THE ORGANIZATION OF SEGMENT-INTERNAL GESTURES

Bryan Gick
Yale University and Haskins Laboratories, New Haven, CT, USA

ABSTRACT

Some theories of the coordination of speech gestures maintain that higher-level phonological categories such as the segment are not needed for the description of observed timing patterns. Likewise, few theories of phonology make substantial reference to gestures. The present paper examines evidence of timing stability between gestures within traditional segmental units. Specifically, relative timing of gestures internal to American English /m/ and /l/ are considered in post-vocalic clusters. Intrasegmental gestural coordination patterns have been relatively well documented in immediately postvocalic, word-final position. However, the question of whether these patterns remain stable when these segments are preceded or followed by additional material has not been answered. Articulometer data are provided for American English /l/. Results indicate that although traditional segments indeed show distinct patterns of internal coordination, surrounding segmental material nevertheless has a significant and patterned effect on this coordination. A resulting model for segment-internal and -external gestural organization is proposed.

1. INTRODUCTION

A number of recent studies have shown patterned variations in the timing of tautosegmental gestures in speech articulation [1, 2, 3, 4, 5]. Additionally, several other studies have investigated gestural timing relations across different segments in consonant clusters [6, 7, 8]. In these studies, however, only the "consonantal" (C-) gestures[1] of multi-gesture segments (e.g., the tongue tip fronting gesture of /l/) have been measured. No segment-internal timing patterns (particularly between C-gestures and "vocalic" (V-) gestures, such as the tongue dorsum backing gesture of /l/) have been investigated in consonant clusters to date.

Byrd's phase window framework [9] "predicts that the relative timing of gestures constituting a 'segment' will be less affected by contextual variables than that of other gestures not constituting a 'segment'." The present study begins to address this prediction by observing the effect of varying surrounding material on segment-internal gestural timing in /l/. The question will be addressed of whether segments are phased to surrounding material as units that retain stable internal timing, whether gestures are phased to surrounding gestures in patterns independent of segmental affiliation, or whether some combination of these two factors are involved. The answer to this question will bear crucially on the choice between possible models of gestural coordination.

1.1. Previous Work

Previous studies indicate that prevocalic and postvocalic C-gestures follow fundamentally distinct timing patterns within clusters [6, 7, 8]. The one previous study giving within-segment data that may be applicable to the present question [1] reports does not report on allophones in prevocalic clusters. In addition, of all the multiple-gesture segments, /l/ has been the most studied for internal timing phenomena in general. Thus, in order to incorporate and expand on these past findings, the present study focuses on allophones of /l/ in postvocalic consonant clusters.

1.1.1. Segment-internal 'Lag' in Postvocalic /l/. Krakow [1, p. 51] reported results on segment-internal timing in postvocalic allophones of /m/. She found that the C-gesture (lip closure) was achieved significantly later than the V-gesture (velum lowering; average "lags" ranged from about 90ms to about 260ms). A similar lag has since been observed by other researchers for /l/ [2, 3, 4]. In these studies, the average lag between the V-gesture (tongue dorsum backing) and the C-gesture (tongue tip fronting[2]) in final allophones ranged from 5ms to 110ms across speakers.

1.1.2. 'Chain Shift' in C-gestures. Several experiments concerning the 'C-center' [6, 7, 8] have shown that, while C-gestures in onset clusters are timed with respect to a single central point (the 'C-center'), only the left edge of the leftmost C-gesture in coda clusters is timed to the vowel. Other C-gestures in a coda cluster appear to follow in sequence. In a separate study, Munhall et al. [10] found that increasing the number of segments in a coda cluster had a shortening effect on acoustic vowel duration. They refer to this as 'compensatory shortening'.

It is possible to reconcile these findings by positing a 'chain-shift' of C-gestures to the left. At first glance, this could simply be interpreted as a shortening of the vowel (Fig. 1a). This possibility will be referred to as Hypothesis 1 (H1). Under this hypothesis, the timing relationship between tautosegmental gestures ('tip lag' in the case of /l/) is predicted to remain stable regardless of variation in surrounding segments.

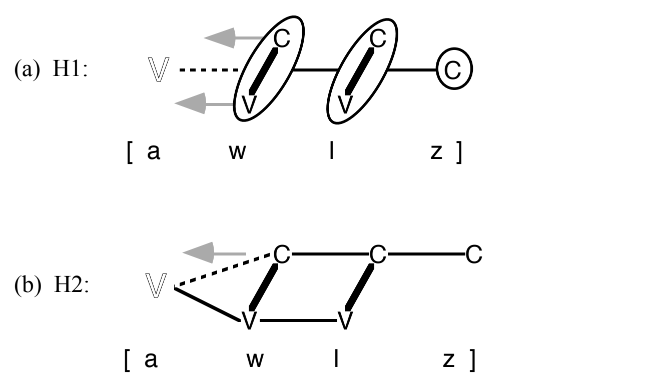

Figure 1. Two models of chain shift in the English word 'owls'. (a) shows the model predicted by H1, and (b) by H2. The outlined 'V' represents the peak vowel of the syllable, in this case /a/. All lines indicate phasing relationships. Bold lines link component gestures within segments. Dashed lines indicate the interval subject to shortening by chain shift.

However, as no previous works have tested V-gestures in coda clusters, H1 is not the only plausible hypothesis. It is also possible that the 'chain shift' in C-gestures applies *only* to C-gestures, leaving the timing of V-gestures (with respect to the peak vowel as well as to each other) unaffected. In this case, since the timing relationship between the peak vowel and the following coda V-gestures would be unchanged, any decrease in the acoustic duration of the peak vowel could be considered phonetic (possibly the result of simple overlap with the leftmost C-gesture). This alternative will henceforth be referred to as Hypothesis 2 (H2). Thus, H2 predicts that lag will be reduced by the introduction of additional segmental material in the coda, but crucially that the temporal reduction of the normal lag will be the same across all segments in the coda (Fig. 1b). There are other conceivable hypotheses, but these two are the only ones that assume a stable, linear behavior within the categories of C-gestures and V-gestures. More complicated results may require further postulation, but for the moment other possibilities will be ignored.

In Figure 1, C-gestures and V-gestures are represented on different 'tiers' based on their categorically distinct properties of timing and magnitude [1, 2, 3, 4, 5]. The similarity of this representation to those of autosegmental phonological theories such as Clements and Keyser's 'CV phonology' [11] is incidental, although there surely exist parallels and implications beyond the scope of the present work.

1.1.3. Krakow 1989. Krakow [1] reports a very small number of data that may be relevant to distinguishing between H1 and H2. Krakow's study looked at the time lag between achievement of velum lowering and bilabial contact in /m/ allophones. While most allophones were final singletons, as in 'home', 'seam' and 'palm' (Krakow's study was not concerned with the present issue), one of the stimuli used was 'helm'. From these, internal duration measures for the /m/ can be compared between the single /m/'s and the /m/ in the /lm/ cluster.

The results of this comparison are not conclusive. The lag in the /m/ of 'helm' appears to fall well within the range of the singleton /m/'s. This is consistent with H1. However, as this comparison is tangential to Krakow's goals in this experiment, only averages are given, with no statistical comparison. More importantly, though, the utterances were not controlled for syllable position. Complete stimuli were 'home E', 'seam ore', 'seem E', 'palm aid', and 'helm it'. It is therefore not clear what syllable positions these final /m/'s are in—whether they are in the coda, or whether they have been partly or fully resyllabified onto the following syllable. Since internal timing varies dramatically across syllable positions, (a major point of Krakow's work), it is impossible to draw any firm conclusions on the basis of this data.

2. EXPERIMENT

As a test to distinguish between our two models of segment-internal gestural organization, an experiment was conducted using point-parameterized midsagittal movement data collected from a single speaker.

2.1. Methods

2.1.1. Subjects. The articulations of one female native speaker of American English were measured in this study. The speaker is a colleague at Haskins Laboratories, but was unaware of the nature of the experiment.

2.1.2. Stimuli. This study controlled for resyllabification by using /h/ as an onset filler for the following word. /h/ is ideal for this position, as it does not interfere with oral articulators, yet still has been shown to establish a syllable boundary for gestural purposes [4]. Stimuli consisted of sentences of the form "I say x hotter again," where x is one of: "hall," "harl," "howl," "hile," "hoyle," or "helm." The first of these, "hall," a coda singleton, provides the baseline timing for postvocalic /l/; the middle four items show the effect of an intervening segment[3] on this basic timing; and the final item, "helm," shows whether this effect is the same in different positions in the coda cluster. Ten repetitions of each word were collected. Stress on the preceding vowel was effectively maintained by alternating only the word in question.

2.1.3. Data Collection and Analysis. The subject read stimuli aloud from randomized written lists presented in groups of twelve items per page. The subject was instructed to use her normal speech at a comfortable rate. The first and last members of each list were discarded, leaving ten items per page.

Data were collected using the EMMA (electromagnetic midsagittal articulometer—see [12]), a three-coil transmitter system at Haskins Laboratories, New Haven, CT. Small receivers were attached to the subject's tongue, lips, mandible, maxilla and nose bridge, the latter two used for correction of head movement in the midsagittal plane. Voltages induced in the receivers by three fixed electromagnets situated around the subject's head were used to determine location of the receivers in the midsagittal plane. Movement data was sampled at 500Hz. Of the receivers recorded, the following were used in this experiment: tongue tip (TT), tongue body (TB), tongue dorsum (TD), and upper and lower lip (UL, LL).

The measure used for the dorsal gesture of /l/ was TD backing. Locations of closure achievement were automatically selected by calculating velocity signals (first derivative) of the relevant movement signals, and locating the point at 3 percent short of zero velocity (of the total velocity range for that articulator in that dimension). For the tongue tip gestures, the speaker exhibited both fronting and raising, so locations were determined from tangential velocity of both TT fronting and TT raising. One repetition of "hall" and three repetitions of "harl" were omitted because of unclear or obscured movement maxima.

Additional measures were made to test for the 'compensatory shortening' effect reported in [10]: Duration from acoustic VOT (between initial /h/ and peak vowel) to achievement of TB raising for /j/ and to vertical lip aperture (UL - LL) for /w/. Comparisons were made for these measures between the pairs "hie/hile," and "how/howl." Ten repetitions each of each token were collected for this test, using the same procedures described above.

Analysis of variance (ANOVA) was used to test for significant variance in the different allophones of /l/.

2.2. Results

2.2.1. Replication of Tongue Tip Lag. The lag effect reported in [2, 3, 4] was replicated in the present experiment. For singleton coda /l/'s in "hall," our speaker showed an average lag of 126.42ms from achievement of TD backing to achievement of TT fronting/raising.

2.2.2. Effect of Adding Segmental Material. The bar graph in Figure 2 and Table 1 show ANOVA results for the lag effect in singleton tokens (“hall”, leftmost bar) and 2-consonant codas (“harl,” “howl,” “helm,” “hoyle,” and “hile”). Error bars indicate standard error.

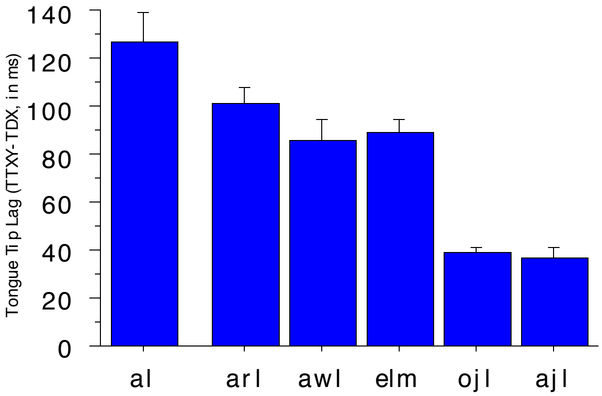

Figure 2. Tongue tip lag in postvocalic /l/ allophones.

	Mean Diff.	Crit. Diff	P-Value	
ajl, al	-89.755	21.112	<.0001	S
ajl, arl	-63.938	22.644	<.0001	S
ajl, awl	-49.026	20.549	<.0001	S
ajl, elm	-51.983	20.549	<.0001	S
ajl, ojl	-2.123	20.549	.8365	
al, arl	25.817	23.156	.0296	S
al, awl	40.729	21.112	.0003	S
al, elm	37.772	21.112	.0007	S
al, ojl	87.632	21.112	<.0001	S
arl, awl	14.912	22.644	.1919	
arl, elm	11.955	22.644	.2940	
arl, ojl	61.815	22.644	<.0001	S
awl, elm	-2.957	20.549	.7737	
awl, ojl	46.903	20.549	<.0001	S
elm, ojl	49.860	20.549	<.0001	S

Table 1. ANOVA results for Figure 2. Significant differences are marked with an ‘S’.

Using the results cited in Table 1 to interpret Figure 2, the graph can be seen to depict three statistically distinct degrees of lag: The first, a rather long duration for “hall” (around 126ms); a second, shorter one for “harl,” “howl,” “helm” (around 95ms); and a third, much shorter one for “hoyle” and “hile” (around 40ms).

2.2.3. Replication of Munhall et al. The additional measures collected to test for the ‘compensatory shortening’ effect [10] showed that the duration from acoustic VOT to achievement of TB raising for /j/ was an average of 40.16ms shorter in “hile” than in “hie” ($p < .0001$), and the duration from acoustic VOT to achievement of lip aperture minimum was an average of 22.46ms in “howl” than in “hile” (not significant: p=.1249).

3. DISCUSSION

H1 predicted that there would be no significant difference in intra-gestural lag between coda singletons and clusters. H2 predicted both the presence of this difference and that its magnitude would be the same regardless of whether the additional coda segment occurred before the /l/ (as in “harl” and “howl”) or after the /l/ (as in “helm”). The above results support H2.

The greatly reduced degree of tip lag duration seen in “hoyle” and “hile” must also be accounted for. Using Haskins’ MAVIS software as a visual aid to observe 2-dimensional dynamic reconstructions of the vocal tract during these articulations, it was apparent that the difference between these two cases and the others is the /j/—it is the only preceding element that requires the tongue dorsum to move far forward and up, opposite from where it needs to be for the following /l/. This simply means that it takes much more time for the V-gesture of the /l/ to be achieved. This is an interesting result, as once again, the global timing of the C-gesture remains unaffected by the lateness of the tautosegmental V-gesture, resulting in a greatly reduced lag between the two gestures. Interesting though this finding may be, it is tangential to the present analysis. These /j/ tokens will therefore be set aside for the remainder of the discussion.

A final prediction of the model in Figure 1b is that the degree of reduction in segment-internal lag should be equivalent in duration to the ‘compensatory shortening’ effect on the vowel. Comparing the results in 2.2.2. and 2.2.3. shows that this is true for this speaker—the across-token mean decrease resulting from the addition of one segment to the coda is about 34ms for lag, and about 31ms for compensatory shortening (the difference between these two is much less than error). It would have been preferable to have tokens of /l/ to compare directly (tokens of “hell” to compare with “helm” would have been ideal for this, but were not collected), but our model predicts C-gestures and V-gestures to behave the same whether they are components of /l/ or of /w/. The addition of this token to the stimulus set as well as several more subjects will undoubtedly help in clarifying some of these issues in future studies.

4. CONCLUSION

While the small amount of data reported by Krakow [1] initially seemed to support H1, suggesting that segments are the only units of gestural coordination needed to account for variation within syllable codas, results from the present more controlled experiment speak to the contrary. This experiment supports the model predicted by H2, indicating that both intra- and inter-segmental gestural relationships are important factors in English speech articulation. While more studies with more speakers are certainly needed before a final ruling can be made on these issues, these findings add to the growing body of literature demonstrating the importance of gestures as units of representation in accounting for complex patterns of syllable-based allophonic variation.

ACKNOWLEDGMENTS

The author is grateful to Doug Whalen for data collection and discussion, to Dani Byrd and Elisha Danford for assistance with data analysis, and to our subject. This work was supported by NIH grants DC-02717 and HD-01994 to Haskins Laboratories.

NOTES

1. See [4] for a discussion of the distinction between the terms "consonantal/vocalic gesture" and "C-gesture/V-gesture".

2. Sproat & Fujimura [2] measure tongue mid lowering for their consonantal gesture in favor of the more commonly used tongue tip fronting.

3. Gick [4] found that prevocalic and postvocalic glides pattern with liquids and nasals in terms of their gestural componency and temporal patterning.

REFERENCES

[1] Krakow, R. A. 1989. *The Articulatory Organization of Syllables: A Kinematic Analysis of Labial and Velar Gestures*. PhD dissertation, Yale University.

[2] Sproat, R. and O. Fujimura. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *Journal of Phonetics,* 21, 291-311.

[3] Browman, C. P. and L. Goldstein.1995. Gestural syllable position effects in American English. In F. Bell-Berti & L. J. Raphael (Eds.) *Producing Speech: Contemporary Issues. For Katherine Safford Harris*. Woodbury, NY: American Institute of Physics. 19-34.

[4] Gick, B. In press. Articulatory correlates of ambisyllabicity in English glides and liquids. In *Papers in Laboratory Phonology VI: Constraints on Phonetic Interpretation.* Cambridge: Cambridge University Press.

[5] Gick, B. In press. A gesture-based account of intrusive consonants in English. *Phonology.*

[6] Browman, C. P. and L. Goldstein. 1988. Some notes on syllable structure in articulatory phonology. *Phonetica*, 45, 140-155.

[7] Honorof, D. N. and Browman, C. P. 1995. The center or edge: How are consonant clusters organized with respect to the vowel? In K. Elenius and P. Branderud (Eds.) *Proceedings of the XIIIth International Congress of Phonetic Sciences*. Vol. 3, pp. 552-555. Stockholm: KTH and Stockholm University.

[8] Byrd, D. 1995. C-centers revisited. *Phonetica*, 52, 285-306.

[9] Byrd, D. 1996. A phase window framework for articulatory timing. *Phonology,* 13, 139-169.

[10] Munhall, K., C. Fowler, S. Hawkins and E. Saltzman. 1992. 'Compensatory shortening' in monosyllables of spoken English. *Journal of Phonetics*, 20, 225-239.

[11] Clements, G. N. and S. J. Keyser. 1983. *CV Phonology: A Generative Theory of the Syllable*. Cambridge, MA: MIT Press.

[12] Perkell, J., M. Cohen, M. Svirsky, M. Matthies, I. Garabieta and M. Jackson. 1992. Electromagnetic midsagittal articulometer (EMMA) systems for transducing speech articulatory movements. *Journal of the Acoustical Society of America,* 92, 3078-3096.