317

Journal of Phonetics (1980) 8, 269–275

Silence as a cue to the perception of syllable-initial and syllable-final stop consonants

Lawrence J. Raphael

Herbert H. Lehman College, CUNY and Haskins Laboratories, U.S.A.

and

Michael F. Dorman

Arizona State University and Haskins Laboratories, U.S.A.

Received 20th October 1978

Abstract:

To assess the role of silence as a cue for stop manner, we varied the silent closure interval between the stops in utterances of the form $CVC_1 /\!/ C_2V$ and asked listeners to transcribe the utterances. At short closure intervals, the listeners generally did not report the syllable-final stops; only the syllable-initial stops were reported. At longer closure intervals both the initial and final stops were reported. However, the interval necessary to report the syllable-final stop varied as a function of place of articulation and as a function of whether the syllable-final and syllable-initial stops shared place of articulation. When place of articulation was shared, an extremely long closure interval was necessary.

Introduction

In order to produce a stop consonant a speaker must close his vocal tract. This closure, in turn, is reflected in the acoustic signal by a period of relative silence.[1] The co-occurrence of articulatory closure and silence suggests that silence should be an important condition for the perception of stop consonant manner.

There are, in fact, several studies which have shown that silence can provide a sufficient condition for the perception of stop manner. In an early study, Bastian, Eimas & Liberman (1961) demonstrated that the insertion of a sufficiently long interval of silence after /s/ in a naturally-produced "sore" causes listeners to hear "store". In a recent study, Summerfield & Bailey (1977) reported that when silence was imposed between the friction and steady-state vowel formants of synthetic /si/, listeners heard /ski/. A similar manipulation with /su/ caused listeners to perceive /spu/. Thus, there is evidence that silence can be a sufficient

[1] Halle, Hughes & Radley (1957) noted that no energy above 300 Hz is to be found during the "silent" period. The articulatory closure, however, may be accompanied by vocal pulsing during part or all of its duration.

0095-4470/80/030269+07 $02.00/0

condition for the perception of stop manner. For a recent review of this evidence, see Liberman & Studdert-Kennedy (in press).

In contrast, relatively little is known about the role of silence as a necessary condition for the perception of stop manner.[2] What little is known has come indirectly from the identification of syllable-final stops in sequences of two stops. Lisker (1957) commented (in a footnote) that when tape-recordings of [rægɪd] *ragged* and [ræbɪd] were cut and recombined as [rægbɪd] *rag-bid* and [ræbgɪd] *rab-ged*, listeners reported hearing only one stop – the one cued by the post-closure transitions, i.e. [b] and [g] respectively. Abbs (1971), in an unpublished study, described a similar effect. Thus, it appears that silence can be a necessary condition for the perception of the first stop in a sequence of two stops.

We may account for this outcome in the following manner. First, we should note that unreleased final stops are, at best, difficult to identify (Halle, Hughes & Radley, 1957). Further, the closure interval and resulting silence are of greater duration in the production of a sequence of two stops than in the production of an intervocalic stop. (For example, in a corpus of 144 utterances, we have found a mean closure interval of 40 ms in the context VC//V; an interval of 74 ms in the context V//CV; and 127 ms in the context VC_1 $//C_2V$.) We should then note that both Lisker and Abbs used a silent interval appropriate for a single intervocalic stop. Given that interval, it is not surprising that only one stop, the syllable-initial one, was heard. It is reasonable to assume that listeners will report hearing two stops, that is both the syllable-final and syllable-initial one, if the duration of silence in the acoustic signal is increased.

This assumption is supported by the results of a study which assessed the interval of silence necessary to report two identical stops, as in *top pick*, when silence was inserted between the syllables of *topic*. Pickett & Decker (1960) found that when the closure interval in *topic* was lengthened to approximately 250ms, listeners reported hearing *top pick*. Thus, when the silent interval was sufficiently long, both the syllable-final and syllable-initial stop were reported.

The purpose of the present study was to assess directly the duration of silence necessary for the identification of syllable-final and syllable-initial stops in sequences of stops with the same or with different places of articulation. To do this we varied the duration of the silent interval between the stops. In addition, we assessed whether the duration of the silent interval needed for the recognition of one of the stops in a sequence was affected by the identity of the other. Finally, we assessed whether voicing during the closure period affected the interval necessary for stop identification.

Method

A male speaker recorded the syllables /bab/, /bad/, /bag/ and /ba/, /da/, /ga/ into computer memory using the Haskins Laboratories' Pulse Code Modulation System. To produce the stimuli for one test condition we first truncated the closure voicing following the syllable-final stops to 125 ms. We then created all nine possible combinations of CVCs followed by CVs. The duration of the closure voicing between the syllables was systematically varied in 25 ms steps between 0 and 50 ms and in 15 ms steps from 50–125ms. Five tokens of each stimulus in this "voiced closure" condition were recorded. The entire sequence was then randomized and recorded on audio tape with a 3-second interval between successive stimuli. To produce the stimuli for the second test condition, we first truncated voicing following the syllable-final stops to a duration of 15 ms. (We chose this duration after

[2] Halle, Hughes & Radley (1957) state that "silence" is a necessary cue for the perception of a stop sound. . . " They do not, however, offer experimental evidence to substitute their claim.

listening to the isolated CVCs – with 15 ms of closure voicing the syllable-final stops were still highly identifiable.) A series of stimuli was then generated in this "voiceless closure" condition which were analogous, with respect to closure duration, to those of the "voiced closure" condition.

Twelve beginning phonetics students at Lehman College of the City University of New York were told they would hear, and were asked to transcribe, disyllables of the form /bVC CV/, /bV CV/ and /bVC V/, in which the consonants were /b/, /d/, or /g/. The listeners were asked to leave a space between the syllables in their transcriptions so that when only one stop was reported it would be clear to which syllable, the first or the second, they had assigned it.

The listeners were tested in two groups of six. One group heard the stimuli of the "voiced closure" condition first, then the stimuli of the "voiceless closure" condition. The second group heard the two conditions in the reverse order. The stimuli were reproduced in a large sound-attenuated room via a tape recorder and loudspeaker at 72dB SPL.

Results

The results for the identification of syllable-final and syllable-initial stops in the "voiced" and "voiceless closure" conditions were virtually identical. We will therefore present only the data from the "voiceless closure" conditions.

In Fig. 1(a) we have plotted the identification functions for the syllable final-stops. The function for each stop represents the averaged identification of that stop when it was followed by both of the other stops. We shall describe the identification of stops with the same place of articulation later in this section.

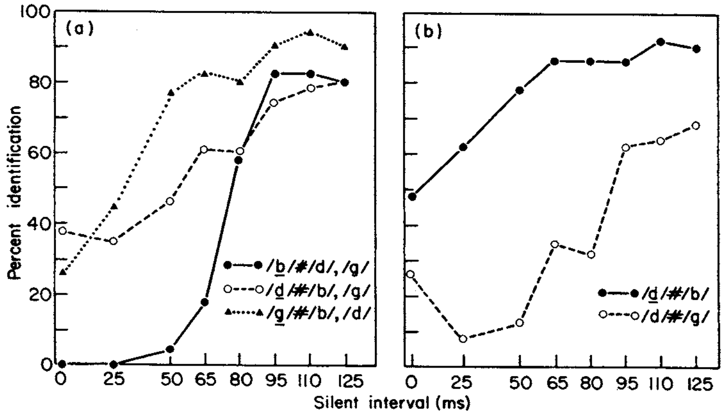

Figure 1 (a) Identification of syllable-final stops (C_1) in CVC_1 C_2V sequences.
(b) Identification of syllable-final /d/ before /b/ and /g/.

In order to identify /b/ at the 50% correct level, listeners needed 77 ms of silence following the syllable-final formant transitions. Recognition of /g/ was accomplished with only 27 ms of silence, while /d/ required 54 ms. Since the identification of /d/ varied as a function of the identity of the following stop, we have plotted in Fig. 1(b) the identification functions for /d/ in the /bad ba/ and /bad ga/ sequences. When /d/ was followed by /g/, recognition of /d/ was 50% correct at 89 ms of silence; when /d/ was followed by /b/, recognition of /d/ was 50% correct when there was no silence at all. In contrast to the contextual constraints on the identification of /d/, the identification of /b/ and /g/ was little affected by the identity of the following stop.

We should emphasize that at the brief silent intervals subjects usually reported hearing no syllable-final stop – it was not the case that subjects reported hearing a stop for which the cues were inappropriate. That is, if a stop was reported at all, it was usually the correct one.

We turn now to the averaged results for the identification of syllable-initial stops. These are shown in Fig. 2(a). We see that identification of /d/ required essentially no silent interval preceding the formant transitions, while /b/ and /g/ required no more than 25 ms of silence. Even in the instance in which recognition was impaired, i.e. at 0 ms interval, performance was depressed on the average by no more than 35%. However, as shown in Fig. 2(b), recognition of /g/ was markedly affected by the identity of the preceding stop: /g/ was heard 90% of the time when preceded by /b/, but only 44% of the time when preceded by /d/. In contrast, recognition of /b/ and /d/ was little affected by the identity of the preceding stop consonant.

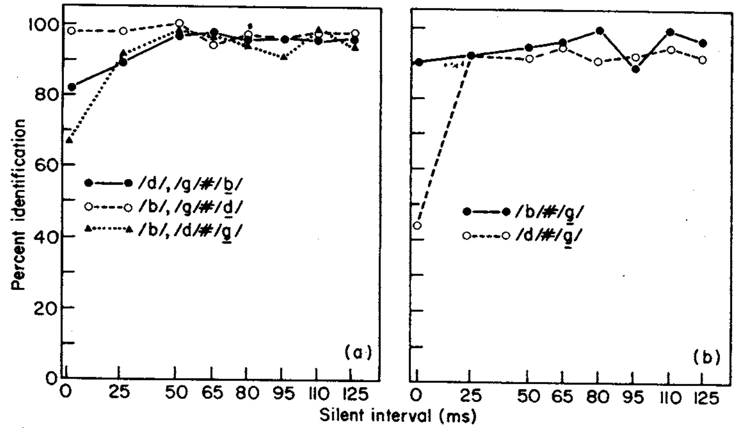

Figure 2 (a) Identification of syllable-initial stops (C_2) in CVC_1 C_2V sequences.
(b) Identification of syllable-initial /g/ following /b/ and /d/.

We should point out that at the zero ms silent interval listeners reported hearing a stop 97% of the time. Since stops were almost always heard, errors were not due to perceptual loss, as was the case for syllable-final stops. Rather, the errors were due to misidentifications.

Finally, we turn to the results for sequences of stops with the same place of articulation. The averaged identification of syllable-final stops is shown in Fig. 3(a); for syllable-initial

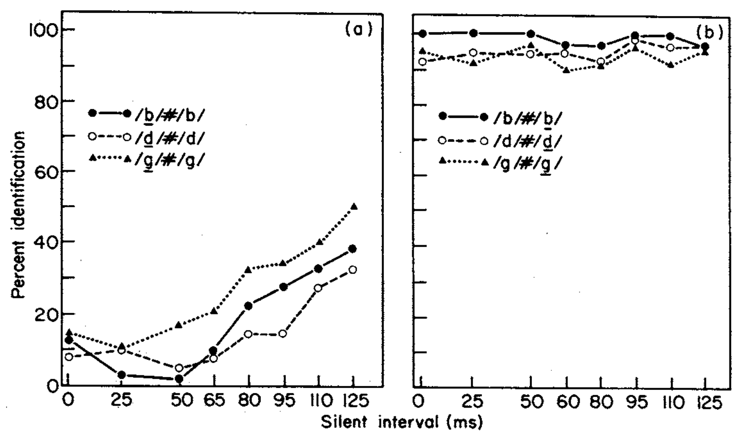

Figure 3 (a) Identification of syllable-final stops in sequences of identical stops.
(b) Identification of syllable-initial stops in sequences of identical stops.

stops in Fig. 3(b). The general outcome is similar to that for sequences of stops with different places of articulation: syllable-final stops require a long silent interval of recognition, whereas syllable-initial stops do not. However, a much longer silent interval is needed for recognition of syllable-final stops in the case of the same-place stop sequences than in the case of different-place stop sequences. Only /g/ reached 50% correct at the longest silent interval used in the experiment. At the same interval, recognition of /b/ was 39% correct and /d/ 33% correct. As might be expected, listeners correctly identified syllable-initial stops more than 90% of the time at all intervals of silence.

Discussion

As we indicated at the outset, silence should be an important condition for the perception of stop manner. Since silence is a necessary acoustic correlate of stop production, if that silence is removed from the signal, an important cue for stop manner is removed, and the perception of a stop should be impaired.

Indeed, we find that it is; in the CVC CV contexts here investigated, silence must follow the syllable-final stops if those stops are to be identified. We suggest that the importance of silence in these instances results from the longer closure duration and silent interval correlated with the production of a two-stop sequence *vis a vis* the production of a single stop. That is, when the closure interval is removed from two-stop sequence the temporal cues are generally appropriate for a single stop – and that is all listeners report. Because implosive transitions provide relatively weak cues for stop perception, the syllable-final stop is the one not reported. The importance of this outcome for theories of stop consonant recognition is that recognition routines for stop manner must take into account acoustic information distributed over some length of time: spectral information limited to the duration of formant transitions cannot fully account for the recognition of stop manner.

The difference in closure interval required for listeners to identify syllable-final [b] [d] and [g] appears to be a function of several variables. One may be the duration of the syllable-final formant transitions. In the present experiments, the duration of the /b/ transitions was shorter than either the /d/ or /g/ transitions, and we find that /b/ needed the greatest duration of following silent interval to be identified. However, the /d/ and /g/ transitions were of similar duration, yet the identification functions for final /d/ and /g/ differed. Another variable may be the similarity in place cues between the syllable-final and syllable-initial stops. We note that for those instances in which there was a large difference in interval needed to identify a syllable-final (or syllable-initial) stop as a function of the following (or preceding) stop, recognition was impaired the most by the stop with the adjacent place of articulation [see Figs 1(b) and 2(b)]. However, further experiments using other vocalic contexts will be necessary to precisely define the nature of this interaction. At all events, it is clear that there is not a fixed interval of silence necessary for identification of syllable-final stops. This implies, of course, that a "feature detector" tuned only to a specific closure interval cannot fully account for the recognition of stop manner.

The identification of syllable-final stops in the context of two stops with the same place of articulation deserves special consideration. We find that in this context the syllable-final stops are not reliably identified even when the closure interval is extended to 125 ms. Indeed, as Repp (1976) has shown, an interval of greater than 200 ms is necessary for accurate identification.[3] To account for this we should look first to articulation to see if the closure interval between two stops with the same place of articulation is of greater duration than the closure interval between two stops with different places of articulation. We find that it is not: for example, in a corpus of 60 utterances of 15 speakers spectrographic analysis reveals that the average closure intervals in "I pai*d B*ailey" and in "I pai*d g*aily" did not differ from that of "I pai*d D*aley." The average intervals were, in fact, identical' 106 ms. How then can we account for the very long interval needed for perception of the syllable-

[3] Repp (1976) has reported that listeners need an interval of approximately 200 ms to consistently report the syllable-final stop in sequences of identical voiced stops.

final stops in the case of a sequence of two stops with the same place of articulation? The most likely account is that in connected discourse, speakers mark the difference between a single stop and a sequence of two stops with the same place of articulation by producing acoustic cues in addition to that of the closure interval. For example, the vocalic nucleus before stop closure may be stressed, resulting in changes in pitch contour, amplitude envelope, or vowel duration. Or, perhaps, the intensity of the closure voicing is increased. At all events, if any of these cues are produced, the stimuli used in the present study were relatively impoverished. (They were spoken and edited so that the only cue differentiating the one- from the two-stop sequences would be closure interval duration.) It may well be the case, then, that because our stimuli did not contain the normally distinctive cues, our listeners needed an extremely long closure to differentiate the one- from the two-stop condition. To confirm this possibility a more detailed acoustic analysis of the two-stop context is needed. We have currently begun such an analysis.

In contrast to the large changes in the identification of syllable-final stops as a function of the duration of the closure interval, the identification of syllable-initial stops was little affected by the duration of the interval. Moreover, when errors in identification were made, they were substitutions rather than omissions. Thus, silence had not only a quantitatively but also a qualitatively different effect on the recognition of syllable-initial and syllable-final stops. We might infer from this outcome that rather different recognition routines mediate the identification of initial and final stops.

We should note, however, a situation in which no stop at all is heard in syllable-initial position when the preceding silent interval is too brief. In one of the several experiments which led to those reported here, the disyllable /bɛd gjɛ/ was produced and the closure interval then varied over the range 0 – 110 ms. When these stimuli were presented for identification, the listeners reported hearing /bɛjɛ/ at silent intervals of up to 90 ms. This outcome seems reasonable if we consider that syllable-initial /gj/ transitions and syllable-initial /j/ transitions are acoustically quite similar. It is not unreasonable to find, in this context, that a substantial closure interval is necessary to differentiate the stop- from the semivowel-initiated syllable. Our interest in this finding is that the phonetic interpretation of the information carried on syllable-initial transitions is, at least in certain instances, dependent on acoustic information preceding the release of stop closure. Thus, it can be the case for syllable-initial stops, as it is for syllable-final stops, that recognition routines must take into account acoustic information distributed over time.

Finally, we note that the presence of voicing throughout the closure interval did not affect the interval necessary to identify either the syllable-final or syllable-initial stops. Since in natural discourse voicing does not occur consistently throughout the closure interval, we should expect that the presence of continuous closure voicing would be of little added benefit to the presence of a silent interval. Our results support that expectation.

The research reported here was supported by grant HD-01994 to Haskins Laboratories from the National Institute of Child Health and Human Development. We thank Tony Levas and Suzi Pollock for assistance in collecting and tabulating the data.

References

Abbs, M. (1971). A study of cues for the identification of voiced stop consonants in intervocalic position. Doctoral dissertation, University of Wisconsin (unpublished)

Bastian, J., Eimas, P. & Liberman, A. (1961). Identification and discrimination of a phonemic contrast induced by silent interval. *Journal of the Acoustical Society of America* 33, 842(A).

Halle, M., Hughes, C. & Radley, J-P. A. (1957). Acoustic properties of stop consonants. *Journal of the Acoustical Society of America*, 29, 107–116.

Liberman, A. M. & Studdert-Kennedy, M. Phonetic perception, (in press). In *Handbook of Sensory Physiology* Vol. VIII, *Perception*, (Held, R., Leibowitz, H. & Teuber, H-L., Eds). Heidelberg: Springer-Verlag.

Lisker, L. (1957). Closure duration and the voiced-voiceless distinction in English. *Language* 33, 42–49.
Pickett, J. & Decker, L. (1960). Time factors in perception of a double consonant. *Language and Speech* 3, 11–17.
Repp, B. (1976). Perception of implosive transitions in VCV utterances. *Haskins Laboratories, Status Report on Speech Research, SR–48*, 209–233.
Summerfield, A. & Bailey, P. (1977). On the dissociation of spectral and temporal cues for stop consonant manner. *Journal of the Acoustical Society of America* **61**, S–46(A).