# CONTEXTUAL INFLUENCES ON STOP CONSONANT ARTICULATORY POSTURES IN CONNECTED SPEECH

James Dembowski* and John R. Westbury†
*State University of New York at New Paltz, †University of Wisconsin-Madison

**ABSTRACT**

Phonetic contexts are assumed to produce consistent allophonic variations in articulatory posture across speakers within a language. However, recent studies of multiple talkers producing connected speech have suggested that contextual influences on articulatory posture may be more varied and less predictable than commonly assumed. This study examined contextual variability in articulatory postures for bilabial, alveolar, and velar stop consonants, at the releases of all stops in 37 English sentences, produced by 12 normal young adults. Data included 557 tokens for each speaker, distributed across 190 phonetic contexts. Cluster analysis was used to find postural similarities among context conditioned phones (within place categories), and then phonetic contexts were examined to see if and how they reflected postural similarities. Results showed logical phonetic associations among contextually conditioned phones, but not the same associations from one speaker to another. Across speakers, liquids and fricatives appeared to affect stop posture more than vowels.

## 1. INTRODUCTION

Students of speech production commonly assume that phonetic contexts influence articulatory postures for consonants in consistent and predictable ways. Over time, certain of these assumptions have achieved an acceptance akin to the force of law. For example, velar stops have long been described as having two primary variants, front and back, depending on the articulatory requirements of surrounding sounds [4]. This description is not without empirical support. Both Houde [3] and Kent & Moll [5] found a front-back velar consonant distinction in their respective cineradiographic studies. However, their analysis employed limited contexts and speakers. A more recent study [2] employed multiple speakers, greater and more varied contexts, and meaningful connected speech tasks rather than citation-form nonsense syllables. This study generated some results inconsistent with earlier analyses: to wit, no consistent front-back distinction was apparent among /k/ articulations across 18 contexts produced by three speakers. Thus the concept of two primary velar allophones appeared to be a methodological artifact of having studied velars in limited and deliberately contrastive vowel contexts. A stronger test of whether contextual variations (in velars or any other segments) represent distinct allophones, used across speakers of the same language, requires an examination of the distributions of articulatory postures over many and varied contexts, for relatively many speakers. This report summarizes an attempt to do that.

## 2. METHODS

### 2.1. Analytic Approach

The methods of this study represent something of a reversal of the approach commonly used to examine phonetic context effects on articulation. Most early studies began by selecting phonetically contrastive utterances (usually VCV or CVC syllables), and then looked at articulatory postures in production of those utterances to see if and how they reflected the chosen contrasts. For the present study, *all* stop consonant productions in a set of connected speech tasks from a large database were examined without regard to any prior concern for their phonetic or acoustic contrasts. Articulatory postures for the stop productions were analyzed first to find those which were similar, and then phonetic contexts were reviewed to uncover phonetic similarities among posturally similar stop articulations.

### 2.2. Speakers and Tasks

Data were obtained from 12 of 57 speakers contained in the University of Wisconsin X-Ray Microbeam Speech Production Database (XRMB-SPD) [6]. The speech tasks for each speaker included a set of 37 meaningful English sentences, some produced multiple times, and others produced only once. Release bursts associated with all bilabial, alveolar and velar stops in these sentences were identified from the sound pressure waveform. The speech sample for each talker included 557 stop tokens, distributed across 190 phonetic contexts. Articulatory postures were represented by eight articulator landmarks: four radio-opaque pellets aligned along the midline of the tongue, one each at the mid-sagittal vermillion border of the upper and lower lips, and two on the jaw. Reference pellets allowed data to be expressed relative to a head-based coordinate system in which the abscissa was the maxillary occlusal plane and the ordinate was given by the plane normal to the occlusal plane and passing through the tips of the central maxillary incisors.

### 2.3. Articulatory Posture Analysis

Viewing articulatory data with respect to an anatomically based coordinate system, or some set of anatomic landmarks, is a common approach to seeking spatial patterns. Traditionally, allophonic variations in consonants have been described in terms of "high," "low," "front," or "back" spatial variations within the mouth, such as the front-back velar distinction addressed above. However, attempts to use traditional spatial quadrant divisions (high front, low front, high back, low back) to describe systematic spatial patterns among the multiple articulator points and phonetic contexts of the present study presented two problems. First, when axes were imposed on individual pellet distributions in order to make front, back, high and low distinctions, a boundary dividing, say, high from low data, produced divisions in which two data points very close to each other in space might fall just on either side of the high-low boundary, and then be considered categorically "different," while two points relatively distant in space could be classed as categorically identical. A second problem was that multiple articulator points for a given phonetic event did not occupy comparable positions in a simple anatomical quadrant space, thus forcing a rather arbitrary decision about which pellet might be "most important" or "most representative" of articulation for that event. For example, one tongue pellet might be high for a given consonant, while another was low.

These problems were addressed by employing cluster analysis as an alternative to the usual high-low-front-back descriptors for articulator points in particular segments. Cluster analysis derives data groupings based on measures of distance among points rather than on the imposition of boundaries. Thus, for any one pellet, data points close together are grouped together before they are classified with points further away. Additionally, cluster analysis is a multivariate procedure that, with centering and scaling to adjust for differences in variance among pellet distributions, allowed entering positions of all eight articulator pellets for a given segment together into the analysis, without concern for whether one pellet was more or less representative of segment articulation than any other. Thus, differences in quadrant-related positions among pellets for a given segment ceased to be an issue in deciding how to group articulatory postures.

Cluster analysis was performed on all contextually unique "allophones" within each of the three stop-place categories. Note that, for lack of a better term, the label "allophone" is used somewhat untraditionally to mean any contextually unique stop phoneme. Also, the "contextually unique" condition means not only a unique set of flanking phones for a given stop segment under study, but also indicates identical phoneme strings occurring in different sentences. For example, the segments of the word "school" were considered to occur in *two* "allophonic" contexts because the word appeared in two different sentences. Voiced and voiceless segments were grouped together as "allophones" of each place. That is, /p/ and /b/ segments were combined within the bilabial place, /t/ and /d/ within the alveolar place, and /k/ and /g/ within the velar place. Contextually unique stops which were produced only once by a speaker were represented by the one available set of articulator positions, but segments for which there were multiple tokens were represented by the mean *(x,y)* values across all available tokens. Each speaker's pellet data were centered with respect to each pellet's "place" centroid, and scaled by within-pellet x- and y- dimension standard deviations.

The clustering algorithm used here employed Euclidean distance as a dissimilarity measure, and employed the "compact" (i.e., "complete linkage") method [1]. The cluster analysis was performed separately for each place category within each individual speaker. The goal of the analysis was to identify spatially homogeneous postures in order to see if they corresponded to phonetically homogeneous contexts. Each speaker's within-place analysis sought to divide "allophone" postures into eight groups. Given the number of contextually unique segments in each place category (45 to 95), this cut of the cluster analysis insured that most groups contained several cases (with few "groups" of one), and yet the number of cases in each group was not so inclusive as to frustrate attempts to find some common phonetic thread among them. Subsequently, the within-speaker clusters were compared to see if any cross-speaker patterns emerged.

## 3. RESULTS AND DISCUSSION

### 3.1. Within-speaker Clusters

Data divisions produced by cluster analysis seemed to reflect some broad (though imperfect) correspondence with phonetic context patterns for "allophone" groups within speaker and place category. From one speaker to another, however, the phonetic correspondences were not necessarily the same. Table 1 exemplifies selected velar clusters from two speakers. Two clusters are shown for speaker 16 (s16), and three for speaker 20 (s20). For s16, Cluster A included the combination /sku/, from the word "school" produced in two different sentences (and thus classified as two different "allophone" tasks). The same cluster also included the combination /lkʊ/ from the phrase "creole ***c***ooking." One might surmise that /lkʊ/ clustered with the two /sku/ tasks because there is an articulatory similarity in the alveolar constrictions for /l/ and /s/, as well as a similarity in the supposed high-back-round vowels /ʊ u/. Of course, it seems transparently logical that the two /sku/ contexts should be classified together. Without actually looking at data, one might be tempted to assume that these similar contexts would cluster together and, in fact, they did so for most speakers. But not all. Some speakers did *not* show the expected pattern. This was the case for speaker 20, for whom three different clusters held the three contexts that grouped together for s16.

| Cluster | Speaker | |
|---|---|---|
| | S16 | S20 |
| A | /sku/ school<br>/sku/ school<br>/lkʊ/ creole cooking | /sku/ school<br>/rks/ dark suit<br>/æks/ wax<br>/eks/ make sense<br>/ɛks/ excluded<br>/oks/ jokes |
| B | /rks/ dark suit<br>/skm/ ask me<br>/æks/ wax<br>/eks/ make sense<br>/zkr/ does creole<br>/ɛks/ excluded<br>/skl/ excluded | /sku/ school<br>/skm/ ask me<br>/ukɛ/ you can<br>/rkf/ dark fibers<br>/ngr/ in greasy<br>/ngr/ ingredients<br>/zkr/ does creole<br>/ŋkr/ shaving cream<br>/dkə, nkə/ and comp...<br>/əgo/ the gorgeous |
| C | | /lkʊ/ creole cooking<br>/əkʌ/ a country<br>/skl/ excluded<br>/ŋkð/ think that's |

Table 1. Selected velar clusters for two speakers.

For s20, one occurrence of the /sku/ context clustered with other contexts which all had /s/ in common (Table 1, s20's Cluster A). However, some of these other contexts had phonemes which might conventionally be considered "back" (/o/ and probably /r/), while other phonemes are conventionally considered "front" (/e ɛ æ/). The second occurrence of /sku/ clustered with several contexts containing /r/, along with segments generally classified as "alveolar," but varying in manner (Cluster B). This group also contained two contexts with "high-back" vowels (/ukɛ/ and /əgo/).

For s20's Cluster C, the cluster containing /lkʊ/, the association among contexts is not transparent. Two contexts have /l/ in common, and a similarity in supposed forward, close constrictions for /l s ð/ might account for this particular cluster pattern.

To extend the comparison between speakers, note that even though one other context with which /skl/ (in “ex***c***luded”) groups in s20's data contains /l/, no other context contains /s/. In contrast, for s16, all the contexts with which /skl/ groups contain /s/or /z/, and none contain /l/ (Table 1, Cluster B, s16). Comparing only the two clusters containing /skl/ across the two speakers, one might think that for s16 alveolar fricatives exert a coarticulatory influence which is less pronounced for s20, but then a look at the first group of /sku/ (“school”) related contexts for s20 (Cluster A) shows that such a generalization would be overly simple. The main point of these comparisons is that within each speaker there appeared to be a phonetic logic to the postural clusters, but the logic was not necessarily the *same* from one speaker to another.

Even so, variation in the pattern of clusters did not preclude some cross-speaker consistency. To find consistent clusters across speakers, each cluster of each speaker was compared with each cluster of every other speaker, within place categories. These comparisons identified all groups of two or more “allophones” co-occurring in a pronounced majority of speakers (where “pronounced majority” was somewhat arbitrarily defined as 8 out of 12, or 67% of speakers). A thorough presentation of cross-speaker patterns is beyond the scope of this report. However, the most pronounced patterns can be summarized and exemplified.

| Flapped Alveolars | |
|---|---|
| A | /ɑɾɚ/ water<br>/ʌɾɑɪ/ what I<br>/rdɚ/ order<br>/ʌɾɚ/ butterfly<br>/ɑɾə/ lot of<br>/ɪɾəl/ little |
| B | /ɑɪdə/ provide a<br>/ɑɪɾəm/ item |
| C | /iɾə/ me to<br>/idi/ ingredients |

Table 2. Cross-speaker context classes for flapped alveolars.

### 3.2. Cross-speaker Patterns

In the present data, the most pronounced cross-speaker contextual influences came from “liquids” /l r/ and from alveolar, palatal, or lingua-dental fricatives. This is interesting from a historical perspective in that studies of context effects on stop consonants have tended to focus on vowel influences, an inevitable consequence of employing CVC or VCV utterances. The effects of English consonants on consonants have been relatively little addressed, despite the fact that English is rich with consonant clusters and is not primarily a language of repetitive stop-vowel contrasts. To the extent that vowel influences emerged in the present study, they were most obvious in the bilabial stop data, and seemed less robust and more complex for lingual stops. One lingual category where vowel influence did appear strong was flapped alveolars, though, of course, English alveolar flaps only appear in intervocalic contexts, or in vowel-stop-syllabic liquid contexts (where a schwa may be considered to come between the stop and the liquid), so there are no adjacent obstruents to provide an influence. Interestingly, however, where flaps were formed before syllabic liquids (e.g., “water,” “little”), these vowel-flap-liquid contexts clustered together across speakers, and clustered separately from vowel-flap-vowel contexts, again suggesting the distinct contextual influence of liquids. This is shown in Table 2 which presents the three classes of contexts which appeared across a majority of speakers producing potentially flapped alveolars. Group A is dominated by liquids, especially /r/. Group B appears distinguished by the diphthong /ɑɪ/ in front of the flap, while Group C is characterized by alveolars in the context /i/.

<table>
<tr><th colspan="2">Place</th></tr>
<tr><th>Alveolar</th><th>Velar</th></tr>
<tr><td rowspan="2">/ztə/ things to<br>/str/ street<br>/ntr/ country<br>/ətɔr/ dormitory<br>/ʌtʃ/ much<br>/ntʃ/ second children<br>/rdɑ/ your dark<br>/rdu/ or do<br>/ədɔr/ the dorm...<br>/rdʒ/ large<br>/utʃ/ do Charlie’s<br>/stj, stʃ/ moisture<br>/ɑʊtu/ about too<br>/ʌtʃ/ much<br>/itʃ/ each<br>/ɚtʃ/ urchins<br>/#tʃ/ #cheap<br>/ədʒ/ damage<br>/rdʒ/ gorgeous<br>/zdʒ/ always jokes</td><td>/ngr/ in greasy<br>/ngr/ ingredients<br>/zkr/ does creole<br>/ŋkr/ shaving cream<br>/tgr, ʔgr/ outgrew</td></tr>
<tr><td>/əkr/ across<br>/ogr/ program<br>/mgr/ problem grows</td></tr>
</table>

Table 3. Selected cross-speaker context classes (lingual stops).

In most of the present data the dominant influence, in addition to liquids, seemed to come from fricatives. This primary pattern is exemplified in one relatively large class of alveolar contexts which grouped across speakers, shown as the “alveolar” column in Table 3. This group shows a variety of vowel contexts, but a preponderance of /r/ and palatal fricative contexts. The contextual influence of /r/ potentially extends across multiple segments. In the alveolar group example, the /ədɔ/ and /ətɔ/ contexts seem anomalous until one recognizes that they come from the phrase “the dormitory” and that /r/ follows the post-stop vowels. Velars showed two distinct groups dominated by /r/. Both are shown in the “velar” column of Table 3.

### 3.3. Explaining Context Effects

One possible explanation for the powerful influence of liquids and fricatives is that there may be a hierarchy of postural constraint among segment articulations, and those which are most constrained posturally exert the greatest contextual influence. Liquids are acoustically distinguished from vowels primarily by the relative frequency of the third formant (F3: low for /r/, high for /l/). Within the vocal tract, the spatial extent of nodes and anti-nodes in the standing wave pattern for F3 is more limited than for lower formants. Thus, relative to the demands of vowel production, a speaker may have less allowable variability in where to squeeze the vocal tract tube for liquids in order to produce an appropriate formant pattern. Similarly, the demands on degree (and possibly on place) of constriction for fricatives may be greater than for vowels and stops, again limiting allowable articulatory variability for these segments. Thus, when the acoustic requirements for segments impose relatively strict postural constraints, those segments may exert proportionally great contextual influence over segments whose acoustic requirements allow greater postural variability in their production.

### 3.4. Additional Considerations

Finally, consider that the phonetic influence of flanking segments is unlikely to be the only contextual influence on articulation of a given sound, and that the speech production system appears to offer substantial flexibility in how a speaker achieves acoustic-phonetic goals. Dialectical patterns, prosodic patterns, and individuals' idiosyncratic articulations (arising from some combination of individual anatomy and learned patterns of movement over the course of speech acquisition) must influence how context-conditioned "allophone" postures cluster -- and how they *do not* cluster. Of the 190 contextually unique "allophones" examined here, only about 60% clustered together into distinct groups across speakers. The approximately 40% of remaining allophones clustered into different combinations across the twelve speakers, suggesting that the speakers' language, and the acoustic properties of the vocal tracts through which the language is expressed, offer speakers some articulatory alternatives for the production of specific segments in specific contexts, and that speakers choose to exploit those alternatives.


**ACKNOWLEDGMENTS**

Research support was provided by USPHS/NIH Grant DC00820, and a collaborative research agreement between the University of Wisconsin-Madison and ATR Human Information Processing Research Laboratories, Kyoto, Japan.


**REFERENCES**


[1] Aldenderfer, M.S. and Blashfield, R.K. 1984. *Cluster Anlaysis*. Newbury Park: Sage Publications.

[2] Dembowski, J., Lindstrom, M.J., and Westbury, J.R. 1998. Articulator point variability in the production of stop consonants. In Cannito, M.P., Yorkston, K.M. and Beukelman, D.R. (eds.), *Neuromotor Speech Disorders* (pp. 27-46). Baltimore: Paul H. Brookes.

[3] Houde, R.A. 1967. *A Study of Tongue Body Motion During Selected Speech Sounds*. Unpublished doctoral dissertation, University of Michigan.

[4] Jacobson, R., Fant,G., and Halle, M. 1969. *Preliminaries to Speech Analysis*. Cambridge, MA: The MIT Press.

[5] Kent, R.D. and Moll, K. L. 1972. Cineflourographic analysis of selected lingual consonants. *Journal of Speech and Hearing Research*, 15, 453-473.

[6] Westbury, J.R. 1994. *X-Ray Microbeam Speech Production Database User's Handbook*. Madison: University of Wisconson-Madison, Waisman Center.