Journal of Experimental Psychology:
Human Perception and Performance
1991, Vol. 17, No. 1, 289–302

0096-1523/91/$3.00

749

Effects of Preceding Context on the Voice-Onset-Time Category Boundary

Bruno H. Repp and Hwei-Bing Lin
Haskins Laboratories
New Haven, Connecticut

The category boundary on a voice-onset-time (VOT) continuum ranging from *bin* to *pin* shifts when the stimuli are preceded by different carrier phrases. By using a variety of precursor phrases, by varying the temporal interval between precursor and test word, and by selectively eliminating either voicing information or spectral structure from the precursors, the present experiments show that the context effect is caused by the presence or absence of voicing in the precursor's final segment. The effect decreases with temporal separation but persists over several seconds. In addition, the "baseline" VOT boundary for isolated test words interspersed in a test sequence shifts depending on what precursor stimuli occur in the same sequence. The perception of VOT thus seems to be sensitive to both close and distant manifestations of laryngeal activity, always in an assimilative fashion, which suggests an integrative perceptual mechanism with a long time constant.

The research described in this article was stimulated by an unexpected finding. In recent experiments (Repp & Lin, 1989), we investigated the effect of preceding context on the discriminability of small differences in voice onset time (VOT). The subjects in one of our experiments (Experiment 3) were also given an identification test in which stimuli from a *bin–pin* continuum varying in VOT were either presented in isolation or preceded by one of two carrier phrases, *Take this* or *Take the*. (These stimuli are described in more detail later.) The results of this identification test, which have not been previously reported, are shown in Figure 1. There was a clear shift of the category boundary in favor of *pin* responses following *Take this*, whereas the boundary following *Take the* was in the same place as that for isolated test stimuli. The average category boundaries, estimated by probit analysis and averaged over subjects, were at VOT values of 35.2 ms (no precursor), 26.6 ms (*Take this*), and 36.7 ms (*Take the*). The difference among these three boundaries was highly significant, $F(2, 24) = 83.99$, $p < .001$, but there was no significant boundary difference between the no-precursor and the *Take the* conditions when compared separately.

This type of context effect has not, to the best of our knowledge, been described previously in the literature. We decided to investigate its cause, which was likely to tell us something new and important about the perception of the phonetic voicing distinction, a topic of long-standing interest in speech research (see, e.g., Berg, 1986; Lisker & Abramson, 1970; Pisoni, 1977; Summerfield, 1981). Phonetic context effects are of theoretical significance because they reveal either listeners' tacit knowledge of systematic contextual variation in speech production (cf. Repp, 1982) or interactions in the auditory processing of the speech signal (cf. Diehl & Kluender, 1989), both of which are essential ingredients of any model of speech perception.

Most experiments on the perception of the voicing distinction in syllable-initial stop consonants have used isolated syllables or words, so the critical segment was in fact utterance-initial. A few studies, however, have demonstrated effects of preceding context. The most relevant of these was conducted by Summerfield (1981), who showed that the simulated speaking rate of a precursor phrase affects the perceptual boundary on a VOT continuum, with slower precursor rates leading to an increase of responses in the voiced category. Summerfield also showed that this effect derived primarily from the temporal variations in the final syllable of the precursor (however, see Kidd, 1989) and that it decreased substantially as the silent interval between precursor and test syllable was lengthened beyond a duration acceptable for a stop consonant closure. The duration of that closure itself had a rate-dependent effect on the perceived stop voicing, with relatively shorter closures leading to more voiced responses (again, see Kidd, 1989, for different results). Also, the precursor effects were unidirectional compared with a no-precursor baseline: The VOT boundary shifted toward longer values when the precursor rate was slow, but it stayed in place when the precursor rate was fast. The precursor phrase ended with the word *the* in Summerfield's experiments. His findings are thus consistent with the absence of a boundary shift in our (moderately fast) *Take the* precursor condition relative to the no-precursor baseline (Figure 1), but they cannot explain the shift in our *Take this* condition in which the precursor was produced at the same intended rate.

This conclusion may be premature, however. The precursors in our experiment had been excerpted from complete phrases of *Take this bin* and *Take the pin*, and they included the original silent closure intervals. Even though the boundary difference obtained in the perceptual test was contrary to what

---

This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. Hwei-Bing Lin was also at the Department of Linguistics, University of Connecticut. She is now at Speech and Hearing Sciences, The Graduate Center, City University of New York.

Correspondence concerning this article should be addressed to Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, Connecticut 06511-6695.

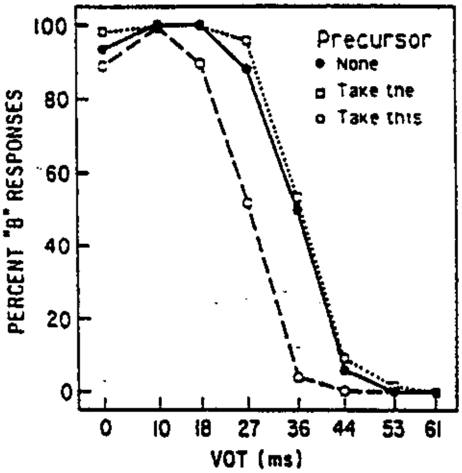

Figure 1. Percentage of "B" responses as a function of VOT for stimuli from a *bin–pin* continuum in three precursor conditions. Average data of 13 subjects.

might be predicted if there had been coarticulatory information predictive of stop consonant voicing, it is worth noting that the closure durations differed between the *Take this* (115 ms) and the *Take the* (84 ms) precursors. Because longer closures are generally associated with voiceless stop consonants (e.g., Lisker, 1957), this difference could easily account for the boundary shift in favor of *pin* responses following *Take this*, even though that phrase was originally followed by *bin*. In fact, Summerfield (1981) found such an effect of closure duration in his Experiment 3. Our first order of business, therefore, was to rule out this relatively trivial explanation of our finding.

Experiment 1: The Role of Closure Duration

In Experiment 1, we compared the effects of the two precursors at equal closure durations. To study the perceptual contribution of closure duration itself, we used four different silent intervals that spanned the normal range of stop consonant closures in this context.

In addition, we included a third precursor, referred to as *Take this-*. Its final /s/ noise had originally been produced in word-initial position (viz., in the word *spin*) but had been excerpted and substituted for the original /s/ noise of *Take this*. Because English phonology permits only /p/ following a word-initial /s/, it was of interest to see whether subjects' labeling responses would shift even more toward *pin* in the *Take this-* context, particularly at the shorter closures that characterize fricative-stop clusters (cf. Wingate, 1982). When we used the same stimuli in a VOT discrimination task (Repp & Lin, 1989), we found much more interference from the *Take this-* precursor than from the *Take this* precursor. This made us suspect that the location of the word boundary was ambiguous in the former context, leading to occasional perception of *Take the spin*. It is conceivable that, at short closure durations, even *Take this* with a proper word-final /s/ permits some word boundary ambiguity; if so, this would provide an alternative explanation of the boundary shift observed in our pilot study (Figure 1). Such an effect should, however, disappear at longer temporal separations.

The predictions for Experiment 1, then, were as follows: If the previously observed VOT boundary difference following *Take this* and *Take the* was due entirely to the difference in closure durations, it should be absent at all closure durations. If it was due to word boundary ambiguity in the *Take this* context, it should be present at short closures but disappear at longer ones. This tendency should be even stronger in the case of the *Take this-* precursor. In addition, if closure duration provides a secondary cue to stop voicing, *pin* responses should increase with closure duration, up to a point.

Method

Subjects. Thirteen paid volunteer subjects from the Yale University community participated. They were all native speakers of American English and reported having normal hearing.

Stimuli. The *bin–pin* continuum was constructed from natural productions of these words in the context of *Take this* by a female speaker. The utterances were recorded in a sound-insulated booth, digitized with a 20-kHz sampling frequency, low-pass filtered at 9.6 kHz and edited to separate the final words from the carrier phrases. Using a waveform editor, a VOT continuum was created by replacing initial periodic segments of *bin* with aperiodic segments of equal duration of *pin*, proceeding in steps of two glottal cycles. The six stimuli used in Experiment 1 had VOTs of 10, 18, 27, 36, 44, and 53 ms (rounded to the nearest ms). These stimuli were the same as in our pilot study (Figure 1), except that the two extreme stimuli of the earlier continuum were omitted for reasons of economy.

The precursors, too, were the ones used previously, spoken by the same person. The phrase *Take this* had originally been produced before *bin*, whereas *Take the* had been excerpted from an utterance ending in *pin* to avoid continuation of voicing into the closure. (Possible effects of anticipatory coarticulation are considered later; we have already noted that the direction of the observed perceptual boundary shift was contrary to what might be predicted on that basis.) The *Take this-* precursor, previously used in Experiment 2 of Repp and Lin (1989), was acoustically identical to the *Take this* precursor except for the spliced-in final fricative noise, which had been excerpted from the utterance *A crazy spin*. This word-initial /s/ noise was longer than the word-final one it replaced (148 ms vs. 94 ms) and probably also had spectral and amplitude characteristics that reflected its provenance.

The duration of the silent closure interval following each precursor was varied in four steps: 45, 80, 115, 150 ms. Each of the six *bin/pin* stimuli was presented in isolation and preceded by each of the three precursors at each of the four closure durations, which resulted in 78 stimuli that were recorded five times in different random orders, with interstimulus intervals of 2 s.

Procedure. The subjects listened binaurally in a quiet room over TDH-39 earphones at a comfortable intensity. The task was to identify the initial consonant of each test word as *B* or *P*.

Results and Discussion

The average labeling functions (percentage of "B" responses as a function of VOT) are shown separately for each closure duration in the four left-hand panels of Figure 2. It is evident, first, that the *Take the* precursor condition yielded results very similar to those for isolated stimuli: this replicates our earlier findings. Second, there was a large boundary shift in

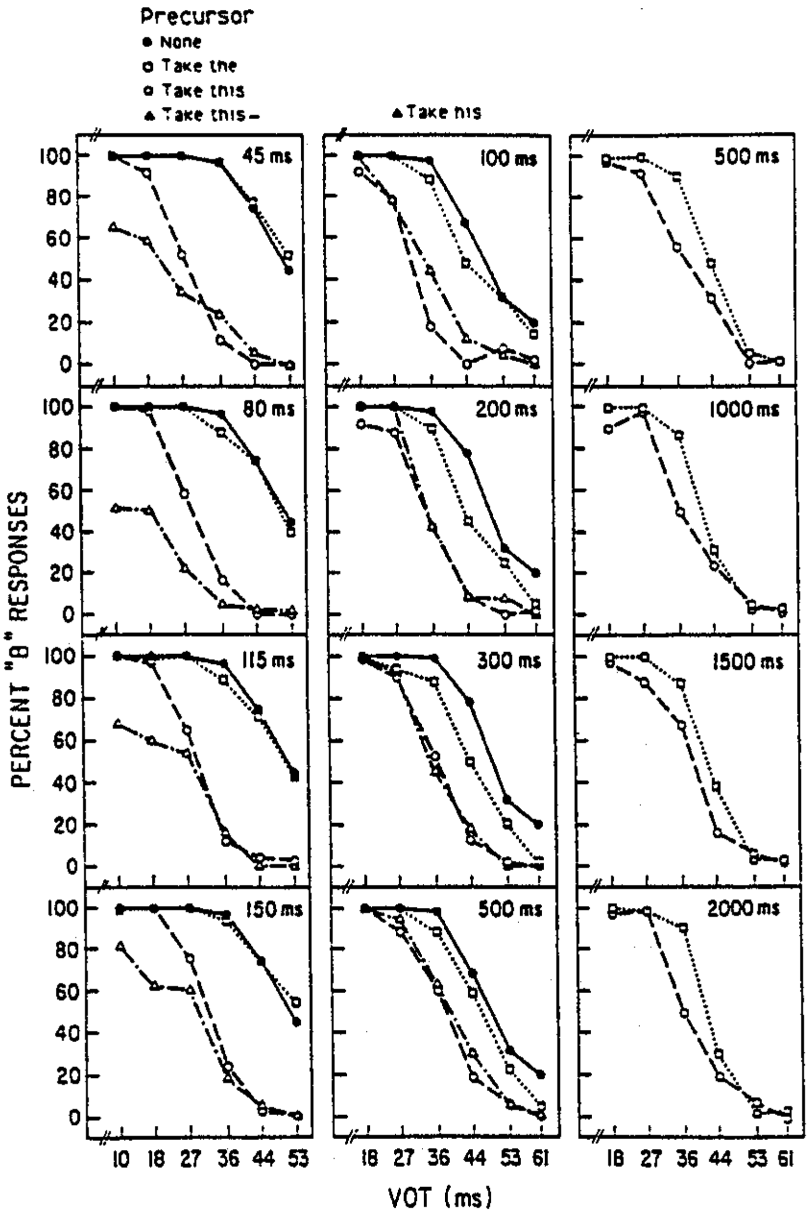

Figure 2. Results of Experiment 1 (left-hand panels), Experiment 2 (center panels), and Experiment 3 (right-hand panels). Each panel represents a particular closure duration.

the *Take this* precursor condition, favoring voiceless stop percepts at all closure durations. The shift was much larger than in the pilot experiment (Figure 1). Because the average boundary location in the *Take this* condition (29 ms of VOT) was similar to that previously obtained (27 ms of VOT), the larger boundary shift resulted from an unexpected increase in voiced stop responses in the *Take the* and the no-precursor conditions.[1] A third finding is that, as predicted, the *Take this-* precursor introduced an additional bias to report voiceless stop consonants at short VOTs. Finally, it is obvious that variations in closure duration had surprisingly little effect.

We conducted an analysis of variance (ANOVA), with precursor condition and closure duration as factors, on sub-

---

[1] The average category boundary for isolated stimuli was at the unusually long value of 52 ms of VOT, as compared with 35 ms previously. The reason for this large difference is not clear, as the range of VOTs was merely narrowed, not shifted, with respect to the pilot study (Figure 1). The category boundary was at a relatively long VOT for a bilabial continuum even in the pilot experiment. One reason for this bias may lie in the method of stimulus construction, which retained the acoustic properties of *bin* at and following the onset of voicing. These properties included the amplitude and fundamental frequency contours, which are known to be secondary voicing cues (see Abramson & Lisker, 1985; Darwin & Seton, 1983). The real reason, however, became apparent in later experiments of this series.

jects' proportions of *bin* responses (averaged over VOT) in the three precursor conditions. The ANOVA revealed highly significant effects of condition, $F(2, 24) = 142.84$, $p < .0001$, $MS_e = 26.30$, and of closure duration, $F(3, 36) = 8.79$, $p < .0003$, $MS_e = 3.25$, as well as an interaction between the two factors, $F(6, 72) = 7.33$, $p < .0001$, $MS_e = 2.52$. The latter two effects reflect the fact that there was an increase in voiced responses with increasing closure duration, but only in the *Take this* and *Take this-* conditions. This resulted in a reduction of the boundary shift with increasing closure duration, though it was still very large at the longest interval used.

These results definitely eliminate the hypothesis that our original observation of a boundary shift was caused by differences in silent closure duration associated with the precursors. In fact, closure duration had an unexpectedly small effect. In the *Take the* context, it did not affect the subjects' responses at all, contrary to Summerfield's (1981) finding of a decrease in voiced responses at intervals of 100 ms–250 ms in a comparable situation. In the *Take this* and the *Take this-* contexts, there was, on the contrary, an increase in voiced responses with closure duration. This should probably not be interpreted as a direct effect of closure duration as a cue to stop consonant voicing (cf. Kidd, 1989), but should rather be interpreted as a reduction with increasing temporal separation of some effect exerted by the precursors on the test syllables.

What, then, was responsible for this effect? One possibility was considered earlier: ambiguity of the word boundary location. We expected, however, that this factor would come into play only at relatively short closure durations, those characteristic of tautosyllabic fricative-stop clusters (e.g., less than 100 ms). The finding that the boundary shift was still so large at the longest closure duration (150 ms) seems to provide evidence against this explanation. However, the additional effect exerted by phonetic ambiguity in the *Take this-* precursor, obtained as expected at short closures, also persisted at the longest closure in only slightly attenuated form. Thus, a role of word boundary ambiguity cannot be ruled out completely. In any case, the difference between the *Take this* and the *Take this-* conditions demonstrates that phonetic properties of the precursor's final segment can affect stop consonant perception over separations of at least 150 ms. Perhaps, then, the difference between the *Take this* and the *Take the* conditions is also attributable to properties of the segments immediately preceding the closure (e.g., to the fact that /s/ is voiceless, whereas /ə/ is voiced). Before examining this hypothesis in detail, we conducted two additional experiments on the effects of temporal separation.

Experiment 2: Longer Closure Durations

The results of Experiment 1 raise the question of how long the temporal separation can be before the precursor effect disappears. This question has some bearing on the explanation of the effect: Word boundary ambiguity becomes an unlikely explanation once the closure interval is so long as to become effectively a pause separating the words. Experiment 2 increased the temporal separations up to 500 ms, which far exceeds the range of stop closure durations encountered in natural speech.

In addition, the experiment included a new precursor, *Take his*, which replaced *Take this-*. The experiment was intended to provide preliminary information about whether the voicelessness of the final fricative in *Take this* caused the boundary shift. If so, *Take his*, which ends with the voiced fricative /z/, should not cause any boundary shift.

Method

Subjects. Eight subjects from the same subject pool used in Experiment 1 participated. Four of them had also served in Experiment 1.

Stimuli. The stimuli were the same as in Experiment 1, with the following modifications. To better accommodate the category boundaries in all conditions, the range of VOTs was shifted by omitting the 10-ms VOT stimulus and by adding a 61-ms VOT stimulus. Four longer closure durations were used: 100, 200, 300, and 500 ms. Instead of the *Take this-* precursor, a newly recorded precursor, *Take his*, was used, produced by the same female speaker, originally preceding *bin*. Number of stimuli, test arrangement, and procedure were the same as in Experiment 1.

Results and Discussion

The labeling functions are shown in the center panel of Figure 2, separately for each closure duration. The shift in the range of VOTs apparently had no effect on the no-precursor labeling function. In contrast to Experiment 1, the *Take the* precursor led to a slight reduction in "B" responses relative to the isolated test words. As before, however, the *Take this* precursor caused a much larger boundary shift to the left, which diminished somewhat with increasing silent interval duration but, astonishingly, was still clearly present at the 500-ms interval. Finally, contrary to the predictions, the effect of the *Take his* precursor was very similar to that of the *Take this* precursor.

The statistical analysis of the precursor data confirmed the main effect of precursor condition, $F(2, 14) = 16.27$, $p < .0003$, $MS_e = 22.79$, and its interaction (i.e., its decrease) with closure duration, $F(6, 42) = 3.79$, $p < .005$, $MS_e = 2.50$. There was also a main effect of closure duration, $F(3, 21) = 5.95$, $p < .005$, $MS_e = 3.38$, which reflects the fact that "B" responses increased overall because of the progressive reduction of the *Take this* and *Take his* precursor effects. A separate comparison of the *Take this* and *Take his* conditions showed no difference between them, and no interaction with closure duration, only a main effect of closure duration, $F(3, 21) = 11.52$, $p < .0002$, $MS_e = 2.89$, due to a systematic increase in "B" responses.

This experiment yielded two surprising findings. First, the boundary shift was still present at a temporal separation of 500 ms. This rules out an explanation in terms of word boundary ambiguity, though the possibility remains that such ambiguity adds to the effect at short closure durations. Second, the *Take his* precursor had the same effect as the *Take this* precursor, which seems to suggest that the boundary shift was not caused by the voicelessness of the final fricative in *Take this*. However, a closer look at the waveform of the *Take his* precursor clarified this result. It emerged that the precursor's final fricative noise, 77 ms in duration, was largely voiceless, even though it had preceded *bin* in the original utterance: Glottal pulses ceased about 20 ms into the noise. Such final fricative devoicing at a juncture preceding another voiced segment appears to be common, though it is not well documented with data in the phonetic literature, which is unanimous in stating that devoicing occurs when the following segment is voiceless. (For data on British English, see Haggard, 1978.) Our results thus show only that the *phonological* voicing status of the final fricative (/z/ or /s/) is irrelevant to the precursor effect. The hypothesis that the precursor effect derives from the *phonetic* voicing status of the final segment is alive and well but remains to be tested properly (see Experiment 4).

Experiment 3: Very Long Closure Durations

The unexpected finding that the precursor effect was still present at 500 ms of temporal separation aroused our curiosity, and we decided to extend the interval durations even further in the hope that the temporal range of the phenomenon could be delineated.

Method

Subjects. Ten subjects from the same subject pool participated. Five had served in Experiment 2; 3 of them in Experiment 1 as well.

Stimuli. We used the same stimuli as in Experiment 2, but only the *Take the* and the *Take this* precursors. Isolated test words were likewise omitted to shorten the test and to avoid confusion in view of the rather long temporal separations between precursors and test stimuli. The four closure durations were 500, 1,000, 1,500, and 2,000 ms. There were five blocks of 48 stimuli each. The procedure remained unchanged.

Results

The labeling functions are shown in the right-hand panel of Figure 2. A small boundary shift was evident that did not change much with closure duration. Even at a 2-s separation there was still a small average effect. The ANOVA showed a significant main effect of condition, $F(1, 9) = 9.02$, $p < .02$, $MS_e = 14.70$, but no main effect of or interaction with closure duration.

Further Discussion of Experiments 1–3: The Time Course of the Precursor Effect

To get a better overview of the time course of the precursor effect, a summary of the results of Experiments 1–3 is presented in Figure 3. The figure shows average category boundaries (in milliseconds of VOT, determined by probit analysis of the average data) as a function of the temporal separation between precursor and test word (on a log scale).[2] The function on top shows that the VOT boundary in the *Take the* context decreased somewhat across the three experiments, from about 51 ms to 43 ms of VOT. Recall that the *Take the* labeling function coincided with the function for isolated test words in Experiment 1 but began to separate from it in Experiment 2. Experiment 3 did not include isolated stimuli for comparison, but the trend appears to continue. The function in the center represents the *Take this* boundaries, and here there is a clear tendency for the VOT boundary to increase with closure duration, both between and within experiments, from about 27 ms to 39 ms of VOT. At the bottom of the figure, the difference between the *Take the* and the *Take this* boundaries is plotted, together with 95% confidence intervals. The boundary difference can be seen to extend up to separations of 2 s and perhaps beyond. Considering how little change there was over the longer intervals, it might be speculated that there are really two effects: one decaying within about 1 s and the other extending over much longer intervals.

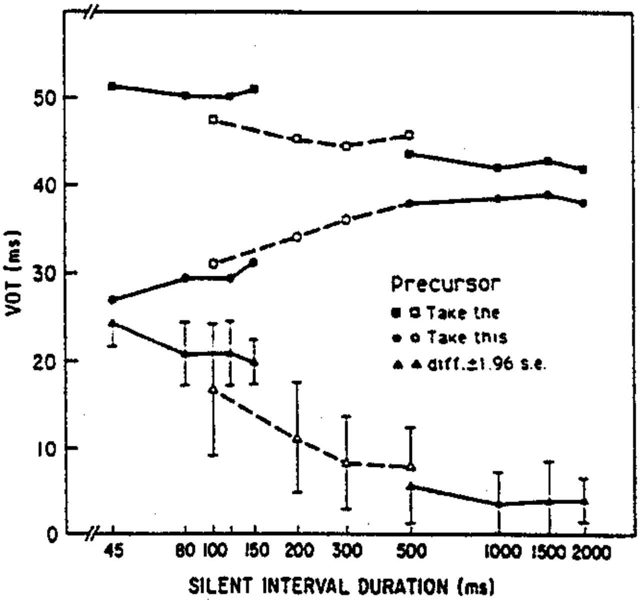

Figure 3. Summary of results from Experiments 1–3: VOT boundary locations as a function of temporal separation between precursor and test word in the *Take the* and *Take this* precursor conditions. The function at the bottom shows the difference between the boundaries in the two precursor conditions, with 95% confidence intervals.

A puzzling feature of these data is that the two precursor conditions converge onto a category boundary that is some 10 ms shorter than that for isolated test words in Experiments 1 and 2. Either there was a bias to perceive voiceless stop consonants following any precursor, regardless of how long the temporal separation, or some unknown factor introduced a bias to hear voiced stop consonants in isolated stimuli. This issue is addressed in later experiments.[3]

---

[2] The individual results of those subjects who participated in two or all three of the experiments are consistent with the pattern shown in Figure 3.

[3] A listener might expect isolated words to be pronounced at a slower rate than phrase-final words, so that the same stimuli might sound relatively faster in isolation than following a precursor. If so, the boundary for isolated stimuli should be at a shorter VOT (cf. Summerfield, 1981), which is the opposite of what we found. We also note that, although the boundary obtained for isolated stimuli in our

Although the difference between the *Take this* and the *Take the* precursor effects cannot be attributed to differences in speaking rate (if there were any unintended differences in rate, they were much too small to have any substantial effect), its decrease with increasing closure duration is remarkably similar to that of the speaking rate effect studied by Summerfield (1981; see his Figure 4, p. 1082). He found that the effect "was substantially reduced at an interval of 250 msec and reached an asymptote at an interval of about 1 sec" (p. 1082) and that it contained a "small long-lasting component" (p. 1082); the longest interval actually used in his study was 8 s. This time course is reminiscent of the decay times of the two components of auditory storage suggested by a large literature (reviewed by Cowan, 1984). Summerfield also showed that the speaking rate effect was primarily due to variations in the final syllable of the precursor (Summerfield, 1981, Experiment 4); in fact, the effect of selective rate variations in earlier syllables (more distant in time) was similar in size to the long-term component of the effect of overall rate variation. The speaking rate effect and the present precursor effect then perhaps have more in common than one might have thought: Both depend primarily on properties of the syllables or segments immediately preceding the closure and both seem to be mediated by auditory memory for these properties. Still, it seems that they cannot be one and the same phenomenon. In the case of the speaking rate effect, the relevant acoustic property of the final segment is its duration or rate of spectral change. In the present case, it may be the presence versus absence of voicing.

Experiment 4: A Variety of Precursors

Experiment 4 examined the role of voicing in the final precursor segment by using a much wider variety of precursor phrases. Three of these phrases ended in a voiceless fricative or affricate, /s/, /č/, or /f/. The latter two consonants cannot form a syllable-initial cluster with /p/, so if ambiguity of word boundary location is a factor at all, it should not operate with them. Three other precursors ended in a voiced fricative or affricate, /z/, /ǰ/, or /v/. Because of the problem with final fricative devoicing in Experiment 2, these precursors were produced both in a natural and in a deliberately voiced version. Two additional precursors were included to test whether it was specifically fricative voicing/voicelessness or general presence versus absence of voicing that influenced VOT perception. These precursors ended in a vowel, /i/, and were produced either in a normal voice or in a whisper. If whispered precursors ending in a vowel had the same effect as precursors ending in a voiceless fricative, this would show the effect to be caused by general voicelessness of a preceding segment. To guard against artifacts due to token differences, multiple productions of each precursor were used. Moreover, each precursor was used in two different versions, one originally produced before /b/, the other before /p/. This served to test the influence of anticipatory coarticulatory information on voicing judgments. Such effects may have been present in our earlier experiments, even though they were in opposition to the precursor effect of primary interest.

Method

Subjects. The 10 subjects of Experiment 3 were recalled for this experiment.

Stimuli. A new set of utterances was recorded by a different female speaker. She produced a variety of three-word phrases beginning with *Take* and ending with either *bins* or *pins*. The intervening words were: *six, these, such, large, tough, five,* and *three.*[4] The first six words formed three pairs contrasting in the voicing of the final segment (/s/ vs. /z/, /č/ vs. /ǰ/, /f/ vs. /v/). A similar contrast was created for the precursor ending in /i/ by having the speaker produce the whole phrase either voiced or in a whisper. Each of the precursor phrases ending in a voiced fricative or affricate preceding *bins* was produced in an additional version, with the speaker (a graduate student of speech science) consciously trying to maintain voicing through the labial stop closure. Each of the phrases was produced five times in a random sequence. Thus there was a total of 95 utterances (eight precursors, each followed by either *bins* or *pins*, plus three deliberately voiced precursors followed by *bins*, each produced five times). They were recorded in a sound-proof booth, sampled at 20 kHz, low-pass filtered at 9.6 kHz, and edited to remove the final *bins* or *pins* and the preceding stop closure. (Somewhat unintentionally, a significant amount of closure voicing was retained in tokens of *Take three.*)

One token each of *bins* and *pins* was used to construct a new VOT continuum, following the same procedures as in Experiment 1. The VOTs of the six stimuli used in the experiment were approximately 14, 23, 32, 40, 49, and 58 ms. Each of these stimuli was preceded by each precursor and also occurred five times in isolation, for a total of 600 stimuli, which were arranged in five blocks of 120 such that each unique stimulus occurred exactly once in each block. (That is, different blocks used different tokens of the same precursors.) There was a different random order within each block. The duration of the silent interval between precursor and test word (the stop closure) was fixed at 120 ms. Interstimulus intervals were 2 s, with longer intervals from time to time.

Procedure. The procedure remained the same as previously: The subjects identified the initial consonant of each test word as "B" or "P."

Results and Discussion

The results are summarized in Figure 4, which plots the average VOT boundaries for all precursor conditions. The left-most bar represents the boundary for isolated test words,

---

pilot study (Figure 1) is closer to the asymptotic boundary following precursors in Experiment 3 (Figure 3), peak performance in a discrimination task using the very same isolated stimuli (Repp & Lin, 1989) occurred at an even shorter VOT. Such a dissociation of labeling and discrimination performance is unusual in the categorical perception literature and suggests that the boundary for isolated stimuli was influenced by as yet undetermined factors.

[4] We chose meaningful phrases to increase the naturalness of the materials, at the cost of relinquishing some control over their phonetic composition. It was assumed that the segments preceding the final consonant of the precursor would not exert any effect on VOT perception.

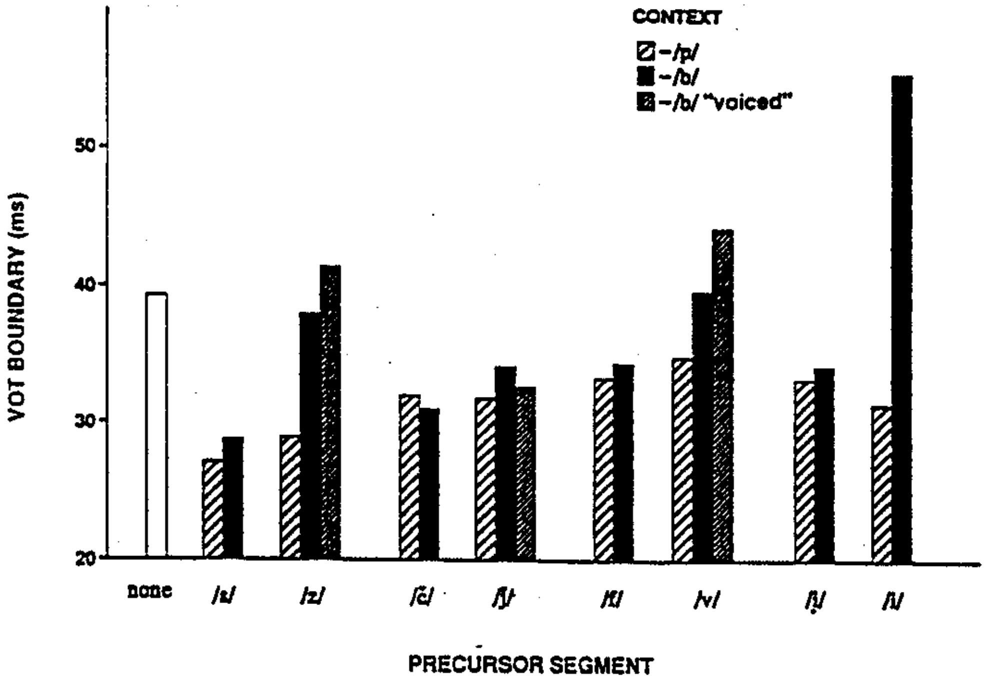

Figure 4. VOT boundaries as a function of final precursor segment and its original context in Experiment 4.

against which the various precursor effects may be compared, with due caution.[5]

The first result to note is that all phonologically voiceless precursors (/s/, /č/, and /f/, as well as whispered /i/) shifted the VOT boundary to shorter values, regardless of their original context. The shift following /s/ was somewhat larger than the others. Second, phonologically voiced precursors from the original *pins* context—/z(-p)/, /ǰ(-p)/, /v(-p)/—had exactly the same effect. Third, phonologically voiced precursors from the original *bins* context had very divergent effects, depending on the segment: /ǰ(-b)/ behaved like a voiceless segment, shifting the boundary down; /z(-b)/ and /v(-b)/ left the boundary unchanged, but deliberately voiced /z(-b)/ and /v(-b)/ caused a small upward shift; and /i(-b)/ caused a very large upward shift. In other words, there were large effects of anticipatory coarticulation in phonologically voiced precursor segments, except for /ǰ/, whereas phonologically voiceless precursor segments showed no coarticulation effects at all.

These patterns were shown to be statistically reliable in analyses of variance on the response percentages. In one analysis, isolated stimuli and the deliberately voiced precursor conditions were excluded, so that the remaining 16 conditions could be treated as representing the following three crossed factors pertaining to the final segments of the precursors: place of articulation (four levels), phonological voicing status (two levels), and original context (two levels). Because all main effects and interactions were significant, we conducted separate analyses on the effects of phonologically voiceless and voiced segments. For voiceless segments, the main effect of place of articulation reached significance, $F(3, 27) = 3.38$, $p < .04$, $MS_e = 14.44$, owing to the shorter VOT boundaries following /s/, but there were no significant effects involving original context (i.e., coarticulation). For voiced segments, there were highly significant effects ($p < .0001$) of place of articulation, of original context, and of their interaction, due to the abnormal behavior of /ǰ/ and to the very large original context effect for /i/. Another analysis was conducted comparing only the voiced and deliberately voiced fricative precursors. The effect of deliberate voicing was significant $F(1, 9) = 5.94$, $p < .04$, $MS_e = 3.64$, but in addition the main effect of place of articulation and the interaction were highly significant ($p < .0001$), again because /ǰ/ behaved like a voiceless fricative.

These results, in conjunction with the acoustic analysis reported shortly, contribute considerably to our understanding of the precursor effects. First, they suggest that any voiceless segment preceding the stop closure causes a VOT boundary shift in favor of /p/ responses. This rules out the hypothesis (already called into question in Experiment 3) that the effect of a preceding /s/ was entirely due to the possibility of its forming a word-initial cluster with /p/. The effect of a preceding /s/, however, was somewhat larger than that of /č/, /f/, or whispered /i/. Thus there may have been some additional contribution of word boundary ambiguity, although the difference could also have been due to the specific acoustic properties of /s/. However, the boundary shift

---

[5] It should be noted that the boundary for isolated stimuli in this experiment was at a much shorter VOT value than in Experiments 1 and 2, was closer to that obtained in the pilot study (Figure 1), and was consistent with the asymptotic boundary in the precursor conditions of Experiment 3 (Figure 3). See also Footnotes 1 and 3 and later experiments.

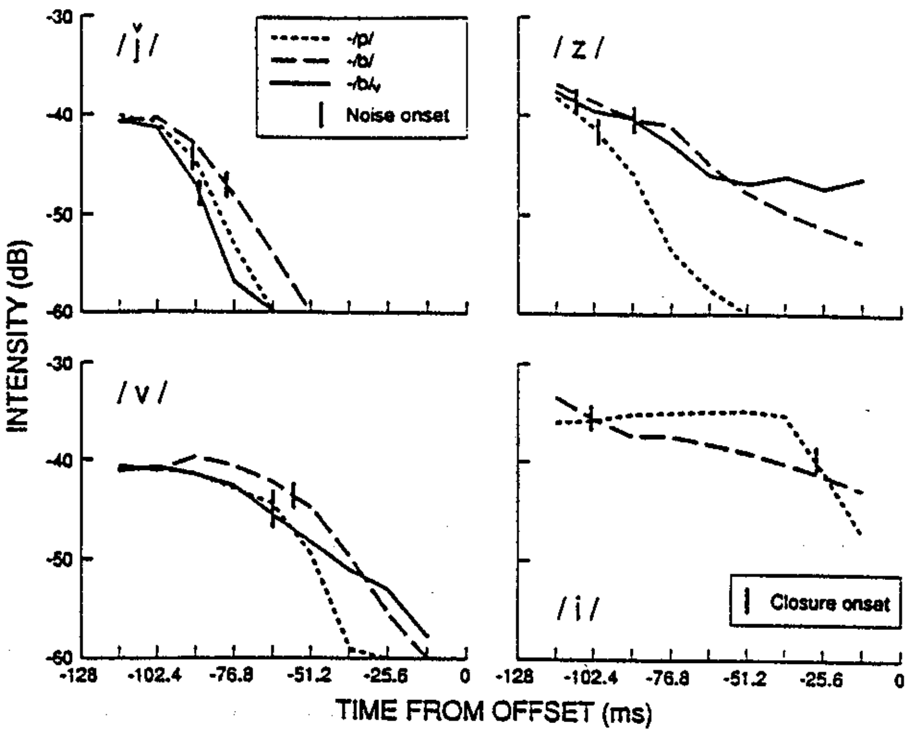

Figure 5. Intensity of the fundamental frequency during the final 128 ms of precursors ending in phonologically voiced segments, averaged over the five tokens of each type. Each data point derives from a window extending 12.8 ms to either side. The solid function represents the deliberately voiced condition. The vertical bars signify the average onset of aperiodic frication detectable in the waveform: for /i/, they indicate the beginning of the stop consonant closure.

seemed to be primarily caused by the phonetic voicing status of the final segment of the precursor.

Precursors ending in a phonologically voiced fricative or affricate originally preceding /p/ had the same effect as phonologically voiceless precursors. The presumable reason for this was phonetic devoicing of the final segment. This was confirmed by an acoustic analysis of the precursors. The presence and strength of voicing during the final 128 ms of each precursor token were determined by measuring the amplitude of the spectral peak corresponding to the fundamental frequency (if any) in the discrete Fourier transforms of nine successive 25.6-ms waveform segments overlapping by 50%. The resulting curves of voicing amplitude as a function of time, averaged across the five tokens of each precursor, are shown in Figure 5 together with an approximate indication (based on waveform measurements) of where the fricative noise (or, in the case of /i/, the closure) began. It is evident that in /ǰ(-p)/, /z(-p)/, and /v(-p)/ (dotted lines), voicing ceased soon after the onset of frication. Moreover, the abnormal results for /ǰ/ preceding /b/ were obviously due to the fact that the speaker never maintained voicing through this affricate (Figure 5, upper left panel). In /z(-b)/ and /v(-b)/, however, voicing continued with diminishing strength through the fricative noise,[4] and in the deliberately voiced segments the amplitude of the fundamental was elevated further at the end. Clearly, therefore, the presence and degree of voicing in the final frication were responsible for the VOT boundary shifts following the various precursors. It is worth noting, though, that the most relevant acoustic parameter was apparently neither the voicing amplitude at precursor offset—in which case /z(-b)/ should have caused a larger upward shift of the VOT boundary than /v(-b)/ (cf. Figure 4)—nor, in the case of devoicing, the distance of voicing offset from the end of frication [in which case /v(-p)/ should have caused a smaller downward shift than /ǰ(-p)/ and /z(-p)/, not to mention /s/, /f/, and /č/]. Rather, the acoustic parameter most predictive of the perceptual results was the slope of the decrease in voicing amplitude: The more gradual the decrease, the more likely were subjects to perceive voiced stop consonants following the precursor.

This observation is further supported by the measurements for /i/ (Figure 5, lower right-hand panel). The perceptual effect of /i(-p)/, which equalled that of a voiceless fricative, seems paradoxical in view of the facts that voicing persisted right up to the end of the precursor (which included about 25 ms of closure voicing) and that the amplitude of its fundamental was higher than in voiced fricatives (and than in /i/ (-b)/, except at the very end). What /i(-p)/ had in common with the devoiced fricatives and affricates, however, was the

---

[4] The data for /z(-b)/ are not representative of individual tokens: Two were devoiced, like /z(-p)/, and three were more strongly voiced, like /z(-b)/$_v$.

rapid falloff in voicing amplitude, even though it occurred much later in the stimulus. The /i(-b)/ precursor, on the other hand, had a much more gradual decrease in voicing amplitude, which took place during the approximately 100 ms of voiced closure that happened to have been included and which apparently caused the striking increase in voiced stop responses, despite the intervening 120 ms of silence. (That closure voicing at the beginning of a closure separating two stop consonants can influence the perceived voicing of the second stop consonant has been shown previously in Dutch by Berg, 1986.)

This identification of a rate parameter as a possible cause of the precursor effect raises the interesting possibility that the fairly large speaking rate effects observed by Summerfield (1981) were in fact partially mediated by the rate of change in voicing amplitude at precursor offset. The common finding is that voiced stop consonant responses to the critical VOT continuum increase as the precursor rate is slowed down. In Summerfield's experiments, changes in speaking rate were simulated by changing the time base of a speech synthesis program, which naturally also changed the amplitude curve associated with the fundamental at precursor offset (assuming that the amplitude of the final vowel was not constant). Diehl, Souther, and Convis (1980) used a similar method with precursors ending in /z/ and found only a very small effect: judging from their spectrograms, the voicing decay at precursor offset may not have been changed substantially. Miller, Green, and Schermer (1984) used natural speech precursors produced at different rates but kept the word *the* preceding the test words constant: they found only a very small rate effect. In an older, unpublished experiment, Summerfield (1975) varied only the steady-state portions of a synthetic precursor and also obtained only rather small rate effects on VOT perception. Kidd (1989), who used natural-speech precursors ending in *the*, found somewhat larger effects of precursor rate. However, the only truly large VOT boundary shifts (about 10 ms) contingent on variations in precursor rate were obtained by Summerfield (1981), whose precursors may have varied in the rate of voicing decay at offset. Therefore, this local phonetic factor may have contributed to the rate effect in that study.

At this stage in our research, the inevitable question arose of whether the observed phonetic effects are caused by acoustic properties alone or whether such effects depend on the precursors being perceived as speech. The remaining experiments were intended to address this vexing issue that permeates contemporary speech perception research.

Experiment 5: Noise Precursors

It is notoriously difficult to dissociate genuinely speech-specific factors from purely auditory ones. A situation in which the same stimulus can be interpreted by listeners as either speech or nonspeech (cf. Best, Morrongiello, & Robson, 1981; Tomiak, Mullennix, & Sawusch, 1987) would be most suitable for this purpose but very difficult to contrive in the present precursor paradigm. We asked, instead, to what extent certain acoustic properties of the preceding speech segment—namely, its duration, intensity, and amplitude envelope—affect VOT perception at an auditory level. We investigated this by replacing the speech precursors with nonspeech noises sharing these properties but having no spectral structure (cf. Gordon, 1988). Because all precursors here were aperiodic noises, we expected that they would all cause a downward VOT boundary shift relative to a no-precursor baseline and that any differences among precursors in the magnitude of these shifts would reveal the role of the acoustic properties mentioned. This experiment, then, investigated the possible role of factors other than the presence of voicing near the end of the precursor.

Method

Subjects. Nine new subjects from the same general pool participated.

Stimuli. Stimuli, test arrangement, and procedure were exactly the same as in Experiment 4, except that each speech precursor was replaced with amplitude-modulated broadband noise. This noise was generated from each individual speech precursor by randomly reversing the polarity of sampling points with a probability of .5 (Schroeder, 1968). Thus, each noise had exactly the same duration, overall intensity, and amplitude envelope as the speech from which it was derived, but it lacked spectral structure and was not identifiable as speech.[7]

Results and Discussion

The results are easily summarized. The average VOT boundaries for all 19 precursor conditions, as well as for isolated stimuli, were within 3 ms of each other, and the analysis of variance revealed no significant differences whatsoever. Thus, the noise precursors did not affect the VOT boundary at all.

One possible explanation of this result is that the listeners dissociated the nonspeech precursors from the following speech, attributing them to a different sound source (cf. Gordon, 1988). This interpretation would negate relatively peripheral auditory interactions as an explanation of the earlier VOT boundary shifts. An alternative explanation, however, is that the broadband noise precursors lacked precisely the acoustic property that was responsible for the speech precursor effects—namely, periodicity. Although we attributed the downward shift of the VOT boundary caused by precursors ending in voiceless fricatives (Experiment 4) to the lack of periodicity at precursor offset, perhaps this effect was caused by the early and abrupt cessation of voicing rather than by its absence later on. If so, precursors totally devoid of periodicity would naturally be ineffective. However, this explanation conflicts with the finding in Experiment 4 that the effect of whispered precursors was similar to that of precursors ending in a voiceless fricative.

---

[7] Following the random sample-reversal procedure, the noise precursors had a flat spectrum, like white noise. However, because they were converted into sound at the same time as the digitized test words, which had been input with high-frequency pre-emphasis, they underwent high-frequency de-emphasis at the D/A conversion stage. Thus their actual spectrum had a falling slope (about 6 dB per octave above 1000 Hz and less below).

A third possibility is that precursors ending in voiceless segments simply have no effect at all, that such an effect was wrongly suggested by the unstable boundary for isolated test stimuli, and that all true precursor effects are upward shifts of the VOT boundary due to the presence of voicing at precursor offset. This interpretation, however, is at odds with the results of Experiments 1–3, which show a decrease over time in the effect of a precursor ending with /s/.

A final explanation is that the noise precursors in this experiment did have an effect, but that the boundary for isolated stimuli was at too short a VOT value. Indeed, the VOT boundary for isolated test words was at 29 ms of VOT, 10 ms shorter than the boundary for the very same stimuli in Experiment 4 and in fact close to the /s/ precursor boundaries in that experiment. Had the boundary for isolated stimuli been in the same place as in Experiment 4, we could have concluded that the preceding noise caused a boundary shift, just like the voiceless fricative precursors. It is becoming increasingly clear that the general experimental context has a substantial effect on the VOT boundary for isolated test words and that comparisons with this baseline are therefore problematic. Apparently, the presence of nonspeech noise precursors in the present test caused the baseline boundary to move toward a much shorter VOT.

Experiment 6: Speech and Noise Precursors

Experiment 6 attempted to circumvent this problem by including both speech precursors ending in voiceless fricatives and nonspeech precursors in the same test. It was expected that this manipulation would place the baseline boundary between the locations observed in Experiments 4 and 5. The question was whether the boundaries in all precursor conditions would be at shorter VOTs.

Method

Subjects. Eleven subjects from the same general pool participated, several of whom had also served in earlier experiments.

Stimuli. In addition to isolated test words, the *Take six*, *Take such*, and *Take tough* precursors of Experiment 4 (from the *bins* context only) were used, as well as their amplitude-modulated noise versions from Experiment 5. Six precursor conditions multiplied by six test stimuli plus isolated test words resulted in 42 stimuli altogether, which were presented five times in random order with different tokens in each block. The procedures were the same as in the previous experiments.

Results and Discussion

Our expectation concerning the boundary for isolated stimuli was indeed confirmed: The boundary was at about 33 ms of VOT, almost halfway between the boundaries observed in Experiments 4 and 5. However, the category boundaries in all precursor conditions fell within 2 ms of each other and did not differ significantly from each other. This was surprising in view of the differences among the three speech precursors obtained in Experiment 4. Even more surprising, however, was the fact that there was absolutely no difference between the precursor and baseline boundaries.

To highlight the contrast with the previous findings, the results of Experiment 6 are plotted in Figure 6 together with the corresponding data from Experiments 4 and 5. The uniformly longer boundaries in Experiment 6 than in Experiment 5 could simply reflect the difference in subject groups; how-

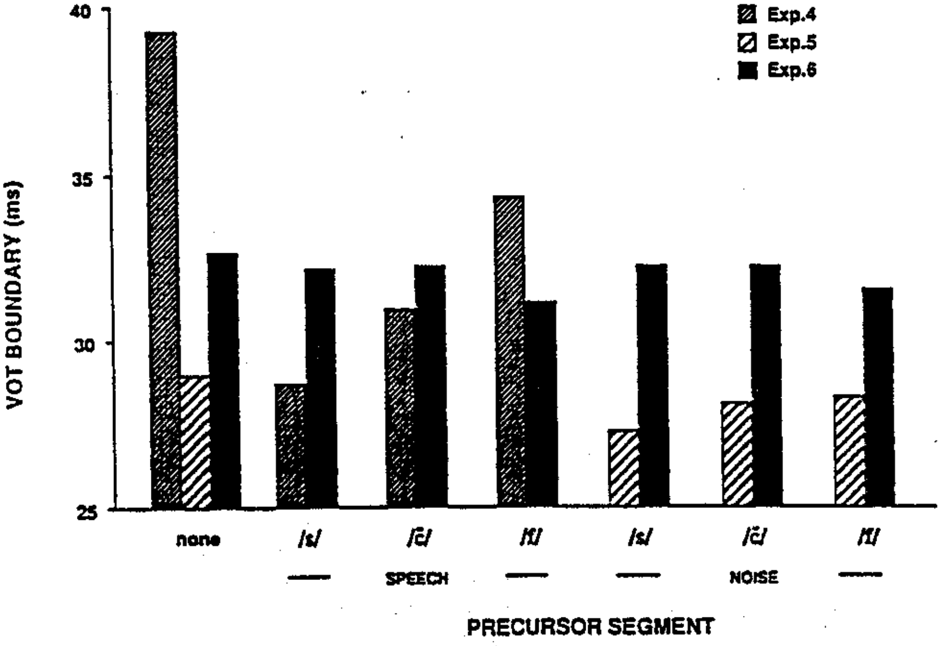

Figure 6. Results of Experiment 6 in relation to selected results from Experiments 4 and 5: VOT boundaries as a function of precursor condition.

ever, the ineffectiveness of the speech precursors in Experiment 6 relative to Experiment 4 is puzzling. One possible explanation for these results is that the presence of nonspeech precursors in the test led subjects to use an attentional strategy of dissociating *all* precursors from their following context. Such a voluntary segregation of components of a speech stream seems implausible. Alternatively, if the absence of differences among the speech precursors is disregarded, it is still possible that the boundary for isolated stimuli was at "too short" a VOT in Experiment 6 and that all precursors "really" caused a downward shift in the boundary. A final attempt to circumvent this baseline problem was undertaken in Experiment 7 by using a separate test containing only isolated test stimuli.

Experiment 7: Filtered Speech Precursors

Experiment 7 had two purposes: First, to obtain an independent baseline estimate and to assess directly the influence of various precursor environments on the baseline boundary itself; second, to examine the role of low-frequency energy in the precursor effect. By focusing on the auditory correlate of voicing, this experiment was complementary to Experiments 5 and 6. Experiment 7 used precursors with final voiced segments as well as precursors with final voiceless segments, so a precursor effect could be established by comparing these two types without reference to a baseline (as in Experiments 1–4).

The experiment included a separate baseline test and four conditions with precursors; each also included isolated stimuli. The first condition partially replicated Experiment 4: the precursors chosen were *Take five* and *Take these*, from either the original *bins* or the *pins* context. Thus, the comparison was between phonetically voiced and voiceless versions of the same phonological structures, which had yielded significantly different boundaries in Experiment 4. We expected these results to be replicated here: the question was where the boundary for the embedded isolated stimuli would be located relative to that obtained in the separate baseline test.

In the second condition, the precursor phrases were high-pass filtered (HPF) to remove the fundamental frequency and its lowest harmonics. The result was intelligible speech of telephonelike quality. Although, as is well known, physical absence of the fundamental frequency does not prevent its perception in a complex periodic sound (Schouten, 1970), this condition served to investigate whether low-frequency energy as such is involved in the precursor effect. If it is, the difference between phonetically voiced and voiceless precursors should be reduced or eliminated by high-pass filtering.

In the third condition, the precursor phrases were low-pass filtered (LPF), so that practically only the fundamental frequency remained. The precursors retained some of the time and amplitude structure of the originals, but they were barely intelligible as speech. If low-frequency energy as such is important, particularly its amplitude contour at precursor offset, then these precursors should still produce effects similar to those of the originals.

The fourth condition was intended to investigate the possibility that low-frequency voicing information is effective only when it occurs in the context of amplitude-modulated high-frequency energy, as in speech. To this end, new precursors were constructed by combining the LPF stimuli of the third condition with nonspeech noises derived from the same original precursor phrases (LPF + N). These noises (similar to those used in Experiment 5) were first high-pass filtered to prevent auditory masking of the low-frequency information by the noise. We thought that addition of this telephone-quality noise to the voicing might enhance the effectiveness of the voicing information by providing a more precise definition of voice offset time and by making the stimulus somewhat more speechlike (though not quite intelligible). On the other hand, we realized that, to the extent that there is an independent effect of voicelessness in the precursor, addition of the noise might cause a boundary shift to shorter VOTs.

All four conditions included isolated *bins/pins* stimuli to further examine shifts in their boundary across conditions. The hypothesis suggested by Experiments 4–6 was that the overall relative amount of voicing in the stimulus ensemble affects the baseline boundary. If so, the boundary should be at the longest value in Condition 1 and at the shortest in Condition 4, with the results for Conditions 2 and 3 depending on whether low-frequency energy as such is important.

Method

Subjects. Twelve new subjects from the same general pool participated.

Stimuli. In Condition 1 (intact speech precursors), the carrier phrases used were *Take five* and *Take these* from the *bins* and *pins* contexts.[8] For Condition 2 (HPF precursors), the precursor phrases were digitally filtered with a lower cutoff frequency of 1000 Hz and stop-band attenuation of 50 dB, which ensured a steep slope. For Condition 3 (LPF precursors), the upper cutoff frequency was set at 500 Hz, and the stop-band attenuation was 50 dB. For Condition 4 (LPF + N precursors), amplitude-modulated noise versions of the precursors were prepared as in Experiment 5; subsequently, they were high-pass filtered (like the speech precursors for Condition 2) and digitally added to the LPF precursors of Condition 3.

Procedure. The baseline sequence of isolated test words was presented first, followed by the four experimental conditions and, finally, a repetition of the baseline sequence. The order of the four conditions was counterbalanced across subjects. Otherwise, the procedure was the same as in the previous experiments.

Results and Discussion

The results are shown in Figure 7 in terms of average VOT boundaries. Consider first the results for isolated stimuli. The boundaries in the pretest and posttest were at 23.1 and 25.6 ms of VOT, respectively. The difference was nonsignificant, $F(1, 11) = 3.47$, $p < .09$. These boundaries are shorter than in any preceding experiment, and in fact are close to what one would expect for stimuli along a bilabial VOT continuum.

---

[8] The deliberately voiced versions (see Experiment 4) were not used because they differed somewhat in intonation and overall amplitude from the "natural" versions, contextual variants of which sounded very similar except for the presence or absence of final fricative devoicing.

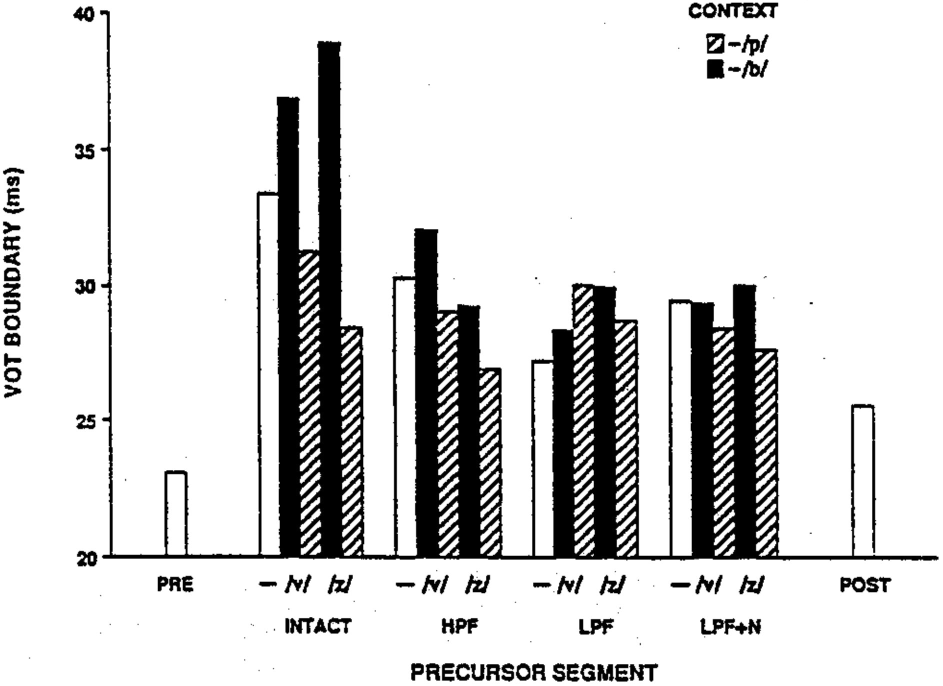

Figure 7. VOT boundaries as a function of various precursor conditions in Experiment 7.

This finding alone suggests that all previous "baseline" boundaries were artificially elevated because of some influence of general test environment.

Indeed, the boundaries for the isolated stimuli embedded in the present precursor test sequences were also at longer VOTs, with the longest boundary in Condition 1 (intact precursors). A one-way ANOVA across the six VOT boundaries for isolated stimuli showed a highly significant difference, $F(5, 55) = 7.77$, $p < .0001$, $MS_e = 20.54$. Subsequent Newman-Keuls tests showed the boundary in the intact condition to be significantly ($p < .01$) longer than the pre- and posttest boundaries and also longer than the boundary in the LPF condition. Furthermore, the HPF and LPF + N boundaries were significantly longer than the pretest boundary. Only the LPF boundary did not differ significantly from the pre- and posttest boundaries. Essentially, each boundary fell into the range of the boundaries following the precursors in that condition (see Figure 7).

With regard to the precursor conditions, it is evident from Figure 7 that only the intact condition (Condition 1) generated a substantial precursor effect. A 2 × 2 ANOVA (Voicing × Place of Articulation of the precursor-final segment) on these data revealed a significant effect of voicing, $F(1, 11) = 20.24$, $p < .001$, $MS_e = 38.39$, with longer VOT boundaries following voiced than voiceless segments and a significant Voicing × Place interaction, $F(1, 11) = 17.41$, $p < .002$, $MS_e = 4.26$, due to the voicing effect being larger for /z/ than for /v/. These data are in good agreement with those of Experiment 4 (Figure 4). The effect of precursor voicing was much smaller in the HPF condition, but was still reliable, $F(1, 11) = 6.51$, $p < .03$, $MS_e = 12.58$. Instead of an interaction, there was a main effect of place, $F(1, 11) = 7.40$, $p < .02$, $MS_e = 9.50$, owing to longer boundaries following /v/ than following /z/. All HPF boundaries were in the same range as the boundaries following voiceless intact precursors. The boundaries in the LPF and LPF + N conditions were also in the same range. There were no significant LPF precursor effects. However, a small effect of voicing emerged again in the LPF + N condition, $F(1, 11) = 5.72$, $p < .04$, $MS_e = 5.83$.

In summary, these data show that only the voicing information in intact speech precursors has a substantial effect on the VOT boundary: the presence of low-frequency energy in the precursor seems to be necessary but not sufficient (though a small effect was obtained with HPF precursors). Because high-pass filtering merely degrades phonetic voicing information, whereas low-pass filtering largely eliminates it and the addition of noise restores it only minimally, the results suggest that it is the phonetic voicing information in the precursor-final segment, rather than the simple auditory property of low-frequency energy, that causes the precursor effects.

The results also demonstrate convincingly that the boundary for isolated stimuli is unstable and shifts into the range of boundaries caused by the precursors in any particular condition. This shift is invariably toward longer VOTs.

General Discussion

The present series of experiments demonstrates two effects of preceding phonetic context on the perception of stop consonant voicing: an effect of the immediately preceding phonetic segment and an effect of the whole stimulus ensemble in the test sequence.

The effect of the immediately preceding segment is assimilative in nature and appears to derive from the segment's phonetic voicing status. Experiments 2 and 4 show clearly that the phonological voicing category of the precursor segment is irrelevant: When voiced fricatives are devoiced, they exert the same precursor effect as voiceless ones. Also largely irrelevant is the closure duration intervening between precursor segment and test word (Experiment 1: see also McCasland, 1977) and the capability of the precursor segment to form a tautosyllabic cluster with the initial consonant of the test word (Experiment 4). Experiment 4 suggests that the most relevant phonetic property may be the degree of abruptness of voicing cessation in the precursor, and Experiment 7 indicates that only intact speech produces a sizable precursor effect. Perhaps the most interesting finding is the relatively slow temporal decay of the effect (Experiments 2 and 3): Although it decreased in size over the first 300-500 ms, a residual precursor effect extended over several seconds of temporal separation.

These findings imply that listeners base their decision about the voicing status of an ambiguous word-initial stop consonant on a wide temporal interval that includes the segment (or its offset) preceding the stop closure. Such a temporal integration account of voicing perception was elaborated by Chasaide (1985) and also hinted at by Berg (1986). Acoustic cues to the voicing of an intervocalic stop are commonly present before its closure begins, certainly within words (Lisker, 1986), but probably also across word boundaries. The coarticulatory effects observed in Experiment 4 attest to this. The precursor effect to be explained, however, does not rest on proper cues to stop consonant voicing; rather, it reflects voicing properties of the preclosure segment that are only partially conditioned by the following stop consonant. In Repp's (1982) classification, this is more like a context effect than a cue integration effect (or trading relation).

Such context effects are usually explained in one or both of two ways (see Diehl & Kluender, 1989; Repp, 1982): They may reflect a perceptual compensation for certain articulatory or acoustic regularities in speech production or they may be due to a purely auditory processing interaction. Because our knowledge of the processing of complex auditory signals is still quite limited (see, e.g., Yost & Watson, 1987), auditory interactions can never be ruled out with certainty. Experiment 5, at least, shows that variations in the precursor amplitude envelope are irrelevant, and Experiment 7 shows little effect of low-frequency energy as such. Also, the precursor effects obtained over long temporal separations (Experiments 2 and 3) are most likely not due to auditory processing interactions. They do reflect memory for the precursor, however, and we have noted earlier that the temporal decay of the precursor effect is consistent with the two components of auditory memory discussed by Cowan (1984). Clearly, the precursor information that causes the VOT boundary shift is not retained in categorical form, otherwise temporal separation would not have reduced the precursor effect substantially. The memory involved, however, could be for articulatory gestures and states rather than for complex auditory properties, and the dynamic and static components of such an articulatory memory might correspond to the two time constants of the decay function.

If an explanation of the precursor effect were to be found in a regularity of speech production of which listeners have tacit knowledge, that regularity would have to be the devoicing of final voiced consonants when they precede a word beginning with a voiceless consonant. Although such devoicing may occur even preceding a voiced consonant (cf. Experiments 2 and 4), it is not the rule in that context (Haggard, 1978; Ladefoged, 1975). Thus, the occurrence of a voiced fricative at the end of a precursor phrase is a clear indication that the test word begins with a voiced consonant. A voiceless fricative, on the other hand, is not predictive of the voicing of the following segment. These facts suggest that a context effect should be exerted by voiced precursor segments but not by voiceless ones. The present results, however, suggest that both have an effect (see Figures 3 and 7). Thus, they do not provide strong evidence that the precursor effects are specifically linked to the phenomenon of final fricative devoicing.

The results of Experiment 6, in which both speech and nonspeech precursors were used, remain somewhat puzzling. Not only were the effects of the two types of precursors equivalent, but differences among the effects of the speech precursors obtained in Experiment 4 were no longer present. It is probably most parsimonious to consider these results merely as a failure to replicate these differences among precursors rather than as evidence that the subjects, under the influence of the nonspeech precursors, managed to disengage the speech precursors perceptually from the following test words. Previous related research (Repp, 1985) has suggested that phonetic context effects are highly resistant to attentional manipulation. On the other hand, it might be noted that one author (BHR), who served as a pilot subject for some of the experiments, never showed any effects of preceding context on his VOT boundary. There were considerable variations in the magnitude of precursor effects for individual subjects in the experiments, but there was hardly ever a subject who did not show any effect at all. These observations suggest that the precursor effect may depend on subjects' (in)ability to listen analytically to speech.

The present series of experiments revealed a second kind of context effect, due to the test environment, which was unexpected and was initially attributed to the stimuli varying in VOT. Only in Experiment 7 did it become clear that the test stimuli were not abnormal at all and had a VOT boundary at the value expected for labial stop consonants (about 25 ms) when presented by themselves in a separate test. Whenever they were interspersed among stimuli preceded by precursor phrases, their VOT boundary shifted dramatically in favor of voiced responses. The magnitude of this shift varied across experiments, and some of this variation surely must be attributed to variability among subject groups. Nevertheless, there was clearly a relation to the nature of the precursors used in a test: When all precursors were intact speech (Experiments 1-4), the VOT boundary for the "baseline" stimuli was always above 35 ms; this was also true for the asymptotic boundary of stimuli separated from speech precursors by long silent intervals (Experiment 3). Only in the intact precursor condition of Experiment 7 was there a somewhat shorter boundary, 33 ms. A boundary at that value was also obtained when the test contained a mixture of speech and nonspeech precursors (Experiment 6), whereas boundaries of 30 ms or less were obtained in conditions using only nonspeech precursors (Experiment 5) or filtered speech precursors (Experiment 7). Even in those conditions, the boundary still seemed somewhat elevated with respect to that for isolated stimuli in a separate test.

The mechanism causing these boundary shifts remains obscure. Although in Experiments 1–3 the "baseline" boundary was at a rather high value, the data of Experiments 4–7 are consistent with the hypothesis that the baseline boundary fell in the midrange of the other boundaries obtained in the experiment. Thus, the phenomenon could be related to that of adaptation level (Helson, 1964): Subjects were not able to maintain separate perceptual criteria for test stimuli presented in and out of context, and they treated the baseline stimuli (which were relatively infrequent in each test) as if they had been preceded by a precursor—perhaps the one that occurred last. It might be noted here that, although the present test stimuli had a "normal" VOT boundary when presented in a separate test, they did contain residual cues favoring the "voiced" category (see Footnote 1), which may have assumed a more prominent role in the precursor test environment. A clearer elucidation of the causes of this phenomenon requires further research, though it is not clear whether anything important about speech perception will be learned thereby. The finding is of methodological importance, however, as it draws attention to the problem of finding an appropriate baseline condition in studies of context effects.

References

Abramson, A. S., & Lisker, L. (1985). Relative power of cues: Fo shift versus voice timing. In V. Fromkin (Ed.), *Phonetic linguistics* (pp. 25–33). New York: Academic Press.

Berg, R. J. H. van den. (1986). The effect of varying voice and noise parameters on the perception of voicing in Dutch two-obstruent sequences. *Speech Communication, 5,* 355–367.

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29,* 191–211.

Chasaide, A. Ní. (1985). *Preaspiration in phonological stop contrasts.* Unpublished doctoral dissertation, University College of North Wales, Bangor, Wales.

Cowan, N. (1984). On short and long auditory stores. *Psychological Bulletin, 96,* 341–370.

Darwin, C. J., & Seton, J. (1983). Perceptual cues to the onset of voiced excitation in aspirated initial stops. *Journal of the Acoustical Society of America, 74,* 1126–1135.

Diehl, R. L., & Kluender, K. R. (1989). On the objects of speech perception. *Ecological Psychology, 1,* 121–144.

Diehl, R. L., Souther, A. F., & Convis, C. L. (1980). Conditions on rate normalization in speech perception. *Perception & Psychophysics, 27,* 435–443.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43,* 137–146.

Haggard, M. (1978). The devoicing of voiced fricatives. *Journal of Phonetics, 6,* 95–102.

Helson, H. (1964). *Adaptation-level theory: An experimental and systematic approach to behavior.* New York: Harper & Row.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15,* 736–748.

Ladefoged, P. (1975). *A course in phonetics.* New York: Harcourt Brace Jovanovich.

Lisker, L. (1957). Closure duration and the intervocalic voiced-voiceless distinction in English. *Language, 33,* 42–49.

Lisker, L. (1986). "Voicing" in English: A catalogue of acoustic features signaling /b/ versus /p/ in trochees. *Language and Speech, 29,* 3–11.

Lisker, L., & Abramson, A. S. (1970). The voicing dimension: Some experiments in comparative phonetics. In *Proceedings of the Sixth International Congress of Phonetic Sciences* (pp. 563–567). Prague, Czechoslavakia: Academia.

McCasland, G. P. (1977). English stops after /s/ at medial word-boundary. *Phonetica, 34,* 218–228.

Miller, J. L., Green, K., & Schermer, T. M. (1984). A distinction between the effects of sentential speaking rate and semantic congruity on word identification. *Perception & Psychophysics, 36,* 329–337.

Pisoni, D. B. (1977). Identification and discrimination of the relative onset time of two component tones: Implications for voicing perception in stops. *Journal of the Acoustical Society of America, 61,* 1352–1361.

Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin, 92,* 81–110.

Repp, B. H. (1985). Perceptual coherence of speech: Stability of silence-cued stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 11,* 799–813.

Repp, B. H., & Lin, H.-B. (1989). Effects of preceding context on discrimination of voice onset times. *Perception & Psychophysics, 45,* 323–332.

Schouten, J. F. (1970). The residue revisited. In R. Plomp & G. F. Smoorenburg (Eds.), *Frequency analysis and periodicity detection in hearing* (pp. 41–54). Leiden, The Netherlands: Sijthoff.

Schroeder, M. R. (1968). Reference signal for signal quality studies. *Journal of the Acoustical Society of America, 44,* 1735–1736.

Summerfield, Q. (1975). Cues, contexts, and complications in the perception of voicing contrasts. *Speech Perception* [Report series from the Department of Psychology, The Queen's University of Belfast], *2*(4), 99–129.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074–1095.

Tomiak, G. R., Mullennix, J. W., & Sawusch, J. R. (1987). Integral processing of phonemes: Evidence for a phonetic mode of perception. *Journal of the Acoustical Society of America, 81,* 755–764.

Wingate, A. H. (1982). A phonetic answer to a phonological question. *UCLA Working Papers in Phonetics, 54,* 11–27.

Yost, W. A., & Watson, C. S. (Eds.). (1987). *Auditory processing of complex sounds.* Hillsdale, NJ: Erlbaum.

Received June 27, 1988
Revision received May 4, 1990
Accepted May 4, 1990 ■