4871

CLOSURE DURATION AND RELEASE BURST AMPLITUDE CUES TO STOP CONSONANT MANNER AND PLACE OF ARTICULATION*

BRUNO H. REPP
Haskins Laboratories

The perception of stop consonants was studied in a constant neutral [s–l] context. Truncated natural [p], [t], and [k] release bursts at two intensities were preceded by variable silent closure intervals. The bursts, though spectrally distinct, conveyed little specific place information but contributed to the perception of stop manner by reducing the amount of silence required to perceive a stop (relative to a burstless stimulus). Burst amplitude was a cue for both stop manner and place; higher amplitudes favored "t," lower amplitudes favored "p" responses. The silent closure interval, a major stop manner cue, emerged as the primary place cue in this situation: Short intervals led to "t," long ones to "p" responses. All these perceptual effects probably reflect listeners' tacit knowledge of systematic acoustic differences in natural speech.

INTRODUCTION

Silent closure duration is an important cue to the perception of stop consonant manner – that is, of phonetic distinctions that rest on the perceived presence versus absence of a stop consonant (e.g., Dorman, Raphael and Liberman, 1979; Bailey and Summerfield, 1980; Repp, 1984). The question of principal interest in the present study was whether different amounts of closure silence are needed to perceive stop consonants having different places of articulation. Specifically, it was hypothesized that, because labial stops generally have longer closure durations than alveolar and velar stops in natural speech (e.g., Menon, Jensen and Dew, 1969; Suen and Beddoes, 1974; Bailey and Summerfield, 1980; Stathopoulos and Weismer, 1983), longer intervals might be needed for their perception, too.

This hypothesis makes two semi-independent predictions: (1) Given *unambiguous* cues to stop consonant place of articulation, more silence will be needed to perceive "p" than "t" or "k"; that is, perception of stop manner, as cued by closure duration, may depend on perceived place of articulation. (2) Given *ambiguous* place cues and sufficient silence to perceive a stop consonant, short closure silences will lead to "t" or "k" responses while long silences will lead to "p" responses; that is, closure duration is a direct cue to place of articulation. The first of these predictions is difficult to test because the different acoustic configurations needed to unambiguously specify place of articulation may have psychoacoustic effects on perception of the closure silence,

* This research was supported by NICHD Grant HD01994 and BRS Grant RR05596 to Haskins Laboratories. A short version of this paper was presented at the 105th meeting of the Acoustical Society of America in Cincinnati, Ohio, May 1983.

which are difficult to dissociate from phonetic effects due to perceived place of articulation. The second prediction, however, can be tested easily by varying silence duration in a constant acoustic environment.

In a previous study addressing these issues, Bailey and Summerfield (1980) used synthetic speech stimuli consisting of an initial [s] noise followed by a variable silent interval and by a vocalic portion with or without initial formant transitions. Two findings are relevant here. When either the second formant of a steady-state vowel or the vocalic formant transitions were varied so as to unambiguously cue the perception of "p," "t," or "k," the amount of silence required to perceive the stop consonant did not vary significantly with place of articulation, except that it was reduced for "k" cued by formant transitions. Bailey and Summerfield attributed this latter effect to auditory energy summation caused by the proximity of the second and third formants at vowel onset; that is, they assumed a psychoacoustic rather than phonetic basis for the effect. The other finding was that, when the place-of-articulation cues in the vocalic portion were ambiguous, so that (given sufficient silence) the same acoustic pattern elicited more than one type of stop response, "p" responses were clearly preferred at longer closure durations, while "t" or "k" responses predominated at short closures. The first, negative finding suggests that stop manner perception is largely independent of perceived place of articulation. The second finding, however, suggests that the listeners' internal perceptual criteria for place of articulation do include closure duration as an important acoustic dimension.

The principal aim of the present study was to replicate Bailey and Summerfield's findings, using natural-speech stimuli that, instead of variable formant frequencies or transitions, included release bursts appropriate for each place of articulation. A second aim was to examine the specific contribution of the release burst itself to stop manner perception. As a rule, alveolar and velar stops following [s], in contrast to labial stops, do not need any closure silence to be perceived as long as an intact natural release burst is present (Repp, 1984). This difference in silence requirements might be due to the higher amplitude and longer duration of alveolar and velar bursts (Zue, 1976), and it might disappear when the overall amplitudes of these bursts are reduced to resemble those of labial bursts. In addition to examining this question, the present experiment also investigated to what extent burst amplitude affects perception of stop manner and place, following Ohde and Stevens (1983) and Repp (1984).

Two methodological decisions require justification. First, to exclude cues to stop place of articulation in the signal portions surrounding the critical cues of closure duration and release burst, these cues were embedded in a constant [s–l] context. Preliminary observations suggested that [l] resonances contain only weak (if any) formant transition cues to preceding stop consonants, so this segment seemed ideally suited for the purpose. However, this resulted in some consonant clusters ([stl] and [skl]) that are unfamiliar to English speakers and listeners. It was assumed, however, that these clusters would not be difficult to produce or perceive, and the results tend to justify this assumption. Second, in order to make closure duration a salient cue to stop manner at all three places of articulation, it was necessary to reduce the natural release bursts, since full alveolar and velar release bursts are generally sufficient cues for perception of a stop consonant. This was done by waveform truncation and resulted in residual bursts that were spectrally distinct but, as it turned out, conveyed surprisingly little place information. The present study thus primarily addresses the question of the role of closure duration as a cue when other place-of-articulation cues are highly ambiguous.

METHOD

Stimuli

A number of repetitions of *slat, splat, stlat,* and *sclat* were recorded by a male speaker of American English, low-pass filtered at 4.8 kHz, and digitized at 10 kHz. One good utterance of each syllable was selected and manipulated further by computer waveform editing procedures. The release bursts (i.e., the aperiodic signal portion preceding the first glottal pulse) of *splat, stlat,* and *sclat* (originally 17, 43, and 43 msec in duration, respectively) were excerpted and trimmed to 10 msec duration. This was done by eliminating the final low-amplitude portions of the labial and alveolar bursts. The velar burst, on the other hand, had several amplitude peaks, the last and most pronounced of which happened to occupy the last 10 msec; therefore, this final portion was taken as the truncated burst. Two versions of each truncated burst were created by changing their amplitudes by 10 dB: The labial burst was amplified by that amount while the alveolar and velar bursts were attenuated. This was done because the labial burst had less high-frequency energy than the other two bursts (see below). Each of these six bursts was spliced onto the [læt] portion (365 msec long) derived from *slat;* thus, the voiced portion immediately following each burst was constant and contained no distinctive cues to place of stop articulation. A seventh, burstless stimulus was included as a baseline. All seven stimuli were preceded by a constant [s]-noise (226 msec long) derived from *slat,* and by a variable closure interval. Closure intervals were varied from 0 to 100 msec in 20-msec steps, for a total of 35 stimuli that were recorded in five different random orders.

Subjects and procedure

Ten subjects, including nine paid student volunteers and the author, listened to the stimulus tape over TDH-39 earphones at a comfortable intensity (approximately 76 dB SPL for vowel peaks) and identified the stimuli in writing as beginning with "sl," "spl," "stl," or "scl." Instructions alerted subjects to the unfamiliar consonant clusters.

RESULTS AND DISCUSSION

Figure 1 compares the labeling function (percentage of stop responses, regardless of place of articulation, as a function of closure duration) for burstless stimuli with the

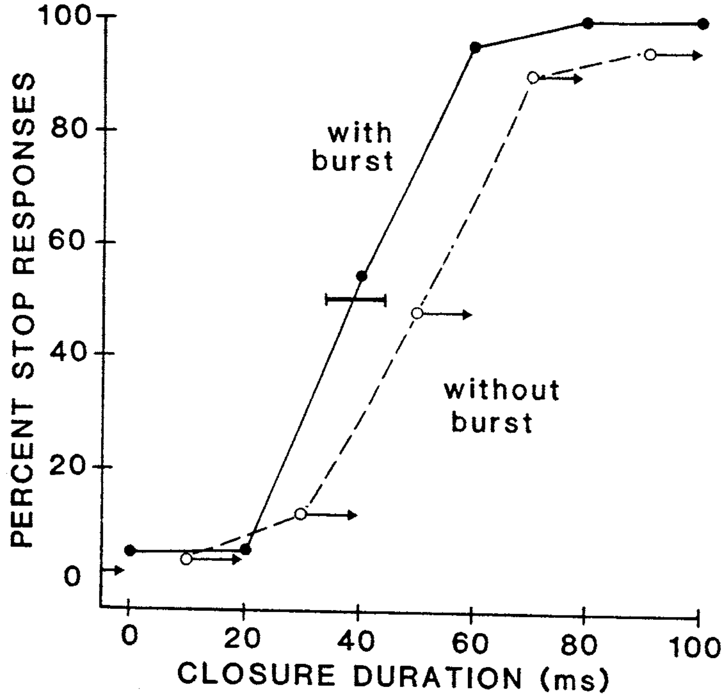

Fig. 1. Effect of presence versus absence of release burst. The solid function is the average of all six burst conditions; the horizontal bar indicates the range of 50% cross-overs. Closure durations in the no-burst conditions are nominal; actual durations are indicated by arrows.

average labeling function for the six types of stimuli with bursts. As indicated in the figure by the horizontal bar at the 50% point, the average phonetic boundaries for the six burst conditions varied over a 10-msec range, from 34.5 to 44.5 msec of closure silence. The boundary for the burstless stimuli was clearly longer – at a nominal 50.5 msec of silence (i.e., measured to the onset of the nonexisting burst), or at an actual 60.5 msec of silence (as indicated by the arrows in the figure). This difference was exhibited by all subjects and was significant in a one-way analysis of variance on the total percentage of stop responses, after applying a correction for the conversion to nominal closure duration and after omitting the data for the author who showed the largest difference, $F(1,8) = 16.6, p < 0.01$. Thus, the truncated release bursts made a significant contribution to stop manner perception (cf. Repp, 1984); that is, the boundary was shortened by

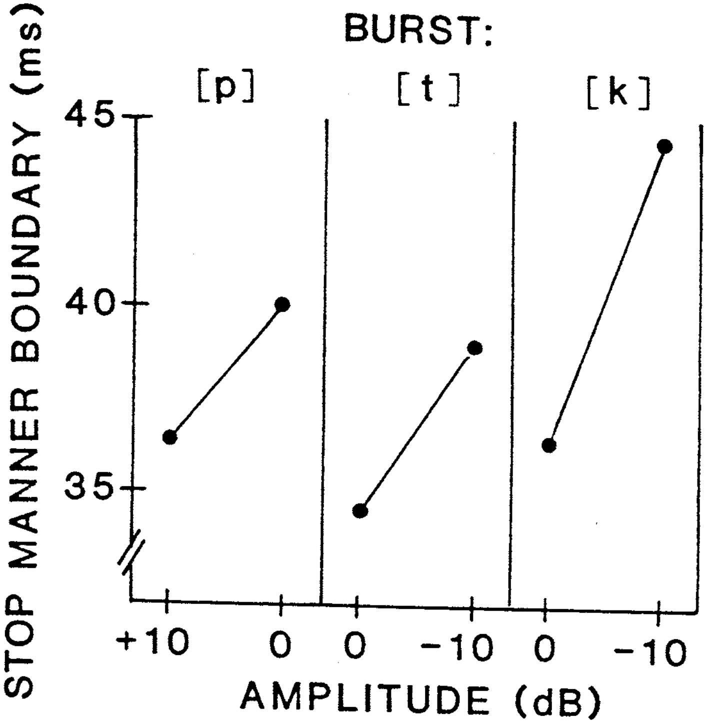

Fig. 2. Comparison of category boundaries in six burst conditions: Effects of burst category and amplitude.

more than the 10 msec expected if the presence of a burst merely had prolonged the effective closure duration.

Figure 2 shows the effects of burst category (intended place of articulation) and amplitude on the stop manner boundary, still combining all kinds of stop responses. Burst amplitude clearly had an effect: Amplification of the labial burst increased stop responses (i.e., shortened the boundary), while attenuation of the alveolar and velar bursts decreased stop responses. Thus, burst amplitude could be traded against closure silence in stop manner perception (cf. Repp, 1984). The effect of burst amplitude was significant in an analysis of variance, $F(1,9) = 8.4$, $p < 0.05$. The main effect of burst category was nonsignificant, and so was the interaction.

A comparison across the three burst categories is difficult because amplitude differences are confounded with spectral differences. Overall rms amplitudes were determined after redigitizing the stimuli without preemphasis. Unexpectedly, the amplitude of the labial burst turned out to be 3 dB higher than that of the alveolar and velar bursts, which were equal and 6 dB below the amplitude of the [l] onset (the first

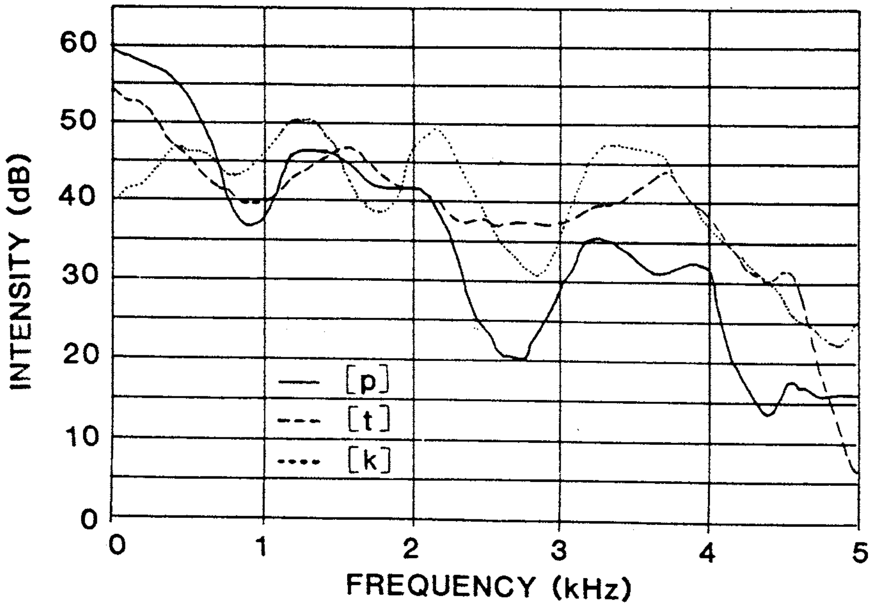

Fig. 3. Spectra of the 10-msec [p], [t], and [k] bursts at their original amplitudes, without pre-emphasis. Spectra were obtained by FFT analysis (program FDI of the ILS package), using a 25.6-msec Hamming window whose left edge preceded the burst onset by 10 msec. The spectra were smoothed by averaging across a 400-Hz rectangular window moving in 20-Hz steps.

10 msec). This was apparently due to a strong low-frequency component in the labial burst waveform. It is likely, however, that the amplitude of higher-frequency components is more important for stop manner perception, as has also been hypothesized by Ohde and Stevens (1983) with regard to place of articulation perception. Figure 3 compares the spectra of the three truncated bursts at their original amplitudes. As expected, the labial burst had less energy than the alveolar and velar bursts in the high-frequency regions above 2 kHz; the average difference is about 10 dB. Thus, amplification of the labial burst by 10 dB resulted in approximately equal levels of high frequency energy across the three burst categories, which is consistent with the very similar stop manner boundaries obtained (see Fig. 2).

So far, stop responses have been treated as a single category. We turn now to an analysis of stop responses by place of articulation. Figure 4 shows conditional percentages of "p," "t," and "k" (i.e., "scl") responses in separate panels as a function of closure duration (from 40 msec up) and burst category, combining the two burst amplitudes. The no-burst condition is also plotted at the *actual* closure durations. It is evident that

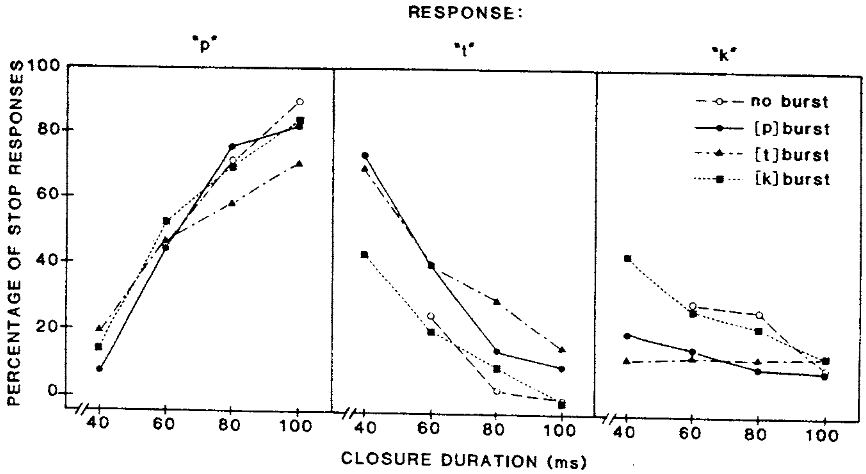

Fig. 4. Percentages of "p," "t," and "k" responses (separate panels) as a function of closure duration and burst condition.

closure duration provided the most important cue to stop place of articulation: At short closures, "t" responses predominated (notwithstanding a possible bias against reporting "stl" clusters) while, at long closure durations, the response was overwhelmingly "p." These trends held almost regardless of the nature of the burst; [p] and [t] bursts, in particular, yielded highly similar results. The results of [k] bursts resembled those for burstless stimuli, perhaps because this late component of the burst did not preserve specific place of articulation information. (Cf. the absence of a pronounced mid-frequency peak in the spectrum – see Fig. 3 – which characterizes velar onset spectra, according to Stevens and Blumstein, 1978.)

The similar perceptual results for [p] and [t] bursts, whose spectra did exhibit the general spectral properties characteristic of these places of articulation (see Fig. 3 and Stevens and Blumstein, 1978; note that the present spectra are not pre-emphasized), may have been due to their short duration. According to Stevens and Blumstein (1978), the most salient place cue is the onset spectrum computed over a window approximately 25 msec long; if so, the 10-msec bursts were presumably integrated with the constant [l] onset following them and thus lost much of their distinctiveness. The present results show very clearly, however, that more than the onset spectrum is involved in place perception: When spectral cues are ambiguous, closure duration takes over as the salient place cue, as also observed by Bailey and Summerfield (1980).

The reason for the effectiveness of the closure duration cue presumably lies in the well-known fact that [p] closures tend to be longer in natural speech than [t] and

[k] closures (although little is known about [stl] and [skl] clusters).[1] An alternative, psychoacoustic explanation might be proposed, however: that the preceding [s]-noise, with its strong high-frequency components, left a trace in sensory memory that was integrated with the onset spectrum following a short closure. Such integration might explain the bias toward "t" responses at short closures, assuming that the predominating response after removal of the preceding [s]-noise would be "p" (or, rather, "b"). Even though research on adaptation in the auditory nerve (e.g., Delgutte and Kiang, 1984) predicts spectral contrast rather than integration, a brief additional test was conducted to address this question. Ten randomized repetitions of the seven stimuli (six with bursts and one without) without the initial [s]-noise were presented for identification as "lat," "blat," "dlat," or "glat" to a new group of nine subjects plus the author. The results were mixed. Two subjects responded randomly. Four subjects identified the burstless stimulus as "lat" but labeled all others predominantly "blat." The remaining four subjects (including the author) distributed their responses more evenly, although accuracy was poor (45% correct for stimuli with bursts; 100% for "lat"). These results show, first, that the relative ineffectiveness of the bursts as place cues in the present experiment was not due to the preceding [s]-noise and closure. Second, although some subjects showed a strong bias toward "b" responses, this bias was not so universal as to lend convincing support to the hypothesis that the striking change from "t" to "p" responses with increasing closure duration in the main experiment was due to spectral integration. More likely, the effect of closure duration has a phonetic origin. That is, listeners *expect* labial stops to have longer closures on the basis of their knowledge of natural speech patterns.

Finally, Figure 5 provides a different breakdown of the data which reveals effects of burst amplitude on perceived place of articulation. The conditional percentage of responses in each stop category, averaged over closure durations from 40 to 100 msec, is shown for each of the six burst category/amplitude conditions. "Correct" responses (i.e., responses reflecting the place of articulation that the burst was intended for) are indicated by the cross-hatched bars. It is evident that correct responses in each stop category decreased as burst amplitude was modified, due to a higher percentage of "p" responses for weak bursts and of "t" and/or "k" responses for strong bursts. This result replicates earlier findings of Ohde and Stevens (1983) with synthetic speech. Despite the relative weakness of the present bursts as specific place cues, it appears that burst amplitude contributed to place as well as manner perception.

1 The speaker of the utterances for this experiment, an experienced linguist, produced five tokens each of *splat, stlat,* and *sclat.* Average closure durations were 111, 112, and 68 msec, respectively, revealing unusually long values for [t]. For *spat, stat, scat,* and for *sprat, strat, scrat,* produced by the same speaker, however, closure durations ranked [p] > [k] > [t]. Clearly, more data are needed to determine whether the [–l] context is an exception to the rule that labial closures are longest in duration.

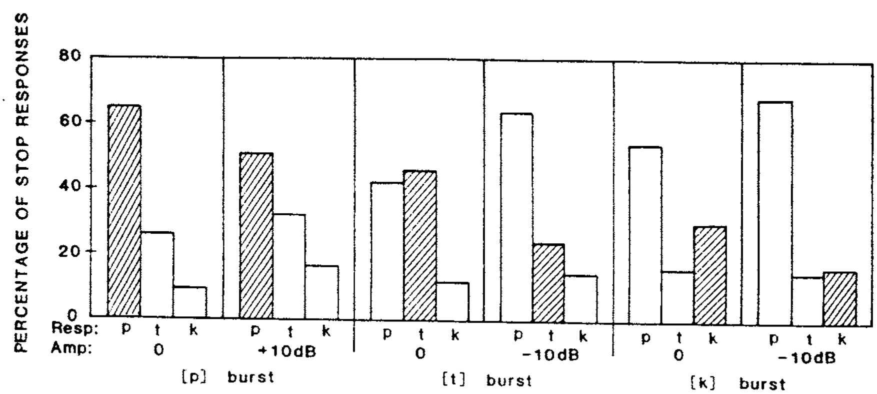

Fig. 5. Response distributions in the six burst conditions, averaged over closure durations.

CONCLUSIONS

The present findings are consistent with many other results suggesting that listeners possess detailed tacit knowledge of the acoustic correlates of phonetic categories (see Repp, 1982, for a review). The perceptual criteria derived from this knowledge apparently specify that labial stops ought to have a longer closure interval than alveolar or velar stops. They also specify that labial stops ought to have weaker release bursts; hence the effect of burst amplitude on place of articulation perception. These perceptual criteria presumably derive from experience with natural speech in its acoustic and articulatory manifestations, and they provide the frame of reference within which speech perception takes place.

REFERENCES

BAILEY, P.J. and SUMMERFIELD, Q. (1980). Information in speech: Observations on the perception of [s]-stop clusters. *Journal of Experimental Psychology: Human Perception and Performance*, **6**, 536-563.

DELGUTTE, B. and KIANG, N.Y.S. (1984). Speech coding in the auditory nerve. IV. Sounds with consonant-like dynamic characteristics. *Journal of the Acoustical Society of America*, **75**, 897-907.

DORMAN, M.F., RAPHAEL, L.J. and LIBERMAN, A.M. (1979). Some experiments on the sound of silence in phonetic perception. *Journal of the Acoustical Society of America*, **65**, 1518-1532.

MENON, K.M.N., JENSEN, P.J. and DEW, D. (1969). Acoustic properties of VCV utterances. *Journal of the Acoustical Society of America*, **46**, 449-457.

OHDE, R.N. and STEVENS, K.N. (1983). Effect of burst amplitude on the perception of stop consonant place of articulation. *Journal of the Acoustical Society of America*, **74**, 706-714.

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. (1984). The role of release bursts in the perception of [s]-stop clusters. *Journal of the Acoustical Society of America*, **75**, 1219-1230.

STATHOPOULOS, E.T. and WEISMER, G. (1983). Closure duration of stop consonants. *Journal of Phonetics*, **11**, 395-400.

STEVENS, K.N. and BLUMSTEIN, S.E. (1978). Invariant cues for place of articulation in stop consonants. *Journal of the Acoustical Society of America*, **64**, 1358-1368.

SUEN, C.Y. and BEDDOES, M.P. (1974). The silent interval of stop consonants. *Language and Speech*, **17**, 126-134.

ZUE, V.W. (1976). *Acoustic Characteristics of Stop Consonants: A Controlled Study*. Lincoln Laboratory Technical Report No. 523 (Lexington, MA).