# Influence of preceding fricative on stop consonant perception

Virginia A. Mann and Bruno H. Repp

*Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510*
(Received 16 August 1979; accepted for publication 21 August 1980)

The effect of a preceding fricative on the perceived place of stop consonant articulation was investigated in a series of experiments. In experiment 1, we preceded synthetic syllables from two [tV]–[kV] continua with fricative noises appropriate to [ʃ] or [s] and showed that more velar stops are perceived in the context of [s]. Experiment 1 also demonstrated a decrease in the magnitude of this perceptual context effect with increased temporal separation of fricative noise and CV portion, and with introduction of a vowel before the noise, which permitted a subjective syllable boundary after the fricative. Experiment 2 showed that although the effect of the fricative on stop perception declines initially with temporal separation, it may persist in reduced form over intervals as long as 375 ms. Experiment 3 replicated the basic fricative context effect using improved stimuli, but failed to replicate the reduction with an intervening syllable boundary obtained in experiment 1, which presumably was due to the vowel preceding the fricative. Experiments 4 and 5 revealed that the fricative context effect depends both on the phonetic category assigned to the fricative and on the specific spectral properties of the fricative noise. Experiment 5 also revealed a reciprocal effect of the stop consonant on fricative identification. We hypothesize that these perceptual effects serve to compensate for coarticulatory dependencies between stop consonants and preceding fricatives.

PACS numbers: 43.70.Dn, 43.70.Ve

## INTRODUCTION

Articulatory gestures are dynamic, overlapping events. As a consequence, the acoustic information for individual phonetic segments is distributed in time, as well as intertwined with and affected by information for neighboring segments. Thus the listener who is recovering the segmental structure of an utterance must not only integrate temporally distributed cues into unitary phonetic percepts, he must also take into account certain dependencies between cues for different phonetic segments. That he does so is revealed not only by his success in dealing with the natural consequences of coarticulation, but also by his perception of synthetic speech in which coarticulatory effects have been deliberately removed or distorted. Just as he is able to achieve perceptual constancy in the face of natural acoustic variability, so he may show perceptual variability when one and the same acoustic segment is placed in different contexts. Although there could be other causes of such "context effects," at least some seem to reflect the listener's expectations of specific contextual dependencies in the acoustic signal originating in coarticulation.

Several perceptual context effects of this kind have been documented. Perhaps the best-known example concerns the perception of release bursts preceding a vocalic segment: When a synthetic noise burst with a center frequency of about 1600 Hz precedes a steady-state periodic portion appropriate to [i] or [u], it leads to the perception of [pi] or [pu]; however, when the same noise is followed by a periodic portion perceived as [ɑ], listeners report hearing [kɑ] (Liberman, Delattre, and Cooper, 1952). Another example concerns an effect of vocalic context on the perception of fricatives. When synthetic fricative noises drawn from a [ʃ]–[s] continuum are followed by various vocalic portions, listeners hear more instances of [s] in the context of rounded vowels such as [u] than in the context of unrounded vowels such as [ɑ] (Kunisaki and Fujisaki, 1977; Mann and Repp, in press). Both of these perceptual context effects are known to correspond to contextual dependencies in the acoustic signal which are consequences of coarticulation.

In the present experiments, we investigated a context effect which, to our knowledge, has not been previously described: the influence of a preceding fricative on the perceived place of articulation of a following stop consonant. (For the sake of simplicity, we often refer to this effect in the following as the "context effect.") In an earlier study (Mann and Repp, in press), we accidentally discovered that listeners gave more "k" responses to a stop ambiguous between [t] and [k] when it followed [s] than when it followed [ʃ]. In experiment 1, we assessed this effect more precisely and also explored how its magnitude is affected by increases in the temporal separation between the fricative noise and the following part of the signal, and by the introduction of a subjective syllable boundary between fricative and stop. Experiment 2 extended our investigation of the effects of temporal separation to longer intervals than those used in experiment 1, and experiment 3 attempted to replicate the syllable boundary effect. In experiments 4 and 5, we investigated whether the effect of the fricative on the stop is primarily due to spectral characteristics of the fricative noise or to the phonetic category to which subjects assign that noise. Experiment 4 also assessed the role of fricative noise amplitude, and experiment 5 examined whether the stop exerts a reciprocal effect on fricative identification.

## I. EXPERIMENT 1

In this experiment, our goal was to demonstrate that [ʃ] and [s] differentially affect listeners' identification of following stop consonants ambiguous between [t] and [k], and to explore some of the factors that might in-

 0001-4966/81/020548-11$00.80  

fluence the magnitude of this effect. There were five different experimental conditions. In the first of these (the CV condition), stimuli drawn from synthetic [tɑ]–[kɑ] and [tu]–[ku] continua were presented for identification. Subjects' responses in this condition provided a baseline measure of stop-consonant perception. In the remaining four conditions, the same CV stimuli had to be identified in the context of a preceding [ʃ] or [s].[1] Here we chose to vary orthogonally two factors: The temporal separation of fricative noise (F) and CV portion (75 or 150 ms)[2], and the presence or absence of a vowel preceding the fricative (VFCV versus FCV). As we will explain below, the second factor was intended to manipulate the presence or absence of a syllable boundary between fricative and stop.

### A. Method

#### *1. Subjects*

Eleven adults participated as subjects. They included eight paid volunteers with limited experience in listening to synthetic speech, a research assistant, and the two authors. The data of one paid subject were excluded because of unusually high variability. The data were pooled across the remaining ten subjects.

#### *2. Stimuli*

All stimuli were generated on the OVE-IIIc serial resonance synthesizer at Haskins Laboratories. They included nine CV syllables constituting a [tɑ]–[kɑ] continuum, nine syllables constituting a [tu]–[ku] continum, two fricative noises appropriate for [ʃ] and [s], and two steady-state vowels, [ɑ] and [u]. The CV stimuli (perceived, in isolation, as beginning with /d/ or /g/) were periodic throughout, 200 ms in duration, and differed in the onset frequencies of the second and third formants ($F_2$ and $F_3$). For stimuli from the [tɑ]–[kɑ] continuum, the initial formant transitions were stepwise linear and 50 ms in duration. Onset frequency of $F_1$ was held constant at 250 Hz; $F_2$ and $F_3$ onset values are listed in Table I. Steady-state frequencies of the first three formants were 771, 1233, and 2520 Hz, respectively. For stimuli from the [tu]–[ku] continuum, the durations of $F_1$, $F_2$, and $F_3$ transitions (all stepwise linear) were 30, 70, and 80 ms, respectively. Onset frequency of $F_1$ was held constant at 200 Hz; $F_2$ and $F_3$ onset values are listed in Table I. Steady-state frequencies of the first three formants were 250, 800, and 2295 Hz, respectively. In all stimuli, fundamental frequency fell linearly from 110 to 80 Hz, and amplitude rose during the first 80 ms and remained steady for the remaining 120 ms.

The two fricative noises were the endpoints of a nine-member continuum used in a previous study of the [ʃ]–[s] distinction (Mann and Repp, in press); they were considered to be reasonably appropriate for their respective categories. Each was characterized by two steady-state resonances (poles) produced by the fricative circuit of the synthesizer (see Table II). Noise duration was 150 ms; amplitude rose during the first 50 ms, remained steady for the next 50 ms, and fell during the final 50 ms. Due to synthesizer characteristics, the peak amplitude of [ʃ] was about 7 dB below that of [s]. The peak amplitudes of the fricative noises were more than 20 dB below those of the vocalic portions.

The two steady-state vowels, [ɑ] and [u], had the same specifications as the steady-state portions of the CV stimuli; they were 100 ms in duration, had a linearly falling fundamental frequency contour (110–80 Hz), and a constant amplitude with initial ramp.

For each of the five conditions, random stimulus sequences were recorded directly from the synthesizer onto magnetic tape. A three-second interval separated individual stimuli, and there were longer pauses between sequences. The *CV condition* employed five sequences of 42 stimuli each. Within each block of 42, half of the stimuli were from the [tɑ]–[kɑ] continuum and half were from the [tu]–[ku] continuum; the nine stimuli from each continuum were presented with unequal frequencies according to a 1–2–3–3–3–3–3–2–1 schedule. (Thus more responses were obtained for the more ambiguous stimuli from the middle of each continuum than for the less ambiguous endpoint stimuli.) Stimuli for the two fricative–stop–vowel (FCV) conditions were formed by following the [ʃ] and [s] noises by each of the CV syllables, thus yielding twice as many stimuli as in the CV condition. In the *F(75)CV condition*, F and CV were separated by a 75-ms period of silence; in the *F(150)CV condition*, the silent gap was 150 ms long. The tapes for each of these conditions contained five random sequences of 84 stimuli. Frequency of stimulus presentation followed the same schedule as in

TABLE I. Formant transition onset frequencies (Hz) for two CV continua.

| Stimulus No. | [tɑ]–[kɑ] $F_2$ | [tɑ]–[kɑ] $F_3$ | [tu]–[ku] $F_2$ | [tu]–[ku] $F_3$ |
|---|---|---|---|---|
| 1 | 1790 | 2709 | 1796 | 2502 |
| 2 | 1770 | 2576 | 1744 | 2379 |
| 3 | 1744 | 2449 | 1695 | 2245 |
| 4 | 1719 | 2338 | 1646 | 2119 |
| 5 | 1695 | 2197 | 1600 | 2000 |
| 6 | 1670 | 2074 | 1554 | 1874 |
| 7 | 1646 | 1943 | 1499 | 1744 |
| 8 | 1623 | 1821 | 1456 | 1622 |
| 9 | 1600 | 1694 | 1404 | 1499 |

TABLE II. Pole center frequencies (Hz) of fricative noises in a [ʃ]–[s] continuum.[a]

| Stimulus No. | Pole 1 | Pole 2 | Used in experiments |
|---|---|---|---|
| 1 | 1957 | 3803 | 1–5 |
| 3 | 2466 | 4148 | 4 |
| 4 | 2690 | 4269 | 5 |
| 5 | 2933 | 4394 | 5 |
| 6 | 3199 | 4655 | 5 |
| 7 | 3389 | 4792 | 4 |
| 9 | 3917 | 5077 | 1–5 |

[a] These values are synthesizer input parameters. Spectral analysis of the synthesizer output showed the actual pole center frequencies to be about 5% lower.

the CV condition. Stimuli for the *VF(75)CV* and *VF(150)CV conditions* were created by immediately preceding the stimuli from the FCV conditions with a steady-state [ɑ] or [u]. The tape for each of these two conditions consisted of four sequences of 168 stimuli each. The first and fourth sequences contained initial [ɑ], the second and third contained initial [u]. The results were pooled across these two initial vowels, whose quality difference seemed to have little influence on listeners' responses.

### 3. Procedure

Each subject participated in two 90-min sessions during which he or she was seated in a quiet room, listening over Telephonics TDH-39 earphones at a comfortable intensity. The CV tape was presented at the beginning of each session, followed by two of the other tapes, chosen to have the same gap duration. Except for the CV tape, order of presentation was counterbalanced across subjects. Appropriate rest periods followed each tape. Each subject gave a total of 15 responses to stimuli drawn from the centers of the CV continua in the CV (heard twice, once in each session) and FCV conditions, 12 responses in the VFCV conditions.

In all conditions, the subjects were asked to identify both consonants. The response choices for the fricatives were "sh" and "s." The response choices for the stops were "d" and "g" in the CV condition, but "t" and "k" in the two FCV conditions; due to the phonology of English, voiceless unaspirated [t] and [k] are transcribed as "d" and "g," respectively, in syllable-initial position, but as "t" and "k," respectively, when preceded by a fricative perceived as belonging to the same syllable. In the VFCV conditions, the subjects were asked to respond with "d" or "g." By these instructions, we effectively induced the subjects to place a syllable boundary between fricative and stop, since syllable-initial [sd], [sg], [ʃd], and [ʃg] clusters violate the phonological rules of English (and more so, it seems to us, than [ʃt] and [ʃk]).

## B. Results

Two subjects tended to make errors in labeling the fricatives (12.8% and 13.1%, respectively). However, exclusion of their data did not change the pattern of results; therefore, they were maintained in the data pool. The average fricative error rate of the other eight subjects was 1.4%.

Figures 1 and 2 show average percentages of alveolar stop ("d" or "t") responses as a function of position along each CV continuum. Figure 1 shows the results for the [tɑ]-[kɑ] series; Fig. 2 shows those for the [tu]-[ku] series. The response functions obtained in the CV condition are represented by dotted lines. The dotted functions in left and right panels duplicate each other, whereas those in upper and lower panels are independent replications. (Recall that the CV tape was heard twice, to provide separate baselines for conditions differing in gap duration.) It can be seen that the probability of a "d" response was high for the first few stim-

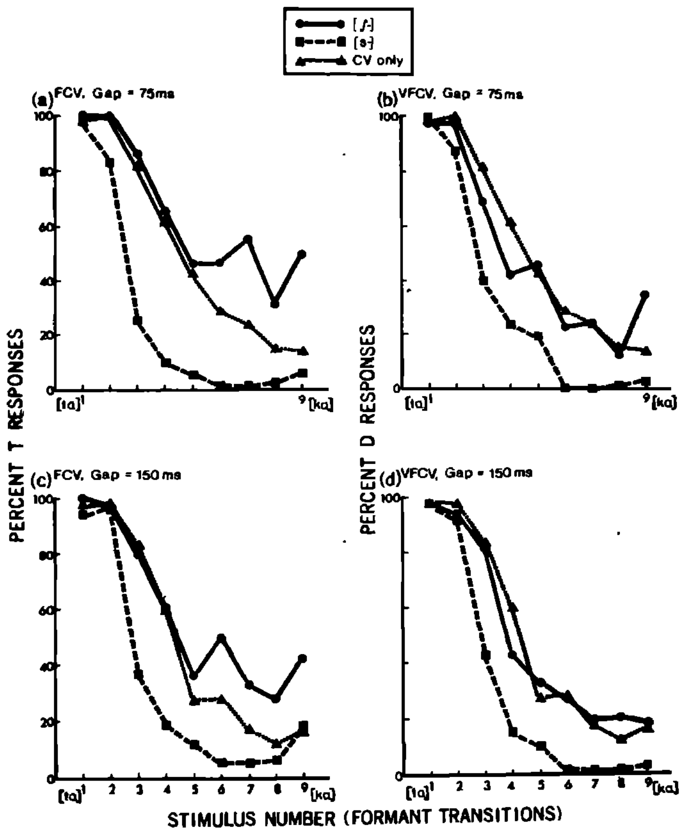


FIG. 1. Effects of [ʃ] and [s] on the perceived place of articulation of a following stop consonant, as a function of gap duration and presence versus absence of a syllable boundary ([tɑ]-[kɑ] continuum, experiment 1).

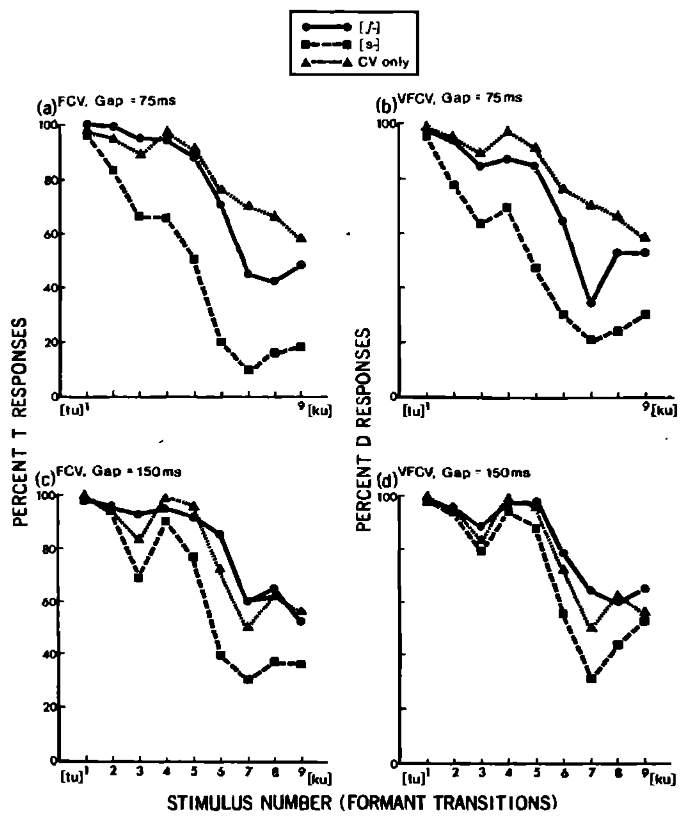


FIG. 2. Effects of [ʃ] and [s] on the perceived place of articulation of a following stop consonant, as a function of gap duration and presence versus absence of a syllable boundary ([tu]-[ku] continuum, experiment 1).

uli in each continuum, and gradually decreased to a minimum for the final few stimuli. Although the two authors, who served as the first subjects, labeled the [tu]–[ku] stimuli very consistently, all other subjects had much difficulty hearing velar stops preceding [u], and gave an overwhelming number of "d" responses. This is evident in Fig. 2: the overall percentage of velar stop responses to stimuli intended to sound like [ku] did not exceed 50%.[3] Despite this unwelcome result, the response distributions on the [tu]–[ku] continuum did change considerably when an initial fricative noise was added to the stimuli; therefore we do report the data here.

The solid lines represent responses to stimuli in which an [ʃ] preceded the CV portion, and the dashed lines represent responses to stimuli with an initial [s]. Comparison of these functions reveals a large effect of fricative context: subjects gave many fewer alveolar stop ("t" or "d") responses in [s] context than in [ʃ] context. This difference was highly significant, both on the [tɑ]–[kɑ] continuum, $F(1,9)=42.9, p<0.0005$, and on the [tu]–[ku] continuum, $F(1,9)=36.3$, $p<0.0005$. The labeling functions for the two fricative contexts were further distinguished by the extent to which they departed from the CV results. Relative to this baseline, the number of alveolar responses was dramatically decreased by the presence of an initial [s] but was not significantly affected by an initial [ʃ].[4]

Comparison of the different panels in Figs. 1 and 2 further reveals the extent to which the magnitude of the context effect was reduced by an increase in the duration of the silent gap, and by the intended placement of a syllable boundary between fricative and stop consonant. In a joint analysis of both CV continua, the effect of gap duration was significant, $F(1,9)=9.8$, $p<0.025$, as was the effect of syllable boundary, $F(1,9)=18.2$, $p<0.005$. While the syllable boundary effect was more pronounced on the [tɑ]–[kɑ] continuum, the gap effect was larger on the [tu]–[ku] continuum. However, the relevant statistical interactions did not reach significance.

Thus experiment 1 showed, consistent with our expectations, that subjects heard more velar stops following [s] than following [ʃ] (or in the absence of any fricative). The difference was remarkably large. How is this effect to be explained? In the Introduction, we indicated that the explanation may lie in coarticulation and the listener's implicit knowledge of articulatory dynamics. However, several alternative hypotheses need to be considered. We postpone a detailed discussion of these alternatives until two further experiments have been described which replicated, extended, and in one respect superseded the results of experiment 1. These two experiments were intended to assess with increased precision the role of temporal separation (experiment 2) and syllable boundaries (experiment 3).

## II. EXPERIMENT 2

Our second experiment was designed to explore further the temporal limits of the context effect observed in experiment 1. Here, we varied the temporal separation from 75 to 375 ms, a value far exceeding normal stop closure durations. We expected to observe a systematic decline and eventual disappearance of the context effect within that range. Clearly, the temporal extent of the effect is an important consideration in any coarticulatory explanation.

### A. Method

#### 1. Subjects

Nine adults participated as subjects. They included six paid volunteers and a research assistant, none of whom had participated in Experiment 1, plus the two authors.

#### 2. Stimuli

The stimuli were very similar to those of experiment 1. The nine members of the [tɑ]–[kɑ] continuum were preceded by either [ʃ] or [s], which was in turn preceded by a steady-state [ɑ]. A silent gap of 75, 150, 225, 300, or 375 ms separated the VF and CV portions. Whereas, in experiment 1, the [ʃ] noise had been 7 dB below the [s] noise, the peak amplitude of the present [ʃ] was adjusted to be about 3 dB higher than that of the [s] noise. The fricative noises were about 21 and 24 dB, respectively, below the peak amplitude of the steady-state [ɑ].

Three sequences of 210 stimuli were recorded, each containing 42 stimuli at each of the five different gap durations. Gap duration varied randomly within each sequence. At each gap duration, half of the stimuli contained initial [s] and half contained initial [ʃ]. Other details were the same as in experiment 1.

#### 3. Procedure

Subjects participated in a single 90-min session in which they listened to the three sequences, paused for a short rest, and then listened to two of the sequences a second time. Thus each subject gave a total of 15 responses to stimuli drawn from the center of the [tɑ]–[kɑ] continuum. By requiring the subjects to label the stops as "d" or "g," we induced placement of a syllable boundary between fricative and stop. This was done so the long silent intervals would not sound unnatural.

### B. Results

Three of the subjects rarely perceived "g" after [ʃ], although they had no such difficulty after [s]. Consequently, they showed exceptionally large context effects at all temporal separations; no systematic decrease was evident, except between the two shortest intervals. The average results of the other six subjects, who showed regular labeling functions, are displayed in Fig. 3. To simplify presentation, responses have been pooled across the nine members of the [tɑ]–[kɑ] continuum. Thus the data are plotted as the average overall percentage of "d" responses as a function of silent gap duration. The solid line represents responses to stimuli which contained [ʃ], the dashed line represents responses to stimuli which contained [s].[5] As in experiment 1, there were more "d" responses in the context of [ʃ] than in the context of [s], $F(1,5)=7.8$, $p<0.05$. At the same time, the magnitude of this context dependency decreased significantly with increased tem-

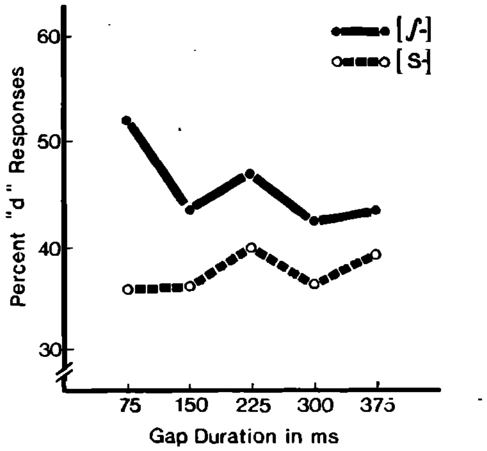


FIG. 3. Effects of [ʃ] and [s] on perceived place of stop consonant articulation, as a function of gap duration (experiment 2).

poral separation between fricative and stop, $F(4, 20) = 3.3$, $p < 0.05$. However, Fig. 3 shows a decline only between 75 and 150 ms, with little change thereafter. Thus the results of these six subjects were essentially similar to those of the three subjects who had difficulty hearing "g" after [ʃ], and inclusion of those three subjects' results would only have reinforced the pattern seen in Fig. 3. One puzzling feature is that although experiment 1 showed that the context effect was primarily due to [s], it was the effect of [ʃ] that changed with temporal separation in experiment 2.

These results show that the context effect may persist over temporal separations of more than 375 ms, even though these separations exceed the durations of natural stop closures, and even though the listeners were led to assume a syllable boundary between fricative and stop, which, according to experiment 1, should have further reduced the context effect to begin with. The data further suggest that there may be two components to the context effect, possibly with different underlying causes. One component rapidly declines with temporal separation and disappears around 150 ms; the other component changes little with temporal separation and persists over intervals beyond 375 ms. We will return to that distinction later in this paper.

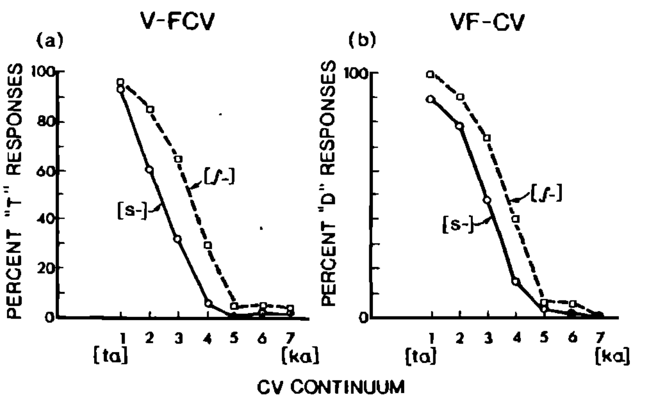


FIG. 4. Effect of syllable boundary instructions (V–FCV versus VF–CV) on the influence of fricatives on stop identification (experiment 3).

## III. EXPERIMENT 3

Experiment 3 had a double purpose—to replicate the basic context effect with an improved set of synthetic stimuli, and to replicate the syllable boundary effect found in experiment 1. In experiment 1, instructions to place a syllable boundary between fricative and stop were confounded with the addition of a vowel preceding the fricative. Although it seemed unlikely that this initial vowel by itself should have been responsible for a reduction in the context effect, it was necessary to rule out that possibility by varying only instructions, holding the stimulus structure constant.

### A. Method

#### *1. Subjects*

Ten subjects participated. They included eight paid volunteers, one of whom had participated in experiment 2, the research assistant, and the second author.

#### *2. Stimuli*

A new, improved [ta]–[ka] continuum was created on the OVE–IIIc synthesizer. This continuum consisted of seven stimuli differentiated only by the $F_3$ transition. The $F_3$ onsets ranged from 2700 to 2100 Hz in 100-Hz steps (plus or minus up to 10 Hz). $F_1$ and $F_2$ started at 285 and 1770 Hz, respectively. The steady-state frequencies of the three lowest formants were 771, 1233, and 2520 Hz. All formant transitions were stepwise-linear and 50 ms in duration. The specified amplitude rise time was likewise 50 ms. Stimulus duration was 250 ms, and fundamental frequency fell linearly from 110 to 80 Hz. Informal listening proved these stimuli to be more naturally sounding and easier to identify than the CV stimuli used in experiments 1 and 2.

The two fricative noises, appropriate for [ʃ] and [s], had the same spectral specifications as those used in experiments 1 and 2, but their duration was 200 ms, they had a triangular amplitude contour which rose during the first 150 ms and fell during the last 50 ms, and they were of approximately equal amplitude. Their amplitude was approximately 12 dB below the steady-state [a].

VFCV stimuli were assembled by following an initial, steady-state [i] vowel of 120-ms duration with either of the two fricative noises, a 75-ms gap, and one of the seven stimuli from the [ta]–[ka] continuum. Thus there were 14 stimuli which were repeated ten times in random order, with interstimulus intervals of 3 s and 6 s after each block of 14. The stimuli in this experiment were digitized at 10 kHz before recording.

#### *3. Procedure*

Each subject listened twice to the same stimulus tape, once with "V–FCV instructions" and once with "VF–CV instructions." Under V–FCV instructions, subjects attempted to place a syllable boundary before the fricative and labeled the stops as "t" and "k." Under VF–CV instructions, they attempted to place the syllable boundary after the fricative and labeled the stops as "d" and "g." In each case, identification of the fricatives as "sh" or "s" was also required. The order of instructions was counterbalanced across subjects.

That the subjects were at least moderately successful in carrying out these instructions was indicated by some listeners' expressions of surprise when, on completion of the experiment, they learned that the same stimuli had been presented twice. Two subjects—the second author and his research assistant—were informed of this fact in advance; nonetheless, they found it possible to maintain a V–FCV or VF–CV "mode" of listening without difficulty and without involuntarily switching back and forth between the two modes.

### B. Results

The pooled results of the ten subjects are shown in Fig. 4. It is evident that the labeling functions for the stop consonant continuum were more regular than in experiment 1, as expected. The basic context effect—fewer alveolar stop responses following [s]—was replicated, $F(1, 9) = 42.8, p < 0.001$. However, there was no indication of any effect of instructions, $F(1, 9) = 0.2$. Even the two experienced listeners, who were convinced that they carried out the instructions successfully, both showed a slightly *larger* context effect in the VC–CV condition. Exclusion of trials on which the fricative was misidentified (3%) did not make any difference.

Thus this well-controlled experiment failed to replicate the "syllable boundary effect" of experiment 1. By default, that effect must be ascribed to the presence versus absence of an initial vowel (VFCV versus FCV). This interpretation is given some support by a comparison of the present results with those of experiment 4, described in more detail below. In that experiment, which immediately followed upon experiment 3, the same subjects listened to stimulus sequences that included the stimuli of experiment 3 with the initial [i] portion removed. Indeed, the context effect was larger in FCV than in VFCV stimuli, although not significantly so. Thus it seems that the context effect is sensitive to stimulus components preceding the fricative noise, but not to the cognitive variable of subjective syllable division.

## IV. DISCUSSION: THREE HYPOTHESES

With the findings of experiments 1–3 in mind, we now proceed to discuss several alternative hypotheses that may be invoked to account for the effect of a preceding fricative on stop consonant perception.

Three hypotheses need to be considered as either alternatives or supplements to the possible perception–production relation hypothesized by us at the outset. The first of these is a response bias account. According to this hypothesis, listeners are, for some reason, biased towards "sk" (or "s–g") responses and against "st" (or "s–d") responses. ([ʃ] + stop clusters do not occur in English in syllable-initial position and therefore should not be subject to any response bias.) Obviously, such a bias would account for the stability of the context effect over relatively long temporal separations (experiment 2); however, it cannot by itself explain the reduction that was observed as gap duration was increased (experiments 1 and 2). One prediction that should hold for any type of response bias is that the label given to the stop should depend on the perceived category of the fricative, not on the precise acoustic stimulus structure.

A second alternative explanation of the context effect studied here is that the fricative noise imposes a psychoacoustic transformation (e.g., through auditory contrast) on the formant transitions of the CV syllables. For example, the high-frequency noise characteristic of [s] might lower the preceived onset frequency of the third-formant transition, thus increasing the perceived compactness of the vocalic onset spectrum—a cue for velar place of articulation (Stevens and Blumstein, 1978). Such a contrast effect would be consistent with the finding that increased temporal separation of fricative and stop reduced the magnitude of the context effect; the persistence of the context effect at temporal separations as long as 375 ms may or may not be compatible with an auditory contrast explanation. Unlike the response bias explanation, the auditory contrast hypothesis predicts variations in the magnitude of the context effect with changes in fricative noise spectrum, regardless of perceived category.

Yet another hypothesis needs to be considered: The offset spectrum of the fricative noise provides a cue to place of stop articulation which is perceptually integrated with the formant transition cues. If this hypothesis were true, we would not be dealing with a true context effect at all, but rather with a demonstration of perceptual integration of temporally separated cues for the same phonetic percept. A number of studies have demonstrated that natural fricative noises contain cues to the place of a following stop occlusion (Uldall, 1964; Malécot and Chermak, 1966; Schwartz, 1967; Bailey and Summerfield, 1980). To explain the present results, the steady-state [s] noise must have provided a cue for velar, rather than alveolar, place of occlusion. Studies by Uldall (1964) and Schwartz (1967) suggest that, on the contrary, steady-state [s] noise favors [t] percepts. Malécot and Chermak (1966), on the other hand, report data from a systematic study with synthetic speech, which lead to the opposite conclusion: When listeners had to identify syllable-final stop consonants from frequency changes in [s] noise alone, "k" responses were more frequent than "t" responses following a steady-state noise. (For more reliable [t] percepts, an upward transition was required in the fricative noise.) Malécot and Chermak also cite parallel observations made in spectrograms of natural speech. Thus what we might call the "noise-offset-cue hypothesis" cannot be rejected *a priori* as a possible account for the fricative–stop context effect. However, while this hypothesis may be compatible with the observed effect of gap duration, it fails to account for any context effects beyond temporal separations of 200 ms, since this seems to be the upper limit of temporal cue integration for stop place of articulation (Repp, 1978; Repp *et al.*, 1978). In addition, the noise-offset-cue hypothesis predicts, as does the auditory contrast hypothesis, that listeners should be quite sensitive to the spectral characteristics of the fricative noise.

Experiments 4 and 5 were designed to provide information crucial to deciding between some of the alternative hypotheses discussed above. The question of interest was whether the effect of the fricative on the following stop was primarily a function of fricative noise spectrum or of perceived fricative category. To that end, we examined the effects of several different fricative noises on the [tɑ]–[kɑ] distinction. A finding that the number of velar stop responses is a continuous function of fricative noise spectrum, regardless of how the fricative is labeled, would be consistent with auditory contrast as the basis of the effects observed in experiment 1. It would also be consistent with an explanation couched in terms of fricative offset spectrum being integrated with the transitional cues to stop consonant place of articulation. However, if assigned phonetic category should prove the major determinant, a response bias explanation would be favored. The articulatory hypothesis, outlined in the Introduction, could accommodate either outcome; therefore the following experiments addressed only the other three hypotheses. We will return to the issue of perception–production relationships in the General Discussion.

## V. EXPERIMENT 4

Experiment 4 had two purposes: (1) to investigate whether changes in fricative–noise spectrum *within fricative categories* would change the magnitude of the context effect, and (2) whether such within-category effects, if obtained, would diminish as the amplitude of the fricative noises is reduced. The presence of within-category effects would support the auditory contrast and noise-offset-cue hypotheses, and sensitivity of these effects to noise amplitude would further point towards an auditory origin. Since fricative–noise amplitudes had varied considerably among our earlier experiments, a controlled study of this factor was called for in any case.

### A. Method

The subjects of experiment 3 continued on to experiment 4 in the same session. The stimuli were nearly the same, with three differences: (1) There were no initial vowels (i.e., the stimuli were of the FCV type); (2) instead of two fricative noises, there were four; (3) the noises existed in two versions, one "normal" (about 12 dB below the steady-state vocalic portion, as in experiment 3), the other attenuated by 10 dB. The four noises were stimuli 1, 3, 7, and 9 from the nine-member continuum (Table II), with the improvements mentioned in connection with experiment 3, and digitized at 10 kHz. Stimuli 1 and 3 sounded [ʃ]-like, stimuli 7 and 9 sounded [s]-like; thus within-category effects could be assessed by comparing 1 with 3, and 7 with 9. Between-category effects, although confounded with a larger spectral difference, could be seen by comparing 1 and 3 with 7 and 9.

Two stimulus tapes were prepared, one with the normal noises and one with the attenuated noises. Each tape contained 10 random sequences of 28 stimuli, resulting from the combination of the four fricative noises with the seven CV stimuli, separated by 75 ms of silence. The order of the two tapes was counterbalanced across subjects.

### B. Results

Figure 5 shows the stop identification results, with a summary of the data in the right-hand panel. As expected, there was a large effect of fricative category (fricative noises 1 and 3 versus 7 and 9), $F(1,9)=50.5$, $p<0.001$. In addition, however, there was a significant within-category effect of fricative spectrum (fricative noises 1 and 7 versus 3 and 9), $F(1,9)=11.5$, $p<0.01$, but this effect was confined to the "sh" category (fricative noises 1 versus 3), as indicated by a significant interaction, $F(1,9)=10.0$, $p<0.02$. Fricative noise amplitude had no effect at all; in fact, the effect of noise spectrum within the "sh" category was slightly larger for the attenuated noises.

The fricative noises were quite consistently identified as either "sh" (stimuli 1 and 3) or "s" (stimuli 7 and 9), regardless of amplitude, although "errors" were less frequent for the endpoint tokens. The error percentages for stimuli 1, 3, 7, and 9 were 0, 4.1, 4.9, and 1.5 at the normal amplitude, and 1.9, 4.9, 2.4, and 1.0 at the lower amplitude. Interestingly, [ʃ]-like stimuli were more often mislabeled as "s" in conjunction with a "k" response (2.1%) than in conjunction with a "t" response (0.6%), while the reverse was true for mislabelings of [s]-like stimuli as "sh" (0.8% versus 1.5%). Exclusion of these

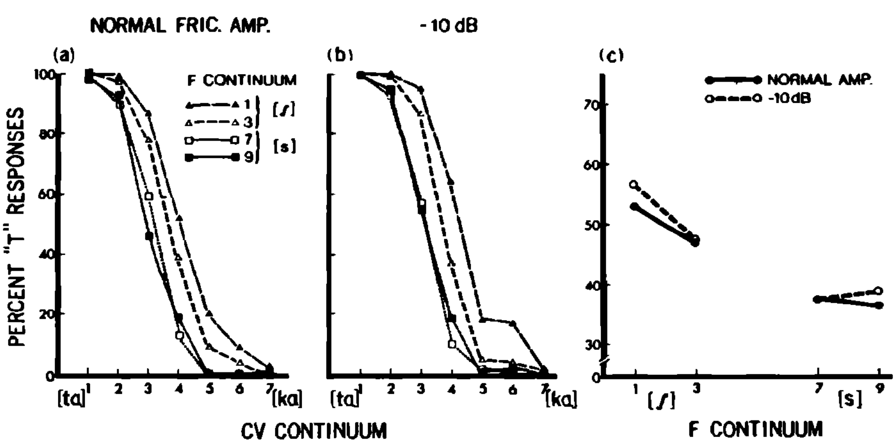


FIG. 5. Effect of four different fricative noises on stop identification, for two different fricative noise amplitudes [panels (a), (b)], summarized in panel (c).

errors from the data did not change the pattern of results.

### C. Discussion

The results of this experiment suggest that the effect of the fricative on stop perception is, at least in part, a continuous function of fricative noise spectrum. This supports the auditory contrast and noise-offset-cue hypotheses. Why effects of fricative spectrum were confined to the "sh" category is not clear at this point. Extrapolating from the trend within the "sh" category [Fig. 5(c)], it seems possible that the fricative context effect is entirely continuous, independent of perceived fricative category. However, the present experiment did not include any ambiguous fricative noises (such as stimulus 5 on the $F$ continuum) which are needed to examine that issue. Experiment 5 served that purpose.

Our failure to find any effects of noise amplitude is reassuring as far as amplitude variability in earlier experiments is concerned. It does not have any theoretical implications, as a positive finding might have had. The error pattern suggested that the stop may have a reciprocal effect on fricative identification. This possibility was further investigated in experiment 5.[6]

## VI. EXPERIMENT 5

Experiment 5 reexamined the nature of within-category effects of fricative noise spectrum on stop identification and, in addition, tested whether the way a given fricative noise is labeled affects identification of the following stop. To that end, this study included several fricative noises ambiguous between [ʃ] and [s]. Thus it tested whether, in addition to the spectral effects already demonstrated, there is a categorial component to the context effect.

### A. Method

#### *1. Subjects*

Ten adults served as subjects. They included eight paid volunteers, five of whom had participated in experiment 2, and both authors.

#### *2. Stimuli*

The 15 FCV stimuli used in this experiment were formed by pairing each of 5 fricative noises with each of 5 CV stimuli, separated by a constant 75-ms period of silence. The stimulus components were similar to those employed in experiment 2.[7] The fricative noises included the two unambiguous continuum endpoints and three noises ambiguous between [ʃ] and [s]. The ambiguity of these three additional noise stimuli was known from an earlier study of the [ʃ]–[s] distinction (Mann and Repp, in press); they were stimuli 4, 5, and 6 from the noise continuum listed in Table II. The CV stimuli were drawn from the [tɑ]–[kɑ] continuum employed in experiments 1 and 2. They included the two endpoint stimuli and the three stimuli (4, 5, and 6) most ambiguous between [tɑ] and [kɑ] (cf. Fig. 1).

Five randomized sequences of 55 stimuli were recorded directly from the synthesizer. Within each sequence, stimuli which contained two unambiguous components were presented once, stimuli which contained one ambiguous component were presented twice, and those which contained two ambiguous components were presented three times.

#### *3. Procedure*

Each subject participated in a single 1-h session, in which the test tape was presented twice. Thus each subject gave a total of 30 responses to stimuli in which both components were (more or less) ambiguous. The task was to identify the fricative–stop cluster as "st," "sk," "sht," or "shk."

### B. Results

The results for stop consonant identification are shown in Fig. 6. The left panel displays standard labeling functions for the CV stimuli, a separate function for each preceding fricative noise. It is obvious that stop identification was strongly affected by the preceding fricative, as in the previous studies, $F(4,36)=13.2$, $p<0.0005$. The frequency of "t" responses declined steadily as the fricative noise became more [s]-like. To examine whether that decline can be accounted for entirely by the changes in noise spectrum, we turn to the right panel of Fig. 6, which displays the overall percentages of "t" responses contingent on whether the fricative was identified as "sh" or "s."[8] There we see that, although "t" responses declined within each fricative category, there was a sizeable break at $F$ stimulus 5, the noise most consistently ambiguous between [ʃ] and [s]: When a subject called this particular noise "sh," "t" responses were more likely to follow than when it was called "s." This categorical effect was shown by seven of the ten listeners and, due to the very large effects shown by some subjects, was significant, $F(1,9)=7.0$, $p<0.05$.

In contrast to experiment 4, the present data show declines in "t" responses within both fricative categories. In fact, the decline within the "s" category (not obtained in experiment 4) was the more consistent across subjects, $F(2,18)=9.3$ $p<0.005$; that within the "sh" category did not reach conventional levels of

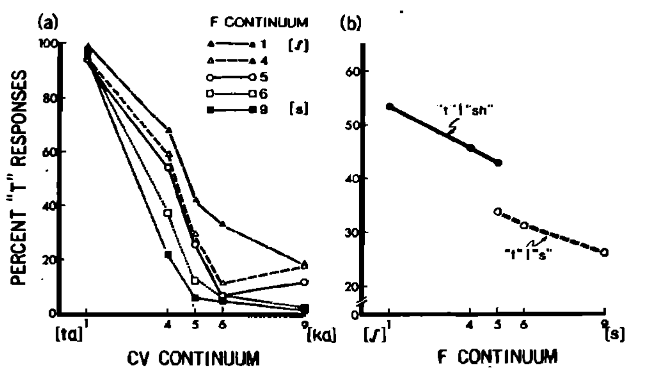


FIG. 6. Effect of five different fricative noises on stop identification [panel (a)], and the continuous and categorical components of that effect [panel (b)].

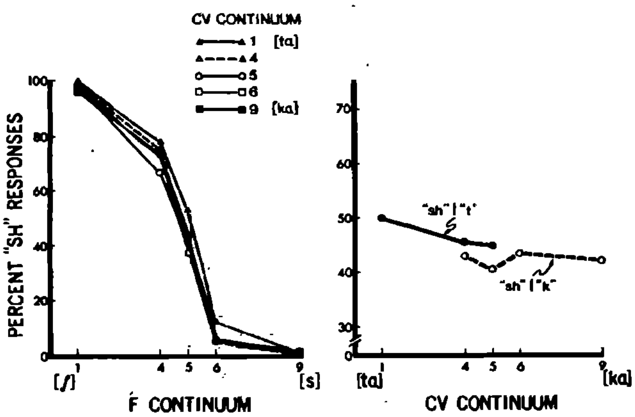


FIG. 7. Effect of five different CV portions on fricative identification [panel (a)], and the continuous and categorical components of that effect [panel (b)].

significance due to large individual differences in trends, $F(2,18)=3.2$, $p<0.10$.

These results demonstrate that, for the majority of listeners, the fricative context effect has both a categorical and a continuous component. However, there are some listeners who seem to lack the categorical component, whereas others show an effect which is almost entirely categorical.

Examination of the fricative identification results revealed an effect of the CV formant transitions, $F(4,36)=6.7$, $p<0.0005$. As Fig. 7(a) shows, however, this effect was very small compared to that of the fricative on stop identification; it was primarily due to CV stimulus 1 which led to somewhat more "sh" responses than the other CV portions. Of course, this reciprocal effect exhibits the same contingencies as that of the fricative on stop identification: "sh" goes with "t," and "s" with "k." To see whether there is any categorical component to the reciprocal context effect, fricative identification was analyzed contingent on stop identification. The results are shown in Fig. 7(b). There were two consistently ambiguous CV stimuli (4 and 5), giving a good opportunity to observe categorical effects. However, the effects obtained were rather small (due in part to the small size of the total effect) and fell short of significance, $F(1,9)=4.5$, $p<0.10$. The decline in "sh" responses within the "t" category was significant, $F(2,18)=7.9$, $p<0.0005$; there was no significant effect within the "k" category.

## VII. GENERAL DISCUSSION

Let us summarize first the results of experiments 1–5. Experiment 1 showed that listeners report more instances of "k" when CV stimuli ambiguous between [tV] and [kV] are preceded by an [s]-like noise than when they are preceded by an [ʃ]-like noise. This context effect was replicated in all subsequent experiments and thus has been established beyond any doubt. Experiments 1 and 2 indicated that, apart from a moderate initial decline, the effect is relatively insensitive to the temporal separation of F and CV, and may persist up to 375 ms of separation and beyond. It also seems to be unaffected by changes in fricative noise amplitude, within the range investigated here (experiment 4), and by placement of a subjective syllable boundary between fricative and stop (experiment 3). There is some evidence, however, that it may be reduced when an initial vowel is added to FCV stimuli (experiment 1 in view of experiment 3, and experiment 3 compared with experiment 4).

Experiments 4 and 5 showed that the effect of the fricative on stop perception depends on perceived fricative category as well as on fricative noise spectrum. There is also a smaller, reciprocal effect of the stop on fricative perception (experiment 5), "s" responses being more frequent in conjunction with "k" responses, and—although the evidence is weak here—this effect may likewise have two components. The pattern of within-category effects of fricative noise spectrum on stop identification varied from study to study, even though highly similar stimuli were involved.[9]

It is tempting to relate the fast-decaying and slow-decaying components suggested by the results of experiment 2 to the continuous and categorical components, respectively. So far, however, we have no evidence that within-category effects decay more rapidly with temporal separation than the categorical effect; we hope to examine this possibility in a future experiment.

The finding that the context effect studied here is partially categorical and partially continuous makes it impossible to rule out any of the candidate theoretical explanations proposed earlier. Rather, we seem to need different explanations for the two components of the effect. Neither auditory contrast nor noise-offset cues to place of stop occlusion can explain the observed categorical effect; however, either of them may account for the somewhat elusive effects of noise spectrum within fricative categories. On the other hand, the categorical effect does fit a response bias hypothesis.

What could be the origin of such a bias? One possibility is that the bias arises from unequal frequencies of [s]+stop sequences in the language. To check this out, we added up the frequencies of all words beginning with these consonant clusters in a standard word-frequency count (Kučera and Francis, 1967). We found that not only is [skɑ-] less than one-fifth as frequent as [stɑ-], but [sk-] is, in general, less than one-third as frequent as [st-]. (The situation is reversed for [stu-] and [sku-], however.) If frequency of occurrence had influenced our subjects' responses, then fewer velar stops should have been reported in the context of [s-(ɑ)]. Clearly, our finding that [s] leads to an increased number of velar stop percepts, regardless of the following vowel, does not favor such an account.[10] We also have evidence (Repp and Mann, forthcoming) that subjects show no preference for "k" to go with "s" responses when presented with isolated fricative noises and asked to guess the following stop. Thus the categorical effect, though bias-like in appearance, cannot derive from a simple response preference.

Let us reconsider now our original suggestion of a perception–production relationship. It seems plausible to us that the increased number of velar stop responses in the context of [s] may arise from listeners' implicit knowledge of certain variations (as yet not confirmed by acoustic measurements) in the transitions for velar and alveolar stops depending on the articulation of the preceding fricative, reflecting a forward shift in the place of tongue-palate contact following [s] relative to the place for [ʃ]. Similarly, the effect of the stop on fricative identification may be due to the absence of appropriate transitions at the offset of the fricative noise. Since coarticulatory effects, such as the ones just referred to, are invariably assimilatory in character, perceptual compensation for them will favor contrasts; hence, listeners' preference for reporting combinations of fricatives and stops with opposite polarities on the place-of-articulation dimension. That the perceptual effect is partially categorical, partially continuous, may indicate that the compensatory mechanism operates both before and after phonetic category decisions. Thus it seems that reference to articulation still provides the most parsimonious explanation of the present results.

At present, we do not have any solid measurements demonstrating that there are, in fact, coarticulatory dependencies of stop consonants on preceding fricatives. However, we do have some perceptual evidence for the existence of such coarticulatory effects (Repp and Mann, forthcoming). In the case of the reciprocal effects of stops on fricative identification, the existence of coarticulation in the form of spectral changes in the fricative noise is well known, but we do not know whether these transitions affect the identification of the fricative itself. This possibility is not implausible, however, since the end of the noise seems to be more important for fricative identification than its beginning (Mann and Repp, in press).

It is possible that the perceptual effects investigated here reflect a very general principle of speech perception: a tendency to differentiate successive phonetic segments along the place-of-articulation dimension (cf. Repp, 1978). Such a tendency would compensate for the general assimilatory nature of coarticulation, regardless of whether or not coarticulation occurs in a specific instance. We suspect, however, that the tie between speech perception and production is much more specific, perceptual context effects occurring only when corresponding coarticulatory dependencies are in fact observed. Our present results may be an instance of such a relationship.

## ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia Mann. Some of our results were reported at the 96th and 97th Meetings of the Acoustical Society of America in Honolulu, Hawaii, December 1978, and in Cambridge, Massachusetts, June 1979, respectively; and at the 10th International Congress of Phonetic Sciences in Copenhagen, August 1979. We thank Alvin Liberman for his advice and encouragement at all stages of our research, and Alice Healy and David Isenberg for helpful comments on an earlier version of this paper.

[1] Although syllable-initial [ʃt] and [ʃk] do not occur in English, neither perception nor production of these clusters—as far as we can see—poses any difficulties to native speakers of English.

[2] It was necessary that these stimulus portions be separated by at least 50 ms of silence in order to assure reliable perception of a stop consonant (see, e.g., Bailey and Summerfield, 1980).

[3] This result was not entirely unexpected. Several of our colleagues have told us of their difficulty in synthesizing [ku] without bursts (cf. Dorman *et al.*, 1977, concerning the importance of the burst in natural tokens of [ku]).

[4] As can be seen in Figs. 1 and 2, an initial [ʃ] somewhat increased the number of alveolar responses on the [tɑ]–[kɑ] continuum, relative to the baseline, but had the opposite effect on the [tu]–[ku] continuum. This was particularly true for the $F(75)$CV condition where the statistical interaction reached significance, $F(1,9)=5.96, p<0.05$.

[5] Errors in fricative identification were not tabulated in this experiment. To the extent that they occurred, they were likely to be independent of gap duration.

[6] We should mention that we have some relevant data from another study that conflict with the present results in that they showed a within-category effect for "s" but not for "sh." In that study (conducted for a different purpose and described in detail as experiment 2 in Mann and Repp, in press), noises drawn from the full nine-member [ʃ]–[s] continuum were followed by a variable silent interval and one of two fixed CV portions, called [tɑ] and [tu]. Since the formant transitions in the CV portions had not been chosen to be optimal for alveolar stops, and since the nine subjects were given the option to identify the stops as either "t" or "k," a large number of "k" responses was obtained, especially to stimuli containing [tu]. As in the present studies, "k" responses were considerably more frequent following [s]-like noises than following [ʃ]-like noises. More detailed examination of the data showed that "k" responses remained at a constant level as long as the fricative was identified as "sh," but increased rapidly as noises identified as "s" became more and more [s]-like. In conjunction with experiment 4, these results show that within-category effects of fricative spectrum may differ quite radically from study to study—an impression that experiment 5 confirmed.

[7] Experiment 5, described last for expository reasons, actually preceded experiments 3 and 4, so that the improved stimuli were not yet available.

[8] A contingent percentage was first obtained for each individual subject and stimulus combination, and these values were subsequently averaged over subjects and over the five stimuli on the CV continuum to obtain the data points in Fig. 6(b). If the responses are pooled before calculating contingent percentages, artifacts are likely to arise. Earlier presentations of these data (see Acknowledgments), in which we concluded that there were no within-category effects, were based on the incorrect analysis procedure.

[9] It is possible that the absence of a spectral effect within the [s] category in experiment 4 was due to digitization of the stimuli, which removed energy above 5000 Hz.

[10] Although experiment 2 cast some doubt on the finding (experiment 1) that the context effect is primarily due to [s] rather than to [ʃ], we have recently replicated that finding (Repp and Mann, forthcoming). Therefore it seems legitimate to focus on the contribution of [s] and to ignore that of [ʃ]. Moreover, since neither [ʃt] nor [ʃk] occur as initial clusters in English, there is no language-related reason why listeners should prefer one over the other.

Bailey, P. J., and Summerfield, A. Q. (1980). "Information in speech: Observations on the perception of [s]-stop clusters," J. Exp. Psychol. HPP 6, 536–563.

Christie, W. M., Jr. (1978). "Some cues for syllable juncture perception in English," J. Acoust. Soc. Am. 55, 819–821.

Dorman, M. F., Studdert-Kennedy, M., and Raphael, L. J. (1977). "Stop consonant perception: Release bursts and formant transitions as functionally equivalent, context-dependent cues," Percept. Psychophys. 22, 109–122.

Kučera, H., and Francis, W. N. (1967). *Computational Analysis of Present-Day American English* (Brown U. P., Providence, RI).

Kunisaki, O., and Fujisaki, H. (1977). "On the influence of context upon perception of voiceless fricative consonants," Annual Bulletin of the Research Institute of Logopedics and Phoniatrics 11, 85–91 (University of Tokyo).

Liberman, A. M., Delattre, P. C., and Cooper, F. C. (1952). "The role of selected stimulus variables in the perception of the unvoiced stop consonants," Am. J. Psychol. 65, 497–516.

Malécot, A., and Chermak, A. (1966). "Place cues for /ptk/ in lower cutoff frequency shifts of contiguous /s/," Lang. Speech 9, 162–169.

Mann, V. A., and Repp, B. H. (in press). "Influence of vocalic context on perception of the [ʃ]–[s] distinction," Percept. Psychophys.

Repp, B. H. (1978). "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants," Percept. Psychophys. 24, 471–485.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. (1978). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol. HPP 4, 621–637.

Repp, B. H., and Mann, V. A. (forthcoming). "Perceptual assessment of fricative–stop coarticulation."

Schwartz, M. F. (1967). "Transitions in American English /s/ as cues to the identity of adjacent stop consonants," J. Acoust. Soc. Am. 42, 897–899.

Stevens, K. N., and Blumstein, S. E. (1978). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. 64, 1358–1368.

Uldall, E. (1964). "Transitions in fricative noise," Lang. Speech 7, 13–14.