511

Perception & Psychophysics
1985, 37, 579-587

Phonetic information is integrated across intervening nonlinguistic sounds

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

and

ARTHUR G. SAMUEL
Yale University, New Haven, Connecticut

When the fricative noise of a fricative-vowel syllable is replaced by a noise from a different vocalic context, listeners experience delays in identifying both the fricative and the vowel (Whalen, 1984): mismatching the information in the fricative noise for vowel and consonant identity with the information in the vocalic segment appears to hamper processing. This effect was argued to be due to integration of the information relevant to phonetic categorization. The present study was intended to eliminate an alternative explanation based on acoustic discontinuities. Noises and vowels were again cross-spliced, but, in addition, the first 60 msec of the vocalic segment (which comprised the consonant-vowel transitions) either had a nonlinguistic noise added to it or was replaced by that noise. The fricative noise and the majority of the vocalic segment were left intact, and both were quite identifiable. Mismatched consonant information caused delays both for original stimuli and for ones with the noise added to the transitions. Mismatched vowel information caused delays for all stimuli, both originals and ones with the noise. Additionally, syllables with a portion replaced by noise took longer to identify than those that had the noise added to them. When asked explicitly to tell the added versions from the replaced, subjects were unable to do so. The results indicate that listeners integrate all relevant information, even across a nonlinguistic noise. Completely replacing the signal delayed identifications more than did adding the noise to the original signal. This was true despite the fact that the subjects were not aware of any difference.

Phonetic information is spread throughout the acoustic signal. This is true even in the case of fricative-vowel syllables, where it might seem that there are two invariant cues. In such syllables, there are two distinct acoustic segments: a noise that can be identified in isolation as the fricative and a vocalic segment that can independently specify the vowel. Nonetheless, there is vowel information in the fricative noise (Whalen, 1983; Yeni-Komshian & Soli, 1981) and fricative information in the vocalic formant transitions (Harris, 1958; Mann & Repp, 1980; Whalen, 1981). Thus, one of the most promising cases of context-independent phonetic cues turns out to vary contextually.

Evidence from cross-splicing studies also indicates that subjects are sensitive to the information that is spread through the syllable. In a series of reaction time studies (Whalen, 1984), listeners were presented with edited fricative-vowel stimuli containing mismatches between information in the fricative noise and information in the vocalic segment. Listeners were slower to identify both the consonants and the vowels of the syllables with mismatches, suggesting an attempt to integrate that information, even though the information was not necessary to identify the phones. This was true whether the mismatch was between information about place of articulation in the transitions and in the noise or between vowel information in the noise and in the vocalic segment itself. It was also true whether the subjects were identifying the vowel or the fricative.

The present experiments were designed to clarify the interpretation of that work. In particular, there was a possibility that some relatively uninteresting psychoacoustic discontinuity in the previous stimuli accounted for the reaction time data. That is, since the stimuli were (digitally) edited, there could have been abrupt changes in the spectrum at the cut, possibly causing a purely auditory disruption of processing. This possibility was less likely in one experiment (Whalen, 1984, Experiment 5), in which the fricative noise was separated from the vocalic segment by 60 msec of silence (thus distancing the two spliced portions), and yet the delay caused by mismatch-

This research was supported by NICHD Grant HD-01994 to Haskins Laboratories. A portion of this work was presented at the 107th meeting of the Acoustical Society of America, Norfolk, VA, May 1984. We thank Suzanne Boyce for running the subjects for Experiments 1 and 2, and Donna Kat for help with Experiment 3. Michael Studdert-Kennedy, Quentin Summerfield, and an anonymous reviewer provided helpful comments.

D. H. Whalen's mailing address is: Haskins Laboratories, 270 Crown Street, New Haven, CT 06510.

Copyright 1985 Psychonomic Society, Inc.

ing information remained. However, it is conceivable that the inserted silence failed to displace an auditory trace of the fricative noise. If this trace did not match the vocalic segment, subjects could have perceived a discontinuity. Thus, the data do not completely rule out an auditory discontinuity account of the reaction time results.

The present experiments attempted to replicate the slowing effect of mismatches in cases in which it was clear that an auditory discontinuity account could not hold. To that end, the temporal progression of the syllable was left intact (that is, no silence was introduced), but the location of the digital splice coincided with the imposition of a nonlinguistic noise. This noise (either a naturally produced cough or a synthesized buzz) occurred during the vocalic formant transitions, comprising the first 60 msec of the vocalic segment. The first experiment tested a weak version of the auditory distraction theory with mismatches of formant transitions. On this account, any nonlinguistic noise will slow processing, and the added noises will not increase the effect of mismatched transitions. The phonetic integration theory favored by Whalen (1984) would, instead, predict that, as long as the transitions were not masked completely, they would add their own delay in addition to that caused by the noise. The strong version of the auditory theory would predict that as well, since there would be two acoustic discontinuities in the signal. To test the strong version of the acoustic theory, Experiment 2 used the mismatch of vowel information, which persists even when the 60 msec is replaced completely by the noise. The prediction was that if the previously obtained delays were auditory distractions at the boundaries, then the mismatch effects should disappear when the acoustic segments were not contiguous. If, however, listeners were integrating phonetic information, that information would still be available and the effect should persist.

EXPERIMENT 1

Experiment 1 examined a mismatch of information for fricative place of articulation, between the information in the vocalic formant transitions and that in the noise itself. We will call this a mismatch of consonant information, even though the transitions (as the name implies) provide information about both the consonant and the vowel. The nonlinguistic noise (the natural cough or the synthetic buzz) was introduced in one of two ways. For both matched and mismatched versions, the 60 msec of the vocalic segment that constituted the transitions either had the nonlinguistic noise digitally added (the ''added'' stimuli) or were replaced by the nonlinguistic noise (the ''replaced'' stimuli). The added noise was expected to mask the transitions somewhat, presumably reducing the effect of mismatched information if a mere auditory distraction was the cause. However, if the more global, phonetic interpretation was correct, the mismatch should be just as strong when there was noise added to the signal as when the mismatch was the only complicating factor. The replaced stimuli, however, would not have transitions present, and therefore should show no slowing effect of the cross-splicing.

Two different noises were used in order to reduce the possibility of some unexpected acoustic artifact. We wanted syllables to be perceived as interrupted in a way that allowed what might be called ''phonetic'' restoration (after Warren's, 1970, phonemic restoration). That is, listeners should be able to assume that there was a signal behind the noise, even in the replaced simuli. Both noises were primarily aperiodic but with some periodic shaping, a combination most likely to produce phonemic restoration (Samuel, 1981b).

Method

Natural tokens of the syllables [sɑ], [ʃɑ], [su], and [ʃu] were recorded by a male speaker of English. (The speaker was not the same as in Whalen, 1984.) The tokens were digitized (20-kHz sampling rate, 9.6 kHz low-pass filtered), and test items were selected so that: (1) all fricative noises were of the same duration (160 msec); (2) all vocalic segments were of the same duration (340 msec); and (3) each syllable token was used either for its fricative noise or for its vocalic segment; thus every test syllable had a digital splice in it. Two tokens of each category (e.g., the [s] from [sɑ]) were used.

Two different nonlinguistic noises were used (see Figure 1). One was 60 msec of a naturally produced cough and the other was 60 msec of a buzz consisting of a white noise source filtered at 500 Hz, with some energy (approximately 35 dB lower in amplitude) emerging at the harmonics. The buzz was an average of 8 dB more intense than the portion of speech it was added to or replaced; the cough was an average of 5 dB more intense.

Five copies of each digitized syllable were made. One of these (the ''original'') was intact except for the digital splice between the fricative noise and the vocalic segment (as described above). Two ''added'' and two ''replaced'' versions were constructed: In the ''added'' versions, the cough noise or buzz noise was added

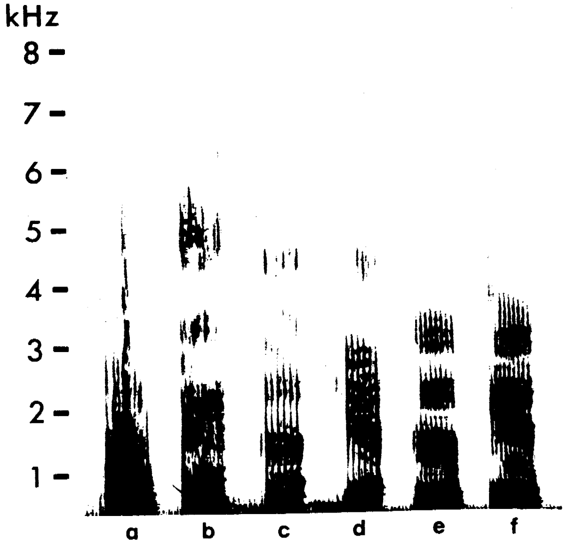

Figure 1. The two nonlinguistic noises and samples of the relevant portions of the speech signal: (a) the 500-Hz buzz, (b) the naturally produced cough, and the transitions from (c) [sɑ], (d) [ʃɑ], (e) [su], and (f) [ʃu].

Table 1
Construction of the Stimuli

Syllable Heard As	Matched (Experiments 1 and 2)		Consonant Mismatch (Experiment 1)		Vowel Mismatch (Experiment 2)	
	Noise	Vocalic	Noise	Vocalic	Noise	Vocalic
sɑ	s[ɑ] +	[s]ɑ	s[ɑ] +	[ʃ]ɑ	s[u] +	[s]ɑ
ʃɑ	ʃ[ɑ] +	[ʃ]ɑ	ʃ[ɑ] +	[s]ɑ	ʃ[u] +	[ʃ]ɑ
su	s[u] +	[s]u	s[u] +	[ʃ]u	s[ɑ] +	[s]u
ʃu	ʃ[u] +	[ʃ]u	ʃ[u] +	[s]u	ʃ[ɑ] +	[ʃ]u

Note—Each column presents the syllables used to construct the stimulus syllables. The portion of each syllable enclosed in brackets was digitally excised.

digitally to the first 60 msec of the vocalic segment. In the "replaced" versions, the first 60 msec of the vocalic segment were completely replaced by the cough or buzz.

For all three types of stimuli ("original," "added," and "replaced"), half of the syllables had vocalic segments matched with the fricative noise (e.g., the [u] from [su] paired with an [s] noise) and half had mismatched ones (e.g., the [u] from [ʃu] paired with an [s] noise). Note that when the nonlinguistic noise replaced the first 60 msec of the vocalic segment, there was very little left to be mismatched. That is, even though the rest of the vocalic segment came from an inappropriate syllable, the transitions were, by design, mostly completed by 60 msec. Thus, there should not have been much of a phonetic mismatch in the "replaced" stimuli. The first column of Table 1 shows the construction of the matched stimuli; the second column shows the construction of the mismatched stimuli. The match/mismatch factor, the five noise conditions (original, and added and replaced for two types of noise) and the two tokens of the four fricative and vowel categories result in 80 stimuli.

In each of two conditions, 400 syllables were presented over headphones; the randomized sequences contained five repetitions of each of the 80 stimuli. The interstimulus interval was 2,500 msec. Subjects were asked, in one condition, to identify the vowel ("a" or "u") as quickly as possible. In the other condition, they were asked to identify the consonant ("s" or "sh") as quickly as possible. The order of these conditions was balanced across subjects, as was the determination of which button was to be pushed by the dominant hand. Responses under 100 msec were counted as mistakes, and the equipment gave up waiting for an answer after 2,500 msec. Missing responses and mistakes in identification accounted for 5.3% of the consonant judgments and 3.8% of the vowel judgments. These trials were not included in the reaction time analyses.

The subjects were 20 Yale students who were paid for their participation; all were native speakers of English with no reported hearing difficulties.

Results and Discussion

Figure 2 shows the reaction times in Experiment 1 for the first two factors of interest, collapsed over the category identified. Overall, mismatches of consonant information, as seen in the left pair of bars, slowed identifications a significant 16 msec [$F(1,19) = 9.97$, $p < .01$]. The presence of noise also slowed reaction times [$F(4,76) = 8.19$, $p < .001$], as shown in the three bars to the right. Adding the noise caused an 8-msec delay, and replacing the noise caused an additional 12-msec delay.

Figure 3 shows the interaction of consonant information mismatch and extraneous noise. In each pair of bars, the open bar shows the mean reaction time to stimuli with matched consonant information. The cross-hatched bar shows the responses to stimuli with mismatched consonant information. The most important result is apparent in the

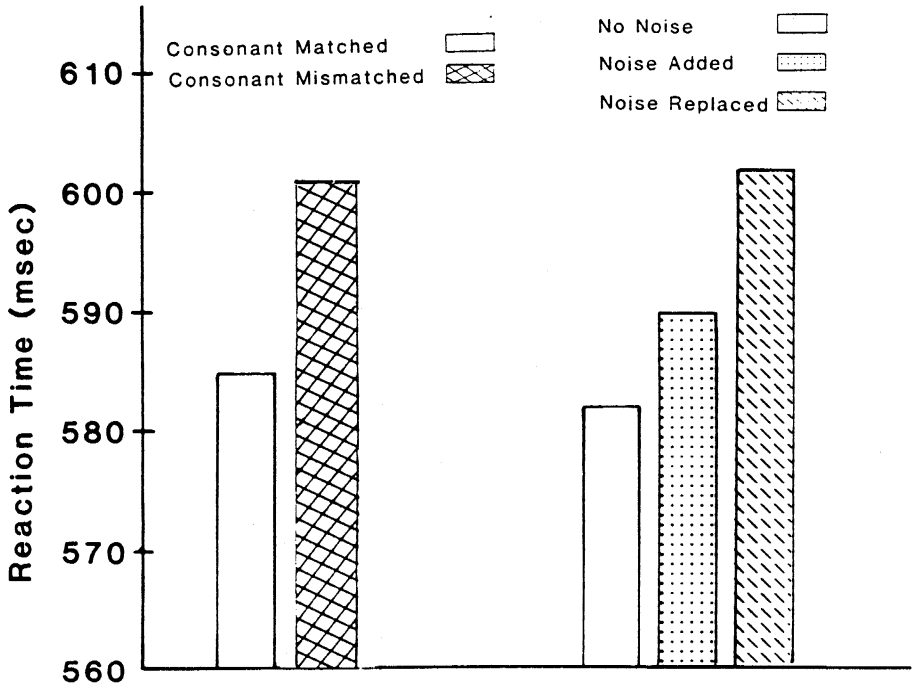

Figure 2. Identification times for stimuli with matched or mismatched consonant information (left pair of bars) and for stimuli with no noise, noise added, or noise replaced (right trio of bars) (Experiment 1).

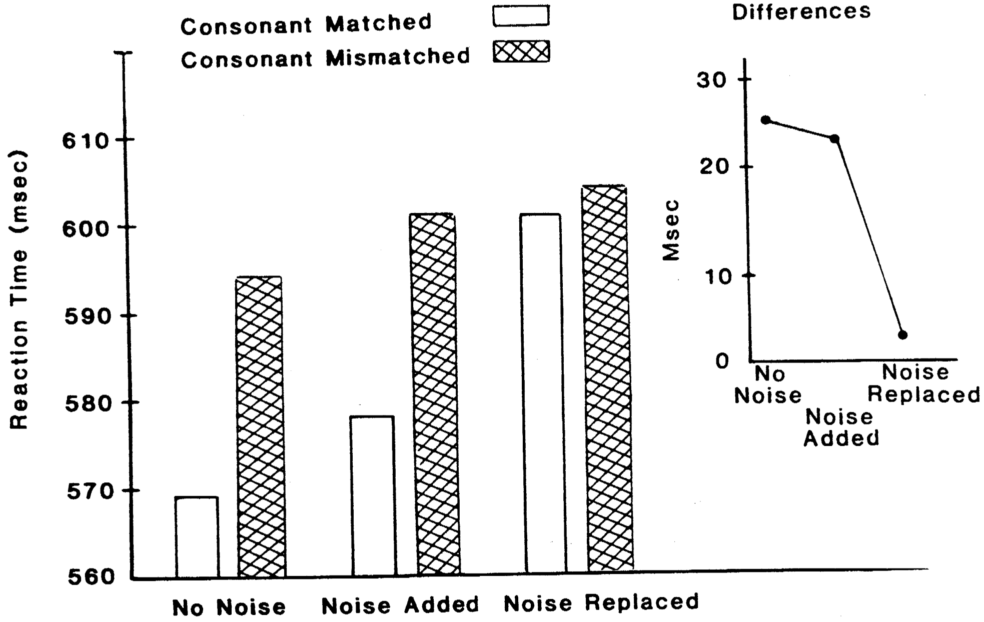

Figure 3. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) consonant information with no noise (leftmost pair), noise added (middle pair), or noise replaced (rightmost pair) (Experiment 1).

middle pair of bars. Even though both matched and mismatched stimuli include acoustic discontinuities (at the boundaries of the nonlinguistic noises), the effect of the mismatch is still robust [F(1,19) = 22.32, p < .001, for just the "added" stimuli]. The comparison of these bars with the two leftmost shows that the addition of the noise slowed judgments an average of 8 msec; the mismatch of transitions added 24 msec whether the noise was present or not.

As can be seen from the rightmost pair of bars, and from the plot of the differences between bars on the right, the difference between matched and mismatched stimuli is negligible in the replaced stimuli (a nonsignificant difference of 2 msec). Not only is there an interaction between added/replaced and match/mismatch [F(1,19) = 12.76, p < .01], but a separate analysis of the replaced data alone shows no effect of mismatch [F(1,19) = 0.32, n.s.]. As predicted, there is not enough transitional information left after 60 msec for the effects of a mismatch to appear.

The effect of mismatch was the same whether the consonant or the vowel was identified [F(1,19) = 1.97, n.s., for the interaction]. Reaction times also did not vary as a function of the type of nonlinguistic noise [F(1,19) = 0.61, n.s.]; nor did type of noise interact with any other factors.

More errors were made on mismatched items than on matched ones [5.3% vs. 3.7%, respectively; F(1,19) = 6.72, p < .05]. The higher error rate goes with the longer reaction time: There is no speed/accuracy tradeoff for these stimuli. There was an interaction of consonant/vowel identification with noise type [F(4,76) = 7.81, p < .001], due mainly to the behavior of the items without noise: They generated more errors than the added and replaced items when the consonant was identified, and fewer when the vowel was identified. A further interaction of these two factors with match/mismatch [F(4,76) = 2.74, p < .05] shows that the difference is largely due to the high error rate on mismatched items without noise when the consonant was identified—precisely the case in which the mismatch would be the most noticeable.

The previously obtained slowing of reaction time with mismatched information (Whalen, 1984) was found even when explicitly nonlinguistic discontinuities were present. The effect on identification was not weakened by any masking of the transitions that might have occurred: The phonetic relevance of the transitions was still perceived. Thus, the noises were not too intense to test the theory. The weak version of the auditory theory is thus disproved, since there are two delays present, and they do not interfere with each other. The strong version of the theory, that there are two auditory discontinuities at work (the transitions and the nonlinguistic noises), is still viable, however. Experiment 2 examines a situation in which this interpretation is not possible.

Note that the identification times for the replaced stimuli are essentially the same as they are for the mismatched added stimuli (see Figure 3): The delay caused by a mismatch is the same as the delay caused by the absence of the original signal. One interpretation of this is that appropriate transitions facilitate identification, and that mismatched transitions are no worse than having no transitions at all. Alternatively, the similarity in mean reaction times might be coincidental. Experiment 2 provided an opportunity to test these alternatives while examining the effect of mismatching vowel information.

EXPERIMENT 2

Experiment 2 mismatched the vowel information in the fricative noises with that of the vocalic segment. Manipulations similar to those of Experiment 1 were carried out, but with a different expectation: Mismatches of phonetic information should show up even in the replaced stimuli. This expectation was based on the fact that the vowel mismatch does not depend just on the first 60 msec of the vocalic segment, but is instead present throughout the fricative noise, on the one hand, and the vocalic segment, on the other.

Method

The syllable pieces of Experiment 1 were again used in Experiment 2, although the combinations used for the mismatched stimuli were different. The matched stimuli were identical (see the first column of Table 1). The mismatched syllables are outlined in the third column of Table 1. The transitions were always appropriate to the fricative, that is, the consonant information was matched. The same five noise conditions of Experiment 1 were used in Experiment 2: no noise, cough or buzz added to the first 60 msec of the vocalic segment, and cough or buzz replacing those 60 msec.

The stimuli were presented as before, with the two conditions of consonant identification and vowel identification, each presented as a separate block. Missing responses and mistakes in identification accounted for 4.7% of the consonant judgments and 3.1% of the vowel judgments. These trials were excluded from further analysis.

The subjects were 20 Yale students who were paid for their participation. All were native speakers of English with no reported hearing difficulties. Half of them had participated in Experiment 1.

Results and Discussion

Figure 4 presents the results of mismatching vowel information and for including noise in the stimuli, collapsed over the category identified. The two bars at the left indicate that mismatching vowel information had a significant slowing effect of 24 msec [$F(1,19) = 46.90$, $p < .001$]. The three bars on the right indicate that adding noise slowed judgments by 29 msec and that replacing part of the syllable with noise slowed judgments by an additional 15 msec [$F(4,76) = 29.73$, $p < .001$]. All three of these categories were significantly different from each other.

Figure 5 shows the results by both match and noise condition. In each pair of bars, the open bar shows the mean reaction time to stimuli with matched vowel information. The cross-hatched bar shows the responses to stimuli with mismatched vowel information. Unlike Experiment 1,

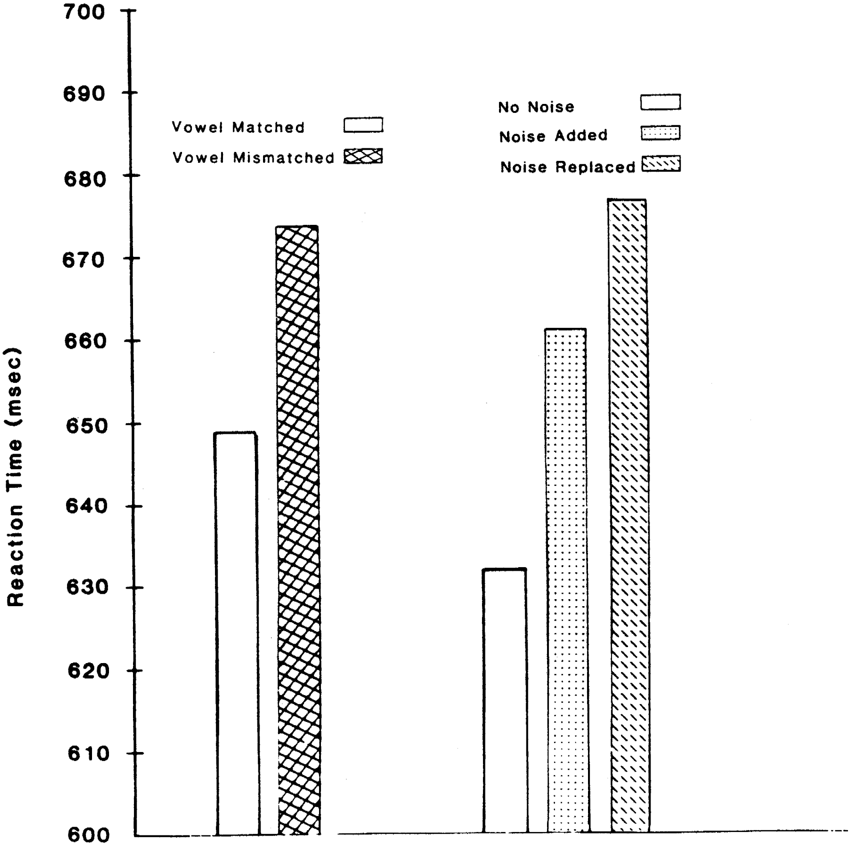

Figure 4. Identification times for stimuli with matched or mismatched vowel information (left pair of bars) and for stimuli with no noise, noise added, or noise replaced (right trio of bars) (Experiment 2).

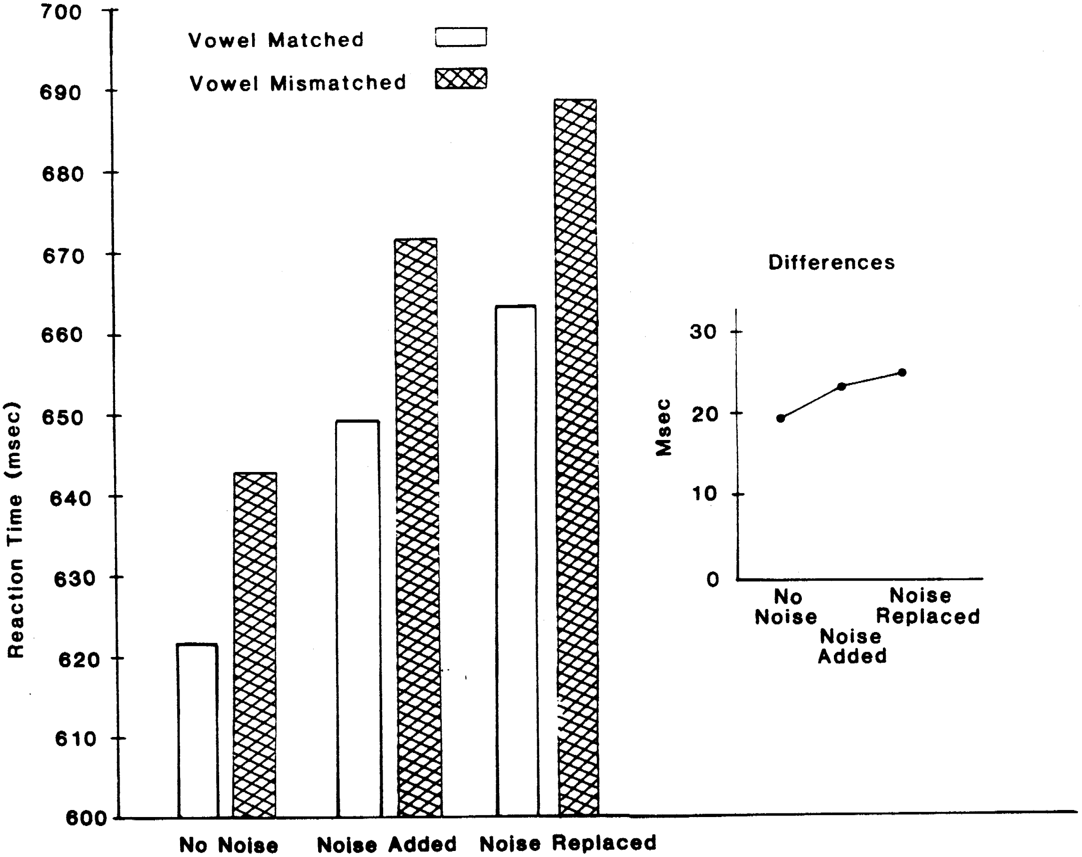

Figure 5. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) vowel information with no noise (leftmost pair), noise added (middle pair), or noise replaced (rightmost pair) (Experiment 2).

vowel mismatches caused delays in each case; the effect of mismatches did not differ across these conditions [F(4,76) = 0.27, n.s.]. If anything, these delays increased with the presence of noise, as is shown by the plot on the right. This plot shows the differences between the matched and mismatched stimuli for the no-noise, noise-added, and noise-replaced stimuli, respectively, from left to right.

There was one interaction between the match/mismatch factor and the category identified (consonant or vowel). The mismatch effect was approximately twice as large when the vowel was identified (15 msec for consonant identification, 31 msec for vowel identification) [F(1,19) = 6.78, p < .05]. A separate analysis of the consonant identification data alone shows that the effect of mismatch was still significant [F(1,19) = 9.57, p < .01].

The main effect of noise type was not significant [F(1,19) = 1.12, n.s.], nor did it enter into any significant interactions.

As in Experiment 1, more errors were made on syllables with mismatches [4.9% vs. 2.9%; F(1,19) = 13.32, p < .01]. The effect was more pronounced when the consonant was identified [a difference of 3.6% vs. 0.4% when the vowel was identified; F(1,19) = 8.55, p < .01], despite the fact that the information mismatched was labeled "vowel" information. However, as mentioned before, it takes two phonetic segments for a mismatch to occur, and the greater susceptibility of the consonant judgments is probably due to the greater closeness of [s] and [ʃ] than of [ɑ] and [u].

As in Whalen (1984), mismatching the rather weak vowel information in the fricative noise with the more powerful information in the vocalic segment slowed phonetic judgments. Even though a nonlinguistic noise indicated that the signal had been corrupted, listeners were still affected by mismatches between two temporally separated portions of the utterance. Experiment 2 is particularly interesting because the information critical to the mismatch was not removed when the nonlinguistic noise replaced the transitions (as it was in Experiment 1). The "replaced" stimuli in Experiment 2 demonstrated that even when *all* tokens include significant acoustic discontinuities, the disruption due to mismatching phonetic information persists: The identification delays are due to an impairment of the processes that deal with the relevant information, not to any simple distractions caused by auditory discontinuities. Thus, the strong version of the acoustic theory is disproved as well.

One difference between the two experiments is the absolute amount of time it took for the phonetic decisions.

Subjects were, on the whole, 68 msec slower in Experiment 2 than in Experiment 1. An analysis (with the factors used before plus the factor of experiment) of the 10 subjects who participated in both experiments shows that the difference is a real one; the effect of experiment was reliable [F(1,9) = 8.91, p < .05]. The only interaction of the experiment factor was with mismatch and noise, which was expected, since the effect of mismatches disappeared for the replaced versions in Experiment 1 but not in Experiment 2. In the first experiment, 30% of the stimuli had detectable mismatches of phonetic information; in the second, 50% did. This increase of conflicting information probably resulted in more cautious identifications, which slowed down responses.

The fact that the two delaying effects, mismatches of vowel information and the inclusion of the two types of noise, were independent allows us to choose between two explanations proposed for the results of Experiment 1. In that experiment, it seemed either that mismatched transitions slowed identification or that the availability of appropriate information speeded identification. The similarity of identification times for syllables with mismatched information to those for syllables in which the noise replaced the transitions left both possibilities open. As can be seen in Figure 5, mismatched information slowed identifications whether nonlinguistic noise was present or not. These results indicate that both the mismatches and the nonlinguistic noise have an interfering effect on identification times.

EXPERIMENT 3

The first two experiments showed that subjects were sensitive to whether the signal was intact or not: in both, replaced stimuli produced significantly slower reaction times than did added stimuli. One possible explanation for this effect is that the replaced items were heard as interrupted or discontinuous and that this distracted the subjects enough to slow them down. A more likely explanation, given that phonetic integration occurs across the noise, is that the perceptual system expects to find the signal even when nonlinguistic noises are present, and that perceptual processing is slowed when this expectation is not met. Experiment 3 tested whether there were noticeable differences between added and replaced stimuli that would support the "distracting" hypothesis. The test involved explicitly asking the subjects to discriminate between added and replaced stimuli. If the subjects were distracted by the replacement of the signal, then added and replaced stimuli should be discriminable.

Method

The stimuli were the "added" and "replaced" items used in the first two experiments. Ninety-six tokens were used in Experiment 3, representing the crossing of four factors: (1) buzz versus cough as extraneous noise, (2) added versus replaced, (3) matched, consonant mismatched, or vowel mismatched, and (4) tokens. The last factor, tokens, represents the eight examples within each cell of the design and includes two instances each of [sɑ], [ʃɑ], [su], and [ʃu].

The stimuli used in Experiments 1 and 2 were recorded on audiotape and digitized on another computer system, using high-quality audio components and a 12-bit A/D converter. The sampling rate was 20 kHz with 9.6-kHz low-pass filtering.

The entire stimulus set of 96 items was presented twice. The first 48 stimuli spanned all of the factors just described except "added versus replaced." The form of each token ("added" vs. "replaced") was selected randomly. The second set of 48 stimuli included the "other" form ("replaced" if the added form of a token had already been presented, and vice versa). The second pass through the 96 stimuli used the same procedure. Each group of 48 tokens was ordered randomly.

Subjects were told that they would be hearing "sa," "sha," "su," and "shu," with some noise present during each syllable. It was explained that the noise would occur "where the consonant meets the vowel," and that the noise would either replace a small bit of the syllable or be superimposed on it. The subjects were instructed to press one button on a computer terminal if they thought the noise replaced part of a syllable and another button if they thought the noise was superimposed.

The presentation of stimuli was subject-paced: Approximately 1 sec after a subject's response was received, the next stimulus was presented. The entire procedure took approximately 15 min.

Twelve individuals served as subjects in Experiment 3. They were recruited through sign-up sheets posted at Yale University, and were paid for their participation. All were native English speakers with no reported hearing problems. Half had previously participated in another study in which they had made similar judgments.

Results and Discussion

The central question of Experiment 3 is whether listeners can discriminate the "added" and "replaced" versions of the syllables. To answer this question, the percentage of correct responses was calculated for each subject, broken down by matching condition (match, consonant mismatch, vowel mismatch), extraneous noise (buzz or cough), and stimulus form ("added" or "replaced"). These percentages were used to calculate the signal detection parameter d′. This bias-free measure of discrimination performance was computed for each of the six cells defined by the crossing of the three matching conditions and the two extraneous noises. These values were submitted to a two-factor analysis of variance (matching condition × extraneous noise).

The results of this analysis can be summarized very simply: Subjects were utterly unable to discriminate "added" and "replaced" stimuli. In signal detection analyses, a d′ of 0 indicates no discriminability, with increasing values reflecting abilities to discriminate. The obtained grand mean d′ was −0.003, indicating that the "added" and "replaced" stimuli could not be discriminated at all. Given this, it should not be surprising that neither extraneous noise type [F(1,11) < 1] nor matching condition [F(2,22) = 2.84, n.s.] made a significant difference; their interaction was similarly inconsequential [F(2,22) = 1.31, n.s.].

What makes this null result of interest is that the "added" and "replaced" stimuli produced significantly different reaction times in Experiments 1 and 2. We thus have a situation in which a manipulation that is totally unavailable to consciousness produces reliable differences in processing time. The extra acoustic discontinuity produced by the replacement manipulation is sufficient to slow down identification of the speech signal (Experiments 1 and 2), but is not discriminable from the mere addition of noise (Experiment 3).

The inability of listeners to discriminate between the "added" and "replaced" items when they are explicitly asked to do so is reminiscent of results obtained in studies of the phonemic restoration effect (Samuel, 1981a). An important difference to note, however, is that in studies of restoration, care is taken to remove all local cues to a phone; if the stretch of speech immediately before or immediately after the replacement locus is played, the relevant phone will not be heard. In the present study, both the fricative and the vowel are perfectly intelligible in isolation; only the transitions are replaced (or have noise added). Thus, there is not enough evidence to tell whether the present results reflect some sort of restoration. A better analogy might be to the classic categorical perception findings (cf. Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967). In those studies, listeners also fail to discriminate between acoustically different tokens (ones within a phonemic category). Moreover, just as in the present study, reaction time analyses of identification times reveal differences between these indiscriminable items (Pisoni & Tash, 1974). The reaction time analyses thus provide insights into the processing of speech that cannot be revealed in overt discrimination tasks.

EXPERIMENT 4

In Experiment 3, listeners proved to be utterly unable to discriminate between two types of stimuli that had produced reliably different reaction times in Experiments 1 and 2. The conditions of Experiment 3 were similar to those of the first two experiments: A single token was presented on each trial, and subjects were required to identify each token as a member of one of two predetermined categories. In Experiment 4, the conditions were modified in order to reveal any discrimination ability that might have gone undetected in Experiment 3. A successive two-alternative forced-choice paradigm was used in which a trial consisted of the presentation of both versions ("added" and "replaced") of a given syllable and the subjects indicated whether the "replaced" version came first or second. If there were any conscious discriminability of the two versions, the subjects should reveal it with this procedure.

Method

The stimuli and equipment were the same as in Experiment 3. The procedure was also the same, except that each token was now paired with its mate. For example, if a trial in Experiment 3 had consisted of a particular /sa/ with cough replacement, the corresponding trial in Experiment 4 consisted of that /sa/ with cough replacement paired with that /sa/ with cough addition. The interstimulus interval was 400 msec.

After the nature of the stimuli was explained to them, the subjects were instructed to indicate whether the "replaced" version came first or second on a given trial by pushing the "1" or the "2" key on a terminal keyboard. After the subject's response, the correct answer ("1" or "2") was displayed on the terminal screen for 500 msec. The feedback was then cleared, and approximately 1 sec later the next pair of tokens was presented. The inclusion of two tokens per trial and the feedback increased the time for the 192 trials to approximately 25 min.

Ten subjects served in Experiment 4. Because we were interested in obtaining an upper bound on discrimination performance, individuals who might be expected to do unusually well were selected: Both authors, a researcher with extensive experience with this type of stimuli, and two graduate students with interests and experience in perceptual research participated. The remainder of the subjects came from the same population as that used in the previous experiments.

Results and Discussion

As in Experiment 3, d' scores were computed for each subject for each of the six cells of the stimulus design (buzz or cough × match, consonant mismatch, or vowel mismatch). These scores were submitted to a two-factor analysis of variance. As in Experiment 3, neither the type of noise [$F(1,9) < 1$] nor the matching condition [$F(2,18) < 1$] made any difference; their interaction was similarly negligible [$F(2,18) < 1$]. However, the change in procedure and subjects was not totally ineffective: The grand mean d' was raised to 0.433, a value that was reliably greater than zero [$F(1,9) = 7.79$, $p < .05$].

What does this result indicate with regard to the discriminability of the two stimulus versions? Apparently, the difference underlying the reaction time effect in Experiments 1 and 2 can be brought to a conscious level some of the time for some of the people. However, it must be noted, first, that these testing conditions were quite different from those of Experiments 1 and 2, and, second, that the observed d's are small—even smaller if one excludes from the mean the first author's score of 1.7. The fact that the only person to achieve a respectable score was the one who had constructed the stimuli and spent many hours listening to them suggests that the analogy previously drawn between the present case and that of categorical perception is apt: In both cases, extensive practice can bring the previously unavailable distinction under conscious control (see Samuel, 1977, for the results of practice on categorical perception). Overall, the results of Experiments 3 and 4 support the position that the subjects in Experiments 1 and 2 were unaware of the difference between "added" and "replaced" stimuli, yet the latter led to reliably slower identification times than the former.

GENERAL DISCUSSION

The phonetic mismatch effects of Whalen (1984) were successfully replicated, using stimuli containing nonlinguistic noises that should have encouraged the perceptual system to block integration of portions of the signal. The present study also shows that having the original signal behind the noise is less disruptive than replacing the signal altogether. This indicates that the perceptual system looks for coherence even within competing noise. The results of this search for coherence are not available to consciousness, as was shown in Experiments 3 and 4.

It appears, then, that listeners are indeed sensitive to all phonetic information given them, and that delays caused by mismatches, even those that cannot be readily heard, are due to increased phonetic processing. This is true even when the relevant portions of the signal are not contiguous; similar effects have been found in vowel-to-vowel coarticulation across stop consonant closures (Martin & Bunnell, 1982). Even when subjects are given every excuse for failing to integrate, as when a nonlinguistic noise occurs in the middle of the signal, they still do integrate. The mismatch adds just as much time to the perceptual process whether the extraneous noise is present or not. This indicates that the previously obtained result is not simply a short-term psychoacoustic disruption but is sustained over a relatively long stretch. Whether the information stored is acoustic or (weakly) categorical remains to be seen.

REFERENCES

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

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

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

MARTIN, J. G., & BUNNELL, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance*, **8**, 473-488.

PISONI, D. B., & TASH, J. (1974). Reaction times to comparisons within and across phonetic categories. *Perception & Psychophysics*, **15**, 285-290.

SAMUEL, A. G. (1977). The effect of discrimination training on speech perception: Noncategorical perception. *Perception & Psychophysics*, **22**, 321-330.

SAMUEL, A. G. (1981a). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General*, **110**, 474-494.

SAMUEL, A. G. (1981b). The role of bottom-up confirmation in the phonemic restoration illusion. *Journal of Experimental Psychology: Human Perception and Performance*, **7**, 1124-1131.

WARREN, R. M. (1970). Perceptual restoration of missing speech sounds. *Science*, **167**, 392-393.

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

WHALEN, D. H. (1983). Vowel information in postvocalic fricative noises. *Language and Speech*, **26**, 91-100.

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

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

(Manuscript received December 10, 1984;
revision accepted for publication June 7, 1985.)