Perception & Psychophysics
1989, 45 (4), 323-332

Effects of preceding context on discrimination of voice onset times

BRUNO H. REPP and HWEI-BING LIN
Haskins Laboratories, New Haven, Connecticut

When discriminating pairs of speech stimuli from an acoustic voice onset time (VOT) continuum (for example, one ranging from /ba/ to /pa/), English-speaking subjects show a characteristic performance peak in the region of the phonemic category boundary. We demonstrate that this "category boundary effect" is reduced or eliminated when the stimuli are preceded by /s/. This suppression does not seem to be due to the absence of a phonological voicing contrast for stop consonants following /s/, since it is also obtained when the /s/ terminates a preceding word and (to a lesser extent) when broadband noise is substituted for the fricative noise. The suppression is stronger, however, when the noise has the acoustic properties of a syllable-initial /s/, all else being equal. We hypothesize that these properties make the noise cohere with the following speech signal, which makes it difficult for listeners to focus on the VOT differences to be discriminated.

One of the most reliable findings in speech perception research is the so-called "category boundary effect" for stimuli varying in voice onset time (VOT) (Wood, 1976): subjects find it easier to discriminate syllables that fall on opposite sides of the phonemic category boundary on a VOT continuum than stimuli that, although acoustically different to a similar degree, are drawn from within a phoneme category. Such a peak in the discrimination function along an acoustic speech continuum is one of the hallmarks of categorical perception (Studdert-Kennedy, Liberman, Harris, & Cooper, 1970). Two alternative explanations of this effect have been proposed (see reviews by Howell & Rosen, 1985, and Repp, 1984). According to one, there is a psychoacoustic discontinuity along the VOT dimension that gives rise to the discrimination peak, and that also determines the location of the phonemic category boundary (usually between 20–40 msec of voicing lag for English-speaking listeners). According to the other explanation, untrained listeners base their discrimination judgments less on auditory qualities than on the phonemic categories assigned to the stimuli. In this view, the discrimination peak is not due to a psychoacoustic discontinuity but to subjects' attention to a higher-level, discrete representation of the input, and the location of the category boundary is determined by language-specific factors, not universal auditory ones.

There is some evidence that attention to phonological categories plays a role in VOT discrimination tasks. For example, listeners trained to pay attention to the auditory properties of stimuli in low-uncertainty tasks tend to show a reduced category boundary effect or none at all (Carney, Widin, & Viemeister, 1977; Kewley-Port, Watson, & Foyle, 1988; Sachs & Grant, 1976; Samuel, 1977; Soli, 1983; however, see also Macmillan, Goldberg, & Braida, 1988), and speakers of languages whose stop consonant voicing contrasts differ from English may show a category boundary effect at a different VOT than is characteristic for English listeners (Williams, 1977). However, there is also a history of experimentation concerning possible psychoacoustic discontinuities on VOT continua (reviewed by Howell & Rosen, 1985; Repp, 1984; Rosen & Howell, 1987a, 1987b), which has culminated in the finding that nonhuman animals show enhanced sensitivity to VOT differences in the region of the English phoneme boundary (Dooling, Okanoya, & Brown, 1988; Kuhl, 1981; Kuhl & Padden, 1982). The most recent contributions to this issue stem from Kewley-Port et al. (1988) and Macmillan et al. (1988). Kewley-Port et al. found that human listeners trained in a low-uncertainty task exhibited no discrimination peak along a VOT continuum; they concluded that the peak has a phonological origin. However, Macmillan et al., who sampled the continuum more finely in an otherwise similar experiment, found a clear peak at about 18 msec of VOT. Thus the category boundary effect in that region of a VOT continuum does seem to have a psychoacoustic underpinning. At the same time, it is also clear from Macmillan et al.'s work that subjects do make use of "context coding" or labeling in high-uncertainty tasks such as the frequently used ABX paradigm. Such attention to phonological categories might magnify the psychoacoustic discrimination peak (if psychoacoustic and phonological effects are additive) or substitute a peak of different origin (if psychoacoustic and phonological effects are mutually exclusive, resulting from attention to independent internal representations).

This research was supported by NICHD Grant HD-01994 and BRS Grant RR-05596 to Haskins Laboratories. Hwei-Bing Lin was also at the Department of Linguistics, University of Connecticut, Storrs, CT 06268. We are grateful to Leigh Lisker and Neil Macmillan for helpful comments on an earlier version of the manuscript. Send correspondence to: Bruno H. Repp, Haskins Laboratories, 270 Crown Street, New Haven, CT 06511-6695.

The present series of experiments began as an attempt to eliminate the contribution of phonemic labeling to VOT discrimination performance in a high-uncertainty discrimination task. The method, explained below in more detail, entailed preceding the test syllables with a fixed precursor that neutralized the phonological voicing contrast on the following stop consonant. This procedure also raised the question, however, of the extent to which the precursor might interfere with the auditory processing of VOT through some form of forward masking or interference in auditory memory. Several additional experiments were conducted to address this issue. It was expected that an investigation of the relative sensitivity of the VOT category boundary effect to a preceding context would provide new information about its origins in phonemic labeling and/or in the auditory representation of VOT.

EXPERIMENT 1

Well-known methods for manipulating the category boundary effect in speech discrimination include the substitution of analogous nonspeech stimuli, the use of listeners with different instructions or auditory skills, and the use of discrimination tasks with varying memory demands. In Experiment 1, however, the same stimuli were presented to the same listeners in the same task with the same instructions. The critical manipulation concerned the presence or absence of an immediately preceding phonetic context. In one condition, the VOT differences to be discriminated were often phonemically distinctive, whereas in the other they were not. This was achieved by first presenting stimuli from a standard [pa]-[p^ha] (phonemically, /ba/-/pa/) continuum varying in VOT, and by then preceding these stimuli with a constant [s] noise plus silence appropriate for a labial closure interval. In English there is no phonemic voicing distinction for stops after tautosyllabic /s/, and although stop consonants are produced without aspiration in these clusters, the conventional orthographic transcription—and the phoneme category assigned by linguists—is /p,t,k/, not /b,d,g/ (see Lisker, 1984). This fact was exploited previously by Sawusch and Jusczyk (1981) in a study of the auditory as opposed to linguistic origins of selective adaptation and contrast effects along a VOT continuum. For our present purpose the transcription is relevant insofar as it preempts the "voiceless" symbols and thus impedes a categorical distinction between unaspirated and aspirated stops following /s/, at least for listeners without phonetic training. To the unsophisticated listener, both [spa] and [sp^{h}a] are /spa/.

The predictions were thus very straightforward: When asked to discriminate stimuli from the [pa]-[p^ha] series, subjects should exhibit the typical category boundary effect; but for the [spa]-[sp^{h}a] stimuli, the discrimination peak should disappear if it was due to subjects' attention to phonemic categories.[1] If, on the other hand, the category boundary effect is partially or entirely due to a psychoacoustic discontinuity (such as a "temporal order threshold" for VOT—see Pisoni, 1977, but also Rosen & Howell, 1987b), the category boundary effect should persist. This could be either because listeners always make auditory discriminations onto which phonological categories are merely grafted, or because listeners' attention is drawn to auditory differences in the absence of phonological contrast.

Two methodological precautions were taken against possible complications in this simple design. First, a category boundary effect might be obtained for the [spa]-[sp^{h}a] series because subjects fail to integrate the [s] noise with the rest of the syllable. This might occur because, in the rapid successive presentation of stimulus triplets for discrimination, the [s] noises may form a separate acoustic stream (Bregman, 1978; Cole & Scott, 1973). Indeed, Diehl, Kluender, and Parker (1985) have argued that such streaming occurred in the abovementioned study by Sawusch and Jusczyk (1981), and that it may have invalidated some of their conclusions. Sawusch and Jusczyk used an interstimulus interval (ISI) of 300 msec. To counteract stream formation, a relatively long (a 1-sec) ISI was used in the present discrimination task. Since, in addition, there was a 3-sec response interval after each stimulus triplet, auditory streaming was considered quite unlikely under these conditions. That listeners would be able to deliberately ignore the /s/ seemed unlikely in view of Repp's (1985) demonstration that—for untrained listeners at least—it is very difficult to intentionally segregate an initial [s] noise from a following phonetic context.

Second, it is possible that an [s] noise preceding [pa]-[p^ha] stimuli interferes with VOT perception at a strictly auditory level, through some form of forward masking or by increasing the load on auditory memory and thereby making small stimulus differences less accessible. Although it would be surprising if such interference removed the psychoacoustic category boundary effect completely, a lowering of discrimination performance and a consequent reduction of the boundary peak might occur. To assess this possibility, a third condition was included in the experiment, in which a burst of white noise preceded the [pa]-[p^ha] stimuli. The white noise was chosen to be at least equal in energy to the [s] noise across the whole frequency spectrum, so that its auditory interference with VOT perception would be at least equal to that caused by the [s] noise. The subjects, however, were expected to hear these stimuli as a nonspeech noise followed by /ba/ or /pa/, so a phonological category boundary effect should be obtained, though perhaps attenuated by auditory interference, which indirectly would hamper labeling accuracy. Any additional reduction in the category boundary effect in the [s] noise precursor condition relative to the white noise precursor condition might then be attributed specifically to the neutralization of the phonological contrast, and hence to subjects' attention to linguistic stimulus attributes.

Method

Subjects. Twelve Yale undergraduates were paid to participate. They were all native speakers of American English.

Stimuli. The [pa]-[p^ha] CV continuum was constructed on the Haskins software synthesizer in its serial configuration. Using conventional procedures, VOT (the duration of the initial aperiodic excitation) was varied from 0 to 70 msec in 10-msec steps, resulting in eight stimuli. The syllables did not have any release bursts. In the [spa]-[sp^ha], or [s]-CV, continuum the stimuli were prefixed with a 58-msec natural-speech [s] noise.[2] A 60-msec silent interval intervened between the constant [s] noise and each synthetic syllable. In the noise-CV condition, a 58-msec noise burst preceded each syllable by 60 msec. This noise burst was excerpted from broadband noise recorded from a General Radio 1390-A noise generator, and its amplitude was adjusted until its flat spectral envelope (obtained by Fourier analysis) completely subsumed the typical ʃ-shaped envelope of the [s] noise, obtained at its most intense point. Consequently, the two noises were similar in energy around 4 kHz, but the white noise had stronger low-frequency components than the [s] noise. In addition, the white noise had an abrupt onset and offset, whereas the [s] noise was naturally tapered. These differences were thought to enhance any auditory interference due to the white noise, relative to that caused by the [s] noise. The test of the phonological hypothesis was thus rather conservative.

All stimuli were digitized at a 10-kHz sampling rate with appropriate low-pass filtering at 4.9 kHz. In each of the three stimulus conditions, all two-step (20-msec VOT) pairings of the stimuli were presented in an AXB format. This led to 6 (stimulus pairs) × 4 (arrangements within an AXB triad) = 24 stimulus triplets, which were recorded three times in random order on magnetic tape. The ISIs were 1 sec within triplets, 3 sec between triplets, and 10 sec between blocks.

Procedure. The tapes were played back binaurally over TDH-39 earphones in a quiet room at a comfortable intensity. Each subject listened first to the CV series, which was preceded by four easy practice trials. The task was to listen carefully to the onsets of the syllables, and to write down "A" or "B," depending on whether the second stimulus in an AXB triplet matched the first or the third, which were known to be always different from each other. Subsequently, half the subjects listened to the [s]-CV series and then to the noise-CV series, and the other half listened in the reverse order. They were told that exactly the same differences were to be discriminated, but that all the syllables would be preceded by a constant [s] or noise burst, which was to be ignored.

Results and Discussion

The results, averaged across subjects, are shown in Figure 1. As expected, the discrimination function for the CV continuum exhibited a pronounced peak, suggesting a category boundary at approximately 24 msec of VOT. By contrast, in the [s]-CV condition, there was no peak at all: the discrimination function was fairly flat and performance was poor, though above chance (except for the 50/70 stimulus pair). Finally, in the noise-CV condition, there was a peak, but it was lower and narrower than the peak in the CV condition, mainly because of a large reduction in correct responses for the 10/30 pair. Note that, with the exception of the 50/70 stimulus pair, the performance decrements were restricted to the region of the original peak, even though there was some room for decrements elsewhere.

A repeated-measures analysis of variance was conducted on the response percentages, with the factors stimulus pair and condition. Both factors had significant main

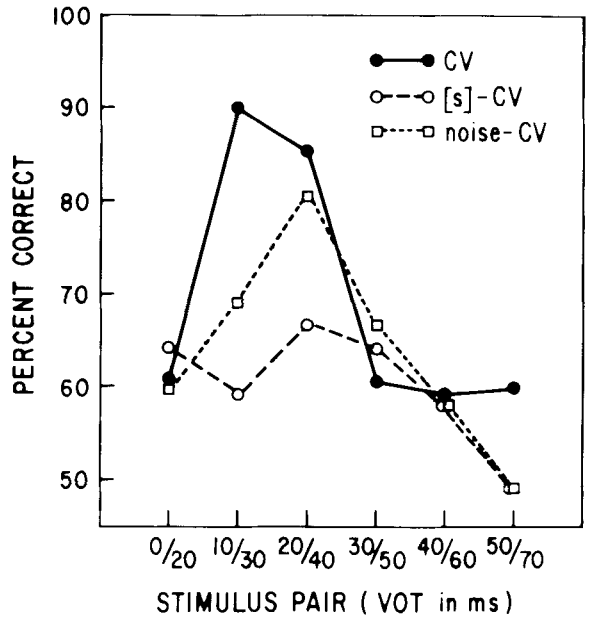

Figure 1. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 1.

effects [$F(5,55) = 9.71, p < .0001$, and $F(2,22) = 6.33$, $p = .0067$, respectively], and their interaction was significant as well [$F(10,110) = 3.02, p = .002$]. Separate analyses of pairs of conditions revealed that, although the difference between the CV and noise-CV conditions was statistically reliable, that between the noise-CV and [s]-CV conditions was not, due to considerable variability among the subjects.[3]

It is clear from these results that a preceding nonspeech noise interferes with the discrimination of VOT near the category boundary. If this interference takes place in auditory processing, then the [s] noise presumably generates a similar auditory disturbance. The complete disappearance of the category boundary effect in the [s]-CV condition is consistent with the phonological hypothesis, which claims that when no phonemic contrast is perceived, there is no category boundary effect. However, the absence of a statistically reliable difference between the two precursor conditions raises the possibility that both types of noise had their effects at auditory levels of processing. Alternatively, it is possible that the subjects interpreted (consciously or unconsciously) the white noise as a fricative (for example, [ʃ]), or perceptually "restored" a fricative noise potentially hidden in the white noise (see Warren, 1984), which also would have attenuated the difference between the two precursor conditions. In that case, the results would be entirely consistent with the phonological hypothesis.

In addition to this ambiguity of interpretation, the comparison between the [s]-CV and noise-CV conditions was inherently problematic because the noises differed in a number of acoustic properties. These difficulties are endemic to studies using nonspeech substitutes for speech sounds. It was decided, therefore, to conduct a second

experiment in which the difference between two precursor conditions was in the context preceding the /s/.

EXPERIMENT 2

In Experiment 2, the syllables carrying the VOT variation were preceded by exactly the same [s] noises in both precursor conditions, but in one of the conditions, the preceding context made the /s/ the initial phoneme of the test word, while in the other condition, the context made it appear to be the final phoneme of a preceding word. In the first context the phonemic voicing contrast was thus neutralized, whereas in the second it was not. The prediction of the phonological hypothesis was therefore that the category boundary effect should be absent in the first condition but not in the second. The auditory interference hypothesis, on the other hand, predicted similar performance in the two conditions, worse than in a control condition without the preceding [s].

Method

Subjects. Twelve new Yale undergraduates were paid to participate. All were native speakers of American English.

Stimuli. This experiment, and all of the following ones, used stimuli constructed entirely from natural speech. A female speaker was recorded saying the phrases *A crazy spin*, *Take this bin*, and *Take this pin*. The speech was digitized at 20 kHz with low-pass filtering at 9.8 kHz. With the help of a waveform editing program, the *bin* and *pin* syllables were excerpted from the *Take this* context, and a VOT continuum was fashioned by replacing initial waveform segments of *bin* with corresponding amounts of aperiodic waveform from *pin*, proceeding in steps of two glottal cycles. This resulted in stimuli with VOTs of 10, 18, 27, 36, 44, 53, and 61 msec (rounded to the nearest msec). An additional stimulus with zero VOT was created by excising the 10-msec release burst of *bin*.

The *pin* portion of *A crazy spin* was excised and discarded, leaving *A crazy s*. To make the [s] noises in the two precursors identical, the [s] of *A crazy s*, 148 msec in duration, was excerpted (without deleting it) and substituted for the shorter (94-msec) [s] noise in *Take this*. When the precursors were recombined with *bin/pin*, this was found to result in more naturally sounding stimuli than the reverse substitution. Thus the [s] noises in both precursors had phonetic properties characteristic of syllable-initial /s/. The *Take this* precursor used was derived from the *bin* context; the other *Take this* carrier phrase was discarded.

Each of the three stimulus conditions contained five blocks of 24 AXB triads presenting two-step discriminations (here corresponding to differences of about 17 msec of VOT). In the CVC condition, the *bin-pin* stimuli occurred in isolation with ISIs of 1 sec within triads. In the *A crazy s*, or #[s-]CVC, and *Take this*, or [s-]#CVC, conditions, they were preceded by the respective constant precursors.[4] The silent interval between the end of the [s] noise and the CVC word was 77 msec, which equalled the original closure interval in *A crazy spin*. The ISIs within triads were 500 msec, to compensate for the longer stimulus durations. The ISIs between triads were 3 sec.

Procedure. The procedure was the same as in Experiment 1, with the CVC condition first and the order of the other two conditions counterbalanced across subjects.

Results and Discussion

The results are shown in Figure 2, which is analogous to Figure 1. In the CVC condition, a pronounced peak in discrimination performance was obtained, suggesting

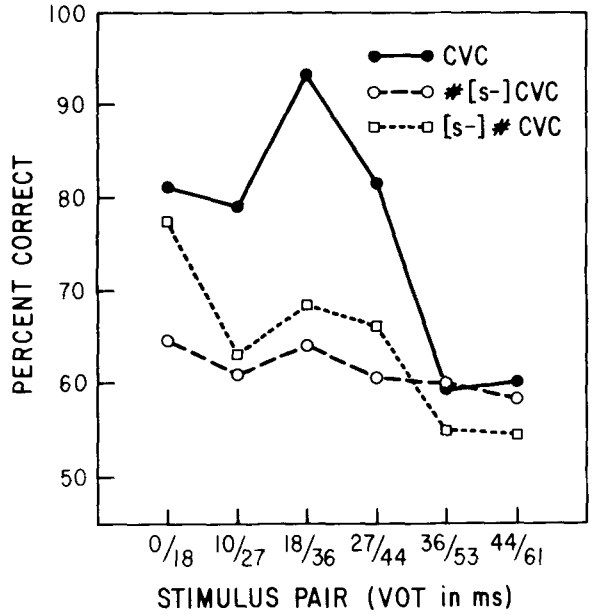

Figure 2. Discrimination performance as a function of stimulu pair in the three conditions of Experiment 2. The precursors ar *A crazy s* (circles) and *Take this* (squares).

a category boundary around 28 msec of VOT.[5] The onl unusual feature of this discrimination function is th elevated performance for the first stimulus pair. This i likely to be an artifact of stimulus construction. It wil be recalled that the zero VOT stimulus was generated b a different method—removal of the release burst—whic gave it a "softer" onset that some subjects found ver distinctive, while others did not seem to notice it. As fo the #[s-]CVC and [s-]#CVC conditions, it is evident tha the discrimination peak was severely reduced or abser in both, and that there was little difference between them except perhaps for the anomalous first stimulus pair.

The statistical analysis revealed significant effects c stimulus pair [$F(5,55) = 12.40$, $p < .0001$] and of cor dition [$F(2,22) = 34.79$, $p < .0001$], as well as a sig nificant interaction [$F(10,110) = 4.62$, $p < .0001$ When the CVC condition was omitted, however, only th effect of stimulus pair was significant [$F(5,55) = 5.0$ $p = .0007$]. The #[s-]CVC and [s-]#CVC results wer thus statistically equivalent.

These results replicate the earlier finding that a prece ing [s] noise severely reduces or even eliminates the VO category boundary effect. However, there was no ind cation of any effect of linguistic structure. A precedi /s/ had the same detrimental effect, whether it initiat the test word or whether it terminated the preceding wor This seems to disconfirm the phonological hypothesis a support the auditory interference hypothesis, although t total disappearance of the boundary peak is somewhat su prising from an auditory perspective.

This interpretation of the results assumes that the phon logical structure was perceived in accordance with t context but proved irrelevant to VOT discrimination.

is possible, however, that the subjects never perceived the intended difference in word-boundary location and heard *Take this bin* as *Take the spin*, even though the instructions said that the precursor was *Take this*. If this seems implausible, it is still possible that VOT discrimination is performed at a prelexical level of phonological coding that depends solely on the phonetic properties of the speech segments and precedes the assignment of lexical word boundaries. The fact that aspiration noise following a stop closure is a word-boundary cue (Christie, 1974) does not contradict this hypothesis: the discrimination of relatively long VOTs was not affected by precursors. Thus the phonological hypothesis is still alive.

EXPERIMENT 3

To the extent that the phonetic properties of the speech segments forced a particular phonological structure on the speech signal, which took precedence over contextual constraints, Experiment 2 did not achieve its purpose. It merely confirmed the basic finding that a preceding [s] noise with syllable-initial phonetic properties eliminates the category boundary effect. Would an [s] precursor that unambiguously terminates a preceding word still interfere with VOT discrimination? And if so, is the interference due to the /s/ at all? Perhaps *any* preceding context would interfere with VOT perception.

In Experiment 3, we examined these two questions. Like Experiment 2, it included three discrimination conditions, one in which the test stimuli occurred in isolation and two in which they were preceded by a carrier phrase. In one instance, the carrier phrase was unambiguously *Take this*, while in the other it was *Take the*.

Method

Subjects. Thirteen new Yale undergraduates were paid to participate. All were native speakers of American English.

Stimuli. The stimuli and test sequence in the baseline CVC condition were the same as in Experiment 2. In the *Take this*, or [-s]#CVC, condition, the original syllable-final [s] noise (94 msec long) and the original silent closure duration following it (115 msec long) were reinstated, which made this context quite unambiguous. A new carrier phrase was recorded for the *Take the* condition. To avoid closure voicing, it was produced as *Take the pin* by the same female speaker, with a silent closure duration of 84 msec. The stimuli from the *bin–pin* continuum were substituted for the original *pin*.

Procedure. As in Experiments 1 and 2, the baseline condition was always presented first, and the order of the two precursor discrimination conditions was counterbalanced across subjects.

Results

The results are shown in Figure 3. The CVC condition again yielded a pronounced peak in the short VOT range. This time, both precursor conditions also showed peaks, but they were somewhat lower and shifted towards longer VOTs. The peak for the *Take the* precursor stimuli was at a longer VOT than that for the *Take this* precursor stimuli.[6]

The ANOVA yielded a significant effect of stimulus pair [$F(5,60) = 31.01$, $p < .0001$], and a significant stimulus pair × condition interaction [$F(10,120) = 8.31$, $p < .0001$], but no significant main effect of condition. In a comparison of only the two precursor conditions, the two significant effects remained significant [$F(5,60) = 14.60$, $p < .0001$, and $F(5,60) = 4.56$, $p = .0014$, respectively]. Thus there was reliable evidence only for shifts in peak location (which we will not attempt to explain here)—not for a general performance decrement caused by precursors.

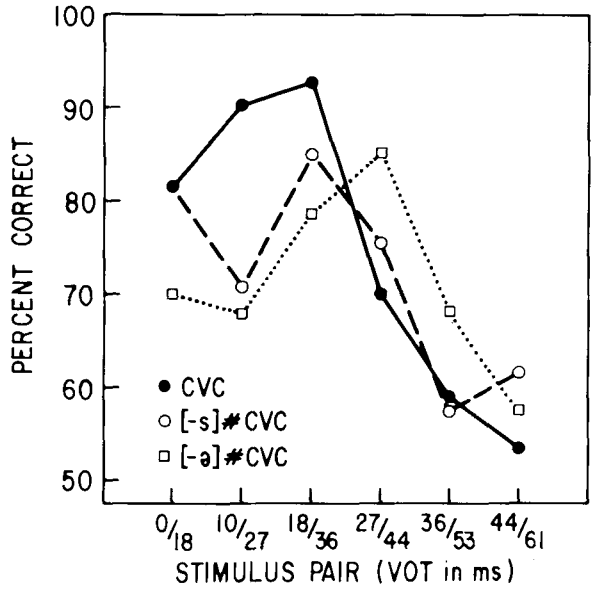

Figure 3. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 3. The precursors are *Take this* (circles) and *Take the* (squares).

The results of this experiment, when compared with those of Experiment 2, show that an /s/ that unambiguously belongs to a preceding word interferes much less with VOT discrimination (if at all) than does an /s/ that has phonetic characteristics appropriate for word-initial position, all else being equal. This is entirely consistent with the phonological hypothesis. However, the auditory interference hypothesis is by no means ruled out. For one thing, the silent closure duration following the syllable-final [s] noise (itself a cue to a syllable boundary) was longer than that following the syllable-initial [s] noise. Naturally, this may have reduced any auditory interference. Then there were also acoustic differences between the two [s] noises in duration, amplitude envelope, and spectral detail that could have played a role. Moreover, the comparison between syllable-initial and syllable-final [s] noises was made across experiments, which is always problematic. Therefore, another experiment was conducted.

EXPERIMENT 4

The purpose of Experiment 4 was to independently assess the roles of [s] noise characteristics and silent closure duration in the precursor effect on VOT discrimination.

Method

Subjects. Twelve Yale undergraduates, some of whom had participated also in Experiment 3, were recruited and paid for their services.

Stimuli. There were five conditions, three of which replicated those of earlier experiments: isolated CVC stimuli; CVC stimuli preceded by *Take this*, with the original 94-msec syllable-final [s] noise plus a 115-msec silent closure (as in Experiment 3); and CVC stimuli preceded by *Take this*, with the spliced-in 148-msec syllable-initial [s] noise plus a 77-msec closure (as in Experiment 2). The two additional conditions represented the other two possible combinations of [s] noise and closure duration. Because of the increased number of conditions, the two extreme stimulus pairs (0/18 and 44/61 msec of VOT) were dropped to reduce test length, leaving only four two-step VOT contrasts. Each condition thus contained 16 different AXB triads, which were repeated five times.

Procedure. As usual, the CVC condition was presented first, and the order of the other four conditions was counterbalanced across subjects. The subjects were told that the precursor was always *Take this*; the phonetic differences among the precursor conditions were not mentioned in advance.

Results and Discussion

The results are shown in Figure 4. The discrimination function for the isolated CVC stimuli replicates that obtained in Experiments 2 and 3. The function for the precursor condition with syllable-final [s] and long closure also resembles that found in Experiment 3, showing but slight interference. The function for the precursor condition with syllable-initial [s] and short closure is quite different in shape from that found in Experiment 2, for unknown reasons. However, it does replicate the much greater interference obtained in that condition. The remaining two conditions, with mismatched [s] noises and closures, yielded results rather similar to the syllable-final [s] plus long closure precursor condition.

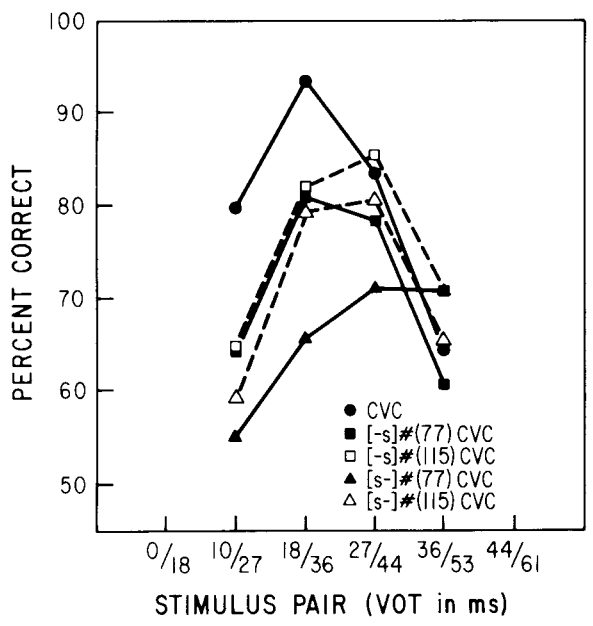

Figure 4. Discrimination performance as a function of stimulus pair in the five conditions of Experiment 4. The symbols [s-] and [-s] represent fricative noises from originally syllable-initial and syllable-final position, respectively. The numbers in parentheses are the closure durations. All precursors were intended to be perceived as *Take this*.

Alternatively, the data can be summarized as follows: all precursors interfered with VOT discrimination, though only at the shorter VOTs. Within the four precursor conditions, the combination of syllable-initial [s] and short closure led to much more interference than any of the other combinations. Thus, for the syllable-initial [s] noise, lengthening of the closure reduced interference considerably; for the syllable-final [s] noise, there was only a slight reduction. Similarly, at the short closure duration, a change in [s] noise made a large difference; at the long closure duration, only a small one.

The statistical reliability of these effects was examined in two ANOVAs. The first analysis included only the four precursor conditions, with noise type, closure duration, and stimulus pair as factors. Apart from the expected main effect of stimulus pair [$F(3,33) = 12.79$, $p < .0001$], there was a significant main effect of noise type [$F(1,11) = 9.47$, $p = .0105$], a marginally significant main effect of closure duration [$F(1,11) = 4.85$, $p = .05$], a marginally significant interaction between closure duration and stimulus pair [$F(3,33) = 2.96$, $p = .0465$], and a significant triple interaction [$F(3,33) = 3.68$, $p = .0217$]. The noise type × closure duration interaction was not significant. The triple interaction reflects the finding that the discrimination function in the long-noise/short-closure condition had a different shape than the functions in the other three conditions. The marginal significance levels indicate considerable variability among subjects.

A second ANOVA compared the isolated CVC condition with the precursor condition closest in performance level (syllable-final noise, long closure). The main effect of condition (the difference in average performance level) was not significant, but the stimulus pair × condition interaction (the difference in shape of the discrimination functions) was highly significant [$F(3,33) = 6.39$, $p = .0016$].

In summary, these results confirm the earlier finding that preceding [s] noises interfere with VOT discrimination as long as the VOTs compared are relatively short (40 msec or less), but not if they are relatively long. In addition, the results show that the interference depends both on [s] noise type and closure duration: The phonetic constellation appropriate for a syllable-initial /s/ leads to more interference at short VOTs than any other noise-closure combination, thus reducing substantially the peak in the discrimination function.

These results are still compatible with both a phonological and an auditory interference account. From the perspective of the phonological hypothesis, they show that only the complete phonetic pattern characterizing syllable-initial /s/ leads to (overt or covert) /s/-stop cluster formation and phonological neutralization of the stop voicing contrast. From the auditory perspective, the two types of [s] noise generated different amounts of auditory interference because of their different acoustic properties

(duration, amplitude, etc.), and this differential interference was reduced at longer temporal separations because of a ceiling effect on discrimination performance.

In a parallel study, we (Repp & Lin, 1988, Experiment 1) collected identification data for the very same stimuli, with five different closure durations ranging from 45 to 150 msec. The subjects labeled the stop consonants as "B" or "P" less accurately when the syllable-initial [s] noise preceded them, with a strong bias towards "P" responses, than when the syllable-final [s] preceded them. This difference decreased only slightly as closure duration increased, and it was still present at the longest closure. This pattern diverges from the present discrimination results, which already show a close convergence at a closure duration of 115 msec. Thus, as closure duration increased, discrimination performance exceeded what would be predicted on the basis of phonemic labeling in the syllable-initial [s] precursor condition. Here is a suggestion that the category boundary effects in that condition, at least, did not derive directly from attention to phonological categories, though it is also possible that covert labeling in the AXB task did not follow the same pattern as overt labeling in the identification test (cf. Repp, Healy, & Crowder, 1979).

In a final attempt to distinguish between the two alternative accounts of the precursor interference effects, in Experiment 5 we returned to the method of nonspeech analogues, in defiance of its inherent problems.

EXPERIMENT 5

In Experiment 5, the entire precursors of Experiment 4 were replaced with broadband noises having exactly the same durations, overall amplitudes, and amplitude envelopes (cf. Gordon, 1988; Salasoo & Pisoni, 1985). Only spectral structure was eliminated. Thus, Experiment 5 also partially replicated Experiment 1, where a more primitive kind of nonspeech noise precursor had been used. To shorten the experiment, only the short-closure conditions of Experiment 4 were included. The prediction was that, if the different amounts of interference generated by the two kinds of [s] noise in Experiment 4 were due to differences in their acoustic properties (other than spectral differences), then the two nonspeech noises likewise should generate different amounts of interference, similar to those produced by the [s] noises. If, on the other hand, the difference between the two [s] noise conditions in Experiment 4 was due to spectral or specifically phonetic factors (that is, /s/-stop cluster formation at some prelexical level in perception), then the two nonspeech noises should have equivalent effects, similar in magnitude to the effect of the syllable-final [s] noise in Experiment 4 or even smaller.

Method

Subjects. Twelve subjects from the same pool participated. Some of them had also been subjects in Experiment 4.

Stimuli. Each of the two entire *Take this* precursors was converted into envelope-matched broadband noise using a computerized procedure first described by Schroeder (1968), which randomly reverses the polarity of digital sampling points with a probability of .5. The resulting noise has a flat spectrum, but it retains the amplitude envelope of the speech.[7] It thus sounds vaguely speechlike, but it is not identifiable as an utterance. The stimuli from the *bin-pin* continuum were presented in isolation and preceded by either of the two noise precursors, with intervening silent intervals of 77 msec. The stimulus sequences were the same as those of the corresponding conditions in Experiment 4.

Procedure. As in previous experiments, the two precursor conditions were presented after the isolated CVC condition, in counterbalanced order. The subjects were told that the words would be preceded by a noise, which they should ignore.

Results and Discussion

The results are shown in Figure 5. It can be seen that the noise precursor derived from *Take this* with syllable-final [s] interfered minimally with VOT discrimination, but that the other precursor, which had the amplitude envelope of *Take this* with syllable-initial [s], did reduce performance at the shorter VOTs. This pattern was quite similar to that obtained with the corresponding speech precursors, though the absolute amount of interference was less. There was also considerable variability among the subjects. In the ANOVA including all three conditions, there was a significant main effect of condition [$F(2,22) = 3.87, p = .0364$] and a significant condition × stimulus pair interaction [$F(6,66) = 2.42, p = .0354$], both of which were obviously due to the more effective precursor condition. The stimulus pair main effect was, as always, highly significant.

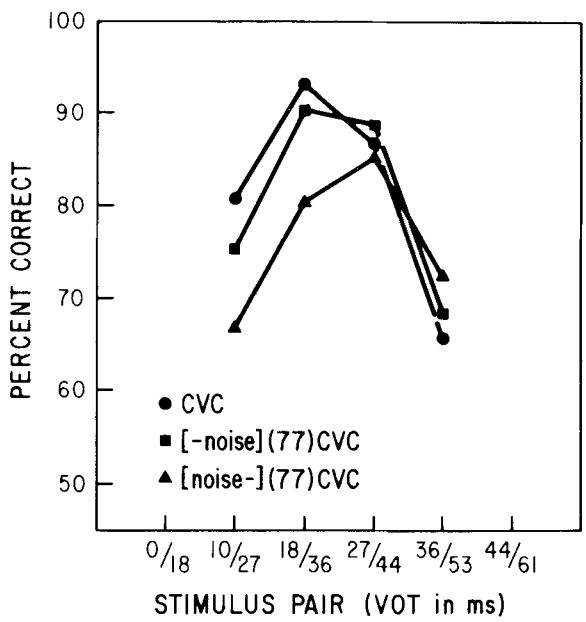

Figure 5. Discrimination performance as a function of stimulus pair in the three conditions of Experiment 5. The [-noise] precursor was derived from *Take this* with syllable-final [s], and the [noise-] precursor was derived from *Take this* with (spliced-in) syllable-initial [s]. Filled symbols are used to match the corresponding speech precursor conditions in Figure 4.

These results suggest that the different amounts of interference caused by syllable-final and syllable-initial [s] noises (Experiment 4) are at least partially due to acoustic differences between the two noises. Since spectral differences were eliminated in the nonspeech precursors, and the duration differences between the original [s] noises were less well defined in the nonspeech analogues because of the absence of spectrally marked segment boundaries, this leaves differences in absolute amplitude and amplitude contour at noise offset as possible factors. This weakens the phonological account of the differences observed in Experiment 4. Also, the possibilities (mentioned in connection with Experiment 1) of hearing the nonspeech noise as a fricative or perceptually restoring a hidden fricative noise seem less plausible here, where the whole precursor phrase was transformed into noise. As in Experiment 1, however, nonspeech noise (Experiment 5) interfered less with VOT discrimination than did [s] noise (Experiment 4). This may also reflect acoustic differences—namely, the different spectral composition of the noises. Why a noise with predominantly high-frequency components (the natural [s]) should interfere more with the auditory processing of VOT than a broadband noise is not clear, however. Alternatively, the difference may have been caused by a reduced perceptual coherence of the nonspeech precursors with the following speech, compared to all-speech stimuli. A similar explanation was proposed by Gordon (1988) when he failed to find an effect of amplitude-modulated noise precursors on the perception of speech stimuli differing in closure duration. This explanation presumes that part or all of the interference takes place at an auditory level beyond the periphery, where the allocation of perceived sources plays a role. An acoustic factor disrupting source continuity may have been the presence of relatively strong low-frequency aperiodic energy in the nonspeech precursors, which was absent from the following speech.

GENERAL DISCUSSION

The present series of experiments started with an attempt to suppress the category boundary effect on a VOT continuum by preceding the stimuli with /s/ and thus neutralizing the phonological voicing contrast (Experiment 1). This manipulation was highly successful, in that the discrimination peak indeed disappeared. However, a control condition with a nonspeech noise precursor also yielded some interference. Experiment 2 showed that the interference caused by an [s] noise was not affected by whether a word boundary preceded or followed the /s/. Experiment 3 suggested that an [s] noise with syllable-initial phonetic properties interferes more than one with syllable-final properties, which Experiment 4 confirmed, though only when the intervening closure duration was relatively short. Experiment 5 indicated that this difference was due to acoustic differences among the [s] noises, since amplitude-matched nonspeech noise precursors showed a similar difference, though less interference in absolute terms.

The results of several of these experiments could be interpreted as lending support to the hypothesis that the VOT discrimination peak (the "category boundary effect") originates at a phonological level of speech processing, not at the level of basic auditory sensitivities. Kewley-Port et al. (1988) recently arrived at the same conclusion when they observed that the discrimination peak was absent in a low-uncertainty discrimination task with trained subjects. Their conclusion is challenged, however, by Macmillan et al. (1988), who have found a discrimination peak in a similar experiment, in which they sampled the VOT continuum more finely and concluded that there is a psychoacoustic boundary on a VOT continuum. The present results are consistent with that interpretation as well. This ambiguity of interpretation pervades Experiments 1-4, but Experiment 5 tends to favor a psychoacoustic account for the precursor effects studied here. That is, the results suggest that the effect of a preceding /s/ on discrimination performance was caused not so much (or not at all) by the neutralization of the phonological voicing contrast in the following stop consonant as by interference with the auditory processing of VOT. This interference may then be considered a possible reason for why phonological neutralization of voicing contrasts in /s/-stop clusters is common in the languages of the world (see Note 1, however).

The mechanism of this interference is not known. It would require a whole series of further studies to disentangle the many possibilities. The discrimination of voice onset time may rely not only on purely temporal differences but also on differences in F1 onset frequency (Soli, 1983), in the relative strength of aspiration (Repp, 1979), and in the amplitude envelope at voicing onset (Darwin & Seton, 1983). A preceding noise could interfere with the processing of any or all of these. How such interference could result in the complete elimination of the psychoacoustic boundary at around 20 msec of VOT is still not clear.

One way in which a noise precursor might affect auditory processing could occur through peripheral forward masking or adaptation. It is not clear, however, why adaptation of nerve fibers sensitive to the high frequencies characteristic of [s] noises should interfere with the perception of spectral and temporal signal properties in the low-frequency region, which VOT discrimination mainly relies on (voicing onset, F1 onset frequency). A more plausible interpretation is that the presence of a precursor simply increased the complexity of the stimuli and thus made it more difficult for listeners to focus on the acoustic properties relevant to the task. This interference may have taken place largely in auditory memory, rather than in peripheral auditory processing. This hypothesis is supported by the finding that nonspeech noise precursors generally interfered less with VOT discrimination than did speech precursors. Although speech and nonspeech

precursors differed in spectral content and thus were not fully equated in their acoustic properties, the main difference between them may have been that the speech precursor "cohered" with the following word, whereas the nonspeech precursor did not do so to the same extent. A parsing of the auditory input into likely sources probably precedes storage in auditory memory (cf. Bregman, 1978; Crowder, 1983), and a listener's knowledge of what constitutes a likely speech source may influence this parsing.

REFERENCES

BREGMAN, A. S. (1978). The formation of auditory streams. In J. Requin (Ed.), *Attention and performance VII* (pp. 63-76). Hillsdale, NJ: Erlbaum.

CARNEY, A. E., WIDIN, G. P., & VIEMEISTER, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *Journal of the Acoustical Society of America*, **62**, 961-970.

CHRISTIE, W. M., JR. (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America*, **55**, 819-821.

COLE, R. A., & SCOTT, B. (1973). Perception of temporal order in speech: The role of vowel transitions. *Canadian Journal of Psychology*, **27**, 441-449.

CROWDER, R. G. (1983). The purity of auditory memory. *Philosophical Transactions of the Royal Society (London)*, **302B**, 251-265.

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., & PARKER, E. M. (1985). Are selective adaptation and contrast effects really distinct? *Journal of Experimental Psychology: Human Perception & Performance*, **11**, 209-220.

DOOLING, R. J., OKANOYA, K., & BROWN, S. D. (1988). Speech perception by budgerigars (*Melopsittacus undulatus*): Synthetic VOT stimuli. *Journal of the Acoustical Society of America*, **83**, (Suppl. 1), S51.

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

HOWELL, P., & ROSEN, S. (1985). Natural auditory sensitivities as universal determiners of phonemic contrasts. In B. Butterworth, B. Comrie, & O. Dahl (Eds.), *Explanations of linguistic universals* (pp. 205-235). The Hague: Mouton.

KEWLEY-PORT, D., WATSON, C. S., & FOYLE, D. C. (1988). Auditory temporal acuity in relation to category boundaries: Speech and nonspeech stimuli. *Journal of the Acoustical Society of America*, **83**, 1133-1145.

KUHL, P. K. (1981). Discrimination of speech by nonhuman animals: Basic auditory sensitivities conducive to the perception of speech-sound categories. *Journal of the Acoustical Society of America*, **70**, 340-349.

KUHL, P. K., & PADDEN, D. M. (1982). Enhanced discriminability at the phonetic boundaries for the voicing feature in macaques. *Perception & Psychophysics*, **32**, 542-550.

LISKER, L. (1984). How is the aspiration of English /p,t,k/ 'predictable'? *Language & Speech*, **27**, 391-394.

MACMILLAN, N. A., GOLDBERG, R. F., & BRAIDA, L. D. (1988). Resolution for speech sounds: Basic sensitivity and context memory on vowel and consonant continua. *Journal of the Acoustical Society of America*, **84**, 1262-1280.

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

REPP, B. H. (1979). Relative amplitude of aspiration noise as a voicing cue for syllable-initial stop consonants. *Language & Speech*, **27**, 173-189.

REPP, B. H. (1984). Categorical perception: Issues, methods, findings. In N. J. Lass (Ed.), *Speech and language: Advances in basic research and practice*. (Vol. 10, pp. 243-335). New York: Academic Press.

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., HEALY, A. F., & CROWDER, R. G. (1979). Categories and context in the perception of isolated steady-state vowels. *Journal of Experimental Psychology: Human Perception & Performance*, **5**, 129-145.

REPP, B. H., & LIBERMAN, A. M. (1987). Phonetic category boundaries are flexible. In S. N. Harnad (Ed.), *Categorical perception* (pp. 89-112). New York: Cambridge University Press.

REPP, B. H., & LIN, H.-B. (1988). *Effects of preceding context on the voice-onset-time boundary*. Manuscript submitted for publication.

ROSEN, S., & HOWELL, P. (1987a). Auditory, articulatory, and learning explanations of categorical perception in speech. In S. N. Harnad (Ed.), *Categorical perception* (pp. 113-160). New York: Cambridge University Press.

ROSEN, S., & HOWELL, P. (1987b). Is there a natural sensitivity at 20 ms in relative tone-onset-time continua? A reanalysis of Hirsh's (1959) data. In M. E. H. Schouten (Ed.), *The psychophysics of speech perception* (pp. 199-209). Dordrecht: Martinus Nijhoff.

SACHS, R. M., & GRANT, K. W. (1976). Stimulus correlates in the perception of voice onset time (VOT): II. Discrimination of speech with high and low stimulus uncertainty. *Journal of the Acoustical Society of America*, **60** (Suppl. 1), S91.

SALASOO, A., & PISONI, D. B. (1985). Interaction of knowledge sources in spoken word identification. *Journal of Memory & Language*, **24**, 210-231.

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

SAWUSCH, J. R., & JUSCZYK, P. (1981). Adaptation and contrast in the perception of voicing. *Journal of Experimental Psychology: Human Perception & Performance*, **7**, 408-421.

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

SOLI, S. D. (1983). The role of spectral cues in discrimination of voice onset time differences. *Journal of the Acoustical Society of America*, **73**, 2150-2165.

STUDDERT-KENNEDY, M., LIBERMAN, A. M., HARRIS, K. S., & COOPER, F. S. (1970). Motor theory of speech perception: A reply to Lane's critical review. *Psychological Review*, **77**, 234-249.

WARREN, R. M. (1984). Perceptual restoration of obliterated sounds. *Psychological Bulletin*, **96**, 371-383.

WILLIAMS, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, **21**, 289-297.

WOOD, C. C. (1976). Discriminability, response bias, and phoneme categories in discrimination of voice onset time. *Journal of the Acoustical Society of America*, **60**, 1381-1389.

NOTES

1. It should be kept in mind that these predictions concern the discriminability of small VOT differences in the region of the original discrimination peak. Few listeners would fail to discriminate the extreme tokens of [spa] and [sp^{h}a], despite the phonological neutralization.

2. This noise was excised from a male speaker's production of the word *spectacular* in a sentence context. This rather short noise was used for no better reason than that it happened to be readily available in digitized form when the stimuli were constructed. As will be seen, however, it served its purpose well.

3. This variability was apparently not due to effects of test order, which were nonsignificant in a separate analysis. That analysis also employed an arcsine transformation of the response proportions, which left the pattern of the results unchanged. Subsequent analyses did not include these two refinements.

4. In the abbreviations for the conditions, the number sign (#) stands for a linguistic word boundary, and the dash (-) following the [s] represents the fact that the [s] noise had syllable-initial phonetic properties.

5. The shift in the peak to a longer VOT value relative to that in Experiment 1 could be due to any of the many acoustic differences be-

tween the synthetic and natural stimuli: presence versus absence of a release burst, different amplitude envelopes, different vowels, different syllable structure, and duration. It is well known that the VOT category boundary does not "stand still," but is influenced by a variety of extraneous variables (see Repp & Liberman, 1987).

6. A corresponding difference in phoneme boundaries for the same stimuli was obtained in a labeling test administered to the same subjects. This boundary shift, which became the subject of a separate investigation (Repp & Lin, 1988), will not be discussed further here. Suffice it to note that it cannot have been due to coarticulatory cues to the originally following /p/ in the *Take the* precursor, because this should have caused a boundary shift in the opposite direction. Repp and Lin (1988) also showed that the difference in closure duration was not responsible.

7. Actually, the noises presented to the subjects had a sloping rather than a flat spectrum. This was because the speech had been digitized with high-frequency pre-emphasis (of about 6 dB per octave above 1 kHz), and the digital noise files and the speech had to be (re)converted into sound in the same stimulus sequence, for which the computer demanded compatible specifications. Thus, even though the nonspeech noises in the computer files had a flat spectrum, high-frequency de-emphasis was applied at output. This could have been circumvented by first redigitizing the speech without pre-emphasis, but it was not considered enough of a problem to warrant the effort.

(Manuscript received April 22, 1988;
revision accepted for publication September 30, 1988.)