

Effects of syllable onset on the timing of pitch accent in Belgrade Serbian

Robin Karlin*

Abstract. In this paper, I present the results of an acoustic study on Serbian, a pitch-accent language with sonorant-sonorant onset clusters like /mr/ and /ml/. I show that peak timing in falling accents is not affected solely by syllable onset duration, as suggested by the segmental anchoring hypothesis, but rather is determined by an interaction between syllable onset complexity and syllable onset duration, indicating a gestural representation of tone.

Keywords. tone; pitch accent; phonetics-phonology interface; Serbian

1. Introduction.

1.1 THEORETICAL BACKGROUND. There are two major schools of thought on at what point in phonological representation (or derivation) the alignment of tone contours to the segmental string occurs: autosegmental (featural) approaches, and gestural approaches. In the autosegmental approach (Pierrehumbert 1980; Bruce 1977), time is only nominally present in the underlying representation: segments and tones are represented as points that occur in a linear order, and the fact that F₀ trajectories occur at the same time as segments is represented by association lines between a tonal tier and a segmental tier (Goldsmith 1990). More detailed timing relationships are the purview of phonetic mapping rules.

One of the more widespread and received versions of phonetic mapping rules for tone is the “segmental anchoring hypothesis” (Ladd et al. 1999; Ladd & Schepman 2003) which posits specific “anchoring points” in the segmental string that the tonal targets from the underlying autosegmental representation align with. Segmental anchoring hypothesizes that both the beginning (onset) and the end (target) of pitch movements are anchored (Prieto 2011). Under this hypothesis, the pitch excursion is affected by the duration of the interval between the anchor points at the beginning and the end.

In contrast with phonological association, which is limited to the language’s specific tone-bearing unit (TBU—typically a unitary vowel, syllable, or mora), there is quite a lot of freedom in what tones can be aligned with: a pitch onset or target can be “anchored to” (occur around the same time as) various segmental landmarks, including, among others, the right or left edge of a mora (Morén & Zsiga 2006), the end of a syllable onset (or beginning of a rime, Atterer & Ladd 2004), or the end of a vowel (Welby & Loevenbruck 2006). Furthermore, there is no explicitly defined restriction on what pitch movements can be anchored to—for example, pitch movements do not actually have to be anchored (that is, phonetically aligned) with the segmental structure they are underlyingly associated with. The most common example of this is “peak delay” (Silverman & Pierrehumbert 1990; Xu 2001), where peaks of H(igh) tones phonetically occur after the syllable they are phonologically associated to; phonetic alignment occurring before the phonologically tonal syllable is less common, but has been documented in cases such as the Valjevo dialect of Serbian (Zec & Zsiga under review; Karlin 2018).

*I would like to thank Andrej Bjelaković and Biljana Čubrović for their help in collecting this data, as well as Abby Cohn, Draga Zec, Lisa Zsiga, and my fellow p-lab members for their invaluable input at all stages of this project. All remaining errors are my own. Robin Karlin, Cornell University (rpk83@cornell.edu).

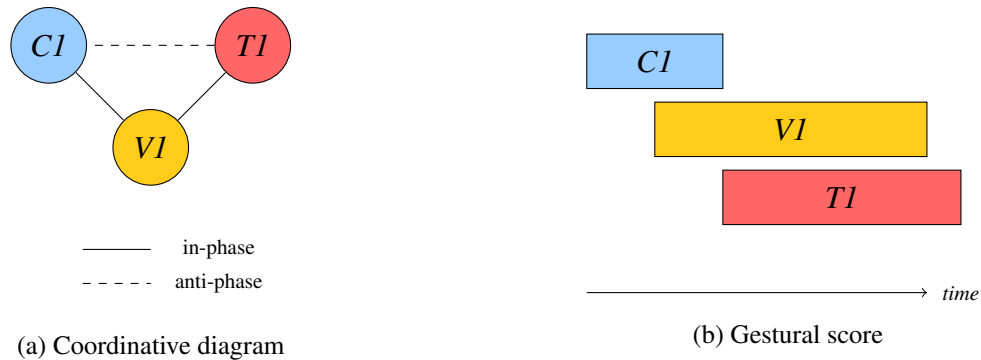


Figure 1: Coordinative diagram (left) and schematic for the resulting gestural score (right).

On the other hand, Articulatory Phonology models include timing included in the representation in great detail, including overlap of gestures both within and across “tiers”. This adheres to the central tenets of Articulatory Phonology, where articulatory gestures are the fundamental unit of contrast, rather than features, and include precise information on timing. For tone, that means that tonal gestures are represented in a coordinative structure alongside segmental gestures, and are specified with a duration.

However, there has not yet been a thorough investigation into what coordinative patterns are available for tone. Some researchers have proposed that lexical tone is coordinated in a c-center structure (Gao 2008; Yi 2014; Karlin 2014), which is a coordinative structure that has previously been investigated and proposed for consonant clusters. In the original c-center pattern, there are two consonant gestures; but in the tonal version, the second consonant gesture is replaced by a tone gesture. The consonant gesture and tone gesture are in anti-phase coordination (that is, 180deg offset from each other); however, they both are in-phase coordinated with the vowel gesture, which pulls them temporally closer to each other.

The c-center structure and resulting gestural score are illustrated in Figure 1. Here, the left edge of the box represents the onset of the gesture, which for segmental gestures, is actually before the acoustic beginning of the segment, while the right edge represents the target achievement. Since tone gestures are measured using F0 (an acoustic measure), the gestural landmarks for F0 occur at the same time as the acoustic landmarks. In a c-center structure with tone, the consonant gesture starts first, and the tone gesture starts last, while the vowel gesture is in the center of the two onsets—hence, c-center. The acoustic result of this gestural timing is that the F0 excursion starts near the acoustic start of the syllable onset.

In addition to the inclusion of time in the representation, there are a couple of main differences between mapping via segmental anchoring and underlying gestural coordination. For example, under some conceptualizations of Articulatory Phonology, only the onsets of gestures can be controlled and timed; this is unlike the segmental anchoring hypothesis, which again, argues that both the onset and the target of a pitch excursion are timed. Furthermore, working in parallel with the disagreement on target timing, there is also not full agreement in Articulatory Phonology on where gestures derive their individual timing; inherent stiffness has been proposed to play a role (see Saltzman & Munhall 1989; Browman & Goldstein 1990; Saltzman 1986 for some discussion

of the contribution of stiffness to gestural duration), but, as pitch excursions can stretch or contract depending on the amount of segmental material between anchor points (Prieto 2011; Arvaniti et al. 1998), there must be some other mechanism at work for tone.

Finally, the two theories vary in where they permit alignment. As described previously, segmental anchoring, allows many possibilities of underlying representation to phonetic alignment; in contrast, the c-center hypothesis makes very strict predictions with where the pitch onset will occur relative to the syllable onset. Given the wide range of possibilities for onsets of F0 excursions, the c-center hypothesis seems overly strict; however, the extent to which targets can be more freely timed (given some non-ballistic model of tone gesture duration) and still be coordinated in a c-center structure has not been thoroughly investigated.

1.2 LANGUAGE BACKGROUND. Serbian is an interesting case study for investigating the timing of pitch excursions, because it has lexically specified tone (“pitch-accent”), and also permits sonorant-sonorant syllable onset clusters, such as /ml/ and /mr/. These clusters allow an investigation of the timing of the onset of the pitch excursion, as well as the timing of the H(igh) target.

The term “accent” has been used in the Serbian literature to describe different phenomena in the prosodic system of Serbian, but most typically refers to a joint stress-length-pitch phenomenon of prominence. According to traditional accounts, Serbian is a pitch-accent language with four accent types that contrast on stressed syllables (Lehiste & Ivić 1986):

- Short falling (ǎ¹)
- Long falling (â)
- Short rising (à)
- Long rising (á)

The names of the accents are indicative of the bundle of prosodic characteristics that have been included in the term “accent”. The length descriptors refer to the phonological length of the vowel in the “accented” syllable: short accents have a short vowel, and long accents have a long vowel. The pitch descriptor refers, generally speaking, to the pitch contour of the “accented” syllable: falling accents start high and fall, while rising accents start low and rise. All words in Serbian, excluding function words such as clitics and prepositions, have one primary stress, and thus one accent. Some minimal pairs do exist, as in *òran* ‘disposed’ vs. *ôran* ‘plowed’ (length contrast only) or *pàra* ‘steam’ vs. *pàra* ‘dime’ (pitch contrast only); however, minimal accentual pairs are somewhat rare, and except in dictionaries, Serbian orthography does not mark any aspect of the accentual system.

I follow the representation of Serbian pitch accent presented by Inkelas & Zec (1988), who argue that Serbian accent is most fruitfully treated as a H(igh) pitch that determines the location of stress. Specifically, the H is lexically associated to some syllable, and stress is located one syllable to the left. Thus for rising accents, the stressed syllable is lower in pitch than the following syllable. Falling accents occur when the H is assigned to the first syllable—since the stress cannot move one syllable to the left, H and stress occur on the same syllable, which results in a falling contour from the stressed syllable to the following syllable (see Inkelas & Zec 1988 for further discussion of the phonological and phonetic support for this analysis).

¹In this list I am providing the accentual symbols according to the Serbian tradition; for the rest of this chapter, I will provide IPA when necessary alongside the orthography.

2. Methodology. In the following sections, I present the results of an acoustic experiment conducted on the falling accent of Belgrade Serbian. This experiment uses acoustic data to both directly examine the predictions of the segmental anchoring hypothesis, as well as a proxy for articulatory data to probe some of the predictions of Articulatory Phonology.

2.1 PARTICIPANTS. Data for Experiment 1 was collected in late summer 2016 at the Faculty of Philology at the University of Belgrade in Belgrade, Serbia. Eleven native speakers of Belgrade Serbian (ages 19 - 39; 3 male, 8 female) participated in this experiment. Included in this dataset are five speakers; the others were removed from the dataset due to frequent mispronunciations or a failure to use the correct intonation (thus obscuring the data). The Belgrade speakers were all born and raised in Belgrade, though typically one or both parents were from elsewhere. Some speakers that participated in Experiment 1 had also participated in a pilot study in March of the same year. As all participants were fluent speakers of English, written consent was provided with a consent form in English.

2.2 STIMULI AND TASK. Three real words were used as a base to form target words: *mrâve* (*/mra_Hve/*, ‘ant.PL.ACC’), *mrâmor* (*/mra_Hmor/*, ‘marble.ACC’), and *mrâmora* (*/mra_Hmora/*, ‘marble.GEN’). The syllable onset of the base words was then varied in order to produce a set of five words that were in “perfect rhyme”.² The set of onsets included two two complex onsets (*/ml/*, */mr/*), and three simple onsets (*/m/*, */l/*, */r/*). These syllable onsets provide insight on both phonological effects (i.e., onset complexity) as well as phonetic effects (variation within phonological category). For example, the */r/* in Serbian is typically realized as a tap, and as such is much shorter than the */m/*, even though they are both simple onsets; thus, a comparison of the two would probe effects caused by phonetic characteristics of onsets. The full set of stimuli is provided in Table 1.

Speakers heard a pre-recorded sentence that stated that they either had all examples of one of the base words, or that they had none of the examples of the base words. The target sentence then appeared on the screen, which disagreed with the previous statement, and the participant read it out loud with focus on the target word. Two example interactions are provided in Table 2.

Each participant read a total of 90 sentences with falling accent words, and after cleaning the dataset for production errors and errors in pitch tracking, the remaining dataset has 429 total tokens, approximately evenly distributed across all individual target words.

2.3 DATA ANALYSIS. Data was initially aligned with the Montreal Forced Aligner (McAuliffe et al. 2017), and then corrected by hand in Praat (Boersma & Weenink 2017). Due to wide variation in the production of the carriers “imamo” and “nemamo”,³ only the boundaries of the word were corrected; segments were not corrected (and segment edges are not used as landmarks in the analysis).

F0 was collected using Praat’s “Get Pitch” function, and smoothed with a bandwidth of 10 Hz. The corrected text grids and F0 tracks were then processed with a Matlab script. Pitch track landmarking was done using a Matlab script that first found pitch extrema located within certain boundaries—for example, no earlier than the acoustic beginning of the word, and no later than the

²This was necessary due to the lack of accentual marking in Serbian orthography.

³Productions for *imamo* ranged from [imamo] to [im:o] to [im::] to [iamo]; similarly, *nemamo* appeared as [ne:mamo], [nem:o], [nem::], and [neamo].

Length	Orthography	Phonology	Gloss
Falling			
Long	mrave	/ˈmra: _H ve/	“ant.ACC.PL”
	mlave	/ˈmla: _H ve/	<i>nonce</i>
	mave	/ˈma: _H ve/	<i>nonce</i>
	lave	/ˈla: _H ve/	<i>nonce</i>
	rave	/ˈra: _H ve/	<i>nonce</i>
Short	mramor	/ˈmra _H mor/	“marble”
	mlamor	/ˈmla _H mor/	<i>nonce</i>
	mamor	/ˈma _H mor/	<i>nonce</i>
	lamor	/ˈla _H mor/	<i>nonce</i>
	ramor	/ˈma _H mor/	<i>nonce</i>
	mramora	/ˈmra _H mora/	“marble.GEN.SG”
	mlamora	/ˈmla _H mora/	<i>nonce</i>
	mamora	/ˈma _H mora/	<i>nonce</i>
	lamora	/ˈla _H mora/	<i>nonce</i>
	ramora	/ˈma _H mora/	<i>nonce</i>

Table 1: Target words used in the experiment, organized by accent type.

Context:	Nemamo ni na jednoj slici bube. “We don’t have insects in any picture.”
Response:	Nije tačno! Imamo mrave . “That’s not true! We have ants.”
Context:	Imamo sve slike izlomljenih minerala. “We have all the pictures of broken minerals.”
Response:	Nije tačno! Nemamo mramora . “That’s not true! We don’t have marble.”

Table 2: Two examples of carrier sentences and their responses used in the experiment

second syllable nucleus for the F0 peak of a word with a falling accent⁴. These values were then used to bound where further landmarks could be located. An example of F0 marking on an actual token from Belgrade is given in Figure 2. There are two landmarks used for analysis:

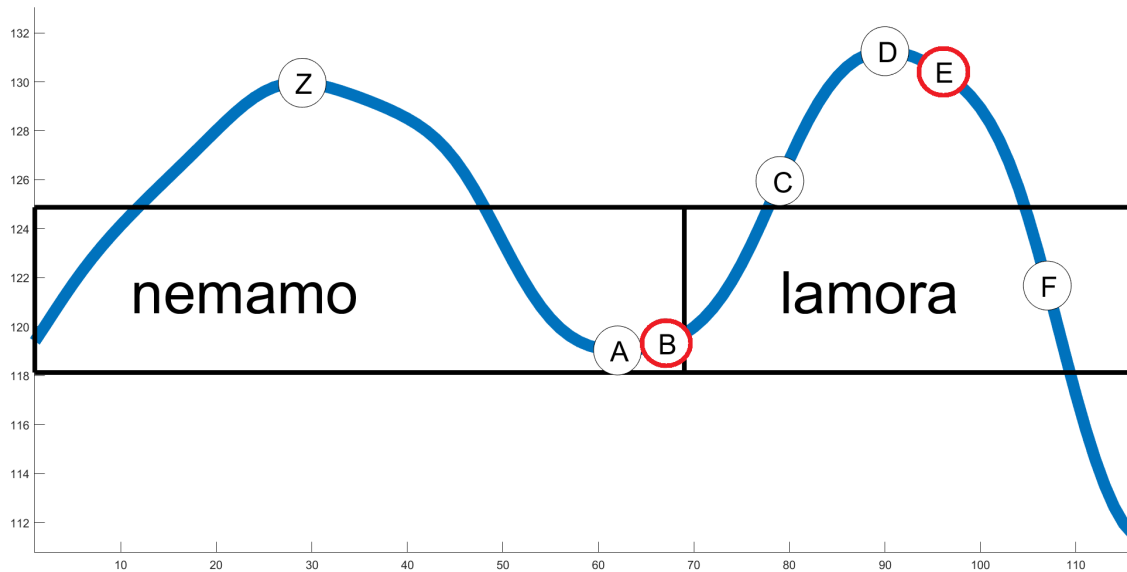


Figure 2: An example landmarked pitch trajectory.

- **Excursion onset (B):** The excursion onset was marked at the first point after the F0 valley where F0 speed reached 20% of the maximum onset speed; used as the start of the F0 excursion.
- **Peak offset (E):** The peak offset was marked at the first point after the F0 peak where F0 speed crossed the 20% maximum offset speed. This can be construed either as the release of the H F0 gesture, or as the onset of an L F0 gesture. Used as the point of target achievement.

As the absolute pitch peak is less stable and prone to small fluctuations, peak timing was measured using the F0 offset rather than the actual target F0 peak; similarly, analyses that involve the start of upward F0 movement references the F0 onset, rather than the F0 valley. The 20% speed threshold ensures that only F0 changes of sufficient magnitude are labeled.⁵ These landmarks also make it possible to directly compare plateau-like peaks and true peaks.

Statistical analyses were performed in R (R Core Team 2017), using the lme4 package (Bates et al. 2014) for linear mixed effects models. Random intercepts for `Subj` (participant) are included in all linear models presented below. Order is not included as a random effect, as the target words were presented in random order in each round. The three fixed effects of interest, from most coarse-grained to most fine-grained (indeed, continuous), are `Complexity` (complexity of the

⁴There were some adjustments made for Valjevo pitch landmarks as opposed to Belgrade pitch landmarks, as Valjevo pitch landmarks consistently occur earlier.

⁵In no cases do the results wildly change in direction or significance when using F0 peak instead of F0 offset; these speed-based landmarks simply have a less variable distribution than the extremum-based landmarks.

syllable onset, i.e. complex or simple), *Identity* (identity of the syllable onset, i.e. /r, l, m, mr, ml/), and *OnsDur* (duration of the syllable onset).

Models were built and compared incrementally, starting with the null model, which includes just participant (*Subj*) as a random effect. Models were compared with likelihood ratio tests, using the *anova* function from the *lmerTest* package (Kuznetsova et al. 2015). Homoskedasticity and normality of the residuals were assessed graphically. In this paper I will be using an α -level of 0.01; p-values below 0.05 but greater than 0.01 are considered “marginally significant” and the corresponding effects taken as a suggestion for further exploration.

3. Results.

3.1 EFFECT OF CARRIER. There was no effect of carrier (*nêmano* vs. *îmano*) on the timing of the peak offset ($p = 0.54$).

3.2 SEGMENTAL CHARACTERISTICS. All syllable onsets had mean durations that were statistically significantly different from each other: /r/ ($M = 41.8$ ms, $SD = 8.1$ ms) < /l/ ($M = 66.1$ ms, $SD = 13.3$ ms) < /m/ ($M = 89.6$ ms, $SD = 15.8$ ms) < /mr/ ($M = 125.4$ ms, $SD = 18.9$ ms) < /ml/ ($M = 136.3$ ms, $SD = 21.4$ ms), all $p < 0.0001$ using a Tukey HSD test. These differences are illustrated in Figure 3. In this figure, each onset is represented by a different color, and this color scheme will be used throughout the paper.

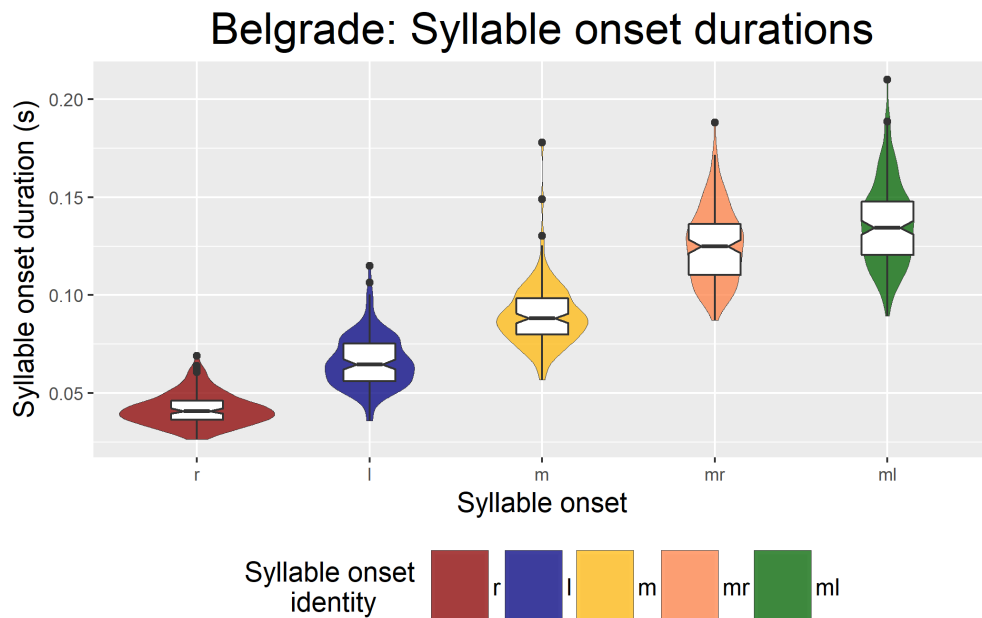


Figure 3: Durations of syllable onsets.

3.3 PEAK OFFSET TIMING. These differences in syllable onset duration are reflected in peak timing, or the timing of the peak relative to the left edge of the word (*PeakOffset*). Generally speaking, the longer the syllable onset, the later the peak: adding syllable onset duration to a linear model predicting peak timing significantly improves the model ($p < 0.0001$; see Table 3). This

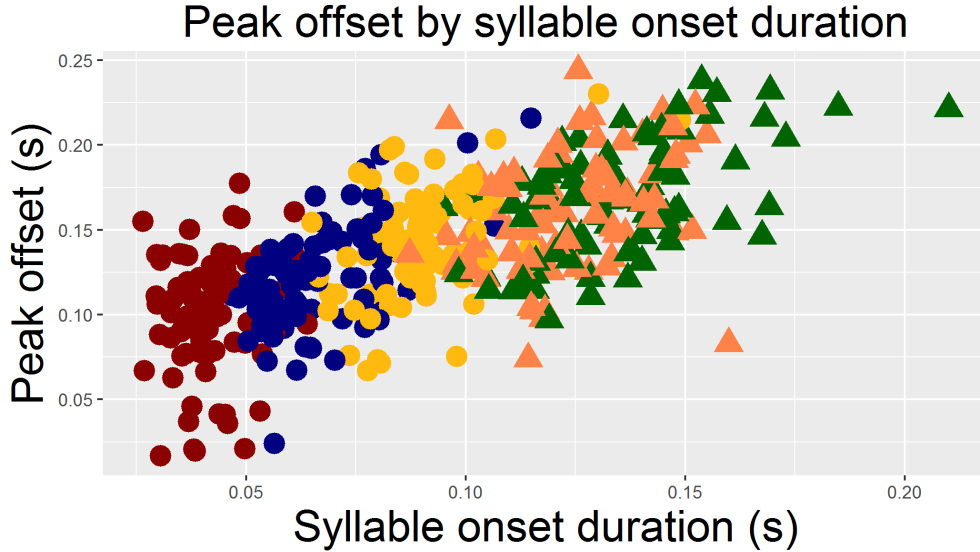


Figure 4: A scatter plot showing the relationship between syllable onset duration and peak offset timing.

is illustrated in Figure 4, where syllable onset duration is on the X axis and peak timing on the Y axis. There is a general trend of increase in lag between the left edge of the word and the peak as one moves from short syllable onsets to long syllable onsets. This is generally consistent with the segmental anchoring hypothesis, in that it appears that peaks are aligned to some point in the rime; because the rime is delayed by longer syllable onsets, so is the peak.

Model for PeakOffset	χ^2	DegF	p^\dagger
1 + (1 Subj)	—	—	—
OnsDur + (1 Subj)	293.67	1	< 0.0001**
OnsDur + Complexity + (1 Subj)	7.57	1	0.006*
OnsDur + Complexity + Complexity:OnsDur + (1 Subj)	2.83	1	0.093

[†]As compared to model immediately above

[°] < 0.05, * < 0.01, ** < 0.001

Table 3: Comparison of linear mixed effects models for PeakOffset.

However, this isn't actually a monotonic increase, as peak timing does not fully compensate for the duration of the syllable onset; the addition of Complexity to a model that already has OnsDur significantly improves the model ($\chi^2(1) = 7.57$, $p = 0.006$; see Table 3). This is best seen by looking at the timing of the peak relative to the beginning of the rime (RimeLag), instead of the beginning of the syllable. Rather than a uniform patterning, which is what is predicted given an anchoring point in the rime, there is a sharp differentiation between simple onsets and complex onsets—and there is no statistically significant difference between individual onsets in

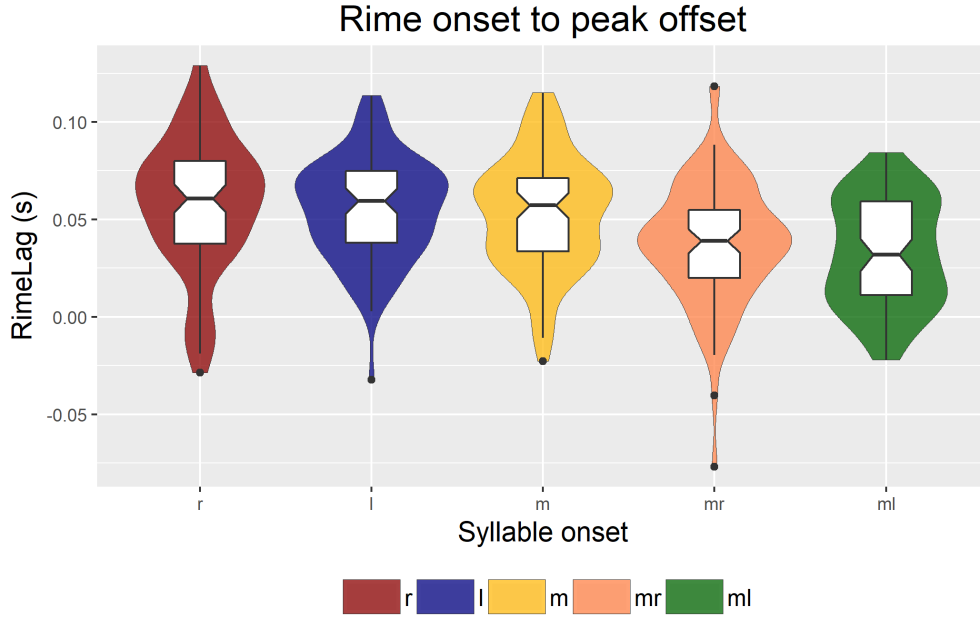


Figure 5: Distance between acoustic left edge of the rime and the peak, separated by syllable onset identity.

each complexity condition. That is, as a group, peaks for simple onsets occur later relative to the onset of the rime, and peaks for complex onsets occur earlier relative to the onset of the rime. This effect is illustrated in Figure 5, where 0 represents the start of the rime. The estimated difference between categories is 20.4 ms (SE = 2.7 ms), where peaks occur closer to the left edge of the rime for complex onsets.

Thus, while peak timing clearly showed effects of the phonetic characteristics of the syllable onset, there is also an effect of the phonological characteristics of the syllable onset. In comparing models (see Table 4), `OnsDur` does provide a better fit than the null model, which is not surprising, because duration and complexity are correlated. However, adding `Complexity` as a fixed effect to a model that already has syllable onset duration significantly improves the model, at $p = 0.006$,

Model for RimeLag	χ^2	DegF	p^\dagger
<code>OnsDur + (1 Subj)</code>	—	—	—
<code>OnsDur + Complexity + (1 Subj)</code>	7.57	1	0.006*
<code>Complexity + (1 Subj)</code>	—	—	—
<code>Complexity + OnsDur + (1 Subj)</code>	2.96	1	0.085

[†]As compared to model immediately above

[°] < 0.05, * < 0.01, ** < 0.001

Table 4: Comparison of linear mixed effects models for RimeLag

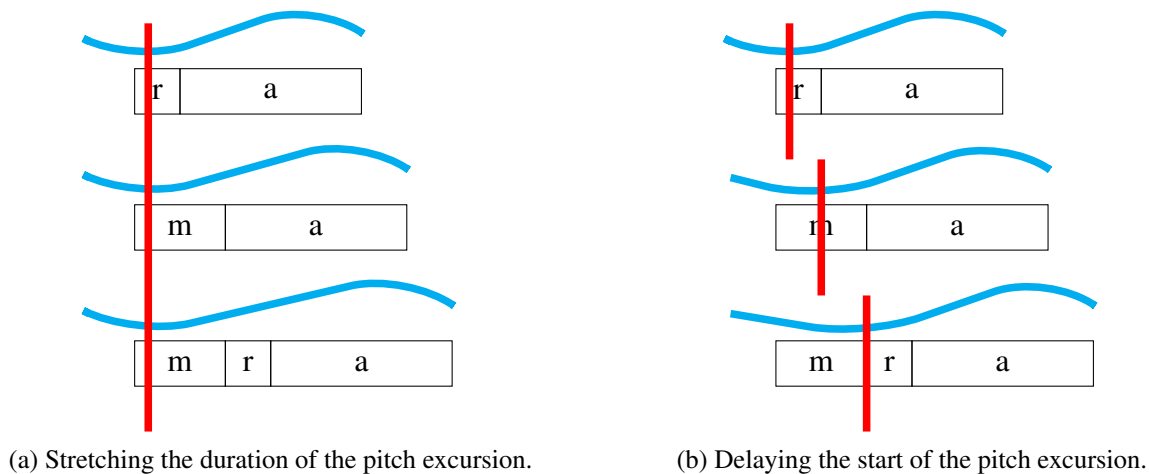


Figure 6: Schematics of the two possible ways to change the timing of peak offset.

while starting with *Complexity* first, and then adding the phonetic duration of the syllable onset, does not improve the model, at $p = 0.085$. Thus, *Complexity* as a phonological category is the major factor that determines where the peak is relative to the beginning of the rime.

3.4 START OF PITCH EXCURSION. In order to achieve a later peak, there are two options: first, as predicted by the segmental anchoring hypothesis, one can lengthen the duration of the pitch excursion. This possibility is illustrated in Figure 6a—the ends are anchored to specific points, and the excursion gets longer as the intervening segmental material increases in duration.

Second, one can keep the same excursion duration but simply delay the start, which is predicted by one particular conceptualization of the c-center hypothesis—specifically, the one that takes gestural onset duration into account, rather than just onset timing. This possibility is illustrated in Figure 6b, where the pitch excursion starts later and later relative to the beginning of the word, but the excursion duration remains the same.

Interestingly, it turns out that in the Belgrade dialect, both strategies are used. However, they're affected by *Complexity* in different ways. First, the longer the syllable onset, the longer the pitch excursion. This is a fairly monotonic increase, as shown in Figure 7, with syllable onset duration on the X axis and excursion duration on the Y axis. A comparison of mixed linear effects models shows that the addition of *Complexity* does not improve the fit of a model that already has *OnsDur*, and there is no interaction between syllable onset duration and complexity (see Table 5).

However, the beginning of the pitch excursion is not anchored to an edge. Rather, the longer the syllable onset, the later the pitch excursion starts (*OnsDur* significantly improves the fit over the null model; $p < 0.0001$). There is also an additional effect of *Complexity* ($p < 0.0001$); however, there is no interaction between *OnsDur* and *Complexity* ($p = 0.64$), which indicates that within each phonological category, changes in onset duration have the same effect (see 6).

Essentially, the consonants are displacing away from the tone gesture in both directions, rather than just to the left or just to the right. This is illustrated in Figure 8, where 0 represents

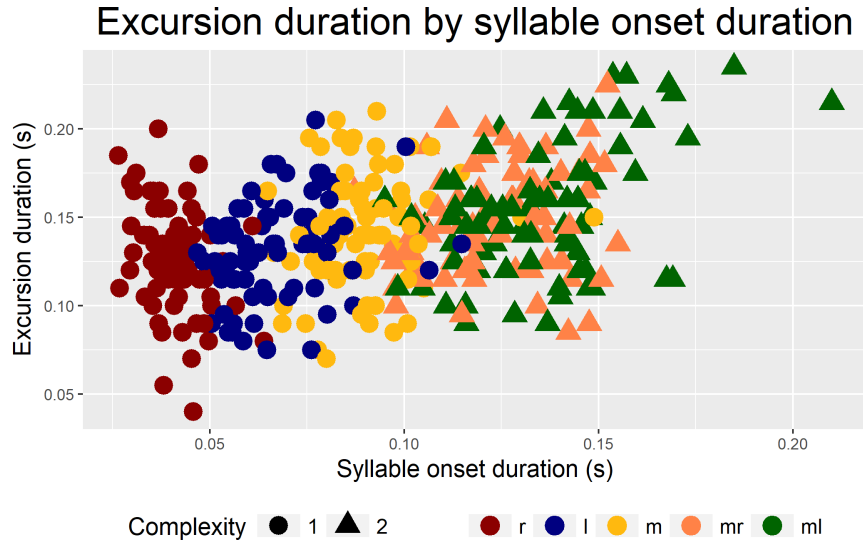


Figure 7: Scatter plot showing the relationship between syllable onset duration and excursion duration.

Model for ExcurDur	χ^2	DegF	p^\dagger
1 + (1 Subj)	—	—	—
OnsDur + (1 Subj)	73.40	1	<0.0001**
OnsDur + Complexity + (1 Subj)	0.01	1	0.91
OnsDur + Complexity + Complexity:OnsDur + (1 Subj)	1.72	1	0.19
† As compared to model immediately above	$^\circ < 0.05, * < 0.01, ** < 0.001$		

Table 5: Comparison of nested models for ExcurDur.

Model for ExcurStart	χ^2	DegF	p^\dagger
1 + (1 Subj)	—	—	—
OnsDur + (1 Subj)	215.67	1	< 0.0001**
OnsDur + Complexity + (1 Subj)	20.11	1	< 0.0001**
OnsDur + Complexity + Complexity:OnsDur + (1 Subj)	0.22	1	0.64
† As compared to model immediately above	$^\circ < 0.05, * < 0.01, ** < 0.001$		

Table 6: Comparison of nested models for ExcurStart.

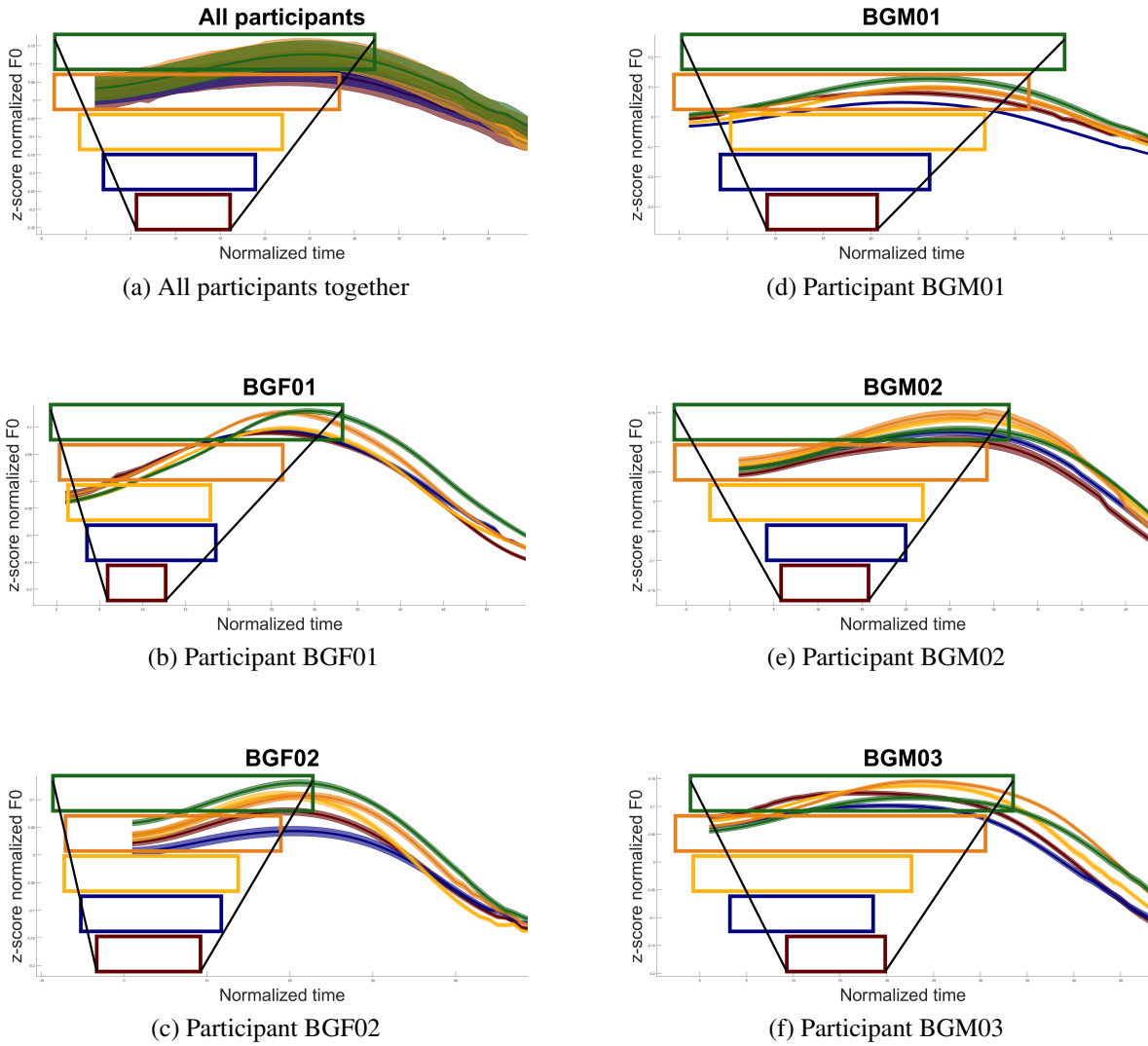


Figure 8: Z-score and time-normalized F0 trajectories (with standard error shading), Belgrade dialect. The acoustic edges of the syllable onset marked by boxes. Zero on the x axis represents the start of the pitch excursion. Both syllable onsets and F0 trajectories are color-coded in the same color scheme as used elsewhere; syllable onsets are arranged in order from /r/ at the bottom to /ml/ at the top.

the beginning of the pitch excursion, not the beginning of the word. Figure 8a contains pitch trajectories and acoustic syllable onset edges for all participants together, while the remaining figures illustrate each participant's patterns separately. For each participant, the edges of the syllable onset displace in both directions from the start of the pitch excursion (highlighted by the black line from /r/ to /ml/). This is reminiscent of the motivation for the name of the “c-center” effect—that is, the center, rather than the edges, of the consonant gestures (here, using acoustic landmarks as proxy for gestures) remains constantly timed to some reference landmark.

4. Discussion. This data shows that the timing of pitch excursions is affected by both phonetic and phonological characteristics of the syllable. On the phonetic side, we saw that there is an effect of the phonetic duration of the syllable onset on the timing of the peak relative to the acoustic left edge of the word. However, we also saw that this is not due to a straightforward anchoring point in the rime, as complexity as a phonological characteristic was the main predictor of temporal distance between the left edge of the rime and the peak. Similarly, we saw that there were straightforward phonetic effects on the duration of the pitch excursion. However, there was an additional effect of complexity on the timing of the beginning of pitch excursion. The result of this combination of factors is a pitch excursion that stretches somewhat with increased intervening segmental material, but in fact serves as some kind of centering gesture.

These findings also suggest that the predictions of segmental anchoring do not hold, and that a gestural model of tone, which necessarily includes duration and timing relationships, may predict more of the data. However, a more specific theory of where tone gestures get their duration is necessary. Previous acoustic-based literature has argued rather thoroughly against a “ballistic” trajectory of tone, instead allowing for the elastic pitch excursions predicted by segmental anchoring. However, Articulatory Phonology is not consistent with its predictions of target timing, which is of course what is being measured when looking at peak timing.

These results leave quite a few open questions for future work. First, in order to specifically test the hypotheses of the Articulatory Phonology approach regarding articulatory timing, an experiment using EMA data is necessary. Even though it is possible to speculate on the gestural organization of consonants and tone based on acoustic data, it would be ideal to actually have the articulatory data.

Second, in staying with acoustic data as a proxy for articulatory data, there are additional effects to probe that would reveal further pieces of the puzzle. One open question is the contribution of duration vs. complexity—specifically, if the apparent effect of complexity on the timing of the peak relative to the beginning of the rime is due to phonological complexity, or if it's actually just some threshold of duration that the complex onsets happen to surpass. One way to look at this would be to find a long simple onset and a short complex onset, such that they overlap in the middle of the duration spectrum, and see if the effects of complexity and phonetic duration still hold.

Relatedly, it would also be interesting to look at the effect of even more complex syllable onsets, such as three-consonant clusters. Such clusters could serve to test if the beginnings of pitch excursions are anchored to some landmark (either acoustic or gestural) on the first consonant, or if it is the full set of consonants that determine timing. Unfortunately, there are no triple sonorant clusters in Serbian, which could prevent an analysis of the timing of the pitch excursion onset, depending on the specific sequence used; however, peak timing could still be examined.

Finally, other dialects of Serbian also offer interesting insights on the possible ways to time tone. For my forthcoming dissertation (Karlin 2018), I also collected this data from speakers from Valjevo, which is located approximately 90 kilometers to the southwest of Belgrade. This dialect has much earlier pitch peaks (Zec & Zsiga under review), which nevertheless exhibit the same effects of syllable onset duration on the timing of pitch peaks. However, the Valjevo dialect does not use the same strategy to create this delay: the beginning of the pitch excursion does not move at all, but rather only the duration of the pitch excursion changes. This is interesting because it lends credence to the idea that peak timing can be achieved in a variety of ways, even between two minimally different dialects.

References

- Arvaniti, Amalia, D Robert Ladd & Ineke Mennen. 1998. Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of phonetics* 26(1). 3–25.
- Atterer, Michaela & D Robert Ladd. 2004. On the phonetics and phonology of “segmental anchoring” of F0: evidence from German. *Journal of Phonetics* 32(2). 177–197.
- Bates, Douglas, Martin Maechler, Ben Bolker, Steven Walker et al. 2014. lme4: Linear mixed-effects models using Eigen and s4. *R package version* 1(7). 1–23.
- Boersma, Paul & David Weenink. 2017. Praat: doing phonetics by computer. <http://www.fon.hum.uva.nl/praat/>.
- Browman, Catherine P & Louis Goldstein. 1990. Gestural specification using dynamically-defined articulatory structures. *Status report on speech research* 95.
- Bruce, Gösta. 1977. *Swedish word accents in sentence perspective*, vol. 12. Lund University.
- Gao, Man. 2008. *Mandarin tones: An articulatory phonology account*: Yale University dissertation.
- Goldsmith, John A. 1990. *Autosegmental and metrical phonology*, vol. 1. Basil Blackwell.
- Inkelas, Sharon & Draga Zec. 1988. Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language* 227–248.
- Karlin, Robin. 2014. The articulatory TBU: Gestural coordination of lexical tone in Thai. *Cornell Working Papers in Phonetics and Phonology*.
- Karlin, Robin. 2018. *The role of the moraic tbu in tone timing: Acoustic and articulatory evidence*: Cornell University dissertation.
- Kuznetsova, Alexandra, Per Bruun Brockhoff & Rune Haubo Bojesen Christensen. 2015. Package lmerTest. *R package version* 2(0).
- Ladd, D Robert, Dan Faulkner, Hanneke Faulkner & Astrid Schepman. 1999. Constant “segmental anchoring” of F0 movements under changes in speech rate. *The Journal of the Acoustical Society of America* 106(3 Pt 1). 1543–1554.
- Ladd, D Robert & Astrid Schepman. 2003. “Sagging transitions” between high pitch accents in English: experimental evidence. *Journal of phonetics* 31(1). 81–112.
- Lehiste, Ilse & Pavle Ivić. 1986. *Word and sentence prosody in Serbocroatian*. MIT Press.

- McAuliffe, Michael, Michaela Socolof, Sarah Mihuc, Michael Wagner & Morgan Sonderegger. 2017. Montreal forced aligner. <http://montrealcorpustools.github.io/Montreal-Forced-Aligner/>.
- Morén, Bruce & Elizabeth Zsiga. 2006. The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory* 24(1). 113–178.
- Pierrehumbert, Janet Breckenridge. 1980. *The phonology and phonetics of English intonation*: Massachusetts Institute of Technology dissertation.
- Prieto, Pilar. 2011. Tonal alignment. In Marc van Oostendorp, Colin J Ewen, Elizabeth V Hume & Keren Rice (eds.), *The Blackwell companion to phonology*, chap. 50, 1185–1203. Blackwell Publishing.
- R Core Team. 2017. *R: A language and environment for statistical computing*. R Foundation for Statistical Computing Vienna, Austria. <https://www.R-project.org/>.
- Saltzman, Elliot. 1986. Task dynamic coordination of the speech articulators: A preliminary model. *Status report on speech research* 9.
- Saltzman, Elliot L & Kevin G Munhall. 1989. A dynamical approach to gestural patterning in speech production. *Ecological psychology* 1(4). 333–382.
- Silverman, Kim & Janet Pierrehumbert. 1990. The timing of prenuclear high accents in English. *Papers in laboratory phonology I* 72–106.
- Welby, Pauline & Hélène Loevenbruck. 2006. Anchored down in anchorage: Syllable structure and segmental anchoring in french. *Italian Journal of Linguistics/Rivista di linguistica* 18. 74–124.
- Xu, Yi. 2001. Fundamental frequency peak delay in Mandarin. *Phonetica* 58(1-2). 26–52.
- Yi, Hao. 2014. A gestural account of Mandarin tone sandhi. *The Journal of the Acoustical Society of America* 136(4). 2144–2144.
- Zec, Draga & Elizabeth Zsiga. under review. Phonological consistency and phonetic variation: Tonal alignment in three Neo-Štokavian dialects .