299

Distribution of acoustic cues for stop consonant place of articulation in VCV syllables

M. F. Dorman

Department of Speech and Hearing Science, Arizona State University, Tempe, Arizona 85281

Lawrence J. Raphael

Haskins Laboratories and Lehman College of the City University of New York, Bronx, New York 10468

(Received 27 March 1979; accepted for publication 2 January 1980)

Most theoretical accounts of the identification of stop consonant place of articulation have focused on how bursts and formant transitions conspire to signal place in CV syllables. In the present series of experiments we have examined the identification of place in VCV syllables and have found that not only do the burst and opening transitions affect the judgement of place, but so also do the closing transitions and the duration of the closure interval. This outcome is consistent with the outcomes of many other experiments in showing that there are multiple acoustic events which bear on the identification of a given phone and that those acoustic events are distributed over time. Theoretical accounts of place identification based on data of this kind may prove more viable than theories based on data from the identification of place in the absolute, syllable-initial position.

PACS numbers: 43.70.Dn, 43.70.Ve

INTRODUCTION

In order to effectively search for invariant acoustic cues to stop consonant place of articulation, investigators must first determine the time interval over which the relevant information in the signal is distributed. Faced with this problem, most have chosen to examine the simplest case—that of a stop in absolute, syllable-initial position. In this position the most salient acoustic correlates of stop production are the release burst and formant transitions. Consequently, theories of place identification have focused on how the burst and formant transitions conspire to signal place of articulation (for example, see Cole and Scott, 1974; Cooper *et al.*, 1952; Delattre *et al.*, 1955; Dorman *et al.*, 1977; Fant, 1969; Fischer-Jørgensen, 1972; Kuhn, 1975; and Stevens and Blumstein, 1978).

We note, however, that absolute syllable-initial position is not the most common position in which a stop occurs. Rather, the more usual position is that in which a stop is preceded by another phone. In one of these contexts—a VCV syllable—the stop gesture not only generates a release burst and (opening) formant transitions following the stop closure, but also generates formant transitions (closing transitions) into the closure and, of course, the closure interval itself. In syllables of this kind we should wonder whether the closing transitions and closure interval affect the identification of place of articulation.

To generate VCV stimuli which would be potentially sensitive to the effects of closing transitions, we have made use of the fact that release bursts can be "spliced out" of a CV syllable and then recombined with a different CV syllable (Fischer-Jørgensen, 1972). In syllables so constructed, the burst can specify one place of articulation and the transitions yet another. Since many of the stimuli created in this manner are ambiguous as to place of articulation, they are potentially sensitive to the effect of closing transitions when placed in the context of a VCV syllable. In the experiments which follow, we compare the identification of place in (i) CV syllables composed of conflicting burst and (opening) transition cues and in (ii) VCV syllables in which the closing transitions signal the same place of articulation as the opening transitions, but in which the burst signals a different place of articulation. Our interest was whether the formant transitions which occur 65–85 ms before the burst release would affect the identification of place of articulation. We have also assessed the effect of closing transitions on the place judgment in VCV's in which the natural closure interval (approximately 80 ms) was shortened to 40 ms. Our interest was to determine whether the closure transitions would have more of an effect on the judgment of place of articulation when they were closer in time to the burst and opening transitions.

I. METHOD

A. Subjects

The listeners were 18 undergraduates at Herbert H. Lehman College of the City University of New York who volunteered for an experiment in speech perception.

B. Stimulus preparation

A male speaker recorded /b/, /d/, and /g/ with the vowels /i/, /a/, and /u/ in the sentence frame, "The little VCV dog." Using procedures described in detail in Dorman *et al.* (1977), the VCV's were excised from the carrier sentence and stored in computer memory. To create stimuli with conflicting burst and transition cues, CV's were first extracted from the VCV's and the bursts and vocalic portions of the syllables stored separately. The stimuli were then recombined in a manner such that each burst was paired with each of the other vocalic portions.

To create VCV stimuli in which the transitions cued one place of articulation and the burst another, the original VCV's were edited and the preclosure, vocalic portions of the utterances were stored in memory. The preclosure, vocalic sections were then combined with the CV stimuli so that in each stimulus the closing and opening transitions specified one place of articulation and the burst a different place. The durations of the closure intervals were set at those found in the original VCV's—these intervals ranged in duration from 66 to 88 ms. (We will refer to this condition as the $V_{80}CV$ condition.)

To create yet another set of stimuli, the stimuli generated for the $V_{80}CV$ condition were edited so that the closure interval for each stimulus was reduced to 40 ms. (We will refer to this condition as the $V_{40}CV$ condition.) Thus, three sets of stimuli were created: a CV set in which the burst and transitions specified different places of articulation; a VCV set in which the closing and opening transitions specified one place of articulation and the burst yet another (and in which the closure intervals averaged approximately 80 ms in duration); and a VCV set in which the closure intervals were exactly 40 ms in duration.

Four tokens of each stimulus in each stimulus set were generated. The stimuli were randomized within each set and then recorded separately on audio tape.

C. Procedure

In the first portion of the experiment, ten subjects listened to the CV and $V_{80}CV$ stimuli. The subjects listened in two groups of five each; the stimulus conditions were randomized across the two listening groups. In the second part of the experiment, eight subjects listened to the $V_{40}CV$ stimuli. In both parts of the experiment, the listeners were instructed to write the identity of the stop consonant on an answer sheet.

To familiarize the listeners with the task, ten stimuli were played to the subjects before each of the stimulus conditions. The experiment was conducted in a large, sound attenuated room. The stimuli were reproduced at a comfortable listening level via loudspeaker.

II. RESULTS AND DISCUSSION

The percentages of burst-cued responses in the CV, $V_{80}CV$, and $V_{40}CV$ conditions are shown in Table I. There we find that fewer burst-cued responses were recorded in the $V_{80}CV$ condition than in the CV condition ($t_{18}=6.02$, $p<0.005$). We also find that fewer burst-cued responses were recorded in the $V_{40}CV$ condition than in the CV condition ($t_{16}=5.60$, $p<0.005$). Thus, the closing transitions contributed significantly to the judgment of place of articulation (see also Repp, 1978). This was so even when the transitions were removed approximately 80 ms in time from the onset of the CV syllable. It would appear that recognition routines for stop consonant place of articulation are sensitive to more information than that transmitted at the onset of the CV syllable. Indeed, it seems as if vocal tract information is tracked continuously over the duration of the entire stop gesture and that a decision about place is made on the basis of information which occurs both before and after the stop closure.

Inspection of Table I also reveals that the closing transitions in the $V_{40}CV$ condition were no more effective in determining place of articulation than were the transitions in the $V_{80}CV$ condition. The outcomes of the two conditions were not, however, identical: as shown in Table II, the distribution of responses differed. In the $V_{40}CV$ condition, the percentage of /d/ responses increased at the expense of the /b/ responses. Quite unexpectedly, this increase in /d/ responses was most marked in stimulus situations characterized by neither /d/ bursts nor /d/ transitions. The percentages of /d/ responses to these stimuli are displayed in Table III. We see that there were few "inappropriate" /d/ responses in the $V_{80}CV$ condition and that the number of "inappropriate" /d/ responses increased dramatically (and significantly) in the $V_{40}CV$ condition ($t_{16}=6.24$, $p<0.0001$).

We can offer an account for this increase in /d/ responses by referring to the vocal tract behavior specified by a very brief closure interval. Umeda (1977) has reported that for stops in intervocalic position the duration of /d/ (av = 31 ms) is shorter than that for either /b/ (av = 60 ms) or /g/ (av = 55 ms). Moreover, Zue and Laferriere (1979) have reported that a flapped /d/ ranges in duration from 10—40 ms. Thus, the brief closure interval in the $V_{40}CV$ condition would be consistent with the production and perception of intervocalic /d/ or /ɾ/. It appears that not only can the closing transitions, bursts, and opening transitions contribute to the identification of place, but so also can the duration of the closure interval. In short, each of the distributed acoustic consequences of stop consonant articulation can be shown to have cue value. Indeed, it appears to be a *general* characteristic of speech perception that the distributed acoustic consequences of an articulatory gesture have cue value (see, for example, Bailey and Summerfield, in press; Dorman *et al.*, in press; Repp *et al.*, 1978; Lisker, 1978).

It is not without precedent that listeners should give /d/ responses to stimuli in which the closure interval is very brief. Port (1976) has reported a similar effect. When the closure interval in "rapid" is reduced, listeners first report "rabid" and then "ratted." These two stimulus situations fit into a class of situations in

TABLE I. Percentage of burst-cued responses as a function of vocalic environment for three stimulus conditions.

	CV	$V_{80}CV$	$V_{40}CV$
/i/	23	09	08
/a/	42	27	24
/u/	41	25	35

TABLE II. Percent response as a function of place of articulation in the $V_{80}CV$ and $V_{40}CV$ conditions.

	$V_{80}CV$	$V_{40}CV$
/b/	42	22
/d/	37	55
/g/	21	23

TABLE III. Percentage of /d/ responses as a function of vocalic environment to stimuli having neither /d/ burst nor /d/ transitions.

	$V_{80}CV$	$V_{40}CV$
/i/	09	41
/a/	14	11
/u/	16	79

which the identification of place is not well predicted by place decisions made on the basis of the short-term spectra which immediately precede and follow a silent interval. Consider, for example, that when silence is interpolated between /s/ and "lit" listeners report "split" (Dorman *et al.*, 1979). Consider, also, the effect of inserting silence between /s/ and steady-state /i/ and /u/: in the former instance listeners report /ski/ while in the latter they report /spu/ (Bailey and Summerfield, in press). In these several situations the identification of place appears to be based on the *change* in spectrum over time, and does not seem to be based on a combination of place decisions made on the basis of the spectra which precede and follow the silent interval.

To summarize, we have found that listeners, when asked to identify place of articulation in VCV syllables, are sensitive (i) to the transitions which precede the closure interval, (ii) to the duration of the closure interval, and (iii) to the bursts and transitions which follow the closure interval. As we indicated earlier, this outcome is consistent with the outcomes of many other experiments in showing that there are multiple acoustic events which bear on the identification of a given phone and that those acoustic events are arrayed in time. It may well prove to be the case that theories of place identification based on data of this kind will prove more viable than those based on data from the identification of place in absolute, syllable-initial position.

ACKNOWLEDGMENT

Research supported by Grant HD-01994 from NIH to Haskins Laboratories.

Bailey, P., and Summerfield, A. (in press). "The Information in Speech: Observations on the perception of [s] and stop clusters," J. Exp. Psychol. Human Perception and Performance.

Cole, R. A., and Scott, B. (1974). "Toward a theory of speech perception," Psychol. Rev. 81, 348–374.

Cooper, F., Delattre, P., Liberman, A., Borst, J., and Gerstman, L. (1952). "Some experiments on the perception of synthetic speech sounds," J. Acoust. Soc. Am. 24, 597–606.

Delattre, P., Liberman, A., and Cooper, F. (1955). "Acoustic loci and transitional cues for consonants," J. Acoust. Soc. Am. 27, 769–773.

Dorman, M., Raphael, L., and Isenberg, D. (in press). "Acoustic cues for a fricative-affricate contrast in word-final postion," J. Phonetics.

Dorman, M., Raphael, L., and Liberman, A. (1979). "Some experiments on the sound of silence in phonetic perception," J. Acoust. Soc. Am. 65, 1518–1532.

Dorman, M., Studdert-Kennedy, M., and Raphael, L. (1977). "Stop consonant recognition: Release bursts and formant transitions as functionally equivalent, context sensitive cues," Percept. Psychophys. 22, 109–122.

Fant, G. (1969). "Stops in CV syllables," Speech Transmission Laboratory QPRS, Royal Institute of Technology, Stockholm 4, 1–25.

Fischer-Jørgensen, E. (1972). "Perceptual studies of Danish stop consonants," Ann. Rep. Inst. Phonetics, Univ. Copenhagen 6, 75–168.

Kuhn, G. (1975). "On the front cavity resonance and its possible role in speech perception," J. Acoust. Soc. Am. 58, 578–585.

Lisker, L. (1978). "Rapid vs rabid: A catalogue of acoustic features that may cue the distinction," Haskins Laboratories Status Report on Speech Research SR-54, 127–132.

Repp. B. (1978). "Perceptual integration and differentiation of spectral cues for intervocalic stop consonants," Percept. Psychophys. 24, 471–485.

Repp., B. Liberman, A., Eccardt, T., and Pesetsky, D. (1978). "Perceptual integration of acoustic cues for stop, fricative, and affricate manner," J. Exp. Psychol. Human Perception and Performance 4, 621–637.

Stevens, K. (1975). "The potential role of property detectors in the perception of consonants," in *Auditory Analysis and Perception of Speech*, edited by G. Fant and M. A. A. Tatham (Academic, New York).

Stevens, K., and Blumstein, S. (1978). "Invariant cues for place of articulation in stop consonants," J. Acoust. Soc. Am. 64, 1358–1368.

Umeda, N. (1977). "Consonant duration in American English," J. Acoust. Soc. Am. 61, 846–858.

Zue, V., and Laferriere, M. (1979). "Acoustic study of medial /t, d/in American English," J. Acoust. Soc. Am. 66, 1039–1050.