The Relationship Between Gestural Timing and Magnitude for American English /l/ Across Speech Tasks

Sarah Harper

University of Southern California
skharper@usc.edu

ABSTRACT

In this study, the effect of speech task on the relative timing of the two lingual gestures involved in the production of /l/ in American English is investigated using articulatory data from the Wisconsin XRMB database. Tokens of word-initial and -final /l/ were taken from Connected (e.g., sentence) and Isolated (e.g., single-word) Speech tasks in the database, with the time and degree of maximum constriction measured for each gesture in each token. Analysis of the timing lag between the intrasegmental movement extrema across conditions indicates that although their sequencing remains constant across speech tasks, exhibiting patterns consistent with the previous literature, the absolute difference in the timing of the gestures differs significantly across conditions, with timing differences closer to zero in Connected Speech than in Isolated Speech. This difference is shown to be related to systematic variation in gestural magnitude across word positions and speech tasks.

Keywords: Articulatory Phonetics, Phonetic Variation, Gestural Timing

1. INTRODUCTION

A substantial body of research on the articulation of multi-gesture consonants has consistently observed an allophonic difference in gestural sequencing patterns across different syllabic positions for these sounds. While some variation in this pattern is observed cross-linguistically [7], for North American English, the nature of this allophonic asymmetry appears to be systematic across a number of consonants containing more than one supralaryngeal gesture. Specifically, studies of /l/ [1, 17], /w/ [6], /r/ [3] and nasal stops [2, 10] have consistently found that the more anterior of the two constrictions involved in their production occupies a more peripheral position within the segment, following the more posterior gesture in syllable-final positions and either preceding or occurring synchronously with the more posterior gesture in syllable-initial positions.

Although attempts to explain why these specific positionally-dependent gestural sequencing patterns emerge have been the focus of much investigation (e.g., [1, 6, 17]), less effort has been devoted to exploring whether these observed patterns are preserved across different speech conditions. Additionally, although some of these proposed accounts rely on asymmetries in gestural magnitude across different word positions to explain the observed timing patterns [1, 3, 17], relatively little is known about the systematic relationship between relative intergestural timing and other aspects of variability in the segment's production. Some previous research has examined the effect of local prosody, such as stress [2] and the strength of an adjacent prosodic boundary [11], on intergestural timing in multi-gesture consonants; however, the effect of more global variation, such as that observed across different styles of speech, has been largely unexplored.

In this study, we extend previous research on the articulation of multi-gesture consonants by examining variation in intergestural timing and gestural magnitude, and the relationship between the two, for /l/ across speech tasks. Recent research on intraspeaker variability in speech production has observed that many acoustic and articulatory attributes of speech vary systematically across different speech tasks [4, 5], presumably as a consequence of greater temporal constraints on articulatory movement in faster, more casual speech [12, 13]. The effects of these temporal constraints on articulation have been observed both in the degree of temporal overlap exhibited by gestures belonging to adjacent segments (e.g., [8]) and the magnitude of individual speech gestures [4]. However, as there has been relatively little work examining task or rate effects on the articulatory properties of gestures in multi-gesture segments (c.f. [9]), it remains largely unknown whether the magnitude of and timing relationships between gestures in these segments exhibit similar effects of speech task.

2. METHODS

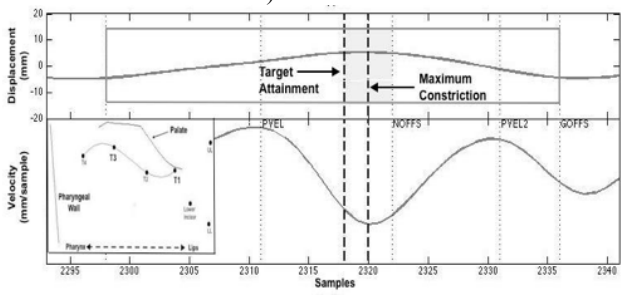
2.1. Corpus and Subjects

Articulatory data for this study was taken from recordings of 17 speakers (7 male, 10 female) in the Wisconsin x-ray Microbeam (XRMB) database [18]. The Wisconsin XRMB database contains both acoustic and kinematic articulatory data, with the

articulatory data comprising measurements of the movement of small pellets (2.5 mm) attached to the tongue, jaw, and lip (Fig. 1 inset). Pellet movements were recorded at a frame rate of 40 Hz and resampled to 145 Hz for all pellets.

A total of 1,433 tokens of word-initial and word-final /l/ were taken from three of the experimental tasks included in the corpus and separated into two conditions for analysis: a *Connected Speech* condition containing data from Sentence and Prose Passage reading tasks (962 tokens), and an *Isolated Speech* condition containing data from a Citation Word reading task (471 tokens). Tokens were taken from all available vowel contexts and, in the Connected Speech condition, from contexts where the preceding or following word (depending on the position of the target /l/ in the word) ended or began with a labial consonant.

Figure 1: Location of movement extremum in the tongue tip gesture for the initial /l/ in 'leaf' (subject JW24). Time of maximum constriction indicated in both positional (top) and velocity (bottom) time series by labelled dotted line**.** Inset: Position of tongue and jaw pellets with palate trace for the initial /l/ in 'leaf' (subject JW24). Elements used in this analysis (T1, T3, Palate, and Pharyngeal Wall) are in bold.



Although data from a total of 57 speakers is included in the database, the 17 speakers in this study were selected for analysis because they had the largest amount of analysable data – at least 25 tokens of /l/ in each word position in Connected Speech tasks, and at least 10 tokens of /l/ in each position in Isolated Speech tasks. The remaining 40 speakers did not have as much analysable data due to not recording certain tasks, tracking failures in one or more lingual pellets, or excessive amounts of /l/ vocalization, and were subsequently excluded from analysis.

2.2. Articulatory Analysis

2.2.1. Temporal Landmark Identification

Although four pellets were glued on the tongue to capture articulatory movement in the XRMB data, only the pellet closest to the tongue tip (T1) and a pellet on the tongue dorsum (T3) were used for this analysis (see Fig. 1). Temporal landmarks associated with the tongue tip and tongue dorsum movement extrema in /l/ (defined below) were identified automatically using an algorithm that calculated a 2-D velocity time series from the T1 and T3 position signals (modified from the *findgest* algorithm by Mark Tiede, Haskins Laboratories).

After using the Penn Phonetics Lab Forced Aligner [19] to perform acoustic segmentation of the XRMB data, tokens of /l/ in the appropriate conditions were located in the articulatory data by finding the articulatory frames corresponding to the acoustically-defined segment start and end points. The articulatory frame corresponding to the acoustic midpoint of each /l/ was used to identify multiple articulatory landmarks for the tongue tip and tongue dorsum gestures in the token. Of the identified landmarks, only one was used for the presented analysis: the movement extremum, defined as the velocity minimum closest to the midpoint frame for the sensor trajectory of interest (T1 or T3) (Fig. 1).

2.2.2. Lag and Constriction Degree Calculations

The measured movement extrema were used to calculate a Maximum Lag (MLag) variable used to assess timing patterns. MLag was defined as the interval between T1 and T3 movement extrema for a given token. Negative MLag values indicate that the movement extremum for the tongue tip gesture temporally precedes the movement extremum for the tongue dorsum gesture in a given token.

Additionally, T1 Aperture (T1A) and T3 Retraction (T3R) measurements were taken at the time of movement extremum for each sensor trajectory. These measurements were taken by calculating the Euclidean distance between the sensor's X-Y coordinate position and the closest point on either the palate trace taken from each speaker (for T1A) or the posterior pharyngeal wall outline approximated for each speaker (for T3R). These aperture and retraction measurements were used as an indication of the magnitude of the gesture associated with the sensor trajectory, with smaller aperture values indicative of greater gestural magnitude.

Tokens with extreme MLag, T1A or T3R values and tokens where high potential for gestural misidentification was expected were manually checked in MView (Mark Tiede, Haskins Laboratories). Tokens lacking an identifiable T1 raising gesture or tokens where it was not possible to identify a T3 velocity minimum unique from that associated with the flanking vowel were excluded from analysis.

3. RESULTS

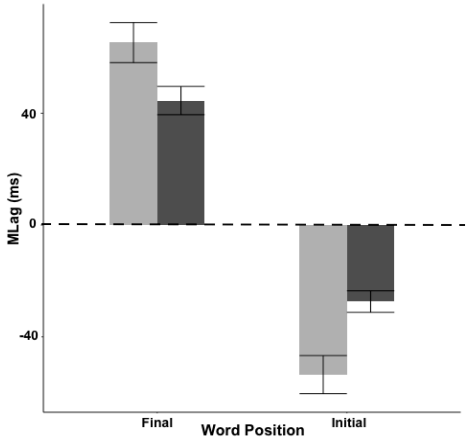
Three separate analyses were conducted using R [15]: two analyses using linear mixed effects models (LMEMs) to separately evaluate the effects of speech task and word position on gestural timing and magnitude, and one analysis using linear regression to evaluate the relationship between gestural magnitude and timing.

3.1. Gestural Timing Across Tasks and Positions

Results of the analysis of task and word position effects on gestural timing are shown in Fig. 2. An LMEM was fit on the dependent variable MLag, with Task (Connected vs. Isolated), Word Position (Initial vs. Final), and Task*Word Position as fixed factors and Subject as a random effect in both models. The results of this analysis shows that gestural timing is affected both by the position of /l/ in the word, in agreement with previous literature on gestural timing in /l/, and by the task in which /l/ is produced.

Word Position was found to significantly affect MLag ($t = -11.83$, $p < 0.0001$), with the direction of this effect mirroring that observed in the previous literature: positive lag values are observed in word-final tokens of /l/ (T3 constriction precedes T1 constriction), while negative lag values are observed in word-initial tokens of /l/ (T1 constriction precedes T3 constriction).

Figure 2: Mean MLag values across Word Positions and Tasks (light grey = Isolated Speech condition, dark grey = Connected Speech condition). Dotted line represents zero ms lag (simultaneous gestures).



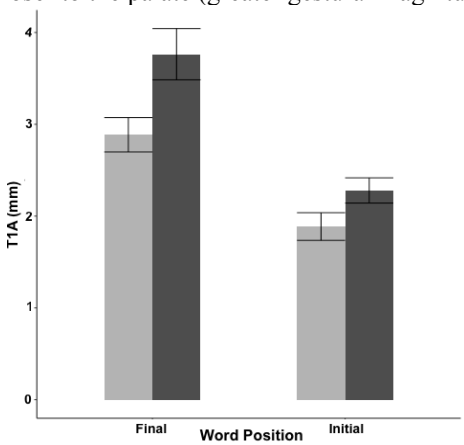
While the same gestural sequencing patterns were observed in the Isolated and Connected speech conditions, substantial differences in relative intergestural timing were observed between the two tasks (as seen in Fig. 2). For both word-initial and word-final tokens of /l/, MLag was significantly closer to zero (simultaneous T1 and T3 constrictions) in the Connected Speech condition than in the Isolated Speech condition ($t = -2.17$, $p < 0.05$). However, as is evident from the significant interaction between Task and Word Position in both models ($t = 2.62$, $p < 0.01$), the extent to which intergestural timing values vary across the two speech tasks differs as a function of word position, with a greater difference between the Isolated and Connected Speech tasks observed for Word-Initial /l/ than for Word-Final /l/.

It is worth noting that the prosodic environment in which tokens of interest occurred was partially confounded with Task, given that all tokens in the Isolated Speech condition were automatically adjacent to a prosodic boundary. To check that the effect of Task on MLag was not purely a consequence of prosodic boundary adjacency, an additional analysis was conducted comparing Word-Final /l/ in the Isolated Speech condition to the subset of all Word-Final Connected Speech tokens that occurred sentence-finally. The results of this analysis confirmed that the Task effects observed here could not be attributed to prosodic boundaries alone, as there was a significant effect of Task on MLag values within this subset of the data ($t = -17.16$, $p < 0.0001$).

3.2. Gestural Magnitude Across Tasks and Positions

For the analysis of task and word position effects on magnitude of the tongue tip and tongue dorsum gestures, separate LMEMs were fit on the dependent variables T1A and T3R, each with the same fixed and random effect structure as the models fit to the intergestural timing data. The results of the T1A analysis largely mirror those of the timing analyses, with both Word Position and Task found to significantly effect T1A (Fig. 3) (Word Position: $t = -2.88$, $p < 0.01$; Task: $t = 3.00$, $p < 0.01$).

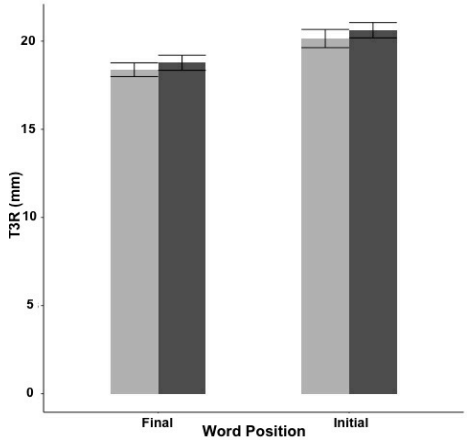
Figure 3: Mean T1A values across word positions and tasks (light grey = Isolated Speech condition, dark grey = Connected Speech condition). Smaller values indicate T1 is closer to the palate (greater gestural magnitude).



The variation in Word Position again follows that observed in previous research [6, 16], with smaller T1A values observed in Word-Initial tokens of /l/ than in Word-Final tokens, suggesting final reduction of the tongue tip gesture. The variation in T1A across the Isolated and Connected Speech conditions indicates that smaller T1A values are observed in the Isolated Speech condition than in the Connected Speech condition, as predicted based on previous research on speech task differences.

Although a significant effect of Word Position was observed on the T3R measurements ($t = 5.61$, $p < 0.0001$), with the direction of this effect following the observation in the literature that the tongue dorsum gesture in /l/ is reduced Word-Initially [1, 6], neither the main effect of Task on T3R nor the interaction between Task and Word Position were found to be significant (Task: $t = 1.75$, $p > 0.05$; Task*Position: $t = -0.363$, $p > 0.1$) (Fig. 4).

Figure 4: Mean T3R values across word positions and tasks (light grey = Isolated Speech condition, dark grey = Connected Speech condition). Smaller values indicate T3 is closer to pharyngeal wall (greater gestural magnitude).



3.3. Relationship Between Timing and Magnitude

The results of the analyses of speech task and word position effects on gestural timing and magnitude suggest that there may be a relationship between these two articulatory measurements, as the direction of the variation in timing mirrors that observed for the T1A measurements for both Word Position and Task. To test whether this apparent relationship was in fact indicative of a dependency between timing and magnitude for /l/, a linear regression model was fit on the data to test whether T1A was a significant predictor of MLag. Due to the difference in timing patterns for word-initial and word-final /l/, the absolute value of both lag measurements were used to allow more direct comparisons between /l/ in each word position.

The analysis found that T1A was a significant predictor of MLag, with a negative relationship observed such that absolute values for both lag measurements increased (became more extreme) as T1A values became smaller (larger gestural magnitude) ($\beta = -1.14$, $p < 0.01$). This finding suggests that there is a systematic, token-by-token relationship between Task and Position.

4. DISCUSSION

Overall, the results of this study confirm that both intergestural timing and gestural magnitude systematically vary across speech tasks in /l/, and additionally suggest that there is a relationship between this variation in timing and magnitude.

Measurements of both intergestural timing lag and tongue tip (T1) constriction degree indicate that the direction of this variation was the same for both measurements, with less extreme lag and constriction degree measurements observed in the Connected Speech condition than in the Isolated Speech condition. These findings mirror those from previous research on articulatory variation across speech tasks and rates, e.g. [4, 9], in that they are indicative of gestural reduction or 'undershoot' in faster, more casual connected speech styles [12]. Although similar patterns of reduction were not observed for tongue dorsum (T3) retraction in the Connected Speech condition, further research will be necessary to determine whether this is due to asymmetries in the extent to which task affects the multiple gestures in /l/ or due to other factors, such as coarticulation with surrounding vowels.

The findings of this study also indicate that there is a direct, token-by-token relationship between timing and magnitude. This observed relationship follows naturally from previous research illustrating that the duration of both articulatory gestures and their relative timing intervals are reduced as speech rate increases [8, 14].[1] As explained by models in which a set of planning oscillators or 'clocks' determine the interval over which a given articulatory gesture is active, the fact that individual articulatory gestures shorten in faster speech inherently leads to a decrease in both the magnitude of the gesture and the timing interval between adjacent gestures [16], with the extent of the durational decrease predicted to directly determine the extent of the magnitude and timing reduction. The failure to observe reduction of the tongue dorsum retraction gesture in the Connected Speech condition is again the one finding of this study that does not fit within this predicted relationship between duration, timing and magnitude, and raises the possibility that additional task-specific factors beyond the overall effect of gestural duration may be at play in this data.

5. REFERENCES

[1] Browman, C.P., Goldstein, L. 1995. Gestural syllable position effects in American English. In Bell-Berti, F.R., Lawrence, J., editors, *Producing Speech: Contemporary Issues*. AIP Press, New York., 19-33.

[2] Byrd, D., Tobin, S., Bresch, E., Narayanan, S. 2009. Timing effect of syllable structure and stress on nasals: A real-time MRI examination. *J.Phon.* 37(1), 97-110.

[3] Campbell, F., Gick, B., Wilson, I., Vatikiotis-Bateson, E. 2010. Spatial and temporal properties of gestures in North American English /R/. *Lang. Speech* 53, 49-69.

[4] Farnetani, E., Faber, A. 1992. Tongue-Jaw Coordination in Vowel Production: Isolated Words vs. Connected Speech. *Speech Communication* 11, 401-410.

[5] Ferguson, S.H., Kewley-Port, D. 2007. Talker differences in clear and conversational speech: acoustic characteristics of vowels. *J.Speech Lang. Hear. Res.* 50(5): 1241-1255.

[6] Gick, B. 2003. Articulatory correlates of ambisyllabicity in English glides and liquids. In Local, J., Ogden, R., Temple, R., editors, *Papers in Laboratory Phonology VI*. Cambridge: Cambridge University Press, 222-236.

[7] Gick, B., Campbell, F., Oh, S., Tamburri-Watt, L. 2006. Toward universals in the gestural organization of syllables: A cross-linguistic study of liquids. *J.Phon.* 34, 49-72.

[8] Hardcastle, W.J. 1985. Some phonetic and syntactic constraints on lingual coarticulation during /kl/ sequences. *Speech Communication* 4, 247-263.

[9] Harper, S., Goldstein, L., Narayanan, S. 2017. Stylistic effects on the acoustic and articulatory properites of English rhotics. Presented at *PaPE* 2017, Cologne, Germany.

[10] Krakow, R.A. 1989. *The articulatory organization of syllables: A kinematic analysis of labial and velar gestures*. Ph.D. Dissertation, Yale University.

[11] Lin, S.S. 2011. *Production and perception of prosodically varying inter-gestural timing in American English laterals*. Ph.D. Dissertation, University of Michigan.

[12] Lindblom, B. 1983. Economy of speech gestures. In MacNeilage, P.F., editor, *The production of speech*. New York: Springer-Verlag, 217-246.

[13] Moon, S.-J., Lindblom, B. 1994. Interaction between duration, context, and speaking style in English stressed vowels. *JASA* 96, 40-55.

[14] Ostry, D.J., Munhall, K.G. 1985. Control of rate and duration of speech movements. *JASA*, 77(2), 640-648.

[15] R Core Team. 2016. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.

[16] Saltzman, E., Munhall, K.G. 1989. A dynamical approach to gestural patterning in speech production. *Ecological Psychology* 1(4), 333-382.

[17] Sproat, R., Fujimura, O. 1993. Allophonic variation in English /l/ and its implications for phonetic implementation. *J.Phon.* 21, 291-311.

[18] Westbury, J. 1994. *X-ray Microbeam Speech Production Database User's Handbook*. University of Wiconsin, Madison, WI.

[19] Yuan, J., Liberman, M. 2008. Speaker identification on the SCOTUS corpus. *POMA*.

[1] In this paper we have not teased apart the extent to which observed task effects on gestural timing and magnitude can be attributed to general speech rate differences between the tasks versus other stylistic differences. This is left for future research.