THE CENTER OR EDGE: HOW ARE CONSONANT CLUSTERS ORGANIZED WITH RESPECT TO THE VOWEL?

Douglas N. Honorof[†*] *and Catherine P. Browman*[*]
[†]*Yale University and* [*]*Haskins Laboratories, New Haven, Connecticut, U.S.A.*

ABSTRACT

Stable inter-gestural timing patterns were sought for phonotactically-permissible (CC)CVX and XVC(CC) accented monosyllables in American English. Movement evidence for four speakers confirmed the hypotheses of Browman and Goldstein [1] that a pre-vocalic consonant or cluster is organized with respect to a 'tautosyllabic' nuclear vowel by its center (*i.e.*, C-center), but a post-vocalic consonant (or sequence of consonants) by its (first) left edge.

INTRODUCTION

Having examined x-ray microbeam data for one speaker of American English, Browman and Goldstein found evidence for the *C-center* (defined below) [1]. Specifically, they argued that, judging by patterns of articulatory stability, the C-center of a *pre-vocalic* consonant or consonant cluster is more tightly coordinated with the vowel gesture that corresponds to a following acoustic vowel than is either the *left edge* (henceforth, *LE*) of the first pre-vocalic consonant plateau or the *right edge* (henceforth, *RE*) of the last one. However, at least for the monosyllabic target words in their data set, they suggested that it is the LE of the first *post-vocalic* consonant rather than the C-center of the whole sequence of consonants (or the C-center of just the coda consonants) that is most tightly coordinated with the vowel gesture, regardless of whether that vowel and consonant are separated by a word boundary. (See Fig. 1). They did report, however, that those vowel-to-LE measures are even more stable when there is no intervening word boundary (but see [2] for apparent counter-evidence.)

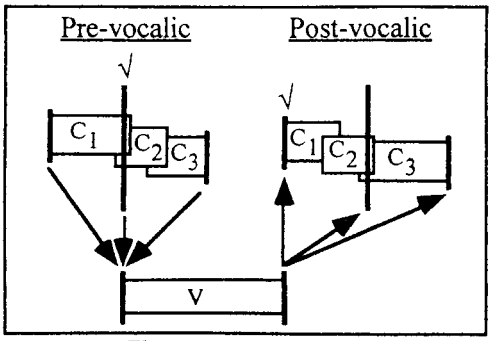

Figure 1: The potential phasing relations considered in the present experiment are indicated schematically with bold vertical lines and arrows. Those phasing relations that are argued for in [1] are indicated with check marks.

One implication of this picture of organization is that increasing the number of consonant gestures in a coda should not reduce the acoustic duration of a 'tautosyllabic' vowel. This implication is at odds with the notion of 'compensatory shortening'. (See [3].)

The experimental results that we report here allow us to address these articulatory and acoustic issues.

METHODS

Design and Stimuli

The present design systematically varies one-, two-, and three-consonant pre-vocalic and post-vocalic consonant 'clusters' in accented, monosyllabic (real and nonsense) English target words. The utterances were designed to disallow rightward re-syllabification on phonotactic grounds. (See Table 1.)

Data Collection Technique and Procedure

All data were collected at the University of Wisconsin x-ray microbeam facility [4]. The utterances were presented to the subjects on a video screen in quasi-random order. The microbeam system then tracked the Cartesian coordinates of gold pellets (2.5-3.0 mm in diameter) affixed to the mid-line surfaces of the subject's articulators as he or she read each utterance aloud at least five times, while acoustic data were simultaneously recorded. Before analyzing the articulatory data, we compensated for any head movement added to articulator movement by using position data

Table 1: List of utterances. Capital letters indicate accent.

Frame: 'I read [past] _____ again.'		
	Position of Target Cs	
Target Cs	*Pre-vocalic*	*Post-vocalic*
[s]	cuff SAYED	CUSS fade
[p]	cuff PAID	CUP fade
[ps]	----------	CUPS fade
[sp]	cuff SPAYED	CUSP fade
[sps]	----------	CUSPS fade
[l]	cuff LAID	----------
[lp]	----------	CULP fade
[lps]	----------	CULPS fade
[pl]	cuff PLAYED	----------
[spl]	cuff SPLAYED	----------

gathered on reference pellets affixed to the nose and upper incisor. The data were automatically rotated to the occlusal plane.

Subjects

Four college students participated in the present study. All were natives of Wisconsin between the ages of 18 and 20, three female, one male. All were of normal speech and hearing ability, with the exception of Subject STR2 who had a 30 dB notch at 8 kHz in the right ear only, which we do not believe to have affected his performance on the task.

Measurement Procedure

The sampling rates for the x-ray data differed from pellet to pellet (40-160 Hz), and sometimes for a single pellet from utterance to utterance. Therefore, when smoothing articulatory data, we set the number of points in our (software) triangular filters according to channel-specific sampling rates so that window sizes were always brought as close as possible to 25 msec. The x-ray data were then re-framed by interpolation to 200 Hz (T=5 msec/frame).

In all cases we followed Browman and Goldstein in defining consonant and vowel gestures on the oral tier only [1] (see also [5]). Wherever possible, we labeled the relevant articulatory trajectory for each target consonant gesture by automatic algorithm, first finding the relevant extremum, then marking as the LE the first frame whose displacement amplitude fell within a spatial noise level (SNL) equal to five percent of the mean of the speaker's mean range of displacement across all displacement trajectories analyzed for the present experiment (including those from utterances later excluded from analysis). We marked the right edge (RE) as the last frame whose amplitude fell within that same SNL.

Lip Aperture

In order to label 'p' in a principled way, we computed a trajectory that we call *lip aperture* (henceforth, *LA*) by subtracting vertical displacement of the lower lip pellet from that of the upper lip pellet (positioned on the lower and upper vermilion borders, respectively). Thus, a minima or *valley* in LA presumably corresponds to attainment of mid-line closure of the lips.

We had hoped to measure 'p' as the LE and RE of a basin around that valley. Indeed, in general this is what we did (though for one token we were only able to do so by reducing the SNL by .1 mm) However, in many cases, edges of contiguous 'p#f' and 'f#p' sequences were 'blurred' in LA, perhaps due to conflicting demands being made on the lower lip by the two closure gestures. Although the failure of the articulators to return consistently to their 'neutral' positions between these two contiguous gestures presents no theoretical difficulties, in cases where labial gestures were contiguous, we were forced to label 'p' by automatically picking the 'p' edge that was not contiguous with 'f', and then calculating the contiguous edge in terms of an utterance-type-specific ratio of the known edge to the relevant anchor point. These ratios were calculated on the basis of various averages of the ratios for non-contiguous edges to relevant anchor points for the tokens whose 'f'-contiguous labels were found automatically ('f#sp', 'f#spl', 'sps#f', 'ps#f' and 'lps#f').

Tongue Tip Constriction Degree

Due to the non-zero slope of the hard palate in the region where alveolar consonants are articulated, peaks in mid-line vertical displacement of the tongue tip pellet (positioned 7 to 9 mm back of the actual tip) do not always co-occur in time with the actual moment of tightest constriction as measured for that pellet. Therefore we measured the relative

amplitude of tongue tip displacement for 's' in a trajectory that we call (inverted) *tongue tip constriction degree* (henceforth, *TTCD),* which is simply the x and y coordinates of the tongue tip pellet rotated to the slope of a relevant segment of a mid-line pellet tracing of the palate by the formula:

$$TTCD = - TTX * SIN\theta + TTY * COS\theta$$

where θ is the slope of the palate segment found by linear regression.

We also labeled 'l' in TTCD. Our justification for doing so is found in Sproat and Fujimura's suggestion that the tongue tip gesture for 'l' may be consonantal, but the tongue body gesture, vocalic [6]. Because the present research concerns the timing of consonant gestures with respect to vowel gestures, we ignored the movement of the tongue body for 'l'.

Right Anchor Point

We labeled the trans-vocalic *right anchor point* (henceforth, *RAP*, *i.e.*, the attainment of target for post-vocalic 'd'—see [1]) by automatic algorithm, choosing the first positive-to-negative zero crossing in the region of interest of the first derivative of smoothed TTCD (SNL=10% of the subject's mean velocity range for that pellet across utterances). Our algorithm required two consecutive frames to be within the zero region, and none without, before the edges of the zero region were declared.

Left Anchor Point

We used a trans-vocalic *left anchor point* (henceforth *LAP*) similar to that used in [1]. That is, we identified absolute peak vertical displacement of a pellet affixed to the tongue dorsum (58 to 64 mm back of tongue tip). This measure is effectively equivalent to the C-center of the singleton onset ('k'). The slope of the soft palate remains nearly horizontal in the region where tongue dorsum constrictions are articulated for all four subjects, so we found it unnecessary to rotate the tongue dorsum data before labeling the LAP.

C-centers

Again, following [1], we computed the C-center as the mean of the temporal midpoints of the plateaus (or basins) surrounding the spatial peaks (or valleys) associated with the consonant gestures in a cluster. However, we were not always able to distinguish the neighboring edges of 's' and 'l' in TTCD when both occurred in one cluster, *i.e.*, 'spl' and 'lps'. Nor were we always able to distinguish the RE of the first 's' from the LE of the second 's' in 'sps' clusters. Therefore, in such cases, we chose to label only the LE of the first TTCD gesture and the RE of the second. We then counted the resulting 'plateau' as a single gesture for the purposes of computing the C-center. (However, we chose not to analyze the post-vocalic 'sp(C)' and 'lp(C)' utterances of Subject SCH2, who appeared to have had difficulty producing many of them under the experimental conditions.)

For an 'f#p' or 'p#f' contiguous utterance, we computed the C-center with reference to the normalized 'p' edge as discussed above.

Acoustic Vowel Duration

We also measured the acoustic duration of [ʌ] before post-vocalic target consonants. To this end we segmented the waveform by placing two labels without particular reference to articulatory labels, one at the start of aspiration following 'k' and another at the instant of gross spectral change corresponding to the first following post-vocalic consonant. However, we were not always able to identify discrete acoustic boundaries between the vowel and post-vocalic 'l'.

RESULTS AND DISCUSSION

For the articulatory data, separate ANOVA were run for each subject for each position (pre-vocalic and post-vocalic) on one factor with three levels: RE, LE, and C-center.

For the pre-vocalic consonants, in each case the measure for each of the three levels was subtracted from the RAP, as in [1]. For all subjects, the pre-vocalic C-center organization was more stable (*i.e.*, had a smaller standard deviation from the group mean with a lower Levene's p-score) than either of the other measures, as was found in [1]. By subject, $F(2,87)=7.81$, Levene's $p<.01$; $F(2,87)=4.08$, Levene's $p<.03$; $F(2,81)=9.67$, Levene's $p<.01$; $F(2,84)=10.09$, Levene's $p<.01$.

For the post-vocalic consonants, the measure for each of the three levels had the measure for the LAP subtracted from it, as in [1]. This time, for all four subjects, the local LE organization proved more stable than the C-center or RE, again as the analysis of data from the single subject in [1] suggested. By subject, $F(2,99)=27.03$, Levene's $p<.0001$; $F(2,42)=6.04$, Levene's $p<.01$; $F(2,93)=17.56$, Levene's $p<.0001$; $F(2,102)=19.41$, Levene's $p<.0001$.

The acoustic vowel duration for [ʌ] in the accented syllable did not show a significant or consistent trend toward increasing or decreasing with cluster complexity; Levene's p-values ranged across speakers from >.17 to >.89. While this finding would not be easily explained by 'compensatory shortening' (see [3]), it is expected given the results of our articulatory analyses in which the LE was most stable.

We interpret the pre- and post-vocalic results as strongly supporting the suggestion in [1] that there is a difference in pre- and post- vocalic organization (in American English monosyllabic words[1]), at least for labial consonant gestures and consonant sequences involving labial gestures. Nevertheless we refrain from drawing rigid conclusions until our findings can be confirmed for other constriction locations, and until comparable results obtained from point-source tracking and linguapalatal devices have been scrutinized, which we hope to do in the future.

ACKNOWLEDGEMENTS

We would like to thank Dani Byrd for her helpful comments. The present study was supported by NIH Grant HD-01994 and NIH Grant DC-00121 to Haskins Laboratories.

REFERENCES

[1] Browman, C. P. & Goldstein, L. (1988), "Some notes on syllable structure in articulatory phonology", *Phonetica,* vol. 45, pp. 140-155.

[2] Byrd, D. (in press), "C-centers revisited", *Phonetica.*

[3] Munhall, K., Fowler, C., Hawkins, S. & Saltzman, E. (1992), "'Compensatory shortening' in monosyllables of spoken English", *Journal of Phonetics,* vol. 20, pp. 225-239.

[4] Nadler, R. D., Abbs, J. H. & Fujimura, O. (1987), "Speech movement research using the new x-ray microbeam system", *Proceedings of the XIth International Congress of Phonetic Sciences,* vol. 1, Tallinn, Estonia: Academy of Sciences of the Estonian S.S.R. Institute of Language and Literature, pp. 221-224.

[5] Fujimura, O. & Lovins, J. B. (1978), "Syllables as concatenative phonetic units", in Bell, A. & J. B. Hooper (Eds.): *Syllables and Segments,* New York: North-Holland Pub. Co., pp. 107-120.

[6] Sproat, R. & Fujimura, O. (1993), "Allophonic variation in English /l/ and its implications for phonetic implementation", *Journal of Phonetics,* vol. 21, pp. 291-311.

[1]For a discussion of how these claims relate to P-centers, weight units and moraic structure, phonetic affixes and extrasyllabicity, compensatory lengthening, allophonic variation, and issues of universality, see [1].