Perceptual assessment of fricative–stop coarticulation

Bruno H. Repp and Virginia A. Mann

Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06510

(Received 27 May 1980; accepted for publication 15 October 1980)

The perceptual dependence of stop consonants on preceding fricatives [Mann and Repp, J. Acoust. Soc. Am. 69, 548–558 (1981)] was further investigated in two experiments employing both natural and synthetic speech. These experiments consistently replicated our original finding that listeners, report velar stops following [s]. In addition, our data confirmed earlier reports that natural fricative noises (excerpted from utterances of [stɑ], [skɑ], [ʃtɑ], and [ʃkɑ]) contain cues to the following stop consonants; this was revealed in subjects' identifications of stops from isolated fricative noises and from stimuli consisting of these noises followed by synthetic CV portions drawn from a [tɑ]–[kɑ] continuum. However, these cues in the noise portion could not account for the contextual effect of fricative identity ([ʃ] versus [s]) on stop perception (more "k" responses following [s]). Rather, this effect seems to be related to a coarticulatory influence of a preceding fricative on stop production: Subjects' responses to excised natural CV portions (with bursts and aspiration removed) were biased towards a relatively more forward place of stop articulation when the CVs had originally been preceded by [s]; and the identification of a preceding ambiguous fricative was biased in the direction of the original fricative context in which a given CV portion had been produced. These findings support an articulatory explanation for the effect of preceding fricatives on stop consonant perception.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

In a recent paper (Mann and Repp, 1981), we described a perceptual dependency of stop consonants on preceding fricatives: A stop ambiguous between [t] and [k] was more likely to be labeled "k" when preceded by [s] than when preceded by [ʃ] or by no fricative at all. This perceptual context effect was demonstrated in a series of experiments with synthetic speech. The present experiments employed both natural and synthetic speech to investigate further the possible origins of this effect.

We proposed in our earlier paper that the influence of fricative context on stop perception reflects listeners' perceptual compensation for a coarticulatory influence of fricatives on following stop consonants; an influence which results in a relative forward shift of velar and/or alveolar place of stop occlusion following [s]. Of course, the most direct ways of confirming the existence of such a coarticulatory effect would be to observe ongoing articulation and to measure its consequences in the acoustic signal. We are engaged in such efforts and hope to report their outcome in a separate paper. The present experiments, however, took a more indirect approach. Their purpose was to provide *perceptual* evidence for coarticulation by excerpting portions from natural utterances and examining how listeners identify them, both when presented in isolation and when recombined with (more or less ambiguous) synthetic stimulus portions. Such perceptual assessment of coarticulation, while it cannot replace direct articulatory and acoustic measurements, has the special advantage of revealing whether a given coarticulatory effect has any perceptual significance.

Several previous studies have attempted to assess coarticulation by excerpting acoustically defined segments from natural utterances and presenting them to listeners for identification. For example, Fant *et al.* (1970) and Lehiste and Shockey (1972) used this method to find evidence for effects of different initial vowels on the opening transitions (and of different final vowels on the closing transitions) of stops in VCV utterances; it was used by Benguerel and Adelman (1975) and by Yeni–Komshian and Soli (1979) to find perceptually significant traces of vowel quality in preceding consonants; and by Ali *et al.* (1971) to determine the detectability of vowel nasality due to following nasal consonants. This technique has serious drawbacks, however. When listeners are required to identify phonetic segments whose primary cues have been deleted from the speech signal, the task becomes one of *inference* or guessing rather than perception. On the other hand, when listeners merely report the phonetic segments they actually perceive, performance is often too accurate to be sensitive to small variations in signal parameters.

We have used the "method of isolation" with some success in the present studies (experiments 1B and 2A); however, we have relied, in addition, on a second, novel method which we find especially attractive—the "method of substitution" (experiments 1B and 2C). Instead of omitting a portion of the signal, we replace it with a phonetically ambiguous, synthetic stimulus of similar overall structure. We then test for the presence of perceptually significant coarticulatory traces in the remaining natural signal portion by gauging their power to bias perception of the ambiguous synthetic stimulus towards the phonetic category corresponding to the replaced segment. Thus the synthetic substitute may serve as an indicator of coarticulatory effects, and useful results may be obtained where the method of isolation would yield only chance-level guessing or near-perfect identification.[1]

Below we report two experiments. The first employed natural fricative noises that were excerpted from fricative–stop–vowel (FCV) utterances. By presenting these noises in isolation and in conjunction with synthetic CV portions, we examined the role of coarticulatory cues to stop identity in the fricative noise portion. The second experiment employed natural CV portions from

 0001-4966/81/041154-10$00.80

the same FCV utterances. By presenting these stimuli in isolation and in conjunction with synthetic fricative noises, we endeavored to determine whether CV portions contain coarticulatory traces of the fricative that originally preceded them. Our experiments provide clear perceptual evidence that such traces exist, thus corroborating our hypothesis (Mann and Repp, 1981) that the perceptual influence of preceding fricatives on stop consonant perception has a basis in coarticulation.

I. EXPERIMENT 1

Experiment 1 had three conditions (A, B, C). Those methodological aspects common to all three are described below; specific features are described later under individual headings.

A. General method

1. Subjects

Ten subjects participated. They included seven paid volunteers (some of whom had taken part in earlier experiments employing similar stimuli), a research assistant, and the two authors. Since experience did not seem to influence the basic pattern of results, the data were pooled across subjects in this and subsequent experiments.

2. Stimuli

A male, phonetically trained, native speaker of American English spoke the utterances [sa], [ʃa], [sta], [ska], [ʃta], [ʃka] repeatedly in random order as part of a list containing a number of other utterances. The recordings were made in a soundproof booth using a Shure dynamic microphone and a calibrated Ampex AG–500 tape recorder. Subsequently, the utterances were digitized at 10 kHz and stored in separate files using the Haskins Laboratories Pulse Code Modulation (PCM) system. Three good tokens of each of the six utterances were selected for use in the experiments. The fricative noise was excerpted from each stimulus and stored separately. Acoustic parameters of these noises are given in the Appendix.

In conditions A and C, some of the natural fricative noises were combined with digitized synthetic CV portions drawn from a [ta]–[ka] continuum that had been created on the OVE IIIc synthesizer at Haskins Laboratories. There were seven CV stimuli, distinguished only by the onset frequency of the third formant (F3) which decreased from 3222 Hz for the most [ta]-like stimulus to 1902 Hz for the most [ka]-like stimulus in steps of approximately 215 Hz (plus or minus up to 10 Hz). All stimuli had 50-ms stepwise-linear formant transitions (F1: From 285 to 771 Hz; F2: From 1770 to 1233 Hz; F3: To 2520 Hz) followed by 200 ms of steady-state resonances, a linearly falling fundamental frequency (110 to 80 Hz), and a flat amplitude contour with a 50-ms ramp at onset and a 30-ms ramp at offset. These stimuli were perceived as /da/ or /ga/ in isolation but as /ta/ or /ka/ when preceded by a fricative noise, due to the phonotactic principles of English.

3. Procedure

The subjects listened to the stimulus tapes (described below) in a quiet room at a comfortable intensity, using an Ampex AG–500 tape recorder and Telephonics TDH-39 earphones. The conditions were presented in a single session in fixed order (A, C, B), separated by brief rest periods.

B. Condition A: Replication of basic context effect

The purpose was to replicate the basic finding that listeners are biased to hear "k" rather than "t" in the context of a preceding [s], as compared with a preceding [ʃ] or a null context. To avoid the problems inherent in synthesizing appropriate fricative noises (Mann and Repp, 1981), we used natural fricative noises in conjunction with a synthetic [ta]–[ka] continuum.

1. Method

Listeners first heard a sequence of isolated CV syllables (the seven stimuli from the [ta]–[ka] continuum ten times in random order) which they identified as beginning with "d" or "g." Subsequently, they listened to the same syllables preceded by a fricative noise, plus a 75-ms silent interval. The noises were those excepted from [ʃa] and [sa], and there were three tokens of each. As there were six physically different noises, there were 42 different stimulus combinations which were presented five times in random order. The subjects identified both the fricative ("sh" or "s") and the stop ("t" or "k").

2. Results and discussion

Figure 1 shows the results. Because of the rather wide spacing of the stimuli on the synthetic [ta]–[ka] continuum, listeners' category boundaries were quite sharp, so that the present test of effects of fricative context was conservative. Of the seven CV syllables, only stimulus 4 was ambiguous in isolation, and it was the only one whose perception was affected by a preceding fricative. However, that effect was exactly as predicted: A preceding [ʃ] had no effect relative to the isolated-CV baseline, whereas a preceding [s]

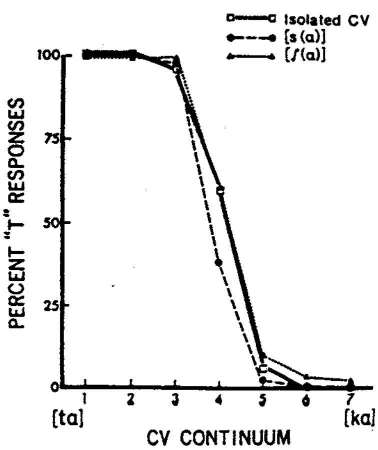

FIG. 1. Effects of preceding [ʃ] or [s] (without cues to stop manner) on stop consonant perception (experiment 1A).

lowered the percentage of "t" responses. This small effect was sufficiently consistent across subjects to be highly significant in a standard repeated-measurements analysis, $F(1,9)=20.4$, $p<0.005$, and it also reached significance when the variation between fricative noise tokens was taken as the error estimate, $F(1,4)=11.3$, $p<0.05$.[2]

Thus we successfully replicated the basic effect of a preceding [s] on stop consonant perception. By replicating the effect with natural fricative noises, we have eliminated any doubts deriving from our earlier use of synthetic noise stimuli. However, the possibility still exists that the natural [s] and [ʃ] noises were not equally neutral as potential cues to place of articulation of a following stop. The next condition addressed this point.

C. Condition B: Identification of stops from F(CV) portions

In part, this condition examined how accurately listeners can identify alveolar and velar stop consonants upon hearing fricative noises excepted from FCV utterances. That cues to stop place of articulation are contained in fricative noises which precede a stop closure has been reported by several researchers (Uldall, 1964; Malécot and Chermak, 1966; Schwartz, 1967; Bailey and Summerfield, 1980). These cues consist of spectral shifts ("transitions") due to progressive narrowing of the vocal tract towards the stop occlusion (see our Appendix). Malécot and Chermak (1966) and Schwartz (1967) have shown that listeners can identify stops fairly accurately from isolated fricative noises containing appropriate spectral shifts. However, the stop most accurately identified is [p], which was not included in our materials. These studies suggest that [t] and [k] are more difficult to identify from fricative-noise transitions alone. Since we were concerned about the potential role of these cues in the influence of preceding fricatives on stop perception, it was important to determine just how salient these cues were.

In addition to the noises exerpted from FCV utterances, we included the noises used in condition A, which derived from FV utterances. We wondered whether listeners' forced-choice stop responses to these latter noises would exhibit a bias towards "k" following [s]. Such a bias would suggest that these noises were not equally neutral as potential cues to place of stop occlusion; or, considering the fact that these noises really did not contain any such cues (according to our own perception and acoustic analysis—see the Appendix), a response bias contingent on fricative identity would be implicated.

1. Method

The fricative noises were excerpted from natural [ʃa], [sa], [ʃta], [sta], [ʃka], [ska]. As there were three different tokens of each noise, there were 18 stimuli altogether which were presented five times in random order. The subjects' task was to identify the fricative as "sh" or "s" and, in addition, to report (or guess) whether that fricative had been originally followed by "t" or "k." The subjects were told that all noises had been excerpted from FCV utterances; they were not informed about the fact that some derived from FV utterances.

2. Results and discussion

The results are shown in Table I. Considering first only the noises derived from FCV utterances, it is clear that the subjects could identify the stop consonants quite well, being correct on 86% of the trials. They were more accurate in identifying [ta] than [ka], $F(1,9)=11.6$, $p<0.01$. They were also somewhat more accurate with stops following [s] rather than [ʃ], $F(1,9)=8.4$, $p<0.05$, particularly where "k" responses were concerned. Both effects were equally significant with token variability as the error estimate. The second effect could be taken as a bias to respond "k" in conjunction with "s". However, the statistical interaction that would have supported such a bias was not significant. Moreover, the responses to the noises deriving from FV utterances did not suggest such a bias: "k" responses were actually more frequent in conjunction with [ʃ] than with [s], although the difference did not reach significance. Furthermore, we note that "k" responses to F(V) noises were slightly more frequent than "t" responses. This indicates that the better identification of [t] than [k] in F(CV) noises was due to the nature of the acoustic information and not to a simple response preference for "*t*."[3]

Thus we find no evidence of a bias to respond "k" in conjunction with "s" when isolated fricative noises are presented. Apparently, the presence of a full FCV stimulus is necessary to evoke that tendency; therefore, it seems unlikely that we are dealing with a response bias contingent on fricative identity.

D. Condition C: Dissociating two effects of preceding fricatives on stop perception

As a further test of the role of cues to place of stop occlusion in the fricative noise, we juxtaposed fricative noise transitions with CV formant transitions, both of which may serve as cues to place of stop occlusion in FCV stimuli. When conflicting *vocalic* formant transitions are juxtaposed (VC–CV), the CV transitions generally dominate preception; or, if the closure interval is sufficiently long (70 ms or more), two different stop consonants are heard in sequence (Repp, 1978; Dorman *et al.*, 1979). By analogy, we expected the noise transitions to be less salient as cues to stop place of articulation than the CV transitions; the question was

TABLE I. Percentages of "t" and "k" responses to isolated fricative noises.

Stimulus	Response "t"	Response "k"
[ʃ(ta)]	91.3	8.7
[ʃ(ka)]	30.7	69.3
[s(ta)]	94.7	5.3
[s(ka)]	12.7	87.3
[ʃ(a)]	40.7	59.3
[s(a)]	54.0	46.0

whether the noise transitions would have any effect whatsoever. At the silent interval used here (75 ms), we did not notice any tendency to hear two different stops ([stk*a*] or the like).

Whether or not listeners assigned any perceptual weight to the fricative noise cues, we expected to find the basic contextual effect of fricative identity on stop perception (more "k" responses following [s]). By aiming at replicating the context effect using natural fricative noises containing appropriate cues to stop articulation, the study effectively avoided the problem of having to decide whether [ʃ] and [s] noises without such cues are equally "neutral" (cf. condition A). Instead, fricative identity and noise transitions were treated as independent variables in a 2×2 factorial design.

1. Method

The seven stimuli from the [t*a*]–[k*a*] continuum were preceded by [ʃ] or [s] noises excerpted from natural [ʃt*a*], [ʃk*a*], [st*a*], [sk*a*], with 75 ms of silence in between. As there were three physically different noises from each context—12 noises in all—there were 84 stimulus combinations which were presented five times in random order. The subjects identified both the fricatives ("sh" or "s") and the stops ("t" or "k").

2. Results and discussion

Figure 2 shows that, despite the relatively sharp category boundary on the [t*a*]–[k*a*] continuum, there were clear effects of the fricative noise on stop identification. First of all, noise transitions did influence stop identification: There were more "t" responses with transitions deriving from [t] than with transitions deriving from [k], $F(1,9)=26.6$, $p<0.0005$. As predicted, however, the CV transitions were the stronger cue to stop place of articulation, for the noise transitions had relatively small effects when the CV transitions were unambiguous. Second, the basic context effect was replicated: There were more "t" responses following [ʃ] than following [s], $F(1,9)=31.5$, $p<0.0005$. Finally, the two effects did not interact statistically, $F(1,9)=4.7$, $p>0.05$, and thus appeared to be independent. The same results were obtained in an analysis by tokens, since token variability was generally small.[4]

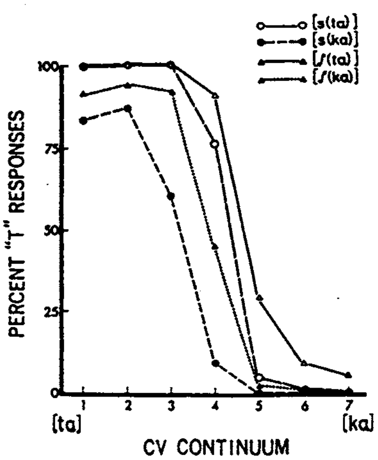

FIG. 2. Effects of preceding [ʃ] or [s] (with cues to stop manner and place of articulation) on stop consonant perception (experiment 1C).

Thus the present data show that the basic context effect of a preceding fricative on stop perception is obtained even when there are cues to place of stop occlusion in the fricative noise portion. This reinforces our earlier conclusion (Mann and Repp, 1981) that there is a context effect due to the fricative *per se*, which is independent of noise properties that directly reflect stop production.

II. EXPERIMENT 2

The results of experiment 1 suggest that the contrasting effect of preceding [ʃ] and [s] on stop perception reflects neither a simple response bias nor an effect of cues to stop place of articulation contained in the fricative noise. By ruling out these alternatives, we have gained indirect support for our hypothesis that the effect derives from perceptual compensation for a coarticulatory influence of a preceding fricative on stop consonant production. In our second experiment, which also comprised three conditions,[5] we attempted to obtain direct evidence for such a coarticulatory dependency by examining in several different ways how listeners respond to natural CV portions that had been originally produced in the context of either [ʃ–] or [s–].

A. General method

1. Subjects

Twelve subjects participated in conditions A and B which were run in a single session in a fixed order (B, A). There were nine paid volunteers, two of whom had been subjects in experiment 1, plus a research assistant and the two authors, all of whom had been subjects in the earlier experiment. These last three subjects participated in two identical sessions whose results were averaged before they were combined with the results of the other subjects, who participated only in a single session. Condition C was conducted at a later time and used a partially different group of ten subjects (seven new paid volunteers, the research assistant, and the two authors).

2. Stimuli

The same natural utterances of [ʃt*a*], [ʃk*a*], [st*a*], and [sk*a*] that had supplied the fricative noises of experiment 1 also provided the CV portions for the present experiments. There were three physically different tokens of each CV stimulus, and each was employed in two versions, one including the release burst and one without the burst. The stimuli with bursts consisted of the total signal portion following the silent closure interval in the source utterances. The burstless stimuli were obtained by deleting all energy preceding the first clear pitch pulse; the deleted portion usually included a small amount of aspiration following the release burst. All in all, there were 24 distinct CV portions. Details of their acoustic structure are reported in the Appendix.

In condition C, these CV portions were preceded by synthetic fricative noises from a nine-member [ʃ]–[s]

continuum; in condition B, just the endpoints of that continuum were used. The fricative noises were distinguished by the center frequencies of two poles generated by the fricative circuit of the OVE IIIc synthesizer. (No zero was specified.) These frequencies are listed in Table II; they increased in roughly equal steps from stimulus 1 ([ʃ]-like) to stimulus 9 ([s]-like). All noises were 200 ms in duration and had approximately equal amplitudes, with a triangular amplitude contour which peaked after 150 ms. They were digitized at 10 kHz.

B. Condition A: Identification of stops from (F)CV portions

This condition provided the most direct perceptual test for the existence of coarticulatory variations in the production of stops following [ʃ] and [s]. In this study, which used the "method of isolation," the subjects' task was to identify the initial (stop) consonants in isolated CV portions, with and without bursts. To the extent that any confusions would occur along the place dimension, we expected these errors to reflect any coarticulatory variation in the CV formant transitions (and perhaps in the release burst) introduced by the original fricative context. Specifically, since coarticulation is generally assimilative, a stop following [s] might exhibit transitions reflecting a more forward place of articulation than a stop following [ʃ], because [s] has a more forward place of articulation than [ʃ]. Therefore, if such coarticulatory effects exist, we expected errors in stop identification to be biased towards a forward place of articulation when the CV portion had originally been preceded by [s]. It was considered possible that this effect, if obtained, would be more pronounced for (intended) [k] than for [t], since velar place of articulation might have more freedom to shift than alveolar place of articulation (as evidenced by the existence of two major allophones of velar stops in English). Also, judging from our earlier perceptual results and from our introspections on fricative-stop articulation, coarticulatory shifts in stop place of articulation should be primarily due to [s]. This implies that stops originally preceded by [ʃ] should be more accurately identified than those originally preceded by [s].

TABLE II. Pole frequencies of fricative noises (Hz).[a]

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

[a] The values given are synthesizer input parameters. Measurements of the acoustic output suggested that the actual pole center frequencies were about 5% lower. Some irregularities in step size were caused by our use of prespecified frequency values in conjunction with the limited frequency resolution of the synthesizer. Any effect these irregularities might have had on our results in experiment 2C worked in favor of the null hypothesis.

1. Method

The 24 CV portions (two intended stops, two original fricative contexts, three tokens, with and without burst) were presented five times in random order. The listeners had to identify the initial consonant by forced choice between four alternatives: "b," "th," "d," "g." It was explained that "th" represented the initial sound in *that*; this fricative, whose place of articulation is—roughly speaking—intermediate between "b" and "d," is easily perceived in the absence of any fricative noise and was in fact a frequent response choice.

2. Results and discussion

The listeners' responses are summarized in Table III. First of all, it is immediately evident that stimuli with bursts were much more accurately identified than burstless stimuli. When bursts were present, misidentifications occurred only with [t*a*], and they were primarily "th" responses. These responses, however, were more frequent to [(s)t*a*] than to [(ʃ)t*a*], which is in accord with our hypothesis that [(s)t*a*] has a more forward place of articulation than [(ʃ)t*a*].

This hypothesis is further supported by the pattern of responses to burstless stimuli, which were much less accurately identified. First, we see that [k*a*] was more often "correctly" identified as "g" than [t*a*] as "d," $F(1, 11) = 10.8$, $p < 0.01$—an unexpected result that was apparently due to the large number of "th" responses to [t*a*] stimuli.[6] Second, more "errors" occurred in response to [(s)–] stimuli than to [(ʃ)–] stimuli, $F(1, 11) = 19.4$, $p < 0.002$. Since virtually all errors were in the direction of a more forward place of articulation (except for the rare "g" responses to [t*a*]), the result implies that [(s)–] stimuli had a more forward place of production than [(ʃ)–] stimuli, as predicted.

There were some marked differences between individual stimulus tokens. In particular, one of the three burstless [(s)k*a*] tokens evoked the response pattern characteristic of the [(ʃ)k*a*] tokens. This indicates a fair amount of articulatory variability from utterance to utterance. However, with token variance as the error term, the differences just reported were still significant at the $p < 0.05$ level.

These results confirm our hypothesis of a forward shift in place of stop articulation following [*s*], and,

TABLE III. Identification of stops in isolated (F)CV portions.

	Without burst				With burst			
	Response (%)							
Stimulus	"b"	"th"	"d"	"g"	"b"	"th"	"d"	"g"
[(s)t*a*]	28.0	52.2	17.2[a]	2.5	···	21.1	78.9	···
[(ʃ)t*a*]	5.6	44.7	43.3	6.4	···	8.6	86.6	4.7
[(s)k*a*]	13.6	15.3	21.4	49.7	···	···	···	100.0
[(ʃ)k*a*]	2.2	2.8	13.3	81.7	···	···	···	100.0

[a] Note: "Correct" responses are in bold type.

moreover, are in accord with our perceptual results in suggesting that the shift is, indeed, primarily due to [s]. We cannot tell from these results whether the release bursts conveyed any information about coarticulatory shifts since, in most cases, the presence of a burst seemed to be sufficient for correct identification; therefore, whatever spectral variations occurred in the burst portion were not revealed in listeners' responses. However, the vocalic formant transitions must have varied with the preceding fricative in the manner predicted (see the Appendix), and this variation was, moreover, perceptually significant. Thus we have now support for an articulatory effect that parallels the perceptual context effect observed in our earlier studies.

C. Condition B: Identification of stops in F + (F)CV stimuli

In this condition, the CV stimuli of condition A were presented in the context of an actual preceding [ʃ] or [s]. Thus, in addition to recreating (approximately) the context in which the stops had been originally produced, we had the opportunity to observe any effect of preceding synthetic fricative noises on the perception of stops cued by natural CV portions.

1. Method

The 24 natural CV portions were preceded by either a [ʃ] noise or a [s] noise, the endpoint stimuli of a synthetic noise continuum (see Table II), plus a 75-ms silent gap. The resulting 48 stimuli were presented five times in random order. The subjects' task was to identify the fricative as either "sh" or "s," and the stop as either "p," "t," or "k." Note that, in the context of a preceding fricative, the stops were now to be given voiceless category labels, in conformity with the phonology of English. In contrast to condition A, "th" responses did not seem appropriate here, as [sθ] and [ʃθ] clusters are extremely uncommon and not readily perceived.

2. Results and discussion

The results are displayed in Table IV, separately for stimuli preceded by synthetic [ʃ] and stimuli preceded by synthetic [s]. The fricatives were generally correctly identified (2.7% errors). Without the "th" response category, the stops in stimuli with bursts were now identified with nearly perfect accuracy, and burstless [t*ɑ*] was now identified more accurately than burstless [k*ɑ*], as originally predicted, at least when preceded by [ʃ]. (However, see footnote 6.) Otherwise, the responses to burstless stimuli replicated the pattern found in condition A: The stops in [(ʃ)–] stimuli were identified more accurately than the stops in [(s)–] stimuli, and confusions for [(s)–] stimuli tended more towards a forward place of articulation than confusions for [(ʃ)–] stimuli, $F(1, 11) = 7.2$, $p < 0.05$, this effect being most pronounced for [k*ɑ*]. In addition, there was a clear effect of the preceding synthetic fricative noise: "t" responses were more frequent after [ʃ], while "k" responses were more frequent after [s], $F(1, 11) = 34.7$, $p < 0.001$. Thus the present experiment replicated both the coarticulatory effect (due to the excerpted fricative) and the corresponding perceptual effect (due to the substituted fricative) on stops in a single design. The marked token differences observed in condition A were also replicated; however, all statistical effects held up when token variance was taken as the error estimate.

TABLE IV. Stop identification in natural (F)CV portions preceded by synthetic [ʃ] or [s].

	Without burst			With burst		
	Response (%)					
Stimulus	"p"	"t"	"k"	"p"	"t"	"k"
[ʃ] + [(s)tɑ]	10.5	**83.8** [a]	5.6	0.3	**99.7**	···
[ʃ] + [(ʃ)tɑ]	3.0	**91.7**	5.3	···	**98.3**	1.7
[ʃ] + [(s)kɑ]	4.7	66.4	**28.9**	···	1.7	**98.3**
[ʃ] + [(ʃ)kɑ]	1.1	51.1	**47.8**	···	4.4	**95.6**
[s] + [(s)tɑ]	10.5	**74.4**	15.0	0.3	**99.1**	0.6
[s] + [(ʃ)tɑ]	5.0	**79.4**	15.6	···	**96.4**	3.6
[s] + [(s)kɑ]	3.0	31.7	**65.3**	···	···	**100.0**
[s] + [(ʃ)kɑ]	···	20.0	**80.0**	···	···	**100.0**

[a] Note: "Correct" responses are in bold type.

D. Condition C: Fricative identification in F + (F)CV stimuli

In this condition, we employed the "method of substitution" to see whether the coarticulatory traces of the preceding fricatives in the natural CV portions would bias the perception of ambiguous synthetic fricative noises in the direction of the original fricative. Thus this condition was analogous to experiment 1C, which showed that cues contained in natural fricative noises that had been excised from FCV utterances influenced stop perception when synthetic CV portions were added. There is an important difference, however: The cues to place of stop articulation in the fricative noise of an FCV utterance are quite pronounced and, as we showed in experiment 1B, generally sufficient to identify the stop from the fricative noise alone. On the other hand, any cues to place of fricative articulation contained in the CV portion are subtle and indirect; our informal observation is that they are not sufficient to identify a missing fricative. Therefore we expected that any influence of the CV portion on fricative perception would be rather small.

1. Method

The 24 CV portions were preceded by nine synthetic fricative noises forming an [ʃ]–[s] continuum (Table II), plus a 75-ms gap. The resulting 216 stimuli were presented four times in random order. The subjects' task was to identify the fricative as "sh" or "s" and the stop as "p," "t," or "k."

Since seven of the ten subjects in condition C were newly recruited, this part of experiment 2 also served as a semi-independent replication of the error patterns in stop identification observed in conditions A and B. In addition, we re-examined a question that received conflicting answers in our earlier studies (Mann and Repp, 1981): Whether, and in which way, stop identification is influenced by the precise spectral properties of the preceding (steady-state) fricative noise.

2. Results and discussion

The fricative identification results are shown in Fig. 3, separately for stimuli with and without bursts at CV onset. Although the differences between the various identification functions were relatively small, the statistical analysis (conducted on percent "sh" responses averaged over all members of the fricative noise continuum) revealed several reliable effects. First, "sh" responses were more frequent to burstless stimuli than to stimuli containing bursts, $F(1,9) = 12.5$, $p < 0.01$. Second, "sh" responses were more frequent to stimuli containing [t*a*] than to stimuli containing [k*a*], $F(1,9) = 8.8$, $p < 0.05$. Third, and most interestingly, "sh" responses were more frequent to stimuli containing [(ʃ)−] CV portions than to stimuli containing [(s)−] CV portions, $F(1,9) = 20.5$, $p < 0.01$; this was the effect of original fricative context we were looking for. However, there was also a triple interaction, $F(1,9) = 14.0$, $p < 0.01$. To clarify this interaction, separate analyses were conducted on stimuli with and without bursts.

The separate analyses revealed, for burstless stimuli, only an effect of original fricative context, [(ʃ)−] versus [(s)−], $F(1,9) = 5.7$, $p < 0.05$; however, stimuli with bursts showed not only the same effect in more pronounced form, $F(1,9) = 14.5$, $p < 0.01$, but also an effect of (intended) stop, [t*a*] versus [k*a*], $F(1,9) = 9.5$, $p < 0.05$, and an interaction between these two effects, $F(1,9) = 10.3$, $p < 0.02$. Figure 3 shows that the interaction derives from the effect of original fricative context being larger for [k*a*] than for [t*a*]. Analyses using token variance as the error term yielded the same pattern of results, with somewhat reduced levels of significance; the effect of original fricative context, which was of greatest interest to us, remained significant overall ($p < 0.01$), and separately for stimuli with bursts ($p < 0.05$).

These results show that acoustic variations at the onset of the CV portion, induced by the articulation of a preceding fricative noise, are sufficient to create a slight but significant bias towards perception of the original fricative category when an ambiguous noise cue is present. This bias was larger when the CV portion included a burst; thus the burst may convey part of the coarticulatory information. The finding that "sh" responses were somewhat more frequent with [t*a*] (and "s" responses with [k*a*]) replicates an effect of stop consonant identity on fricative perception which we had observed in one of our earlier studies (Mann and Repp, 1981: Experiment 5). The effect mirrors the now-familiar influence of the fricative on stop perception: In both cases, "s" tends to go with "k," and "sh" with "t." That the effect was reliably observed only in stimuli with bursts probably relates to the fact that only these stimuli permitted accurate identification of the intended stops. We have no explanation at present for our finding of an overall increase in "sh" responses in the absence of bursts.

As in conditions A and B, stop identification was much more accurate when bursts were present: [k*a*] was hardly ever misidentified (0.2% "t" responses), but the stop in [(ʃ)t*a*] was misidentified as "k" slightly more often (5.8%) than the stop in [(s)t*a*] (1.4%). Burstless stimuli, on the other hand, generated a large number of errors, including a small percentage (2.1) of "p" responses. The response pattern for burstless stimuli warrants some closer scrutiny; it is plotted as percent "k" responses in Fig. 4, with the synthetic noise continuum along the abscissa.

The figure shows that "k" responses were more frequent to [k*a*] than to [t*a*], $F(1,9) = 120.2$, $p < 0.001$, and that original fricative context had an effect with [k*a*]—"k" responses being more frequent to [(ʃ)k*a*] than to [(s)k*a*]—but not with [t*a*]. This was reflected in a significant interaction, $F(1,9) = 19.7$, $p < 0.01$, in addition to a significant main effect of original fricative context, $F(1,9) = 41.0$, $p < 0.001$. However, an effect of original fricative context on [t*a*] was reflected in "p" responses (not shown in Fig. 4), which were more frequent to [(s)t*a*] than to [(ʃ)t*a*]. This pattern of results replicates condition B.

Consider now the effect of the actual fricative noise

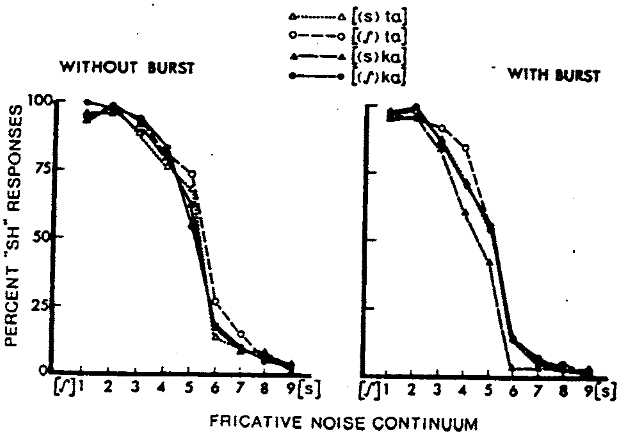

FIG. 3. Effects of CV portions from different fricative contexts on fricative perception (experiment 2C).

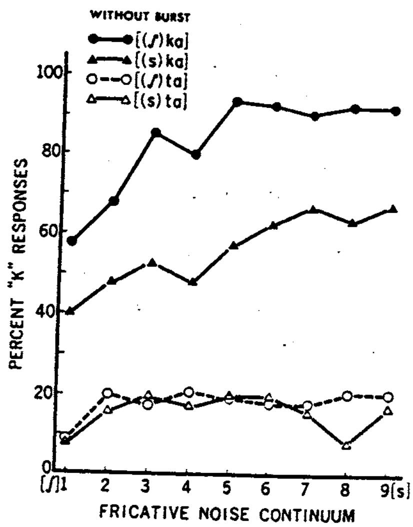

FIG. 4. Effects of (synthetic) fricative noise spectrum on stop identification in CV portions excerpted from different fricative contexts (experiment 2C).

on stop identification: The percentage of "k" responses increased significantly as the synthetic noises changed from [ʃ]-like to [s]-like, $F(8, 72) = 5.8$, $p < 0.001$. This is the familiar effect of fricative context on stop identification. For unknown reasons, this effect was essentially restricted to [ka], as reflected in a significant interaction, $F(8, 72) = 6.3$, $p < 0.001$. It is also evident that the increase for [ka] occurred almost exclusively on the left half of fricative noise continuum, viz., within the "sh" category. In this respect, the data replicate experiment 4 of Mann and Repp (1981), which had combined the same synthetic fricative noises with synthetic CV portions from a [ta]–[ka] continuum. However, the present data were not sufficient to determine with any degree of confidence whether, for the ambiguous fricative noises in the middle of the [ʃ]–[s] continuum, the perceived fricative *category* had any separate influence on stop perception (cf. Mann and Repp, 1981: Experiment 5). The present pattern of results admits that possibility; in any case, it supports our earlier conclusion (Mann and Repp, 1981) that spectral properties of the fricative noise contribute significantly to the effect of the fricative on stop perception.

III. SUMMARY AND CONCLUSIONS

The present series of experiments increases our understanding of the perceptual context effect discovered by Mann and Repp (1981)—the tendency to perceive velar stops following [s]. The effect itself must be considered firmly established, as it has been obtained consistently, not only in all-synthetic stimuli (Mann and Repp, 1981) but also in combinations of natural fricative noises with synthetic CV portions (Experiments 1A and 1C) and in combinations of synthetic fricative noises with natural CV portions (Experiments 2B and 2C).

Experiment 1C successfully ruled out the hypothesis that the context effect is due to supposedly neutral fricative noises acting as direct cues to stop place of articulation. While there are demonstrable perceptual effects of direct place cues in the fricative noise, these effects are independent of the influence of fricative identity on stop perception. Our results also ruled out the possibility that a simple bias to respond "k" in conjunction with "s" underlies the context effect (Experiment 1B).

In experiment 2, we obtained clear evidence that fricative articulation produces perceptually significant changes in the following CV portions. Thus we have established an empirical basis for the hypothesis that the perceptual context effect represents a form of compensation for coarticulatory shifts. It is true that our data reflect the articulation of only a single speaker; it remains to be seen whether fricative–stop coarticulation is a universal phenomenon. At the very least, however, our data show that such coarticulation *can* occur.

We are aware, of course, that the demonstration of coarticulatory interactions between fricative and stop production by no means proves that they are the *cause* of the corresponding perceptual effect. Indeed, the perceptual effect may represent a general tendency to differentiate successive phonetic segments on the place-of-articulation dimension—a tendency which would parallel the general assimilatory nature of coarticulation but may not be related to the specific coarticulatory interactions between the segments in question. Experiments to prove a specific connection between perception and production are difficult to design but perhaps not impossible, and we are presently giving this issue a good deal of thought.

Our studies leave open several additional questions about the nature of the context effect of interest. For example, there is the question of whether the effect of the fricative on the stop is a function of perceived fricative category or of fricative noise spectrum. Our earlier experiments (Mann and Repp, 1981) suggested that both factors are involved, and our present experiment 2C reaffirmed a strong role of fricative noise spectrum. To the extent that future studies will replicate an effect of perceived fricative category, two separate mechanisms may be needed to explain the perceptual context effect. Perhaps, both mechanisms serve to compensate for coarticulatory effects; but it is conceivably that only one of them does.

The perceptual context effect and the associated coarticulatory shifts demonstrated here are by no means isolated or exotic phenomena. Just as coarticulation between successive phonetic segments is probably even more common than the considerably available evidence suggests, perceptual context effects appear to be the rule rather than the exception. For example, stop perception is affected not only by preceding fricatives but also by liquids (Mann, in press) and other stops (Repp, 1978). There are not only proactive context effects in perception but also retroactive ones, such as the influence of following vowels on fricative perception (Mann and Repp, 1980). The parallel to the well-known bidirectionality of coarticulation is obvious. We believe that, as the evidence for perceptual and articulatory interdependencies between phonetic segments continues to increase, static and mechanistic approaches to the problem of speech perception—still in vogue, but beset with increasing difficulties—will have to make way for more dynamically oriented theories.

ACKNOWLEDGMENTS

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories, and by NICHD Postdoctoral Fellowship HD05677 to Virginia A. Mann. Portions of the results were reported at the 99th Meeting of the Acoustical Society of America in Atlanta, April 1980. We thank Alvin Liberman, Michael Studdert-Kennedy, and Winifred Strange for helpful comments.

APPENDIX

Here we report acoustic measurements of the natural-speech stimuli used in our experiments. All spectral measurements were made by visual inspection of successive spectral cross sections, provided by a Federal Scientific UA-6A spectrum analyzer and displayed as point plots on a Hewlett-Packard 1300A scope. All spectra were computed over 25.6-ms windows in 12.8-

ms steps; they were smoothed and pre-emphasized. Maximum resolution was 40 Hz. The precise position of the windows with respect to stimulus onset (or offset) could not be controlled; we simply took the first (last) cross section that yielded clear spectral peaks as the stimulus onset (offset). The effect of this uncertainty in temporal alignment on the measurements was considered negligible.

A. Fricative noises

There were 18 stimuli to be measured: Three tokens of [ʃ] noise and three tokens of [s] noise from each of three original contexts, [−a], [−ta], and [−ka]. We examined the last 10 sections (128 ms) of each stimulus, starting with the last and proceeding backwards. From each spectrum, we determined the location of the major energy peaks below 5 kHz, as well as the lower cutoff frequency—the point below which there was either no energy at all or only small, isolated peaks. (This latter measure may have been dependent on input amplitude and therefore should not be taken absolutely: However, it is highly relevant to a comparison of noises from different contexts.) Having determined these measures, we averaged them across the three tokens of each noise in each context, omitting all value that were spurious or inconsistent within or across tokens. A graphic representation of these average parameters for [ʃ] and [s] noises in [−(ta)] and [−(ka)] context is provided in Fig. 5.

Figure 5 represents spectral peaks as connected circles and the lower cutoff frequencies as simple lines. The figure shows that the major resonances (poles) were fairly steady state and not much influenced by the place of stop occlusion. Obviously, the parameter sensitive to stop occlusion was the lower cutoff frequency, particularly in the last 50 ms. In the context of [ta], the lower edge of the spectrum shifted rapidly upward, whereas, in the context of [ka], [s] showed a small downward shift, and [ʃ] showed a large downward shift followed by a small upward shift. At stimulus offset, the cutoff frequencies differed by 600–800 Hz between [−(ta)] and [−(ka)] stimuli. In addition, tokens of [s(ka)] showed scattered patches of energy below the cutoff frequency over the last 50 ms; if those peaks, one of which was as low as 300 Hz (not shown in Fig. 5), had been included in the cutoff frequency estimate, the dip in the cutoff function for [s(ka)] in Fig. 5 would, of course, have been much more dramatic. There is an indication in Fig. 5 that the earlier portion of the [ʃ] noise was also affected by context: In [ʃ(ka)], but not in [ʃ(ta)], there was initially an energy minimum between the two lower spectral peaks.

Tokens of [s(a)] and [ʃ(a)]—not shown in Fig. 5 for reasons of clarity—were highly similar in spectral structure to the other noises, except that they did not show any pronounced changes in lower cutoff frequency at offset. Their average cutoffs at offset were just about halfway between those for [−(ta)] and [−(ka)] stimuli.

Thus our data suggest that fricative noises preceding a stop closure are characterized by a rapid loss of low-frequency energy preceding [t] and by a relative increase in low-frequency energy preceding [k], these changes taking place within the last 50 ms or so. The major spectral peaks, on the other hand, do not seem to shift with place of stop occlusion, at least in the range below 5 kHz. Since our observations are based on a very small number of utterances of a single speaker, we should not draw any conclusions except that we have described the acoustic basis for the perceptual effects observed in experiments 1B and 1C. However, our data seem to agree with earlier informal reports in the literature (Malécot and Chermak, 1966; Uldall, 1964).

The durations of our fricative noises (averaged across tokens) were as follows: [s(a)], 211 ms; [ʃ(a)], 216 ms; [s(ta)], 208 ms; [s(ka)], 204 ms; [ʃ(ta)], 158 ms; [ʃ(ka)], 157 ms. Thus it appears that our speaker shortened his [ʃ] noises considerably more than his [s] noises when a stop consonant followed.

B. CV portions

For each of the 12 CV portions (3 tokens each of [(s)ta], [(ʃ)ta], [(s)ka], and [(ʃ)ka]), we traced the major spectral peaks (formants) through the first 10 spectral sections that yielded a clear formant pattern.

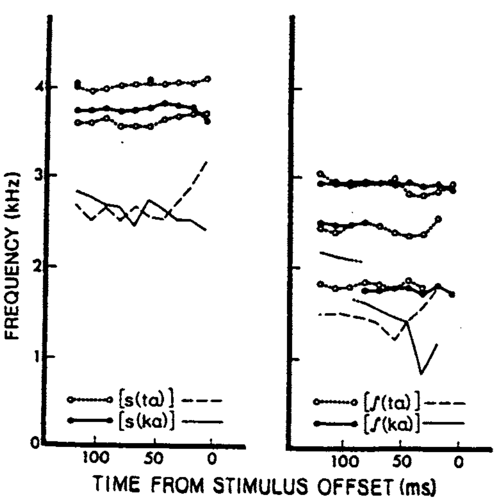

FIG. 5. Average spectral structure of fricative noises in different stop contexts (experiment 1).

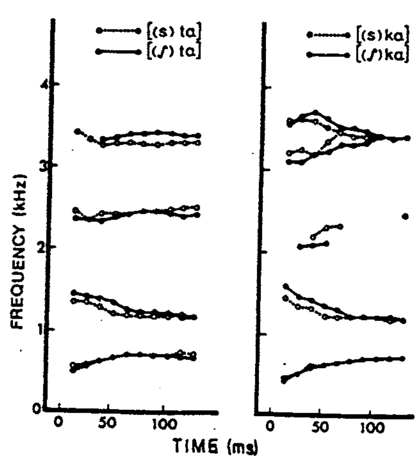

FIG. 6. Average spectral structure of CV portions in different fricative contexts (experiment 2).

Thus we did not include the release burst whose spectrum was too irregular (especially in [ta]) to permit useful comparisons, given the limited amount of data. The formant trajectories, averaged across tokens, are displayed in Fig. 6.

It can be seen that, although there had been clear perceptual differences between (burstless) CV stimuli from different fricative contexts, the acoustic effects of the preceding fricative were rather small: The second formant had a somewhat higher frequency (by up to 100 Hz) following [ʃ] than following [s], and this difference seemed to persist throughout the transitional phase (about 50 ms). There are indications of a higher onset of F3 in [(s)ka] than in [(ʃ)ka], but this formant was weak and often altogether absent in [ka]. The differences observed, though small, are in agreement with a forward shift in place of stop occlusion following [s], since a forward shift implies a greater separation of F2 and F3 onsets (cf. the greater separation for [ta] than for [ka]). The "split F4" for [ka] appears to be an idiosyncratic feature of the speaker who produced these utterances.

We have now completed the acoustic analysis of a much larger corpus of utterances from several speakers and have found consistent evidence for coarticulatory shifts in CV formant transitions following [s] versus [ʃ]. We hope to report these results in a future publication.

The durations of our CV stimuli ranged from 440 to 540 ms, although the major energy was contained within the first 300 ms or so. The durations of the burst-cum-aspiration portions—which were removed to obtain the burstless versions—varied from 18 to 35 ms. On the average, they were slightly longer for [ka] (25.2 ms) than for [ta] (21.5 ms); there was little difference between fricative contexts.

We did not measure stimulus amplitudes since an earlier study of ours (Mann and Repp, 1981: Experiment 4) suggested that the relative amplitude levels of the noise and CV portions have little influence on perception. Suffice it to say that, when substituting synthetic for natural stimulus portions, we tried to maintain approximately the original amplitude relationships.

[1]Perceptual coherence between synthetic and natural signal portions is not difficult to achieve, especially when—as in the present studies—they have different sources of excitation (aperiodic fricative noise versus largely periodic CV portion) and, moreover, are separated by a silent closure interval. However, we have also been successful in combining natural and synthetic voiced portions, separated by silence (Mann, in press), and synthetic noises with natural voiced portions, immediately adjoined (Mann and Repp, 1980).

[2]Errors in fricative identification were virtually nonexistent, except for a single subject (14%) whose exclusion would not have changed the results.

[3]Only two subjects made any errors in fricative identification (2% and 7%, respectively).

[4]As in condition A, only one subject committed a large number of fricative identification errors (22%); nevertheless, he showed the pattern of Fig. 2. Exclusion of his data would not have changed the results.

[5]There was a fourth condition whose results essentially duplicated those obtained in the third. This condition was presented between conditions B and A. For details, see Repp and Mann (in press).

[6]The result was unexpected because velar stops were thought to be more susceptible to articulatory shifts than alveolar stops. However, this hypothesis is difficult to test perceptually, because the probability of confusions along the place dimension depends on the perceptual distances between the few alternative categories available. Most likely, "th" is closer to "d" than "d" is to "g." Therefore, a small forward shift in the articulation of [tɑ] will result in a large number of "th" responses, whereas a larger forward shift in the articulation of [kɑ] might result in only a moderate number of "d" responses. As will be seen, omission of the "th" category in condition B led to the "expected" better identifiability of alveolar stops.

Ali, L., Gallagher, T., Goldstein, J., and Daniloff, R. (1971). "Perception of coarticulated nasality," J. Acoust. Soc. Am. 49, 538–540.

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

Benguerel, A. P., and Adelman, S. (1975). "Coarticulation of lip rounding and its perception," in *Structure and Progress in Speech Perception*, edited by A. Cohen and S. G. Nooteboom (Springer-Verlag, New York).

Dorman, M. F., Raphael, L. J., and Liberman, A. M. (1979). "Some experiments on the sound of silence in phonetic perception," J. Acoust. Soc. Am. 65, 1518–1532.

Fant, G., Liljencrants, J., Malac, V., and Borovickova, B. (1970). "Perceptual evaluation of coarticulation effects," Quarterly Progress Report on Speech Research (Speech Transmission Laboratory, Royal Institute of Technology, Stockholm) 1, 10–13.

Lehiste, I., and Shockey, L. (1972). "On the perception of coarticulation effects in English VCV syllables," Working Papers in Linguistics (Department of Linguistics, Ohio State University) 12, 78-86.

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. (in press). "Influence of preceding liquid on stop perception," Percept. Psychophys.

Mann, V. A., and Repp, B. H. (1981). "Influence of preceding fricative on stop consonant perception," J. Acoust. Soc. Am. 69, 548–558.

Mann, V. A., and Repp, B. H. (1980). "Influence of vocalic context on perception of the [ʃ]–[s] distinction," Percept. Psychophys. 28, 213–228.

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

Repp, B. H., and Mann, V. A. (in press). "Perceptual assessment of fricative–stop coarticulation," Haskins Laboratories Status Report on Speech Research.

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.

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

Yeni-Komshian, G., and Soli, S. (1979). "Extraction of vowel information from fricative spectra," in *Speech Communication Papers Presented at the 97th Meeting of the Acoustical Society of America*, edited by J. J. Wolf and D. H. Klatt (Acoustical Society of America, New York).