Perceptual order and the effect of vocalic context on fricative perception

VIRGINIA MANN

University of California, Irvine, California, and Haskins Laboratories, New Haven, Connecticut

and

SIGFRIED D. SOLI

*The House Ear Institute, Los Angeles, California
and Haskins Laboratories, New Haven, Connecticut*

This three-part study demonstrates that perceptual order can influence the integration of acoustic speech cues. In Experiment 1, the subjects labeled the [s] and [ʃ] in natural FV and VF syllables in which the frication was replaced with synthetic stimuli. Responses to these "hybrid" stimuli were influenced by cues in the vocalic segment as well as by the synthetic frication. However, the influence of the preceding vocalic cues was considerably weaker than was that of the following vocalic cues. Experiment 2 examined the acoustic bases for this asymmetry and consisted of analyses revealing that FV and VF syllables are similar in terms of the acoustic structures thought to underlie the vocalic context effects. Experiment 3 examined the perceptual bases for the asymmetry. A subset of the hybrid FV and VF stimuli were presented in *reverse*, such that the acoustic and perceptual bases for the asymmetry were pitted against each other in the listening task. The perceptual bases (i.e., the perceived order of the frication and vocalic cues) proved to be the determining factor. Current auditory processing models, such as backward recognition masking, preperceptual auditory storage, or models based on linguistic factors, do not adequately account for the observed asymmetries.

Studies of speech perception have revealed that the acoustic cues that support perception of individual phonetic segments are overlapped and interwoven. Even in fricative-vowel (FV) syllables such as *sa*, *sha*, *su*, and *shu*, where the vocalic portion is associated with the vowel and the frication portion with the fricative consonant, the vocalic portion contains perceptual information for the fricative as well as for the vowel (e.g., Mann & Repp, 1980). Likewise, the frication portion may contain information for the vowel (e.g., Yeni-Komshian & Soli, 1981). This spread of segmental information is attributed to the co-articulation of articulatory gestures for adjacent segments and has led some researchers to suggest that speech perception proceeds as if by reference to the way in which speech is produced (see Liberman & Mattingly, 1985; Mann & Repp, 1980).

Previous studies of segmental perception in FV utterances have illustrated that listeners readily make use of acoustic information spread across the entire utterance.

These studies provoke questions about the spread of segmental information in other combinations of segments and about how listeners integrate and weight this information in their perceptual judgments. The present research attempts to address these questions by comparing the effect of vocalic context on the perception of fricatives in FV and vowel-fricative (VF) stimuli. This comparison allowed us to determine how the order of vowel and fricative segments in an utterance influence the integration and weighting of information spread over the utterance.

FV and VF syllables have several advantages for the proposed comparisons. The vocalic (periodic) and the frication (aperiodic) portions of a syllable can easily be separated, and either portion can be presented in isolation as a perceptual stimulus. Their relatively long durations and high amplitudes also permit acoustical analyses that can help to identify perceptual cues in each portion of the utterance. Previous perceptual research has shown that the vocalic portion of an FV utterance contains information about the place of articulation of the syllable-initial fricative (Harris, 1955; Mann & Repp, 1980; Whalen, 1981). Research with VF utterances has shown that the vocalic portion contains information about the voicing of the syllable-final fricative (Denes, 1955; Derr & Massaro, 1980; Soli, 1982). Several investigators have also shown that perceptual information for some vowels (viz., the high vowels [i] and [u]) is contained in the aperiodic noise spectra of both syllable-initial and syllable-final fricatives (Soli, 1981; Whalen, 1983; Yeni-Komshian & Soli, 1981).

This research was supported in part by NICHD Grant HD01994 to Haskins Laboratories, by a grant to the second author from the Computer Science Center, University of Maryland, and by funding from the University of California, Irvine. Early portions of this research were reported at the 103rd and 105th meetings of the Acoustical Society of America in Chicago, May 1982, and in Cincinnati, May 1983. We are grateful for the valuable comments on an earlier version of this paper by K. S. Harris and A. M. Liberman. Correspondence should be addressed to Virginia Mann, Department of Cognitive Sciences, School of Social Science, University of California, Irvine, CA 92717.

There have been few acoustic and perceptual comparisons of the spread of segmental information in FV and VF utterances. Kunisaki and Fujisaki (1977) developed perhaps the most sensitive method for examining how vocalic information and frication information contribute to fricative perception. They began with a synthetic continuum of fricative noises varying in spectral peak frequencies, so as to range from [ʃ] to [s] when presented in isolation. These fricative noises were appended to synthesized vocalic segments to produce VFV syllables, and comparisons were made between the fricative labeling functions obtained in various vocalic contexts. Kunisaki and Fujisaki observed that, in the context of a *following* vowel, [u] or [o], the fricative labeling boundary shifted to produce more [s] responses; however, a *preceding* vowel had no effect on the boundary.

Mann and Repp (1980) replicated Kunisaki and Fujisaki's (1977) findings with FV and VF syllables, while extending their technique through the use of naturally produced vocalic segments rather than synthesized segments. From subjects' responses to these hybrid stimuli, Mann and Repp distinguished three vocalic context effects on fricative identification. First, in the [u] vocalic context, the fricative labeling boundary shifted to produce more [s] responses. Second, more [s] responses were also given in the context of vocalic segments that originally contained [s] (i.e., [sa] and [su]). Third, more [s] responses were also given in the context of vocalic segments produced by a male speaker, relative to responses to a female speaker. They concluded that vocalic segments carry perceptual information about vowel quality (i.e., lip-rounding), consonant place of articulation, and speaker characteristics, all of which influence fricative perception.

In the present study, we have further examined the influence of vocalic cues for vowel quality and consonant place of articulation on fricative perception. We use Mann and Repp's (1980) term *vocalic context effects* to characterize both types of cues collectively. Our emphasis is on the influence of segmental order on vocalic context effects. Hybrid stimuli composed of synthetic fricative segments and natural vocalic segments were used to ensure that the vocalic cues available in natural speech would be present in our stimuli. The first experiment documents the asymmetric vocalic context effects in FV and VF utterances; the second experiment examines the acoustic bases for the perceptual asymmetries; the third experiment separates the effects of acoustic structure and order of presentation.

EXPERIMENT 1

Experiment 1 was designed to determine whether the *vowel-quality* and *place-of-articulation* effects would extend to VF stimuli and whether they would be equivalent to the effects that occur in FV stimuli. The methodology was drawn from Mann and Repp's (1980) study in which listeners had labeled the fricatives in hybrid FV stimuli constructed by removing the frication from naturally pro-

duced FV syllables, replacing it with synthetic fricative noises from along an [s]-[ʃ] continuum. Our innovation was to replace the noises in VF syllables, as well as in FV ones.

Method

Stimuli. The perceptual stimuli consisted of hybrid FV and VF syllables with synthetic fricatives and vocalic portions from natural utterances. Citation form utterances [sa, fa, su, fu, as, af, us, uf] were spoken by 2 male talkers, 1 a trained phonetician (A.S.A.) and 1 one of the authors (S.D.S.). Two male talkers were used because our intent was to probe to the generality of the results and not to replicate the male versus female speaker effects noted by Mann and Repp (1980). Ten repetitions of each utterance were recorded in random order, and, although the full set was used in the acoustic analysis, the perceptual study was limited to the first three tokens.¹ After sampling the tokens at 10 kHz, the Haskins 224 waveform editor was used to excise the vocalic portions from each of these tokens, beginning with the onset of the first pitch period in FV utterances. The vocalic portions of VF utterances were excised ending with the offset of the last pitch period. We refer to the vocalic portions of the FV syllables as [(s)a], [(f)a], and so forth, and to those of the VF syllables as [a(s)], [a(f)], and so forth, thus identifying the vowel and noting the fricative place of articulation.

A continuum of nine fricatives ranging from [ʃ] to [s] was synthesized by varying the center frequencies of the two poles in the fricative circuit of the Haskins OVE IIIc synthesizer. The first member of the noise continuum (Stimulus 1) had spectral peaks at 2.4 and 2.8 kHz, as determined by LPC analyses (Markel & Gray, 1976); the ninth member (Stimulus 9) had peaks at approximately 3.9 and 4.8 kHz. The lower frequency peak increased by approximately 100 Hz per stimulus in Stimuli 1-4, and by 200 Hz per stimulus in Stimuli 5-9. The higher frequency peak increased by approximately 250 Hz per stimulus over the entire series. The rate of increase in peak frequencies was chosen to approximate equal-sounding stimulus intervals. Each synthetic fricative was 200 msec in duration with an amplitude contour that increased linearly over the first 60 msec, remained flat over the next 80 msec, and decreased linearly over the last 60 msec. A symmetrical contour was chosen so that the noises would sound equally appropriate as syllable-initial and as syllable-final fricatives. Nine hybrid syllables were formed from each natural vocalic segment by adjoining to it each member of the synthetic fricative continuum. A 9-msec gap separated the noise and vocalic portions in the hybrid FV syllables, and a 12-msec gap was used in the hybrid VF syllables. These gap durations were based on the mean duration of silence between the fricative and voicing in the natural utterances used in the perceptual study (i.e., the three tokens of each syllable). Three stimulus tapes were prepared, all recorded at 7½ in. per second. The first tape presented each of the nine isolated synthetic fricatives 10 times in a randomized order, with a 2.5-sec interstimulus interval and a 7.5-sec interval between blocks of nine stimuli. The second and third tapes each consisted of the hybrid FV and VF syllables. Each tape contained the stimuli for a single speaker, and over the course of the tape, each of the 216 hybrid stimulus (i.e., 9 fricatives × 8 syllables × 3 tokens) occurred eight times. There was a 2.5-sec interstimulus interval and a 7.5-sec interval between blocks of 108 stimuli.

Subjects. The subjects were 10 undergraduate students at the University of Maryland who volunteered to participate in the experiment for course credit. All were native speakers of English and reported no history of speech or hearing disorders.

Procedure. The stimulus tapes were presented binaurally using a Revox B77 tape recorder, TDH-39 headphones and a sound-attenuated testing chamber. The tape containing the identification trials for the isolated fricatives was presented first to all subjects.

The two tapes containing the hybrid syllables for each talker were then presented in counterbalanced order across subjects. In all cases, the subjects were instructed to determine whether each stimulus contained "s" or "sh" and to write their answers on a response sheet.

Results

The mean percentage of "sh" responses is displayed in Figure 1 as a function of the stimulus number of the synthetic fricatives. Responses to the FV stimuli appear in panel a, and responses to the VF stimuli appear in panel b. Each panel offers the labeling function for the isolated fricatives and for each of the four vocalic contexts. Responses have been combined across the 10 subjects, across the three tokens of each vocalic segment, and across the 2 speakers. Thus, for the isolated fricatives, each point summarizes 100 responses; for the hybrid stimuli, each point summarizes 480 responses.

The labeling functions in Figure 1a reveal the influence of the FV vocalic segments on the identification of initial fricatives. Note first that the functions for all of the hybrid stimuli are shifted to either the left or the right of the function for the isolated fricatives. These shifts indicate that all of the vocalic segments from FV utterances contained perceptual information that interacted with the cues in the spectrum of the fricative. The effects of cues associated with the place of articulation contrast may be seen by comparing the functions for the [(s)a] and [(f)a] contexts and for the [(s)u] and [(f)u] contexts. The functions for both [(f)a] and [(f)u] are shifted to the right, relative to [(s)a] and [(s)u], indicating that more "sh" responses were given to the stimuli originally produced with an initial [f] than to the stimuli produced with an initial [s]. The effects of place cues were larger for [a].

The effects of cues associated with differences in vowel quality may be seen by comparing the labeling functions for [(s)a] and [(s)u] and for [(f)a] and [(f)u]. The labeling responses to the FV and VF stimuli are also summarized in Table 1 in terms of the differences in mean percentage of "sh" responses due to the vowel-quality and place contrasts. The function for [(s)a] represents more "sh" responses than does the function for [(s)u], while the function for [(f)a] represents many more "sh" responses than does the function for [(f)u]. Thus, fewer "sh" responses (and more "s" responses) were given in the context of the vowel [u]. Recall that anticipatory lip-rounding before [u] lowers the frequency of the major spectral peaks in the fricative (Kunisaki & Fujisaki, 1977; Soli, 1982; Yeni-Komshian & Soli, 1981). When frequency is lowered, the labeling boundary also shifts to a lower frequency, narrowing the "sh" category and broadening the "s" category. Thus, the FV results replicate both the place-of-articulation and the vowel-quality (i.e., lip-rounding) effects reported previously (Kunisaki & Fujisaki, 1977; Mann & Repp, 1980; Whalen, 1981, 1983).

The labeling functions for the hybrid VF syllables are displayed in Figure 1b. These functions are more closely clustered around the function for the isolated fricatives than are those in Figure 1a, indicating that identification

Figure 1. Group labeling functions for FV stimuli (panel a) and VF stimuli (panel b) plotted as a function of synthetic frication noises (Stimuli 1-9). The legend identifies the vowel quality and original fricative place of articulation associated with each of the natural vocalic segments that defined the vocalic context in the hybrid stimuli.

of final fricatives was influenced less by place of articulation and vowel quality in the vocalic segments. Table 1 compares the differences between the mean number of "sh" responses in the various FV and VF stimuli. The differences in overall "sh" responses between the [a(s)] and [a(f)] functions and between the [u(s)] and [u(f)] functions were very small. The perceptual effects of place of articulation cues in the VF vocalic segments were only about one fourth the magnitude of the comparable effects in the FV segments. In addition, the effects of vowel-quality cues in the VF segments, while somewhat larger, were also reduced to approximately two thirds the magnitude of the comparable effects for the FV segments.

Table 1
Effects of Vowel Quality and Place of Articulation: Differences in Mean Percentages of "sh" Responses to FV and VF Stimuli

Vocalic Cue	Difference in "sh" Responses
FV Stimuli	
Vowel quality	
[(s)a]–[(s)u]	9.0
[(f)a]–[(f)u]	27.8
Place of articulation	
[(s)a]–[(s)a]	29.7
[(f)u]–[(s)u]	10.6
VF Stimuli	
Vowel quality	
[a(s)]–[u(s)]	11.3
[a(f)]–[u(f)]	13.8
Place of articulation	
[a(f)]–[a(s)]	7.1
[u(f)]–[u(s)]	3.4

The effects of vocalic cues in the FV and VF syllables and their consistency across talkers were evaluated by statistical analyses. The dependent variable for each subject was the total number of "sh" responses summed across the continuum of fricatives (Stimuli 1–9) separately for each FV context, each VF context, and each talker. These data were submitted to analysis of variance with talker (A.S.A. vs. S.D.S.) treated as a grouping factor and place of articulation (alveolar [s] vs. palatal [ʃ]), vowel quality (unrounded [a] vs. rounded [u]), and syllable position (FV vs. VF) as repeated measures. The overall effects of place of articulation and vowel quality on fricative labeling were statistically significant [$F(1,18) = 206.60$, $MS_e = 2,464.9$, $p < .0001$, and $F(1,18) = 76.87$, $MS_e = 2,117.0$, $p < .0001$, respectively]. However, both effects exhibited strong interactions with syllable position [for place of articulation, $F(1,18) = 51.61$, $MS_e = 837.2$, $p < .0001$; for vowel quality, $F(1,18) = 36.87$, $MS_e = 1,102.5$, $p < .0001$], due to the much larger magnitude of the effects in the FV stimuli.

The place-of-articulation and vowel-quality effects also exhibited a three-way interaction with syllable position, indicating that the difference between initial and final effects of place-of-articulation cues was significantly larger than the difference between the initial and final effects of vowel rounding cues [$F(1,18) = 14.27$, $MS_e = 65.0$, $p < .001$]. There was no main effect of talker ($F < 1.00$), and although the effect of vowel quality was stronger for talker A.S.A. [$F(1,18) = 8.90$, $MS_e = 245.0$, $p < .008$], as was the effect of place of articulation [$F(1,18) = 4.83$, $MS_e = 57.6$, $p < .04$], there were no three-way interactions between talker, vowel quality, and order, or between talker, place of articulation, and order. Thus, while we find some talker differences in the potency of vowel-quality and place-of-articulation cues, there is no indication that the contrast between preceding versus following cues is talker-dependent.

Discussion

As seen in Figure 1 and Table 1, both preceding and following vocalic segments have some effect on listeners' perception of a fricative as "s" or "sh." However, the effect of the following segments tends to be considerably stronger than the effect of the preceding segments. Significant effects of vowel quality (i.e., the vowel [a] vs. the lip-rounded vowel [u]) occur for both FV and VF, but these effects were approximately 50% greater for the FV stimuli. Likewise, significant effects of fricative place of articulation (i.e., whether the original utterance contained [s] vs. [ʃ]) occur for both FV and VF, but these effects were almost four times as great for the FV stimuli.

There are two logically distinct explanations of the present results. One is that the *acoustic structures of the vocalic segments*, and thus the underlying articulatory processes themselves, are somehow asymmetrical. Another is that the *perceptual or memory processes* of the listener somehow favor later occurring information, making listeners biased towards vocalic information in the FV stimuli, relative to that in the VF ones. If we look to asymmetries in the acoustic structure of FV and VF utterances, then at least two hypotheses need to be considered. The VF stimuli may differ from FV ones in having been produced with *less coarticulation*: less anticipation of the fricative constriction and less perseveratory lip-rounding for [u] (cf. Soli, 1981; Yeni-Komshian & Soli, 1981). Alternatively, VF coarticulation may be as extensive as FV coarticulation, but there may be *less evident acoustic consequences*. For example, VF coarticulation may occur as the amplitude of voicing is dropping near the end of the vocalic portion of the VF syllable, in which case the salience of the coarticulatory cues for the final fricative would be reduced. To evaluate these hypotheses, Experiment 2 offers an acoustical analysis of the natural speech syllables used in constructing the stimuli of Experiment 1.

If, instead, we look to the listening process rather than to the acoustic structure of the stimuli, the greater effect of the following vocalic segments would have to involve the *heard order* of the fricative noise and vocalic segment such that the later occurring vocalic segment plays a greater role. The literature on speech perception offers some indication that the heard order of acoustic cues can influence their interpretation (see Abbs, 1971; Sharf & Beiter, 1974; Sharf & Hemeyer, 1972); the final transitions in VC syllables seem to be more potent cues to consonant place of articulation than do the initial ones in CV syllables. To evaluate the possibility that heard order is a factor in the present asymmetry, in Experiment 3, we repeated Experiment 1 but played the stimuli *backward* so that the acoustic structure of each stimulus was heard in reverse. If heard order is an important factor, then following vocalic segments should have a greater effect than should preceding ones, whether we compare the FV and VF stimuli used in Experiment 1 or the FV_{reverse} and VF_{reverse} stimuli that are the result of playing the test tape backward.

EXPERIMENT 2

Experiment 2 examined the acoustic structure of the entire corpus of natural FV and VF utterances from which Experiment 1 had been drawn to determine whether either of two hypotheses about the perceptual information in the vocalic segment could account for the weaker perceptual effects in the VF stimuli. The first hypothesis proposes that the vowel and consonant segments in VF syllables are produced with less perseveratory lip-rounding during the fricative, relative to the magnitude of anticipatory lip-rounding during the fricative of a FV syllable; it is evaluated primarily through an analysis of fricative noise spectra. The second hypothesis proposes that anticipatory and perseveratory lip-rounding are more or less equivalent but that, in VF utterances, the timing of the perseveratory rounding gesture may be more variable or may coincide with a drop in the amplitude of voicing; it is evaluated by an analysis of the vocalic segments.

A number of investigators have observed that anticipatory lip-rounding before a rounded vowel lowers the frequencies of the major spectral peaks in the fricative (Bondarko, 1969; Heinz & Stevens, 1961; Mann & Repp, 1980). Rounding lowers the frequencies of the spectral peaks in the fricative and produces a dip in the spectrum, usually between 4 and 5 kHz (cf. Soli, 1981; Yeni-Komshian & Soli, 1981). This effect of rounding corresponds to the observed shift in the perceptual labeling boundary shifts toward a lower frequency in the context of the rounded vowel [u]. Thus, it has been argued that listeners draw upon lip-rounding information in the vocalic portion of the stimulus, as well as upon the spectral information in the fricative noise. The place effects have been attributed to the vocalic formant transitions that occur as the fricative constriction is released and the more open vowel configuration is assumed. These formant transitions following the release of a consonant have long been known to serve as a relevant source of perceptual information for stop and fricative place of articulation (e.g., Dorman, Studdert-Kennedy, & Raphael, 1977; Harris, 1955; Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967).

Method

The material for Experiment 2 was the entire set of 10 tokens of each utterance for each talker, including the three utterances from which the perceptual stimuli had been drawn and the remaining acceptable utterances.² Recall that this set was comprised of the following utterances: [sa, ja, su, ju, as, af, us, uf]. The number of tokens per utterance never dropped below six for any of these utterances.

Acoustic analysis techniques. The waveforms for the selected tokens were lowpass filtered at 5 kHz and sampled at 10 kHz into a laboratory computer. The digitized waveforms were first-differenced, multiplied by a sliding 25.6 Hamming window, and submitted to spectral analyses. The 128-point LPC spectra (Markel & Gray, 1976) were computed at a 5-msec frame rate using a 14-term model. The peaks from the spectra were picked and tracked to estimate formant frequencies and amplitudes for each frame. Computer-generated running spectra and formant frequency and amplitude plots were examined visually to detect gross errors in the

formant estimates. These errors were relatively infrequent and were limited to the low-amplitude regions of the utterances between the friction and voicing. Obvious tracking errors were corrected by hand editing.

The running spectra were aligned, for FV utterances, at voice onset and, for VF utterances, at voice offset, as determined from waveform measurements. In FV utterances, mean fricative spectra for all tokens of each utterance for each talker were computed for the frame 100 msec before vowel onset ($\Delta t = 100$ msec) and for the frame immediately before vowel onset ($\Delta t = 0$ msec). In VF utterances, mean spectra were computed for the frames 100 msec after vowel offset ($\Delta t = 100$ msec) and immediately after vowel offset ($\Delta t = 0$ msec). These mean spectra from the rounded and unrounded vowel contexts for FV utterances were then contrasted with the comparable spectra from the VF utterances.

The formant tracks were aligned in the same manner as the running spectra, and mean formant frequencies and amplitudes were computed for the first, second, and third formants (F1, F2, and F3) in each frame. A 99% confidence interval was also computed for each mean frequency value using the sampling distribution of the chi-square statistic. The number of values used to compute the mean for each formant within each frame could vary due to tracking errors that dropped values from the frame. The number of frames per token, a value dependent on the token's duration, was also variable. Thus, the number of observations for each mean formant frequency and the degrees of freedom for the estimated confidence interval were determined separately for each formant within each frame.

Results and Discussion

Effects of lip-rounding. The results of the spectral analyses for the natural [s] fricatives are shown in Figure 2 and for the [ʃ] fricatives in Figure 3. Focusing first on the results for [s], each panel in Figure 2 displays the mean spectrum in the context of unrounded [a] and rounded [u].

The [s] spectra for Talker A.S.A. clearly exhibit the effects of both anticipatory and perseveratory lip-rounding during the fricative. The major spectral peak near 3.8 kHz in the unrounded context is lowered to around 3.2 kHz in the rounded context for both the FV and VF utterances. Rounding also produced a consistent dip in the mean spectrum at about 4 kHz. These effects of rounding are clearly evident at $\Delta t = 0$ and 100 msec. The effects of anticipatory and perseveratory rounding for Talker S.D.S., although slightly different than for Talker A.S.A., are also evident in Figure 2. For Talker S.D.S., rounding produced a decrease in the fricative peak frequency from approximately 4.2 to 3.5 kHz and less of a dip in the spectrum. (Previous spectral analyses [Soli, 1981] employing a 10-kHz bandwidth revealed that the dip in the spectrum for this talker's fricatives is centered at approximately 5 kHz and thus would not be fully evident with the 5-kHz bandwidth used in the present analyses.) For both talkers, both the magnitude and the timing of anticipatory and perseveratory lip-rounding effects on [s] are generally similar and symmetrical in FV and VF utterances.

The results of the analyses for the [ʃ] spectra seen in Figure 3 lead to similar conclusions. For Talker A.S.A., the major spectral peak in the [a] context is about 3.2 kHz and drops to about 2.3 kHz in the [u] context. (Rounding also produced a dip in the spectrum at 4 kHz, which was

Figure 2. Mean LPC spectra for [s] friction noises from FV and VF utterances in unrounded [a] and rounded [u] vowel contexts for Talkers A.S.A. and S.D.S. at $\Delta t = 0$ msec and $\Delta t = 100$ msec, respectively. Each panel contrasts spectra from unrounded and rounded vowel contexts for a single talker at a fixed Δt . The spectra from the rounded vowel context are marked with diagonal hatching. The top row (panels a–d) displays these contrasts for FV utterances; the bottom row (panels e–h) displays these contrasts for VF utterances. Talker A.S.A.'s spectra are in the two columns on the left; Talker S.D.S.'s spectra are in the two columns on the right. The first column for each talker contrasts the unrounded and rounded contexts in FV and VF utterances at $\Delta t = 0$ msec; the second column displays these contrasts at $\Delta t = 100$ msec.

Figure 3. Mean LPC spectra for [f] friction noises from FV and VF utterances at $\Delta t = 0$ msec and $\Delta t = 100$ msec. Each panel contrasts spectra from unrounded and rounded vowel contexts for a single talker at a fixed Δt . The spectra from the rounded vowel context are marked with diagonal hatching. The top row (panels a–d) displays these contrasts for FV utterances; the bottom row (panels e–h) displays these contrasts for VF utterances. Talker A.S.A.'s spectra are in the two columns on the left; Talker S.D.S.'s spectra are in the two columns on the right. The first column for each talker contrasts the unrounded and rounded contexts in FV and VF utterances at $\Delta t = 0$ msec; the second column displays these contrasts at $\Delta t = 100$ msec.

less clearly defined than was the dip in the [s] spectra.) For Talker S.D.S., lip-rounding reshaped the spectrum by shifting the highest amplitude components from the frequency region above 2.5 kHz to the region below 2.5 kHz in the [u] context. Although the effects of anticipatory and perseveratory rounding were somewhat different for the two talkers, for each talker, the magnitude and timing of these effects (as estimated from the comparisons at $\Delta t = 0$ and 100 msec) were much the same in the FV and VF utterances.

Our primary concern is with the order of fricative and vowel and its effect on acoustic structure, and we note that the symmetry between FV and VF utterances is remarkably similar for the two talkers, implying that it generalizes across talkers. For both [s] and [ʃ], there is consistent evidence that vowel lip-rounding is no less coarticulated in VF utterances than in FV utterances. The

timing of the rounding gesture during the frication appears fairly symmetrical, at least up to 100 msec from the vowel.

The symmetrical effects of lip-rounding seen in the frication analyses indicate that the perceptual asymmetries observed in Experiment 1 are not likely to be the product of coarticulatory asymmetries in the FV and VF utterances. We may now turn to the second issue, namely, whether the vocalic segments differ in their acoustic salience. The formant statistics for the vocalic segments are presented as spectrogram-like displays in Figures 4 and 5. Figure 4 compares the FV and VF utterances containing [u], and Figure 5 compares those utterances containing [a]. The means for the most relevant temporal and spectral characteristics of the second and third formant trajectories are also summarized in Table 2. The four panels in the left halves of Figures 4 and 5 display the

Figure 4. Formant statistics for 100 msec of the fricative and the entire vocalic portion of FV and VF utterances spoken with the rounded vowel [u]. The four panels on the left display results for Talker A.S.A.; the four panels on the right display results for Talker S.D.S.

Figure 5. Formant statistics for 100 msec of the fricative and the entire vocalic portion of FV and VF utterances spoken with the unrounded vowel [a]. The four panels on the left display results for Talker A.S.A.; the four panels on the right display results for Talker S.D.S.

Table 2
Mean Formant Transition and Steady-State Characteristics for
F2 and F3 in the FV and VF Vocalic Segments

Utterance	Talker	F2				F3
		Onset Frequency	Offset Frequency	Duration	Slope	Onset Frequency
FV Utterances						
[(s)a]	A.S.A.	1443	1257	79	-2.36	2588
	S.D.S.	1482	1296	99	-1.88	2754
[(f)a]	A.S.A.	1629	1272	74	-4.81	2255
	S.D.S.	1786	1296	408	-4.52	2422
[(s)u]	A.S.A.	1506	729	225	-3.44	2363
	S.D.S.	1600	886	192	-3.72	2686
[(f)u]	A.S.A.	1595	768	239	-3.46	1932
	S.D.S.	1791	886	228	-3.97	2373
VF Utterances						
[a(s)]	A.S.A.	1441	1248	109	1.77	2637
	S.D.S.	1497	1292	130	1.58	2789
[a(f)]	A.S.A.	1668	1248	143	2.94	2329
	S.D.S.	1727	1238	200	2.44	2524
u(s)]	A.S.A.	1586	778	76	6.71	2412
	S.D.S.	1556	915	135	4.74	2490
[u(f)]	A.S.A.	1436	793	68	9.48	2182
	S.D.S.	1619	905	114	6.28	2422

Note—Frequencies are expressed in hertz, durations are in milliseconds, and slopes are in hertz/millisecond.

results for Talker A.S.A., and the four panels in the right halves of these figures display the results for Talker S.D.S. Each display includes 100 msec of the fricative and the entire vocalic portion of the utterance. The formant bands are centered on the mean formant frequencies, with their bandwidths representing 99% confidence intervals on these means. The edges of the bands have been smoothed slightly with a triangular, 1-2-1 window. In regions where the confidence interval could not be computed because the formant track was lost or where the interval exceeded 1 kHz, the width of the displayed interval was arbitrarily set to 1 kHz, and the lines bounding the formant bands were not drawn. These regions occurred most often at the boundary between the fricative and vocalic portions of the utterance. Mean formant amplitude ranges are represented by the density of the vertical lines within the formant bands. These lines represent a 60-dB range. Regions where the vertical lines are absent correspond to mean formant amplitudes of 0 dB or less; regions where the lines are most closely spaced correspond to amplitudes of 60 dB or more.

The formant statistics for pairs of utterances that contrast unrounded [a] with rounded [u] are displayed in corresponding panels of Figures 4 and 5. Comparisons of the [sa] and [su] utterances for both talkers (Figures 4a and 5e, and Table 2) reveal that the mean formant frequencies at voice onset differ somewhat, although their confidence intervals indicate substantial overlap in the distributions. Because of this overlap, the differences in mean onset frequencies in the two vowel contexts were not statistically significant. The formants in [sa] also had higher amplitudes and were less variable than in [su]; however, the largest difference for both talkers was in

the F2 transitions. In [su], these transitions averaged more than 200 msec in duration and extended over almost the entire vocalic segment, whereas in [sa] they were less than 100 msec in duration.

Initial transitions in CV syllables following consonantal release are usually associated with the tongue and jaw movements of the opening gesture and rarely exceed durations of 100 msec at normal speaking rates (e.g., Gay, 1978). Thus, the lengthy transitions in [su] are unlikely to have been produced by the movements of these articulators. We suggest instead that the F2 transitions in [su] might be produced by the actual lip-rounding gesture that begins during the fricative and continues over most of the vocalic portion of the utterance. The same types of transitions with slightly greater durations are also seen in both talkers' productions of [fu] (see Table 2).

Lip-rounding transitions are also evident in the VF vocalic segments for both talkers (see Figures 4b, 4d, 4f, and 4h). The onset and offset frequencies of these transitions are similar to those for the FV transitions; however, the dynamic characteristics of the FV and VF transitions differ in certain respects. The VF transitions average about half the duration and have slopes that are about twice the magnitude (see Table 2). This comparison indicates that the unrounding gesture in VF utterances begins later and is shorter in duration and more rapid than is the rounding gesture in FV utterances. The analyses of the fricatives suggest, however, that the amount of rounding during the fricative is about the same for both initial and final fricatives.

A second important difference between the FV and VF transitions is in the formant amplitudes during the vocalic portion of the utterance. The FV transitions maintain rela-

tively stable amplitudes throughout, whereas the VF transitions lose amplitude rapidly as they rise in frequency—ending up 40–50 dB less than the average amplitude of the initial transitions (see Figure 4). This amplitude difference, which would almost certainly decrease the salience of the lip-rounding cues in the VF vocalic segments, is due in part to the diminishing amplitude of voicing in anticipation of the upcoming voiceless fricative. Thus, the asymmetrical perceptual effects of vowel quality could not be caused by differences in the precision or the amount of coarticulation in FV and VF utterances. Instead, they are more likely to have been caused by differences in the timing of the lip-rounding gesture in relation to the amplitude of voicing. These timing differences would function to reduce the salience of the cues provided by lip-rounding transitions in VF utterances.

Effects of fricative place of articulation. We confined our analysis of the effects of the fricative place contrast in the FV and VF vocalic segments to those that contained [a], since the perceptual effects of the place contrast were largest for the stimuli containing these segments. The formant statistics for [sa] and [ja] for the two talkers are displayed in Figures 5a, 5c, 5e, and 5g. The mean F2 onset frequencies were 200–300 Hz higher for [ja] with approximately equal transition durations (see Table 2). Even though the magnitude of this frequency difference is not large on an absolute scale, the confidence intervals on the means are so narrow that the formant bands do not overlap until the transitions are almost completed. The slopes of the transitions in [ja] were also about twice as great as in [sa] for both talkers. In other words, for both talkers there were small but extremely reliable differences in F2 onset frequencies and slopes associated with the place contrast, which are likely to have provided vocalic cues for fricative place of articulation. F3 onset frequencies were also about 300 Hz lower in [ja] than in [sa]; however, the wider confidence intervals and lower amplitudes for F3 make it a less likely vocalic place cue for the fricative.

The F2 transitions in the VF vocalic segments have offset frequencies, amplitudes, and confidence interval bandwidths similar to those for the FV utterances (see Figure 5). Thus, the perceptual asymmetries in the place effects cannot readily be attributed to a reduction in the amplitude of the VF transitions, as seen for vocalic segments containing [u]. In Figure 5, it may also be seen that the vocalic transitions in the VF segments are no more variable than are those in the FV segments. Any measurable asymmetry appears to lie in the durations and slopes of the transitions. For both talkers, the F2 transitions in [as] were twice as long as in [ja]. The slopes of the initial transitions were also about one third greater than the slopes of the final transitions.

Aspects of the slope and duration of the vocalic formant transitions, then, are the only acoustic cue that might explain the greater effects of the following vocalic segments. If the greater slope and duration of these transitions is responsible for the asymmetry between the FV and VF stimuli studied in Experiment 1, then we should

expect that the FV stimuli would persist in giving the greater effect, even when they are played backward such that they are heard as VF. Experiment 3 offered a test of this prediction.

EXPERIMENT 3

Experiment 3 was a perceptual study that placed acoustic differences in the FV and VF stimuli in opposition to the heard order of the fricative and vocalic portions. This was accomplished by requiring the subjects to listen to a tape that contained a subset of the FV and VF stimuli from Experiment 1, played in reverse. In referring to the stimuli in this experiment, we have adopted a convention of referring to the *heard order* of the fricative and vowel. Thus, when the tape was played backward, the subjects heard the original FV stimuli as VF_{reverse} and the original VF stimuli as FV_{reverse}.

The logic of comparing forward and backward presentation has been successfully applied by Sharf and Beiter (1974), who, noting that the consonants in VC syllables are recognized more accurately than are those in CV syllables (Sharf & Hemeyer, 1972), went on to show that formant transitions make a greater contribution to consonant place of articulation when they follow a vocalic segment than when they precede it. Our question was whether heard order would be the more important factor in the asymmetric effects of preceding and following vocalic segments.

Backward presentation results in slightly aberrant sounding stimuli, but it has the advantage of allowing us to manipulate heard order while controlling for any effects of acoustic structure. To the extent that differences in the acoustic structure of FV and VF syllables were responsible for the asymmetry observed in Experiment 1, we should expect that, in the backward-listening condition, the VF_{reverse} stimuli should exhibit stronger context effects than should the FV_{reverse} stimuli. This is because the former have the acoustic structure of the FV syllables that gave the greater effect in Experiment 1. The opposite result is predicted by the view that heard order is important, since if the advantage of the original FV stimuli rests on the fact that the vocalic segments followed the fricative noises, then the FV_{reverse} stimuli should show stronger effects because they present the vocalic segment after the frication. According to this account, the weaker effects of the VF segments from Experiment 1 should actually become stronger, whereas the strong effects of the FV segments should become weaker.

Method

Subjects. The 11 volunteer subjects were undergraduate students at the University of California, Irvine, who received course credit for their participation in the experiment. All were native speakers of English and reported no known history of hearing or speech disorders.

Stimuli and Procedure. Pilot testing conducted in preparation for Experiment 3 indicated that the effect of reversing the stimuli was comparable for the 2 talkers used in Experiment 1. On the ba-

sis of this indication and the negligible effects of talker in Experiment 1, we decided to use only the stimuli from Talker A.S.A., the talker who had produced the largest effects of vowel quality and place of articulation in Experiment 1. The stimuli were further limited to two of his three tokens, randomly chosen with no effort to select or reject particular tokens.

The FV vocalic segments {each of the nine fricative noises combined with each of the two tokens of [(s)a], [(f)a], [(s)u], and [(f)u]} and the VF segments {each of the nine noises combined with each of the two tokens of [a(s)], [a(f)], [u(s)], and [u(f)]} were recorded in three blocks of randomly ordered identification trials with a 4-sec interstimulus interval and a 10-sec interblock interval.³ Over the three blocks, each of the 144 stimuli (16 syllables $\times$ 9 fricative noises) occurred a total of three times.

The test tape was presented to individual listeners binaurally using a Technics 1200 tape recorder, TDH-39 earphones, and a sound-attenuated testing chamber. The tape was first played in the *forward* direction and after a brief break, the tape was played in *reverse*. Presentation of five randomly selected backward stimuli was used to prepare the subjects for listening to the backward test stimuli. In both conditions, the listeners were instructed to identify the consonant in each stimulus as "s" or "sh," writing their responses as in Experiment 1.

Results

The mean number of "sh" responses is displayed in Figure 6, where responses are averaged across subjects

Figure 6. The effect of perceptually reversing the FV and VF stimuli employed in Experiment 1. Group responses are plotted as a function of synthetic frication noises (Stimuli 1-9). The left panels present group responses to stimuli heard in the original order (replicating Figure 1); the right panels present responses to stimuli that were heard in reverse. In both cases, the panels are marked according to the heard order of the stimuli (i.e., the *reversed FV* stimuli had originally been heard as VF stimuli, and the *reversed VF* stimuli had originally been heard as FV stimuli). The legend in the left panels identifies the vowel quality and original fricative place of articulation associated with each of the natural vocalic segments that defines the vocal context, and is the same as that for Figure 1.

and tokens and appear as a function of the synthetic fricative noise (Stimuli 1–9). Each point represents the average of 66 responses. Responses to stimuli in which the vocalic segment followed the frication noise appear in the top panels; responses to stimuli in which the vocalic segment preceded the frication noise appear below. The left side of Figure 6 contains a replication of Experiment 1, summarizing the mean responses to stimuli heard in the forward direction. As in Figure 1, the effects of vowel quality and fricative place of articulation are stronger for the FV stimuli than for the VF ones.

The right panels are for stimuli heard in reverse. The top panel summarizes the pattern of responses when VF stimuli were heard as FV_{reverse}; the bottom panel is for FV stimuli heard as VF_{reverse}. To the extent that acoustic structure was a factor in the asymmetry seen in Experiment 1 (and replicated on the left side of Figure 6), the context effects should be greater in the VF_{reverse} panel since that panel summarizes responses to the acoustic stimuli (the FV syllables) that had given the larger effect in Experiment 1. To the extent that heard order was a factor, the context effects should be greater in the FV_{reverse} panel since that panel summarizes responses to stimuli in which the vocalic segment was heard after the fricative noise. It is quite clear that the context effects are greater in the upper panel; it would appear that heard order of the stimuli is the more important factor.

These observations are validated by an ANOVA in which the dependent variable was the total number of “sh” responses summed across tokens; the repeated measures were listening context (forward or backward presentation of the tape), syllable position (heard order of fricative and vowel, i.e., FV or VF), consonant place of articulation ([s] or [ʃ]), and vowel quality (unrounded [a] or rounded [u]). The overall effects of place of articulation and vowel quality were significant [$F(1,10) = 182.1$, $MS_e = 67.09$, $p < .0001$, and $F(1,10) = 18.44$, $MS_e = 195.16$, $p < .002$, respectively]. There was no overall effect of syllable position; the subjects tended to give as many “sh” responses to the backward stimuli as to those heard in the normal (forward) direction. The effects of both vowel quality and place of articulation interacted with syllable position [$F(1,10) = 9.62$, $MS_e = 50.20$, $p < .01$, and $F(1,10) = 10.91$, $MS_e = 13.82$, $p < .008$, respectively], indicating that both effects were greater when the vocalic segments followed the fricative noises. However, there was no interaction between listening context and either vocalic property ($p > .4$) and no three-way interaction involving heard order. The greater effects of a following vocalic segment were essentially the same for the original FV syllables as for the VF syllables played backward as FV_{reverse}.

GENERAL DISCUSSION

Experiment 1 revealed asymmetric vocalic context effects in the perception of fricatives in VF and FV syllables. Labeling of fricatives in FV stimuli was influenced

more by vocalic context than was that in VF stimuli; this asymmetry was seen for both vowel-quality (i.e., lip-rounding) and place-of-articulation cues. In Experiment 2, acoustic analyses of the natural FV and VF utterances used in Experiment 1 enabled us to determine whether less coarticulation occurs in VF utterances than in FV utterances, or whether coarticulation in VF utterances is more variable or differs in timing. Any of these factors could make coarticulatory cues less salient. In Experiment 3, the stimuli from Experiment 1 were played backward to dissociate the segmental order in the listening task from the order in which the segments were originally produced.

Experiment 1 replicated the asymmetric vocalic context effect reported in previous research (Kunisaki & Fujisaki, 1977; Mann & Repp, 1980). If these results occurred because of coarticulated vowel rounding and consonant place of articulation, one would expect to find less acoustic evidence of coarticulation in VF utterances. The acoustic analyses of the frication noise spectra showed, however, that the effects of vowel rounding are much the same for FV and VF utterances. The analyses of the vocalic portions of the utterances also showed that the formant transition cues for lip-rounding are no more variable in VF utterances than in FV utterances. In fact, evidence of anticipatory and perseveratory lip-rounding was comparable in FV and VF utterances.

The only major difference in the vocalic portions of the utterances was in the timing of the lip-rounding gesture in relation to the onset and offset of voicing. The transitions in the VF utterances, produced by the unrounding gesture, were more rapid and occurred later in the vocalic segment while the amplitude of voicing was dropping, which could have reduced their perceptual salience. If such a reduction in perceptual salience were the cause of the asymmetry, weak context effects should have been obtained for the FV_{reverse} stimuli that contained the same vocalic segment as the original VF stimuli; however, this was not the case. One could even argue that a reduction in the perceptual salience of the transitions in VF stimuli should actually produce *larger* context effects since they are *unrounding* cues.

Our acoustic analysis also failed to reveal any major differences between FV and VF utterances in formant transitions that could serve as consonant place-of-articulation cues. There were small differences in the durations and slopes of the FV and VF transitions that could have influenced their perceptual salience. If this were the case, one would again predict that the context effects for the FV_{reverse} stimuli should have been as weak as those for the original VF stimuli, but this was not the case.

In Experiment 3, which required the subjects to label the fricatives in hybrid FV and VF stimuli presented in forward and reverse order, the order of the vocalic and frication segments in the listening task, not the order in which they were originally produced, emerged as the primary determinant of the asymmetric vocalic context effects. In other words, the effects of vocalic context are only partially predicted by the coarticulatory interactions

between the fricative and the vowel during production. The direction of the effect of lip-rounding and consonant place of articulation is as one would expect, but there appear to be additional perceptual memory and/or linguistic factors contributing to the magnitude of these effects.

Perceptual memory processes, such as backward recognition masking (e.g., Massaro, 1972), might be used to account for asymmetric vocalic context effects. The vocalic segment in an FV stimulus could mask the preceding fricative segment. Likewise, the fricative segment in a VF stimulus could mask the preceding vocalic segment. Some previous research on backward recognition masking in speech has shown that when a segment has sufficient duration (e.g., 30 msec for vowels) and is intelligible in isolation, a backward masker with similar spectral properties presented with 0 msec delay degrades recognition by only a few percentage points (Dorman, Kewley-Port, Brady, & Turvey, 1977; Pisoni, 1972). According to this research, a backward recognition masking explanation of the present results would be limited by the fact that our vocalic segments and the frication noises had very different spectral properties, were longer than 30 msec, and were separated by 9- to 12-msec gaps. Thus, ideal conditions for backward recognition masking may not have existed.

However, Massaro (1977) has argued that the presence of ceiling effects in performance may sometimes prevent backward recognition masking from exhibiting an influence, because subjects who are performing near ceiling level require very little information to perform accurately. This line of reasoning challenges the research of Pisoni (1972) and Dorman et al. (1977). Insofar as their research involved near-ceiling levels of performance, any conclusions about backward recognition masking may be unwarranted. The present study had the potential to be more sensitive to the effects of backward recognition masking, because it included ambiguous stimuli. Consistent with Massaro's (1977) line of reasoning, we find that frication noises that are ambiguous in isolation tend to be asymmetrically affected by vocalic context (i.e., Stimuli 4-6 on the isolated curve depicted in Figure 1). Nonetheless, we also find that some of the endpoint stimuli that are unambiguously classified as "s" show the same asymmetric effects (i.e., Stimuli 8 and 9; see Figures 1 and 2). Hence, the asymmetric effect of vocalic context is not strictly dependent upon the ambiguity of the frication noises, and if inherent ambiguity is critical, then backward recognition masking is not a sufficient explanation of our results.

Another, somewhat related, perceptual memory explanation might involve a rapid decay in precategorical auditory store (PAS). Since phonetic decisions are based on temporally distributed acoustic cues, earlier cues may have to be temporarily stored in some precategorical form, pending arrival of later cues. The effect of a rapid decay of the precategorical information would be a greater perceptual emphasis on the later cues (e.g., the vocalic segment in an FV utterance or the frication in a VF utterance). PAS with rapid decay has been previously described

in the literature and has received some support (e.g. Crowder, 1971, 1973; Darwin & Baddeley, 1974; Massaro, 1970, 1972). However, Penny (1989) has recently noted that PAS must persist for upward of 60 sec in the absence of other stimuli and have a capacity of six to seven "items" (i.e., syllables), if it is to account for the relevant research findings. According to this view, PAS cannot explain the present asymmetric context effects that concerned single syllables less than 500 msec in duration. Shilling and Soderquist (1990) have also pointed out other factors that limit the adequacy of the PAS model.

Linguistic factors might also be used to account for asymmetric vocalic context effects. Such factors are particularly appropriate to consider, given Repp's (1981) and Tomiak, Mullennix, and Sawusch's (1987) observations that vocalic context effects reflect special proclivities of a phonetic mode of perception as opposed to some more general properties of auditory perception (see also Best, Morriongiello, & Robson, 1981; Mann & Liberman, 1983). One linguistic factor that parallels the asymmetry between FV and VF stimuli is the bias in the direction of phonological rules. Javkin (1979) has observed that phonological rules that assimilate a preceding segment to a following one are almost twice as common as those in the opposite direction. Another linguistic factor relates to the status of CV and VC syllables in language. Ohala and Kawasaki (1984) have observed that the CV junction may be more salient than the VC junction, since the most salient modulations in a syllable occur near the CV junction. Moreover, languages which contain CV syllables are far more common than those with VC syllables. Note, however, that some context effects operate across syllable boundaries and that acoustic information spread over an entire utterance is routinely integrated into a single phonetic percept (Mann, 1980; Mann & Repp, 1980, 1981; Repp, 1978; Repp, Liberman, Eccardt, & Pesetsky, 1978).

In summary, the present results demonstrate the combined influence on vocalic context effects of coarticulatory interactions during production and of perceptual interactions during the listening task. At present, no single model can adequately account for the perceptual interactions that cause a preceding vocalic segment, relative to a following one, to have a weaker effect. Additional research that further dissociates the effects of produced order from the effects of heard order can provide an empirical basis for the development of more adequate models.

REFERENCES

- ABBS, M. H. (1971). *A study of cues for the identification of voiced consonants in intervocalic contexts*. Unpublished doctoral dissertation, University of Wisconsin, Madison.
- BEST, C. T., MORRONGIELLO, B., & ROBSON, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, *29*, 191-211.
- BONDARKO, L. V. (1969). The syllable structure of speech and distinctive features of phonemes. *Phonetica*, *20*, 1-40.
- CROWDER, R. G. (1971). The sound of vowels and consonants in immediate memory. *Journal of Verbal Learning & Verbal Behavior*, *10*, 587-596.

- CROWDER, R. G. (1973). Representation of speech sounds in precategorical acoustic storage. *Journal of Experimental Psychology*, **1**, 14-24.
- DARWIN, C. J., & BADDELEY, A. D. (1974). Acoustic memory and the perception of speech. *Cognitive Psychology*, **6**, 41-60.
- DENES, P. (1955). Effect of duration on the perception of voicing. *Journal of the Acoustical Society of America*, **27**, 761-764.
- DERR, M. A., & MASSARO, D. W. (1980). The contribution of vowel duration, F_0 contour, and frication duration as cues to the /juz/-/jus/ distinction. *Perception & Psychophysics*, **27**, 51-59.
- DORMAN, M. F., KEWLEY-PORT, D., BRADY, S., & TURVEY, M. T. (1977). Vowel recognition: Inferences from studies of forward and backward masking. *Quarterly Journal of Experimental Psychology*, **9**, 483-497.
- DORMAN, M. F., STUDDERT-KENNEDY, M., & RAPHAEL, L. (1977). Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues. *Perception & Psychophysics*, **22**, 109-122.
- GAY, T. (1978). Effects of speaking rate on vowel formant movement. *Journal of the Acoustical Society of America*, **63**, 223-230.
- HARRIS, K. S. (1955). Cues for the discrimination of American English fricatives in spoken syllables. *Language & Speech*, **1**, 1-7.
- HEINZ, J. M., & STEVENS, K. N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, **33**, 589-596.
- JAVKIN, H. R. (1979). *Phonetic universals and phonological change* (Tech. Rep. No. 4). Berkeley: University of California, Phonology Lab.
- KUNISAKI, O., & FUJISAKI, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute of Logopedics & Phoniatrics*, **11**, 85-91.
- LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36.
- MANN, V. A. (1980). Influence of preceding liquid on stop-consonant perception. *Perception & Psychophysics*, **28**, 407-412.
- MANN, V. A., & LIBERMAN, A. M. (1983). Some differences between phonetic and auditory modes of perception. *Cognition*, **14**, 211-235.
- MANN, V. A., & REPP, B. H. (1980). Influence of vocalic context on perception of the [j]-[s] distinction. *Perception & Psychophysics*, **28**, 213-228.
- MANN, V. A., & REPP, B. H. (1981). Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, **69**, 548-558.
- MARKEL, J. D., & GRAY, A. H. (1976). *Linear prediction of speech*. New York: Springer-Verlag.
- MASSARO, D. W. (1970). Preperceptual auditory images. *Journal of Experimental Psychology*, **85**, 411-417.
- MASSARO, D. W. (1972). Preperceptual images, processing time, and perceptual units in auditory perception. *Psychological Review*, **79**, 124-125.
- MASSARO, D. W. (1977). Rate of perceptual processing. *Psychological Research*, **39**, 277-283.
- OHALA, J. J., & KAWASAKI, H. (1984). Prosodic phonology and phonetics. *Phonology Yearbook*, **1**, 113-127.
- PENNY, C. G. (1989). Modality effects and the structure of short-term verbal memory. *Memory & Cognition*, **17**, 398-422.
- PISONI, D. B. (1972). Perceptual processing time for consonants and vowels. *Haskins Laboratories Status Report on Speech Research*, **SR-31/32**, 83-89.
- REPP, B. H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, **24**, 471-485.
- REPP, B. H. (1981). Two strategies in fricative discrimination. *Perception & Psychophysics*, **30**, 217-227.
- REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception & Performance*, **4**, 621-637.
- SHARF, D. J., & BEITER, T. (1974). Identification of consonants from formant transitions played forward and backward. *Language & Speech*, **17**, 110-118.
- SHARF, D. J., & HEMEYER, T. (1972). Identification of place of consonant articulation from vowel formant transitions. *Journal of the Acoustical Society of America*, **51**, 652-658.
- SHILLING, R. D., & SODERQUIST, D. R. (1990). Auditory backward recognition masking: Effect of interaural phase on masker efficacy. *Perception & Psychophysics*, **47**, 332-336.
- SOLI, S. D. (1981). Second formants in fricatives: The acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, **72**, 976-984.
- SOLI, S. D. (1982). Structure and duration of vowels together specify fricative voicing. *Journal of the Acoustical Society of America*, **72**, 366-378.
- TOMIAK, G. R., MULLENNIX, J. W., & SAWUSCH, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America*, **81**, 755-764.
- WHALEN, D. H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]-[j] boundary. *Journal of the Acoustical Society of America*, **6**, 275-282.
- WHALEN, D. H. (1983). Vowel information in postvocalic fricative noises. *Language & Speech*, **26**, 91-100.
- YENI-KOMSHIAN, G. H., & SOLI, S. D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, **10**, 966-975.

NOTES

1. The study only included utterances that contained the consonant and vowel that had been requested and that contained no background noise, vocal-fry, lip-smacking, or peak-clipping.
2. As in the perceptual study, we omitted tokens that contained background noise, vocal-fry, lip-smacking, or peak-clipping, and those in which the speaker had failed to articulate the consonant/vowel that had been requested by the randomization list.
3. Pilot testing had suggested that we use longer ISIs and interblock intervals than had been used in Experiment 1, so that we could allow the subjects sufficient time to respond to the $VF_{reverse}$ and $FV_{reverse}$ stimuli.

(Manuscript received June 7, 1990;
revision accepted for publication December 12, 1990.)