THE TEMPORAL BASIS OF COMPLEX SEGMENTS

Jason A. Shaw, Karthik Durvasula & Alexei Kochetov

Yale University, Michigan State University & University of Toronto
jason.shaw@yale.edu, durvasul@msu.edu & al.kochetov@utoronto.ca

ABSTRACT

We examined the temporal basis of the phonological distinction between complex segments and segment sequences. Our working hypothesis was that the gestures of complex segments are coordinated with reference only to gesture onsets while segment sequences are coordinated with reference to the offset of the first gesture and the onset of the second. We evaluated this hypothesis using kinematic recordings of (1) palatalized labials, [p^j], in Russian, as an example of a complex segment; (2) Russian [br] sequences; and (3) labial-glide sequences in American English: [bj], [mj], [vj], [pj]. Results indicated that Russian [br] shows sequential timing, the same pattern of coordination as all four labial-glide sequences in English; the timing of Russian [p^j] was different. In line with our hypothesis, the labial and palatal gestures of Russian [p^j] were coordinated by gesture onsets. Our results are consistent with distinct modes of coordination for complex segments and segment sequences (150 words).

1. INTRODUCTION

A large number of speech segments in the world's languages are complex in that they involve coordination of multiple articulatory gestures. The aim of this paper is to examine whether there is a temporal basis to the phonological distinction between complex segments and simplex segment sequences. We formulate a specific hypothesis for the temporal basis of complex segments and evaluate it using kinematic recordings of palatalized labials in Russian and comparable segment sequences in American English.

Our working hypothesis is that the gestures of complex segments are coordinated with reference only to gesture onsets while segment sequences are coordinated with reference to the offset of the first gesture and the onset of the second. This distinction is schematized in Figure 1. Panel (a) shows complex segment timing while panel (b) shows segment sequences. Our working hypothesis is roughly equivalent (caveat below) to in-phase and anti-phase coupling in Articulatory Phonology, whereby the gestures of complex segments are in-phase and the gestures of sequences are anti-phase [4]. The caveat is that we assume that landmark-based coordination relations can be stated with consistent lags, as per the phonetic constants in the models of [7]. For example, two gestures can be timed such that the onset of movement control is synchronized with a consistent +/- lag. Possible instantiations are shown in panels (c) and (d). Panel (c) shows complex segment timing with a positive lag; panel (d) shows gestures timed as segment sequences with negative lag. Notably, owing to the influence of the +/- lag, the surface timing of (c) and (d) is identical despite being coordinated based on different articulatory landmarks.

Allowing for the theoretical possibility that gesture landmarks are coordinated with a consistent +/- lag influences our approach to hypothesis testing. From this theoretical perspective, measures of gestural overlap alone may under-determine temporal control structures, as illustrated in Figure 1(c) and (d). The same surface timing relation could be derived from differentiate combinations of coordination relations and lag values: (1) in-phase timing with a positive lag (c), anti-phase timing with a negative lag (d) or even an intermediate timing relation, e.g., "c-center" timing however derived, with no lag. However, these competing hypotheses about temporal control structure can be differentiated by considering relations between temporal intervals.

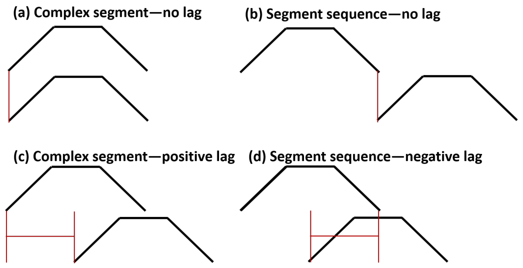
(a) Complex segment—no lag
(b) Segment sequence—no lag
(c) Complex segment—positive lag
(d) Segment sequence—negative lag

Figure 1: Hypothesized gestural coordination patterns for complex segments (a), (c) and segment sequences (b), (d)

Our strategy for differentiating hypotheses is to consider how the lag between gesture onsets varies with gesture duration. The basic strategy follows [8] in that we evaluate how temporal coordination conditions covariation between intervals. To make this concrete, consider the labial and palatal gestures of a palatalized labial consonant in which the palatal gesture begins at the midpoint of the labial gesture, a surface pattern consistent with both (c) and (d). As the duration of the labial gesture increases, we observe whether the lag between gesture onsets also increases or whether it stays the same. Covariation between labial gesture duration and intergestural onset lag is predicted only for segment sequences (b,d) and not for complex segments (a,c). The reasoning is as follows, if the gesture onsets are timed directly, even if there is a positive lag, then variation in labial gesture duration will be entirely independent of the interval between the labial onset and the palatal onset. If, on the other hand, the palatal gesture is timed to some gestural landmark later in the unfolding of the labial gesture, e.g., gesture offset as in (d), then increases in labial gesture duration will delay the onset of the palatal gesture, increasing the lag between gesture onsets.

Although we focus in this paper on labial and palatal gestures, we view the hypothesis about the temporal basis of complex segments vs. segment sequences as potentially general across the range of segment types that are "complex" under our definition. This includes, for example, aspirated stops involving coordination of laryngeal and supra-laryngeal gestures as well as nasals specified for both velum lowering and a supra-laryngeal constriction. Operationalized within the context of specific dependent variables, the hypothesis can be stated as follows:

1. Sequential segment timing: the lag between the onsets of gestures increases with the duration of the first (temporal precedence) gesture.
2. Complex segment timing: the lag between the onsets of gestures is not affected by the the duration of the gestures.

In the remainder of this paper we test the hypothesized temporal basis in three comparable data sets. The case of complex segments comes from Russian palatalized labials, which are compared to labial-glide sequences in American English from both a publicly available dataset and data from an experiment we conducted.

2. RUSSIAN DATA FROM KOCHETOV [5]

The source of our Russian data is Kochetov [5]. In that paper, Electromagnetic Articulography data was reported for one male, AK, and three female speakers, AS, NT, and DK. Here, we obtained the data from the paper, and reanalyzed a subset of the data from the three female speakers. These speakers produced a common set of materials including word-initial [p^{j}] and [br] sequences. We selected these sequences because in Russian [p^{j}] is unambiguously a complex segment and [br] is unambiguously a sequence of segments [10].

We analyzed 4-5 repetitions of four items from each speaker. Two items /tat#pʲapɨ/ 'тат пяпы' and /ta#pʲapɨ/ 'та пяпы' had [pʲ] word-initially and two items /brat#pʲatava/ 'брат пятого' and /brat#padaja/ 'брат падая' had /br/ word-initially.

The gestures from each token were parsed using the `findgest` algorithm in **mview**, a Matlab-based program for data visualization and analysis developed by Mark Tiede at Haskins Laboratories [9]. Gesture onsets and offsets were determined with reference to the velocity signal of the primary articulator: lip aperture was used to parse labial consonants; tongue blade was used to parse palatal gestures; tongue tip was used to parse the rhotic trill.

Kochetov [5] found that secondary articulations (palatal gestures) have a shorter onset lag than separate consonants. Here we evaluate whether gesture lag varies with the duration of the labial gestures. There are of course many ways to define GESTURE DURATION. Here and throughout we used the interval from gesture onset to gesture offset. However, the results remain the same qualitatively with respect to our main predictions even under other definitions of the term, including gesture duration as gesture onset to achievement of constriction or gesture duration as gesture onset to constriction release.

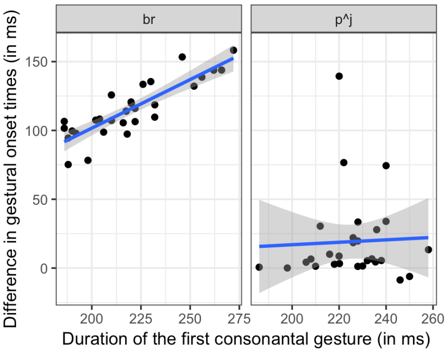

Figure 2: Correlations for the Russian data

Figure 2 shows the results for three Russian speakers, where each point represents a single measurement for a single speaker. The results show the expected positive correlation in the case of [br] for each speaker, but crucially not in the case of [p^{j}]. As labial gesture duration varies across tokens for [p^{j}], the lag between the labial gesture and palatal gesture is largely unaffected.

We excluded one of the speakers (AS) from further statistical analysis, because this speaker showed quite a lot of variability in the [p^{j}] (see outliers in Figure 2). To the rest of the gestural lag measurements, we fitted a linear mixed-effects model. Random intercepts for speakers and items were included in the model. Fixed factors were GESTURE DURATION, SEQUENCE [p^{j}, br] and the interaction term. Nested model comparison based on AIC revealed that variance explained by the interaction term justified the increased complexity it adds to the model; this model indicates that the effect of gesture duration on lag is not uniform across [p^{j}] and [br] (Table 1). SEQUENCE was coded with [p^{j}] as the reference category. The significant positive interaction indicates that the effect of G.D. on lag is greater for [br] than for [p^{j}]. In line with our predictions, gesture lag varies with the duration of the labial gesture in the case of [br], the segmental sequence, but not in the case of [p^{j}], the complex segment.

The temporal difference between the palatal glide gesture in secondary palatalization and in segmental palatal glides is consistent with our hypothesized temporal basis of complex segments vs. segmental sequences.

Table 1: Mixed effects model for the Russian TB gestures in palatal(ised) consonants [G.D. = GESTURE DURATION, Seq = SEQUENCE]

Fixed Eff.	Est.	Std. Err.	t-val	p(>|t|))
Inter.	-9.2	33.2	-0.3	0.78
G.D.	0.09	0.2	0.6	0.54
Seq	-72.8	49.3	-1.5	0.15
G.D.:Seq (br)	0.82	0.2	3.6	<0.001

3. ENGLISH DATA FROM THE X-RAY MICROBEAM DATABASE

To provide an additional point of comparison to the Russian data, we also investigated labial-palatal sequences in American English. The Wisconsin X-Ray Microbeam Speech Production Database includes data from American English speakers completing a range of speech production tasks, including word lists, sentences, and read passages [12] and is comparable to EMA data [2]. One of the word lists in the database, Task 33, includes the word 'beautiful'. Since this word begins with [bj], which is typically analyzed as a sequence of a labial stop followed by a palatal glide (or vowel) in English and not as a complex segment, it offers a useful baseline for comparison with the Russian data. There was just one token of this word per speaker, and we measured 20 speakers using the same methods as for the Russian analysis described above. Lip aperture, defined as the euclidean distance between sensors on the upper and lower lips, was used to track the labial gesture; a sensor on the tongue blade (labeled 'T2') was used to track the palatal gesture. The results are shown in Figure 3, where each point represents a single speaker. There is a positive correlation between [b] duration and the lag between gestures [b] and [j]. This is the same result we observed for Russian [br] clusters and it differs, as expected, from the Russian [p^{j}] words.

Statistical significance of the trend in Figure 3 was confirmed with a linear regression model with gesture lag as the dependent variable and GESTURE DURATION as the independent variable [$\hat{\beta}$=0.93,t=4.65,p<0.001]. The results indicate that the lag-time between the labial gesture and the TB gesture are positively correlated with the duration of the initial labial gesture.

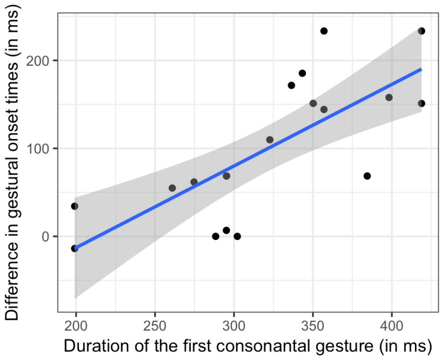

Figure 3: Correlations for the English X-ray microbeam data

4. ENGLISH DATA FROM ONGOING EXPERIMENT

Since the relevant data from the X-ray Microbeam corpus consisted of just one token of [bj] per speaker, we collected new EMA data from two more American English speakers to augment our baseline for segmental sequences. We collected 30 repetitions per subject of four words begining with labial-palatal sequences: 'muse', 'butte', 'pew', 'view'.

We used an NDI Wave electromagnetic articulograph system sampling at 100Hz to capture articulatory movement. Three sensors were placed on the sagittal midline of the tongue at the tongue tip (TT), tongue blade (TB), and tongue dorsum (TD). Additional sensors were placed on the upper and lower lip, just above and below the vermillion border. A sensor was also attached just below the lower incisor to track jaw movement. References sensors were attached to the nasion and left/right mastoids.

Stimulus display was controlled using E-prime version 2.0. The four target words were included in a list with seven filler words and displayed on a monitor in the carrier phrase ‘It’s a ______ perhaps’. Each of the eleven words was displayed once per block in random order. Participants were instructed to read the sentence at a comfortable speaking rate.

In post-processing, data was computationally corrected for head movements using the reference senseors and smoothed using Garcia’s robust smoothing algorithm [3]. Gesture identification followed the same procedure described above for the Russian data. Lip aperture was used to parse labial gestures for [m], [b], [p], and [v]; the TB sensor was used to parse the palatal gesture.

Shown below are the results for the two English speakers from our experiment, where each point represents a single measurement for a single speaker (Figure 4). The results show the expected positive correlation in the case of all the consonantal sequences tested [bj, mj, pj, vj].

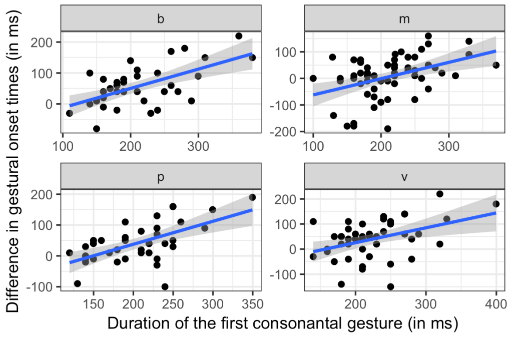

Figure 4: Correlations for the data from the English experiment

Table 2: Mixed effects model for the English TB gestures in palatal consonants [G.D. = GESTURE DURATION, FirstSeg = FIRST SEGMENT]

Fixed Eff.	Est.	Std. Err.	t-val	p(>|t|))
Inter.	-128.3	20.1	-6.4	<0.001
G.D.	0.64	0.09	7.4	<0.001
FirstSeg (b)	51.5	13.1	3.9	<0.001
FirstSeg (p)	39.6	13.1	3.0	0.003
FirstSeg (v)	25.8	12.4	2.1	0.04

We fitted a linear mixed-effects model with random intercepts for speakers and items to the gesture lag between the labial and palatal gestures. As with the preceding analyses, the main fixed effect of interest was GESTURE DURATION. We also included the FIRST SEGMENT [m, p, b, v] as a fixed factor in the analysis, as the mean lag-times might differ systematically based on the identity of this consonant. Nested model comparison based on AIC did not justify inclusion of the interaction term between GESTURE DURATION (G.D.) and FIRST SEGMENT. This indicates that the effect of labial duration on lag was uniform across segments. A summary of the fixed effects is presented in Table 2. The reference category for the FIRST SEGMENT factor was [m].

The significant positive coefficient for G.D. indicates that lag increases with increases in C_1 duration, as expected for segmental sequences. FIRST SEGMENT also had a significant effect on lag. Relative to [m], the reference category, lag was longer for each of the other consonants, [b] > [p] > [v] > [m], which may be related in part to differences in tongue position across voicing specifications [1, 11], since the spatial position of the tongue has been shown to influence CV lag [6]. Regardless of the degree of lag, however, the crucial result is that lag-times vary systematically with the duration of the first consonant in all cases tested. This is consistent with the pattern found for [br] in Russian and for the 20 speaker sample of English [bj]; crucially, only the complex segment, Russian [p^j], is different.

5. CONCLUSION

We tested a hypothesis about the temporal basis of complex segments vs. segment sequences in three data sets. For the Russian data, across multiple repetitions for each speaker of [p^j], there was no positive correlation between the duration of the labial gesture and temporal lag between labial and palatal gestures. This result conforms to our hypothesis for complex segments. The same Russian speakers’ [br] sequences had a positive correlation between labial gesture duration and the lag between labial and rhotic gestures. This conforms to our hypothesis for segment sequences. English speakers, including a large sample (n=20) producing a small number of repetitions and a small sample (n=2) producing a large number of repetitions (n=30), showed the pattern hypothesized for segment sequences for all combinations of labial-palatal gestures. Overall, the hypothesized temporal basis for complex segments vs. segmental sequences makes the correct predictions for this data and offers the potential to generalize across a wide range of complex segments, including those that are not always thought of as complex in the same way (e.g., aspirated stops) and those for which the proper characterization is otherwise contentious (e.g., prenasalized stops, affricates).

6. REFERENCES

[1] Ahn, S. 2018. The role of tongue position in laryngeal contrasts: An ultrasound study of English and Brazilian Portuguese. *Journal of Phonetics* 71, 451–467.

[2] Byrd, D., Browman, C. P., Goldstein, L., Honorof, D. 1999. Magnetometer and x-ray microbeam comparison. *Paper presented at the 14th International Congress of Phonetic Sciences.*

[3] Garcia, D. 2010. Robust smoothing of gridded data in one and higher dimensions with missing values. *Computational statistics & data analysis* 54(4), 1167–1178.

[4] Goldstein, L. H., Nam, H., Saltzman, E., Chitoran, I. 2009. Coupled oscillator planning model of speech timing and syllable structure. *Frontiers in Phonetics and Speech Science: Festschrift for Wu Zongji* 239–249.

[5] Kochetov, A. 2006. Syllable position effects and gestural organization: Evidence from russian. *Papers in Laboratory Phonology VIII* 4–2, 565–588.

[6] Shaw, J., Chen, W.-R. 2018. Variation in the spatial position of articulators influences the relative timing between consonants and vowels: evidence from cv timing in mandarin chinese. *Paper presented at the 16th Conference on Laboratory Phonology*.

[7] Shaw, J. A., Gafos, A. I. 2015. Stochastic time models of syllable structure. *PLoS One* 10(5), e0124714.

[8] Shaw, J. A., Gafos, A. I., Hoole, P., Zeroual, C. 2011. Dynamic invariance in the phonetic expression of syllable structure: a case study of moroccan arabic consonant clusters. *Phonology* 28(3), 455–490.

[9] Tiede, M. 2005. Mview: software for visualization and analysis of concurrently recorded movement data.

[10] Timberlake, A. 2004. *A reference grammar of Russian*. Cambridge University Press.

[11] Westbury, J. R. 1983. Enlargement of the supraglottal cavity and its relation to stop consonant voicing. *Journal of the Acoustical Society of America* 73, 1322–1336.

[12] Westbury, J. R. 1994. X-ray microbeam speech production database user's handbook.