*Perception & Psychophysics*
*1977, Vol. 22 (2), 109-122*

# Stop-consonant recognition: Release bursts and formant transitions as functionally equivalent, context-dependent cues

M. F. DORMAN
*Arizona State University, Tempe, Arizona 85281*
*and Haskins Laboratories, New Haven, Connecticut 06510*

M. STUDDERT-KENNEDY
*Queens College and Graduate Center of The City University of New York, New York, New York 10036*
*and Haskins Laboratories, New Haven, Connecticut 06510*

and

L. J. RAPHAEL
*Lehman College and Graduate Center of The City University of New York, New York, New York 10036*
*and Haskins Laboratories, New Haven, Connecticut 06510*

Three experiments assessed the roles of release bursts and formant transitions as acoustic cues to place of articulation in syllable-initial voiced stop consonants by systematically removing them from American English /b,d,g/, spoken before nine different vowels by two speakers, and by transposing the bursts across all vowels for each class of stop consonant. The results showed that bursts were largely invariant in their effect, but carried significant perceptual weight in only one syllable out of 27 for Speaker 1, in only 13 syllables out of 27 for Speaker 2. Furthermore, bursts and transitions tended to be reciprocally related: Where the perceptual weight of one increased, the weight of the other declined. They were thus shown to be functionally equivalent, context-dependent cues, each contributing to the rapid spectral changes that follow consonantal release. The results are interpreted as pointing to the possible role of the front-cavity resonance in signaling place of articulation.

The present paper deals with an aspect of the problem of perceptual constancy—the invariance problem—in speech recognition. At the level of phoneme recognition, the problem is manifest in the variety of acoustic signals that may be categorized as the same phoneme (Liberman, Cooper, Shankweiler, & Studdert-Kennedy, 1967; Studdert-Kennedy, 1974, 1976). Here we are concerned with the variation in acoustic cues for a given stop consonant as a function of the following vowel. Many studies have demonstrated that formant transitions are generally sufficient cues for stop-consonant recognition. Since the shape of these transitions varies with the following vowel, accounts of stop-consonant recognition have generally emphasized the role of context-conditioned cues (perhaps relational invariants) within the consonant-vowel syllable.

We thank Alvin Liberman, Paul Mermelstein, and, especially, Gary Kuhn for their careful comments on an early draft of this paper. We thank Agnes McKeon for her expert preparation of the figures. We also thank Suzi Pollack and Tony Levas for help in running subjects and tabulating data. This research was supported in part by NICHD Grant HD-01994 and BRSG Grant RR-05596.

Recently, however, Cole and Scott (1974a, 1974b) have suggested that stop consonants before different vowels may be recognized in terms of a context-independent acoustic cue (or simple invariant), namely, the burst produced at the release of stop-consonant occlusion.

In the following experiments, we explore these cues in some detail with natural speech. We assess, first, the extent to which separable components of the complex of acoustic cues for initial, voiced stop consonants—the release burst, the devoiced and the voiced formant transitions—are sufficient cues for the perception of place of articulation. We ask, second, whether one of the components—the burst—is a functionally invariant cue for stop-consonant recognition. Finally, we discuss the implications of our results for an account of stop-consonant recognition.

## Acoustic Segmentation of Stop-Consonant-Vowel Syllables

Acoustic analysis of /bV,dV,gV/ syllables, reveals five qualitatively distinct segments before a stable vowel formant pattern is reached (cf. Fischer-

Jørgensen, 1954; Halle, Hughes, & Radley, 1957; Fant, Note 1; Fischer-Jørgensen, Note 2): (1) a period of occlusion (usually silent, though occasionally voiced); (2) a transient explosion (usually less than 20 msec) produced by shock excitation of the vocal tract upon release of occlusion; (3) a very brief (0-10 msec) period of frication, as articulators separate and air is blown through a narrow (though widening) constriction, as in the homorganic fricative; (4) a brief period (2-20 msec) of aspiration, within which may be detected noise-excited formant transitions, reflecting shifts in vocal-tract resonances as the main body of the tongue moves toward a position appropriate for the following vowel; (5) voiced formant transitions, reflecting the final stages of tongue movement into the vowel during the first few cycles of laryngeal vibration. Since we are only concerned with stop consonants in the present study, we shall not consider the role of the first segment (occlusion) which serves to distinguish stops from vowels and other consonants. Furthermore, since the explosion and frication, even if separable on an oscillogram or spectrogram, are probably not discriminable by ear, we shall treat them in what follows as a single burst of energy, lasting some 2-30 msec.

The fourth segment (aspirated and devoiced formant transitions), although usually distinguishable on an oscillogram with a high resolution time scale, is not always readily apparent on a spectrogram (see Figure 1). Investigators have therefore tended to discount it as an acoustic cue[1] and to concentrate attention on the burst and on the voiced formant transitions. The present paper attempts to redress the balance by treating this segment as a separable component of the cue complex.

**Bursts and Transitions as Cues for Stop Consonants**

Research with synthetic speech has revealed that both bursts and voiced formant transitions may serve as separate cues to place of articulation of initial /b,d,g/. Many studies have shown that transitions of the second and third formants are sufficient cues for the place distinction (for example, Delattre, Liberman, & Cooper, 1955; Liberman, Delattre, Cooper, & Gerstman, 1954), and these are, in fact, the standard cues used in speech synthesis. It is important to note that—since the acoustic shape of formant transitions varies as a function of the following vowel—formant transitions are necessarily context-dependent cues for stop consonants. The same is true of velar bursts. Hoffman (1958) found that, while bursts centered at frequencies above 3,000 Hz acted as cues for /d/, burst cues for /g/ lay near the second formant of the vowel and were, therefore, context-dependent (cf. Liberman, Delattre, & Cooper, 1952). Hoffman could find no burst that would serve as a powerful cue for /b/, but this may have reflected, in part, the deficiencies of his synthesizer, rather than of natural speech.

In fact, attention has recently turned to the question of how cues isolated in synthetic speech experiments act and interact in naturally produced speech. Cole and Scott (1974b) have argued that, while formant transitions do provide essential phonetic information for the consonant phonemes, "the major role of transitional cues is to provide information about the temporal order of phonetic segments within a syllable" (p. 349). [See, also, Day (1970) and Liberman, Mattingly, and Turvey (1972) for discussions of the role of formant transitions in providing information about temporal order in speech.] Cole and Scott (1974b) go further to suggest that, for /b,d,g/ (1974a) or /b,d/ (1974b) in stressed syllable-initial position, the invariant place cue lies in the initial noise energy (burst and aspiration) before the onset of laryngeal vibration.

The latter claim drew apparent support from an experiment. Using a tape-splicing procedure to remove formant transitions from /bi, bu, di, du, gi, gu/,

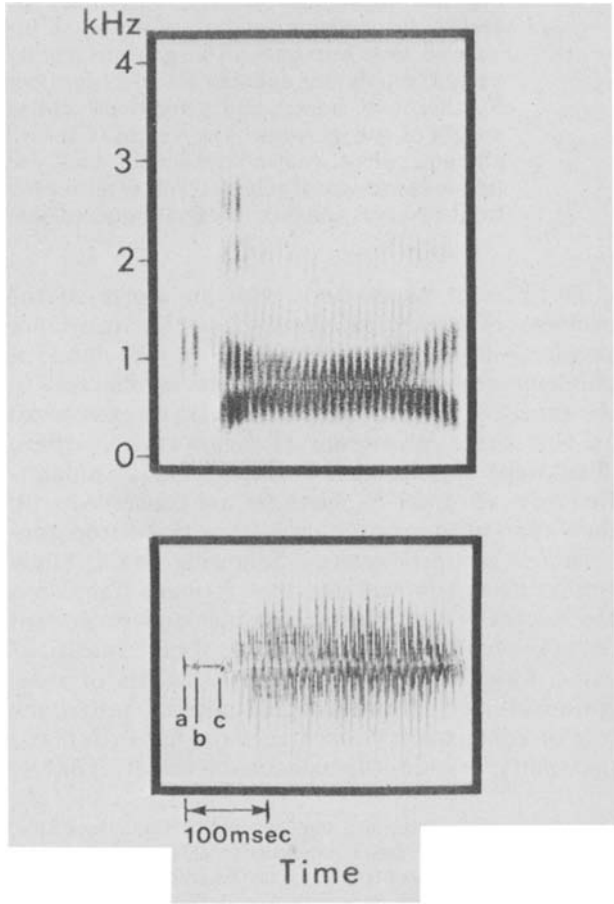


**Figure 1. A spectrogram of the syllable /gʌd/, spoken by Speaker 2 (top). An oscillogram of the same utterance is shown at the bottom: burst (a-b) and aspiration (b-c) duration and the onset of voicing (c) are indicated by vertical lines.**

thus leaving burst and aspiration followed by steady-state vowel, Cole and Scott (1974a) found that recognition of the syllables remained essentially unimpaired. Moreover, when the initial energy from /bi/ was transposed to /u/, or the initial energy from /bu/ was transposed to /i/, recognition was again unimpaired. This relation was also reported for /di/ and /du/. However, for the /gi/ to /u/ transposition, only 21% correct responses were reported, and 90% of the error responses were /b/. The /gu/ to /i/ transposition fared better with 82% correct responses. Cole and Scott (1974a, p. 101) concluded that "stop consonants may be recognized before different vowels ... in terms of invariant acoustic features.

Implicit in this conclusion is the assumption that bursts are not only invariant, but sufficient cues to place articulation. For, if they are not sufficient, it matters little whether or not they are invariant. However, it has been known for a number of years, both from synthesis experiments (Hoffman, 1958; Liberman et al., 1952) and from the acoustic analysis of natural speech (Fischer-Jørgensen, 1954; Halle et al., 1957; Fant, Note 1; Fischer-Jørgensen, Note 2), that, while release burst spectra vary systematically with the following vowel for initial velar stops, they are largely invariant for initial labial and apical stops. The most novel aspect of Cole and Scott's (1974a) conclusion is, therefore, that burst cues are *sufficient* for recognition of stop-consonant place of articulation. Several considerations suggest that these claims may merit careful attention.

First, Cole and Scott (1974a) made no attempt to separate the release burst from the context-conditioned voiceless aspiration. If we examine the spectrograms of Figure 2 in Cole and Scott (1974a; p. 104), we see obvious acoustic differences between the transposed portions of syllable pairs. Had listeners been asked to identify the vowels of these transposed portions, they might well have been able to do so, thus demonstrating that the experimenters had transposed not consonants but whispered consonant-vowel (CV) syllables. In fact, Winitz, Scheib, and Reeds (1972), in an experiment closely related to that of Cole and Scott (1974a), have reported precisely this result for the (admittedly longer) burst and aspiration portions of initial /p,t,k/.

A second reason to question Cole and Scott's conclusion is that they transposed energy for the voiced stops between only two vowels. Since most dialects of English contain approximately 16 distinctive vowel nuclei, transpositions over two vowels represent a rather meager test of their hypothesis. Indeed, Fischer-Jørgensen (Note 2) has shown for Danish /b,d,g/ that bursts are effective cues for /b/ and /g/ before /i/ and /u/ but not before /a/, while for /d/ a burst is an effective cue before /i/

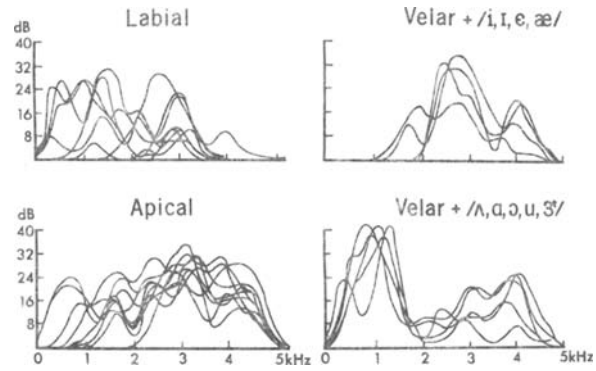


**Figure 2. Spectra of bursts from syllable-initial /b,d,g/ spoken before nine vowels by Speaker 2. The velar spectra have been divided into front and center-back vowel series.**

but not before /a/ or /u/. Thus, there is already evidence from a language other than English that bursts are not adequate cues for the distinction among these stop consonants in certain vowel environments.

A third, and perhaps the most important, consideration is that articulatory gestures associated with initial labial, apical, and palatovelar stop consonants before a variety of different vowels give rise to systematic variations in syllabic acoustic structure that make the hypothesis of any single sufficient cue (whether burst or transition) across all environments extremely unlikely. Every researcher who has worked on speech synthesis is familiar with the fact that a "good" rendering of a particular phonetic segment may require different acoustic patterns in different phonetic environments. For example, good initial, voiced apical stops are more readily synthesized with a burst before high, front vowels, but with extensive voiced formant transitions before back vowels. Furthermore, even though isolated cues may serve a valid experimental function, natural speech typically displays a complex of cues with varying acoustic salience and therefore, we may suspect, varying perceptual weight in different environments. It will simplify the description and interpretation of our experimental results if we here spell out the most important acoustic variations and some possible perceptual consequences. For more detail than we can give here, the reader is referred to Fant (1959), Fischer-Jørgensen (1954), Flanagan (1972), Halle et al. (1957), Heinz (1974), Klatt (1975), and Fischer-Jørgensen (Note 2).

**Release burst energy.** The energy (duration × intensity) in the transient release and its following frication varies as a function of several factors, including the cross-sectional area of the constriction just after release, the resonant cavity in front of the point of release, and, perhaps, the release gesture itself. Thus, /b/, for which there is essentially no front cavity and for which the release gesture is rapid

(Fujimura, 1961; Kuehn, 1973), usually displays a weak transient and virtually no frication, while /g/, for which the cross-sectional area between tongue and palate is relatively large, for which the front cavity is narrowly tuned, and for which tongue release is relatively slow, displays the longest burst of the three stops, including on occasion, as Fischer-Jørgensen (1954) noted, a "double" release transient (see Figure 1) [perhaps due to a suction effect (Fant, Note 1)]. Burst energy for /d/, with a smaller cross-sectional area between tongue and alveolar ridge and a more broadly tuned front cavity than for /g/, but with a release velocity roughly the same as for /b/, falls midway. We might then predict increasing energy in—and therefore perceptual importance of—the burst as the point of occlusion moves back in the mouth.

Cutting across all three places of articulation, however, are possible variations in burst energy due to coarticulation with the following vowel. A major contrast is between front unrounded vowels, such as /i,ɪ,ɛ/ and center-to-back rounded vowels, such as /ɜ,ɔ,u/. For /b/, increased cross-sectional area of the constriction just after release may give rise to a longer, and so more effective, release burst before rounded than before unrounded vowels. For /d/, elongation of the front cavity before rounded vowels is likely to yield lower burst intensity than before unrounded vowels. For /g/, the effect of front cavity elongation before rounded vowels may be counteracted by increased cross-sectional area of the palatiolingual constriction and narrower front-cavity tuning than before unrounded vowels. Thus, if we assume that acoustic energy at least partially determines auditory salience and perceptual weight, we might expect the release burst to play a more important role before rounded than before unrounded vowels for /b/ and /g/, but exactly the reverse for /d/.

**Release burst spectrum**. Spectral sections taken through the release burst of /b/ in nine vocalic environments generally show a broad curve with peaks over low frequencies, below approximately 2,000 Hz (see Figure 2); the low-frequency peaks tend to be stronger before rounded than before unrounded vowels. For /d/, the spectral curve is broad and of a relatively high intensity, with peaks generally over higher frequencies, above approximately 2,000 Hz (see Figure 2); the peaks tend to shift upward before unrounded vowels and to be somewhat stronger than before rounded vowels. Apart from these minor rounding dependencies, /b/ and /d/ bursts are relatively unaffected by the following vowel. We may note, however, that these bursts do not occupy invariant positions on the frequency scale in relation to their following vowels; the apical burst is spectrally continuous with $F_2/F_3$ of the high front vowels, but spectrally distinct from $F_2$ of the back rounded vowels; for the labial burst these relations tend to be reversed. The spectrum of the velar burst, on the other hand, is generally narrower and of a relatively high intensity, with its main peak close to $F_3$ of a following front vowel and close to $F_2$ of a following back vowel, reflecting the changes from the front articulation of /gi/ to the articulation of /gu/. Thus, while labial and apical bursts are largely invariant on the frequency scale, but variable in relation to following vowel, velar bursts are more or less invariant in relation to the following vowel but variable on the frequency scale. [For a more comprehensive description of burst spectra in different vocalic environments, see Zue (1976).] The possible perceptual implications of these facts will become clear when we report our results.

**Formant-transition range and energy**. At least three articulatory factors underlie variations in formant-transition structure. First are variations in the extent of transitions as a function of place of articulation and following vowel. For bilabials, transitions are longer (and so presumably more effective cues) before unrounded than before rounded vowels. For apical stops, the distance between point of occlusion and vowel-target configuration varies, so that we might expect both devoiced and voiced transitions to be more effective cues to /d/ before back vowels, where transitions are relatively long, than before front vowels, where they are relatively short. Similarly, for velars the determining factor is degree of similarity between the velar tongue constriction and that of the following vowel; in general, close vowels (such as /i/) will have relatively little transition, and open vowels (such as /a/), a more marked transition.

A second factor affecting formant-transition structure is the onset of voicing relative to onset of the release burst [i.e., VOT (Lisker & Abramson, 1964)]. An increase in the time taken for consonantal release (i.e., in release burst duration) leads to an increase in the time taken for development of a transglottal pressure drop sufficient to initiate voicing, and so to an increase in VOT. If VOT is increased, transitions into the following vowel may be largely complete at voicing onset, so that the duration of devoiced transitions relative to voiced transitions is increased. Since release burst duration (and so VOT) typically increases from labial to apical to velar points of articulation (Lisker & Abramson, 1964; Table 1), we may reasonably predict corresponding increases in the perceptual weight attached to devoiced transitions.

Finally, speakers differ in vocal-tract shape and dimensions, as well as in articulatory habits (see, for example, Bell-Berti, 1975; Ladefoged, DeClerk, Lindau, & Papcun, 1972), and even two phonetically identical utterances of the same speaker are prob-

**Table 1**
**Release Burst and "Aspiration" Durations, and Voice Onset Times* in Milliseconds, for /b,d,g/ Followed by Nine Vowels for Two Speakers****

| | Speaker 1 | | | Speaker 2 | | |
|---|---|---|---|---|---|---|
| Syllable | Release Burst | Aspiration | VOT | Release Burst | Aspiration | VOT |
| /bid/ | 4 | 5 | 9 | 6 | 4 | 10 |
| /bɪd/ | 5 | 1 | 6 | 9 | 9 | 15 |
| /bɛd/ | 5 | 5 | 10 | 9 | 4 | 13 |
| /bæd/ | 3 | 2 | 5 | 7 | 8 | 15 |
| /bʌd/ | 5 | 2 | 7 | 9 | 4 | 13 |
| /bad/ | 3 | 1 | 4 | 11 | 4 | 15 |
| /bɔd/ | 3 | 6 | 9 | 6 | 6 | 12 |
| /bud/ | 7 | 4 | 11 | 10 | 14 | 24 |
| /bɝd/ | 4 | 5 | 9 | 10 | 13 | 23 |
| Mean | 4.3 | 3.4 | 7.7 | 8.6 | 7.3 | 16.0 |
| /did/ | 7 | 1 | 8 | 25 | 12 | 37 |
| /dɪd/ | 7 | 5 | 12 | 15 | 6 | 21 |
| /dɛd/ | 6 | 7 | 13 | 12 | 13 | 25 |
| /dæd/ | 8 | 6 | 14 | 13 | 8 | 21 |
| /dʌd/ | 6 | 6 | 12 | 10 | 5 | 15 |
| /dad/ | 5 | 6 | 11 | 7 | 8 | 15 |
| /dɔd/ | 5 | 7 | 12 | 8 | 7 | 15 |
| /dud/ | 6 | 6 | 12 | 5 | 10 | 15 |
| /dɝd/ | 7 | 7 | 14 | 10 | 15 | 25 |
| Mean | 6.3 | 5.7 | 12.0 | 11.7 | 9.3 | 21.0 |
| /gid/ | 7 | 12 | 19 | 25 | 10 | 35 |
| /gɪd/ | 21 | 3 | 24 | 17 | 18 | 35 |
| /gɛd/ | 7 | 11 | 18 | 22 | 14 | 36 |
| /gæd/ | 12 | 6 | 18 | 29 | 7 | 36 |
| /gʌd/ | 14 | 9 | 23 | 18 | 13 | 31 |
| /gad/ | 11 | 6 | 17 | 21 | 25 | 46 |
| /gɔd/ | 13 | 7 | 20 | 20 | 15 | 35 |
| /gud/ | 8 | 11 | 19 | 20 | 15 | 35 |
| /gɝd/ | 12 | 11 | 23 | 20 | 21 | 41 |
| Mean | 11.7 | 8.4 | 20.1 | 21.3 | 15.3 | 36.7 |

**Voice onset time (VOT) is the sum of release burst, affrication, and aspiration durations.*
***The values for Speaker 1 are averages of two tokens of each syllable; for Speaker 2, the values are for one token of each syllable.*

ably never identical acoustically. If we add chance variations in relative effectiveness of bursts and transitions, due to such factors as distance between speaker and listener (or between speaker and microphone), we must conclude that predictions of the perceptual weight attached to the several acoustic cues to place of articulation can be, at best, statistical, and that the likelihood of any single cue being the sole determinant of the percept in all contexts is extremely low.

As will be seen, the results of the following three experiments support this conclusion. Experiment 1 assesses the role of bursts and formant transitions in the recognition of natural speech by systematically removing them from American English /b, d, g/ spoken before nine different vowels by a single speaker. Experiment 2 replicates Experiment 1 with a different speaker. These two experiments are thus concerned with whether the manipulated cues are *sufficient* for recognition. Experiment 3, on the other hand, is concerned with whether the release burst is functionally invariant; it assesses the invariant cue value of the release burst for the second speaker by transposing it from each consonant-vowel-consonant (CVC) syllable across all vowels for each class of stop consonant.

## EXPERIMENTS 1 AND 2

### Method

**Experiment 1**

Nine CVC syllables were recorded by a male speaker in a carrier phrase, "The little CVC dog," with stress on the CVC. Two tokens of all combinations of initial /b,d,g/, followed by /i,ɪ,ɛ,æ,ʌ,a,ɔ,u,ɝ/, with a constant syllable-final /d/, were recorded. In addition, phrases of the type "The little VC dog" were recorded, where V was again one of the nine vowels above and C was again /d/. The phrases were digitized with an effective frequency response of 160-7,000 Hz, by means of the Haskins Laboratories pulse code modulation system (Cooper & Mattingly, 1969), and the test syllables were excised and edited. Two parallel sets of 45 experimental signals were then constructed by the following steps:

(1) Each syllable was left in its original form.

(2) From each CVC, the burst was removed. A burst was defined as an utterance initial, high-amplitude (relative to the surrounding signal) component of the signal (see Figure 1). The duration of the burst was determined for each syllable on a high-resolution oscillogram; the values were quite consistent across the two tokens of each syllable. Table 1 lists these durations averaged across tokens. The mean burst duration for /b/ was 4.3 msec, for /d/ 6.3 msec, and for /g/ 11.7 msec.

(3) Each burst was attached to its corresponding VC syllable (for example, the /bid/ burst was attached to /id/), leaving a silent interval between the end of the burst and the first voiced pulse of the vowel, equal in duration to the interval between burst offset and voicing onset in the CVC from which the burst had been removed.

(4) For each CVC, the entire signal up to the first well-defined voicing pulse was removed. Thus, the burst and devoiced formants (i.e., noise excited resonances) were removed (see Figure 1), and the duration of this segment was measured on an oscillogram of each utterance. Table 1 lists the two-token averages of the devoiced formants ("aspiration"), as well as of the entire segment from burst onset to voice onset (VOT) for each syllable. Mean VOT for /b/ was 7.7 msec, for /d/ 12.0 msec, for /g/ 20.1 msec.

(5) Each burst-plus-devoiced-formants was then attached to its corresponding VC syllable.

This procedure permitted us to present five different combinations of the cues to place of articulation (burst, devoiced transition, voiced transition) for each syllable: (a) all three together in the original syllable; (b) burst plus vowel; (c) burst and devoiced transitions plus vowel; (d) voiced transitions plus vowel; (e) devoiced and voiced transitions plus vowel.

Three recordings of each of the 45 signals in each set were generated and randomized into two parallel test sequences of 135 items each. One test was administered to 14 Lehman College undergraduates. The stimuli were played at a comfortable level in a sound-attenuated room, on a Revox 1122 tape recorder, over an audiometric loudspeaker. The other test was administered to nine students and faculty volunteers from Yale University: the stimuli were played at a comfortable level in a sound-attenuated room on an Ampex AG 400 tape recorder over an AR4X loudspeaker at Haskins Laboratories.

The listeners were instructed to write the identity of the initial sound of each syllable. The response categories listed on the answer sheets were /b,d,g,p,t,k,?,Ø/. The ? response was for use when the listener thought that the syllable began with a consonant, but could not decide which one. The Ø response was for use when the listener thought that the syllable began with a vowel.[2] Twenty tokens of the stimuli were played to familiarize the listeners with the task. The listeners were then presented with one of the 135-item test sequences.

**Experiment 2**

Exactly the same procedures of stimulus and test construction as those described above were followed for a second speaker, except that he provided only one token of each syllable and therefore only one test. The sentences were read at a very deliberate rate with stress on the initial consonant of the CVC. Table 1 lists the durations in milliseconds of burst, "aspiration," and VOT for each syllable. The durations are very much longer than (almost double) those of Speaker 1. However, the pattern of increase in burst and VOT durations, from labial to apical to velar stops, is similar to that of Speaker 1.

Eleven Lehman College undergraduates took the test under conditions identical to those of the Lehman College students in Experiment 1.

## Results

### Experiment 1 (Speaker 1)

The two groups of subjects gave very similar results on the two parallel tests. We have therefore combined their data. Figure 3 displays percentage correct identification of initial consonantal place of articulation as a function of vowel nucleus for the five sets of cue combinations (all cues, burst plus vowel, burst and voiceless transition plus vowel, voiced transition plus vowel, voiced and voiceless transition plus vowel) and the three classes of consonant (labial, apical, velar). Responses were scored for place of articulation only, and voicing errors were disregarded. Each data point is based on 69 responses (23 subjects × 3 repetitions). The vowels have been ordered along the horizontal axis to trace a rough path around the rim of the English vowel loop from /i/ through /a/ to /u/, with /ɝ/ appended. The points have been connected by straight lines to facilitate reading of the graphs.

**Labial.** All the original syllables, except /bud/ (85%), were correctly identified more than 90% of the time. The burst was relatively ineffective as a cue and performance hovered around chance (20%) before all vowels, except /u/ (81%) and /ɝ/ (51%).[3] The voiced transition, on the other hand, served almost as well as the full syllable and performance hovered around 90% before all vowels, except /ɔ/ (84%), /u/ (63%), and /ɝ/ (61%), the last two vowels being precisely those for which burst performance was at its best. The addition of the devoiced transition, whether to burst or voiced transition, tended to increase performance by a few percentage points, but this cue clearly carried little perceptual weight.

**Apical.** All the original syllables, except /did/ (87%), /dɛ/ (74%), and /dæd/ (81%) were correctly identified more than 90% of the time. The burst was a moderately effective cue before the front vowels, /i/ (57%) and /ɪ/ (65%), but otherwise carried little weight, and was only marginally aided by addition of the devoiced transition. The full transition (devoiced and voiced portions), on the other hand, was a moderately effective cue (60% or higher) before the back and central vowels, but a weak cue before the front vowels. There seems to be a reciprocal relation between burst and transitions; if the weight of one is high, the weight of the other is low. Wherever the full transition carried any marked weight, removal of its devoiced portion led to an appreciable drop in performance particularly before /u/ and /ɝ/. In general, neither burst nor transition alone maintained performance at the level of the original syllables.

**Velar.** All the original syllables except /gid/ (56%), /gɪd/ (70%), /gɛd/ (61%), and /gæd/ (88%)

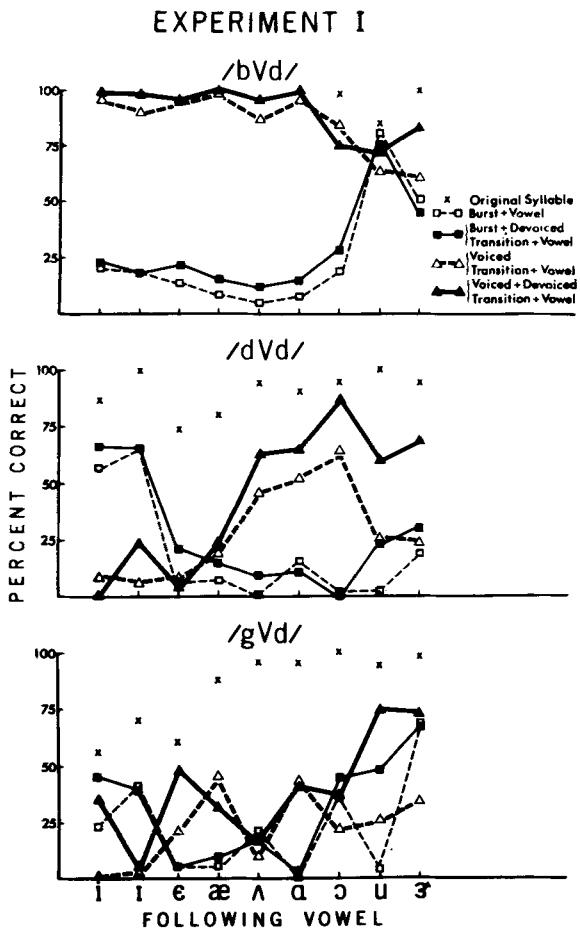


**Figure 3. Percent correct recognition of place of articulation for Speaker 1 as a function of following vowel. The five different combinations of cues to place of articulation are parameters of the curves.**

were correctly identified more than 90% of the time. The burst elicited moderate performances only before /ɪ/ (41%) and /ɝ/ (69%), and was appreciably aided by addition of the devoiced transition only before /u/. For the full transitions, performance was moderate before /ɛ/ (48%), /a/ (41%), /u/ (75%), and /ɝ/ (73%), but weak elsewhere. Just as for /d/, removal of the devoiced portion of the transition had a marked effect before /u/ and /ɝ/. There is again some evidence of a reciprocal relation between burst and transition. Even more obviously than for /d/, no subset of the cues held performance at the level of the original syllables.

**Experiment 2 (Speaker 2)**

Figure 4 displays the results for tokens from the second speaker in the same format as Figure 3. For /b/ and /d/, the pattern of results is similar to that of Experiment 1, apart from a general increase in level of performance; for /g/ the perceptual weight of the burst is clearly greater than it was for Speaker 1. It will be recalled that the duration of the bursts and aspiration segments of Speaker 2's utterances was very much greater than (nearly double) that of the corresponding segments of Speaker 1 (see Table 1).

**Labial.** All the original syllables, except /bud/ (85%), were correctly identified more than 90% of the time. The burst was moderately effective as a cue before all vowels, especially the central to back vowels, /ɔ/ (79%), /u/ (79%), and /ɝ/ (85%), and was as effective as the full syllable for /ɛ/ (97%). The full transition served almost as well as the full syllable for all vowels except /a/ (75%), /u/ (36%), and /ɝ/ (42%), the last two again being the vowels for which burst performance was at its best. The perceptual effect of adding the devoiced transition, whether to burst or voiced transition, was generally small, and not reliable.

**Apical.** All the original syllables were correctly identified more than 90% of the time. The burst was a strong cue before /ɪ/ (100%) and /ɝ/ (91%), moderate before /i/ (72%) and /ɛ/ (79%), but otherwise carried little weight. Addition of the devoiced transition to the burst had no systematic effect. The full transition was almost as effective as the full syllable for central and back vowels, but was a weak cue before the front vowels. Removal of the devoiced portion of the transition tended to lower performance, especially before /u/. Performances on bursts and transitions were reciprocally related before all vowels except /æ/ and /ɝ/.

**Velar.** All the original syllables, except /gid/ (64%), were correctly identified more than 90% of the time. The burst was a moderately effective cue before /ɪ/ (73%), /ɛ/, and /ʌ/ (55%), almost as effective as the full syllable before /i/, /a/, /ɔ/, /u/, and /ɝ/. Addition of the devoiced transition had no systematic effect. The full transition was a moderately effective cue before /ɪ/ (64%) and /ɛ/ (46%), a strong cue before /a/ (91%), but otherwise carried little or no perceptual weight. Removal of its devoiced portion tended to reduce performance, particularly before /ɪ/ and /a/. Burst and transition again tend to be reciprocally related, particularly before central and back vowels, except /a/.

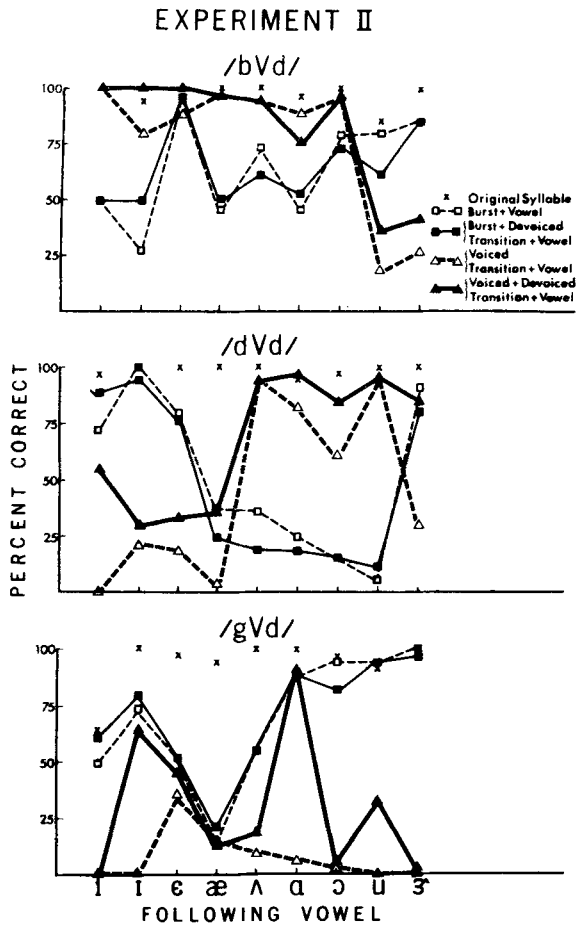


**Figure 4. Percent correct recognition of place of articulation for Speaker 2 as a function of following vowel. The five different combinations of cues to place of articulation are parameters of the curves.**

## Discussion

**Experiments 1 and 2**

The perceptual weight carried by release bursts and formant transitions as cues to place of articulation, varied with consonant, vowel, and speaker. No single cue, or pair of cues, was sufficient for recognition in all contexts. If we take into account variation in acoustic structure, such as those outlined in the intro-

duction, we can make sense of many, though not all, of the results.

**Labial.** As expected, labial bursts were relatively weak cues. For Speaker 2, they were longer in duration and considerably more effective than for Speaker 1. Nonetheless, the patterns of performance are quite similar for the two speakers; apart from an anomalous point at /ɛ/ for Speaker 2, labial bursts tended to be most effective before rounded vowels. Whether this is due to variations in burst energy or to variations in burst frequency position in relation to the following vowel will become clearer when we have reported the results of Experiment 3. Here we note simply that the rank order correlation between burst duration and percent recognition was not significant for either speaker.

Formant transitions, on the other hand, were almost as effective for both speakers as the full complement of cues, before all nine vowels, except /u/ and /ɝ/. The two exceptions are before rounded vowels for which lip constriction necessarily reduces the rise in formant frequency (i.e., the extent of formant transitions) associated with mouth opening.

**Apical.** As expected, apical bursts tended to be longest and most effective for both speakers before front vowels. They were weak before all other vowels (/ɝ/ is an exception for Speaker 2), and seem to have become systematically weaker as rounding (and so front-cavity length) increased, reducing burst energy (see Table 1). However, burst frequency may also be relevant, and we again defer discussion, noting only the lack of significant correlation between burst duration and performance.

For both speakers (particularly Speaker 2), formant transitions were strong cues before central and back vowels /ʌ,a,ɔ,u,ɝ/, where apical transitions are extensive, but weak cues before the front vowels, where transitions are relatively short. Furthermore, as might be predicted from the longer apical than labial VOTs (see Table 1), addition of the devoiced to the voiced transition segments tended to improve recognition of /d/ more than of /b/. However, within the apical series, VOT does not significantly predict the performance gains from addition of the voiceless transitions (the rank order correlation between VOT and performance was not significant).

**Velar.** Speaker differences are most marked for the velar series. The predicted tendency for the burst to be more effective before back, rounded than before front, unrounded vowels was borne out for Speaker 2, despite his somewhat longer front than back vowel bursts. However, for Speaker 1, the burst was simply a very weak cue before all vowels, except /ɝ/. Again, we note the lack of significant correlation between burst duration and performance, and defer comment on these results.

As expected, the relatively short velar transitions were far less effective cues than were labial and apical transitions for both speakers. At the same time, longer VOTs did tend to increase the effectiveness of devoiced transitions. For Speaker 1, the largest performance gains from the addition of devoiced to voiced transitions were for /i/, /ɛ/, /u/, and /ɝ/, the vowels before which aspiration durations were longest; similarly, for Speaker 2 the largest gains were for /ɪ/ and /a/ (see Table 1). However, the rank order correlation between performance and voice onset time was not significant.

Broadly, our results agree with those of Fischer-Jørgensen (Note 2) for Danish initial-voiced stops in these respects: (1) the burst was a relatively effective cue for /b/ before /u/, but not before /a/; (2) the burst was a relatively effective cue for /d/ before /i/, but not before /a/ or /u/; and (3) the burst was a relatively effective cue for /g/ before /i/ and /u/ (Speaker 2 only), but not before /a/ (Speaker 1 only). Our results disagree with those of Fischer-Jørgensen insofar as: (1) the burst was a relatively ineffective cue for /b/ before /i/; (2) the burst was a relatively ineffective cue for /g/ before /u/ (Speaker 1 only); and (3) the burst was a relatively effective cue for /g/ before /a/ (Speaker 2 only).

Our results do not support the implication of Cole and Scott (1974a) that release bursts alone are sufficient cues to the place of articulation of initial-voiced stop consonants. Nor, contrary to our own expectation, did the addition of devoiced transitions to the bursts reliably improve recognition. If we adopt as an arbitrary (and modest) criterion of significant perceptual weight that recognition performance for release-bursts-plus-vowels should drop by no more than 25% below performance for the original syllable, we see that this level was reached for Speaker 1 on only one syllable out of 27 (/bud/), for Speaker 2 on only 13 syllables out of 27 (/bɛd, bɔd, bud, bɝd, did, dɪd, dɝd, gid, gad, gɔd, gud, gɝd/). The role of consonant-vowel (CV) coarticulation in determining burst effectiveness, implicitly denied by Cole and Scott (1974a), is suggested by the preponderance among Speaker 2's 13 syllables, of central-back, rounded vowel syllables for /b/ and /g/, of front unrounded vowel syllables for /d/.

## EXPERIMENT 3

The purpose of this experiment was to test the hypothesis that the initial release burst of /bVd, dVd, gVd/ syllables may be a functionally invariant cue to consonantal place of articulation across a representative set of syllable-nucleus types. The method was to transpose the release burst from each CVC syllable in a series (labial, apical, velar) across all types of

VC syllables in that series. For a fair test of the hypothesis, we needed tokens from a speaker whose release bursts were known to be at least moderately effective cues in their original syllables. We therefore used the 27 CVC (and 9 VC syllables) recorded by Speaker 2 for Experiment 2.

## Method

The experimental signals were constructed in exactly the same way as the burst-plus-vowel signals of Experiments 1 and 2. The burst was removed from all 27 CVC syllables (for durations see Table 1). Each burst was then attached to all nine vowel-/d/ syllables (where the vowels were again /i,ɪ,ɛ,æ,ʌ,a,ɔ,u,ɝ/), leaving a silent interval between burst offset and vowel onset, equal in duration to the devoiced interval for the CVC token being simulated. The result was a set of 81 syllables in each series (labial, apical, velar)—a total of 243.

Three repetitions of each syllable were recorded and randomized into a single test of 729 items. The test was administered to eight Lehman College undergraduates under conditions and instructions identical with those used for the Lehman College students of Experiments 1 and 2.

## Results

Figure 5 displays percentage correct identification of initial consonantal place of articulation as a function of following vowel for the nine bursts in each series. Responses were scored for place of articulation only, and voicing errors were disregarded. To facilitate reading, the results for bursts drawn from syllables containing the four front vowels (/i,ɪ,ɛ,æ/) have been grouped in the upper three graphs; the results for bursts drawn from syllables containing the five central and back vowels (/ʌ,a,ɔ,u,ɝ/) have been grouped in the middle three graphs. The following vowels have been ordered along the horizontal axes to trace a path around the rim of the English vowel loop from /i/ through /a/ to /u/, with /ɝ/ appended, and points have been connected by straight lines to facilitate reading. For untransposed bursts (i.e., bursts placed before the same vowel as that of the syllable from which they were originally drawn) the data point is circled.

Before considering the three series separately, several general points can be made. First, the highest performance for a given vowel is often not elicited by the burst taken from the original syllable containing that vowel. For example, the burst drawn from the syllable /bad/ elicited a lower performance when attached to /ad/ (the circled point over /a/ in the middle labial graph of Figure 5) than did the bursts drawn from any of the other eight /bVd/ syllables. Similar, if less severe, discrepancies appear for many other syllables.

Second, the highest recognition performance

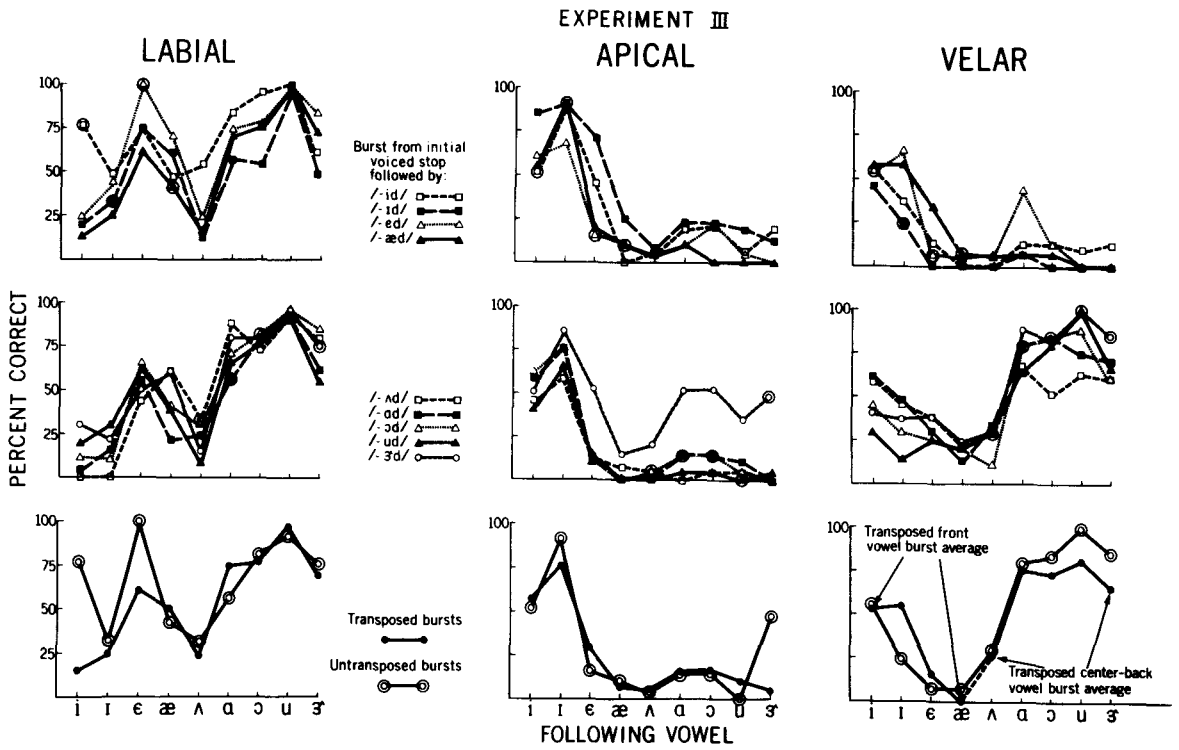


Figure 5. Percent correct recognition of place articulation in burst plus /Vd/ syllables. In the top two rows of figures, each point represents the recognition of syllables composed of a burst taken from one vocalic environment and transposed to each of the other vocalic environments (with the exception of the circled points for the nine untransposed bursts). In the bottom row, average correct recognition scores for syllables in which the bursts were transposed are compared with recognition scores for syllables in which the bursts were attached to the same vowel as that of the syllable from which they were originally taken.

elicited by a particular burst is not always for a syllable containing the same vowel as the syllable from which the burst was drawn. This is most striking in the apical series for which the highest performances elicited by all nine bursts are before /id/ and /ɪd/. Similarly, in the labial series, bursts drawn from all nine syllables, including the front vowel set, elicit their highest performances when attached to back vowel syllables; and in the velar series, bursts from the four central and back vowel syllables, /gad, gɔd, gud, gɝd/, elicit roughly interchangeable performances within their own set.

Both these results suggest a measure of commutability among the bursts of each series. This commutability becomes even more obvious as soon as we notice a third feature of the data, closely related to the first two: the overall form of the performance curves across the vowels is remarkably similar for all bursts within a series, whatever the syllables from which they were drawn. The degree of concordance among the nine curves of each series is a measure of burst commutability or functional invariance. Furthermore, a rather good description of the general curve for each series is provided by simply plotting for each vowel the percentage of correct identification of its "untransposed" burst (circled points). These curves are displayed in the lower three graphs of Figure 5, together with a plot of the mean percentage correct for the transposed bursts. The rank order correlation between these curves, that is, between performances elicited by transposed and untransposed bursts, is then a second measure of burst commutability or functional invariance.

**Labial**

All nine labial curves are roughly parallel: bursts from almost every syllable elicit their highest performance before the central-back rounded vowels, /a,ɔ,u,ɝ/, a moderate performance before /ɛ/ (before /æ/ for the /bud/ and /bɝd/ bursts), and relatively weak performances before /i,ɪ,æ,ʌ/ (except the peak for the /bid/ burst before /i/). Kendall's coefficient of concordance (W) among the nine curves is .79 ($p < .0001$). This significant similarity in pattern of burst effectiveness (or sufficiency) demonstrates that the nine bursts are, to a large degree, functionally invariant. However, Spearman's rho between untransposed and mean transposed curves of the bottom labial graph falls short of significance with a value of .53. This failure is clearly due to the peaks for the untransposed /bid/ and /bɛd/ bursts and suggests that release bursts effective in signaling labiality before /i,ɛ/ may be context-dependent.

**Apical**

All nine apical curves are roughly parallel; bursts from every syllable elicit their highest performances before /ɪ/ and /i/ and, apart from fair performances for the /dɪd/ and /dʒd/ bursts before /ɛ/ and for the /dʒd/ burst before the back vowels and /ɝ/, relatively weak performances elsewhere. Kendall's W among the nine curves is .72 ($p < .0001$). Spearman's rho between the untransposed and the mean transposed curves of the bottom graph (Figure 5) is .60 ($p = .05$), clearly pulled down by the peak for the untransposed /dʒd/ burst. The apical bursts like the labial bursts are to a large degree functionally invariant.

**Velar**

The curves for the velar bursts fall into two distinct groups—front and central-back vowels. The front vowel bursts elicit moderate performances before /i,ɪ,ɛ/ and, apart from a small peak for the /bɛd/ burst before /a/, weak performances elsewhere. The central-back vowel bursts elicit their highest performances before /a,ɔ,u,ɝ/, weak performances elsewhere, though with a tendency for slightly stronger performances before /i,ɪ/. There is thus a small asymmetry: While front vowel bursts do not concord with back vowel bursts before back vowels, back vowel bursts tend to concord with front vowel bursts before front vowels. As a result, Kendall's W among the nine curves, though significant ($p < .001$), is low (.37). However, if we separate the two groups and compute Kendall's W within them, we find for the front vowels, .69 ($p < .05$), and for the central-back vowels, .66 ($p = .01$). The increased coefficients justify separating the bursts into two groups. Accordingly, the average transposed burst curve of the bottom graph (Figure 5) was computed for front vowels and for central-back vowels separately. The result is an excellent fit between transposed and untransposed curves, for which Spearman's rho is .88 ($p < .01$). There is therefore a large degree of functional invariance among the velar front vowel bursts and among the velar central-back vowel bursts.

## Discussion

While the release bursts of initial, labial, apical, and velar stops display a high degree of functional invariance, they do not display a corollary degree of sufficiency. In all three experiments, the release burst was seldom sufficient to maintain performance at the level elicited by the original syllable. Vowel-dependent variations in performance are therefore less aptly characterized as variations in "sufficiency" or "cue adequacy," than as variations in the degree to which the burst may be assumed to contribute to the cue complex in natural speech (cf. Stevens, 1975).

In Experiment 3, identification of the original syllables was almost perfect except for /gid/, which

the listeners identified with 87% accuracy. If we again adopt as an arbitrary criterion of significant perceptual weight that performance on the untransposed burst-plus-vowel should drop by no more than 25% below performance on the original syllables, we arrive at the following set of 14 out of 27 syllables for which the release burst carried weight in judgments of place of articulation in either or both of Experiments 2 and 3: /bid, bɛd, bɔd, bud, bɜ̈d, did, dɪd, dɛd, dɜ̈d, gid, gad, gɔd, gud, gɜ̈d/.

These results bring us into closer agreement with both Cole and Scott (1974a), and Fischer-Jørgensen (Note 2), since the untransposed burst carried significant weight for /b/ before /i/ in Experiment 3. The results also agree very well with those of Liberman et al. (1952). These authors used the relatively crude Pattern Playback II synthesizer to construct schematic stop bursts before seven two-formant monotone vowels, /i,e,ɛ,a,ɔ,o,u/. Identifications reached 75% or higher for /p/ before /i,e,ɛ,ɔ,o,u/, for /t/ before /i,e,ɛ/, and for /k/ before /a,ɔ,o,u/. Considering only the vowels common to both experiments, these results agree with our own in finding bursts to carry weight as labial cues before /i,ɛ,ɔ,u/, as apical cues before /i,ɛ/, and as velar cues before /a,ɔ,u/. The only discrepancy between the two sets of results is in our finding that a release burst carried weight as a velar cue before /i/. This remarkable agreement between the present natural speech study and an experiment carried out with primitive synthetic speech 25 years ago, suggests that the systematic variations in burst effectiveness common to both experiments reflect a robust perceptual process.

The most obvious source of these variations might seem to lie in release burst energy. Unfortunately, we were not able to make reliable intensity measurements of the release bursts in the present study. However, a scan of the syllables for which release bursts proved adequate and of their durations in Table 1 will reveal no obvious correlation, and, as reported above, Spearman's rho between burst duration and performance was not significant for any series. Furthermore, since all schematic bursts synthesized by Liberman et al. (1952) were of equal energy, this factor cannot account for their results. Thus, while variations in burst energy may well account for variations in the overall performances elicited by particular bursts or in the recognition of different tokens of a particular stop-vowel syllable (and so for the different levels of performance elicited by the bursts of Speakers 1 and 2), they cannot account for systematic variations in burst effectiveness across vowels.

The case is no better when we turn to the absolute spectral properties of release bursts. For example, as remarked in the introduction, spectral sections taken through the apical release burst show a broad high-intensity curve, with its greatest weight over frequencies above about 2,000 Hz, largely independent of the following vowel (see Figure 2). We can hardly, therefore, appeal to the absolute spectral properties of the apical burst to explain the fact that the burst carries appreciable weight before high front vowels such as /i,ɪ,/, but essentially no weight before central-back vowels such as /ʌ,a,ɔ,u/.

In fact, the key to the problem may be provided by the work of Kuhn (1975). First, he draws on the acoustic theory of speech production, according to which the resonance of the cavity in front of the point of maximum tongue constriction—that is, "the front cavity resonance"—may be associated with any of the first four formants (Fant, 1960, p. 72). He then shows that "the front cavity seems to be associated with what is perhaps the most intense group of formants: with the $F_3$ group for /i,ɪ,ɛ,æ/, and with the $F_2$ group for /a,ʌ,ʊ,u/" (Kuhn, 1975; p. 430). Next, he demonstrates that a front cavity frequency estimate can be most readily made for the more constricted vowels and for highly constricted consonants, and that for stop consonants the estimate may be derived from the spectral structure of bursts and transitions. Since the front cavity resonance is a function of front cavity length, and since front cavity length is a function of the place of articulation, an estimate of the resonance is tantamount to an estimate of place of articulation. To these facts we may add a variety of evidence from synthetic speech experiments (e.g., Liberman et al., 1952) to suggest that place of articulation is most readily conveyed to stop consonant bursts when their spectral weight lies close to the front cavity resonance of the following vowel. Proximity on the frequency scale may facilitate perceptual integration of the burst with the vowel, so that the listener can track the changing front cavity shape characteristic of a particular place of articulation followed by a particular vowel. As will be seen in the following discussion, this hypothesis can account for many of the variations in burst effectiveness observed in Experiments 2 and 3.

**Labial**

The low-frequency labial bursts carried significant weight (by the criterion defined above) before /ɔ,u,ɜ̈/ in both experiments, and close to significant weight before /ʌ/in Experiment 2 and before /a/ in Experiment 3. For all these vowels the front cavity is strongly associated with the second formant, and the frequency of that formant lies in the region over which the greatest weight of labial burst energy is distributed. The variability in response for /ʌ,a/ may be due to weaker front cavity-to-formant affiliation in less constricted vowels, and the consequent

difficulty for the listener in continuous tracking of the changing front cavity resonance in the absence of a formant transition.

The two other vowels before which labial bursts carried significant weight were /i,ɛ/, for which the front cavity is strongly associated with the third formant. However, the untransposed bursts were notably more effective than the transposed and, as remarked above, this suggests a degree of context dependency. The rapid and relatively extensive lip opening before unrounded vowels and the consequent rapid rise in resonant frequencies may extend the burst frequency range sufficiently high for it to be integrated with $F_3$ of the following vowel. The ineffectiveness of the burst before /æ/ may again be due to weaker front cavity-to-formant affiliation in a less constricted vowel, and the resulting difficulty for the listener. However, the ineffectiveness of the burst before /ɪ/ is unexplained.

**Apical**

Apical bursts carried significant weight before /ɪ/ in both experiments and before /i,ɛ,ɝ/ in Experiment 2, although performance was very weak for most bursts before /ɛ,ɝ/ in Experiment 3. On the assumption that the high-frequency apical burst can be integrated perceptually with the front cavity resonance of $F_3$ for the high front vowels /i,ɪ/, but less readily, if at all, with the less determinate front cavity resonance of the more open vowels /ɛ,æ/, or with the low-frequency front cavity resonance of $F_2$ for the central-back vowels /ʌ,a,ɔ,u/, these results are very much what we would expect. Nonetheless, there are oddities. For example, it is not clear why the /dɪd/ burst (duration 15 msec) should have been more effective before /i/ than was the untransposed burst from /did/ (duration 25 msec) in Experiment 3. Nor is it clear, given the moderate duration of the /dɝd/ burst (10 msec), why it should have been a strong cue (91%) before the low front cavity resonance of $F_2$ for /ɝ/ in Experiment 2 and a moderately strong cue before /a,ɔ,ɝ/ in Experiment 3.

**Velar**

Velar bursts carried significant weight before /i,a,ɔ,u,ɝ/ in both experiments. It will be recalled that the spectral weight of velar bursts tend to lie close to the $F_2$ frequency of the following vowel. Perceptual integration of the burst with the front cavity resonance of $F_3$ for the front vowels should therefore be easiest when $F_2$ and $F_3$ lie close together, as in /i/, precisely as observed. For the central-back vowels /a,ɔ,u,ɝ/, variation in $F_2$ frequency, and so of velar burst frequency, is small. We might therefore expect that velar bursts from all four vowels would be readily commutable and accessible to perceptual integration with the front cavity resonance of $F_2$. Again, this is precisely what was observed. The systematic decline in performance as $F_2$ (and so velar burst frequency) decreased from /ɪ/ to /æ/ (see Figures 4 and 5) suggests that the ineffectiveness of velar bursts before /ɪ,ɛ,æ/ may be due to the increasing separation of burst and front cavity resonance ($F_3$) on the frequency scale. The inadequacy of the burst before /ʌ/ may arise from the relatively weak front cavity-to-formant affiliation for this vowel.

In short, despite several unexplained oddities in the data, the perceptual integration hypothesis provides a remarkably close account of the variations in burst effectiveness in Experiments 2 and 3. At the same time, this account affords insight into the grounds of functional invariance among stop release bursts. Bursts are invariant insofar as they bear the same relation to any particular following vowel. The relation is that of spectral continuity or discontinuity with the main (or front cavity) resonance of the following vowel. If there is continuity (as in an apical burst followed by /ɪ/, for example), the relation contributes significantly to recognition of consonantal place of articulation; if there is discontinuity (as in an apical burst followed by /ʌ/, for example), the relation does not contribute significantly to recognition. The invariance is therefore not a simple first-order invariance based on the absolute frequency and/or amplitude of the bursts. Rather, it is a higher order relational invariance based on spectral relations between burst and following vowel.

The general conclusion that the contribution of the burst to the cues for place of articulation depends on the following vowel is not new. Liberman et al. (1952) remarked of their schematic /p/ and /k/ bursts before schematic vowels that: "the irreducible acoustic stimulus is the sound pattern corresponding to the consonant-vowel syllable" (p. 516). While neither Fant (1959, Note 1) nor Stevens (1975) believes that the perceptual process always requires reference to the vowel, both describe the burst in natural speech as dependent on context for its effect. Stevens (1975) deliberately eschews a description in terms of release bursts and individual formants, since this would imply that these components have independent roles in the cue complex. He emphasizes, rather, "the overall acoustic spectrum immediately following the release" (p. 311). However, regarding the contribution of the burst to this spectrum he writes: "We shall assume that this can be considered as the initiation of the rapid spectrum change at the consonant release, if there is spectral energy in the burst in the vicinity of the major spectral peak for the vowel . . . . Thus the initial burst of energy in syllables beginning with /g/, and the burst for syllables with front vowel preceded by /d/ would be

considered as part of the rapid spectrum change, since major energy concentrations in these bursts occur in frequency regions where the vowel formant transitions are providing cues for place of articulation of the consonant. The d-burst in a syllable with a back vowel, on the other hand, would not be considered as an integral part of the rapid spectrum change . . . . The burst at the onset of the consonant /b/ is relatively weak, and may not play a significant role in shaping the rapid spectrum change." (Stevens, 1975, pp. 312-313).

The present study suggests that, at least for some speakers and listeners, the contribution of the /g/ burst may not be as strong for open vowels as for closed vowels, and that the contribution of the /b/ burst may not always be insignificant. It is precisely to an understanding of such detailed variations that Kuhn (1975) has added by identifying both the burst and "the major spectral peak for the vowel" with the front cavity resonance. In short, Stevens' general description of the conditions under which the burst contributes to the spectral changes following release is consistent both with Kuhn's (1975) front cavity analysis and with our own results.

## GENERAL DISCUSSION

An important feature of the results of Experiments 1 and 2 was the tendency toward reciprocal performances on bursts and transition; where the perceptual weight of one increased, the weight of the other declined. These reciprocal relations follow systematically from the acoustic structure of the syllable. Where transitions are brief (for /b/ before rounded vowels, for /d/ before high front vowels, for /g/ before close vowels), the burst lies near the major spectral peak of the following vowel and contributes significantly to the perceptual outcome; where transitions are extensive (for /b/ before middle, unrounded vowels, for /d/ before central-back vowels), the burst is distinct from the major spectral peak of the following vowel, and contributes little. If we combine this observation with the conclusion of Experiment 3, we are led to recognize that bursts and transitions are acoustically and functionally (that is, perceptually) equivalent: both provide a spectrally continuous change from the consonantal release into the following vowel by which the listener can estimate place of articulation. To say that they are equivalent is not, of course, to say that they are alternative. In natural speech, as we have already emphasized, it must be rare that a listener relies on burst alone or on transition alone, and in Experiments 1 and 2, a single cue was not often sufficient to hold recognition at the level of the original syllable. Bursts and transitions are equivalent and complementary.

Once again, this observation is not new. Over 20 years ago, Cooper, Delattre, Liberman, Borst, & Gerstman (1952) remarked that "bursts and transitions complement each other in the sense that when one cue is weak, the other is usually strong" (p. 603). In a similar vein, Fischer-Jørgensen (1954) commented on synthetic speech studies: "The listener does not compare explosion with explosion and transition with transition, but compares artificial syllables comprising either explosion or transition with natural syllables that always contain both" (p. 56). Finally, Fant (1959; 1960; p. 217) has repeated emphasized that the qualitatively distinct acoustic segments during the first 10-30 msec after release are probably not auditorily discriminable and "should be regarded as a single stimulus rather than as a set of independent cues" (Fant, Note 1, p. 21). And, as we saw above, the acoustic and functional inseparability of burst and transition is explicit in "the rapid spectrum changes" following release that Stevens (1975, p. 311) describes. In short, bursts and transitions are functionally identical.

In conclusion, the results of the present study, and, in particular, the functional equivalence of release bursts and transitions, suggest that the perceptual process may entail continuous tracking of vocal tract resonances. The importance of transitional information for the recognition not only of stop consonants in many contexts, but also of /w,r,l,y/, nasal consonants, fricatives, and perhaps even vowels (Lindblom & Studdert-Kennedy, 1967; Shankweiler, Strange, & Verbrugge, 1977) is attested by an extensive literature (for reviews, see Darwin, 1976; Liberman et al., 1967; Stevens & House, 1972; Studdert-Kennedy, 1974, 1976). We do not doubt that the acoustic invariants for these phonetic segments may eventually be specified, but we see little ground for expecting them to be specified without reference to context.

### REFERENCE NOTES

1. Fant, G. *Stops in CV-syllables*. Speech Transmission Laboratory Quarterly Progress and Status Report, 1969, No. 4, 1-25.
2. Fischer-Jørgensen, E. *Tape cutting experiments with Danish stop consonants in initial position*. Annual Report VII (Institute of Phonetics, University of Copenhagen, Copenhagen, Denmark), 1972.

### REFERENCES

BELL-BERTI, F. Control of pharyngeal cavity size for English voiced and voiceless stops. *Journal of Acoustical Society of America*, 1975, **57**, 456-461.

COLE, R. A., & SCOTT, B. The phantom in the phoneme: Invariant cues for stop consonants. *Perception & Psychophysics*, 1974, **15**, 101-107. (a)

COLE, R. A., & SCOTT, B. Toward a theory of speech perception. *Psychological Review*, 1974, **81**, 348-374. (b)

COOPER, F. S., DELATTRE, P. C., LIBERMAN, A. M., BORST, J. M., & GERSTMAN, L. J. Some experiments on the perception of syn-

thetic speech sounds. *Journal of Acoustical Society of America*, 1952, **24**, 597-606.

COOPER, F. S., & MATTINGLY, I. G. A computer controlled PCM system for the investigation of dichotic perception. *Journal of Acoustical Society of America*, 1969, **46**, 115(A).

DARWIN, C. J. The perception of speech. In E. C. Carterette & M. P. Friedman (Eds.), *Handbook of perception* (Vol. 7). New York: Academic Press, 1976.

DAY, R. S. Temporal-order perception of a reversible phoneme cluster. *Journal of Acoustical Society of America*, 1970, **48**, 95(A).

DELATTRE, P. C., LIBERMAN, A. M., & COOPER, F. S. Acoustic loci and transitional cues for consonants. *Journal of Acoustical Society of America*, 1955, **27**, 769-773.

FANT, G. Acoustic description and classification of phonetic units. *Ericsson Technics*, 1959, **1**, 1-52.

FANT, G. *Acoustic theory of speech production*. 's-Gravenhage: Mouton, 1960.

FISCHER-JØRGENSEN, E. Acoustic analysis of stop consonants. *Miscellanea Phonetica*, 1954, **2**, 42-49.

FLANAGAN, J. L. *Speech analysis synthesis and perception* (2nd ed.). New York: Springer Verlag, 1972.

FUJIMURA, O. Bilabial stop and nasal consonants: A motion picture study and its acoustical implications. *Journal of Speech Hearing Research*, 1961, **4**, 233-247.

HALLE, M., HUGHES, G. W., & RADLEY, J.-P. A. Acoustic properties of stop consonants. *Journal of Acoustical Society of America*, 1957, **29**, 107-116.

HARRIS, K. S. Cues for the discrimination of American English fricatives in spoken syllables. *Language and Speech*, 1958, **7**, 1-7.

HEINZ, J. M. Speech acoustics. In T. A. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12, No. 4). Atlantic Highlands, N.J: Humanities Press, 1974. Pp. 2241-2280.

HOFFMAN, H. S. Study of some cues in the perception of the voiced stop consonants. *Journal of Acoustical Society of America*, 1958, **30**, 1035-1041.

KLATT, D. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech Hearing Research*, 1975, **18**, 686-706.

KUEHN, D. P. *A cinefluorographic investigation of articulatory velocities*. Unpublished PhD thesis, Iowa University, 1973.

KUHN, G. M. On the front cavity resonance and its possible role in speech perception. *Journal of Acoustical Society of America*, 1975, **58**, 428-433.

LADEFOGED, P., DECLERK, J., LINDAU, M., & PAPCUN, G. An auditory-motor theory of speech production. *Working Papers in Phonetics* (University of California, Los Angeles), 1972, **22**, 48-75.

LIBERMAN, A. M., COOPER, F. S., SHANKWEILER, D. P., & STUDDERT-KENNEDY, M. Perception of the speech code. *Psychological Review*, 1967, **74**, 431-461.

LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. The role of selected stimulus variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, 1952, **65**, 497-516.

LIBERMAN, A. M., DELATTRE, P. C., & COOPER, F. S. Some cues for the distinction between voiced and voiceless stops in initial position. *Language and Speech*, 1958, **1**, 153-167.

LIBERMAN, A. M., DELATTRE, P. C., COOPER, F. S., & GERSTMAN, L. J. The role of consonant-vowel transitions in the perception of the stop and nasal consonants. *Psychological Monographs*, 1954, **68**, 1-13.

LIBERMAN, A. M., MATTINGLY, I. G., & TURVEY, M. T. Language codes and memory codes. In A. W. Melton & E. Martin (Eds.), *Coding-processes in human memory*. New York: Halstead Press, 1972.

LINDBLOM, B., & STUDDERT-KENNEDY, M. On the role of formant transitions in vowel recognition. *Journal of Acoustical Society of America*, 1967, **42**, 830-843.

LISKER, L., & ABRAMSON, A. S. A cross language study of voicing in initial stops: Acoustical measurements. *Word*, 1964, **20**, 384-422.

SHANKWEILER, D., STRANGE, W., & VERGRUGGE, R. Speech and the problem of perceptual constancy. In R. Shaw & J. Bransford (Eds.), *Perceiving, acting, knowing: Toward an ecological psychology*. Hillsdale, N.J: Erlbaum, 1977.

STEVENS, K. N. The potential role of property detectors in the perception of consonants. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech*. New York: Academic Press, 1975.

STEVENS, K. N., & HOUSE, A. S. Speech perception. In J. V. Tobias (Ed.), *Foundations of modern auditory theory* (Vol. 11). New York: Academic Press, 1972.

STUDDERT-KENNEDY, M. The perception of speech. In T. Sebeok (Ed.), *Current trends in linguistics* (Vol. 12). The Hague: Mouton, 1974.

STUDDERT-KENNEDY, M. Speech perception. In N. J. Lass (Ed.), *Contemporary issues in experimental phonetics*. New York: Academic Press, 1976.

WINITZ, H., SCHEIB, M. E., & REEDS, J. A. Identification of stops and vowels for the burst portion of /p,t,k/ isolated from conversational speech. *Journal of Acoustical Society of America*, 1972, **51**, 1309-1317.

ZUE, V. W. *Acoustic characteristics of stop consonants: A controlled study*. PhD thesis, MIT, May 1976.

## NOTES

1. Voiceless transitions have been given due weight in studies of voiceless stops (Liberman, Delattre, & Cooper, 1958) and fricatives (Harris, 1958).

2. A relatively open response set provides a sensitive measure of how "stoplike" a signal sounds. In a situation where only /b,d,g/ are permitted as responses, the identifiability of the signals may be overestimated. For example, a signal composed of labial burst and a steady-state vowel, such as /ɪ/, sounds like a click followed by /ɪ/. However, if only /b,d,g/ are permitted as responses, then a subject may well feel that, since the click does not sound like a high-frequency alveolar burst and is not affricated like a velar burst, (s)he should respond /b/. A correct /b/ response would then be made to a signal that does not sound like /b/.

3. The general level of identification of the burst-plus-vowel syllables in Experiment 1 was lower than in either Experiment 2 or Fischer-Jørgensen (Note 2). We have no systematic explanation for the differences between our own experiments, since both speakers spoke their syllables in the same sentence frame at a conversational rate of speech, and were recorded under identical conditions. However, in general, we should expect that when trained speakers record isolated words or syllables in citation form, as did Fischer-Jørgensen (Note 2), the level of identification for burst-plus-vowel syllables constructed from those recordings will be high. On the other hand, in a more nearly normal situation, where an untrained informant produces a word or syllable in sentence context at a conversational rate of speech, as did our speakers, we should expect bursts to be less prominent and therefore less effective cues. Thus, the level of identification of burst-plus-vowel syllables will vary from experiment to experiment depending on the speaker, the manner in which the syllables were produced and, as noted in Footnote 2, the nature of the response set. Nevertheless, we would expect the *pattern* of burst effectiveness across vowel contexts to be similar for different speakers and experiments.

(Received for publication September 29, 1976;
revision accepted February 24, 1977.)