Perception & Psychophysics
1989, 46 (3), 284-292

Vowel and consonant judgments are not independent when cued by the same information

D. H. WHALEN
Haskins Laboratories, New Haven, Connecticut

Perception & Psychophysics
1989, 46 (3), 284-292

Vowel and consonant judgments are not independent when cued by the same information

D. H. WHALEN
Haskins Laboratories, New Haven, Connecticut

Despite many attempts to define the major unit of speech perception, none has been generally accepted. In a unique study, Mermelstein (1978) claimed that consonants and vowels are the appropriate units because a single piece of information (duration, in this case) can be used for one distinction without affecting the other. In a replication, this apparent independence was found, instead, to reflect a lack of statistical power: The vowel and consonant judgments did interact. In another experiment, interdependence of two phonetic judgments was found in responses based on the fricative noise and the vocalic formants of a fricative-vowel syllable. These results show that each judgment made on speech signals must take into account other judgments that compete for information in the same signal. An account is proposed that takes segments as the primary units, with syllables imposing constraints on the shape they may take.

The search for the units of speech perception has been a long one. Many levels have been proposed, from the feature to the word, and experimental evidence has been adduced for each. In one sense, this shows that, in fact, many levels inhere in the speech signal. Yet the large extent to which levels can be reconstructed one from another supports the intuition that only one level is primary. Since segments are not produced in an invariant, temporally discrete fashion, it clearly cannot be the case that we perceive segments one after another with no overlap. In fact, the number and strength of phonetic effects that appear acoustically and perceptually within the syllable are such that the syllable has often been proposed as the unit of speech perception. The syllable encompasses the range over which the largest effects of one segment on another occur, so that segments often seem to be recovered after syllables are recognized (e.g., Massaro, 1975; Savin & Bever, 1970; Segui, Frauenfelder, & Mehler, 1981). As Studdert-Kennedy (1976) has pointed out, however, defining the syllable as the basic perceptual unit does not solve the problems of segmentation and invariance: Although segments unfold across a syllable, they do so in such a way that syllables are no easier to identify or to define than are segments. As an alternative, segments can be viewed as primary, but they cannot be viewed as temporally ordered, invariant, and discrete. To avoid assuming those attributes, Fowler (1984) proposes a definition of segment that includes a temporal extent. This allows any stretch of the speech signal to influence more than one phonetic judgment, without having to abandon the segment as the primary unit.

A radical proposal for the primacy of the segment, and, additionally, for totally separate processing of consonants and vowels, appears in Mermelstein's (1978) study. In two experiments, he failed to find mutual influence of a consonant judgment on a vowel judgment, even though those two judgments depended on the same information. On the basis of this evidence, he proposed that the speech signal is scanned twice—once for consonant information and once for vowel—with no interaction between the two decisions. This far-reaching proposal rests on results from a single continuum. The present study is a replication of Mermelstein's first experiment and an extension of another case in which a consonant judgment and a vowel judgment are based on the same information. In both cases, the two judgments were found to interact, so that information used for one distinction was less likely to be applied to the other. The reasons for the discrepancy from Mermelstein's results will be discussed.

EXPERIMENT 1

Experiment 1 replicates Mermelstein's (1978) first experiment, changing only the number of judgments made by each subject (by a factor of four) and the number of subjects (by a factor of two). Two phonetic distinctions in a VC syllable, æ/ɛ and d/t, can be made to depend solely on the duration of the vocalic segment. That is, longer vowels increase the likelihood of the perception of a voiced stop (see Raphael, 1972), and longer vowels are more likely to be heard as /æ/ than as /ɛ/ (see Ainsworth, 1972). With the addition of an initial b, we obtain the response set "bad," "bat," "bed," "bet." If we vary only the duration of the steady-state portion

This research was supported by NSF Grant BNS-8111470 and NIH Grant HD-01994 to Haskins Laboratories. Part of this research was presented at the 114th Meeting of the Acoustical Society of America, Miami, FL, November, 1987. I thank Gerry McRoberts for subjecting himself to Experiment 3. Carol A. Fowler, Susan Nittrouer, Philip E. Rubin, Michael Studdert-Kennedy, Dominic W. Massaro, and an anonymous reviewer all provided helpful comments. Correspondence may be sent to D. H. Whalen, Haskins Laboratories, 270 Crown St., New Haven, CT 06511-6695.

of the vocalic segment, we should be able to see effects of one decision on another.

Method

To the extent possible, the method of Mermelstein's (1978) first experiment was used. Although a different synthesizer was used, the detailed published parameters were followed, leaving little room for acoustic differences.

Stimuli. Synthetic stimuli were created with the Haskins Laboratories software synthesizer, designed by Ignatius Mattingly. Each syllable consisted of 48 msec of linear initial transitions, a steady state of varying duration, and final linear transitions of 48 msec. The onset values for F1, F2, and F3 were 100, 1000, and 2000 Hz, respectively. The steady-state values of F2 and F3 were 1800 and 2500 Hz, and F1 had one of three values—625, 650, or 675 Hz—yielding syllables with transitions of varying steepness. Varying the F1 was apparently intended by Mermelstein to ensure that each subject's vowel judgment would be ambiguous at one or another setting; the manipulation in itself is not crucial. The offset values for F1, F2, and F3 were 100, 2000, and 3000 Hz, respectively. The bandwidths were 60, 80, and 100 Hz. Two higher formants were fixed at 3500 and 4500 Hz.

The duration of the steady state took the values of 48, 72, 80, 88, 96, 104, 112, 120, 160, and 240 msec. Since the fundamental frequency was 125 Hz throughout, these values always resulted in an integral number of pitch periods (of 8 msec each). Each of the three values of F1 occurred at the 10 durations, resulting in 30 unique stimuli.

Procedure. Ten different randomizations of the 30 stimuli were recorded on one audio tape, and another 10 randomizations were recorded on another tape. The interstimulus interval was 2.5 sec, with a 5-sec pause after every 10, corresponding to the end of a line on the answer sheet. The subjects were instructed to write after each stimulus whether they heard the word as "bad," "bat," "bed," or "bet," and to guess if they were unsure.

In each test session, the first tape was played, then a tape for Experiment 2 was played, and finally the second tape for Experiment 1 was played. The subjects heard the stimuli over headphones in a quiet room.

Subjects. The subjects were 16 undergraduate students (12 female, 4 male) with no reported hearing problems. They were paid for their participation.

Results

Raw percentages for each of the four response categories to each of the 30 stimuli are shown in Appendix 1. To examine the dependency of the vowel judgment on the voicing judgment, however, we need to calculate the percentage of /æ/ judgments when the consonant was /d/ and then again when it was /t/. Similarly, to examine the dependency of the voicing judgment on the vowel judgment, we need to calculate the percentage of /d/ judgments when the vowel was /æ/ and then again when it was /ɛ/. Both /æ/ and /d/ judgments should be preferred as duration increases, so that a positive correlation between them would show only that both were being supported by the same cue. A negative correlation, however, would show that the two judgments are in fact dependent on each other, since a negative correlation shows that the duration does not apply equally to both. Mermelstein (1978) did not analyze his results this way, probably because the individual functions will be incomplete whenever one category or another is not used for a particular stimulus. At the shortest durations, "bet" is heard almost all the time (72.4% in this experiment), while at the longest, "bad" is the sizable favorite (75.4% here). For many subjects, this means that there will be only one answer for some of the stimuli and thus some undefined percentages (0 out of 0). By pooling his responses, Mermelstein was able to avoid this problem, but in doing so, he avoided answering the question at hand as well.

To ensure that the present results and Mermelstein's were equivalent, the responses were first analyzed as they were in Mermelstein (1978). Individual probability functions were generated for the vowel judgments pooled across consonants and for the consonant judgments pooled across vowels. A PROBIT analysis then determined the 50% crossover point for each function. Figure 1 shows, on the top half, the crossover values from all 8 of Mermelstein's (1978) subjects, and from half of the subjects from the present experiment, on the bottom. These 8 subjects were chosen for their resemblance to Mermelstein's, but they were not greatly different from the other 8. The same two observations that Mermelstein made can be justified here: As F1 rose in value, /æ/ judgments tended to increase; similarly, since a higher F1 steepened the F1 transition, voiceless /t/ judgments also tended to increase. Thus, as far as comparison is possible, the present results replicate Mermelstein's.

To examine the interaction of one judgment on another, however, we need to look at the vowel functions separately for the consonants and the consonant functions separately for the vowels. Figure 2 shows these two functions for all of the subjects pooled across the three values of F1. (Each of the six individual plots for the different values of F1 looked extremely similar.) Two results are clear from this display. First, both /æ/ and /d/ judgments increase as duration increases. Second, there is a clear separation of the two functions for steady-state durations less than 150 msec. The differences are in the direction predicted if the judgments are dependent: When the consonant is perceived as /t/, more /æ/s are heard (the solid function in the top panel of Figure 2 is above the stippled one). Similarly, when the vowel is perceived as /ɛ/, more /d/s are heard (the solid function in the bottom panel is above the stippled one). In short, the two judgments depend on each other.

Individual PROBIT functions were often indeterminate since each category tended to lack judgments at one end of the continuum or the other. Accordingly, for a statistical test of the trends evident in Figure 2, four difference scores were calculated for each subject at each level of F1 and duration. Each difference was designed to be positive if the predictions of dependency were correct. Thus, for the vowel effect, the two differences of interest were the number of "bat" judgments minus the number of "bad" (i.e., voiceless consonants should leave more duration for the /æ/ judgment) and "bed" minus "bet" (i.e., voiced consonant leaves less for the /æ/). For the consonant effect, the two differences of interest were "bat" minus "bet" (i.e., an /æ/ judgment should leave less duration for the voiced consonant) and "bed" minus "bad" (/ɛ/ judgments leave more for the /d/). An analysis of variance supported the predictions [$F(1,15) = 5.99$, $p < .05$]. The effects were not different due to changes in F1 [$F(2,30) = 1.50$, n.s.], although they differed with duration [$F(9,135) = 25.06$, $p < .001$]. The disappear-

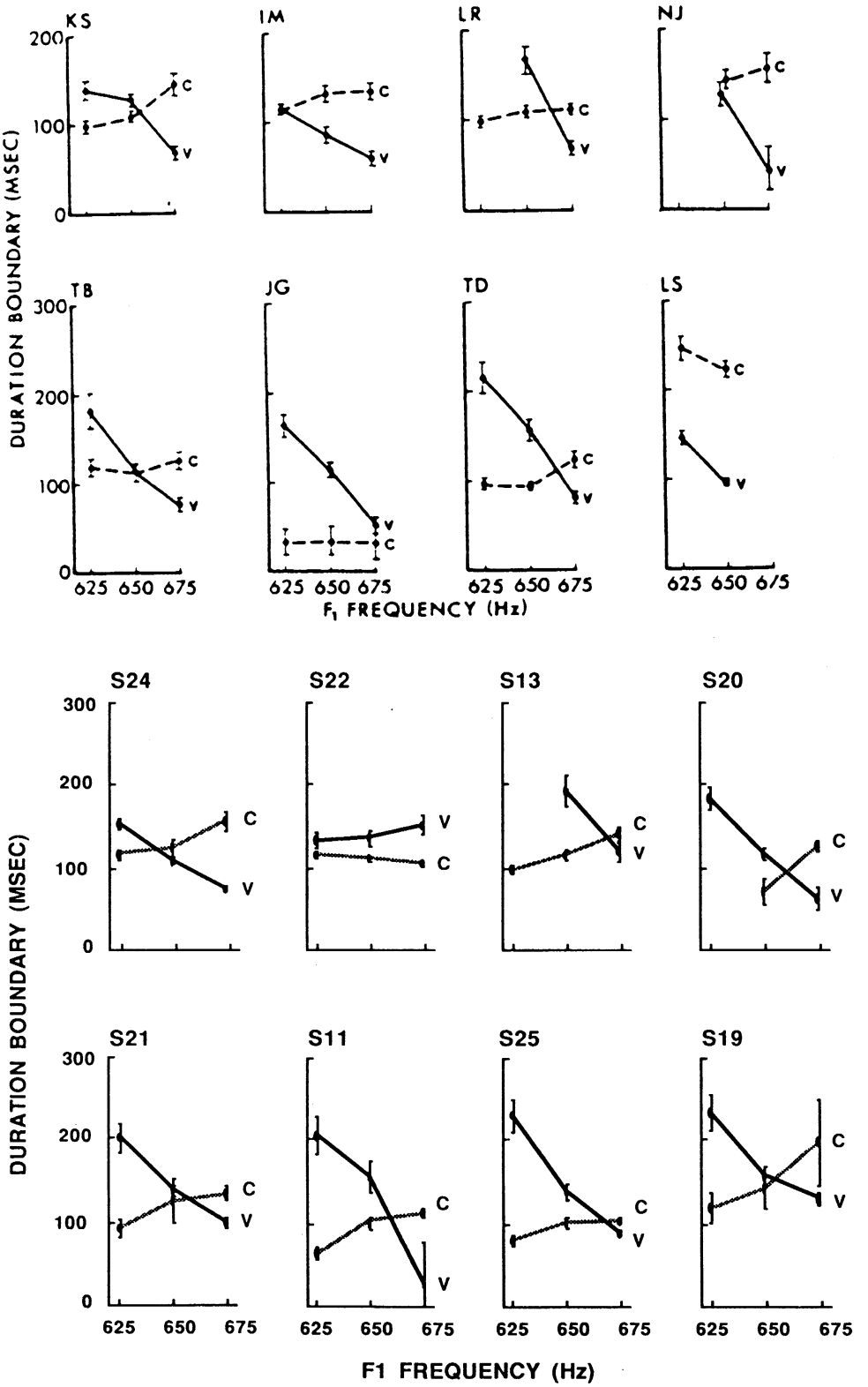

Figure 1. Results from Mermelstein's (1978) Experiment 1, in the top panel, and results from half of the subjects from the present Experiment 1, in the bottom panel. The vowel boundaries are computed across voicing judgment, and the voicing boundaries are computed across vowel judgment.

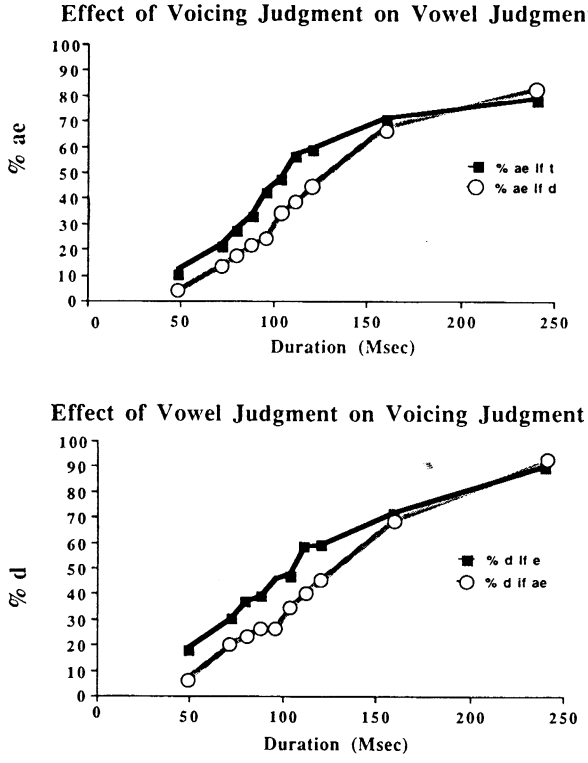

Figure 2. Results from the present Experiment 1 computed in a dependent fashion. Percentages were collapsed across subject and F1 value. In the top panel, the percentage of /æ/ responses when the consonant was /t/ (solid line) is contrasted with that when the consonant was /d/ (stippled line). In the bottom panel, the percentage of /d/ responses when the vowel was /ɛ/ (solid line) is contrasted with that when the vowel was /æ/ (stippled line).

ance of the effect at the longest durations appears to be the root of this effect. The strength of the main effect, however, is clear. These two judgments interact with each other.

Discussion

To the extent that it is possible to compare these results with those published in Mermelstein (1978), there is little difference between the original experiment and the present replication. Analyzing the results as he did gives a similar picture. When, however, those results are analyzed in a way that allows us to ask directly whether the consonant and vowel judgments are independent, we come to the opposite conclusion from that reached by Mermelstein: The two judgments, which depend on the same information in the signal, are dependent on one another.

This discrepancy in analysis has more than one source. As mentioned, Mermelstein does not apply the results of his Experiment 1 directly to the question of independence.[1] Rather, he interprets the F1 effect as bearing on the speaker's ability to account for the temporal constraints on production. He bases his conclusions on his second experiment (which is not replicated here). His second experiment consisted of a directed walk through a two-dimensional phonetic space defined by changes in F1 on one side and changes in duration on the other. After the first few trials of a session, the selection of each later stimulus was based on the subject's judgment of the previous stimulus. F1 frequency was shifted 10 Hz away from the vowel identified, and duration was increased or decreased by 8 msec on the basis of the consonant identified (for /t/ and /d/, respectively). Here again, he found a lack of interaction between the voicing and vowel quality judgments. It seems likely that Mermelstein's second experiment simply maximized trial-to-trial contrast effects, thus washing out any interaction between the two judgments. More importantly, the present Experiment 1 finds positive evidence for an interaction. These reasons seem sufficient for rejecting Mermelstein's conclusion.

While the interaction of the two judgments is clearly present, the boundary shifts that were conditioned by changes in categorization were not as large as might have been expected. By cuing the voicing of a final stop with a voice bar, Raphael (1972) obtained a 63-msec shift in the boundary between voiced and voiceless alveolar stops. The comparable effect in the present experiment was the 21.8-msec shift in the voiced/voiceless boundary when the vowel category changed from /æ/ to /ɛ/. There is no close analogue of the vowel effect in the literature, since Ainsworth (1972) did not use ambiguous formant values. Even so, the 17.3-msec difference in the æ/ɛ boundary based on the d/t judgment is likely to be less than what could be obtained by cuing the vowel category with shifts in F1. Within Experiment 1, the boundary shifts obtained for the extreme values of F1 were 88 and 76 msec for the /d/ and /t/ responses, respectively. Although each judgment has its predicted effect, a decision based on an impoverished signal will not be as strong as one based on more information. Once we accept that the effect should be small, it is not suprising either that the present shifts are in fact small or that Mermelstein (1978) should have failed to find the small effect that was there.

EXPERIMENT 2

The results of Mermelstein (1978) have stood without replication or extension for so long in part because it is difficult to find instances in which two judgments depend on exactly the same information. Another promising case, however, is the coloring of a fricative noise by the quality of an upcoming vowel (see Kunisaki & Fujisaki, 1977; Mann & Repp, 1980; Whalen, 1981). In those studies, it was found that the boundary between /s/ and /ʃ/ was affected by a following vowel, such that rounded vowels gave a lower boundary. The rationale for this shift is that rounded vowels extend the vocal tract and thereby lower the frequency of the noise. In perception, this is compensated for, allowing lower noises to be accepted as /s/ when /u/ follows, and allowing higher noises to be taken as /ʃ/ when /i/ follows. This effect is quite robust and applies

even to vowels that fit poorly, if at all, into the vowel categories of the listener's native language (Whalen, 1981).

Given the general tendency of such context effects to work in both directions, it seemed likely that changes in the fricative noise would affect vowel categorization. Nevertheless, it also seemed advisable to test this notion directly before applying it to a more demanding paradigm. Experiment 2, then, explores the effect of fricative noises on vowel categorization, using natural noises and a synthetic vowel continuum.

Method

To assess the effect of the noise on the vowel, it was necessary to go from an unrounded vowel to a rounded vowel in specifiable steps. A synthetic vowel continuum was used, with /i/ and /u/ as the endpoints. Interpolating systematically between those formant values gives vowels of very un-English qualities, but they were used anyway since Whalen (1981) had found reliable responses for English-speaking subjects to non-English vowels. The subjects were required to use only the two response categories /i/ and /u/.

Stimuli. The syllables /si/, /su/, /ʃi/ and /ʃu/ were recorded on audio tape by a male speaker of American English. These were low-pass filtered at 9.6 kHz and digitized at a rate of 20 kHz. One token of each syllable was selected so that the fricative noises were of approximately the same duration (200 msec for the tokens before /u/, 220 msec for those before /i/). These noises were then extracted for combination with the synthetic vowels. Measurement for the vowel resonance within the noise and for the lowest spectral prominence for the fricative are given in Table 1.

The vowels were synthesized on the Haskins software serial synthesizer. Twelve members of the continuum were created, with formant values of 300, 870, and 2240 Hz for the respective F1, F2, and F3 of the /u/ end, to 270, 2290, and 3010 Hz at the /i/ end. The bandwidths were a constant 80, 80, and 100 Hz. Two versions of each set of steady-state formant values were created. One had an F2 transition somewhat appropriate for /s/ (in every case, starting at 1500 Hz); the other, for /ʃ/ (starting at 2000 Hz). These were used so that there would be more coherence between the (natural) fricative noise and the (synthetic) vowel. The duration of the vocalic segment was 290 msec. F0 began at 110 Hz and dropped linearly throughout the vocalic segment to a final value of 90 Hz. The amplitude was steady for the first 200 msec, then dropped off linearly.

Procedure. Each fricative noise was combined with each of the 12 vocalic segments that had the F2 transition appropriate for that fricative. Five repetitions of the 48 stimuli were randomized and presented to the listeners for a judgment of the vowel category. The responses were written as "ee" for "see" or "she" and as "oo" for "Sue" or "shoe." The subjects were warned that there would be a great many ambiguous cases and that they should guess if they were not sure. The interstimulus interval was 2.5 sec, with 5 sec after every 10, corresponding to a line on the answer sheet.

Table 1
Measurements of Fricative Noise Characteristics and Their Effect on the Vowel Boundary in Experiment 2

Fricative Category	Vowel Context	Vowel Resonance in Noise	Lowest Fricative Pole	Synthetic Vowel Boundary
s	u	1600	3500	1652
s	i	1750	4700	1550
ʃ	u	1735	2300	1533
ʃ	i	1900	2600	1500

Note—All measurements are stated in terms of F2 (in Hz).

Subjects. The subjects were 9 of those who had participated in Experiment 1. (The other 7 subjects of Experiment 1 had heard a preliminary version of this experiment's tape.) Experiment 2 was run between the two blocks of Experiment 1.

Results

Individual subject responses to the vowel continuum were submitted to a PROBIT analysis for each of the four fricative noises. Crossover points were defined in terms of the F2 of the vocalic segment. These values were then submitted to an analysis of variance, with the two factors of fricative category (/s/ or /ʃ/) and original vowel context of the fricative noise (/i/ or /u/). The fifth column of Table 1 shows the mean crossover values for the 9 subjects. Effects both of the vowel context [$F(1,8) = 21.32$, $p < .01$] and of the fricative category [$F(1,8) = 16.42$, $p < .01$] were significant. The interaction was not significant [$F(1,8) = 3.64$, n.s.].

Higher F2 boundaries indicate that more /u/ responses were given. Thus, the /u/ context consistently elicited more /u/ responses. The fricative /s/ also elicited more /u/ responses.

Discussion

The results of Experiment 2 clearly show that it is possible for the acoustic shape of a fricative noise to affect the categorization of a following vowel. One effect is that the original vocalic context of the fricative noise leaves enough of a trace to shift significantly the boundary between /i/ and /u/. These effects include a lowering of the main spectral prominences (Bondarko, 1969; Mann & Repp, 1980) and the introduction of an F2 resonance within the noise (Soli, 1981), both of which were present in the stimuli used here. The fact that these context effects can then be interpreted as vowel information is quite straightforward.

Lower fricative noises, as we have just seen, are associated with rounding, in this case, with /u/. (English /ʃ/ can also be rounded, but the present stimuli showed a difference in /ʃ/ due to vowel category.) The effect of the fricative category on the identification, though, was just the opposite. The low noise of /ʃ/ elicited more /i/ responses than did the high noise of /s/. One possible explanation is that the F2 resonances were being misinterpreted. When the formant resonance in the fricative noise has a higher frequency, more /i/ responses are expected. However, the resonances differ not only with the following vowel, but also with the fricative category: The vowel resonances are higher with /ʃ/ than with /s/. However, this pattern does not fit perfectly with the boundaries obtained. The /ʃu/ formant is lower than the /si/ formant, yet the boundary also is lower rather than higher, as would be predicted. A more likely explanation is that the lack of ambiguity in the fricative category contrasts greatly with the highly ambiguous vowels: To have achieved such a decisive /ʃ/, even an /i/ would have been somewhat rounded, whereas such a distinctive /s/ could only have come about if the /u/s were somewhat spread. Thus, the very lack of ambiguity in the noises forces the opposite

interpretation to the effect on the vowel. Although these explanations cannot be tested fully with the present data, we will see, in Experiment 3, that ambiguous fricative noises that have no formant resonance affect responses as we would expect them to, thus making the issue moot for the present purposes.

While one result of Experiment 2 is puzzling, the other is both straightforward and sufficient to justify the next experiment: Changes in fricative noise can affect the perception of following vowels.

EXPERIMENT 3

Changing a vowel (e.g., from /i/ to /u/) can change the perception of a preceding fricative (in the relevant case, from /ʃ/ to /s/; see Whalen, 1981). Changing a preceding fricative can affect the perception of a following vowel (Experiment 2). With the two of these effects put together, we have another instance in which two phonetic judgments can be made to depend on exactly the same information. If these judgments depend on each other, the interpretation of Experiment 1 will be supported. That is, higher noises should support both /s/ and /i/, leading to a positive correlation between these two categories, whereas interactions between the vowel judgment and the fricative judgment will lead instead to a negative correlation. In addition, if the effect of the vowel on ambiguous fricative noises follows the pattern predicted by previous work, we would have more support for the interpretation given to the results of Experiment 2.

Method

To find dependence of two phonetic judgments competing for the same information, we need to have differing categorizations of exactly the same stimulus. Therefore, in some ways, only a single combination of fricative noise and vocalic formants would be necessary to show an interaction. It also is necessary that the values be ambiguous for all subjects. In Experiment 3, a sampling of a fricative noise, vocalic segment phonetic space was used.

Stimuli. The fricative-vowel syllables were created on the Haskins software serial synthesizer. The vowels were a selection from those of Experiment 2; however, they were synthesized without transitions, giving only a steady state. A pretest was run with a 10-member fricative-noise continuum combined parametrically with 10 of the vocalic segments. The noises consisted of a single fricative pole and a single fricative zero. The poles ranged from 2500 to 3400 Hz in 100-Hz steps. The zeros were placed 1000 Hz lower than the pole. With the characteristics of the synthesizer, this resulted in a noise with a well-defined pole and very little energy in the lower frequencies. In addition, whatever energy was present in the lower frequencies lacked structure, increasing fairly linearly from the low frequencies up to the pole.

The pretest indicated that the 100-member fricative/vowel space was successful at eliciting judgments of the four categories /si, ʃi, su, ʃu/. It was also successful at boring the subjects to the point where their ability to attend to the task was suspect. For the final test, a selection was made to reduce the number of judgments needed. The four extreme (*unambiguous*) syllables, using the combinations of the lowest and highest values of both the noise and the vocalic segments, were used to give the subjects some degree of anchoring of the stimuli. Next, the four middle values of the vocalic formants were combined with three of the middle values of the noise to create the ambiguous stimuli, and 10 of the 12 combinations were actually used. The values are given in Table 2.

Table 2
Difference (in Percent) Between the Contingent Vowel and Fricative Judgments in Experiment 3

Fricative Pole	Vocalic F2 (Hz)	Effect of Fricative on Vowel	Effect of Vowel on Fricative
2900	1386	7.5	7.4
	1515	36.8*	34.5*
	1644	38.5†	41.8†
	1773	13.5	2.5
3000	1515	30.7*	28.7*
	1644	41.8†	37.4†
3100	1386	14.1	14.0
	1515	27.7*	28.4*
	1644	59.7†	53.2†
	1773	30.0*	18.0*

Note—Differences were designed to be positive if the predictions were supported. *t test was significant at the .05 level. †t test was significant at the .01 level.

Procedure. A randomization was made of five iterations of the four unambiguous syllables and 20 iterations of the ambiguous ones. These were recorded on audio tape with an interstimulus interval of 2.5 sec, with 5 sec after every 10, coinciding with the end of each line on the answer sheet. The subjects responded by writing a 1 for /shoe/, a 2 for /sue/, a 3 for /see/, and a 4 for /she/.

Subjects. Ten undergraduate students (6 female, 4 male), with no reported hearing problems, participated. One subject heard more than half of the stimuli as /shoe" and heard only one /see." This indicates that the range used did not include her region of ambiguity. She was therefore excluded from the analysis, and the data from a volunteer from within Haskins Laboratories replaced hers.

Results

Responses to the four unambiguous stimuli were examined only as a check on the subjects' finding the phonetic space acceptable. In all cases, at least three of the four were heard essentially perfectly. Five subjects had few or no /see/ responses to the /si/ stimulus; 1 had no /sue/ responses to the /su/ stimulus; 1 had few /shoe/ responses to the /ʃu/ stimulus; and 3 were essentially perfect on all four.

Raw percentages for the distribution of responses among the four categories for the 10 ambiguous stimuli are given in Appendix 2. To analyze contingencies, however, difference measures were needed. Therefore, responses to the ambiguous stimuli were coded as a pair of contingent percentages: one for the effect of the vowel judgment on the fricative and one for the effect of the fricative judgment on the vowel. For the vowel effect, the percentage of /s/ responses was calculated separately for the instances in which the vowel was /u/ and for those in which it was /i/. The /i/ percentage then was subtracted from the /u/ percentage, since an /u/ judgment is predicted to give more /s/ responses. If there were no /i/ or /u/ responses to a particular stimulus, no percentage could be calculated, and the difference was entered as a zero. This occurred 22% of the time for the vowel

effect. Similarly, for the fricative effect, the percentage of /u/ responses was calculated separately for the instances in which the fricative was /s/ and for those in which it was /sh/. The /sh/ percentage then was subtracted from the /s/ percentage, since an /s/ judgment is predicted to give more /u/ responses. If there were no /s/ or /sh/ responses to a stimulus, no percentage could be calculated, and the difference was entered as a zero. For the fricative effect, this occurred 12% of the time.

The difference scores were submitted to an analysis of variance, with the factors of type of effect (vowel or fricative) and token. The grand mean of 28.3% was significantly different from zero [$F(1,9) = 17.07, p < .01$], indicating that the predictions are supported. There were differences between the vowel effect and the fricative effect [$F(1,9) = 16.37, p < .01$], although the size of the difference was not large (30.0% for the vowel effect vs. 26.6% for the fricative effect). In addition, there were token differences [$F(9,81) = 4.01, p < .01$], which were not statistically different for the two effects [$F(9,81) = 0.93$, n.s., for the interaction].

Individual t tests were run on each of the 10 stimuli for each of the two effects. The results are shown in Table 2. For both the vowel and the fricative effect, three of the stimuli failed to reach significance. Of the ambiguous stimuli, these were three of the four with the more extreme (and therefore less ambiguous) values of both the fricative and the vowel.

Discussion

When a vowel is rounded, a lower than usual frequency in a preceding fricative noise will be heard as /s/. When a /ʃ/ is followed by a vowel, there is a greater likelihood that the vowel will be heard as /i/, since lower fricative noises can be produced by rounding the upcoming vowel somewhat more than usual. The effects can be stated similarly for the expectation of /ʃ/ preceding /i/ and /u/ following /s/. The results of Experiment 3 show that these two effects are not independent: The decision made about the fricative (in an ambiguous syllable) affects the decision made about the vowel, and vice versa. Thus, we have further support for the main conclusion of Experiment 1: When two judgments compete for the same information, the information used for one judgment is not available to the other.

These results also assist in understanding the unexpected effect of fricative category in Experiment 2. In that experiment, the fricative category pushed the vowel judgment in the opposite direction from that predicted by acoustic considerations. If that effect was due to the lack of ambiguity in the noises, Experiment 3 helps make that plausible. If the effect in Experiment 2 was due to the F2 resonance in the noise, Experiment 3 shows that the F2 resonance is not the only aspect of the noise that can affect vowel judgments. Either way, it does seem that there will be a reasonable explanation for the contrary effect in Experiment 2.

GENERAL DISCUSSION AND CONCLUSION

The present results show two instances of a *trading relation* (Repp, 1983) when the trading occurs not between two acoustic parameters, as is usual, but between two phones competing for the same information. In the first case, two categories that depend on the vocalic segment duration are affected by each other. In the second case, several configurations of an ambiguous fricative noise followed by a vocalic segment with ambiguous formant values tended to be heard as the categories that showed the compensatory effects of one on the other. The consonant and vowel judgments, in short, were not independent.

These results put constraints on the possible theories of speech perception. Not only is it impossible for segmental decisions to be made in strict linear order, it is not possible for them to be made without regard to other decisions based on the same acoustic structure. Such interactions have not been a feature of most models, but including them would have different consequences for different theories. As examples, the implications for the fuzzy logical model of Massaro and his colleagues and for the relative prominence theory of Fowler and her colleagues will be detailed.

The need for dependence in phonemic decisions is one that the fuzzy logical model (Massaro & Cohen, 1983a, 1983b; Massaro & Oden, 1980; Oden & Massaro, 1978) has difficulty fulfilling. In this model, phonetic decisions are reached by a probabilistic combination of two independent acoustic features. Although the features are independent, each can take on a different weight in response to the value of the other. The two acoustic features are used to make one phonetic distinction. The most immediate difficulty the fuzzy logical model would have with the present results is that two speech contrasts would have to be calculated from one acoustic feature. That is, even though there were two acoustic features that were manipulated in the present experiments, the effects depended on the setting of only one. For Experiment 1, interactions in the judgments were obtained even when only duration varied. Thus, a fuzzy logical model would have to have two contrasts from one feature, rather than the one contrast from two features it usually has (as for the voiced/voiceless distinction; see Massaro & Cohen, 1983a). Making one decision first and then another independently would give either no correlation between the decisions or a positive one, rather than the negative one obtained. If, on the other hand, the model accommodates the "growing amount of evidence that the prototypes are syllables" (Massaro & Cohen, 1983b:347) by matching to a syllable, the regularities involved would be lost. For the present syllables, vowel duration could be specified as *long* for "bad," *short* for "bet," and *middling* for "bat" and "bed." But the same statements would have to be made for "bead," "beat," "bid," and "bit," for "fad," "fat," "fed," and "fete," and for any number

of other quartets that follow the same principle. Such overt specification quickly puts Massaro's model at a disadvantage in terms of parsimony. If we looked at the dozens of other factors involved in the recognition of syllables, the inadequacy of the fuzzy logical model is even more apparent.

The problems for the relative prominence model (Fowler, 1984; Fowler & Smith, 1986) revolve around the role that syllables should play. This model already has segments that are explicitly based in time and are allowed, if not required, to overlap. Thus, the results of Experiment 3 are prefigured in the previously published papers. The time constraints brought forth by Experiment 1, however, require an explicit syllable that must be fully supported by the segments within it. (If we had only the results of Experiment 1, we would have no reason to think that it is syllables that are important rather than just entire utterances. The relevance of syllable structure in other domains, such as timing [see Cooper, Whalen, & Fowler, 1988], makes it a likely structure here.) How it is that syllables are recognized from the speech stream remains to be determined.

Such dependence in vowel and consonant judgments was assumed by Mermelstein (1978) to be evidence for the syllable as the primary perceptual unit. This is far from a necessary conclusion, however. What is clear is that the syllable is a constraint on the perception, but it is just as likely that what is happening is a mapping from segments onto the syllable's acoustic realization. The mapping reveals the sensitivity of one judgment to another—the fact that the same information is being used for both signifies that both judgments must make sense in relation to each other as well as to the input itself. To say that the syllable is perceived as a whole would seem to indicate that a vast number of probabilities would have to be associated with an equally vast number of possible durations for syllables such as those used here. Even if such syllable templates were stored, the probabilities would depend on each other in a way most easily described in terms of segments. Stipulating that the syllable and its constituent segments are linked in a constrained fashion seems to allow the competition between the two judgments to emerge naturally.

Although the dependence of vowel and consonant judgments is clearly present, this does not exclude instances in which one or another judgment can be made available to consciousness before the influencing environment is available. Many studies have shown an advantage in processing time for syllables over consonants (e.g., Foss & Swinney, 1973; Savin & Bever, 1970; Segui, Frauenfelder, & Mehler, 1981; Swinney & Prather, 1980). These results, however, have recently been shown to be artifactual (Norris & Cutler, 1988), at least for English. When the number of near matches to syllables is equivalent to the number for phonemes, consonants are detected faster than are syllables. These results point to the ability of subjects to extract initial consonant information before the vowel information is available (although the results of Barry, 1984, are in the opposite direction). Similarly, correct identification of stops from gated stop-vowel syllables has been found to precede that of the vowels (Studdert-Kennedy, Kewley-Port, & Pisoni, 1983). Although these results may be due in part to the fact that there is a larger inventory of vowels than stop consonants in English, there are surely going to be cases in which one segment can be correctly identified while another cannot. However, this does not indicate that the information is not used. Even when such a cue not only can specify the consonant category but does (as with a fricative noise occurring with vocalic formant transitions from a different fricative), the other cues are taken into account (Whalen, 1984; Whalen & Samuel, 1985).

Another instance of such conflicting evidence is found in Nittrouer and Whalen (1987). They found that listeners could identify at greater than chance levels the "missing" vowels from fricative noises that were excised from fricative-vowel syllables. To a large extent, this was true whether the fricative identification was correct or not. Yet there was a significant secondary effect that did depend on the correctness of the fricative judgment—one that behaved quite similarly to that found in Experiment 3. Thus, within one experiment, it can be the case that both independent and dependent judgments are being made. It is clearly much easier to account for such data in a theory that allows for dependence than it is in one that denies it: Dependence of consonant and vowel judgments will vary in salience due to changes in task and stimuli, but cannot be denied a role in speech perception.

REFERENCES

AINSWORTH, W. A. (1972). Duration as a cue in the recognition of synthetic vowels. *Journal of the Acoustical Society of America*, **51**, 648-651.

BARRY, W. J. (1984). Segment or syllable? A reaction-time investigation of phonetic processing. *Language & Speech*, **27**, 1-15.

BONDARKO, L. V. (1969). The syllable structure of speech and distinctive features of phonemes. *Phonetica*, **20**, 1-40.

COOPER, A. M., WHALEN, D. H., & FOWLER, C. A. (1988). The syllable's rhyme affects its P-center as a unit. *Journal of Phonetics*, **16**, 231-241.

FOSS, D. J., & SWINNEY, D. A. (1973). On the psychological reality of the phoneme: Perception, identification and consciousness. *Journal of Verbal Learning & Verbal Behavior*, **12**, 246-257.

FOWLER, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics*, **36**, 359-368.

FOWLER, C. A., & SMITH, M. R. (1986). Speech perception as vector analysis: An approach to the problem of invariance and segmentation. In J. S. Perkell & D. H. Klatt (Eds.), *Invariance and variability in speech processes* (pp. 123-136). Hillsdale, NJ: Erlbaum.

KUNISAKI, O., & FUJISAKI, H. (1977). On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin, Regional Institute of Logopedics & Phoniatrics*, **11**, 85-91.

LINDBLOM, B. E. F. (1963). Spectrographic study of vowel reduction. *Journal of the Acoustical Society of America*, **38**, 1773-1781.

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

MASSARO, D. W. (1975). Preperceptual images, processing time, and perceptual units in speech perception. In D. W. Massaro (Ed.), *Understanding language* (pp. 125-150). New York: Academic Press.

MASSARO, D. W., & COHEN, M. M. (1983a). Consonant/vowel ratio: An improbable cue in speech. *Perception & Psychophysics*, **33**, 501-505.

MASSARO, D. W., & COHEN, M. M. (1983b). Phonological context in speech perception. *Perception & Psychophysics*, **34**, 338-348.

MASSARO, D. W., & ODEN, G. C. (1980). Evaluation and integration of acoustic features in speech perception. *Journal of the Acoustical Society of America*, **67**, 996-1013.

MERMELSTEIN, P. (1978). On the relationship between vowel and consonant identification when cued by the same acoustic information. *Perception & Psychophysics*, **23**, 331-336.

NITTROUER, S., & WHALEN, D. H. (1987). Qualitative separateness in children's speech. *Journal of the Acoustical Society of America*, **82**, s84.

NORRIS, D., & CUTLER, A. (1988). The relative accessibility of phonemes and syllables. *Perception & Psychophysics*, **43**, 541-550.

ODEN, G. C., & MASSARO, D. W. (1978). Integration of featural information in speech perception. *Psychological Review*, **85**, 172-191.

RAPHAEL, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, **51**, 1296-1303.

REPP, B. H. (1983). Trading relations among acoustic cues in speech perception are largely a result of phonetic categorization. *Speech Communication*, **2**, 341-361.

SAVIN, H. B., & BEVER, T. G. (1970). The nonperceptual reality of the phoneme. *Journal of Verbal Learning & Verbal Behavior*, **9**, 295-302.

SEGUI, J., FRAUENFELDER, U., & MEHLER, J. (1981). Phoneme monitoring, syllable monitoring and lexical access. *British Journal of Psychology*, **72**, 471-477.

SOLI, S. D. (1981). Second formants in fricatives: Acoustic consequences of fricative-vowel coarticulation. *Journal of the Acoustical Society of America*, **70**, 976-984.

STUDDERT-KENNEDY, M. (1976). Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics* (pp. 243-293). New York: Academic Press.

STUDDERT-KENNEDY, M., KEWLEY-PORT, D., & PISONI, D. B. (1983). Independent consonant and vowel recognition in CV syllables. In A. Cohen & M. P. R. van de Broecke (Eds.), *Abstracts of the Tenth International Congress of Phonetic Sciences* (p. 523). Dordrecht, Holland: Foris.

SWINNEY, D. A., & PRATHER, P. (1980). Phonemic identification in a phoneme monitoring experiment: The variable role of uncertainty about vowel contexts. *Perception & Psychophysics*, **27**, 104-110.

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. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics*, **35**, 49-64.

WHALEN, D. H., & SAMUEL, A. G. (1985). Phonetic information is integrated across intervening nonlinguistic sounds. *Perception & Psychophysics*, **37**, 579-587.

NOTE

1. The relationship between the F1 results and the listener's treatment of temporal variation is intriguing. If, as should be the case (see Lindblom, 1963), a vowel is more centralized when it is produced more rapidly, then longer vowels should be misperceived if they contain the centralized formant values appropriate to a rapid production. The opposite holds both for Mermelstein's (1978) and the present data: Both longer duration and more extreme (less centralized) F1 elicit more /æ/ responses. However, Mermelstein's proposal that this is counterevidence to the relationship of production constraints and perceptual effects presupposes that the subjects were interpreting the durational changes as differences in speaking rate. Since, however, the formant transitions were the same duration for each steady-state duration, it seems more likely that the rate was perceived as constant. In that event, the increased steady-state duration would appear as just that, increased duration, and not as a slower rate. The effect of F1, then, is perfectly straightforward, as is the effect of duration.

APPENDIX 1
Percentage of Responses by 16 Subjects to the 30 Individual Stimuli of Experiment 1

F1 (Hz)	Response Category	Duration of Steady State (msec)									
		48	72	80	88	96	104	112	120	160	240
625	bad	0.9	1.9	3.1	2.8	5.9	8.1	12.5	15.3	38.4	65.9
	bat	4.7	8.1	7.2	8.4	13.8	11.6	13.8	18.1	10.0	3.1
	bed	22.5	32.8	38.8	39.7	44.4	41.3	47.8	44.1	38.1	28.1
	bet	71.9	57.2	50.9	49.1	35.9	39.1	25.9	22.5	13.4	2.8
650	bad	0.0	2.8	6.6	7.8	9.1	17.2	19.1	24.7	50.3	78.1
	bat	9.4	14.1	14.1	18.1	19.7	23.8	26.9	25.0	21.9	7.2
	bed	17.8	23.4	30.3	25.3	28.4	27.8	30.0	29.4	21.3	13.4
	bet	72.8	59.7	49.1	48.8	42.8	31.3	24.1	20.9	6.6	1.3
675	bad	1.3	7.2	7.8	12.2	13.4	18.8	25.9	30.3	53.8	82.2
	bat	16.3	25.3	35.6	37.8	45.9	47.5	45.6	42.2	31.9	9.7
	bed	10.0	17.2	14.4	18.8	15.0	14.4	13.8	12.5	8.4	6.9
	bet	72.5	50.3	42.2	31.3	25.6	19.4	14.7	15.0	5.9	1.3

APPENDIX 2
Percentage of Responses by 10 Subjects to the 10 Ambiguous Stimuli of Experiment 3

Fricative Pole	F2 (Hz)	Response Category				Fricative Pole	F2 (Hz)	Response Category			
		ʃu	ʃi	su	si			ʃu	ʃi	su	si
2900	1386	30.5	6.0	61.5	2.0	3000	1644	12.5	52.5	22.5	12.5
	1515	33.0	23.5	39.0	4.5		1386	10.0	7.0	82.0	1.0
	1644	14.0	69.5	10.5	6.0	3100	1515	19.0	17.5	55.5	8.0
	1773	5.5	86.5	1.0	7.0		1644	7.5	46.0	27.5	19.0
3000	1515	23.5	18.0	48.5	10.0		1773	2.5	67.0	5.5	25.0

(Manuscript received June 27, 1988;
revision accepted for publication February 27, 1989.)