INFLUENCE OF VOCALIC CONTEXT ON PERCEPTION OF THE [ʃ]-[s] DISTINCTION:
I. TEMPORAL FACTORS

Virginia A. Mann and Bruno H. Repp

Abstract. When synthetic fricative noises from an [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 (through anticipatory lip rounding). We replicated the basic perceptual effect and collected acoustic data from one speaker to corroborate the presence of an analogous coarticulatory effect in production. We also found that varying the duration of the fricative noise leaves the perceptual effect unchanged, whereas insertion of a silent interval following the noise reduces it substantially. Subsequently, we tried to determine whether it is mere temporal separation or the perception of an intervening stop consonant that is responsible for this reduction. The results suggest temporal separation as the important factor, which agrees with recent, analogous observations on anticipatory lip rounding.

INTRODUCTION

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). 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. Zue (Note 1) has demonstrated analogous variations in the spectrum of stop consonant bursts with the following vowel.

This coarticulatory effect has a parallel in speech perception: Listeners' identifications of fricative consonants are influenced by vocalic context. Evidence for such a dependency has been scattered through the literature for some time (Delattre, Liberman, & Cooper, 1962; Hughes & Halle,

Acknowledgment: This research was supported by NICHD Grant HD01994 and BRS Grant RR05595 to the Haskins Laboratories, and by NICHD Postdoctoral Training 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, and Alvin Liberman and David Isenberg for helpful comments on an earlier draft of this paper.

[HASKINS LABORATORIES: Status Report on Speech Research SR-59/60 (1979)]
1956; Hasegawa & Daniloff, Note 2), but the clearest demonstration was provided in a recent study by Kunisaki and Fujisaki (1977). Using a continuum of synthetic fricative noises varying from [ʃ] (low pole frequencies) to [s] (high pole frequencies), these researchers found that the category boundary shifted in favor of [s] responses when the noises were followed by synthetic [u] or [o], relative to the boundaries 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.

It is intriguing to assume that listeners have an intrinsic knowledge of articulatory dynamics, and that their phonetic perception is guided by this knowledge. However, speech perception necessarily makes use of the mechanisms of auditory perception, and there are a variety of psychoacoustic factors that may interact with—or, indeed, prevent—the link between perception and production that presumably underlies speech perception at the highest level. With this in mind, we have investigated some of the temporal and spectral stimulus parameters that influence (and sometimes limit) the effects of vocalic context on the perception of fricatives. In the present paper, we will be concerned with temporal parameters; in a subsequent paper (Repp & Mann, Note 3), we will report our investigations of spectral stimulus properties. By determining the roles played by these parameters, we hoped to constrain the possible psychoacoustic explanations of the perceptual context effect. Furthermore, future investigations of analogous parameters in speech production should enable us to draw a closer parallel between perception and production of fricatives.

EXPERIMENT I

The purpose of our first experiment was to replicate the basic finding of Kunisaki and Fujisaki (1977) that the phonetic perception of a fricative noise depends on the nature of the following vowel, and then to determine 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 vocalic portion. While changes in noise duration, within the range employed by us, have no gross effect on phonetic perception, insertion of a silent interval induces perception of a stop consonant (cf. Bastian, Elmas, & Liberman, 1961; Bailey & Summerfield, in press) and thus changes the phonetic structure of the stimulus. Nevertheless, there was no a priori reason to assume that either of these two temporal manipulations would be more effective than the other in reducing the contextual effect of the vowel on the fricative. If listeners assign a phonetic category to the fricative as soon as some of the noise has been processed, then the temporal distance between noise onset and vocalic portion should be the important variable, and it should not matter whether this distance is varied by extending the duration of the noise or by inserting a silent interval after it.
Method

Subjects. The 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 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. Neither experience nor language seemed to affect the pattern of results, so the data of all 12 subjects were combined.

Stimuli. A synthetic fricative noise continuum was created on the OVEIIlc serial resonance synthesizer at Haskins Laboratories, following in part the specifications given by Kunisaki and Fujisaki (1977). 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. Due to the characteristics of the synthesizer, which are intended to mimic natural speech, the noise amplitude increased from stimulus 1 to stimulus 9 by approximately 12 dB. This characteristic of the stimuli was left intact. Stimulus duration was 100 or 250 msec, depending on the condition.

<table>
<thead>
<tr>
<th>Stimulus</th>
<th>Pole 1</th>
<th>Pole 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1957</td>
<td>3803</td>
</tr>
<tr>
<td>2</td>
<td>2197</td>
<td>3915</td>
</tr>
<tr>
<td>3</td>
<td>2466</td>
<td>4148</td>
</tr>
<tr>
<td>4</td>
<td>2690</td>
<td>4269</td>
</tr>
<tr>
<td>5</td>
<td>2933</td>
<td>4394</td>
</tr>
<tr>
<td>6</td>
<td>3199</td>
<td>4655</td>
</tr>
<tr>
<td>7</td>
<td>3389</td>
<td>4792</td>
</tr>
<tr>
<td>8</td>
<td>3591</td>
<td>4932</td>
</tr>
<tr>
<td>9</td>
<td>3917</td>
<td>5077</td>
</tr>
</tbody>
</table>

In addition to the fricative noise continuum, we synthesized two vocalic stimuli with initial formant transitions roughly appropriate for an alveolar voiceless unaspirated stop consonant: [ta] and [t]. (The formant transitions, which approximated those normally following [s] and [ʃ], were required to make the fricative noise and vocalic portions perceptually coherent; see
Repp and Mann, Note 3.) Each of these stimuli 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, 1233, and 2520 Hz for [a], and 250, 800, and 2295 Hz for [u]. [ta] had 50-msec stepwise-linear transitions in the first and second formants with starting frequencies of 500 and 1796 Hz, respectively. [tu] had a 70-msec stepwise-linear transition in the second formant only, with a starting frequency of 1499 Hz. The amplitudes of [ta] and [tu] were adjusted to be approximately equal. They were 12-24 dB higher than those of the fricative noises which, as mentioned above, varied over a 12-dB range.

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].
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, listeners heard reasonable instances of [fa], [sa], [fu], 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, [fta], [sta], [ftu], and [stu] were perceived. Although [ft] clusters do not occur in 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, leading to a basic set of 21 stimuli. This set was replicated once within each sequence in condition 1, whereas in the other conditions the two different vocalic 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. The five conditions were presented in the same fixed order (1-5) to all subjects, with brief pauses in between. Subjects were seated in a quiet room and listened over Telephonics TDH-39 earphones at a comfortable intensity. The tapes were played back 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 Figure 1b. (The function is duplicated in Figure 1d.) It represents the percentage of "sh" responses to the nine isolated noises (condition 1). It can be seen that all listeners reliably identified the end points 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. With the exception of one subject who gave rather inconsistent responses, all individual category boundaries fell in the vicinity of stimulus 5 (mean = 5.22; standard deviation = 0.41), indicating relatively little variation in response criteria between listeners.

Figures 1a and 1b show the effect of immediately following the fricative noises with a vocalic portion. It can be seen that the predicted effect of vocalic context was obtained: Listeners reported more "sh" sounds 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 (1977), was obviously very large and included even stimuli at the [j]-end of the continuum. Comparison with the baseline results for isolated noises (Figure 1b) 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 Figures 1a and 1b 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 Figures 1c and 1d shows that the introduction of a 150-msec gap between the noise and the vocalic portion practically eliminated the effect. Note that conditions 2 and 3 (Figures 1b and 1c) 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 3 had a different effect on perception than the noise-filled interval in condition 2. There was also an indication of a slight overall decrease in "sh" responses (relative to the baseline) in condition 4 (Figure 1c). This may have been due to the short duration of the noises.

The statistical analysis confirmed these observations. A three-way analysis of variance was conducted on the response percentages summed over all noise stimuli, with the factors Context, Noise Duration, and Gap. Context had a highly significant effect, F(1,11) = 55.7, p < .001, and this effect interacted with Gap, F(1,11) = 62.5, p < .001, but not with Noise Duration, F(1,11) = 1.6. 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).
Separate analyses of conditions 2 and 3 and conditions 4 and 5 confirmed that fricative noise duration had no significant effect on the category boundary shift, regardless of whether a gap was present or not. However, reducing the duration of the noise significantly decreased the number of "sh" responses when the 150-msec gap was present, $F(1,11) = 23.2, p < .001$, but not in the absence of a gap, $F(1,11) = 0.5$. Interestingly, the vowel context effect at the 150-msec gap, though small, was still highly significant, $F(1,11) = 17.6, p < .01$. Thus, although the introduction of the gap substantially reduced the context effect, it did not completely eliminate it.

Discussion

Our results partially replicate the findings of Kunisaki and Fujisaki (1977) on the effects of vocalic context on the perception of the distinction between [ʃ] and [s]. Although their data were presented in a somewhat different format, some comparison parameters can be derived from their figures. There was a striking difference in absolute boundary locations between their listeners and ours. Expressing boundary locations in terms of first-pole frequency, we find average boundary locations for our listeners at approximately 2570 and 3060 Hz in [-t(u)] and [-t(a)] context, respectively (cf. Fig. 1a and Table 1), whereas Japanese subjects showed corresponding boundaries at approximately 3200 and 3900 Hz. This large difference suggests language-specific differences in the [ʃ]-[s] distinction. Moreover, it is evident that the Japanese listeners showed a larger context effect (about 700 Hz) than our subjects (about 500 Hz). However, there were enough changes in detailed stimulus structure and method between their study and ours to account for this difference.

Next, we may ask whether the magnitude of the fricative boundary shift as a function of vocalic context corresponds to the magnitude of the analogous spectral shifts in production. Kunisaki and Fujisaki (1977) report acoustic measurements for Japanese. There, the average shift in first-pole frequency between [-a] and [-u] contexts was about 100 Hz for [ʃ] and 300 Hz for [s], although there was considerable variability. Surprisingly, these differences are considerably smaller than the perceptual boundary shifts obtained in the Japanese study (about 700 Hz).

We have been unable to find in the literature systematic spectral measurements of American English fricative noises in the vocalic contexts that we employed. To get some preliminary impression of the magnitude of the coarticulatory effect, we recorded a male native speaker of American English saying [ʃa], [ʃu], [sa], and [su] twelve times in random order. Subsequently, we digitized these utterances at 10 kHz (using the Haskins Laboratories Pulse Code Modulation system) and examined successive spectral cross-sections (12.8 msec time window) of the fricative noises. In each spectrum, we measured the frequency of the lowest peak (which may or may not have represented the first pole) and subsequently averaged these measurements over all cross-sections of a given token. Finally, we averaged over the 12 tokens of each utterance. These means and the associated standard deviations are shown in Table 2. It is evident that, for this single speaker at least, the first spectral peak was lowered by about 250 Hz in [-u] context, relative to [-a] context. This shift is only about half the size of the perceptual boundary shift found for American listeners, which confirms our similar observation on Japanese lis-
Figure 1. Effect of vocalic context on the [ʃ]-[s] contrast in four conditions.
teners. Thus, these comparisons suggest that listeners' intrinsic knowledge of coarticulatory effects in production is not the only factor affecting perception, or that there is perceptual overcompensation.

Table 2
Acoustic context effects of [a] and [u] on preceding [ʃ] and [s].

<table>
<thead>
<tr>
<th>Utterance</th>
<th>Frequency of first spectral peak in fricative noise (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean (12 tokens)</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>[ʃa]</td>
<td>2405</td>
</tr>
<tr>
<td>[ʃu]</td>
<td>2115</td>
</tr>
<tr>
<td>[sa]</td>
<td>3773</td>
</tr>
<tr>
<td>[su]</td>
<td>3563</td>
</tr>
</tbody>
</table>

The present study extended the Kunisaki-Fujisaki study by examining the effects of two temporal variables on the magnitude of the perceptual boundary shift. We found little effect of a change in the duration of the fricative noise from 100 to 250 msec—a range of durations which exceeds that of normal fricative noises in running speech (cf. Klatt, 1974; Umeda, 1977). This suggests that the critical perceptual information is located at the end of the fricative noise, where it adjoins the vocalic portion, rather than at its onset. The finding that the introduction of a silent gap between the noise and the periodic portion nearly eliminated the effect further demonstrates that the perceptual interaction of the two stimulus portions may depend on their temporal contiguity. This is reasonable, since anticipatory lip rounding in production, if it is not fully established prior to the onset of the utterance, 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. If this is true, it might also be the case that the fricative noise must fall within a certain distance from the vowel in order for its perception to exhibit a contextual influence.

There seems little point in further investigating the variable of noise duration. Given that a 250-msec noise is already beyond the range of durations normally encountered in running speech, extending noise duration further, even though it might eventually lead to a decline of the vowel context effect, would provide data of little relevance to the perception of speech. However, the question of how much separation between noise and periodic portions is needed to prevent contextual effects is of greater theoretical interest. This is so because an additional factor may play a role: the perception of an intervening stop consonant when silence is introduced. Is it temporal separation per se that reduces the contextual
effect, or is it the perception of an intervening phonetic segment? Experiment II was designed to answer that question.

**EXPERIMENT II**

Before conducting an experiment that systematically varied gap size, we collected data for stimuli with a gap duration of 75 msec—halfway between the gap sizes used in Experiment I, 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 I, and the same 12 subjects listened to it in a separate session. The results showed a highly significant 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). Indeed, the difference between the 75-msec 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. Since these data suggested a major decrease in the vowel context effect at gap durations shorter than 75 msec, we decided to focus our attention on these short intervals.

**Method**

Subjects. Nine subjects participated in this experiment, including seven paid volunteers, a research assistant, and the two investigators. Half of one subject’s data were rejected since he gave so few “s” responses in the first session.

Stimuli. The stimuli were similar to those used in Experiment I. One change concerned the amplitudes of the fricative noises. In Experiment I, noise amplitude increased strongly from [ʃ] to [s]. Although that inequality had been built into the synthesizer by its manufacturer, presumably to model natural speech, it seemed somewhat extreme to us and, moreover, was not in accord with observations of our own gathered in the meantime (cf. Repp & Mann, Note 3). We therefore modified the amplitude settings in the synthesizer parameters so as to make all fricative noises approximately equal in amplitude. This was achieved by specifying lower amplitudes for the more [s]-like noises, resulting in a relatively constant amplitude difference between fricative noise and vocalic portion of about 24 dB. The fricative noises were 150 msec long and had 50-msec initial and final amplitude ramps. There were eight gap durations: 0, 10, 20, 30, 40, 50, 100, and 150 msec.

The tape recorded for the experiment contained three random sequences of 144 stimuli, separated by 3-sec ISIs. Each sequence contained the 18 combinations of the nine fricative noises with [tə] and [tul], at each of the eight gap sizes. In contrast to Experiment I, 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 both the fricative and the stop consonant (if present) in each stimulus. The response choices were "s," "sh," "st," "sht," "sk" and "shk."
Figure 2. Effect of vocalic context on the [ʃ]-[s] distinction as a function of silent gap duration and stop consonant perception. Individual subject data from Experiment II.
Results

Assuming that the basic vowel context effect would be replicated when no silence intervened between the noise and the vocalic 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 where stop consonants began to be heard?

The results are shown in Figure 2. The data are displayed separately for the nine subjects, in order to show the considerable individual differences. Each of the nine panels contains four response functions: The two steeply rising ones (thin lines) represent the increase in the percentage of stop responses in [-a] and [-u] context as gap duration was increased; the two more nearly horizontal functions (heavy lines) represent the percentage of "sh" responses (averaged over the whole fricative-noise continuum) in [-a] and [-u] context. The difference between the latter two functions is a measure of the magnitude of the vowel context effect, with a 10-percent difference representing a category boundary shift of roughly 200 Hz on the first-pole-frequency dimension of the synthetic noises.

First of all, it is evident that the basic context effect was indeed replicated: All subjects gave more "sh" responses in the [-t] context than in the [-r] 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 (SL) showed a complete disappearance of the context effect at 40 msec of silence; two other subjects (BHR and PP) showed a progressive reduction up to that interval. The remaining subjects showed little change in the magnitude of 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, at least three subjects still exhibited sizeable context effects at the longest gap duration, 150 msec.

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 limits the occurrence of a context effect between vowel and fricative noise. The stop/no-stop boundaries for four of the nine subjects (BHR, VAM, PP, KH) 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. Three of these subjects heard stops earlier (i.e., at shorter gap durations) in [-a] context than in [-u] context, a finding which is in agreement with results obtained by Bailey and Summerfield (in press). The responses of the remaining five subjects were more irregular. One of them (ML) heard stops at all gap durations, including stimuli without any true silence at all. Three subjects (SW, SL, GE) heard stops in all (or nearly all) [-u] stimuli, regardless of gap size, although they tended to hear no stops in [-a] context on at least some trials when gap duration was short. (Note that this context effect runs counter to that for the four subjects with a more regular response pattern.) The remaining subject (JN) showed no such
context effect but a moderate tendency to hear stops even at short gap durations. 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 perception. Inspection of Figure 2 reveals that the onset of stop consonant perception generally is not accompanied by a marked reduction in the magnitude of the context effect. The only possible exception is subject SL for whom the context effect disappeared as soon as all stimuli were perceived as containing stops. However, this subject (and others as well) showed 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.

Discussion

The results of Experiment II justify the conclusion that the perception of an intervening stop consonant has relatively little influence on the effect 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 vocalic portion. 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, the large individual differences and the temporal extent of the context effect for some listeners suggest that it may be difficult to compare temporal parameters between speech perception and production. In perception, and perhaps in production as well, individual strategies may modify whatever basic, underlying phenomenon there may be. In the present case, for example, individually varying tendencies to perceive the fricative noise either as forming a unit with the vocalic portion or as "streaming off" as a separate auditory event may have played a role. Perhaps the context effect could be extended over arbitrarily long temporal separations if listeners made an effort to integrate the fricative and CV portions into a single perceptual unit. The individual differences observed in the present study may in part have derived from differences in strategies of perceptual integration.

The results of Experiment II speak to a question that we address more directly in experiments reported in a separate paper (Repp & Mann, Note 3): Is the context effect of the vocalic portion on the fricative indeed due to vowel quality itself—as we have assumed all along—or is it perhaps due, in part or entirely, to the initial formant transitions of the vocalic portions? Although, following the seminal study of Harris (1958), vocalic formant transitions were believed to be unimportant for the [ʃ]-[s] contrast, recent experiments by Whalen (1979) 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). The formant transitions of [ta] and [tu] in the present experiments were chosen on the basis of the experimenters' intuitions, and it could have been the case that one set of transitions favored "s" (or "sh") percepts more than did the other. However, two observations suggest that the context effects observed in Experiments I and II
were largely due to vowel quality. First, it seems reasonable to argue that, as soon as the formant transitions are interpreted as a cue to place of articulation of an intervening stop consonant (rather than of the initial fricative), they should lose whatever effect they might have had on fricative perception when no stop was heard. If this hypothesis is correct, then any context effects that are observed despite an intervening stop percept—and Experiment II provided ample evidence for such effects—must be due to vowel quality alone.

Second, and more importantly, if the context effect—especially at short gap durations—had been due to the formant transitions in [ta] and [tu] acting as cues to fricative place of articulation, then the transitions of [tu] should have been more appropriate for a forward (dental) place of articulation (thus favoring "s" percepts) than those of [ta] (which favored "sh" percepts, i.e., an alveolar place of articulation). Although both stimuli in isolation sounded to us as beginning with a "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. According to the argument just made, "k" responses should have been less frequent with [tu] than with [ta] if the transitions of [tu] favored a more forward place of articulation ("t" responses, in this instance). In fact, the opposite was observed. 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 subject (SL) showed the opposite pattern, giving "k" responses to [ta] only.8

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. In fact, if the transitions indeed contributed to fricative perception at short gap durations (when no stop was heard), the transition effect may have partially cancelled the vowel quality effect in these subjects, and this may have been the reason why the reduction in the overall context effect at the stop/no-stop boundary was not more pronounced. In order to investigate this possibility, it will be necessary to dissociate the transition and vowel quality effects experimentally, and then to examine the influence of systematic variations in gap size on the two separate context effects. In a separate paper (Repp & Mann, Note 3), we report experiments that achieve such a dissociation (see also Whalen, 1979) and demonstrate independent effects of both transitions and vowel quality on fricative perception. However, we do not yet know exactly how these two separate effects change with variations in gap size. Until we have this information, our conclusion that the vocalic context effect is unaffected by an intervening phonetic segment must remain tentative. Certainly, however, perception of an intervening stop consonant does not prevent effects of vocalic context on fricative perception. Our conclusion stands that temporal separation is the primary factor affecting the size of the context effect investigated here.

REFERENCE NOTES


2. Hasegawa, A., & Daniloff, R. G. Effects of vowel context upon labelling


REFERENCES

Bailey, P. J., & Summerfield, Q. Some observations on the perception of [s]+stop clusters. Journal of Experimental Psychology: Human Perception and Performance, in press.


FOOTNOTES

1 This adjustment was made in the synthesis parameters. Given equal amplitude parameters, [ta] would have emerged from the synthesizer with
considerably higher amplitude than [tu]. Although this difference is intended to mimic natural speech, we found it undesirable to confound such a large amplitude difference (about 10 dB) with the effect of vowel context we were looking for. Thus, we chose the lesser evil of not preserving the natural amplitude relationships of [ta] and [tu]. Essentially, we believe that amplitude variations will have little influence on the context effect under study but, as yet, we have no data to support this prediction.

It may be the case that [ʃ] and [s] in Japanese have different spectra (and correspondingly, different articulatory positions) than in English, causing native speakers of the two languages to place their perceptual boundaries differently, in accordance with their language experience. An articulatory difference is suggested especially for [ʃ], which by some Japanese linguists is considered a compound phone, [sj], equivalent to a palatalized [s] (e.g., Hattori, 1960). Kunisaki and Fujisaki (1977) and Fujisaki and Kunisaki (1978) report data indicating that the average frequencies for the first pole (formant) of Japanese [ʃ] and [s] lie at about 2800 Hz and 4000 Hz, respectively. The average Japanese perceptual boundary occurred around 3500 Hz on this dimension. On the other hand, Heinz and Stevens (1961) report average first-pole frequencies for American [ʃ] and [s] of approximately 2400 Hz and 4800 Hz, respectively, which suggests that the spectra of these fricatives are more distinct in American English than in Japanese, but provides no clue as to why the perceptual boundary is lower for American listeners (viz., at about 2800 Hz). There are various other factors that might have played a role: the stimulus range employed, the relative amplitudes of the fricative noises (cf. Heinz & Stevens, 1961), the nature of the formant transitions in the vocalic portions (cf. Harris, 1958), and the fact that no zeros (anti-formants) were specified for the present fricative stimuli (cf. Fujisaki & Kunisaki, 1978). There were differences in all these respects between the Japanese study and ours, and it would lead too far to attempt to discuss each in detail. However, it should be noted that Hasegawa and Daniloff (Note 2) synthesized a [ʃ]–[s] continuum by a method rather similar to that of Kunisaki and Fujisaki (1977) and found, for American listeners, a perceptual boundary closer to ours, viz., at about 2700 Hz of first-pole frequency. Thus, cross-language differences in perception and production of these fricatives are indicated.

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

When serving as subjects in Experiments I, we had noticed a tendency to hear velar stops on occasion, even though the periodic stimulus portions in isolation were heard 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, 1979). The present experiment also yielded some interesting results pertaining to the perceived place of articulation of the stop consonant, if one was heard. However, we will report these results in a separate paper.

Bailey and Summerfield (in press) showed that the amount of silence needed to hear a stop between a fricative and a vowel decreases as the extent
of the first-formant transition increases. Our [ta] had a much larger first-formant transition than our [tu].

6 We suspect that the tendency of some subjects to hear stops even at short gap durations was caused by the relatively slow amplitude fall-off (50 msec) of the fricative noise. In pilot studies to Experiment II, we used noises with an abrupt offset, and none of the subjects ever heard stops at the shortest gap durations. Otherwise, the results of these pilot studies were similar to those of Experiment II and therefore are not presented in detail.

7 In the pilot studies mentioned in Footnote 6, 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, BHR and GE, also participated in the present study, but only one (BHR, one of the authors) continued to show a slight reduction in the context effect at the stop/no-stop boundary.

8 Interestingly, SL was the only subject in Experiment II showing clear evidence of no context effect at gap durations beyond 40 msec. For her, the context effect could have been entirely due to the transitions.