431

LANGUAGE AND SPEECH, Vol. 26, Part 1, 1983

VOWEL INFORMATION IN POSTVOCALIC FRICATIVE NOISES*

D.H. WHALEN
Haskins Laboratories

When the postvocalic fricative noises of [s] and [ʃ] are excerpted and combined with vocalic segments having inappropriate formant transitions, vowel quality, or both, the fricative percept is determined by the noise. However, there is often a perception of a diphthong in the vowel. This phenomenon was explored for the vowels [ɑ, i, o, u] preceding the fricatives [s] and [ʃ]. In the first of two experiments, all combinations of the vocalic segments and fricative noises were presented for identification of the vowel. The perception of diphthongs occurred much more often on mismatches of vowel quality than of transition, indicating that there is substantial vowel information in the noise. In the second experiment, just the noises of the syllables were presented, with subjects trying to identify the missing vowel. The high vowels [i] and [u] were reliably identified, while identifications of [ɑ] and [o] were at chance. This result agrees with previous studies of initial fricatives (Yeni-Komshian and Soli, 1981). Fricative noises from [i] and [u] were responsible for the large majority of diphthong percepts in Experiment 1. These results illustrate that fricative noises contain considerable information about preceding high vowels.

INTRODUCTION

In the production of a phonetic string, both anticipatory and perseverative coarticulation occur. The resulting intermingling of phonetic cues makes the extraction of acoustic segments that are all the cues for one phone and cues only for that phone almost impossible (Liberman, Cooper, Shankweiler and Studdert-Kennedy, 1967). Two of the most extractable phones are [s] and [ʃ]. These fricatives are realized by an intense noise that is usually distinct from the accompanying segments, and this noise is quite identifiable as to the fricative produced (Harris, 1958; Heinz and Stevens, 1961; Hughes and Halle, 1956; Yeni-Komshian and Soli, 1981). Yet there is also a substantial and perceivable residue of vowel information (LaRiviere, Winitz and Herriman, 1975; Ostreicher and Sharf, 1976; Yeni-Komshian and Soli, 1981). In addition, there is fricative information that remains in the vocalic segment (Mann and Repp, 1980; Whalen, 1981).

The vowel information in initial fricative noises has been shown, for some vowels, to lead to correct identifications of the vowel from the fricative noise alone. However, this information is not highly salient. Not only are the percentages for correct identification of the vowel well below those for identification of the fricative, this vowel

* *This paper is based on a 1982 Yale University Ph.D. dissertation entitled "Perceptual Effects of Phonetic Mismatches." A previous version was presented at the Annual Meeting of the Linguistic Society of America, December, 1981, New York, New York. The research was supported by NICHD Grant HD-01994. I would like to thank Alvin M. Liberman, Leigh Lisker, Bruno Repp, Rayond G. Daniloff, and two anonymous reviewers for comments on this paper.*

information also does not override the information contained in the vocalic segment when the two cues are made to conflict. Indeed such mismatches seldom result in any directly perceivable effect (Whalen, 1982).

The present work examines the effects of coarticulation in vowel-fricative syllables in two ways. In the first, cross-spliced syllables in which vowel quality cues in the fricative noises and in the vowel itself conflict are shown to give rise to a diphthong percept. Experiment 1 examines this in detail for the vowels [ɑ, i, o, u] and the fricatives [s] and [ʃ]. The second method of examining the coarticulation, which has been used in detail for initial fricatives (LaRiviere, Winitz and Herriman, 1975; Yeni-Komshian and Soli, 1981) and in a less detailed way for final fricatives (Ostreicher and Sharf, 1976), assesses the identifiability of the preceding vowel from the noise alone.

EXPERIMENT 1

Method

A male native speaker of English recorded 10 tokens of each of the syllables [ɑs], [ɑʃ], [is], [iʃ], [os], [oʃ], [us], and [uʃ] on magnetic tape. The rounded vowels were not intentionally diphthongized. Lip configuration was maintained into the frication. While this may have been an exaggerated articulation, it was not unnatural. The two major frequency peaks of the fricative noises were 400-800 Hz lower after the rounded vowels. There was not enough systematic difference at this level of analysis to indicate any other correlation to the vowels.

The stimuli were low-pass filtered at 10 kHz and digitized at a sampling rate of 20 kHz. Two tokens of each syllable were chosen so that both the vocalic portion and the noise would be of equal duration in all eight. A vocalic segment duration of 200 msec was found naturally in eight syllables. Seven were shortened by cutting off between 10 and 50 msec from the first part of the vowel; the resulting abrupt onset did not sound unnatural. The eighth modified vocalic portion was lengthened 20 msec by repeating its first pitch pulse three times. The noises were 250 msec in duration; nine were shortened by removing between 10 and 50 msec from near the end of the signal. Figure 1 shows a sample of the separation of noise from vocalic segment.

Once the tokens had been selected and the durations equalized, each noise was combined with each vocalic segment, including the original. This gave four main categories for the 256 stimuli: 1) The vowel was the same as the one the noise was originally produced with (henceforth, "the vowel matched the *original vowel*" or just "the vowel was matched") and the vocalic formant transitions were appropriate to the fricative ("the transitions were matched"): for example, the vocalic segment from one token of [is] combined with the noise from another token of [is]; 2) the vowel was matched but the transitions were mismatched: for example, the vocalic segment from [is] combined with the noise from [iʃ]; 3) the vowel was mismatched and the transitions were matched: for example, the vocalic segment from [is] combined with the noise from [us]; and 4) both vowel and transition were mismatched: for example, the vocalic segment from [is] combined with the noise from [uʃ].

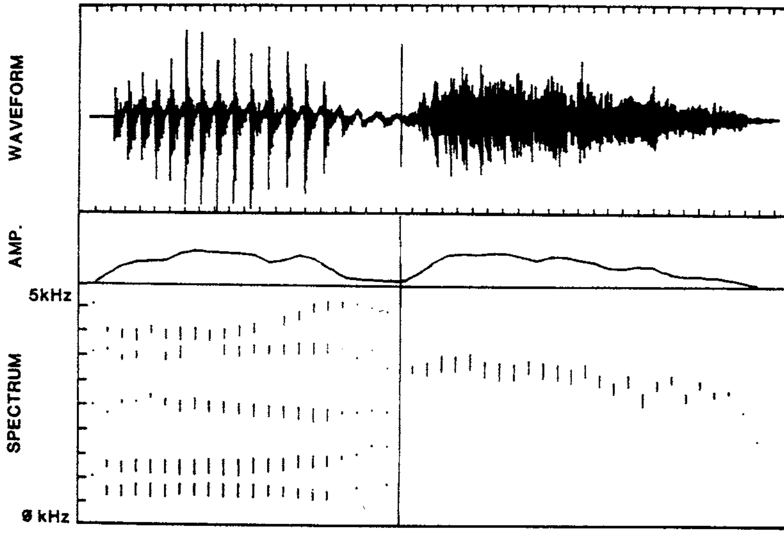

Fig. 1. Representative waveform and spectrum (of [aʃ]), with the vertical line showing where the division between noise and vocalic segment was made.

The stimuli were randomized and recorded on magnetic tape for presentation. The interstimulus interval was 3.5 seconds, with 6 seconds after every ten stimuli.

Ten listeners participated in the experiment. Seven were researchers at Haskins Laboratories who were phonetically trained and/or had extensive experience in speech research. The other three were native speakers of English who had volunteered for experiments at Haskins Laboratories, and were paid for their participation.

Subjects were seated in a quiet room and heard the stimuli over TDH-39 headphones. They were instructed to record their identifications of the vowel on the answer sheet as follows: Non-diphthongized vowels were simply written as "a," "i," "o," or "u," with the phonetic value of each being explained to the naive subjects. Diphthongized vowels were written as a sequence of two of these symbols, whether or not they characterized the exact nature of the offglide.

Results

Each subject provided four judgments of each combination of vowel and original vowel (of the noise). The number of diphthongs perceived by each subject ranged from two to 60 (out of 256 judgments) with a median of 20.5. Misidentifications of the main vowel were excluded from the analysis; they comprised 2.9% of the data.

All four of the vowel categories were given as the second vowel (or offglide). The number of times a particular vowel was identified as the offglide is given in Table 1. There were few reports of [a] and [o] offglides, so these were excluded from the statistical analysis.

TABLE 1

Number of Offglide Percepts, by Fricative Category, for the Four Vowel Qualities

	Fricative was		Total
	s	ʃ	
[ɑ] offglides	1	3	4
[o] offglides	3	6	9
[u] offglides	33	15	48
[i] offglides	137	19	156

Results obtained with initial fricatives would lead us to expect that a mismatch of transition would give rise to diphthong percepts. With some tokens of initial fricatives, joining [ʃ] transitions to a noise from [s] results in the perception of a [j] glide. In the current stimuli, there were 80 syllables in which the vowel quality was matched but the transitions were mismatched. In only one of these cases (the vocalic segment of [oʃ] with the noise of [os]) was a diphthong perceived. With these stimuli, then, the transitions were not the cause of the diphthong percepts.

Of the 204 diphthong responses tabulated, 74.5% occurred when the original vowel and the offglide percept were both [i] or both [u]. If we include those cases where the vowel with which the fricative was produced agreed in relative rounding with the offglide (i.e., [ɑ] giving an [i] offglide and [o] giving an [u] offglide), 93.6% of the cases are accounted for. Thus a large proportion of the responses showed agreement in rounding between the vowel and the original vowel.

Discussion

It is clear that the vowel quality information in the noise is primarily responsible for the diphthong that is perceived. There was one "oi" judgment (mentioned above) when the transition was inappropriate, but overall, mismatch of transition did not seem to be a contributing factor.

The perceptual bias for [i] and [u] judgments is not surprising. These are not only the common offglides of American English, but they are readily articulable in a brief time. (Remember that subjects were to classify offglides that approached [i] and [u] as [i] and [u] rather than being more exact.) To get an [ɑ] percept, for example, there must be tongue and jaw lowering. When there is a fricative to follow, this gesture requires much more time to accomplish than an offglide to, say, [i], since [i] is close to the semi-closed position that [s] or [ʃ] will require. This is the likely reason that listeners rarely reported [ɑ] offglides in the present stimuli.

There were far more [i] offglide percepts than the others. This may be due to the articulatory considerations mentioned above. However, since the offglides are based on vowel information in the noise, it may also be that [i] leaves more information in noises from [i] syllables than other vowels. The next experiment tests for this possibility directly.

EXPERIMENT 2

The results of the first experiment show that the offglides perceived were based on vowel information in the noise. The second experiment examines listeners' abilities to extract that vowel information from the noise alone. If the results match up with those of the first experiment, we will have evidence that the coarticulatory traces of the vowel that can be identified in isolation can give rise to a genuinely perceived vowel-like phone in favorable circumstances.

Method

The fricative noises of Experiment 1 were isolated, and 16 repetitions of each were randomized and recorded on magnetic tape. The inter-stimulus interval was 3.5 seconds.

The 10 subjects of Experiment 1 participated in the experiment. Subjects were seated in a quiet room and heard the stimuli over TDH-39 headphones. They indicated which vowel must have preceded the fricative by depressing one of four buttons, labeled "a," "i," "o," or "u." The phonetic value of each symbol was explained to the naive subjects. The buttons were connected to a computer, which provided immediate feedback for correct responses.

Results

The confusion matrix for the vowel identifications is given in Table 2. For each vowel produced, there are three rows. The first gives the results for [s], the second for [ʃ], and the last for both combined. The values are given as percentages.

Overall, the vowel was correctly identified 41.25% of the time. This was significantly above chance ($t(9) = 4.09$, $p < 0.005$). Of the four vowels, however, only [i] and [u] were identified at above chance levels as measured by a t test (see the second column of Table 3); this was true with both [s] and [ʃ] (first column of Table 3).

The four vowels can be compared on the features of rounding the (relative) height. Subjects correctly identified the roundness of the missing vowel correctly significantly more often than chance (see Table 4; $\chi^2 = 322.04$, $p < 0.001$). Subjects also did better than chance on the height feature (Table 4; $\chi^2 = 48.354$, $p < 0.001$). It appeared that rounding was correctly identified more often than height. Nine of the 10 subjects were more often right on roundness than height ($p = 0.011$ on a sign test).

The two features behaved differently with the different fricatives. When the fricative was [s], more unrounded vowel judgments were given, while [ʃ] elicited more rounded

TABLE 2

Confusion Matrix for Identification of Individual Vowels and Vowel-Fricative Combinations, Experiment 2

	A	I	O	U
[ɑ] s	34.48	42.00	8.15	15.36
[ɑ] ʃ	33.96	11.63	25.79	28.62
[ɑ] s/ʃ	34.22	26.81	16.97	21.99
[i] s	13.44	75.63	5.00	5.94
[i] ʃ	23.82	47.96	13.23	15.99
[i] s/ʃ	18.63	61.80	8.61	10.97
[o] s	25.08	29.47	21.32	24.14
[o] ʃ	16.67	1.57	35.85	45.91
[o] s/ʃ	20.88	15.52	28.59	35.03
[u] s	18.75	24.38	21.25	35.62
[u] ʃ	22.50	2.81	29.69	45.00
[u] s/ʃ	20.63	13.59	25.47	40.31

judgments (Table 5; χ^2 = 322.04). Similarly, the vowel judged to have preceded an [s] was judged as high and [ʃ] as low more often than chance would dictate (Table 5; χ^2 = 48.354).

Discussion

The identifiability of the vowels from the noises agrees well with previous work (LaRiviere, Winitz and Herriman, 1975; Ostreicher and Sharf, 1976; Yeni-Komshian and Soli, 1981). The addition of [o] to the previously studied [ɑ], [i], and [u] allows us to make some tentative comparisons along the features of rounding and height. (Ostreicher and Sharf (1976) use six vowels but do not report the fricative results separately.) These comparisons indicate that rounding is more easily reconstructed from these noises than height. This is presumably the perceptual reflection of the acoustic shaping imposed on the noise by the rounded lips. A relatively lower noise would lead the listener to think that the missing vowel must have been rounded. While the present data tend to bear this out, the higher proportion of round vowel responses to [s] noises confuses the issue.

TABLE 3

Test for Above-Chance Identification of Individual Vowels and Vowel-Fricative Combinations, Experiment 2

Noise	% Correct	% Correct (both [s] and [ʃ]
[ɑ] s	34.48	34.22
[ɑ] ʃ	33.96	
[i] s	75.63*	61.80*
[i] ʃ	47.96*	
[o] s	21.32	28.59
[o] ʃ	35.85	
[u] s	35.62*	40.31*
[u] ʃ	45.00*	

*Significantly better than chance ($p < 0.01$).

Since the [ʃ] noise is lower in frequency than that of [s], the comparison of relative height within [s] or within [ʃ] noises becomes more difficult. In addition, [ʃ] is often produced with some lip rounding even with unrounded vowels, presumably to enhance the low frequencies in the noise (cf. Heffner, 1950, p. 156). This would lead us to expect more *unrounded* vowel judgments from [ʃ]. A study that presented only [s] noises or only [ʃ] noises would help to sort out these complications.

The listeners' responses seem not to have been based on direct perception of a vowel in the noise but rather on educated guesses. Post-session discussions indicate that the only vowels that may have been directly perceived were the [i]s from [s]. Many subjects reported hearing those as a whispered vowel followed by a fricative. Thus the information in these noises, though it reliably influences vowel judgments in the proper direction, is not generally strong enough to build a solid percept in isolation.

TABLE 4

Number of Judgments of Rounded or Unrounded,
High or Low Vowels

		Fricative was produced after a vowel that was: Unrounded	Rounded		Low	High
Vowel identified was	Unrounded	903	373	Low	641	469
	Rounded	451	826	High	633	810

TABLE 5

Number of Judgments of Rounded or Unrounded,
High or Low Vowels

		Fricative produced was: s	ʃ		s	ʃ
Vowel identified was	Unrounded	841	513	Low	471	639
	Rounded	437	762	High	807	636

GENERAL DISCUSSION

The two experiments described show that there is vowel information in the noise portion of final fricatives that is sufficient to give actual vowel (offglide) percepts when the fricative noise is preceded by a mismatched vowel. Considering Experiment 1 by itself, I postulated phonotactic and articulatory reasons for the preponderance of [i] and [u] offglides in the diphthong percepts. Taking Experiment 2 into account, we can see that these are the two vowels that are inherently more identifiable from the noise. Thus the vowels that leave the strongest coarticulatory trace, as measured by identifiability in Experiment 2, are the most common diphthong percepts in Experiment 1. In addition, those noises that prompted the most correct identification of the missing vowel were the noises that gave rise to the majority of the diphthong percepts (156 of 204, as noted above).

The two major effects seen in the present experiments, that high vowels and, to a lesser extent, rounded vowels leave clear coarticulatory traces in final [s] and [ʃ], are clearly based on the possibilities of articulation. Since the narrow constriction necessary for producing [i] and [u] is close to that needed for the fricatives, the two gestures can affect each other more easily than with the more open [o] and [ɑ]. Since the lips are not primary articulators for [s] or [ʃ], they can maintain their rounding through the fricative uninterruptedly. Although [u] is both high and rounded and [i] only high, [i] was recognized more frequently. This may be due to two factors: First, rounding seems to be detectable both in its presence and in its absence, that is, it is as easy to notice the lack of (coarticulated) rounding in a noise as it is to notice the presence of such rounding. Second, [u] was, by design, closer in articulation to another test vowel, [o], than [i] was. This probably reduced the number of correct [u] responses in a way that [i] responses might have been reduced if [e] had been a choice. The result is better than chance identification of [i] and [u] with higher identifications for [i].

The greater identifiability of the high vowels is apparent in the perception of diphthongs in syllables with mismatched cues as well. While the diphthongs of English usually end in a high vowel (thus providing a possible bias in the perception), they may do so for articulatory reasons. In an offglide, we expect less than full vowel quality; yet if there is a consonant following, we must also have a quick movement into the articulation appropriate for it. The high vowels allow this movement much more easily than the low. This, combined with the greater coarticulation discovered for high vowels in the vowel identification test, can account for the preponderance of [i] and [u] offglides in the diphthong percepts.

Together, these results show that it is inappropriate to call the vocalic segment of a syllable *the* vowel (cf. Repp, 1981). Just as there is consonant information in the vocalic segment (for fricatives, see Mann and Repp, 1980; Whalen, 1981), so is there vowel information in the noise of final fricatives. Therefore, not only is the vocalic segment not entirely a vowel, it is not the entire vowel either. While the vowel information in the noise is not sufficient to override information in vocalic segments, Experiment 1 shows us that it can, in the proper circumstances, be perceived as vowel information. Only further experimentation will tell whether it is powerful enough to affect ambiguous vocalic segments, thus demonstrating its cue value in a more traditional manner.

REFERENCES

HARRIS, K.S. (1958). Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, **1**, 1-7.

HEFFNER, R.-M.S. (1950). *General Phonetics* (Madison).

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.

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

LARIVIERE, C.J., WINITZ, H. and HERRIMAN, E. (1975). The distribution of perceptual cues in English prevocalic fricatives. *Journal of Speech and Hearing Research*, **18**, 613-622.

LIBERMAN, A.M., COOPER, F.S., SHANKWEILER, D.P. and STUDDERT-KENNEDY, M. (1967). Perception of the speech code. *Psychological Review*, **74**, 431-461.

MANN, V.A. and REPP, B.H. (1980). Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception and Psychophysics*, **28**, 213-228.

OSTREICHER, H.J. and SHARF, D.J. (1976). Effects of coarticulation on the identification of deleted consonant and vowel sounds. *Journal of Phonetics*, **4**, 285-301.

REPP, B.H. (1981). On levels of description in speech research. *Journal of the Acoustical Society of America*, **69**, 1462-1464.

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.

WHALEN, D.H. (1982). Subcategorical phonetic mismatches slow phonetic judgments. *Haskins Laboratories Status Report on Speech Research*, SR-71/72, 167-198.

YENI-KOMSHIAN, G.H. and SOLI, S.D. (1981). Recognition of vowels from information in fricatives: Perceptual evidence of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, **70**, 966-975.