PHONETIC INFORMATION IS INTEGRATED ACROSS INTERVENING NONLINGUISTIC SOUNDS

D. H. Whalen and Arthur G. Samuel

Abstract. 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 phonetic integration of the information relevant to 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 ms 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. Replacing the signal completely delayed identifications more than 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

†Department of Psychology, Yale University.

Acknowledgment. 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. Michael Studdert-Kennedy provided helpful comments.

the most promising cases of context-independent phonetic cues turns out to vary contextually.

There is also evidence from cross-splicing studies, however, that listeners can detect information specifying the original context of the noise and of the vocalic segment. A series of reaction time studies (Whalen, 1984) indicated that subjects are sensitive to all the information in the syllable. In that work, 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, even though the fricative noise was separated from the vocalic segment by 60 ms of silence (thus distorting the two spliced portions), the delay caused by mismatching 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 attempt to replicate the slowing effect of mismatches in cases where it is clear that an auditory discontinuity account cannot 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 ms of the vocalic segment. If the previously obtained delays were mere auditory distractions, then the mismatch effects should disappear—the auditory disturbance at the boundaries of the noise should be the same for syllables with matched and with mismatched fricative noises and vocalic segments. If, however, listeners do in fact integrate information across the whole syllable, then 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 ms of the vocalic segment which 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 is correct, the mismatch should be just as strong when there is noise added to the signal as when the mismatch is the only complicating factor. The replaced stimuli, however, would not have transitions present, and therefore should show no effect of the cross-splicing.

Two different noises were used 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 stimuli. Both noises were primarily aperiodic but with some periodic shaping, a combination most likely to produce phonemic restoration (Samuel, 1981b).

Procedure

Natural tokens of the syllables [sa], [ʃa], [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 ms).
2. All vocalic segments were of the same duration (340 ms).
3. Each syllable token was used either for its fricative noise or for its vocalic segment—thus every test syllable had an electronic splice in it.

Two tokens of each category (e.g., the [s] from [sa]) were used.

Two different nonlinguistic noises were used. One was 60 ms of a naturally produced cough, while the other was 60 ms of a buzz consisting of a semi-periodic filtering of white noise with peaks at intervals of 500 Hz.

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 digitally to the first 60 ms of the vocalic segment. In the "replaced" versions, the first 60 ms 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 ms 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 ms. 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, while the second column shows the construction of the mismatched stimuli. The match/mismatch factor, the five noise conditions (original; added and replaced for two types of noise), and the two tokens of the four fricative and vowel categories result in eighty stimuli.

<table>
<thead>
<tr>
<th>Syllable Heard As:</th>
<th>Matched (Exp 1 &amp; 2)</th>
<th>Consonant Mismatch (Exp 1)</th>
<th>Vowel Mismatch (Exp 2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;sa&quot;</td>
<td>noise voc.</td>
<td>s[α] + [s]α</td>
<td>s[u] + [s]α</td>
</tr>
<tr>
<td>&quot;fa&quot;</td>
<td>[f][α]</td>
<td>[f][α] + [s]α</td>
<td>[f][u] + [s]α</td>
</tr>
<tr>
<td>&quot;su&quot;</td>
<td>s[u] + [s]u</td>
<td>s[u] + [f]u</td>
<td>s[α] + [s]u</td>
</tr>
<tr>
<td>&quot;fu&quot;</td>
<td>[f][u]</td>
<td>[f][u] + [s]u</td>
<td>[f][α] + [s]u</td>
</tr>
</tbody>
</table>

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

In each of two conditions, subjects heard randomized sequences containing five repetitions of each stimulus over headphones. The inter-stimulus interval was 2500 ms. 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 pushed by the dominant hand. Responses under 100 ms were counted as mistakes, and the equipment was forced to give up waiting for an answer after 2500 ms. Missing responses and mistakes in identification accounted for 5.0% 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 native speakers of English with no reported hearing difficulties.

Results and Discussion

Figure 1 shows the reaction times in Experiment 1 for the first two factors of interest. Overall, mismatches of consonant information, as seen in the left pair of bars, slowed identifications a significant 16 ms, \( F(1,19) = 9.97, p < .01 \). The presence of noise also slowed reaction times, \( F(4,76) = 8.19, p < .001 \), as is seen in the three bars to the right. Adding the noise caused an 8 ms delay, and replacing the noise caused an additional 12 ms delay.
Figure 1. 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 2. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) consonant information with no noise (left-most pair), noise added (middle pair) or noise replaced (right-most pair) (Experiment 1).
Whalen & Samuel: Phonetic Information Across Nonlinguistic Sounds

Figure 2 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 middle pair of bars. Even though both matched and mismatched stimuli include acoustic discontinuities (in the form of the nonlinguistic noises), 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 ms; the mismatch of transitions added 24 ms 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 ms). 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 ms for a mismatch to be detected.

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 did not vary due to the type of nonlinguistic noise, $F(1,19) = 0.61$, n.s., nor did type of noise interact with any other factors.

The previously obtained slowing of reaction time with mismatched information was found even when explicitly nonlinguistic (in a sense, purely auditory) 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. It is still conceivable that there are two auditory discontinuities at work (the transitions and the nonlinguistic noises) and that they do not interfere with each other. Experiment 2 examines a situation where this interpretation is not possible.

Note that the identification times for the replaced stimuli are essentially the same as for the mismatched added stimuli (see Figure 2): 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 provides 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 is based on the fact that the vowel mismatch does not depend just on the first 60 ms of the vocalic segment, but is instead present throughout the noise, on the one hand, and the vocalic segment, on the other.
Procedure

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

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 had participated in Experiment 1.

Results and Discussion

Figure 3 presents the results of mismatching vowel information and for including noise in the stimuli. The two bars at the left indicate that mismatching vowel information had a significant slowing effect of 24 ms, $F(1,19) = 46.90, p < .001$. The three bars on the right indicate that adding noise slowed judgments by 29 ms, while replacing part of the syllable with noise slowed judgments by an additional 15 ms, $F(4,76) = 29.73, p < .001$. All three of these categories were significantly different from each other.

Figure 4 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, 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 ms for consonant identification, 31 for vowel, $F(1,19) = 6.78, p < .05$). 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 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 be-
Figure 3. 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 4. Identification times for stimuli with matched (open bars) or mismatched (cross-hatched bars) vowel information with no noise (left-most pair), noise added (middle pair) or noise replaced (right-most pair) (Experiment 2).
between two temporally separated portions of the utterance. The present experiment 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 process that integrates relevant information, not to any simple distractions caused by auditory discontinuities.

One difference between the two experiments is the absolute amount of time it took for the phonetic decisions. Subjects were, on the whole, 68 ms slower in Experiment 2 than in Experiment 1. An analysis (with the factors used before plus the factor of Experiment) of the ten 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, while in the second, 50% did. This increase of conflicting information probably resulted in more cautious identifications, slowing down responses.

The fact that the two delaying effects, mismatches of vowel information and the addition 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 where the noise replaced the transitions left both possibilities open. As can be seen in Figure 4, 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 have shown that subjects are sensitive to whether the signal is intact or not: In both, replaced stimuli produced significantly slower reaction times than added stimuli. One possible explanation for this effect is that the replaced items are heard as interrupted or discontinuous and that this distracts 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 tests whether there are noticeable differences between added and replaced stimuli that would support the "distracting" hypothesis. The test involves explicitly asking the subjects to discriminate between added and replaced stimuli. If the subjects are being distracted by the replacement of the signal, then added and replaced stimuli should be discriminable.
Procedure

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 /sa/, /fa/, /su/, and /fu/.

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" versus "replaced") was randomly selected. 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 randomly ordered.

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 met the vowel," and that the noise would either replace a small bit of the syllable, or be superimposed on it. 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 one second after a subject's response was received, the next stimulus was presented. The entire procedure took approximately 15 minutes.

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 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 X extraneous noise).
The results of this analysis can be summarized very simply: Subjects are utterly unable to discriminate "added" and "replaced" stimuli. In signal detection analyses, a d' of 0 indicates no discriminability, with increasing values reflecting an ability 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 these 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.

General Discussion

The phonetic mismatch effects of Whalen (1984) were successfully replicated, even with stimuli containing a nonlinguistic noise, inviting the auditory 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 is shown in Experiment 3.

It appears then that listeners are indeed sensitive to all phonetic information given them, and that the delays caused by mismatches, even those that cannot be readily heard, are due to increased phonetic processing. Even when the subject is given every excuse for failing to integrate, as when a nonlinguistic noise occurs in the middle of the signal, she still does 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 psycho-acoustic disruption but is sustained over a relatively long stretch. Whether the information stored is acoustic or (weakly) categorical remains to be seen.

References


