DIFFERENTIAL EFFECTS OF SPEAKER AND VOWEL VARIABILITY ON FRICATIVE PERCEPTION*

KEITH JOHNSON
University of California, Los Angeles

Previous research has shown that listeners' identifications of synthetic fricative noises are influenced by both rounding on an adjacent vowel and by the sex of the speaker who produced the adjacent vowel. In each case, contextual information which indicates a longer vocal tract during the production of the fricative (vowel rounding, or male speaker) results in fewer "sh" responses to the items in an [s] to [ʃ] continuum (i.e., listeners expect generally lower vocal tract resonances from longer vocal tracts). The experiments reported here indicate that these superficially similar perceptual effects interact differently with manipulations of presentation type (blocked *vs.* randomized) and the interstimulus interval. When tokens produced by different speakers were presented blocked by speaker, the perceptual effect of speaker differences was reduced, and the degree of reduction depended on interstimulus interval. However, neither of these manipulations had an impact on the perceptual effect of vowel rounding.

Key words: fricative perception; perceptual normalization; coarticulation; speaker variation

INTRODUCTION

It has been demonstrated that the perception of the fricative noises in an [s]–[ʃ] continuum is affected by rounding on an adjacent vowel (Mann and Repp, 1980; Whalen, 1981), as well as by contextual speaker variation (May, 1976; Mann and Repp, 1980). The experiments described here were conducted to test whether these two perceptual effects represent the same kind of process – that is, whether they respond equally to certain experimental manipulations.

Examples of the effects of coarticulatory rounding on the spectra of [s] and [ʃ] in American English are shown in Figure 1. Panel (a) illustrates the acoustic difference between [s] and [ʃ] in the same unrounded vocalic context. The spectrum for [ʃ] is displaced down in frequency as compared with [s] (Hughes and Halle, 1956; Strevens,

* I appreciate the comments of Van Summers, John Mullennix, John Logan, Steve Goldinger, David Pisoni and other members of the Speech Research Laboratory at Indiana University. Also, many thanks to Peter J. Bailey and Randy Diehl for their helpful comments on an earlier version of the paper. Denise Beike helped collect the data. This research was supported by NIH Training Grant #DC-00012-11 to Indiana University and NIH Training Grant #DC-00029 to UCLA.

Send correspondence to Keith Johnson, Department of Linguistics, UCLA, 405 Hilgard Ave., Los Angeles, CA 90024.

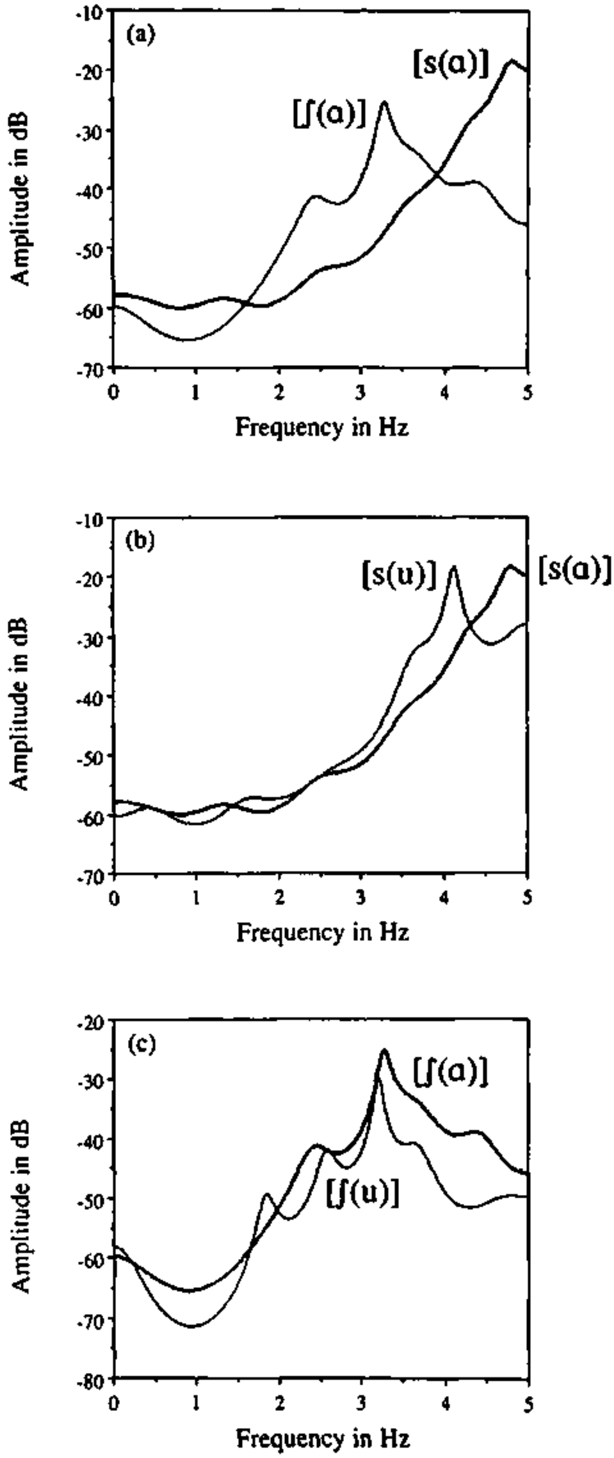

Fig. 1. LPC spectra of the fricative noises [s] and [ʃ] in various vowel environments. All tokens were produced by a male speaker. (a) The [s] of "saw" compared with the [ʃ] of "shah". (b) The [s] of "sue" compared with the [s] of "saw" (c) The [ʃ] of "shoe" compared with the [ʃ] of "shah"

1960; Heinz and Stevens, 1961). Panel (b) shows the effect of coarticulatory rounding on [s], and panel (c) shows the coarticulatory rounding effect on [ʃ]. In both cases the spectrum of the fricative noise produced in the environment of the rounded vowel [u] is displaced down in frequency as compared with the noise produced in the environment of the unrounded vowel [ɑ] especially for [s]. Lengthening the cavity between the lips increases the area of the vocal tract in front of the constriction and hence lowers the resonant frequencies in the fricative noise (Heinz and Stevens, 1961; Shadle, 1985).

Correspondingly, listeners' identifications of fricative consonants in American English are influenced by the quality of the following vowel. Mann and Repp (1980) found that if a synthetic fricative noise from an [s]–[ʃ] continuum is followed by a rounded vowel, listeners respond as if they expected the spectrum of the fricative noise to be transposed down in frequency as compared to the same fricative noise presented in the context of an unrounded vowel (see also Whalen, 1981; Mann and Soli, 1991). Mann and Repp also found that the presence of a silent gap between the fricative noise and the vowel reduced the effect of the vowel on the perception of the fricative.

Another type of variation which affects the acoustic characteristics of fricatives is sex of speaker. Schwartz (1968) found that the spectra of [s] and [ʃ] had, on average, lower frequency prominences for males than for females. Shadle (1985) explored the effect of front cavity area (portion of the vocal tract in front of the fricative constriction) and found that smaller front cavities produced higher frequency prominences in fricative spectra. These findings suggest that sex differences in the acoustics of fricatives are the result of physiological differences. Because men tend to be larger than women,[1] the acoustic resonances of men's vocal tracts tend to be lower in frequency than those of women's vocal tracts. Spectra of [s] and [ʃ] as produced by a male and female speaker are compared in Figure 2. See Badin (1991, Fig. 2) and Schwartz (1968) for comparable comparisons which indicate that the peak in the [s] spectrum for female speakers is above 5 kHz. The acoustic consequence of what is presumably a difference between male and female speakers' front cavity area parallels the difference seen in Figure 1 for fricatives produced in the environment of rounded and unrounded vowels. The generalization is that the spectrum of [s] or [ʃ] produced by a male speaker or in the environment of a rounded vowel has a lower spectral center of gravity as compared with the same fricative produced by a female speaker or in the environment of an unrounded vowel.

Just as there is evidence for effects of contextual vowel rounding on fricative

---

[1] Hiki and Itoh (1986) published calibrated photographs of eight male and seven female Japanese speakers. A two-tailed t-test of volume measurements taken from these photographs (height × width × length of the palate vault) found a reliable difference between the sexes on this rough measure of volume above the occlusal plane ($t = 2.66$, $p < 0.01$). The average volume for males was 39.7 cm^3 and for females was 30.0 cm^3 Howells (1989) took 57 measurements of each of 2509 (1353 male, 1156 female) skulls drawn from populations around the world. He found that males were larger than females on every measure, although, as with the smaller Hiki and Itoh (1986) study, the distributions overlapped.

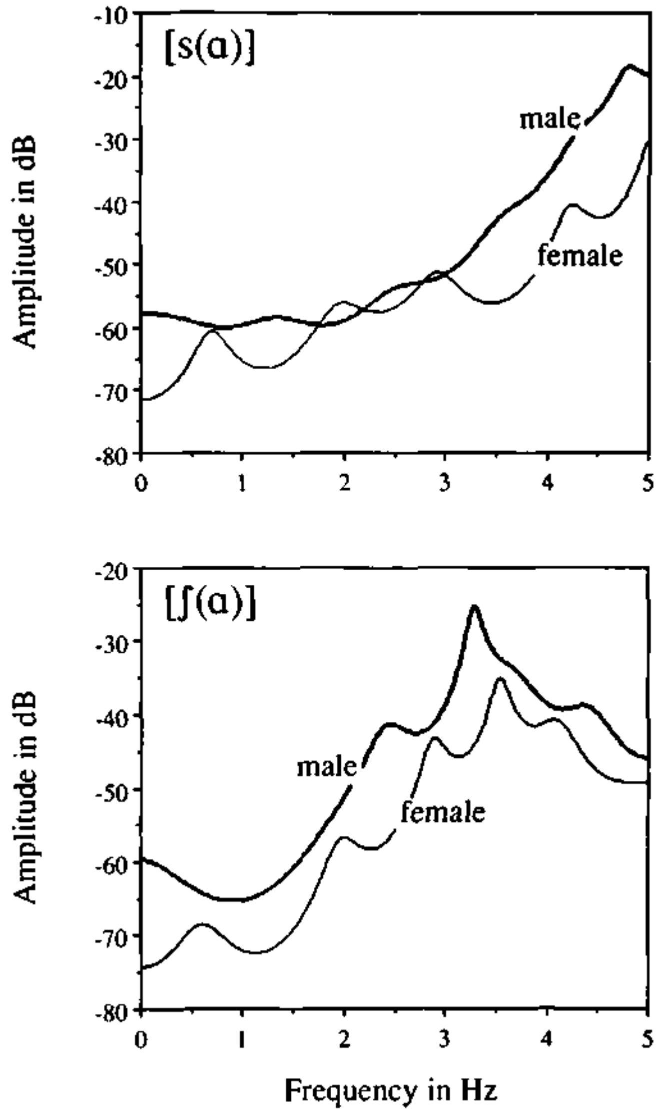

Fig. 2. LPC spectra of fricative noises produced by a male and a female speaker. Top panel: Male and female productions of the [s] in "saw" Bottom panel: Male and female productions of the [s] in "shah"

perception, so there is evidence that listeners' identifications of fricatives are influenced by perceived speaker differences. May (1976) found that, when the formant frequencies of an adjacent synthetic vowel were shifted from values appropriate for a male speaker to values appropriate for a female speaker, the boundary on an [s] – [ʃ] noise continuum shifted toward [s] (i.e., there were more "sh" responses when the contextual vowel had "female" formant values). Mann and Repp (1980) spliced the members of a synthetic fricative noise continuum onto the natural vowels (a) and (u) as produced by a male or female speaker. They found a reliable difference in the boundary between [s] and [ʃ] as a function of the sex of the speaker, with the boundary occurring at a lower spectral peak frequency when the fricative noises were presented in the environment of vowels produced by a male. Interestingly, when a gap was introduced between the fricative noise and the following vowel, the perceptual effects of contextual rounding were substantially reduced (as was mentioned above), but the effect of speaker variability, although significantly smaller, was of the same order of magnitude as it was in the no gap condition (p. 221, Fig. 5). This suggests that the perceptual effects of contextual rounding and speaker differences may occur in different circumstances.

There are several *a priori* reasons to expect the speech perception system to treat vowel and speaker variability differently. First, the effects of these two sources of contextual variability have different degrees of reliability in speech production. The effects of coarticulation are predictable. For instance, in English, vowels preceding nasal consonants are always nasalized, vowel formants always have transitions which reflect the place and manner of articulation of adjacent consonants, and consonants are always pronounced with lip rounding when they precede a rounded vowel. On the other hand, the relationship between acoustic cues and perceived speaker identity[2] is not so reliable. Some speakers who have high F_0, which is an important cue for the perceived identity of the speaker (Johnson, 1990a), have long vocal tracts and thus acoustic/phonetic characteristics which are not typical of other speakers with high F_0. Similarly, some speakers who have short vocal tracts have low F_0. Second, these two sources of variability pose different perceptual problems for the listener. In order for communication to take place listeners *must* categorize the speech signal in terms of linguistic categories (words or syllables or phonemes), whereas speech communication does not require that the listener categorize or uniquely identify the speaker. Third, coarticulatory information in the speech signal is more narrowly localized than is speaker information. For example, the information which a listener might use to decide whether a fricative was pronounced with coarticulatory lip rounding occurs in either the preceding or following brief stretch of sound. On the other hand, speaker information is more widely distributed in the signal; each segment contains acoustic information about the speaker.

To test the hypothesis that the perceptual effects of these two sources of variation differ, this study introduced two manipulations in addition to speaker or vowel variation. Johnson (1990b) found that perceptual vowel normalization is affected by a manipulation of presentation type. When vowels having different fundamental frequencies were randomly intermixed with each other, listeners displayed the type of identification behavior which has been described as F_0 normalization (i.e., they behaved as if they expected the items synthesized with high F_0 to have been produced

---

[2] 'Perceived speaker identity' refers to those physical attributes of the speaker which the listener may extract from the speech signal. Gender is the most obvious such attribute, but it is likely that others are also extracted (accurately or not) during speech perception. See Johnson (1990a, p. 646) for an operational definition.

by a shorter vocal tract than those items synthesized with a low F_0). However, when these same tokens were presented blocked by F_0, the F_0 normalization effect was not present. Johnson hypothesized that this finding reflected the operation of a speaker contrast effect and presented the results of an experiment which seemed to support this interpretation. Whatever the correct interpretation of this result (and the results presented below suggest that speaker contrast is not it), the manipulation of presentation type may be a useful tool in empirically separating effects of speaker and vowel variation in the perception of fricatives. The second parameter manipulated in the experiments reported here was inter-stimulus interval (ISI). The results in Johnson (1990b) suggested that the listener may hold a representation of the speaker in memory from one token to the next in an identification experiment, so increasing the interval between trials may reduce the speaker effect, but not the vowel coarticulation effect.

EXPERIMENT 1: SPEAKER VARIABILITY

Method

Subjects. Forty undergraduate psychology students from Indiana University participated in the experiment in fulfilment of a course requirement. None of the listeners reported any history of speech or hearing difficulty and all were native speakers of American English. The listeners were randomly divided into four groups, as described below.

Materials. A nine step synthetic continuum from [ʃ] to [s] was produced using the Klatt software formant synthesizer (Klatt and Klatt, 1990). The control parameters for the synthesizer are shown in Table 1. The fricative noise continuum was essentially the same as the one used by Mann and Repp (1980), although the step sizes were equal Bark units (whereas Mann and Repp were limited by the frequency intervals available with the OVE synthesizer).

The synthetic fricative noises were concatenated with four different naturally produced vowels. The vowels were spliced (starting from the onset of periodicity, as seen in a digital wave form display) from the words "shah" [ʃɑ] and "saw" [sɑ] produced by a male and a female speaker. Figure 2 shows spectra of the original fricatives in these words. Some acoustic properties of these naturally produced vowels are shown in Table 2 and spectrograms of selected concatenated stimuli are shown in Figure 3. The peak RMS amplitudes of the vowels were equated and then each of the fricative noises was concatenated with the vowels, producing $4 \times 9 = 36$ stimuli. Before being concatenated with the vowels, the peak RMS amplitudes of the synthetic fricative noises were equated at a level 15 dB below the level at which the vowels had been set. This was the average relative amplitude level measured in the natural stimuli. Mann and Repp (1980) found context effects on an [s]–[ʃ] continuum for a wide range of fricative noise amplitudes (6–30 dB attenuation relative to the amplitude of the vowel portion).

Procedure. The experiment was conducted on-line at the Speech Research Laboratory at Indiana University, Bloomington. Stimulus presentation, randomization and response

TABLE 1

Synthesizer control parameters for the synthetic [ʃ] – [s] continuum

	[ʃ] 1	2	3	4	5	6	7	8	[s] 9
F_3	2466	2614	2771	2936	3111	3296	3492	3698	3917
F_4	3108	3294	3491	3698	3918	4151	4396	4657	4932

Step size for F_3 = 0.396 Bark
Step size for F_4 = 0.398 Bark

Bandwidth of each resonance = 10% of the center frequency

Total duration = 140 msec
ramp on = 70 msec
steady amplitude = 35 msec
ramp off = 35 msec
sampling rate = 10 kHz

TABLE 2

Acoustic properties of the naturally produced vowels used in Experiments 1 and 2

	Female		Male		Male	
	"saw"	"shah"	"saw"	"shah"	"sue"	"shoe"
duration	304	324	384	382	334	316
mean F_0	211	214	104	103	108	120
F_1 onset	429	407	464	357	321	321
F_1 vowel	814	764	714	714	286	286
F_2 onset	1357	1714	1286	1538	1500	1736
F_2 vowel	1143	1214	1071	1179	1125	1179
F_3 onset	2857	2821	2571	2393	2257	2236
F_3 vowel	2643	2571	2500	2357	2143	2143
	Experiment 1					
			Experiment 2			

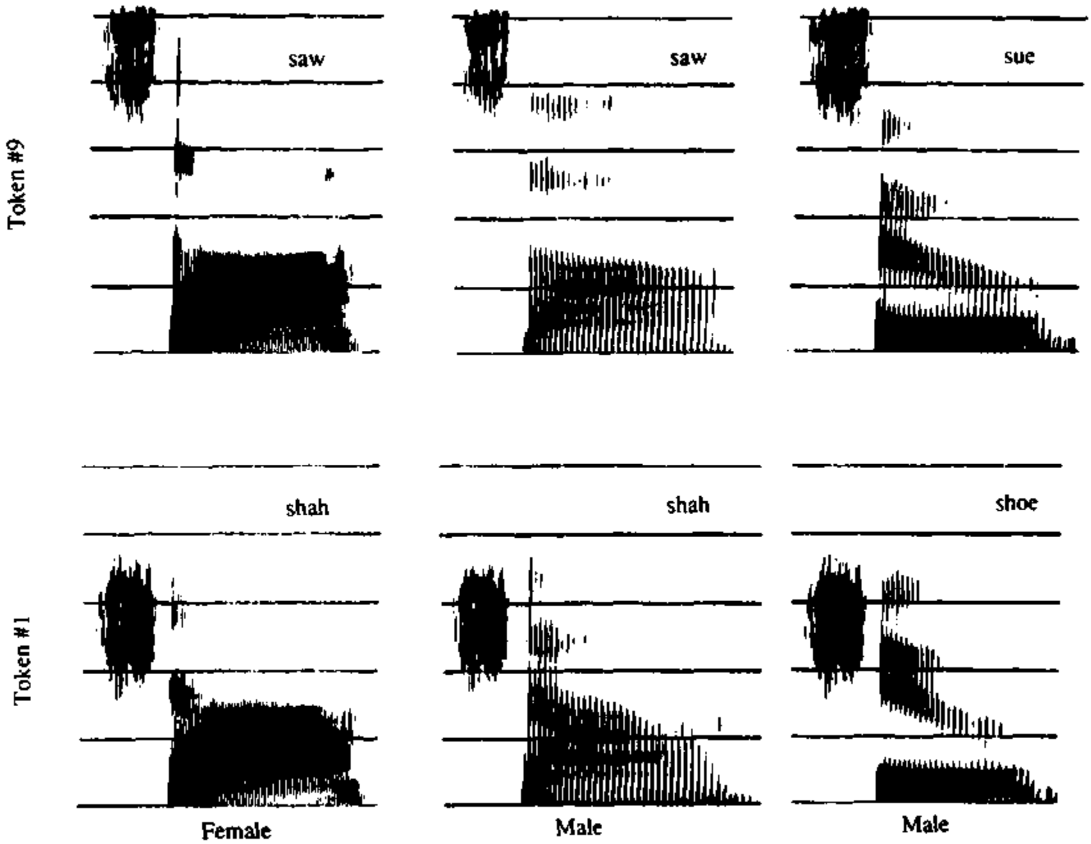

Fig. 3. Spectrograms of representative stimuli from Experiments 1 and 2. The top row of spectrograms shows synthetic fricative noise #9 spliced onto natural vowels produced in the environment of initial [s]; The bottom row shows synthetic fricative noise #1 spliced onto vowels produced in the environment of [ʃ]. The left pair of spectrograms shows [ɑ] produced by the female speaker, the middle spectrograms show [ɑ] produced by the male speaker, and the pair of spectrograms on the right side of the figure shows [u] produced by the male speaker.

collection were performed by a PDP 11/43 computer. Listeners participated in the experiment in groups of four to six listeners. The listeners' task was to identify the fricative in each stimulus by pushing buttons labeled "s" and "sh"

Two groups of 10 listeners identified the fricatives in stimuli blocked by speaker. The listeners in these groups heard a block of stimuli in which the vowels had been produced by one of the speakers and then another block of stimuli in which the vowels had been produced by the other speaker. The order of the blocks was counter-balanced across listeners. The interstimulus intervals (ISIs) for one group in this blocked condition were between 1400 and 2000 msec and those for the other group were between 3400 and 4000 msec. The actual length of the interstimulus intervals were dependent upon response time, with the minimum ISI set to 1000 and 3000 msec for the two groups respectively. Two other groups of 10 listeners responded to blocks of trials in which all 36 stimuli had been randomly intermixed. Thus, for these two groups of listeners, the identity of the speaker was unpredictable from trial to trial. One of the two groups in this mixed condition had ISIs of at least 1000 msec, while the other group had ISIs of at least 3000 msec. All listeners responded to each of the 36 stimuli 10 times, the order of the stimuli being separately randomized for each group of four to six listeners.

Results and discussion

The identification data (number of "sh" responses summed across the continuum) were analyzed in a four way analysis of variance with the between-subjects factors ISI (1000 *vs.* 3000 msec) and presentation type (blocked *vs.* mixed), and the within-subjects factors speaker sex (male *vs.* female) and original consonant context of the vowel ([s] *vs.* [ʃ]).

In addition to the effects of contextual vowel rounding and speaker identity, previous research (Mann and Repp, 1980; Whalen, 1981) has demonstrated that the original consonantal context ([s] or [ʃ]) of the vowel influences the listeners' labelling behavior. If the vowel formant transitions are appropriate for [s], listeners will label more of the stimuli in an [s]–[ʃ] continuum as "s" and, similarly, vowel formant transitions appropriate for [ʃ] result in more "sh" responses. This effect occurred in Experiment 1 [$F(1, 36) = 219.0$, $p < 0.0001$] When the vowel had been produced in the context of [s] the percentage of "sh" responses was 49.9%, *vs.* 66.6% when the vowel had been produced in the context of [ʃ]. The sex of the speaker also had an impact on fricative perception [$F(1, 36) = 126.69$, $p < 0.0001$]. When the vowel had been produced by the male speaker, 52.8% of the responses were "sh" *vs.* 63.8% when the vowel had been produced by the female speaker. This indicates that the listeners expected the fricative spectrum to be generally lower in frequency for the male speaker and thus that more of the stimuli in the synthetic continuum were acceptable tokens of "s" when the voice was male than when it was female.

Presentation type interacted with the speaker sex effect [$F(1, 36) = 25.02$, $p < 0.0001$] This interaction is shown in Figure 4. The speaker effect was smaller in the blocked condition than it was in the mixed condition. This is consistent with the finding of Johnson (1990b) that the F_0 normalization effect in vowel perception was reduced when stimuli were presented blocked by F_0.

There was a marginally significant interaction between speaker sex and ISI [$F(1, 36) = 7.29$, $p < 0.05$]. The speaker effect tended to be larger at the longer ISI. This tendency was due mainly to the blocked condition, as shown in Figure 5. The three-way interaction between speaker sex, ISI, and presentation type was also marginally significant [$F(1, 36) = 3.85$, $p < 0.06$]. In the short ISI conditions, the sex of the speaker had only a small effect on the perceptual boundary between [s] and [ʃ] in the blocked condition, while

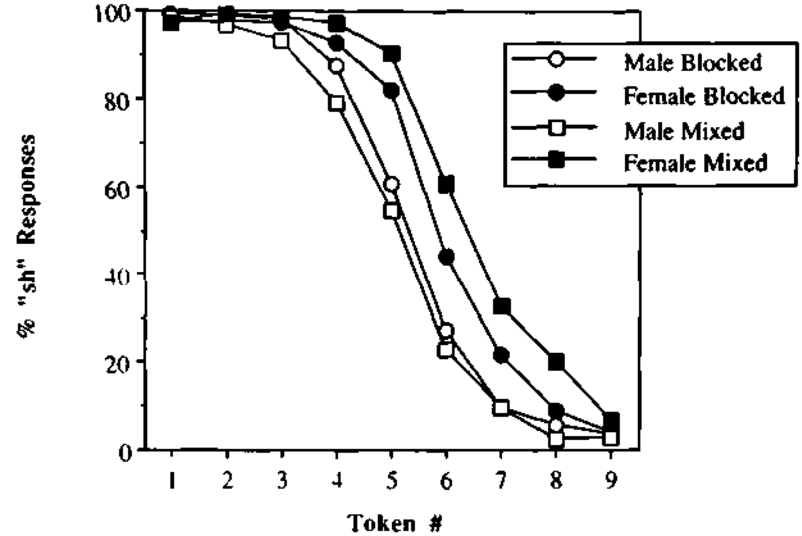

Fig. 4. Results of Experiment 1. The interaction between speaker sex and presentation type. Percent "sh" responses as a function of token number.

the speaker effect was large in the mixed condition. In the long ISI conditions, on the other hand, the sex of the speaker had an effect on the perceptual boundary between [s] and [ʃ] in *both* the blocked and mixed conditions.

The results of Experiment 1 replicate several previous findings: (1) The boundary between [s] and [ʃ] depended on the original consonantal context (Whalen, 1981, showed that this effect is primarily due to the formant transitions in the vowel). (2) The boundary between [s] and [ʃ] depended on the sex of the (contextual) speaker. (3) As would be expected from the results reported by Johnson (1990b), the speaker effect is affected by presentation type: When items were presented blocked by speaker, the speaker sex effect was reduced. An additional finding, which will be considered in more detail in the General Discussion, is that the effect of presentation type was reduced when the interstimulus interval was increased.

EXPERIMENT 2: VOWEL VARIABILITY

Experiment 2 was analogous to Experiment 1, the only difference being that the rounding of the vowel (rather than speaker sex) was manipulated. Of interest were the interactions of presentation type and ISI with the perceptual effect of contextual vowel rounding on the boundary between [s] and [ʃ].

Method

Subjects. Forty undergraduate psychology students from Indiana University

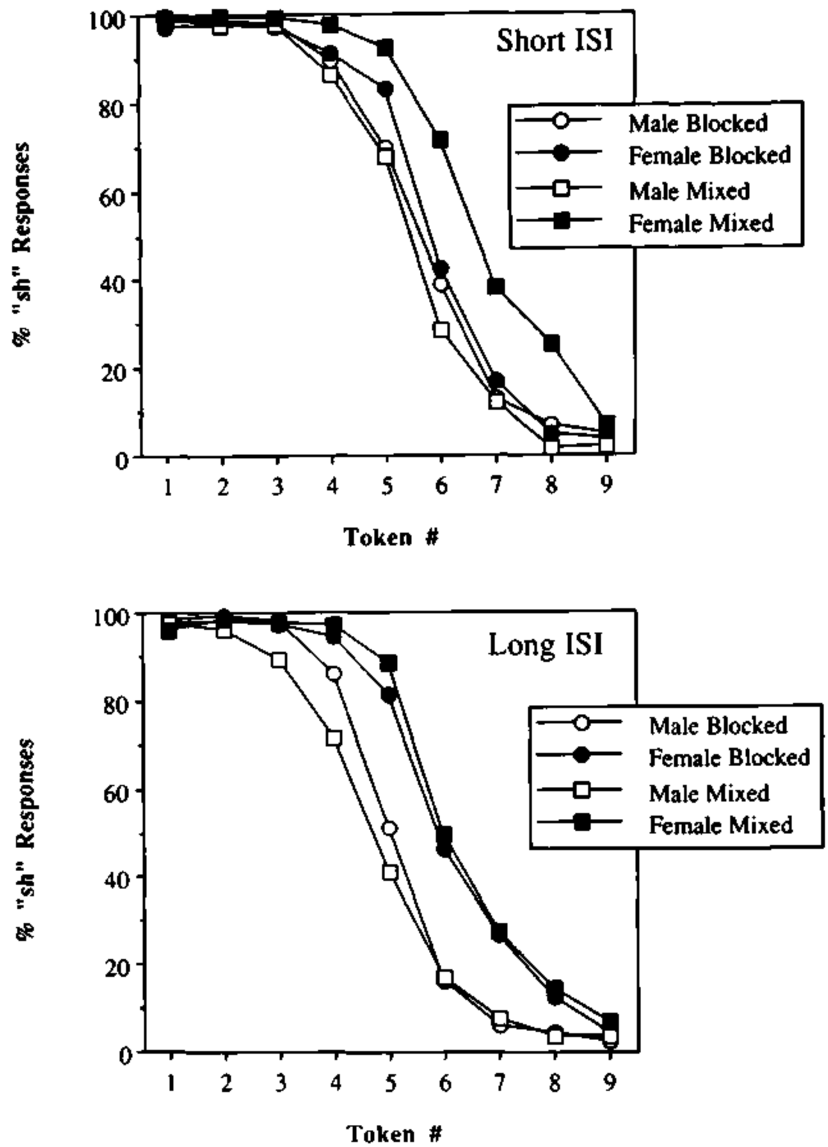

Fig. 5. Results of Experiment 1. The effect of ISI on the speaker sex by presentation type interaction plotted as a function of token number. Top panel: short ISI. Bottom panel: long ISI.

participated in the experiment in fulfilment of a course requirement. None of the listeners reported any history of speech or hearing difficulty and all were native speakers of American English. The listeners were randomly divided into four groups of 10 as discussed below.

Materials. The same nine-step synthetic fricative continuum which had been used in Experiment 1 was used in this experiment. These synthetic fricatives were concatenated with four vowels to produce 4 X 9 = 36 stimuli. The vowels had been spliced from naturally produced utterances of the words "saw" "shah", "sue" and "shoe", and are described in Table 2. The speaker was a male native speaker of American English. Spectra of the original fricatives in these words are shown in Figure 1 and spectrograms of selected concatenated stimuli are shown in Figure 3. As in Experiment 1, the peak RMS amplitude values of the vowels were equated before concatenating them with the synthetic fricatives, and the fricative noises had a peak RMS amplitude which was 15 dB below the peak level of the vowels.

Procedure. All equipment used in conducting the listening sessions was identical to that used in Experiment 1. The listeners were randomly assigned to one of four groups. Two groups heard the stimuli blocked by vowel (stimuli were randomized within blocks), at one of two interstimulus intervals (at least 1000 or 3000 msec between successive stimuli). Two other groups heard all 36 stimuli randomly intermixed with each other, and again the stimuli were presented to one group with short ISIs and to the other with long ISIs.

Results and discussion

The identification data (number of "sh" responses summed across tokens in the synthetic continuum) were submitted to a four-way analysis of variance with the between-subjects factors ISI (1000 *vs.* 3000 msec) and presentation type (blocked *vs.* mixed), and the within-subjects factors original consonant context ([s] *vs.* [ʃ]) and vowel ([a] *vs.* [u]).

There was a main effect for original consonant context [$F(1, 36) = 222.13$, $p < 0.0001$]. When the vowel had originally been produced in the context of [s] the percentage of "sh" responses was 41.7%, and when the original consonant context had been [ʃ] the percentage was 58.9%. This result is consistent with the original consonant context effect found in Experiment 1 and in previous research (Whalen, 1981; Mann and Repp, 1980).

There was also a main effect for vowel [$F(1, 36) = 45.89$, $p < 0.0001$]. Synthetic fricatives in the context of [a] were labelled "sh" 54.6% of the time, while those same fricatives in the context of [u] were labelled "sh" only 46.0% of the time. This result is also consistent with previous research (Mann and Repp, 1980). The only other significant effect was the main effect of ISI [$F(1, 36) = 7.32$, $p < 0.02$] There was a tendency for the synthetic fricative noises to be identified as "sh" more often when the ISI was short (52.9% *vs.* 47.7%, for the short and long ISIs respectively). It is not clear why this difference occurred.

General Discussion

The results of these experiments suggest that the perceptual effect of contextual vowel rounding on fricative perception involves different information in the speech signal than does the effect of speaker sex. Experiment 2 demonstrated that the vowel rounding effect involves only information from within the syllable; experimental manipulations of the larger context (presentation type and ISI) had no effect on the vowel rounding effect. The effect of the original consonantal context was also unaffected by the manipulation of presentation type and ISI. The finding that the perceptual effects of the original consonant context and contextual vowel rounding are similarly unaffected by manipulations of presentation type and ISI suggests that these effects are similar in nature. On the other hand, Experiment 1 demonstrated that the speaker sex effect does interact with the larger context; presentation type and ISI interacted with the speaker effect.

The basic perceptual phenomena (fewer "sh" responses in the context of a rounded vowel or in the context of a male speaker) could be the result of an *auditory contrast* effect: A fricative noise may be perceived as higher in frequency (more [s]-like) in the context of a sound with a lower spectral center of gravity. However, this purely auditory explanation does not provide an explanation of the difference between the contextual vowel rounding and speaker sex effects. The auditory contrast account considers only within-stimulus factors (spectral centers of gravity) and consequently offers no explanation of the interaction of between-stimulus factors such as presentation type or ISI with speaker sex.

An alternative to a purely auditory account of the speaker sex effect views segmental speech perception as *mediated* by the perceptual representation of the speaker. Johnson (1990a) presented data which suggested that listeners used a perceptual frame of reference based on the perceived identity of the speaker for vowel perception. Those experiments found that when the F_0 range in a synthetic precursor phrase was varied, the boundary between "hood" [hʊd] and "hud" [hʌd] showed an effect correlated with the perceived identity of the speaker rather than with the F_0 of the vowel. If, as suggested by the Johnson (1990a) results, the perceptual representation of the speaker mediates segmental speech perception, it may be that the process of developing the speaker representation was affected by the manipulation of presentation type or ISI in Experiment 1.

Johnson (1990b) argued that the interaction between presentation type and speaker sex (in vowel perception) reflects a speaker contrast effect. According to this argument, when items produced by different speakers are randomly intermixed with each other, the perceived identity of the speaker is modified by a contrast effect. Thus, it was argued, a normalization effect which makes reference to a perceptual representation of the speaker will be exaggerated in the mixed presentation. The present data are in general agreement with this account. As in vowel normalization, the process of perceiving fricatives appears to be mediated by speaker information. The fact that the speaker effect is larger in the mixed condition than in the blocked condition suggests (indirectly) that when an experimental manipulation has an impact on the process of developing a representation of the speaker, the listener's perceptual expectations concerning segmental acoustic/phonetic properties are affected.

Taken together with the results reported by Johnson (1990b, Experiment 1), the results of Experiment 1 suggest that there is a basic similarity between perceiving vowels and perceiving fricative noises with regard to the speaker effect; in both cases it interacts with presentation type. This observation is important because it limits the class of theories which can adequately account for the data. If it could be demonstrated that the effect of perceived speaker identity in vowel perception is empirically different from its effect in fricative perception, this could be taken as an argument in favor of theories of vowel normalization which involve reference to formant ratios and the ratio between F_1 and F_0 (Miller, 1989; Syrdal and Gopal, 1986; concerning consonant perception see Jongman and Miller, 1990). The data presented here on fricative perception suggest, however, that a more general theory of normalization is needed; one in which a representation of the speaker mediates both vowel and fricative perception.

Finally, the three-way interaction of ISI, speaker sex, and presentation type in Experiment 1 (Figure 5) suggests that the speaker contrast account of the presentation type effect (Johnson, 1990b) must be modified. It was expected that the speaker sex effect would be reduced in the long ISI mixed condition (as compared with the short ISI mixed condition) because contrast effects in vowel perception are reduced as ISI is increased (Pisoni, 1975). Instead, Experiment 1 found that the effect was present in both the blocked and mixed presentation types when the ISI was long. So, long ISI and mixed presentation both tended to increase the perceptual effect of speaker sex. One way to explain the results of Experiment 1 is to assume that both mixed presentation and long ISI disrupt the process of developing a speaker representation. If this is the correct understanding of the effects of presentation type and ISI, it predicts that the listener's initial expectations about the acoustic difference between male and female fricatives were exaggerated and that with further undisrupted exposure to the speaker (in the short ISI blocked condition) they became more accurate. To test this prediction, data from method of adjustment or goodness-rating tasks are needed in order to determine the internal structure of the listeners' perceptual categories for [s] or [ʃ] under manipulations of speaker, presentation type, and ISI.

(Received March 18, 1991; accepted December 9, 1991)

References

BADIN, P. (1991). Fricative consonants: Acoustic and X-ray measurements. *Journal of Phonetics*, **19**, 397–408.

HEINZ, J.M., and STEVENS, K.N. (1961). On the properties of voiceless fricative consonants. *Journal of the Acoustical Society of America*, **33**, 589–596.

HIKI, S., and ITOH, H. (1986). Influence of palate shape on lingual articulation. *Speech Communication*, **5**, 141–158.

HOWELLS, W.W. (1989). *Skull Shapes and the Map.* Papers of the Peabody Museum of Archeology and Ethnology, Vol. 79. Cambridge, Mass: Harvard University Press.

HUGHES, G.W., and HALLE, M. (1956). Spectral properties of fricative consonants. *Journal of the Acoustical Society of America*, **28**, 303–310.

JOHNSON, K. (1990a). The role of perceived speaker identity in F_0 normalization of vowels. *Journal of the Acoustical Society of America*, **88**, 642–654.

JOHNSON, K. (1990b). Contrast and normalization in vowel perception. *Journal of Phonetics*, **18**, 229–254.

JONGMAN, A., and MILLER, J.D. (1990). Method for the location of burst-onset spectra in the auditory-perceptual space: A study of place of articulation in voiceless stop consonants. *Journal of the Acoustical Society of America*, **89**, 867–873.

KLATT, D., and KLATT, L. (1990). Analysis, synthesis, and perception of voice quality variations among female and male talkers. *Journal of the Acoustical Society of America*, **87**, 820–857.

MANN, V.A, and REPP, B.H. (1980). Influence of vocalic context on perception of the [ʃ]–[s] distinction. *Perception & Psychophysics*, **23**, 213–228.

MANN, V.A., and SOLI, S. (1991). Perceptual order and the effect of vocalic context on fricative perception. *Perception & Psychophysics*, **49**, 399–411.

MAY, J. (1976). Vocal tract normalization for /s/ and /š/. *Haskins Laboratories Status Report on Speech Research*, **SR–48**, 67–73.

MILLER, J.D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustical Society of America*, **85**, 2114–2134.

PISONI, D.B. (1975). Auditory short-term memory and vowel perception. *Memory and Cognition*, **3**, 7–18.

SCHWARTZ, M.F. (1968). Identification of speaker sex from isolated voiceless fricatives. *Journal of the Acoustical Society of America*, **43**, 1178–1179.

SHADLE, C. (1985). *The Acoustics of Fricative Consonants*. Doctoral dissertation, MIT.

STREVENS, P. (1960). Spectra of fricative noise in human speech. *Language and Speech*, **3**, 32–49.

SYRDAL, A., and GOPAL, H. (1986). A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America*, **79**, 1086–1100.

WHALEN, D.H. (1981). Effects of vocalic formant transitions and vowel quality on the English [s]–[š] boundary. *Journal of the Acoustical Society of America*, **69**, 275–282.