Effects of Word-Internal vs. Word-External Tempo on the Voicing Boundary for Medial Stop Closure

Robert F. Port

ABSTRACT

It is known that the perceived voicing of the medial stop in words like rabid and rapid can be controlled by changing the duration of the stop closure. We found in an earlier experiment that increasing the tempo of a preceding carrier sentence shortened the boundary between rabid and rapid along a continuum of silent closure durations, but by far less than the percent decrease in sentence duration. We hypothesized that timing in unaltered portions of the test word reduced the effect of the carrier tempo. This experiment directly compares the effect on this boundary of the tempo of a surrounding carrier when the tempo of the test word itself is changed. A speaker recorded "I'm trying to say rabid to you" at both fast and slow tempos. Rabid was excised from each, and two continua of test words with silent /b/ closures were prepared (from 50 to 200 msec) and embedded in both original sentences. Listeners identified the test words as either rabid or rapid. Results indicate: (1) tempo within the test word had a stronger effect on the boundary than tempo in the carrier, and (2) for a given tempo of carrier, the ratio of the boundary value of closure duration to the duration of the rab syllable was nearly constant for both rab durations.

INTRODUCTION

It has been known since Lisker's 1957 study that if the closure interval corresponding to medial stops, such as those in the words rabid and rapid, has no audible glottal pulsing, English-speaking listeners can hear the stop as either /b/ or /p/ depending on the duration of the closure. Longer closures

*A slightly expanded version of a paper delivered at the 95th Meeting of the Acoustical Society of America, Providence, Rhode Island, 16-19 May, 1978.

+Indiana University, Bloomington, Indiana.

Acknowledgment: Special thanks to Quentin Summerfield for help in preparation of the stimulus tapes. This research was partially funded by a PHS Biomedical Research Support Grant from the Indiana University Foundation.

are heard as voiceless and shorter ones as voiced. Studies of speech production [Lisker, (1957); Peterson and Lehiste, (1960); Port[1]] have shown that the vowel duration also changes, and varies inversely with the stop closure duration when the voicing feature of a following stop is changed. Since it is also well attested that the durations of both these intervals vary with speaking tempo [Peterson and Lehiste, (1960); Gaitenby, (1965); Port (see footnote 1)], the question arises as to what perceptual effects will follow changes in the speaking tempo of an utterance. Specifically, what will happen to the effectiveness of stop closure duration to cue the phonological voicing of a stop?

Of course, there is little reason to suppose that there might be some fixed duration in msec that will serve as the perceptual boundary between cognate pairs such as /b/ and /p/. Rather, we would expect to find that the tempo of the context will affect perceivers' judgments. Indeed, several perceptual experiments have found an effect of speaking tempo on durational timing cues [Pickett and Decker, (1960); Summerfield, (1975)].

In one such study of tempo effects, Port (in press) was able to replicate Lisker's 1957 effect of changing the voicing of a medial stop by manipulation of the stop closure duration. He further demonstrated that the speaking tempo of a carrier sentence would shift this perceptual boundary along the closure duration continuum. In that experiment a single production of the word rabid was cut and spliced so that a continuum of stop closures containing no glottal pulsing was prepared. Subjects identified the stimuli appended to both slow and fast carrier sentences. The results showed a voicing boundary between /b/ and /p/ at about 75 msec of closure duration in the slow carrier. However, when the words were appended to the fast carrier sentence, the perceptual boundary shifted significantly toward shorter values. Thus the tempo of a preceding carrier sentence was able to shift the perceptual voicing boundary along this temporal continuum. On the other hand, a close comparison of the stimuli and the magnitude of the result might give some concern about the effectiveness of tempo. The stimulus sentences, that is, the carrier sentence combined with a test word, shortened by about 30 percent from the slow condition to the fast, while the perceptual boundary shortened only about 10 percent between the two conditions. Why was the effect so small? Notice that the test words in this experiment were made from a single production of rabid. Thus the temporal intervals immediately adjacent to the closure variable were fixed throughout the experiment. If it were the case that the temporal voicing cue is primarily determined by the ratio of the stop closure duration to the preceding or following syllables, then, of course, this might account for the very small effect of a carrier sentence that is external to the test word. In order to directly contrast the effectiveness of timing changes within the test word with timing outside the test word, the following experiment was conducted.

---

Figure 1: Illustration of stimulus construction. The Fast and Slow test words were excised from fast and slow tempo productions of the sentence "I'm trying to say 'rabid' to you." Note that the fast tempo production involved considerable phonetic reduction in the carrier portions of the sentence.
Figure 2: Identification of the test word as rabid (versus rapid) as a function of the stop closure duration. The original word was spoken at either a fast or slow tempo and embedded in either a fast or slow carrier sentence.
METHODS

A speaker recorded the sentence "I'm trying to say rabid to you" at both a fast tempo and a slow tempo. After digitization and 5 kHz filtering of the utterances, stimuli were prepared under computer control as illustrated in Figure 1. The test-word rabid was excised from each sentence by cutting at the end of the initial /r/ constriction and in the middle of the final /d/ closure. Then within each word, cuts were made immediately after closure for the medial /b/ and just before release. The glottal pulsing during the closure was discarded and a continuum was prepared by inserting silent gaps in each test word varying from 50 to 200 msec in steps of 10 msec. With the short stop closures, both test words sounded like rabid and with the long closures, both sounded like rapid. Individual members of each of these continua were then embedded in both the fast and the slow carrier sentences. A listening tape was prepared containing 10 tokens of each of these 64 stimuli in random order and presented to 12 college-educated listeners in Bloomington, Indiana, after the subjects had heard a 10-item practice tape. The listeners checked on a response sheet whether they heard the word rabid or rapid in each sentence.

RESULTS

The results of these identifications are presented in Figure 2. Here the percent identification of each stimulus as rabid, that is, as containing a medial voiced stop, is plotted as a function of the closure duration of the stop. The Fast word in the Fast sentence appears with filled triangles and the Slow word in the Slow sentence has open circles. In the Fast word-Fast sentence condition, stop closures longer than about 70 msec were heard predominantly as /p/, while in the Slow word-Slow sentence condition, the /b-/p/ boundary is more than twice as long at 165 msec. What happened when the word and sentence tempos were crossed? If we start with the Slow word-Slow sentence condition, we find that changing just the carrier sentence from Slow to Fast—that is, from the open circles to the open triangles—moves the boundary about 30 msec, or 20 percent toward shorter values. This replicates our earlier experiment although the effect here is stronger than before, presumably because the test words here are sentence medial rather than sentence final. If, however, we change just the test word from Slow to Fast and leave the carrier sentence the same, (that is, from open circles to filled circles on the graph), we find that the boundary now shifts by 65 msec. This amounts to a shortening of the voicing boundary of about 40 percent—far larger than the effect of a change in the carrier sentence alone. Of course the same effects can be seen if we begin with the Fast word-Fast sentence condition: changing just the carrier moves the boundary significantly, but changing just the test word moves it considerably more.

Individual subject data showed that all 12 listeners were unanimous in providing more /b/ responses in the condition indicated in the pooled data for 5 of the 6 pairs of conditions. In the comparison of the Slow sentence-Slow word condition with the Fast sentence-Slow word condition, however, a single subject disagreed with the others by giving more /b/ responses in the latter condition. A sign test indicates that even this pair of conditions is distinct at p < .01.
These results imply that timing in the two syllables adjacent to the stop closure duration is considerably more important in determining the durational boundary for medial stop voicing than is the speaking tempo in a surrounding carrier sentence. Although this experiment does not permit separation of the effects of the preceding stressed syllable, /rab/, from the effects of the following unstressed syllable, /bid/, there is reason to suppose that the preceding syllable—in particular, the duration of the preceding vowel—plays the most important role in the apparent resistance of the boundary to the effects of the external carrier. First of all, we know that the duration of the preceding vowel in such words also varies as a function of following consonant voicing and, in particular, varies inversely with the stop duration itself (Port*). Second, a closer look at the classic perceptual study by Peter Denes (1955) on word final fricatives suggests an intriguing interpretation of these results. Using a synthetic pair of test words differing in the voicing of the final fricative in the words /use, n./ and /use, v./, Denes varied the duration of both the vowel and the voiceless fricative portion of his test words. He found that short final fricatives were heard as /z/ (even though technically voiceless), while longer durations were heard as /s/. He also found that the voicing boundary along his fricative duration continuum depended on the duration of the preceding vowel. In fact, when his stimulus identifications were plotted as a function of the ratio of the fricative duration to the vowel duration, it appeared that the perceptual boundary could be expressed as a constant ratio of the fricative to the preceding vowel. In particular, the criterion ratio was approximately one. That is, if the fricative was longer than the vowel, it was identified as /s/. If it was shorter than the vowel, it was identified as /z/.

Since, in this experiment, there were two durations of the /rab/ syllable varying over a range of nearly 2 to 1, we decided to replott our identification results as a function of this ratio: stop closure duration divided by the duration of the "vowel" in the preceding syllable. For this purpose we measured the interval from the point of steepest rising F3 slope in the transition from the /r/ into /æ/ up to the point of apparent closure for the /b/ for both Fast and Slow test words (110 msec and 200 msec respectively) and divided this into each closure duration. The replotted results appear in Figure 3 where the vertical axis remains the same as in the preceding figure. The value of 1.0 along the horizontal axis means that the stop closure is equal to the duration of the preceding vocalic interval, and values less than one mean the stop closure is shorter than the vowel. Thus, the value 0.5, for example, means the stop was half as long as the preceding /rab/ syllable. In this display we notice first that the four functions no longer look as different as they did before, and that they are now grouped by carrier sentence rather than by word duration. In particular, it can be seen that in the Fast carrier sentence, the Fast and Slow test words (in filled and open triangles) cross the 50 percent boundary at almost exactly the same ratio—about 0.68. Similarly, in the Slow carrier sentence, the identification functions for the two test words (represented by filled and open circles) have perceptual boundaries very near each other centered around a consonant/vowel ratio of 0.85. Thus for a given tempo of carrier sentence, the voicing

*See footnote 1.
boundary along a continuum of ratios of stop closure to the preceding vowel interval tended to be nearly constant.

In an attempt to evaluate the significance of the effect of the test word tempo at a given tempo of carrier sentence, we examined individual subject data over just that portion of the stimulus continuum where the two test words share the range of C/V ratios. As can be seen in Figure 3, that range lies between .45 and 1.0, although the Fast word has 7 stimuli in this range while the Slow word has 12. Thus, for each subject we asked whether in the Fast carrier sentence he made a larger percentage of /b/ identifications to his 70 Fast word presentations or to the 120 Slow word identifications. Of the 12 subjects, 6 had a larger proportion of /b/ responses in the Fast word condition and 6 in the Slow condition, thereby providing no suggestion of a consistent effect. Although the pooled data in Figure 3 indicate more total /b/ responses to the Fast word than the Slow one when embedded in the Slow carrier sentence over this range, a similar sign test on individual subjects shows that only 7 of the 12 had a higher percentage of /b/ responses for the Fast word. Thus, these results suggest there is no consistent difference between the test words in the boundary value of C/V ratio at a given tempo of carrier sentence. On the other hand, even looking just at this portion of the stimulus continuum, all 12 subjects had a larger number of /b/ responses for both the Fast word and the Slow word in the Slow carrier than in the Fast carrier. In short, the differences between the carrier sentences (triangles vs. circles) in Figure 3 are highly significant, while the differences between the test words (filled vs. open) are probably not. Apparently the relative duration of a vowel and the following stop closure is a highly effective cue to the phonological voicing of the stop (a) over a range of at least 2-to-1 in vowel duration, (b) when the tempo of the surrounding speech remains constant.

**DISCUSSION**

It is tempting to interpret this result, which converges with Denes (1955) and is coherent with the production data showing an inverse relation between vowel duration and consonant constriction duration, as evidence that English listeners do not make voicing judgments by evaluating the stop closure duration and the preceding vowel duration independently. They are, rather, directly comparing the durations of these two intervals.

If further data should continue to support such a hypothesis, it would be of considerable interest. First, this is an example of what might be described as the temporal counterpart of coarticulation, since it suggests that the temporal information for segmental phonological features may not be localizable in any particular interval, but may inherently require several neighboring intervals for specification of its abstract temporal structure. Second, we may note that a temporal phonological cue of this type would have certain practical advantages for a language. For example, in speech production, this ratio would naturally tend to be invariant under changes of speaking tempo and changes in degree of stress. Such a cue might also permit very small timing differences in vowel duration and stop closure duration to combine in a way that yields a more prominent and perceptually useful cue.

Before closing we should comment on one further aspect of these data. If the vowel-to-consonant ratio is the most powerful temporal cue to the voicing
Figure 3: The same results as Figure 2 here plotted as a function of the ratio of the stop closure duration to the preceding "vowel" measured from the point of steepest slope in $F_3$ to /b/ closure. This interval sounds like /rab/ when heard in isolation.
of medial consonants, what accounts for the ability of the external carrier to shift the boundary in the direction that it did? A first guess might be that the shift toward a smaller C/V ratio at the faster tempo reflects a similar shift occurring in speech production. Unfortunately, however, production data agree [Kozhevnikov and Chistovich, (1965); Klatt, (1976); Port*] that stops shorten less than vowels with a tempo increase. Thus differences in the compressibility of stops and vowels would predict a ratio shift in the opposite direction to that observed in Figure 3. On the other hand, if we assume that listeners employ the stop closure duration as a cue by making two separate comparisons, then a fairly simple model emerges. In addition to noting the relative duration of the consonant constriction to the preceding vowel, listeners may also compare the consonant duration to some intervals in the surrounding carrier sentence. When these other intervals are compressed, the stop closure boundary will also shift toward shorter values. Although we do not know what intervals might be involved in such a comparison, we do know that their effects are small relative to the more local timing effects closer to the consonant constriction.

In conclusion, then, this study revealed evidence supporting an early observation of Denes that, when listeners are forced to rely on timing cues for the voicing of a postvocalic consonant, their perception is based largely on the relative duration of the stressed vowel and the following consonant constriction. In addition, this experiment has replicated an earlier study of our own showing that the tempo of speech outside a test word will also have some effect on the perceptual boundary for voicing along a continuum of stop closures. Finally, these results clearly indicate that of these two temporal cues for voicing—one close to the stop closure itself and one more remote—the more local effect is clearly dominant.

REFERENCES


*See footnote 1.