

## Two strategies in fricative discrimination

BRUNO H. REPP

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

Synthetic noises from a [ʃ]-[s] continuum, followed by vocalic portions that influenced the location of the [ʃ]-[s] boundary in an identification test, were presented in AXB and fixed-standard AX discrimination tasks. The majority of naive subjects perceived these fricative-vowel syllables fairly categorically in both tasks; that is, discrimination functions followed the patterns predicted from identification scores and showed shifts contingent on the nature of the vocalic portion. However, two subjects achieved much better discrimination scores than the rest, and so did three experienced listeners and a group of less experienced subjects who had received some discrimination training in the AX task. These listeners, who (judging from their higher accuracy, pattern of responses, and subjective reports) successfully followed the nonphonetic strategy of restricting attention to the spectral properties of the fricative noise, were not influenced by different vocalic contexts. These results support the hypothesis that influences of vocalic context on fricative identification are tied to a phonetic mode of perception.

Several recent studies (Mann & Repp, 1980; Whalen, 1981; Kunisaki & Fujisaki, Note 1) have shown that perception of the [ʃ]-[s] distinction is sensitive to the nature of subsequent vocalic context. Two separate effects may be distinguished. One is due to the quality of the following vowel: Given a somewhat ambiguous fricative noise (often a necessary condition for observing any contextual effects—cf. Harris, 1958), listeners tend to perceive “s” in the context of a rounded vowel (such as [u]) but “sh” in the context of an unrounded vowel (such as [a]). The other effect is due to the nature of the vocalic formant transitions: Listeners tend to perceive “s” when the transitions resemble those normally following [s] frication and “sh” when the transitions resemble those normally following [ʃ] frication. The vowel quality and transition effects are both reliable and pronounced, especially when synthetic fricative noises are spliced together with natural-speech vocalic portions (Mann & Repp, 1980; Whalen, 1981).

One important theoretical question raised by these findings is whether the effects of vocalic context on fricative perception arise at a phonetic (speech-specific) level of processing or whether they are due to some auditory interaction between adjacent stimulus segments. Even though what is known about other contextual effects in speech perception generally suggests a phonetic origin, evidence supporting this contention needs to be adduced for each individual effect, considering the large variety of possible auditory in-

teractions and the sizable group of researchers who seem to believe that such interactions can explain most or all phenomena in speech perception (e.g., Divenyi, 1979; Kuhl & Miller, 1978; Schouten, 1980; Stevens, 1975).

Consider, first, the formant transition effect. In a theory of phonetic perception, it is best described as resulting from the perceptual integration of information conveyed by two separate cues—the fricative noise and the following formant transitions—into a single phonetic percept. The integration is motivated by the fact that both noise and transitions are necessary consequences of producing either [s] or [ʃ] in prevocalic position. On the other hand, if the effect is auditory in origin, it seems implausible that it would arise from perceptual integration, considering the great spectral disparity of the two cues. Rather, the assumptions would be that listeners focus on one cue only (most likely on the noise portion) and that the perception of the relevant auditory properties of the fricative noise is modified by the formant transitions (or vice versa). The auditory mechanisms that could mediate such a perceptual interaction are not obvious, but auditory contrast and nonsimultaneous masking are candidates.

Consider, now, the vowel quality effect. A phonetic explanation for listeners' tendency to hear “s” rather than “sh” in the context of rounded vowels appeals to a well-known coarticulatory effect: Fricative noises preceding rounded vowels characteristically exhibit a downward shift in spectrum, due to anticipatory lip rounding (see Mann & Repp, 1980; Kunisaki & Fujisaki, Note 1). Thus, listeners appear to compensate in perception for a consequence of coarticulation. Of course, such compensation could never occur at a level of processing that has no access to tacit knowledge of articulatory dynamics and con-

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to the Haskins Laboratories. I would like to thank Patti Price for help in conducting the first two experiments and Virginia Mann and Doug Whalen for fruitful discussions and for serving as “expert listeners” in Experiment 2. Other colleagues who have contributed valuable comments include Catherine Best, Alvin Liberman, David Pisoni, and Ovid Tzeng.

textual variations in speech cues. Therefore, to explain the vowel quality effect in auditory terms, we must again assume that the auditory percept of the fricative noise is somehow influenced by the following signal portion (e.g., through some form of spectral contrast).

Since the formant transitions are acoustically dependent on the formant frequencies characteristic of the vowel, the vowel quality and transition effects could reflect a single auditory effect—of the vocalic onset spectrum (cf. Blumstein & Stevens, 1980) on perception of the fricative noise. Thus, this auditory hypothesis has the advantage of parsimony; as we have seen, the vowel quality and transition effects have quite different explanations in a theory of phonetic perception—explanations that are united only by their common appeal to articulation (and its acoustic consequences) as a perceptual guideline.

The present study was conducted to answer the following question: If listeners are led by the task demands to focus on the spectral quality of the fricative noise rather than on its phonetic category, would their responses still be influenced by the periodic stimulus portion following the noise? Presumably, a strictly auditory effect of vocalic context on fricative noise perception would operate whether or not listeners restrict their attention to the noise portion alone. In fact, such a focusing of attention on the primary cue is already implied in the auditory hypothesis, and a further effort on the listener's part should have little if any effect. On the other hand, if the effects of vocalic context are phonetic in nature, they might disappear when listeners focus on the auditory quality of the noise portion (i.e., when they use a perceptual strategy that presumably bypasses the mechanisms specific to phonetic perception).

The extent to which listeners would be successful in adopting such a nonphonetic strategy in judging fricative-vowel stimuli was not known in advance. Many speech stimuli are categorically perceived; that is, untrained listeners perceive the stimuli in terms of phonetic categories even when attempting to make fine auditory discriminations. Typically, however, stimuli that are categorically perceived are distinguished by rather subtle acoustic differences that can be detected only by trained listeners (see, e.g., Carney, Widin, & Viemeister, 1977; Edman, 1979). Fricative-vowel syllables, on the other hand, contain a prolonged noise portion, and it would seem that listeners should be able to detect (sufficiently large) spectral differences in the noise without too much difficulty.<sup>1</sup> Certainly, isolated noises from a [ʃ]-[s] continuum can be discriminated quite easily, even though they can also be labeled phonetically as "sh" or "s" (Healy & Repp, in press).

In the present study, two different discrimination paradigms were used (AXB and fixed-standard AX) that were expected to differ in the extent to which

they facilitated the task of making fine auditory discriminations in the noise portion (cf. Creelman & Macmillan, 1979). In both discrimination tests, fricative noises from a [ʃ]-[s] continuum were presented in the context of several different following vocalic portions. An initial identification test was expected to confirm the earlier finding (Mann & Repp, 1980) that the [ʃ]-[s] boundary shifts with a change in vowel quality or formant transitions. The central question was whether analogous shifts would be observed in the discrimination tasks (as predicted if the stimuli are categorically perceived) or whether selective attention to the auditory properties of the noise portion, especially in the sensitive fixed-standard AX test, would result in a disappearance of vowel context effects.

## EXPERIMENT 1: IDENTIFICATION AND AXB DISCRIMINATION

### Method

**Subjects.** Eight paid student volunteers participated. None of them was experienced in speech discrimination tasks, although some of them had taken part in earlier experiments requiring identification of stimuli similar to those used here.

**Stimuli.** The stimuli consisted of synthetic noise portions followed by natural-speech periodic portions. The fricative noises were generated on the OVE IIIc serial resonance synthesizer at Haskins Laboratories and constituted a nine-member [ʃ]-[s] continuum. The endpoint stimuli were chosen to match approximately in spectrum (below 5 kHz) the [ʃ] and [s] noises of the speaker from whose utterances the periodic portions were taken. The frequencies of the two poles (formants) that characterized each noise are listed in Table 1. Noise duration was 200 msec; the amplitude contour peaked after 150 msec. Overall amplitude was nearly constant across the continuum.

The periodic stimulus portions were excerpted from utterances of [sə], [ʃə], [su], and [ʃu], produced by a male speaker of American English. To indicate the absence of the original fricative noise (but the presence of appropriate formant transitions), these portions will be referred to as [(s)ə], and so on. In an earlier study (Mann & Repp, 1980, Experiment 4), the very same portions had dramatic effects on fricative identification when immediately preceded by synthetic fricative noises from a [ʃ]-[s] continuum similar to the present one. That earlier experiment used three different tokens of each periodic portion, but since token variation was small, a single token of each was deemed sufficient for the present study. Fricative-vowel syllables were constructed by immediately following a synthetic noise with a periodic portion, both having been digitized at 10 kHz and low-pass filtered at 4.9 kHz.

Table 1  
Fricative Noise Stimuli of Experiment 1  
(Pole Center Frequencies in Hertz)

Stimulus	Pole 1	Pole 2
[ʃ] 1	2466	3108
2	2613	3293
3	2769	3488
4	2933	3695
5	3108	3915
6	3293	4148
7	3489	4394
8	3697	4655
[s] 9	3917	4932

There were four identification tests and four AXB discrimination tests, one of each for each periodic portion (a blocked factor). Each identification test contained 10 repetitions of the 9 stimuli resulting from the 9 different noises followed by one particular periodic portion. They were arranged in 5 lists of 18, with ISIs of 3 sec, and 6 sec between lists. Each AXB discrimination test contained 6 repetitions of the 7 2-step comparisons (1-3, 2-4, etc.) in each of 4 AXB arrangements (AAB, ABB, BAA, BBA), resulting in 168 stimulus triads. These were arranged in 6 lists of 28, with ISIs of 500 msec within triads, 3 sec between triads, and 10 sec between lists. The first list of 28 served as practice and was not scored.

**Procedure.** Each AXB test was preceded by the corresponding identification test. The four conditions deriving from the four different periodic portions were distributed over two sessions in counterbalanced order. The subjects listened over TDH-39 earphones in a quiet room. The tapes were played back on an Ampex AG-500 tape recorder. In the identification task, the subjects identified the fricative in each stimulus by writing down "sh" or "s." In the AXB discrimination task, the responses were "A" or "B," depending on whether the second stimulus in a triad was judged to be the same as the first or as the third. The subjects were told to listen carefully for any difference in the noise portion and to guess, if necessary.

## Results and Discussion

**Identification.** Although the identification test was essentially a partial replication of Experiment 4 of Mann and Repp (1980), there were two important differences: (1) The [ʃ]-[s] continuum was more realistic, the endpoints having been modeled on natural speech; and (2) the different periodic portions were blocked rather than randomized. Both changes might be expected to reduce the magnitude of contextual influences on fricative perception: The improved noises were perhaps less ambiguous, and blocked presentation gave listeners an opportunity to adapt to a given periodic portion and to adjust their criteria accordingly. Therefore, it seemed important to demonstrate that vocalic context still influences fricative perception under these conditions.

The results are shown in Figure 1. It is evident that the labeling functions shifted with vocalic context in the expected directions. Listeners were more likely to perceive "sh" in the context of [a] than in the context of [u] [ $F(1,7) = 17.1$ ,  $p < .01$ ], and they were more likely to perceive "sh" when [ʃ]-transitions were present than when [s]-transitions were present [ $F(1,7) = 21.2$ ,  $p < .01$ ]. The interaction between the vowel quality and transition effects was not significant [ $F(1,7) = .5$ ], suggesting that the two effects were independent. The boundary shifts were considerably smaller in magnitude than those observed by Mann and Repp (1980), probably for both of the reasons mentioned (*viz.*, improved fricative noises and blocked periodic portions).<sup>2</sup> However, they were reliable and sufficiently large to predict shifts in discrimination peaks, if categorical perception obtains.

**AXB discrimination.** Preliminary inspection of the results revealed that two of the eight subjects outperformed the others by a wide margin: Their average score was 96% correct. Since these two subjects apparently did something different from the rest, and since their data did not contain any information because of the ceiling effect, their results were excluded. The following results are based on the remaining six subjects only.

The average discrimination functions are shown in Figure 2, separately for each periodic portion, together with predictions derived from the identification results (separately for each subject and then averaged), using the classic low-threshold model of categorical perception (Liberman, Harris, Hoffman, & Griffith, 1957; Pollack & Pisoni, 1971). It is evident that discrimination performance followed the predicted pattern quite closely, except in the [(s)u] condition, in which the match was less good. Discrimination was much better in the boundary region than within phonetic categories, although it was everywhere above chance and usually a good deal better than predicted. There were also indications that the peaks of the discrimination functions shifted as predicted with the nature of the periodic portion, although these shifts did not reach significance here because of the small number of subjects.

At least part of the difference between obtained and predicted discrimination performance may be ascribed to contrast effects in (covert) labeling during the discrimination task (Healy & Repp, *in press*; Repp, Healy, & Crowder, 1979).<sup>3</sup> Therefore, the results of these six subjects indicate fairly strong categorical perception, in agreement with earlier findings of Fujisaki and Kawashima (Note 2) and of May (1979). Apparently, these listeners found it difficult to abandon a phonetic mode of listening and to focus on the auditory quality of the fricative noise; they seemed to make their decisions largely on the basis of the category labels "sh" and "s." It was thought, however, that the more stringent fixed-standard AX

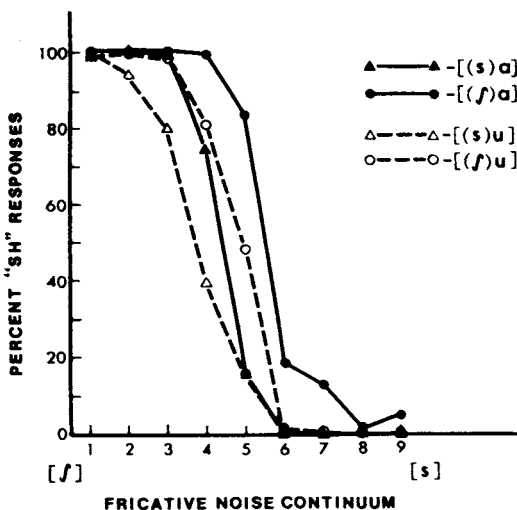


Figure 1. Identification functions for a [ʃ]-[s] noise continuum in four different vocalic contexts (Experiment 1).

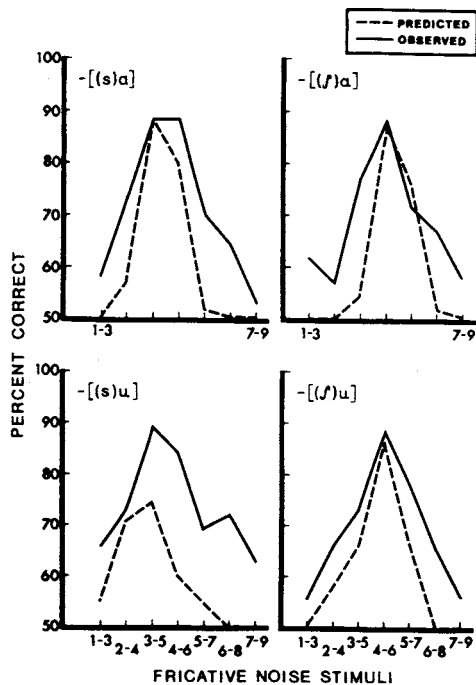


Figure 2. AXB discrimination performance of six "categorical" subjects: Two-step comparisons of [f]-[s] noises in four different vocalic contexts (Experiment 1).

discrimination task might lead subjects to adopt a different strategy, of the kind already evidenced by the two exceptional listeners (and by the author as a pilot subject) in the AXB task. There is little doubt that the high accuracy achieved by these latter subjects reflected a noncategorical, auditory mode of listening.

## EXPERIMENT 2: FIXED-STANDARD AX DISCRIMINATION

### Method

**Subjects.** Ten paid student volunteers participated, seven of whom had previously been subjects in Experiment 1, including the two exceptional listeners. (In addition, a panel of three experienced listeners took the test—see below.)

**Stimuli.** Since the fixed-standard AX task was expected to facilitate discrimination, and since it had to be sufficiently difficult for even the best subjects to produce some errors, a more closely spaced

Table 2  
Fricative Noise Stimuli of Experiment 2  
(Pole Center Frequencies in Hertz)

Stimulus	Pole 1	Pole 2
1	2690	4030
2	2769	4148
3	2850	4269
4	2933	4394
5	3019	4523
6	3108	4655
7	3199	4792

seven-member fricative noise continuum was synthesized. The pole frequencies of the noises are listed in Table 2. The relationship between the two poles was somewhat different in these stimuli than in those of Experiment 1; the present stimuli were more closely related to the continuum used earlier by Mann and Repp (1980), spanning the region of highest ambiguity between [f] and [s]. Only two periodic portions were used, [(f)a] and [(s)u], taken from Experiment 1. Thus, the vowel quality and transition effects were deliberately confounded in this study by choosing the two periodic portions that gave a maximal difference in Experiment 1.

Stimulus 4 on the noise continuum was chosen as the fixed standard. In each stimulus pair, the standard occurred first, followed by a comparison stimulus that could be any of the seven stimuli, with equal probability. Thus, only one-seventh of the stimulus pairs had, in fact, identical noises. There were four different conditions. In two conditions, the standard and the comparison always had the same periodic portion—[(f)a] on one condition and [(s)u] in the other. In the other two conditions, the periodic portions were always different—[(f)a] for the standard and [(s)u] for the comparison in one condition, and the reverse assignment in the other.

Each condition contained 24 repetitions of the 7 possible stimulus pairs, arranged in 6 lists of 28, with ISIs of 500 msec within pairs, 2 sec between pairs, and 10 sec between lists. The first list of 28 served as practice and was not scored; thus, the results are based on 20 responses per pair per subject.

**Procedure.** All four conditions were presented in a single session in counterbalanced order, with the restriction that the condition with equal periodic portions always immediately preceded the condition with the same standard but a different periodic portion in the comparison stimuli. The task was to write down "d" whenever a difference between the noises could be detected, and "s" otherwise. Guessing was discouraged. The subjects were not informed about the true frequency of identical pairs.

### Results and Discussion

For listeners who are successful in this task, "different" responses should show a pronounced minimum for stimulus pairs containing identical noises and a rapid increase as a function of the physical distance of the comparison stimulus from the standard, in both directions. In other words, if listeners operate in an auditory mode, "different" responses plotted as a function of comparison stimulus number should exhibit a V-shaped pattern. Preliminary inspection of the results revealed that, surprisingly, only 2 of 10 listeners showed this pattern. These 2 subjects were precisely those who had performed at the ceiling in the AXB discrimination task (Experiment 1). Therefore, their data were again separated from the rest; they will be considered below.

**Categorical listeners.** Let us first examine the combined results of the other eight subjects, which are plotted in the top panels of Figure 3 ("categorical listeners"). The two conditions with identical periodic portions are on the left; those with different periodic portions are on the right. It can be seen that performance was extremely poor (a horizontal function represents chance performance, implying that "different" responses are equally likely to pairs of identical stimuli and pairs of nonidentical stimuli), decidedly asymmetric—especially when the standard contained the [(f)a] portion—and strongly influenced by the nature of the periodic stimulus portion.

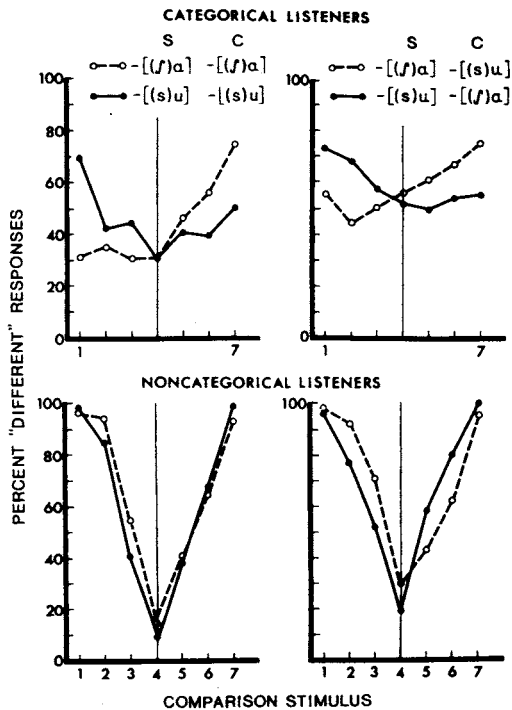


Figure 3. Fixed-standard AX discrimination performance of eight "categorical" subjects (upper panels) and five "noncategorical" subjects (lower panels): Percent "different" responses to pairings of a standard (S, Stimulus 4) with seven comparison (C) stimuli, in four different vocalic context conditions (Experiment 2).

Comparison of the two figure panels suggests that it was primarily the periodic portion of the standard, rather than that of the comparison stimulus, that determined the shape of the response function; this effect (the standard-periodic-portion by stimulus number interaction) was highly significant [ $F(6,42) = 7.5$ ,  $p < .001$ ]. There were more "different" responses when the periodic portions in a pair were different than when they were the same, but this difference fell short of significance [ $F(1,7) = 5.0$ ,  $p < .10$ ].

How is this pattern of responses to be interpreted? The most obvious explanation is that these subjects remained in a phonetic mode, despite instructions to focus on the noise and despite a fixed-standard paradigm. What would the categorical predictions look like in this paradigm? A difficulty arises here, because no identification data were collected for the stimuli used in this experiment.<sup>4</sup> However, the pattern of the data can be rationalized as follows: When the standard stimulus included the [ʃ]a portion, it was categorized as "sh"; consequently, it was difficult to discriminate from more [ʃ]-like stimuli (1-3), but discrimination from more [s]-like stimuli (5-7) improved with their physical distance from the standard, because they approached or crossed the phonetic category boundary. Conversely, when the standard contained [s]u, it was categorized as "s"; con-

sequently, discrimination from more [s]-like stimuli (5-7) was poor, but discrimination from more [ʃ]-like stimuli (1-3) improved with their distance from the standard (cf. upper left-hand panel of Figure 3).<sup>5</sup> The somewhat higher incidence of "different" responses in the two conditions with unequal periodic portions may reflect a response bias induced by the relationship between the irrelevant stimulus portions. This description fits the data sufficiently well to lend plausibility to the claim that this group of subjects remained essentially in a categorical (phonetic) mode of perception even in the fixed-standard AX task.

**Noncategorical listeners.** Consider now the results of the other subjects. As mentioned above, two subjects performed much better than the rest. Their data were augmented by those of three experienced listeners—the author and two colleagues, both of whom are involved in related research on fricative perception. The average results of all five subjects are shown in the bottom panels of Figure 3 ("noncategorical listeners"). Here, we see the expected V-shaped pattern: "Different" responses were least frequent when the standard was paired with itself, and they increased with the physical distance of the comparison stimulus from the standard, with nearly perfect performance when the difference was three steps. This effect of step size was highly significant in an analysis of variance on physically different pairs only [ $F(2,8) = 47.9$ ,  $p < .001$ ]. The figure suggests two further effects: an increase in "different" responses when the periodic portions were different and a shift of the [ʃ]a/[s]u discrimination function relative to the [s]u/[ʃ]a function (lower right-hand panel). Due to the small number of subjects, neither effect reached conventional levels of significance. The first tendency, which was also exhibited by the categorical subjects, may reflect a response bias.<sup>6</sup> The second effect, even though it was not significant, is of interest because its direction is *opposite* to that exhibited by the categorical subjects (upper right-hand panel). Thus, while it is uncertain whether any of the noncategorical listeners were affected by vocalic context, it is clear that they were *not* affected in the same way as the categorical subjects.

**Two listening strategies.** Obviously, the noncategorical subjects used a listening strategy different from that of the categorical subjects. That strategy was the one demanded by the instructions—namely, to focus attention on the auditory (essentially pitch-like) quality of the noise portion. Introspections and comments of the experienced listeners suggested that this strategy may entail a perceptual segregation of the noise portion from the periodic portion—a phenomenon related to auditory streaming (Bregman, 1978; Cole & Scott, 1973). Whether or not phonetic categorization is bypassed in the process, either deliberately or because the noise segregation prevents it, is not known. The author's experience as a listener suggests that some effort and attention are required

to maintain a nonphonetic listening mode; however, another experienced listener commented that she easily and naturally segregated the noise portions. (The same listener shows large effects of vocalic context in an identification task; thus, she is able to integrate information from the two stimulus portions just as easily when it is appropriate to the task.)

That a nonphonetic strategy requires effort and, perhaps, some experience is also suggested by the performance of the categorical listeners. These subjects, even though they had been carefully instructed that subtle differences would occur in the noise portion alone, were apparently not able to follow the instructions effectively. It is a moot point whether an inferior ability to make fine auditory discriminations forced these listeners to remain in a phonetic mode or whether their ability to focus attention on auditory properties of speech stimuli was less developed. However, the second possibility is more plausible. After all, conscious access to auditory qualities of speech, particularly of those relatively brief segments that support phonetic perception, is rarely required of the ordinary speaker/hearer and has traditionally been the exclusive domain of phoneticians and speech scientists. Therefore, it should not be surprising that most naive listeners are not immediately able to perform this feat and instead show a strong tendency to persist in their habitual mode of phonetic perception. If their categorical behavior, especially in the fixed-standard AX task, was nevertheless a bit unexpected, it was only because the fricative-vowel stimuli seemed to offer a relatively easy opportunity to gain access to auditory stimulus properties. The noise portion was steady-state and lasted 200 msec; no training is required for accurate detection of spectral differences when the portion occurs in isolation (Healy & Repp, in press). Presumably, little training would be necessary to transform the categorical listeners of the present study into noncategorical listeners, in contrast with the considerable training that is necessary for subjects to be able to discriminate fine differences in voice onset times or formant transitions of stop consonants (cf. Carney et al., 1977; Edman, 1979). In fact, the ability to focus attention on the noise portion of fricative-vowel stimuli might be discovered rather than learned, as suggested by the extremely accurate performance of two naive listeners. (One of them actually outperformed the three expert listeners.) This question was addressed in the following experiment.

### EXPERIMENT 3: WHAT DOES IT TAKE TO BECOME A NONCATEGORICAL LISTENER?

In two separate but related studies (Experiments 3a and 3b), it was attempted to teach subjects to re-

linquish the phonetic mode of perception and to focus on the auditory properties of the fricative noise.

### Experiment 3a

Since it was expected that not much training would be needed for subjects to discover the proper auditory strategy, the first study examined whether it would be sufficient for subjects to hear and discriminate the fricative noise stimuli in isolation.

### Method

Ten new volunteer subjects were recruited, some of whom had previously participated in experiments involving similar stimuli (but not in the present Experiments 1 and 2). The stimulus tapes were the same as in Experiment 2, with the addition of a new tape containing isolated noise stimuli arranged in the same fixed-standard AX format. To facilitate the task of learning to attend to the noise portions, all subjects listened first to the two conditions (order counterbalanced across subjects) in which standard and comparison noise stimuli were followed by the same periodic portions. Subsequently, they listened to the isolated noises. Finally, the two conditions in which different periodic portions followed the standard and comparison noises were presented (order counterbalanced across subjects), to test whether anything had been learned from discriminating the noises in isolation. All tapes were presented in a single session.

### Results and Discussion

The results are displayed in Figure 4. In the left-hand panel, the functions for the first two conditions replicate the pattern found for the categorical listeners in Experiment 2. In fact, all 10 subjects in the present study fit that pattern; there were no noncategorical listeners. In the right-hand panel of Figure 4, we see that the subjects did rather well with the isolated noises; clearly, these stimuli were not categorically perceived. Despite this success, however, all subjects reverted to a phonetic mode of perception in the remaining two conditions, whose pattern of results again resembles that found in Experiment 2 for categorical listeners.

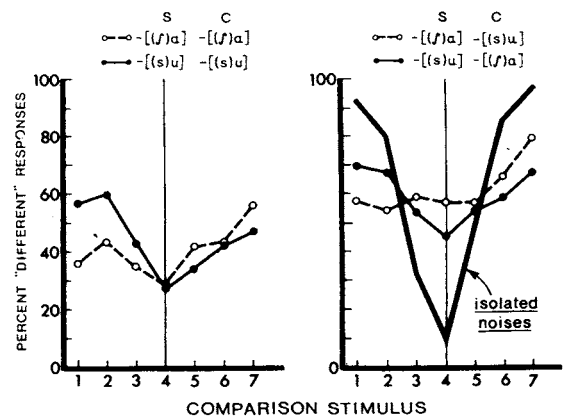


Figure 4. Fixed-standard AX discrimination before (left-hand panel) and after (right-hand panel) discrimination of isolated noises (Experiment 3a).

The statistical analysis of the four vocalic-context conditions confirmed that, as in Experiment 2, the periodic portion of the standard stimulus had a significant effect on the shape of the discrimination function [ $F(6,54) = 4.7$ ,  $p < .001$ ], and that there were more "different" responses to pairs of stimuli differing in their periodic portions than to pairs that had the periodic portion in common [ $F(1,9) = 13.0$ ,  $p < .01$ ]. The latter effect was confounded with practice and may reflect some slight improvement in the course of the experiment, in addition to a response bias. Clearly, however, the subjects did not become noncategorical listeners.

### Experiment 3b

The subjects in Experiment 3a were not able to transfer the discriminatory skill exhibited with isolated noises to the same noises in vocalic context. This suggests that a better awareness of the auditory dimension on which the noises differ is not sufficient to accomplish the task. What may be required, in addition, is the ability to segregate the noise from the following periodic portion and thereby to escape from the phonetic mode of perception. The present study tried out one of several possible methods that might teach listeners this skill.

### Method

Seven of the 10 subjects in Experiment 3a returned for this experiment. In addition, two new volunteers participated. All subjects listened first to a training tape. On this tape, two of the previous conditions were interleaved: On each trial, a pair of isolated noises was followed, after 2-sec, by a pair of exactly the same noises in the [(s)u] context. The subjects were instructed to listen to the isolated noises, to determine the nature of the difference (if any), and then to verify for themselves that exactly the same difference existed between the noises in the syllables. During the first block of 28 trials, the subjects looked at an answer sheet that specified exactly which noise stimuli occurred on each trial. (The nature and arrangement of stimuli were first explained in detail.) During the remaining five blocks, the subjects responded "s" or "d" after listening to both pairs on each trial. They were urged to continue to compare the noise differences in the two pairs.

Following this training condition, the subjects listened to three of the tapes used earlier, in a fixed order. On the first tape, both noise stimuli were followed by [(s)u]; thus, this condition was identical with the training condition, except that pairs of syllables were no longer preceded by pairs of isolated noises. Next, subjects listened to the condition in which the standard noise was followed by [(s)u] and the comparison noise by [(f)α], and finally to the condition in which the standard noise was followed by [(f)α] and the comparison noise by [(s)u].

### Results and Discussion

Preliminary inspection of the results indicated that the new training method was quite successful, but three subjects seemed to benefit much less than the other six (who included the two newcomers). Therefore, the results are displayed separately for "poor" and "good" subjects in Figure 5.

It is evident that the three poor subjects had some

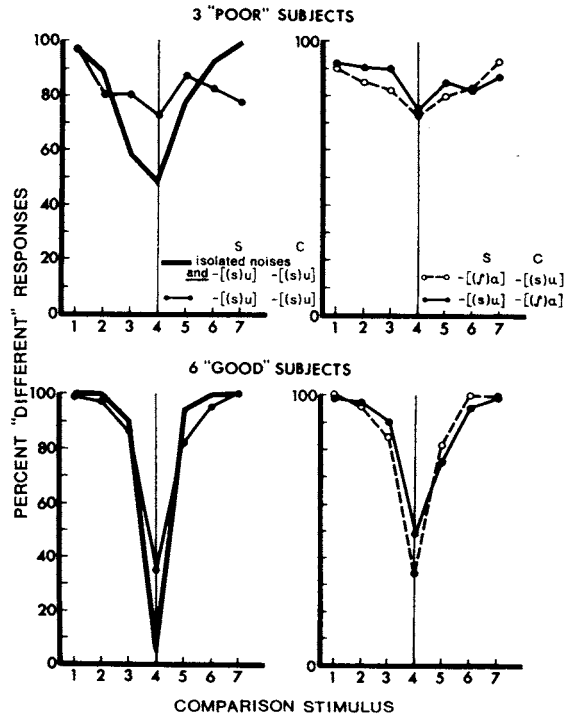


Figure 5. Fixed-standard AX discrimination performance of three "poor" and six "good" subjects after listening to a training tape (Experiment 3b). The heavy lines in the left-hand panels represent the results of the training condition.

trouble in the training task; in particular, they missed out on identical pairs, producing almost 50% false alarms (i.e., incorrect "different" responses). Their performance in the three subsequent tests was extremely poor, due to the even higher false-alarm rates (over 70%). However, there were no clear effects of vocalic context. The high false-alarm rates were almost certainly due to the subjects' knowledge (from the training task and from the preceding instructions) of the true proportion of "same" trials (viz., only one of seven). However, they also indicate that these subjects found it more difficult than the others to discriminate isolated noises. For this reason, they benefited less from the training task that served its purpose only to the extent that the differences between isolated noises could be detected. On the other hand, the apparent absence of vocalic context effects suggests that, rather than persisting in a phonetic mode, these subjects perhaps did learn to segregate the noise portions from their vocalic contexts but then could not easily detect the spectral differences between them (or, rather, their spectral identity). In other words, the epithet "poor," rather than "categorical," seems to be appropriate for these subjects in this experiment.

The six good subjects, on the other hand, were obviously very accurate in the training task and bene-

fited from that experience. Although their false-alarm rates in the vocalic-context conditions were higher than those of the noncategorical subjects in Experiment 2 (presumably, because the present subjects knew about the infrequent occurrence of "same" trials), the discrimination functions were V-shaped and clearly different from those of the categorical subjects in previous experiments. (Note that four of the six good subjects had participated and produced categorical results in Experiment 3a.) In fact, when the average scores were converted into  $d'$  values, they were slightly higher than those of the noncategorical subjects in Experiment 2 (who included the author and two other investigators), indicating remarkable success in the task. There was no significant effect of vocalic context in the two conditions with unequal periodic portions (lower right-hand panel of Figure 5) [ $F(6,30) = 2.1$ ]. Similarly, there was no significant effect of vocalic context in a combined analysis of all nine subjects in the present experiment [ $F(6,48) = 1.5$ ].<sup>8</sup>

In summary, the training task in this experiment was effective and sufficient to convert most listeners from a categorical (phonetic) to a noncategorical (nonphonetic or auditory) mode of listening. The fact that the majority of subjects became as accurate as experienced listeners after only 25 min of self-training is consistent with the suggestion, made above, that the skill of segregating and discriminating fricative noises in vocalic context is discovered rather than slowly learned; that is, it reflects a qualitative change in perceptual processing.<sup>9</sup> Quantitative improvements, such as might occur with further practice, are contingent on that change.

## GENERAL DISCUSSION

### Summary of Results

The present study has four major results:

(1) Fricative identification is influenced by the periodic portion following the fricative noise, even when this portion is held constant over a block of trials. There are independent effects of formant transitions and vowel quality. This replicates the earlier findings of Mann and Repp (1980) and Whalen (1981). The effects were smaller here than in these earlier studies, but this reduction in size may have been due to the use of an improved [ʃ]-[s] continuum as well as to the blocked presentation of stimuli.

(2) Most naive subjects perceive fricative-vowel stimuli rather categorically, and they do so even in a fixed-standard AX task, which was thought to provide a better opportunity for making auditory judgments. This result confirms the earlier findings of Fujisaki and Kawashima (Note 1) and of May (1979).

(3) After appropriate training, most listeners are able to discriminate differences in fricative noise

spectrum accurately and without being influenced by the following periodic portion. (We may disregard an apparent bias to respond "different" when the irrelevant portions of the stimuli in a pair were different, which was perhaps shared by all listeners.) It is important to note that noncategorical listeners were not distinguished from categorical listeners in an identification task; both kinds of subjects showed the expected shifts in labeling functions contingent on vowel quality and formant transitions.<sup>10</sup>

(4) To turn a categorical listener into a noncategorical listener, some practice in segregating the fricative noise from its vocalic context must be provided. Practice with discriminating the noises in isolation is not sufficient.

### Level of Vocalic Context Effects

The fact that noncategorical listeners were not significantly influenced by vocalic context indicates that effects of such context on fricative perception occur at a level that is sensitive to a listener's strategies. Since relatively low-level auditory phenomena—such as auditory masking or contrast—would seem less likely to depend on listening strategies, it appears that the effects of vocalic context are not of this class. However, it may be argued that too little is known about the influence of subjective perceptual organization on auditory interference and contrast, and that the differences between the present two groups of subjects may have resulted from different auditory strategies. For example, different subjects may have centered their attention on different parts of the signal: The noncategorical subjects may have paid attention to the onset of the fricative noise, in which auditory interactions with the periodic portion were absent, whereas the categorical subjects may have focused on the offset of the fricative noise, in which it adjoins the periodic portion and is most susceptible to auditory interference. However, there is some evidence against this hypothesis.<sup>11</sup> Alternatively, a distinction may be made between holistic and analytic auditory processing, with resulting integration vs. segregation of fricative noise and periodic portion. This distinction is useful and valid, but it does not explain by itself why and how vocalic context affects fricative perception in the way it does. In fact, it is fair to say that no convincing auditory explanation for the effects of vocalic context on fricative identification has yet been proposed.

On the other hand, there are numerous studies in the literature that suggest a phonetic origin for various contextual effects in speech perception (e.g., Bailey & Summerfield, 1980; Fitch, Halwes, Erickson, & Liberman, 1980; Mann, 1980; Mann & Repp, 1980, 1981; Repp, Liberman, Eccardt, & Pesetsky, 1978). Several other studies provide direct support for the existence of two distinct modes of processing



speechlike stimuli—one auditory, the other phonetic (e.g., Remez, Rubin, Pisoni, & Carrell, 1981; Bailey, Summerfield, & Dorman, Note 3; Grunke & Pisoni, Note 4). Strong evidence on both counts comes from a recent study by Best, Morriongiello, and Robson (in press), who showed one type of cue integration (viz., integration of silence and formant transitions as cues to stop manner) to be specific to a phonetic mode of perception. Methodologically, the present study is complementary to that of Best et al.: Whereas they showed that certain (speechlike) nonspeech stimuli can be perceived either in an auditory or in a phonetic mode, the present experiments showed that the same is true for certain speech stimuli. In each case, the contextual or cue-integration effect of interest was observed only when listeners responded to phonetic, rather than auditory, properties of the stimuli. In fact, the present experiments demonstrated this for two effects at once—one due to the formant transitions (a cue-integration effect) and one due to vowel quality (a true context effect).

### Conclusions

The present data provide support for the hypothesis that effects of vocalic context on fricative identification are tied to a phonetic mode of perception. They suggest strongly that there are two different strategies of listening to fricative-vowel syllables—one auditory (noncategorical) and the other phonetic (categorical). Regular vocalic context effects occur only in the phonetic mode, presumably because they are mediated by the listener's implicit knowledge of articulation and coarticulation and/or their acoustic consequences. To bring this knowledge to bear on some auditory input is tantamount to being in a phonetic mode of perception: We perceive speech in terms of what our brain knows about it. Similarly, we perceive nonlinguistic auditory attributes with reference to what we know about nonspeech sounds. The frame of reference adopted for a particular stimulus is a joint function of stimulus structure and listener strategy.

The phonetic and auditory modes are available, in principle, for any speechlike stimulus. They may even be used simultaneously. However, since the phonetic mode is the natural way of dealing with speech, and since the auditory properties of speech are often unfamiliar and require the listener to pay attention to fine detail, special laboratory tasks including varying amounts of training may be necessary to elicit the auditory listening strategy. Fricative-vowel syllables differ from, say, stop-consonant-vowel syllables in that some of their auditory properties (e.g., the "pitch" of the fricative noise) are easier to access and discriminate (as compared with, e.g., the "pitch" of formant onsets or the duration of aspiration, whose successful discrimination requires

rather extensive practice—cf. Carney et al., 1977; Edman, 1979). The relative accessibility of an auditory property is largely governed by stimulus factors: Auditory judgments of the pitch of fricative noises are easiest to make when the noises occur in isolation, more difficult in fricative-vowel syllables, and probably even more difficult when the fricatives occur intervocally or are embedded in fluent speech. (The fact that the fricative noises in the present studies were synthetic may also have been a facilitating circumstance.) Task factors, such as interstimulus intervals and stimulus uncertainty, naturally play a role, too. In principle, any auditory property of speech can be detected and discriminated within the limits set by the auditory system, but listeners may have to learn how to gain access to the relevant property. They may have to reorganize their percept in the process (e.g., "segregate" the noise portion from the following vocalic portion), which involves perceptual skills that need to be acquired or elicited by appropriate instructions. The present studies illustrate this point. They also reaffirm the importance of the distinction between auditory and phonetic modes of perception, and they demonstrate that certain contextual effects are specific to the phonetic mode.

### REFERENCE NOTES

1. Kunisaki, O., & Fujisaki, H. On the influence of context upon perception of voiceless fricative consonants. *Annual Bulletin of the Research Institute for Logopedics and Phoniatrics* (University of Tokyo), 1977, 11, 85-91.
2. Fujisaki, H., & Kawashima, T. On the modes and mechanisms of speech perception. *Annual Report of the Engineering Research Institute* (University of Tokyo), 1969, 28, 67-73.
3. Bailey, P. J., Summerfield, Q., & Dorman, M. *On the identification of sine-wave analogues of certain speech sounds* (Status Report on Speech Research, SR-51/52, 1-25). New Haven, Conn: Haskins Laboratories, 1977.
4. Grunke, M. E., & Pisoni, D. B. Some experiments on perceptual learning of mirror-image acoustic patterns. *Research on Speech Perception* (Department of Psychology, Indiana University), 1979, 5, 147-182.
5. Repp, B. H. *Influence of vocalic context on perception of the [j]-[s] distinction: V. Two ways of avoiding it* (Status Report on Speech Research, SR-66). New Haven, Conn: Haskins Laboratories, 1981, in press.
6. Whalen, D. H. *Effects of non-essential cues on the perception of English [s] and [ʃ]*. Paper presented at the 101st Meeting of the Acoustical Society of America, Ottawa, Ontario, May 1981.

### REFERENCES

- BAILEY, P. J., & SUMMERFIELD, Q. Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, 1980, 6, 536-563.
- BEST, C. T., MORRIONGIELLO, B., & ROBSON, R. Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics*, 1981, 29, 191-211.
- BLUMSTEIN, S. E., & STEVENS, K. N. Perceptual invariance and onset spectra for stop consonants in different vowel environments. *Journal of the Acoustical Society of America*, 1980, 67, 648-662.

- BREGMAN, A. S. The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII*. Hillsdale, N.J.: Erlbaum, 1978.
- CARNEY, A. E., WIDIN, G. P., & VIEMEISTER, N. F. Non-categorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, 1977, **62**, 961-970.
- COLE, R. A., & SCOTT, B. Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, 1973, **27**, 441-449.
- CREELMAN, C. D., & MACMILLAN, N. A. Auditory phase and frequency discrimination: A comparison of nine procedures. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 146-156.
- DIVENYI, P. L. Some psychoacoustic factors in phonetic analysis. *Proceedings of the Ninth International Congress of Phonetic Sciences* (Vol. 2). Copenhagen: University of Copenhagen, 1979.
- EDMAN, T. R. Discrimination of intraphonemic differences along two place of articulation continua. In J. J. Wolf & D. H. Klatt (Eds.), *Speech communication papers*. New York: Acoustical Society of America, 1979.
- FITCH, H. L., HALWES, T., ERICKSON, D. M., & LIBERMAN, A. M. Perceptual equivalence of two acoustic cues for stop-consonant manner. *Perception & Psychophysics*, 1980, **27**, 343-350.
- HARRIS, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, **1**, 1-7.
- HASEGAWA, A. *Some perceptual consequences of fricative co-articulation*. Unpublished doctoral dissertation, Purdue University, 1976.
- HEALY, A. F., & REPP, B. H. Context sensitivity and phonetic mediation in categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, in press.
- KUHL, P. K., & MILLER, J. D. Speech perception by the chinchilla: Identification functions for synthetic VOT stimuli. *Journal of the Acoustical Society of America*, 1978, **63**, 905-917.
- LANE, H. The motor theory of speech perception: A critical review. *Psychological Review*, 1965, **72**, 275-309.
- LIBERMAN, A. M., HARRIS, K. S., HOFFMAN, H. S., & GRIFFITH, B. C. The discrimination of speech sounds within and across phoneme boundaries. *Journal of Experimental Psychology*, 1957, **54**, 358-368.
- MANN, V. A. Influence of preceding liquid on stop consonant perception. *Perception & Psychophysics*, 1980, **28**, 407-412.
- MANN, V. A., & REPP, B. H. Influence of vocalic context on perception of the [ʃ]-[s] distinction. *Perception & Psychophysics*, 1980, **28**, 213-228.
- MANN, V. A., & REPP, B. H. Influence of preceding fricative on stop consonant perception. *Journal of the Acoustical Society of America*, 1981, **69**, 548-558.
- MAY, J. G. *The perception of Egyptian Arabic fricatives*. Unpublished doctoral dissertation, University of Connecticut, 1979.
- POLLACK, I., & PISONI, D. B. On the comparison between identification and discrimination tests in speech perception. *Psychonomic Science*, 1971, **24**, 299-300.
- REMEZ, R. E., RUBIN, P. E., PISONI, D. B., & CARRELL, T. D. Speech perception without traditional speech cues. *Science*, 1981, **212**, 947-950.
- REPP, B. H., HEALY, A. F., & CROWDER, R. G. Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception and Performance*, 1979, **5**, 129-145.
- REPP, B. H., LIBERMAN, A. M., ECCARDT, T., & PESETSKY, D. Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance*, 1978, **4**, 621-637.
- SCHOUTEN, M. E. H. The case against a speech mode of perception. *Acta Psychologica*, 1980, **44**, 71-98.
- STEVENS, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. London: Academic Press, 1975.
- WHALEN, D. H. Effects of vocalic formant transitions and vowel quality on the English [s]-[ʃ] boundary. *Journal of the Acoustical Society of America*, 1981, **69**, 275-282.

## NOTES

1. Only three previous studies seem to have used fricatives in a categorical-perception paradigm, and none of them has been fully published. Fujisaki and Kawashima (Note 2) found better-than-chance within-category discrimination of stimuli from a [ʃe]-[se] continuum, but there was a marked peak in the discrimination function at the category boundary. The listeners in this Japanese study perceived fricative-vowel syllables only slightly less categorically than stop-consonant-vowel syllables. This result was replicated with Egyptian listeners by May (1979), who used an [əʃə]-[əsə] continuum. Hasegawa (1976) presented American listeners with a synthetic [ʃ]-[s] noise continuum in two different vocalic contexts, [i-] and [u-]. After demonstrating a shift in fricative labeling contingent on the preceding vowel, he found only rather weak evidence for discrimination peaks in the vicinity of the category boundary. Discrimination performance within phonetic categories was quite good, leading to the conclusion that the stimuli were not categorically perceived. However, the listeners in that study had had some practice in the task; note also that, in contrast with the other studies, the fricatives were in syllable-final position, which, because of the silence following the critical signal portion, may have enhanced auditory memory and thus facilitated discrimination.

2. In that earlier study (Experiment 4 of Mann & Repp, 1980; see their Figure 4), the boundaries varied over as many as five steps on a nine-step fricative noise continuum. However, the boundary shifts are not directly comparable, because the earlier continuum transversed the acoustic stimulus space on a path different from that of the present continuum. The improvement achieved in the present stimuli is evident also in an increased steepness of the labeling functions.

3. These authors demonstrated, following a suggestion by Lane (1965), that predictions derived from an independent single-item identification test underestimate discrimination performance even if discrimination is based exclusively on phonetic labels. This is so because listeners tend to assign different labels to stimuli that occur in close succession (as they do in a discrimination task); therefore, discrimination performance based on these labels will be superior to that predicted from a single-item identification test. Since it was not the purpose of the present study to determine exactly how categorical fricative perception is, the tedious task of having subjects label the stimuli in AXB triads was omitted.

4. An appropriate identification test was not included, because Mann and Repp (1980, Experiment 4) had used similar stimuli previously. However, it turned out that the effects of vocalic context in that earlier identification test were much too large to generate good predictions of the present data. It is quite plausible, however, that the smaller stimulus range used here, together with the particular format of presentation, reduced the magnitude of context effects.

5. In the two conditions in which standard and comparison stimuli have different periodic portions, the minimum percentage of "different" responses might be expected to shift away from the center (Stimulus 4) if discrimination is based on phonetic labels: The noise portion of a comparison stimulus containing [(s)u] must be more [ʃ]-like (and that of one containing [(ʃ)u] must be more [s]-like) than that of the standard for the two stimuli in a pair to be labeled the same way (cf. upper right-hand panel of Figure 3). However, this shift was not very pronounced and was not significant. It was even less evident in Experiment 3a (see below). The fact that the periodic portion of the standard stimulus had a much stronger effect than that of the comparison stimulus suggests that the subjects' difficulty lay primarily in segregating the standard noise from its vocalic context. This may have occurred because

the standard stimulus had to be held in memory for more than 2 sec, whereas the decision process could be set into motion immediately after hearing the noise portion of the comparison stimulus (cf. Footnote 11).

6. When the data were converted into  $d'$  scores, it became evident that, despite the higher percentage of "different" responses in the conditions with different periodic portions, performance was actually somewhat poorer than in the conditions with identical periodic portions.

7. These three subjects also did not do very well in the isolated-noise condition of Experiment 3a. This seems to rule out the alternative explanation that they tended to base their judgments on the syllables rather than on the isolated noises in the present training condition.

8. There was probably a general effect of vocalic context: Performance was more accurate with isolated noises than with noises in any vocalic context. This difference may be ascribed, in part, to interference between stimulus portions in auditory memory. However, performance in the training condition was also favored by shorter noise-to-noise intervals in pairs of isolated noises (the interstimulus interval was 2 sec in all conditions) and by the opportunity to extract additional information from the following pair of syllables.

9. In further support of this claim, it should be noted that the distribution of accuracy levels across subjects was quite bimodal. Listeners were either very accurate or very poor in discriminating noises in context; there was not a single subject who performed at an intermediate level. (Such a level would be expected only if a listener alternated between the two strategies.) Also, one of the categorical listeners in Experiment 3a apparently switched strategies ("caught on") between the last two conditions, which resulted in a sudden and dramatic improvement in performance.

10. This statement is based on a comparison of the identification data for the five categorical and the two noncategorical lis-

teners who participated in both Experiments 1 and 2. In addition, the three expert listeners of Experiment 2 also exhibited strong context effects in several earlier identification tests.

11. The evidence comes from a supplementary reaction-time experiment (described in more detail as Experiment 1 in Repp, Note 5) in which the fricative noise and vocalic portions of natural tokens of [sa], [ʃa], [su], and [ʃu] were separated and recombined in all possible ways and, in addition, the noise portion was shortened by 50 msec in half the stimuli. Ten subjects (the same ones as in Experiment 3a, all of whom had been categorical listeners in that study) listened to a random sequence of these stimuli and identified the fricative consonant by rapidly pressing one of two keys. If these subjects waited for the onset of the periodic portion before making their decision, they should have shown effects of noise duration as well as of vocalic context, having been slowed down by longer noises and by formant transitions or vowels that were different from the original context of the fricative noise. However, the results showed only a marginal noise duration effect and no effect at all of vocalic context, suggesting that the subjects accumulated information right from the onset of the stimulus and reached a decision before the vocalic context could have its effect. (Note that, in this study, the fricative noises were unambiguous cues to fricative identity; there were few errors.) Obviously, the ability to make a phonetic decision at the earliest time possible does not imply an ability to make auditory judgments about a portion of the stimulus, even though that portion carries the information underlying both judgments. (Whalen, Note 6, independently replicated the finding of no vocalic context effects in a nearly identical task and went on to show that such effects do emerge if subjects are asked to identify the fricative *and* the vowel, which forces them to wait for the periodic stimulus portion.)

(Received for publication September 22, 1980;  
revision accepted May 18, 1981.)