Closure Hiatus: Cue to Voicing, Manner and Place of Consonant Occlusion*

Leigh Lisker†

ABSTRACT

Delayed onset of laryngeal vibration following release of an initial stop by about 35 ± 15 msec generates acoustic features eliciting p,t,k responses from speakers of English. These features, by a common misnomer, are referred to as cues to stop voicelessness; in fact, they are cues to voiceless stop aspiration. Medially, before unstressed vowels, English has voiceless stops that are not aspirated, and these lack some of the features of initial /p,t,k/. An important cue to medial /p,k/ before unstressed vowels is an interruption of glottal pulsing during closure, provided this interruption exceeds a certain duration. In experiments replicating and extending earlier studies, a number of naturally produced and synthesized polysyllables were varied in respect to their closure intervals. In part, results replicated earlier findings, but not unambiguously. It appeared that 1) there were significant individual differences in response to stimuli with edited closure intervals; 2) stimuli derived from different tokens of the same phonetic types elicited different responses; 3) the apical flap ([r]) response to very short closure intervals could not be entirely explained by a simple motor theory interpretation.

The recent literature dealing with acoustic cues that separate homorganic stops in English is mostly concerned with stops initially before stressed vowels. With respect to the most important of these—the time of onset of laryngeal pulsing—we are told that /ptk/ is distinguished from /bdg/ along this so-called VOT continuum, in that for /ptk/, the onset of pulsing must be deferred either until a certain time, about 35 msec, after the stop release, or until the articulatory shift from closure to succeeding vowel has been largely completed. The fact that this requirement for initial /ptk/ cannot hold true for phonetic events—linguistically identified with /ptk/—that occur in other contexts, has been relatively unemphasized. If, for example, we say that the word paper includes two instances of /p/, the VOT requirement

---

*This paper was presented orally at the 93rd meeting of the Acoustical Society of America, Pennsylvania State University, 5–8 June, 1977.

†Also University of Pennsylvania.

Acknowledgment: Acknowledgment is made of the support of the National Institute of Child Health and Human Development.

just mentioned must be satisfied only for the initial one, since the medial /p/ does not usually involve much of an interval between release and resumption of pulsing. The acoustic properties of the medial /p/ following its release are in fact usually of a kind that will elicit b judgments, if the signal in which the /p/ is embedded is edited so that the release burst and transition come to be in initial position. Given this fact, how can it be said that any acoustic feature depending on a delay in pulsing onset is the cue to stop voicelessness in English? All that can be claimed is that the presence of such a feature is sufficient to trigger ptk responses; in its absence bdg responses are not necessarily reported.

The attention lavished on the VOT dimension, and the importance attached to particular durations by which pulsing onset lags behind release, reflect the fact that so much of the search for the segmental cues has been focused on the analysis and synthesis of nursery utterances such as ba da ga. This is despite the fact that in a piece of speech all but one phonetic event is noninitial, and it is not generally believed that speech is made up entirely of simple concatenations of CV sequences like those that can occur as complete utterances. For the stops even more than for other classes of phonetic events, at least in English, it is a mistake to gloss over the context-dependent nature of the cues by which /ptk/ and /bdg/ are distinguished. If primary attention had been directed to medial or final position, we should have a somewhat different idea of the acoustic basis for the distinction. Because both /ptk/ and /bdg/ occur initially and both can occur finally without any acoustic signal of release, it would seem impossible to claim that any feature found either before or after closure is a necessary property for the perception of either class of phonemes.

Leaving aside the case of the final stops, let us consider some evidence that an intervocalic occlusion with interruption of pulsing may be interpreted as either /bdg/ or /ptk/ when pulsing resumes immediately upon the release. This evidence comes from an old experiment, (Lisker 1957), since replicated in some recent work by Robert Port (1976), that involved the editing of natural speech recordings so as to vary the duration of a silent interval corresponding to an intervocalic closure. Manipulation of tokens of the words ruby and rupee yielded stimuli that a group of seven phonetically naive subjects labeled as shown in Figure 1. It appears that the duration of silent closure may figure as a significant cue for word identification, specifically for the /p/-/b/ contrast. The rupee-derived stimuli were heard mostly as ruby for closure durations less than 70 msec; ruby, when its buzzed closure was replaced by silence longer than 100 msec, was more often reported as rupee. (The two intermediate curves of the display, giving responses to stimuli composed of cross-combinations of first and second syllables of the source words, will not be commented on now.) The difference in cross-over values for the ruby and rupee curves is a little more than 30 msec, and we may like to think of this difference as the perceptual-phonetic equivalent of whatever other features that precede and follow the silent interval and also operate as cues.

The range of durations tested in this experiment was chosen with a lower limit of 40 msec so as to exclude the possibility that listeners would report hearing a t or d (that is, the alveolar flap) rather than b, while the upper bound of 140 msec was intended to avoid the effect of an abnormally long or
Figure 1: Labeling responses of seven naive listeners to 44 stimuli derived by tapecutting recordings of naturally produced tokens of the target words, and recombining pre- and post-closure intervals in four ways, with silent intervals varying in 10 msec steps from 40 to 140 msec.

Figure 2: Labeling responses of seven listeners to spoken ruby and rupee edited so stop closures of each word were silent intervals varying in duration from 0 to 140 msec, in 10 msec steps. Subjects were instructed to listen for rupee ruby Rudy.
Figure 3: Labelings of edited natural productions of *rabid* and *rapid*. Closure intervals, varying in 15 msec steps from 30 to 150 msec, were either silent or filled with naturally produced glottal buzz. The twelve listeners were phonetically naive.

Figure 4: Responses of three phonetically trained listeners to the same stimuli that provided the data of Figure 3. These subjects were instructed to assign stimuli to one of the five categories listed in the Key.
geminate p. The shift in place and manner judgments at closure of less than 30 msec has been studied in detail by Port, (1976). The conclusion is reasonable that closure duration not only may serve as a cue to stop voicing, but that it must have some minimum duration if it is to signal a stop consonant. At durations too small to be appropriate for the perception of stop manner (and probably for stop production as well), the consonant perceived was a flap. Whether it is because the only flap in English is apico-alveolar, or because there is some purely auditory basis for the perceptual zeroing of the labial place cues, listeners often reported hearing ruddy instead of ruby. While it might be more interesting, especially to the linguist, if the first account were true, I think the second is closer to the mark. The basis for this belief will be made clear later on.

Now I want to turn to some recent experiments, first of all to one performed to see whether the old results would be replicated. Figure 2 represents labeling responses for silent durations ranging from 0 to 140 msec, the upper panel for stimuli derived from a token of ruby, and the lower for those from rupee. The ruby-derivatives were heard mostly as ruby for closures of from 0 to 100 msec. Stimuli derived from rupee were heard as rupee for durations of 70 msec and greater. The /b/-/p/ crossover values are not very different from those measured in the first experiment. Unlike the older results, here no ruby-derivatives achieved better than 70 percent p responses, while no stimuli from rupee were reported as ruby more than 75 percent of the trials. In the case of the flap, there is no duration for which this category included more than 60 percent of the responses, a score achieved only for a rupee with closure interval reduced to 0 msec duration. I think this result does not mean that Port's somewhat different findings cannot be replicated--only that it cannot be guaranteed that every token of an intervocalic /b/ or /p/ can be heard as the flap for very brief closures, even those too short for human vocal tracts to execute.

Figure 3 shows responses to a set of stimuli derived from another favorite pair of words: rabid and rapid. In this set, the closure duration was varied in 15 msec steps, from 30 to 150 msec. For each value of closure duration, two stimuli were prepared, one silent and the other filled with buzz of laryngeal origin. The dashed lines are for the buzz-filled. At least for the particular token of rabid used, rabid → rapid for silent closures of 90 msec and more, a shift occurs that is much more decisive and at a smaller crossover interval than in the case of the second ruby-rupee test. For the stimuli derived from rapid, the shortest silent closure rated no better than 75 percent of the b responses. For both rabid— and rapid-derivative stimuli, the introduction of buzz into the closure interval shifted judgments decisively—80 percent or more—to b. None of the six phonetically untrained listeners reported anything other than rabid or rapid. By contrast, Figure 4 shows labeling responses of three trained listeners. Responses to rabid-derivatives with silent closures, shown in the lower left, fall into five categories: flap, b, geminate b (bb), p and geminate p (pp). With 30 msec of closure, all responses report the flap category, b responses are at a maximum (but no more than 50 percent) for a silent interval of 60 msec, while for 75 msec and longer most responses are p or geminate p (pp). Responses to the stimuli derived from rapid and having silent closures are shown in the lower right. The b responses are even fewer than in the case of the rabid-derived items, and p responses preponderate, starting with a closure of 60 msec. For
the shortest closures the flap responses are also fewer than those elicited by the stimuli having rabid as their source. The upper panel, giving responses to the stimuli with buzz-filled closures, shows only three categories: flap, b and geminate b (bb). Here, too, the stimuli from rabid were most often reported as ratted for the shortest closures. It should be remembered, however, that this is not in agreement with the finding for ruby-rupee, where it was rupee whose derivatives were more often heard as the form with a medial apico-alveolar flap.

The next experiment was undertaken to discover whether the finding for ruby vs. rupee and rabid vs. rapid meant that silent closure could be said to operate as a cue to stop voicing independent of place of articulation. The word-pair used was locker-lager, which in my variety of American English differ only with respect to their stop consonants. Figure 5, in the top panel, suggests a strong place effect: the $\emptyset$ interval between end of implosive transition and release burst reduces the locker token to something ambiguous between locker and lager, while increasing the gap to the longest one tested is no more effective in shifting lager to locker. The subjects were instructed to listen for a nonsense form latter (lætər), but reported only locker and lager. The pooled data of the upper panel, when examined for individual patterns of labeling behavior, revealed that the subjects could be divided, nonarbitrarily, into two groups of three each. Group 1 reported all locker-derivatives as locker in more than 75 percent of their responses, even for closure intervals of $\emptyset$ and 10 msec; moreover locker was reported for the lager-derived stimulus with the longest silent gap. Group 2 showed a bias the other way: locker went to lager for the shortest closures, but lager remained lager 100 percent independent of the closure duration. None of the six subjects was prepared to accept both a shortened /k/ closure as /g/ and an augmented and silenced /g/ closure as /k/. No doubt we can, by pure synthesis, tailor stimuli of a kind to enhance the effectiveness of silent interval as the feature controlling a shift between medial $g$ and $k$, but the present data cast some doubt on the view that closure duration per se functions crucially in natural speech. If, in our figure, we restrict our attention to the range of closure durations recorded in natural speech, say from 30 to 100 msec (see, for example Sharf, 1964), then none of the curves in the display crosses the 50 percent level.

The last experiment is one in which a disyllable ratted (past tense of the verb "to rat") was edited in the usual way and submitted to listeners for labeling. Their responses, shown in Figure 6, are rather surprising; as the silent interval increases, there is a shift from mainly t or d to mainly p judgments, with b not exceeding 27 percent. These data do not tell us where t + d judgments represent flap percepts and where they represent stops (since the subjects had not had enough phonetic experience to make such a discrimination). However, the fact that labials were reported, together with the fact that labials, but not velars, could be converted to flaps, suggests that the place information generated by the apico-alveolar flap articulation is ambiguous, and that this ambiguity has some acoustic basis other than a simple temporal one. Casual inspection of some spectrograms does not make this seem unreasonable. An explanation for the shift from labial to apico-alveolar flap (or simply t to d) judgments that appeals to the fact that only at the latter place can we produce closures of 30 msec and less cannot be turned around, for we cannot claim that closures of 90 msec and more can be produced only at the
bilabial place of articulation.

To summarize: 1) there are significant differences among subjects to the extent in which their labelings of silent closure intervals as /ptk/ or /bdg/ are duration-controlled; 2) response patterns differ, in crossover values and cleanness of category separation, when different tokens of the same words serve as stimulus sources; 3) if we consider the two places of articulation where stops are produced in trochaic words in American English, labial and velar, and particularly if we limit attention to closure durations commonly found in speech, then the nature of the closure interval, silent vs. buzz-filled, seems a more reliable predictor of labeling behavior than does the duration of that interval; 4) the perception of labially produced closures as alveolar flaps when the durations are very short depends, at least partially, on the failure of alveolar flap articulations to produce place cues clearly distinguishable from those produced by bilabial stop articulation.

REFERENCES

