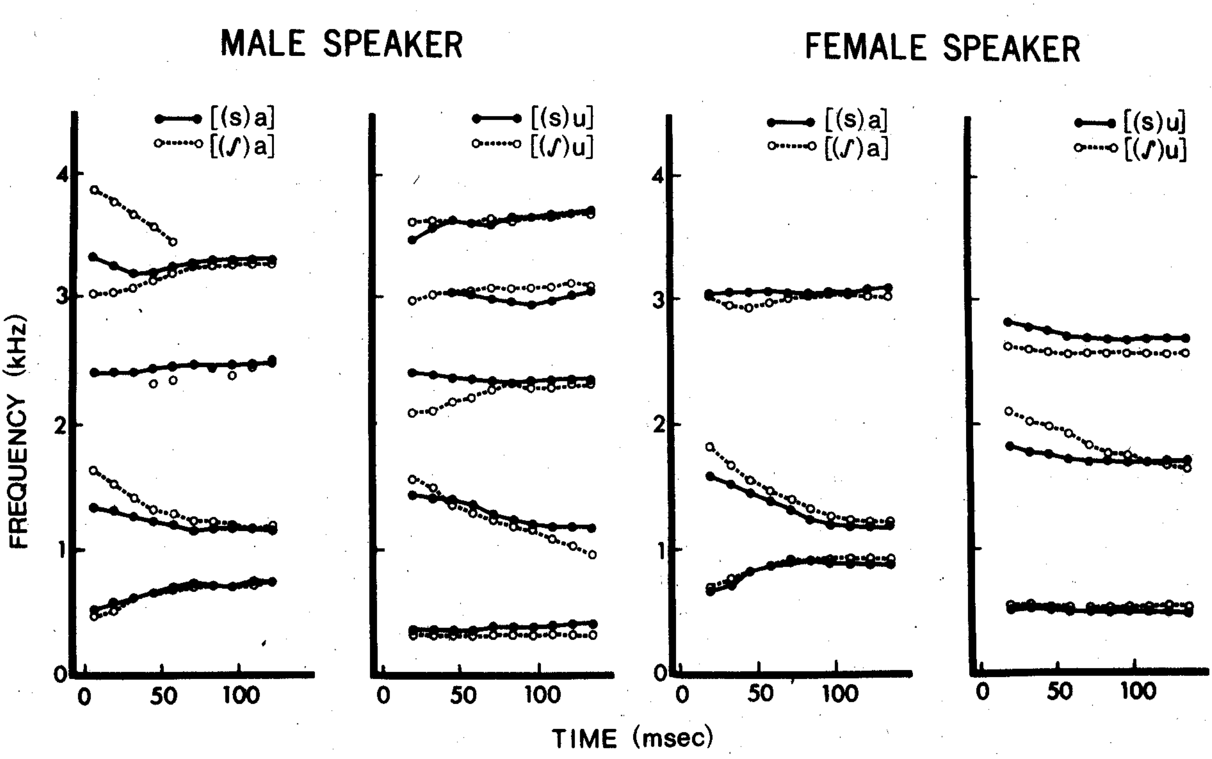

Figure 6. Formant trajectories (averaged over three tokens) over the first 128 msec or so of the natural-speech periodic stimulus portions used in Experiment 4.

windows with respect to signal onset could not be controlled. Thus, the first section included more or less of the silence preceding an utterance. Spectral peaks usually appeared only in the second section, and frequently only the third section provided a clear picture of the formant structure. Therefore, the measurements reported here are conservative with respect to any differences between utterances, since they do not derive strictly from stimulus onset but from 10-20 msec into the periodic portion.

Stimuli of Experiment 4

Spectral peaks were traced through 10 consecutive cross-sections (128 msec) of each isolated periodic portion, starting with the first section that gave a clear formant pattern. It was assumed that most formant transitions would be completed within that period. (The total durations of the periodic portions varied between 300 and 500 msec.) The formant tracings of the three tokens of a given utterance were averaged, omitting spurious peaks not common to all three tokens. In the case of the female speaker, numerous minor peaks due to individual harmonics obscured the course of the formant transitions, especially in [a] stimuli. The formant trajectories in these stimuli were estimated by visual interpolation.

Average formant patterns are shown in Figure 6, separately for the male and female speakers. Five formants could be traced for the male speaker and only three for the female speaker. The results are fairly clear. There seemed to be no systematic differences in F1 between [(ʃ)-] and [(s)-] stimuli. F2 had a higher onset frequency in [(ʃ)-] than in [(s)-] tokens, as predicted. F3 had a lower onset in [(ʃ)u] than in [(s)u], which is also in agreement with the predictions: Higher F2 onset and lower F3 onset reflect a more posterior place of articulation. There was apparently no difference in F3 between [(ʃ)a] and [(s)a], although, for the male speaker, F3 was so weak and inconsistent in [(s)a] that the comparison could not really be made. (Note that F3 was much more pronounced in [(ʃ)a]). In addition, the male speaker showed a curious but consistent pattern in the higher formants of [a] stimuli: While [(s)a] exhibited only one formant, [(ʃ)a] showed two rapidly converging peaks instead. No such difference was observed in [u] stimuli, where both fricative noises were followed by similar, flat F4 and F5 resonances.

Formant Onsets in a Larger Corpus of Utterances

To confirm and extend these measurements, we recorded four additional speakers saying [sa], [ʃa], [su], and [ʃu] 10 times in random order, intermixed with various other utterances. Two of the speakers (A.A. and L.L.) were experienced phoneticians; their utterances had also provided fricative noises for the measurements described in Appendix A. The third speaker (B.H.R.) was one of the authors; he is a native speaker of German and pronounces [s] with a slight lisp, due to irregular dentition. The fourth speaker (S.P.) was a female undergraduate student; her

Table 3
Average Formant Onset Frequencies (in Hertz) Following [ʃ] and [s] Noises

Utterance	F2		F3		F4		F5	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
Speaker A.A. (Male)								
[sa]	*1400*	52	2684	84	3004*	116*	*3772*	232
[ʃa]	*1540*	52	2660	52	3200	52	*4196*	112
[su]	*1444*	72	*2516*	36	3032 3320*	52 32*	*3804*	60
[ʃu]	*1740* 1436	64 76	*2100*	108	3016	28	*4000*	60
Speaker L.L. (Male)								
[sa]	*1400*	60	*2528*	52	*3644*	84	4264	192
[ʃa]	*1708*	52	*2208*	80	*2948*	52	4256	52
[su]	*1448*	52	*2524*	44	*3440*	60	4216	108
[ʃu]	*1764*	72	*2144*	32	*2956*	56	4248 3624*	108 20*
Speaker B.H.R. (Male)								
[sa]	*1388*	56	*2732*	56	3232	92	4000	100
[ʃa]	*1472*	28	*2872* 2192*	80 132*	3232	52	3896*	232*
[su]	1612	48	2520	28	*3208*	40	3780	52
[ʃu]	1684	68	2452	72	*3072*	80	3344* 3996*	72* 72*
Speaker S.P. (Female) ,								
[sa]	*1720*	80	2872	80				
[ʃa]	*2044*	44	2924	40				
[su]	*2076*	60	*3028*	64				
[ʃu]	*2236*	64	*2860*	72				

Note—Values in italics are significantly different from each other ([ʃ] vs. [s]) by t test ($p < .01$).
**Spurious or unreliable peaks.*

formants (F2 and F3 only) proved to be much easier to trace than those of the female speaker in Experiment 4. For each token of each utterance, all major spectral peaks (except F1) were recorded from the first spectral cross-section following the onset of periodicity that showed a clear pattern. (This was sometimes the first, sometimes the second.) Inconsistencies were resolved, as far as possible, by comparisons between tokens and, in some cases, by reexamination. (The measurements of A.A.'s and L.L.'s F2 and F3 onsets were actually done twice, with good agreement.) Means and standard deviations were calculated across the 10 tokens of each utterance, and values for [(ʃ)-] and [(s)-] stimuli were compared by individual t tests. The results are shown in Table 3. Significant differences ($p < .01$) between pairs of values are indicated by italics. The results will be summarized formant by formant.

F2: [(ʃ)-] stimuli consistently showed higher onsets of F2 than [(s)-] stimuli. In one case ([(s)u] vs. [(ʃ)u], Speaker A.A.), five tokens showed a striking difference, while the other five did not, suggesting a rather abrupt transition; therefore, two averages are reported.

F3: F3 had consistently lower onsets in [(ʃ)u] than in [(s)u]. The relationship between [(ʃ)a] and [(s)a] was more variable, just as Figure 6 had suggested. One speaker (L.L.) showed a much lower onset in [(ʃ)a] than in [(s)a]. Two speakers (A.A. and S.P.) showed no clear difference. Speaker B.H.R. showed two peaks in [(ʃ)a] instead of one (usually both at the same time), one strikingly lower, the other somewhat higher than the single peak in [(s)a].

F4: One speaker (L.L.) differed from the other two males by showing extremely large differences in the F4 region: F4 onsets for [(ʃ)-] were 500-700 Hz lower than for [(s)-]. Speaker B.H.R. showed a similar, smaller difference in [u] stimuli only. Speaker A.A. showed a variable pattern: a weak and inconsistent F4 in [(s)a] but not in [(ʃ)a], and an extra peak in [(s)u] but not in [(ʃ)u].

F5: One speaker (A.A.) showed higher onsets of F5 for [(ʃ)-] than for [(s)-], despite large variability in [(s)a] tokens. Speaker L.L., on the other hand, showed no such differences at all. (Spurious peaks at a lower frequency were noted in [(ʃ)u].) Speaker B.H.R. showed weak and inconsistent F5 peaks in [(ʃ)-] utterances, at two different locations in the case of [(ʃ)u], so that no clear picture emerges.

In summary, [ʃ] transitions are characterized by a higher F2 onset, a lower F3 onset (at least in [u]), and, in some speakers, by a lower F4 onset and/or a higher F5 onset than [s] transitions. This pattern, at least as far as the lower formants are concerned, is consistent with the fact that [ʃ] has a more posterior place of articulation than [s]; it confirms our earlier observations depicted in Figure 6.

(Received for publication November 1979;
revision accepted June 30, 1980.)