CAN LINGUISTIC BOUNDARIES CHANGE THE EFFECTIVENESS OF SILENCE AS A PHONETIC CUE?*

Bruno H. Repp

Abstract. This study investigated the influence of three kinds of linguistic boundaries—word boundaries, prosodic breaks, and syntactic breaks—on the perception of a silent interval at the boundary site as a cue to the presence of a labial stop consonant. The experimental technique involved cross-splicing portions of four naturally produced pairs of sentences, as well as presentation of excerpts from these sentences. Although one sentence pair showed a pronounced syntactic boundary effect, the other three (including two that were better controlled for semantic bias) did not, which points to a different, stimulus-specific origin of the effect obtained. Prosodic boundary effects were also generally absent, presumably because the stimuli were constructed such that prosodic variation ceased 78 ms prior to the critical silent interval. Only introduction of a word boundary effected a systematic reduction in stop consonant percepts, although this manipulation was confounded with other contextual factors. On the whole, the data provide little evidence for any direct effects of structural linguistic variables on phonetic segment perception; such effects seem to be restricted to the level of word recognition.

1. Introduction

One fundamental issue in speech perception research concerns the relative importance of physical signal properties ("bottom-up" information) versus the listener's expectations and interpretations ("top-down" processes). There is little doubt that phonotactic, semantic, and pragmatic factors can influence word perception, particularly when the speech signal is ambiguous (see, e.g., Fox, 1984; Ganong, 1980; Massaro & Cohen, 1983). Whenever a listener has internally generated or contextually induced expectations about the likelihood of certain phonological or lexical alternatives, these expectancies will help reduce any uncertainty introduced by insufficient physical information.

It is much less clear whether a listener's apprehension of structural factors that do not affect the likelihood of phonological or lexical alternatives can have repercussions at the level of phonetic segment perception. Specifically, the question is whether linguistic boundaries (syllabic, lexical, or syntactic) can reduce the phonetic coherence of an utterance at the boundary site, with possible consequences for the perceived segmental composition. Such an interaction, if it were to occur, would be theoretically interesting, for it would suggest that higher-level processes of

*Journal of Phonetics, in press.
Acknowledgment. This research was supported by NICHD Grant HD01994 and BRS Grant RR-05596 to Haskins Laboratories. I am grateful to Andrea Levitt and Patti Price for critical comments.

[HASKINS LABORATORIES: Status Report on Speech Research SR-84 (1985)]
lexical access and syntactic analysis can exert a direct influence on the internal representation of the bottom-up information, or at least generate expectations about its detailed acoustic structure. It should be kept in mind, however, that the effects under investigation, unlike the top-down effects studied extensively in research on word recognition (see, e.g., Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980), are rather special phenomena that, even if real, probably play only a very minor role in real speech understanding.

The evidence for such effects, however, is not compelling so far. Previous studies on this topic have been concerned with the function of silence as a phonetic cue. There is much evidence that short periods of silence in speech are not perceived as gaps or interruptions but as carriers of articulatory information about closure of the vocal tracts, as occurs in connection with stop and affricate consonants (see, e.g., Dorman, Raphael, & Liberman, 1979; Repp, Liberman, Eccardt, & Pesetsky, 1978). One particular situation investigated in several recent studies involves the effect of a short interval of silence preceding a fricative noise as a cue to the contrast between a word-initial fricative and affricate (Dechovitz, 1979, 1980, 1981; Price & Levitt, 1983; Rakerd, Dechovitz, & Verbrugge, 1982). The hypothesis tested in these studies was that introduction of a coincident linguistic break might reduce the perceptual effectiveness of the silence, either because the silence could be interpreted as a hesitation associated with the break rather than as an articulatory closure associated with a stop consonant, or because the linguistic boundary has a direct disruptive influence on the coherence of the signal portions preceding and following the silence, so that the presence and precise duration of the closure interval become perceptually irrelevant. Dechovitz (1979, 1980, 1981) claimed to have found such an effect due to syntactic structure alone—i.e., he found a significant reduction of silence-cued affricate percepts when a syntactic boundary at the critical location was created by remote context under semantically neutral and constant local acoustic conditions. These data have not been published, however, and Price and Levitt (1983) have failed to replicate the effect.

All these previous studies, however, found that the introduction of clause- or sentence-final prosody—including a falling intonation contour and final syllable lengthening—reduced the perceptual effect of a following silent interval. Although prosodic changes usually accompany changes in linguistic structure and thus carry considerable lexical and syntactic information, they do involve acoustic changes in the immediate vicinity of the silent interval. Since this may alter some of the local phonetic cues, the observed prosodic effects may not represent an influence of perceived linguistic structure on phonetic perception but may have more direct causes.

The present experiment extended these earlier studies by further investigating the influence on phonetic perception of syntactic and prosodic breaks, and by also considering the possible role of word boundaries. Stimulus materials were chosen in which the critical silence served as a cue for a labial stop consonant following a fricative and preceding a liquid (see Fitch, Halwes, Erickson, & Liberman, 1980; Dorman et al., 1979). The fricative-affricate contrast used previously is characterized by a rather sharp category boundary at a very short silence duration, which raises the possibility of psychoacoustic interactions that are immune to contextual influences. The type of contrast employed here, on the other hand, typically has its category boundary at relatively longer silence durations, so low-level psychoacoustic interactions are unlikely (see Pastore, Szcesniul, & Rosenblum,
1984; Repp, 1985), and it also has a larger region of ambiguity, which makes it more sensitive to influences of all kinds. The critical silence was embedded in plausible, natural sentences, which constitutes an improvement over the somewhat contrived and limited materials used in earlier investigations. Syntactic and prosodic factors were varied independently by swapping the two words surrounding the critical silent interval between syntactically different sentence frames. (Prosodic variation beyond these two words was confounded with syntactic structure.) Prosodic variation in the word immediately preceding the silence included the duration and amplitude envelope of the final [s] noise segment, which—judging from earlier findings and from informal observations during stimulus construction—would certainly have had a strong perceptual effect. Because of this foregone conclusion, it was decided to neutralize this segment and to examine only whether prosodic information beyond the immediately preceding acoustic segment can influence phonetic perception.

It should be pointed out that the role of silence as a cue to stop consonant perception is twofold. If the closure silence is too short (less than about 60 ms in the fricative-liquid context), no stop consonant may be perceived even when other cues are available (e.g., Dorman et al., 1979; Fitch et al., 1980). If the silence is longer (roughly 100-200 ms), a (labial) stop consonant will often be perceived even when there are no other cues (Dorman et al., 1979; Repp, 1985). These two effects may be called "stop suppression" and "stop generation," respectively (Repp, 1985). The stop suppression effect may in part be due to psychoacoustic interactions (such as forward masking) between the closely adjacent signal portions (however, see Pastore et al., 1984), whereas such interactions are much less likely in the case of the stop generation effect. Therefore, if there are any effects of linguistic boundaries on phonetic perception, they are more likely to occur at longer closure intervals, where psychoacoustic interactions play no role. The specific hypothesis tested was that, compared to a no-boundary condition, introduction of a linguistic boundary at the point of the critical silence would decrease the number of stop consonant responses at relatively long closure durations. To the extent that stop suppression is not caused by psychoacoustic interactions, an increase of stop responses might be predicted at short closure durations, because linguistic boundary might then reduce the (negative) cue value of short silences as well.

Following some piloting, two full-size experiments were conducted that were very similar in design. Because stimulus parameters were still not optimal, the first experiment inadvertently focused exclusively on the region of stop consonant suppression, where little sensitivity to linguistic boundaries was expected (and obtained). Therefore, only the results of the second experiment will be reported, which—due to additional stimulus adjustments—successfully encompassed both regions of stop consonant generation and suppression. Where the two designs overlapped, the results of the first experiment were consistent with the findings reported below.

2. Methods

2.1. Subjects

Ten paid volunteers participated. All were Yale undergraduates and native speakers of American English.
2.2. Stimulus Preparation

The stimulus sentences are shown in Table 1. Four pairs of sentences were constructed. The members of each pair contained the same two critical words in succession; the first word ended in [s], whereas the second word either did or did not begin with [b], so that there were two versions of each sentence. In one sentence of each pair (version b), a clause boundary intervened between the two critical words, whereas in the other sentence (version a), the two words formed a syntactic unit. The second critical word, which either did or did not begin with [b], represented fictitious surnames in two instances (Nos. 2 and 3) and real words in the other two (1 and 4). Orthogonal to this distinction, the consonant following the optional [b] was [l] in two words (1 and 3) and [r] in the other two (2 and 4). Because of the two possible versions of the second critical word, there was a total of 16 sentences.

Table 1

<table>
<thead>
<tr>
<th>Stimulus Sentences</th>
</tr>
</thead>
<tbody>
<tr>
<td>a. The royal tomb was protected by six (b)locks of solid gold.</td>
</tr>
<tr>
<td>b. When the clock strikes six, (b)lock the gate.</td>
</tr>
<tr>
<td>a. The girl tried to kiss (B)Radford on the cheek.</td>
</tr>
<tr>
<td>b. After giving his wife a kiss, (B)Radford boarded the train.</td>
</tr>
<tr>
<td>a. Will you please welcome Miss (B)Lackman to the office.</td>
</tr>
<tr>
<td>b. Enraged by a spectacular miss, (B)Lackman quit the game.</td>
</tr>
<tr>
<td>a. To the maid's dismay, worse (b)rooms could hardly be imagined.</td>
</tr>
<tr>
<td>b. What made matters worse, (b)rooms were difficult to find.</td>
</tr>
</tbody>
</table>

These 16 sentences were recorded by a male speaker of American English in a sound-insulated booth using high-quality equipment. The recordings were low-pass filtered at 4.9 kHz and digitized at a 10 kHz sampling rate. Using a waveform editor in conjunction with careful listening, each sentence was divided into four sections that were stored in separate computer files: preceding context (C1), first critical word (W1), second critical word (W2), and following context (C2). All cuts were made at zero crossings. In those sentences in which W2 had an initial [b], the stop closure was edited out and discarded. Thus, W1 ended at the beginning of the stop closure and W2 began at its end. In sentences without a W2-initial [b], the end of W1 and the beginning of W2 coincided, except in two sentences in which a lateral noise burst occurring at an [s-1] juncture was edited out.

For each sentence pair listed in Table 1, each of the two different context frames (C1+C2) existed in two distinct productions. Only one of these was retained—that deriving from sentences in which W2 had been articulated with an initial [b] (an arbitrary choice). The first critical word (W1) existed in four recorded versions; only those two versions that were not followed by a W2-initial [b] were used (another arbitrary choice). In these
two remaining versions, the clause-final [s] noises were much longer in
duration (ranging from 144 to 201 ms across the four sentences) than the
non-clause-final [s] noises (range: 55 to 109 ms). For reasons outlined in
the Introduction, these noises were removed and replaced by a constant [s]
noise excerpted from the same talker's production of the word "spectacular"
(sentence 3b). (For an explanation of this choice, see below and footnote 2.)
This [s] noise, originally only 54 ms in duration, was artificially lengthened
to 78 ms by duplicating a 24-ms central section of the waveform.

Finally, the onsets of the W2 words, which existed in four recorded
versions, were examined and edited. Words articulated with an initial [b] all
had labial release bursts ranging in duration from 12 to 18 ms. These bursts,
which provided strong stop manner cues (see, e.g., Repp, 1984a) were
eliminated, leaving only potential coarticulatory cues in the periodic
stimulus portion. The words without an initial [b] had no bursts and were
retained without change.

In summary, then, for each of the four sentence pairs listed in Table 1,
there were two different context frames C1+C2, each in a single recorded
version; two versions of W1, a clause-final one and a non-clause-final one,
with a common final [s] noise; and four versions of W2, two that had
originally started with [b] and two that had not, and orthogonal to this
distinction, two clause-initial and two non-clause-initial ones.

These components were re-assembled into sentences, with four different
silent closure intervals introduced between the W1 and W2 words: 40, 80, 120,
and 160 ms. All possible combinations of sentence components were employed in
the sentence test, leading to a total of 4 (sentence types) X 2 (contexts) X 2
(W1) X 4 (W2) X 4 (silences) = 256 sentences. They were recorded in 4 blocks
of 64, randomized within each block in groups of 16, with interstimulus
intervals (ISIs) of 3 s and intervals of 10 s between groups. The first and
third blocks contained sentences in which the prosody of W1 was appropriate
for the syntactic context, whereas the second and fourth blocks contained the
sentences in which W1 had the inappropriate prosody. These latter sentences
sounded somewhat odd but not bizarre; they were deemed appropriate for an
assessment of prosodic factors.

In addition to this lengthy main test, four shorter test tapes were
recorded. The first of these was a pretest containing 16 sentences. The
first 8 sentences represented the eight different contexts, with prosodically
appropriate W1 and W2; W2 was either the "stronger" version (i.e., that
originally began with [b]) preceded by the second-shortest silence (80 ms), or
the "weaker" version (that originally began with [l] or [r]) preceded by the
longest silence (160 ms). The second set of 8 sentences contained the
context-W2 combinations not contained in the first set. All 16 sentences were
arranged in a quasi-counterbalanced sequence, with ISIs of 20 s. The purpose
of this pretest was to assess the listeners' response to the test sentences on
first hearing.

The second test contained the W1-silence-W2 word pairs in all 128
possible combinations, without their sentential context. They were recorded
in 4 blocks of 32, with ISIs of 4 s. The purpose of this test was to provide
a baseline for assessing the contribution of the contextual frame, regardless
of its syntactic implications, and to examine prosodic effects in this more
restricted context (cf. Price & Levitt, 1983).
In the third test, the W2 words were preceded only by the constant [s] noise plus silence, to provide a baseline for testing the hypothesis that a word boundary following the [s] reduces the likelihood of silenced-cued labial stop percepts. In this test, the [s] was to be perceived as the initial segment of a nonsense word (e.g., "splock"). The constant [s] noise was taken from a word-initial position (see above) to facilitate this task.  

Finally, the excerpted W2 words were assembled into a single-word test. The 16 W2 words (4 words X 4 versions) were recorded in 4 different random orders with ISIs of 4 s. This test was to provide a baseline against which the effect of closure silence in the other tests could be compared.

2.3. Procedure

The subjects listened to all tests in a single session, using TDH-39 earphones in a quiet room. The tests were presented in a fixed sequence: The pretest was followed by the sentence test, the word pair test, the nonsense word test, and the single word test.

In the pretest, the subjects' task was to write each sentence down verbatim on a blank sheet of paper. Subjects were informed that the sentences were meaningful, that some of them contained proper names, and that the second set of 8 would be very similar—but not necessarily identical—to the first set of 8.

For the sentence test, the subjects were provided with printed answer sheets. Each page listed all the stimulus sentences on top, arranged as in Table 1, without the italics but with two words in each sentence capitalized. The first of those words was a key word in the first clause (e.g., ROYAL) identifying the context; the second was W2. For each item the answer sheet listed the four pairs of possible key words and W2 below each pair, with the initial B in parentheses. The subjects' task was, for each sentence heard, first to circle the appropriate key word and then to indicate, by either circling or crossing out the parenthetical B in the word below, whether W2 did or did not begin with a [b]. Since the sentences came at a fairly brisk rate, the subjects were encouraged to circle the key word before the sentence was over, and to skip the key word if the time seemed too short. Some subjects omitted a few key word responses in the beginning but soon found their rhythm. The only purpose of the key word responses was to keep the subjects' attention on the context and thus to prevent an overly selective listening strategy.

For the word pair test, answer sheets listed for each item the four possible W1-W2 pairs, with the W2-initial B in parentheses. The subjects' task was to find the appropriate word pair and either to circle or cross out the B. For the nonsense word test, the answer sheet listed for each item the four possible choices with a parenthetical P following the initial S (i.e., S(P)LOCK, S(P)RADFORD, S(P)LACKMAN, S(P)ROOMS). Subjects were asked to try their best to consider the stimuli as [s]-initiated nonsense words and to either circle or cross out the P in the correct alternative. Their attention was drawn to the unfamiliar [sr] cluster as a possible beginning of a nonsense word. Finally, the answer sheet for the single word test listed the four possible W2 choices for each item, and subjects located the correct alternative and either circled or crossed out the parenthetical word-initial B.
3. Results and Discussion

3.1. General Contextual Effects

Averaging over different versions of W1 and W2, Figure 1 shows the results in terms of percent labial stop responses to W2 onset, separately for each sentence pair (S1-S4), as a function of silent closure duration. The various response functions compare sentences with (S-b) and without (S-a) a syntactic break preceding W2, word pairs (WP), and nonsense words (NW). The percentage of "b" responses to single W2 words (SW) is indicated by the arrows at the right-hand side of each panel.

The first thing to note is that the percentage of labial stop percepts increased as closure duration increased. Repeated-measures analyses of variance on the separate tests showed that this expected effect was extremely significant and also interacted strongly with the Sentence factor, as is evident from the different slopes of the response functions (all effects at least p < .001). A visual comparison with the single-word (SW) percentages shows that labial stop responses at the longer closure durations exceeded those to single W2 words by a considerable margin (the stop generation effect), whereas the opposite relationship held at the shortest silent interval (the stop suppression effect).

The next finding to note in Figure 1 is that the response functions for word pairs (WP) were not systematically different from those for sentences (S-a and S-b combined); thus, having some sentential context around the W1-silence-W2 constellation did not influence the subjects' criterion for reporting a "b." By contrast, the percentages of labial stop responses were much higher in [s]-initiated nonwords (NW) than in the other conditions, where a word boundary separated the [s] from the following context. (The exception is Sentence 3, where a ceiling effect may have prevented a difference from emerging.) In a combined analysis of the WP and NW conditions, the main effect of Condition was highly significant, $F(1,9) = 51.18$, p < .0001, and so were its interactions with Sentences, Closure Duration, and both of these factors (all p < .0004 or less, mainly due to the different pattern for sentence 3). The interaction with Closure Duration reflected the fact that the effect was smallest at the shortest silence duration; there was no tendency toward a reversed effect in the stop suppression region, which suggests some psychoacoustic limit at short silences. A response bias against the unfamiliar "sr" clusters in nonwords could have operated in sentences 2 and 4, but not in sentence 1. Thus, unless the immediate context preceding the [s] (i.e., W1) had some direct influence on subjects' criteria, apart from introducing a word boundary, these results may be interpreted as supporting the hypothesis that the linguistic factor of word juncture attenuated the cue value of longer silences as a positive stop manner cue.

3.2. Syntactic Effects

Turning now to the comparison of syntactic conditions, it is evident from Figure 1 that there was a large and consistent difference between the two versions of sentence 1, with the syntactic boundary version (S-b) receiving fewer "b" responses. However, none of the other three sentences showed such a consistent difference. This pattern of results was reflected in a highly significant Sentence X Context interaction, $F(3,27) = 10.7$, p < .0001, whereas the main effect of Context was not significant. Separate analyses of variance for individual sentences showed a significant effect of Context for sentence
Figure 1. Percent stop responses, separately for the four sentences (S1-S4), as a function of closure duration. Separate response functions are shown for sentences without a syntactic boundary (S-a), with a syntactic boundary (S-b), for isolated word pairs (WP), and for nonsense words (NW). The arrows represent the percentages for single words (SW).
Repp: Linguistic boundaries

1, F(1,9) = 11.51, p < .01, but no significant effects for any of the other sentences. Since sentences 2 and 3, because of the use of proper names as W2, were semantically better controlled than sentences 1 and 4, these results do not support the hypothesis of a syntactic influence on phonetic perception. Rather, they suggest that there was something peculiar about sentence 1.

The most likely possibility is that the two alternatives of W2 were not equally plausible in the two semantic contexts of sentence 1, "six blocks of gold" being more acceptable than "six locks of gold," and perhaps also "lock the gate" being preferred to "block the gate." This possibility was assessed by presenting versions a and b of sentences 1 and 4 in written form to 20 staff members of Haskins Laboratories, with the request to choose the W2 alternative that "fits better into the sentence frame (i.e., that makes the sentence more meaningful, more likely, or more appealing)." To counteract order effects, two versions of this short test were used, with reversed orderings of the sentences and of the W2 alternatives for each sentence. The results revealed that "block(s)" indeed was considered relatively more plausible in sentence 1a (8 out of 20 responses) than in sentence 1b (0 responses). A similar asymmetry was obtained for sentence 4: "brooks" was preferred in sentence 4a (10 responses) relative to sentence 4b (1 response). Although sentence 4 did not show a significant "syntactic" effect in the sentence test, there was a tendency in that direction (Figure 1). Therefore, the "syntactic" effect in sentence 1 is attributed very tentatively to a semantically conditioned response bias.

3.3. Prosodic Effects

The absence of consistent syntactic effects implies also that prosodic variation in the sentence frame preceding W1 had no systematic effect (with the possible exception of sentence 1). In addition, however, it was quite obvious from the data that W1 prosody itself had very little effect. The effect of appropriate vs. inappropriate prosody (with respect to syntactic structure) should have been revealed in a Context X W1 interaction in the sentence test. This interaction was nonsignificant. There could also have been an effect due to W1 intonation per se (clause-final vs. non-clause-final), regardless of its context appropriateness. The W1 main effect, however, was likewise nonsignificant in both the sentence and word pair tests. Moreover, no individual sentence showed any pronounced prosodic effect. This was surprising, since earlier studies (Dechovitz, 1979; Rakerd et al., 1982) had found strong prosodic effects, and the present technique of cross-splicing might have been expected to introduce artifactually large effects.

One important possibility to consider is that clause-final and non-clause-final versions of W1 simply did not differ much, apart from the original difference in final [s] duration (see Methods section), which had been neutralized. To examine this issue, temporal measurements were obtained from the W1 waveforms and are shown in Table 2. It is clearly evident that clause-final versions (b) of W1 had substantially longer durations and lower terminal fundamental frequencies than non-clause-final versions (a). The durational differences extended over all acoustic segments of the W1 syllable, including of course the final [s] prior to its neutralization (not shown in Table 2). Thus there was a clear basis for potential prosodic effects due to W1.
Table 2

<table>
<thead>
<tr>
<th>Sentence</th>
<th>W1 Duration (ms)</th>
<th>Terminal F₀ (Hz)</th>
</tr>
</thead>
<tbody>
<tr>
<td>six</td>
<td>[s] [I] [k]</td>
<td>Total</td>
</tr>
<tr>
<td>1. a.</td>
<td>75</td>
<td>172</td>
</tr>
<tr>
<td>b.</td>
<td>135</td>
<td>289</td>
</tr>
<tr>
<td>kiss</td>
<td>[kh] [I]</td>
<td>Total</td>
</tr>
<tr>
<td>2. a.</td>
<td>46</td>
<td>97</td>
</tr>
<tr>
<td>b.</td>
<td>63</td>
<td>150</td>
</tr>
<tr>
<td>miss</td>
<td>[m] [I]</td>
<td>Total</td>
</tr>
<tr>
<td>3. a.</td>
<td>39</td>
<td>89</td>
</tr>
<tr>
<td>b.</td>
<td>106</td>
<td>197</td>
</tr>
<tr>
<td>worse</td>
<td>(not segmentable)</td>
<td></td>
</tr>
<tr>
<td>4. a.</td>
<td></td>
<td>151</td>
</tr>
<tr>
<td>b.</td>
<td></td>
<td>233</td>
</tr>
</tbody>
</table>

Definitions: [s] = fricative noise
[I] = voiced portion
[k] = silent closure interval
[kh] = release burst and aspiration
[m] = nasal murmur
Terminal F₀ = average F₀ of the last three complete pitch periods

The absence of any systematic prosodic effects then presumably has to do with the presence of a constant [s] noise between the prosody-carrying portion of W1 and the critical silent interval. This constant signal portion may have acted as a buffer against prosodic influences, and if so, it must be concluded that these influences are quite local in nature. In earlier studies using the fricative-affricate contrast, the distinctive prosodic information continued right up to the beginning of the silence. As was already pointed out above, there was little doubt that the [s] noise, had it been allowed to vary according to its natural production characteristics in clause-final and non-clause final position, would have had a strong influence on subjects' likelihood of reporting labial stop percepts. Such an effect would have been expected on the basis of fricative noise duration alone (Repp, 1984b; Summerfield, Bailey, Seton, & Dorman, 1981).  

4. Summary and Conclusions

In the present study it was attempted to create a perceptual discontinuity at the point of a critical silent interval by purely linguistic means in a relatively natural speech processing situation. The effect of word boundaries was studied, as well as the effects of (slightly removed) prosodic and syntactic breaks, following earlier studies by Dechovitz (1979, 1980, 1981), Rakerd et al. (1982), and Price and Levitt (1983).
There was a clear effect of introducing a word boundary. Although this effect was confounded with the presence vs. absence of preceding word context and therefore must be interpreted with care, it does suggest the possibility that within-word silence is more tightly integrated into the speech stream than is between-word silence. The reason for this may lie in subjects' expectations based on experience with real speech, in which interword intervals tend to be less reliable indicators of phonetic distinctions than intraword silences.

In contrast to several previous studies, there were no effects of prosodic discontinuity. The most likely explanation for this is the fact that the fricative noise immediately preceding the critical silence was not allowed to vary, so that the distinctive prosodic information ended 78 ms before the silent interval. If this interpretation is correct, it indicates that prosodic effects of the kind demonstrated by Price and Levitt (1983) and Rakerd et al. (1982) are extremely local in character and are probably caused by the duration of the acoustic segment preceding the silence, which acts as a secondary stop manner cue. Similarly restricted effects have been observed in related experiments on the perception of vowel duration in sentence context (Luce & Charles-Luce, 1985; Nooteboom & Doedeman, 1980) and on the perceptual consequences of varying speaking rate (e.g., Summerfield, 1981). Rather than constituting a direct influence of suprasegmental variation on segmental perception, these effects may be mediated by changes in local acoustic signal properties serving as segmental cues.

There were no consistent effects of syntactic structure per se on phonetic perception. The anomalous results for one sentence pair were probably due to a semantic bias. These negative results confirm the conclusions of Price and Levitt (1983) and cast further doubt on the replicability of Dechovitz's (1979, 1980, 1981) unpublished findings showing a "purely syntactic" effect on phonetic perception. It seems likely that syntactic processes operate exclusively at a level beyond that of segmental phonetic classification.

References


Performance, 6, 110-125.

Footnotes

However, the effects studied here do not require phonetic ambiguity, as do most other contextual effects in speech perception. Rather, these purely structural effects, if extant, should disrupt the perceptual contribution of closure silence even at its optimal, least ambiguous setting. (That there is often some ambiguity even at that setting is due to the fact that closure duration is only a secondary cue to stop manner; see Repp, 1984a.)
In the author's judgment, word-final [s] noises were not acceptable as word-initial segments, whereas the word-initial [s] seemed acceptable both as a word-initial or word-final segment. In any case, in the sentences and word pairs lexical and semantic constraints were assumed to exert sufficient pressure on listeners to consider the [s] as W1-final, even if its acoustic characteristics were more appropriate for a word-initial position.

Except for a small reversed effect for sentence 2, $F(1,9) = 10.38$, $p < .02$, which interacted strongly with one of the two W2 factors, $F(1,9) = 69.05$, $p < .0001$, being due entirely to the clause-initial version of W2. The reason for this interaction is not known.

It is conceivable that potential effects of syntactic structure were attenuated in the sentence test because the repetition of the same sentences and listeners' knowledge of the critical phonetic contrast gave rise to selective listening strategies. However, the original positive findings of Dechovitz (1979, 1980, 1981) were obtained with even more repetitive materials, and at least some degree of attention to preceding context was maintained by the requirement of key word responses in the sentence test. Moreover, in the pretest both sentences 1 and 4 showed an effect of syntactic structure at the longer closure duration ("b" responses were given only when there was no syntactic break preceding W2), whereas sentences 2 and 3 showed no effects. Thus there was no syntactic effect in the semantically unbiased sentences even on first hearing.

Price and Levitt (1983) found no prosodic effect in a cross-splicing experiment similar to the present one, but this may have been due to an unusually clear-cut phonetic contrast cued by a small amount of closure silence, a situation that was not duplicated here.

Two additional stimulus variables were lodged in the critical W2 word: one contrasting the strong and weak versions of W2, and the other contrasting the clause-initial and non-clause-initial versions. The effects of these factors followed a highly varied and token-dependent pattern of results and are of only marginal interest here. Details may be obtained from the author.