Laboratory Phonology
Journal of the Association for Laboratory Phonology

Karlin, R. 2022. Expanding the gestural model of lexical tone: Evidence from two dialects of Serbian. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 13(1): 3, pp. 1–43. DOI: https://doi.org/10.16995/labphon.6443

Open Library of Humanities

Expanding the gestural model of lexical tone: Evidence from two dialects of Serbian

Robin Karlin, Waisman Center, University of Wisconsin – Madison, Madison, WI, USA, rkarlin@wisc.edu

There is mounting evidence suggesting that temporal information is necessary in representations of lexical tone. Gestural models of tone provide a natural entry point to linking abstract association with physical realization, but remain underdeveloped. We present the results of two acoustic production studies on two dialects of Serbian, a lexical pitch accent language. In the Belgrade dialect, pitch accents are aligned relatively late in the tone-bearing unit, while in the Valjevo dialect, pitch accents are phonetically retracted, sometimes into the preceding syllable. We varied the phonetic duration of syllable onsets of candidate tone-bearing units in falling (experiment 1) and rising (experiment 2) pitch accents, and measured the effects on the timing of F0 excursions. Consistent interactions between F0 excursions and the segmental content indicate that the phonological system of abstract tone association is the same in both dialects, despite differences in temporal alignment. We argue that this apparent mismatch between association and alignment can be expressed straightforwardly in the Articulatory Phonology framework by allowing tone gestures to coordinate with other gestures in all the ways that segmental gestures can, rather than restricting tone to c-center coordination.

1. Introduction

The question of how to generate the phonetic timing of tone contours from a temporally impoverished phonological representation has long been a topic of debate in phonological theory (e.g., Arvaniti, Ladd, & Mennen, 2006b; Laniran, 1992; Myers, 1999; Prieto, 2011; Xu & Wang, 2001). In gestural models, it is impossible to represent any atom of speech without including a temporal component, which directly specifies the phonetic result. However, gestural models of tone are still in their infancy, due in large part to the small number of studies on just a handful of tonal languages.

We present data from two acoustic studies on two varieties of Serbian, one with late alignment of accentual peaks and one with early alignment, and show that they share an abstract phonological system but have distinct phonetic realizations. Articulatory Phonology is well equipped to capture these facts: Both abstract phonological relationships and phonetic realization can be derived from coordinative relationships, where the presence of a coordinative relationship between two gestures indicates phonological association and the precise nature of that coordinative relationship predicts the phonetics. We expand on the gestural model of tone and show that tonal coordination is more diverse than what has been previously hypothesized, and propose that tone gestures can be coordinated in all the same manners as segmental gestures.

1.1. Timing and the tone-bearing unit in models of tone

In Autosegmental theories of tone, tones reside in a separate tier from segments, and are 'associated' to the segmental content of a word (Goldsmith, 1990). The phonological association of a tone to a segment (or segmental structure) is not underlying, but rather part of the derivation from underlying to surface form. The unit that tone targets can be associated to is language-specific, and is referred to as the *tone-bearing unit* (henceforth, TBU); the vowel, mora, and syllable have been proposed for various languages (DiCanio, Amith, & García, 2014; Ladd, 2008; Morén & Zsiga, 2006; Yip, 2002, inter alia). Phonological distribution plays a major role in determining the TBU for a particular language. For example, the four Mandarin tones can occur on any open syllable, regardless of the number of tone targets in the lexical tone, indicating that the syllable is the TBU in Mandarin (Yip, 1989). In contrast, Thai simple tones H(igh) and L(ow) can occur on words with just one sonorant mora, while contour tones (i.e., HL and LH tones) can only occur on words with two sonorant moras, suggesting that the mora is the TBU in Thai (Morén & Zsiga, 2006).

In addition, phonetic information can also play a role in the analysis of a given language's TBU. Specifically, the association of a tone to a TBU implies some sort of overlap in time (Sagey, 1986), and the (acoustic) co-occurrence of F0 contours associated with tone targets and segmental units has been used as evidence for phonological association between the two (see e.g., Kanerva, 1990;

Myers, 1999, for discussion on Chichewa alignment versus association). The precise nature of that overlap is not included in autosegmental representations; that is, an HL tone does not predict a particular alignment of the H target with a particular point of the segmental TBU. Instead, timing details are relegated to phonetic mapping rules. For example, it has been hypothesized that a given language may align their tones to the left or right edge of the TBU. This ordinarily assumes that the target of a tone is the relevant landmark for alignment; targets can be achieved at the left or right edge of the TBU (e.g., Morén & Zsiga, 2006; Prieto, D'Imperio, & Fivela, 2005). In a falling (HL) tone, these alignments would produce early and late peaks, respectively (see **Figure 1** for illustration). These mapping rules are idiosyncratic to individual languages and thus it is difficult—if not impossible—to predict the phonetic form from underlying tones.

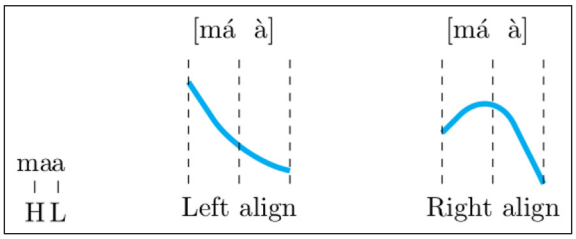


Figure 1: Examples of two phonologically identical HL tones that produce two phonetically distinct contours, with left alignment (left) and right alignment (right).

However, it is not uncommon for pitch targets to occur outside their TBU, thus calling into question the assumption that phonetic alignment of tone targets and phonological association of tones are tightly paired (see discussion in Arvaniti, Ladd, & Mennen, 2000; Ladd, 2008; Myers, 1999; Prieto, 2011; Roettger, 2017, inter alia). Late pitch targets have been attributed to so-called 'peak delay,' a phenomenon where pitch targets are achieved after the TBU they are associated to (Xu, 2001). Peak delay has been documented in several languages, including lexical tone languages (see e.g., Myers, 1999 for Chichewa; Xu, 2001 for Mandarin; Morén & Zsiga, 2006 for Thai), as well as in intonation-only languages (see e.g., Silverman & Pierrehumbert, 1990 for English; Arvaniti, Ladd, & Mennen, 1998 for Greek). Peak delay is so ubiquitous that it is encoded in the PENTA model of speech melody (Xu, 2005; Xu & Wang, 2001), where the model predicts that all pitch turning points occur after the end of the TBU due to inertia.

There are comparatively fewer cases of early tonal target achievement, but there are documented cases. In tonal crowding, peaks occur early relative to the TBU; however, this term specifically describes circumstances where pitch targets are shifted to the left due to tonal pressures from the right, such as the addition of a boundary tone (Arvaniti, Ladd, & Mennen, 2006a). This is distinct from early peaks that occur without additional time pressure, such as those described by Zec and Zsiga (to appear) in Valjevo Serbian (for more detail, see Section 1.3), and Bruce (1977) for

Stockholm Swedish. In Swedish, both accent I and accent II are characterized by an H target (F0 peak); the two accents are distinguished by timing, where accent I peaks occur earlier relative to the stressed syllable than accent II peaks. The case of Swedish accent I is quite extreme, as accent I "occurs as early as in the pre-stress syllable, even if this syllable belongs to a preceding word" (p. 46; 49)—that is, the accentual peak can occur so early that it leaves the phonological word entirely. Thus, not only is it difficult to predict the phonetic form from underlying form without phonetic mapping rules, it is also difficult to accurately predict phonological relationships from acoustic alignment alone. Importantly, it is unclear if the timing of tone targets is the main metric to consider in this relationship.

In recent years, it has become increasingly clear that temporal information should be included in phonological representation. Contrastive alignment within a syllable was hypothesized to not be possible in lexical tone languages (Hyman, 1988; Odden, 1995, as cited in Remijsen & Ayoker, 2014). However, recent studies on Shilluk (Remijsen & Ayoker, 2014), Dinka (Remijsen, 2013), and Yoloxóchitl Mixtec (DiCanio et al., 2014) have provided examples of such contrastive alignment, suggesting that more detailed timing information is necessary for tonal representation, at the very least in those cases.

Further evidence for the necessity of timing information in tonal representation comes from systems of intonation. Contrasts in alignment in intonation have led to the introduction of the star convention in Autosegmental-Metrical theory (Ladd, 2008; Pierrehumbert, 1980). Pierrehumbert (1980) described two distinct rises in English: an early rise, denoted as L+H*, and a late rise, denoted as L*+H. In this use of the star convention, the star indicates which tone is associated to the stressed syllable. Thus, the early rise is created by a leading L tone with the transition to the H on the stressed syllable, while in the late rise, the L is aligned to the stressed syllable, and the transition to H trails after. This has been adopted in some analyses of pitch accent, which frequently interacts with stress; for example, Smiljanić (2002) proposed two pitch accents for Serbian, L+H* and L*+H, which represent early (falling accent) and late (rising accent) F0 rises, respectively; Myrberg (2010) (after Bruce, 1977) addressed the alignment differences in the Swedish accentual system using this convention as well, proposing HL* for accent I and H*L for accent II. In proposing a bitonal HL* pitch accent, this model accounts for a peak seemingly occurring outside the domain of its TBU: The L of the HL* pitch accent is anchored to the stressed syllable in accent I, which leaves H to occur earlier. However, the star convention is currently used for a wide variety of other purposes in Autosegmental-Metrical accounts, including status as a pitch accent, rather than a phrase accent (Ladd, 2008); non-contrastive phonetic alignment (cf. discussions in e.g., Atterer & Ladd, 2004; Prieto, 2011); as well as phonological association without phonetic alignment (see e.g., the proposal in Arvaniti et al., 2000).

Furthermore, it is unclear that targets are the only crucial point to consider in the relationship between TBUs and tone. Tone targets have been preferred as points of alignment in accounts of

lexical tone, but there is accumulating evidence that the onsets of F0 movement are reliably timed (within a particular tone of a given language) as well. For example, the early versus late rises in English described by Pierrehumbert (1980) include early versus late onsets of F0 excursion, rather than early versus late peak timing alone. Speakers' control of this timing was verified in the experiment presented by Pierrehumbert and Steele (1989) where speakers imitated a continuum from L + H* to L* + H and shifted the preceding low and the peak together in time, rather than only shifting the H peak. The well-documented phenomenon of 'segmental anchoring' also indicates consistent timing of F0 onsets, describing where an F0 excursion is seemingly 'anchored' to segmental landmarks both at its onset and at its peak (Arvaniti et al., 1998; Ladd, Faulkner, Faulkner, & Schepman, 1999; Prieto & Torreira, 2007), and stretches to accommodate additional segmental material between the two anchors. Articulatory work on Mandarin and Thai tone also suggests consistent timing of the onsets of F0 excursions (e.g., Gao, 2008; Karlin, 2014; Yi, 2014, see Section 1.2 for more detail).

1.2. Gestural approaches to tone

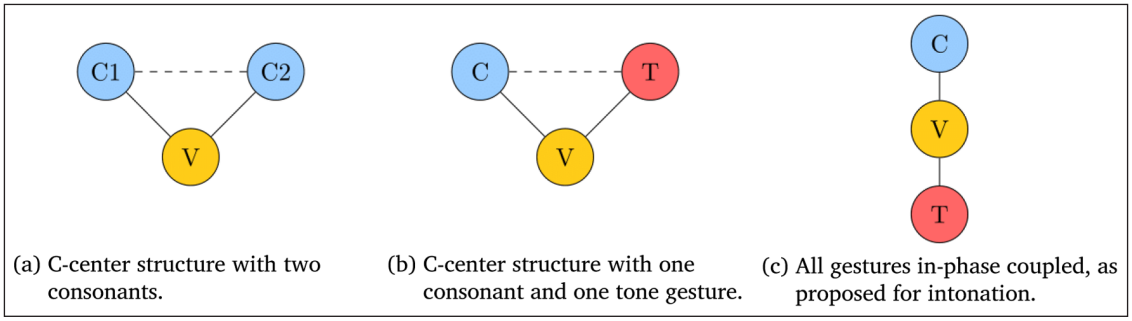
Gestures are, importantly, specified in both space and time, and provide natural points for alignment. Recently, tone has fruitfully been treated as a gesture (e.g., Gao, 2008; Karlin, 2014; Mücke, Nam, Hermes, & Goldstein, 2011; Prieto, Mücke, Becker, & Grice, 2007; Yi, 2014, 2017), where tone gestures use a similar convention to those in Autosegmental theory, i.e., there are High and Low targets for tones (among others), rather than target contours. The space dimension is mapped in F0 space, as that is the linguistically relevant variable, rather than some laryngeal posture. Thus, a High tone gesture would have some (relatively) high F0 target, and a Low tone gesture would have some low F0 target, similar to a target of closed lips (–2 mm lip aperture) for the bilabial closure in an /m/, as opposed to specific positions of each lip.

Timing information comes from two sources, and the acoustic linearization can then be derived from the combined spatial and temporal details included in a gestural constellation. First, any individual gesture is specified with some stiffness, which is an abstraction of duration analogous to spring stiffness (Browman & Goldstein, 1986; Fuchs, Perrier, & Hartinger, 2011). Gestures with high stiffness values have a shorter duration, while gestures with low stiffness values have longer durations. The second source of timing is the relative timing of gestures to each other, which is achieved through the coordination of two or more gestures into a 'constellation.'

In more restrictive versions of AP, such as the coupled oscillator model (Goldstein & Pouplier, 2014; Nam, Goldstein, & Saltzman, 2009) there are two basic modes of coordination: in-phase coordination, which is when two gestures start at the same time, and anti-phase coordination, which is when two gestures are 180° out of phase with each other. In this version of the theory, only gestural onsets can be coordinated; all other points fall out naturally based on various temporal properties of the gesture. To date, most work on tone in the AP framework

has focused on the coordination of gestural onsets. The most explored gestural coordination for tone is the 'c-center structure,' first described for consonant clusters in certain languages (Browman & Goldstein, 1988; Goldstein, Chitoran, & Selkirk, 2007; Marin, 2013; Marin & Pouplier, 2010). This structure utilizes in-phase and anti-phase coordination of gestural onsets; there are two consonant closure gestures in-phase coordinated with a vowel, and anti-phase with each other (illustrated in **Figure 2a**). In this case, the two consonantal gestures repel each other, since they want to be at opposite sides of their cycle, but they are simultaneously pulled together because they both want to be in-phase with the vowel. The result is that the onset of the vocalic gesture is coordinated with the gestural center of the consonantal gestures, instead of with the onset of one consonant gesture, as is typical in CV syllables (Browman & Goldstein, 1988). This pattern has been found in tone languages as well, such as Mandarin (Gao, 2008; Yi, 2014, 2017; Zhang, Geissler, & Shaw, 2019) and the first tone gesture in Thai contour tones (Karlin, 2014). In these tone languages, instead of a second consonant gesture, there is a tone gesture, and that tone gesture causes the same shifts in timing as a consonant gesture would (illustrated in **Figure 2b**).

This same structure has not been verified in intonation languages thus far (Mücke, Grice, Becker, & Hermes, 2009; Mücke et al., 2011; Prieto et al., 2007). Instead, the timing of pitch gestures in intonational languages patterns with in-phase coordination: The pitch gesture, consonant gesture, and vowel gesture all start at the same time (illustrated in **Figure 2c**). Based on these findings, Mücke et al. (2011) proposed that lexical tone languages exclusively use the c-center structure for tone coordination, while intonation uses in-phase coordination. The argument is that tone gestures are directly integrated with the lexical representation of the word, and as such have the same status as segmental gestures; they can thus influence the structure of the syllable constellation itself. In contrast, the argument continues, intonation is not part of the representation of the word (though for counterarguments in the framework of exemplar theory, see Martinuzzi & Schertz, 2021; Schweitzer et al., 2015); consequently, the gestures do



(a) C-center structure with two consonants.

(b) C-center structure with one consonant and one tone gesture.

(c) All gestures in-phase coupled, as proposed for intonation.

Figure 2: Coordinative schematics for c-center, with either two onset consonants or a simple onset with tone gesture. Solid lines indicate onset-to-onset (in-phase) coordination; dashed lines represent onset-to-target (anti-phase) coordination.

not directly coordinate with the segmental gestures, but rather are 'overlaid' after the gestural constellation for the syllable is already established.

However, the body of literature addressing pitch as a potential articulatory gesture remains relatively small, and thus the c-center hypothesis for tone is largely untested. Only a few languages have been investigated: Prior to the development of the c-center hypothesis for tone, the author is only aware of work on Mandarin and some speculation on Thai (Gao, 2008); since then, there has been additional work on Mandarin (Yi, 2014, 2017; Zhang et al., 2019) and Thai (Karlin, 2014), which has further confirmed that tones in these two languages utilize the c-center, and ongoing work on Tibetan (Geissler, 2019), which paints a more complicated picture. This is problematic for three reasons. First, it is a rather large leap to generalize from two languages (Mandarin and Thai) to thousands of tone languages. Second, Mandarin and Thai are not related to each other but they do come from similar types of tone systems (i.e., they are both Asian tone languages). Third, the leap was made after discovering that a handful of European intonational systems did not have the c-center, which is problematic for similar reasons.

Furthermore, as discussed in the previous section, evidence from studies in the Autosegmental(-Metrical) framework suggests that gestural targets, as well as gestural onsets, are also a potential point of temporal stability between tones and their TBUs. Gafos (2002) has argued for a less restricted set of coordinative landmarks, including the onset, target attainment, release, and c-center (midpoint of gestural plateau). This provides a larger possible set of alignments than exclusively onset-to-onset (in-phase) or onset-to-target (anti-phase) coordination, but still produces constrained predictions as to points of timing stability between gestures. Crucially, this allows for the possibility that the goals of gestures may be prioritized for timing by the abstract phonology (Turk & Shattuck-Hufnagel, 2020). Target coordination is slightly more complicated than onset-to-onset coordination, as it does require estimates of how long each gesture will take to complete; however, there is evidence that speakers do in fact make such estimates. For example, Shaw and Chen (2019) demonstrated that the onset timing of tongue body movements is influenced by the distance from one gestural target to the next, with larger movements initiating earlier. This suggests that speakers are estimating how much time they need to achieve a gestural target at a critical time and move the onset of the gesture accordingly. There is also evidence from a study on Thai tonal coarticulation that speakers are estimating the amount of time needed to reach upcoming tone targets, comparing to an estimate of the segment durations in the TBU, and adjusting both the rate of change in F0 and segmental durations accordingly (Karlin, 2018).

This suggests that a gestural model more similar to that proposed in Gafos (2002) is necessary not only for segments, but also for tone. Looking ahead, the c-center hypothesis for tone also strictly constrains the predicted patterns of tone gesture onsets. Crucially, it does not predict that tone gestures could conclude before the onset of the TBU, as described for Swedish accent I and Valjevo Serbian rising accents. This failure to predict is true even considering gestural

coordination, rather than acoustic simultaneity: Even though the gestural onset of a lip closure gesture for an /m/ well precedes the acoustic start of an /m/, fully completing a H tone gesture before the acoustic start of a syllable onset strongly indicates that the tone gesture is not coordinated in C_2 position of a c-center structure. However, coordinating the H gesture target with the syllable onset would predict early peaks. This paper aims to fill these gaps and expand the gestural model of tone by comparing two dialects of Serbian, an Indo-European language with lexical pitch accent.

1.3. Language of study: Serbian

Serbian `(srb)` is a South Slavic (Indo-European) language with lexical pitch accent. 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 has four accent types that contrast on stressed syllables (Lehiste & Ivić, 1986): short falling (e.g., *lùka* 'onion.GEN);[1] short rising (e.g., *lùla* 'smoking pipe'); long falling (e.g., *Lûka* 'Luka [*name*]); and long rising (e.g., *lúka* 'harbor'). All words in Serbian, excluding function words such as clitics and prepositions, have one primary stress, and thus one accent (Zec, 2005).

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' (stressed) 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: Syllables with falling accents start high and fall moving into the following syllable, while syllables with rising accents start low and rise moving into the following syllable. Schemata for the four accent types are provided in blue in **Figure 3** (Lehiste & Ivić, 1986, and references therein), as they would be produced in trisyllabic words with 'accent' on the first syllable.

There are currently two major autosegmental proposals for the representation of the Serbian accentual system. For this paper, we take as a starting point the analysis presented by Inkelas and Zec (1988), who argue that Serbian accent is most fruitfully treated as a H(igh) pitch that determines the location of stress. In this analysis, the location of the H is lexically specified, and stress is located one syllable to the left. Thus, for falling accents, the TBU is the stressed syllable, while for rising accents the TBU is the post-stress syllable. The four Serbian accents thus can be broken down and described as in **Table 1**. This analysis accounts for the distribution of accents in Serbian: Rising accents can occur on any (non-final) syllable, but falling accents can only occur initially, since the stress cannot move one syllable to the left of an initial H. It also accounts for

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

the contour of rising accents, where the stressed syllable is lower in pitch than the following syllable: Falling accents occur when the H is assigned to the first syllable, and thus the contour falls from the stressed syllable. However, this account does not explicitly predict the phonetic alignment of the F0 contours relative to the TBU.

Accent	Vowel	Stress	Pitch	Phonology	Orthography	Gloss
Short falling	short	initial	initial	/ˈlu$_H$ka/	lùka	‘onion.GEN’
Short rising	short	initial	second	/ˈlula$_H$/	lùla	‘smoking pipe’
Long falling	long	initial	initial	/ˈlu:$_H$ka/	Lûka	‘Luka (name)’
Long rising	long	initial	second	/ˈlu:ka$_H$/	lúka	‘harbor’

Table 1: A breakdown of the features of the four Serbian accents as they occur on initial syllables, following Inkelas and Zec 1988. The lexical H is noted as a subscript after the syllable it is assigned to.

The main alternative analysis utilizes the star convention from Autosegmental-Metrical theory to express the contrast between rising and falling accents, and crucially associates both pitch accents to the stressed syllable, indicating that the TBU is always the stressed syllable. Smiljanić (2002) analyzed the Serbian[2] tone system in terms of pitch melodies: Rising accents are L*+H, while falling accents are L+H*. That is, rising accents have an L anchored to the stressed syllable, followed by an unanchored H, while falling accents have an H anchored to the stressed syllable, which is preceded by an unanchored L. This analysis was based on observations of the alignment of pitch extrema with acoustic segment boundaries, but fails to account for the phonological distribution of accents. Under this account, there would have to be a specific restriction in the phonology that forbids falling accents (L+H*) on non-initial syllables (Zec & Zsiga, to appear).

Of particular scientific interest are the Belgrade and Valjevo dialects of Serbian, which are typically assumed to have the same system of tonal contrast, but with different temporal alignment. Zec and Zsiga (2016) describe ‘variable retraction’ of the Valjevo rising accents in phrase-initial position:[3] For two of the three Valjevo speakers, the F0 peak occurred on the post-stress syllable 50% of the time, and was otherwise retracted to the stressed syllable. For the remaining speaker, the F0 peak consistently occurred on the stressed syllable. In the Belgrade dialect, F0 peaks of rising accents are consistently realized on the post-tonic syllable.

[2] As spoken in Belgrade.

[3] In phrase-final position, both Belgrade and Valjevo dialects shift a H peak onto the preceding syllable due to tonal crowding.

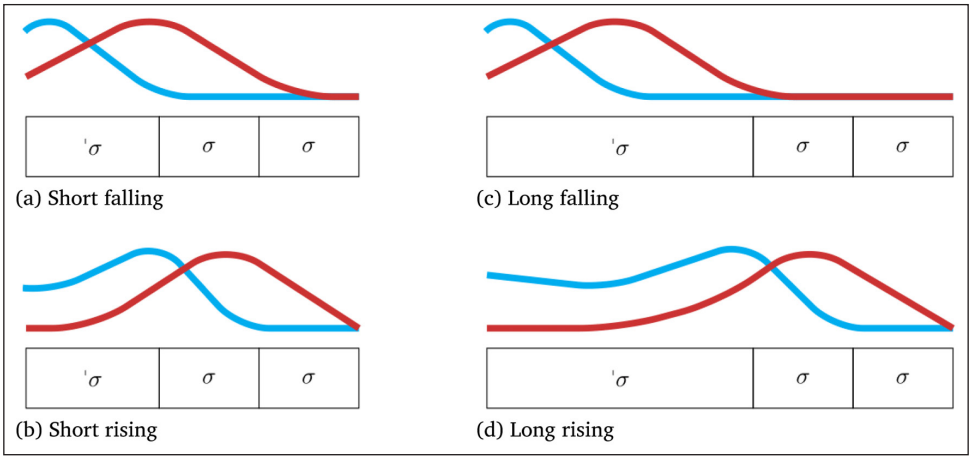


Figure 3: Schemata of the F0 movements for the four accent types on trisyllabic words, comparing Belgrade (red) and Valjevo (blue) dialects. Contours adapted from Zec and Zsiga 2016.

1.4. This study

In this study, we consider three hypotheses to capture the differences between the Belgrade and Valjevo dialects of Serbian. In the first set of hypotheses, the dialects share a common system of TBUs. The difference between the two hypotheses is which syllable serves as the TBU—most crucially, in rising accents.

> *Hypothesis 1*: Under the Inkelas and Zec (1988) analysis, Belgrade F0 peaks occur near the end of the TBU, while in the Valjevo dialect, F0 peaks occur at the very beginning or even before the beginning of the TBU. The main difference between falling and rising accents in both dialects is that the stressed syllable is the TBU for falling accents, and the post-stress syllable is the TBU for rising accents.

> *Hypothesis 2*: In contrast, the Smiljanić (2002) associates both accents with the stressed syllable, and the accentual contrast is merely one of alignment (L* + H versus L + H*) in both dialects. Furthermore, within each accent, Belgrade dialects have late alignment, and Valjevo dialects have early alignment.

Alternatively, it could be the case that the Belgrade and Valjevo systems have diverged regarding which syllable is the TBU. This hypothesis corresponds to a scenario where the retracted peaks in Valjevo Serbian are reflective of a shift in TBU, rather than solely a phonetic shift.

> *Hypothesis 3*: The Valjevo dialect has undergone a shift in the association of tones to TBUs, where all lexical H tones are associated to stressed syllables, and the contrast between rising and falling accents is one of alignment alone (a Smiljanić-type system), while Belgrade Serbian is under an Inkelas and Zec-type system.

In order to determine what the TBU is, we will examine the temporal interactions and points of temporal stability between segments and tones by varying the segmental content of the potential TBUs. The c-center hypothesis for tone only generates Hypotheses 2 and 3—specifically, the c-center structure for tone cannot generate Valjevo rising accents if they are realized fully before the segmental TBU, which is the case in Hypothesis 1, but could potentially produce contrastive alignment within one TBU, which is the case in Hypotheses 2 and 3. An expanded gestural model for tone, on the other hand, is able to generate all three scenarios. Thus, if the patterns of temporal stability and interaction support Hypothesis 1, it is necessary to expand the gestural model of tone.

In order to address these questions, we present the results of two acoustic studies on these dialects of Serbian. In the first study, we examine the interaction between segment timing and the timing of falling accent contours, which are uncontroversially associated to the initial syllable. In the second study, we compare the timing of falling accents, which provide a baseline of expected interactions between TBUs and tones, to the timing of rising accents contours, which are more controversial, and can distinguish between the hypotheses.

2. Experiment 1: Falling accents

In this experiment, we focus on the falling accent contour and establish the relationship between the segmental content of a TBU and F0 timing. The falling accents are rather uncontroversial compared to rising accents; both the Inkelas and Zec (1988) and Smiljanić (2002) analyses agree that a H(igh) tone of some sort is associated to an initial stressed syllable. Furthermore, although the Valjevo dialect exhibits peak retraction in both accent contours, in falling accents the peak still occurs within the first syllable. Thus, the behavior of the falling accents can serve as a basic case to help interpret the behavior of the rising accents (in Section 3). We use acoustic data as a proxy for articulatory information, following Selkirk and Durvasula (2013) and Ruthan, Durvasula, and Lin (2018).

2.1. Methods

2.1.1. Participants

Data was collected from a total of 24 participants (13 Belgrade, 11 Valjevo). Speakers of the Belgrade dialect were all born and raised in Belgrade, though typically one or both parents were from elsewhere. Speakers of the Valjevo dialect had all been raised in Valjevo, but were currently studying in Belgrade. Since living in Belgrade for an extended amount of time affects the realization of accent, the Valjevo speakers were all young university students who still had family ties in Valjevo and frequently visited home. One Valjevo participant was excluded after data collection as they had already accommodated their accentual production to the Belgrade dialect, as determined by trained Serbian phoneticians (one native Valjevo speaker, one native Belgrade speaker).

Several additional exclusions had to be made. The bulk of exclusions were due mainly to frequent errors, either improper focus placement (seven participants, including broad focus, narrow focus on the carrier word, and list intonation), which made it impossible to find accentual peaks in the target word, or errors in segmental or accentual production or both (three participants). Accentual errors were largely due to the use of non-words in the study, some of which share orthographic representation with existing words with different accents. There were two particularly problematic words: First, the word *rave* with a long rising accent (rather than the long falling accent desired) is a slang term that means 'whore.ACC.PL'; second, the word *lave* with a long rising accent (again, instead of the desired long falling accent) means 'lava.GEN'. Participants were typically consistent with their pronunciation; there were participants that always used the wrong accent, even after reminders, and participants who always remembered the correct and novel accent. Participants who used the wrong accent were excluded.

The final dataset includes data from 13 participants: five speakers of Belgrade Serbian (ages 19–39; 3M, 2F) and eight speakers of Valjevo Serbian (ages 19–22; 2M, 6F).

2.1.2. Target words

Target words were formed from three base words, where the first syllable onset of these words was varied to create a set of five rhyming words, using three simple onsets /r, l, m/ and two complex onsets /mr, ml/: *mrâve* /ˈmra:$_H$ʋe/ 'ant.ACC.PL' (derived words: *râve*, *lâve*, *mâve*, *mlâve*); *mràmor* /ˈmra$_H$mor/ 'marble' (derived words: *ràmor*, *làmor*, *màmor*, *mlà*mor); *mràmora* /ˈmra$_H$mora/ 'marble.GEN' (derived words: *ràmora*, *làmora*, *màmora*, *mlàmora*).[4] These onsets vary in phonetic duration, allowing investigation of phonetic effects of different syllable onsets.

Since Serbian typically does not mark any aspect of accent in its orthography, participants received a list of words that would be used in the study, which marked the accents using dictionary conventions and grouped them together to make clear what accents they had. They were informed that the nonce words were supposed to be a 'perfect rhyme' with the real word it looked like, i.e., that they had the same pitch accent. They were also told that the nonce words were supposed to refer to other things in the same lexical category as the real word—e.g., since *mràmor* was a type of stone, *mlàmor*, *màmor*, *làmor*, and *ràmor* were other types of stone. Participants were allowed to reference a sheet with accents and definitions through the study, though none had to.

[4] The full experiment also included a set of short rising words (base word *mravinjak* 'anthill') as well as a set of long rising words (base word *Monu* 'Mona.ACC [clothing brand]'); however, they are excluded from the present analysis as the location of syllable onset variation was not informative for determining the phonological association of rising accents.

2.1.3. Carrier phrases

In order to reduce boredom and consequent list intonation, there were two carrier phrases in this study: *nêmamo X* /ˈneːₕmamo/ 'we don't have X' and *ìmamo X* /ˈima(ː)ₕmo/ 'we have X'. In the full experiment there were 50 phrases total (5 accent types × 5 syllable onsets × 2 carrier phrases); this paper analyzes just the falling accents, which were 30 of these phrases.

2.1.4. Task

During the experiment, participants first heard a spoken prompt, which was recorded by a native speaker of Belgrade Serbian. The prompt either claimed that there were no instances of a lexical category (e.g., *Nemamo mineral ni na jednoj slici* 'We don't have a mineral in any picture') or that they had all instances of a lexical category (e.g., *Imamo slike za sve minerale* 'We have pictures for all of the minerals'). The participant then responded in disagreement with a written response that was presented on the screen (e.g., *Nije tačno! Imamo mramor*! 'That's not true! We have marble!'). In order to prevent overlap, rushing, and list intonation, the written response only appeared on the screen after the context prompt ended.

The 50 sentences were put in random order and then split into two groups of 25, creating two blocks to prevent fatigue (thus, one round of the experiment had two blocks with 25 trials each). After the two blocks were completed, the 50 sentences were randomized again, instead of repeating the first random order. The experiment repeated for three rounds, for a total of 150 sentences.

The experiment was presented in PsychoPy. Due to difficulties triggering clean recordings with native PsychoPy tools, the experiment was recorded in Audacity as one sound file with a Samson GoMic.

2.2. Data labeling and analysis

2.2.1. Segmentation

Data was initially aligned with the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017), and then corrected by hand in Praat (Boersma & Weenink, 2017). Marking syllable onset boundaries was typically straightforward: /m/ was demarcated by the onset and offset of nasal murmur, and /l/ was marked based on the confluence of formant transitions in and out of vowels, decreased formant amplitude, and overall decreased intensity compared to vowels. Marking word-initial /r/ was slightly more complex: The most typical realization of /r/ in Serbian is as a tap, and as such the end of /r/ was marked at the end of the tap closure. The closure period was reliably preceded by a brief fall in F3, which was used to mark the onset of the /r/. This accounts for the bracing movement of the tongue before the tap closure (Gibson, Sotiropoulou, Tobin, & Gafos, 2017) and also creates consistency with the segmentation of /mr/ clusters, which are acoustically composed of a brief vocalic interval between the nasal portion and the tap closure.

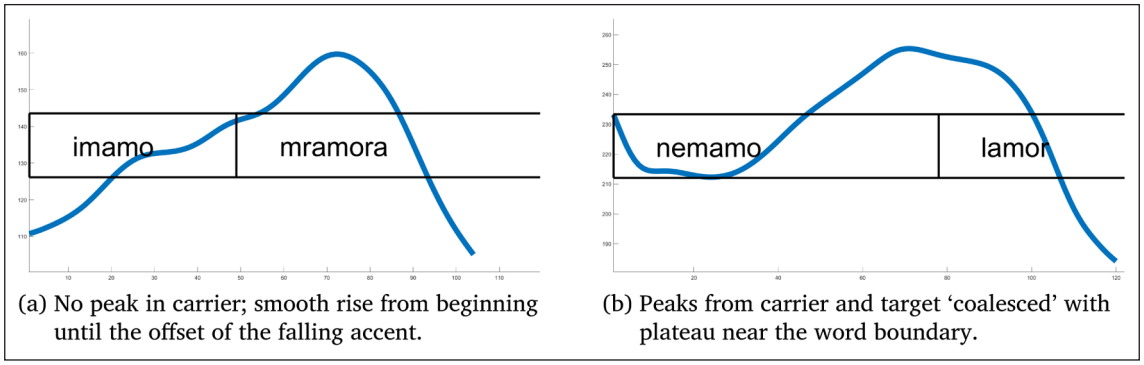
2.2.2. Pitch landmarks

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 semi-automatized using a Matlab script that found F0 extrema located within certain segment-based boundaries. Boundaries were refined by iteratively hand-checking pitch landmarking to ensure that the correct local extrema were found. Due to the dialect-based timing differences, the boundaries were ultimately defined slightly differently for Belgrade and Valjevo dialects—for example, the leftmost boundary was the acoustic onset of the target word for both dialects, but the rightmost boundary was the beginning of the second syllable nucleus for the Belgrade dialect, and the beginning of the second syllable onset for the Valjevo dialect. Dialect-specific boundaries ensured that the algorithm was provided a sufficiently large region to search for target word contours, while still avoiding potential extraneous peaks associated with the carrier phrase or phrase-final intonation. The resulting pitch landmarks were then used to bound where further landmarks (i.e., maximum onset velocity, F0 valley, and F0 gesture onset) could be located.

After pitch landmarking was completed, each trial was verified by hand for reasonable tracking. Several trials had to be excluded due to irreconcilably incorrect pitch tracking, or pitch tracks that did not clearly show minima or maxima. Only F0 trajectories with clear 0 velocity points between the carrier and target words can be used to determine the start and duration of the pitch excursion, as otherwise it is impossible to determine where the upward pitch excursion for the target word begins. F0 minima were more prone to unreliability than pitch maxima due to intonational differences within and between participants. Various intonational patterns completely obscured this minimum, including focus on the carrier word rather than on the target word. Two examples of such tokens from the Belgrade dialect are given in **Figure 4**; compare the tokens with clear minima in **Figure 5**.

In the Belgrade dialect, out of a possible 450 trials, 427 had suitable peak marking; out of those 427 trials, 401 also had suitable minima. In the Valjevo dialect, out of a possible 720 trials, 693 had suitable peak marking; out of those 690 trials, 456 also had suitable minima. Some participants had more data excluded than others, but with one exception, remaining data was balanced across syllable onset and accent within participant.

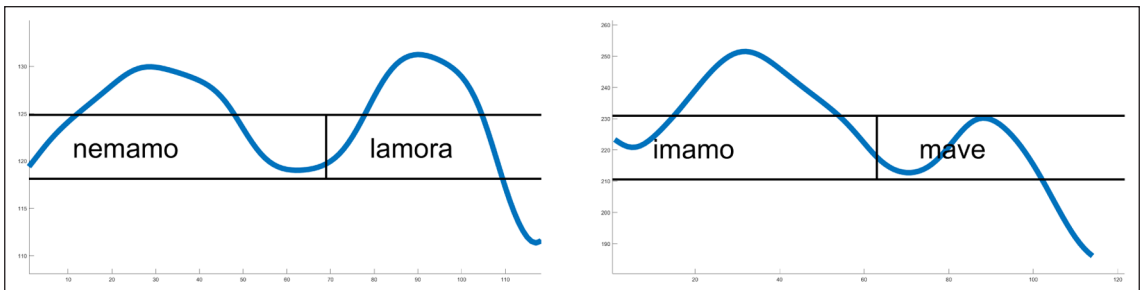
As the absolute F0 peak is less stable and prone to small fluctuations, peak timing was measured using the gestural release rather than the actual target F0 peak (Hoole, Mooshammer, & Tillmann, 1994). The H gesture release was marked at the first point after the accentual F0 peak where F0 speed achieved 20% of the maximum release speed. Similarly, analyses that involve the start of upward F0 movement references the F0 onset, rather than the F0 valley. This was marked at the first point after the F0 valley where F0 speed achieved 20% of the maximum onset speed. This type of landmarking is typical for Articulatory Phonology trajectories and has been used to landmark pitch gestures as well as segmental gestures (Gao, 2008; Karlin, 2014; Katsika, Krivokapić, Mooshammer, Tiede, & Goldstein, 2014; Yi, 2014).



(a) No peak in carrier; smooth rise from beginning until the offset of the falling accent.

(b) Peaks from carrier and target 'coalesced' with plateau near the word boundary.

Figure 4: Examples of F0 shapes that were removed for the analysis of F0 onsets.



(a) Even peaks from carrier to target, clear minimum in between.

(b) Carrier peak higher than target peak, but still clear minimum in between.

Figure 5: Examples of F0 shapes that allow analysis of F0 onsets.

2.2.3. Statistical analyses

Data was analyzed with linear-mixed effects models in R (R Core Team, 2019), using the package lme4 for model building (Bates, Maechler, Bolker, Walker, et al., 2014) and lmerTest to get p values for model components (Kuznetsova, Brockhoff, & Christensen, 2015). Models had fixed effects of syllable onset duration (duration of the varied syllable onset) and accent length (short versus long). Dialects are first examined separately, and then together to examine the effect of dialect. Random intercepts were included for participant. Models were built incrementally, and likelihood ratio tests used to compare models; final models include only factors that significantly improved model fit. Models were built starting with syllable onset duration, then adding accent length and their interaction; models with both dialects then added dialect and its interactions with syllable onset duration and accent length. Post-hoc tests were done using least means squared tests with a Tukey adjustment using the emmeans package (Lenth, 2019). In analyses of the timing of the onset of the F0 excursion, the Aikake Information Criterion (AIC) was used to evaluate if phonetic duration or complexity of the syllable onset was a better predictor, where lower values indicate better model fit.

Dependent variables are H offset (timing of accentual peak relative to the beginning of the word), the timing of the onset of the pitch excursion (interval between the start of the upward

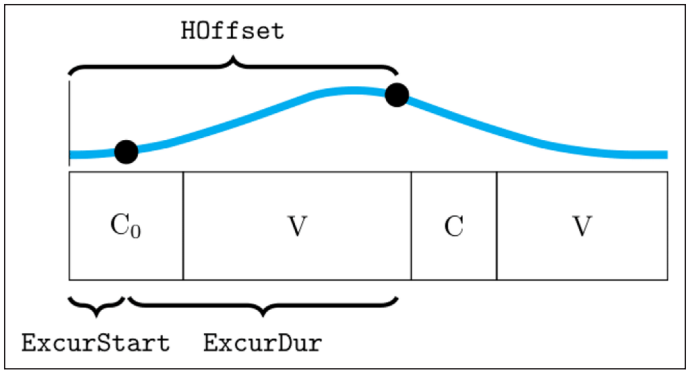


Figure 6: A schema of the dependent variables used in analysis. The blue line is a schematized short falling accent, with black dots to mark the start (leftmost) and H offset (rightmost) of the pitch excursion.

pitch trajectory and the beginning of the word), and the duration of the pitch excursion (interval between H offset and start of pitch excursion). A schematic of these variables is provided in **Figure 6**.

2.3. Syllable onset duration

As different syllable onsets were chosen in order to compare effects of syllable onset complexity versus phonetic duration, we will first present the acoustic duration characteristics of the onsets used, which differ in duration as intended. Including dialect as a predictor of syllable onset duration did not significantly improve model fit ($\chi^2(1) = 0.93, p = 0.33$), so the data for syllable onset duration is pooled; only syllable onset identity is used as a predictor of syllable onset duration. There is a significant effect of syllable onset identity on syllable onset duration. All syllable onsets are significantly different from each other ($p < 0.0001$ for all comparisons): /r/ (45.4 ± 2.28 ms) < /l/(66.8 ± 2.28 ms) < /m/ (92.0 ± 2.28 ms) < /mr/ (127.6 ± 2.27 ms) < /ml/ (137.2 ± 2.28 ms).

2.4. Results

2.4.1. Target achievement (H gesture offset)

In both dialects, H offset occurs later with in words with longer syllable onset duration; these patterns are illustrated in **Figure 7**. The summary of the full model is presented in **Table 2**.

In models for the Belgrade dialect alone, syllable onset duration as a single fixed effect significantly improves the model; words with longer syllable onsets have later H offsets ($\beta = 0.72$ ms, $SE = 0.04$ ms). The addition of accent length also significantly improves the fit of the model; H offsets occur significantly later in words with short vowels (147.0 ± 6.36 ms) than in words with long vowels (130.0 ± 6.55 ms, $p < 0.0001$). The interaction between accent length and

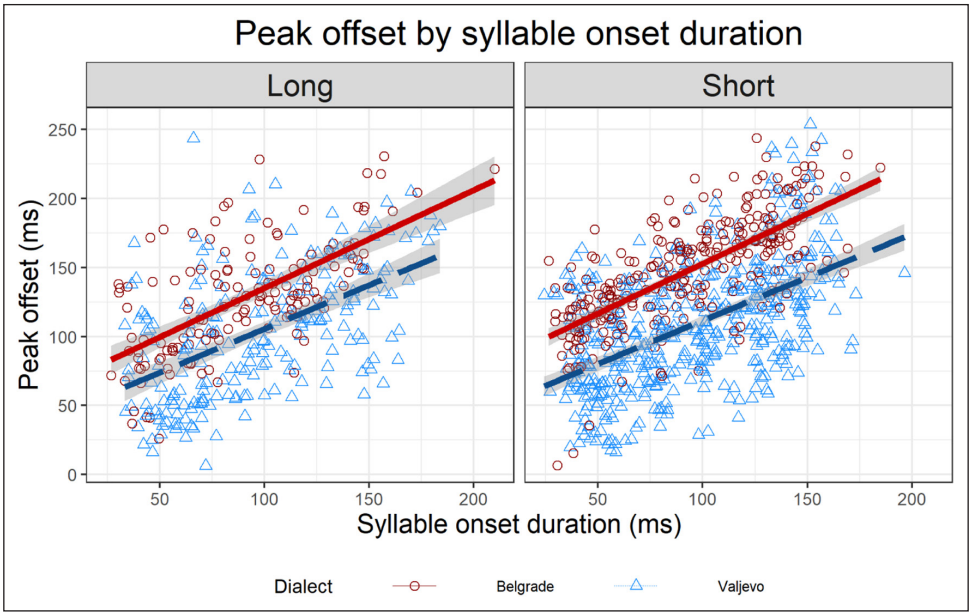


Figure 7: Scatter plots comparing the relationships between varied syllable onset durations and H offset in falling words (both dialects).

syllable onset duration is not significant ($\chi^2(1) = 0.13, p = 0.72$), indicating that syllable onset duration affects H offset timing in both short and long falling accents equally.

Patterns are similar in the Valjevo dialect. The addition of syllable onset duration significantly improves the fit of the model; words with longer syllable onsets have later H offsets ($\beta = 0.66$ ms, $SE = 0.03$ ms). The addition of accent length also significantly improves the fit of the model, where again accentual peaks occur significantly later in short falling words (109.0 ± 10.0 ms) than in long falling words (103.0 ± 10.1 ms, $p = 0.01$). As the difference in both dialects is quite small, it is unlikely that this is under deliberate control. The interaction between accent length and syllable onset duration is not significant ($\chi^2(1) = 0.10, p = 0.75$).

In all, both dialects show effects of syllable onset duration on the timing of H achievement. In a model that includes both dialects, adding dialect as a fixed effect significantly improves the fit of the model; as described previously in the literature, pitch peaks overall occur earlier in the TBU in the Valjevo dialect (104.0 ± 8.54 ms) than in the Belgrade dialect (142.0 ± 10.80 ms). Adding the interaction between dialect and syllable onset duration does not significantly improve the model ($\chi^2(1) = 0.92, p = 0.34$), indicating that in both dialects, accentual peaks occur later with increased syllable onset duration. Adding the interaction between dialect and accent length does significantly improve the model; as previously noted, there is a smaller difference in H offset between short and long falling accents in the Valjevo dialect as compared to the Belgrade dialect. The three-way interaction between dialect, accent length, and syllable onset duration does not significantly improve the fit of the model ($\chi^2(1) = 0.24, p = 0.63$).

	β	*SE*	*t* value	*p* value
(Intercept)	67.36	10.29	6.55	<0.0001***
OnsetDuration	0.68	0.02	31.55	<0.0001***
ShortAccent	17.69	2.79	6.35	<0.0001***
Valjevo	–29.94	12.88	–2.32	0.04*
ShortAccent:Valjevo	–11.88	3.55	–3.34	0.0009***

Table 2: Summary of the LME model for H offset: `HOffset ~ onsetDuration + accentLength + dialect + accentLength:dialect + (1|Participant)`. *Reference levels: accentLength = long, dialect = Belgrade.*

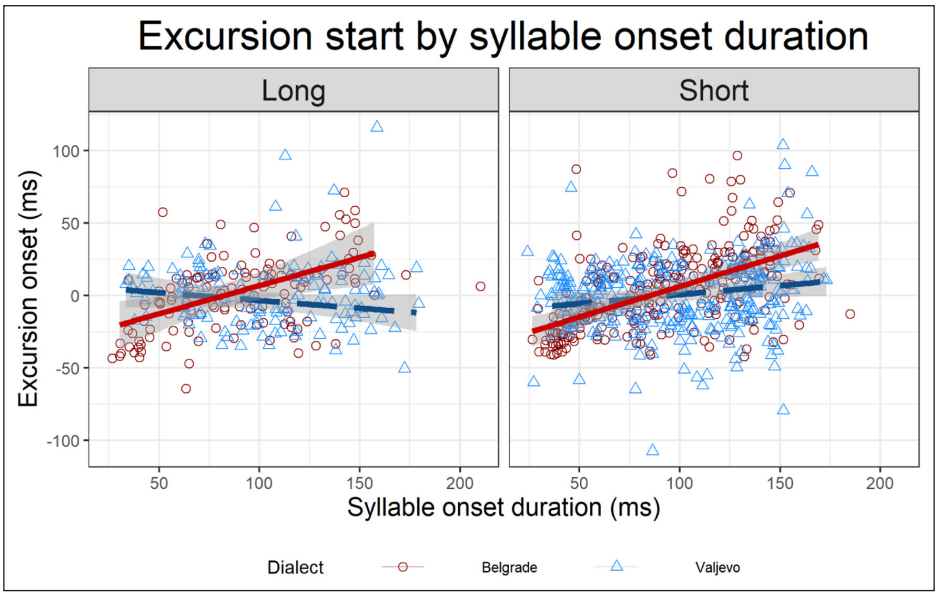


Figure 8: Scatter plots comparing the relationships between varied syllable onset durations and start of the pitch excursion in falling words (both dialects).

2.4.2. Excursion characteristics

Start of excursion The timing of the start of the pitch excursion patterns differently in each dialect; this is illustrated in **Figure 8**. In the Belgrade dialect, the addition of syllable onset duration significantly improves the fit of the model; pitch excursions start later in words with longer syllable onsets ($\beta = 0.40$ ms, $SE = 0.03$ ms). Phonetic duration of the syllable onset is a better predictor ($AIC = 3{,}511.7$) of the timing of the start of the pitch excursion than syllable onset complexity ($AIC = 3{,}622.9$), indicating that coordination of the H gesture onset is gradiently determined by duration of the syllable onset, rather than having two modes in simple versus complex onsets. The addition of accent length does not significantly improve the fit of the model ($\chi^2(1) = 1.58, p = 0.21$), nor does the interaction between accent length and syllable

onset duration ($\chi^2(1) = 0.71, p = 0.40$). The effect of syllable onset duration on the start time of the pitch excursion is the same for both long and short falling words.

The Valjevo dialect behaves somewhat similarly, though the pitch excursion moves less in response to syllable onsets than in the Belgrade dialect. The addition of syllable onset duration significantly improves the fit of the model in the Valjevo dialect; pitch excursions start slightly later in words with longer syllable onsets ($\beta = 0.11$ ms, $SE = 0.03$ ms). Phonetic duration of the syllable onset ($AIC = 4{,}175.6$) and syllable onset complexity ($AIC = 4{,}175.8$) are equally good predictors; as the effect of syllable onset duration on F0 onset timing is quite small, this equivalency is likely due to the correlation between complexity and duration. Adding accent length does not significantly improve the fit of the model ($\chi^2(1) = 0.0001, p = 0.99$), nor does the interaction between accent length and syllable onset duration ($\chi^2(1) = 2.15, p = 0.14$).

In a model that considers both dialects together, adding syllable onset duration significantly improves model fit. Adding dialect does not improve the model fit ($\chi^2(1) = 0.51, p = 0.48$). However, adding the interaction between dialect and syllable onset duration does significantly improve model fit; syllable onset duration has less effect on the timing of the start of the pitch excursion in the Valjevo dialect than in the Belgrade dialect ($\beta = -0.28$ ms, $SE = 0.04$ ms, $p < 0.0001$). No other factors or their interactions significantly improve the fit of the model. The summary of the full model is provided in **Table 3**.

Excursion duration The patterns for excursion duration in both dialects are illustrated in **Figure 9**. The duration of the pitch excursion is affected by the syllable onset duration in the Belgrade dialect. Adding syllable onset duration to the model significantly improves the fit; words with longer syllable onsets also have longer pitch excursions ($\beta = 0.29$ ms, $SE = 0.04$ ms). Adding accent length as a second fixed effect significantly improves the fit of the model ($\chi^2(1) = 12.87, p = 0.0003$); pitch excursions in short falling words are significantly longer (147.0 ± 8.1 ms) than pitch excursions in long falling words (131.0 ± 7.9 ms, $p < 0.0001$). Thus, accentual peaks occur later in short falling words due to the duration of the pitch excursion, rather than the start of the pitch excursion. The interaction between accent length and syllable onset duration does not significantly improve the fit ($\chi^2(1) = 0.09, p = 0.77$).

	β	***SE***	***t* value**	***p* value**
(Intercept)	–33.88	5.36	–6.32	<0.0001***
OnsetDuration	0.40	0.03	13.83	<0.0001***
Valjevo	22.79	7.07	3.23	0.003**
OnsetDuration:Valjevo	–0.28	0.04	–7.22	<0.0001***

Table 3: Excursion start timing: `excursionStart ~ onsetDuration + dialect + dialect:onsetDuration + (1|Participant)`. *Reference levels: onsetDuration = 0, dialect = Belgrade.*

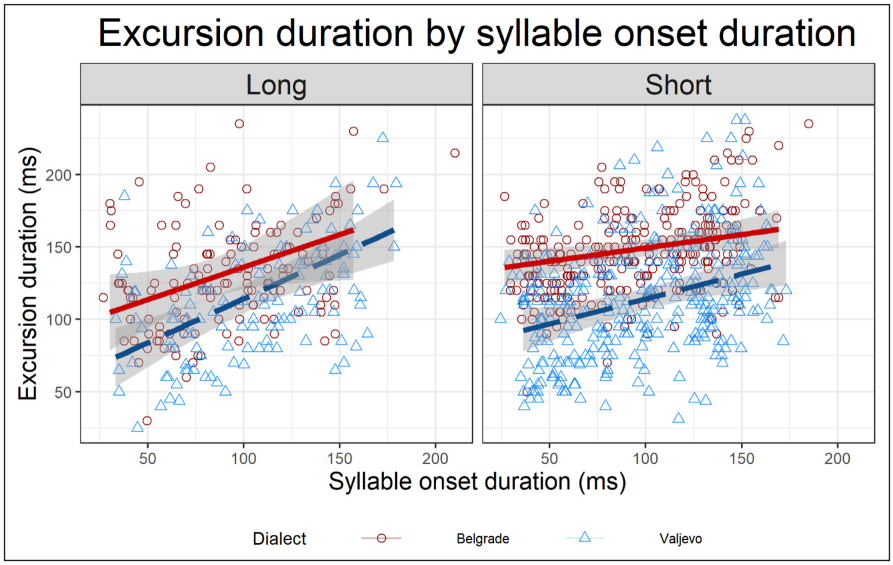


Figure 9: Scatter plots comparing the relationships between varied syllable onset durations and excursion duration in falling words (both dialects).

In the Valjevo dialect, syllable onset duration significantly improves the fit of the model; words with longer syllable onsets have longer pitch excursions ($\beta = 0.49$ ms, $SE = 0.04$ ms). The addition of accent length significantly improves the fit of the model; short accents have longer excursions (114.0 ± 7.7 ms) than long accents (114.0 ± 8.0 ms). The interaction between accent length and syllable onset duration does not significantly improve the model fit ($\chi^2(1) = 1.67, p = 0.20$); pitch excursions stretch with syllable onset duration equally in both short and long falling accents.

When considering both dialects together, the addition of dialect marginally improves the fit of the model ($\chi^2(1) = 7.37, p = 0.007$); excursions in the Valjevo dialect are shorter than excursions in the Belgrade dialect ($\beta = -46.77$ ms, $SE = 12.03$ ms, $p = 0.0008$). Both dialects increase the duration of their pitch excursions with longer syllable onsets; however, the addition of the interaction between dialect and syllable onset duration significantly improves the model, where Valjevo pitch excursions increase more with increased syllable onset duration ($\beta = 0.21$ ms, $SE = 0.05$ ms, $p < 0.0001$). The interaction between accent length and dialect also significantly improves the model fit; the difference between excursion durations in the long and short accents is smaller in the Valjevo dialect than in the Belgrade dialect ($\beta = -8.73$ ms, $SE = 3.99$ ms, $p = 0.03$). No other interactions significantly improve the model. The summary of the full model is provided in **Table 4**.

2.5. Interim summary

This experiment provided a baseline for the behavior of pitch accents and their TBUs in Serbian. In both dialects, H offset is later when the lexical H is associated to a syllable with longer syllable

	β	*SE*	*t* value	*p* value
(Intercept)	104.72	9.15	11.44	<0.0001***
OnsetDuration	0.29	0.04	7.89	<0.0001***
ShortAccent	15.50	2.85	5.43	<0.0001***
Valjevo	–46.77	12.03	–3.89	0.0008**
ShortAccent:Valjevo	–8.73	3.99	–2.19	0.03*
OnsetDuration:Valjevo	0.21	0.05	34.22	<0.0001***

Table 4: Excursion duration: `excursionDuration ~ onsetDuration + accentLength + dialect + dialect:accentLength + dialect:onsetDuration + (1|Participant)`. *Reference levels: accentLength = long, dialect = Belgrade.*

onset. However, the two dialects achieve this in different ways. In the Belgrade dialect, pitch excursions both start later and get longer with increases in syllable onset duration. In the Valjevo dialect, the pitch excursions start only slightly later with increased syllable onset duration, and the bulk of the H offset timing comes from stretching the pitch excursion with increased syllable onset duration. These patterns of 'stretching' suggest that both dialects have some form of H gesture target coordination, but differ on how the onset of the H gesture is coordinated with the syllable. As the association of falling accents is uncontroversial, this indicates that the two dialects express the same phonological association with different coordinative schemes. Despite these small differences, however, in both dialects it is true that the timing of the TBU provides timing information to the accentual pitch excursion. As rising accents may differ in phonological representation between the two dialects, we can compare this baseline of how falling accents behave in each dialect to the behavior of rising accents in each dialect in Experiment 2.

3. Experiment 2: Rising accents

In order to probe the association of the H tone in rising accents in Belgrade and Valjevo Serbian, we examine the effects of syllable onset on the coordination and timing of the rising accent in two rising accent variations: First, when only the stressed syllable varies in onset complexity, and second, when only the post-tonic syllable (proposed TBU) varies in onset complexity.

3.1. Methods

3.1.1. Participants

Participants were recruited according to the same criteria as in Experiment 1. Data was collected from 19 participants (10 Belgrade, nine Valjevo). Some data had to be excluded due to consistent production errors, either segmental or accentual (three participants) or due to an alternative

lexical accent[5] of the target word *omladinu* (two speakers). In all, the dataset includes eight (4F, 4M) Belgrade speakers and six (5F, 1M) Valjevo speakers. Four of the Belgrade speakers and three of the Valjevo speakers had previously participated in Experiment 1.

3.1.2. Target words

The target words for Experiment 2 were very similar to those used in Experiment 1, but focus solely on short accents. This serves to control the duration of the experiment and keep the statistical models more simple, and also makes for a more straightforward comparison between rising and falling accents, as the length descriptor refers to the length of the stressed vowel, not the length of the vowel with the H tone.

As in Experiment 1, the target words were formed from three real words, with one syllable onset varied (using /r, l, m, mr, ml/) to make four additional nonce words. The three base words are ***mr***ràmora 'marble.GEN' (derived words: ***r***àmora, ***l***àmora, ***m***àmora, ***ml***àmora); ***mr***àvinjak[6] 'anthill' (derived words: ***r***àvinjak, ***l***àvinjak, ***m***àvinjak, ***ml***àvinjak); and ò***ml***adinu 'youth.ACC' (derived words: ò***r***adinu, ò***l***adinu, ò***m***adinu, ò***mr***adinu).[7] The template *mràmora* is a short falling accent, which means that the association of the lexical H to the first syllable is uncontroversial; this provides verification of the effects of the TBU on tone. The template *mràvinjak* is a short rising accent, where the varied syllable onset is on the stressed syllable and the lexical H is potentially on the post-tonic syllable. The template *òmladinu* is also a short rising accent, but the varied syllable onset is on the post-tonic syllable. The *òmladinu* template is thus the crucial test case; if the lexical H is associated to the post-tonic syllable, the varied syllable onset will have an effect on the timing of the accentual contour similar to the *mràmora* template.

3.1.3. Carrier phrases

In order to prevent some boredom and make sure the participants were paying attention throughout, two different stimuli frames were used: *Da li želite X?* 'Do you want X?' and *Jeste li rekli X?* 'Did you say X?'. These two stimuli prompted slightly different response phrases: *Neću*! *Daj mi Y, molim te* 'I don't want that! Give me Y, please' and '*Nisam! Daj mi Y, molim te* 'I didn't [say that]! Give me Y, please', respectively. This context ensured that focus would be put on the target word.

---

[5] Producing *omladinu* with a falling accent is linked to older speakers and those from regions other than Belgrade; typically younger speakers, especially those from Belgrade, have a rising accent on this word.

[6] Note that <v> in Serbian represents the approximant [ʋ].

[7] The syllabification of *òmladinu* is unambiguous: First, the principle of maximal syllable onset would encourage a syllabification of /o.mla.di.nu/ rather than /om.la.di.nu/; second, the inclusion of the root *mlâd* 'young' is transparent, which further encourages a syllabification of /o.mla.di.nu/.

3.1.4. Task

As in Experiment 1, participants first heard a context prompt (recorded in advance by a native speaker of Valjevo Serbian who has been living in Belgrade for 20 years) and then read the given response. In order to prevent overlap, rushing, and list intonation, the written response only appeared on the screen after the context prompt ended. The context prompt asked the participant if they wanted or had asked for a certain object; the object in the question was a semantically plausible replacement for the target word, and had the same accent and syllable number as the target words. The same prompts were used for all words in an accentual group, so it was not possible to fully anticipate what the response was, as there are always five possible responses for any given prompt. Two example questions and responses are presented in **Figure 10** (where the target word was presented in upper case for the experiment as well, in order to encourage a focused reading).

Context:	Da li želite drveta? "Do you want [pieces of] wood?"
Response:	Neću! Daj mi MRAMORA, molim te. "I don't! Give me [pieces of] MARBLE, please."
Context:	Jeste li rekli 'očevinu'? "Did you say 'inheritance'?"
Response:	Nisam! Daj mi OMLADINU, molim te. "I didn't! Give me YOUNG PEOPLE, please."

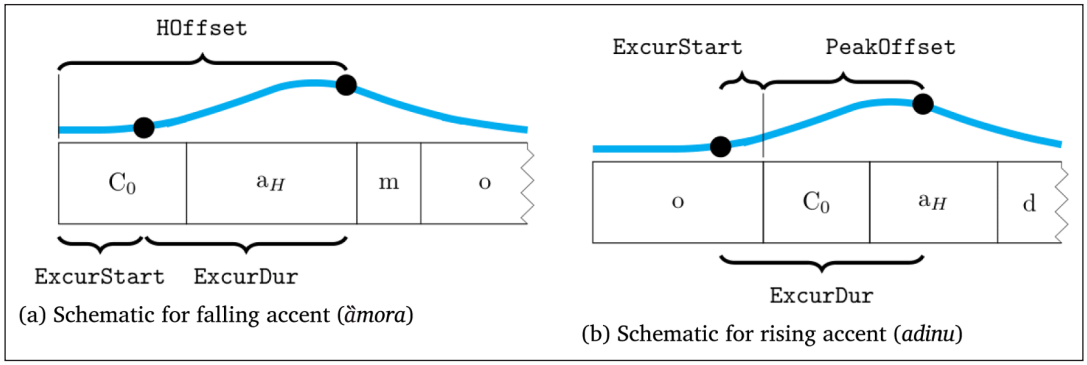
Figure 10: Example contexts and responses.

There were 15 target phrases total (three accent types × five syllable onsets), with two questions for each target phrase. As in Experiment 1, the order of presentation was fully randomized for every round of the experiment. For this experiment, the 30 prompt questions were put in random order and then split down the middle to make two blocks (thus, two blocks with 15 sentences each). After the two blocks were completed, the 30 sentences were randomized again, instead of repeating the first random order. The sentences were repeated five times, for a total of 150 trials.

The experiment was presented using PsychoPy. Participants were recorded in a quiet room, using either a TASCAM DR-100mkII microphone (four participants, all Belgrade speakers), a Sennheiser noise-canceling headset (four participants, all Valjevo speakers), or a Shure head-mounted microphone (four Belgrade speakers, two Valjevo speakers).

3.1.5. Data labeling and analysis

Data was processed and segmented as in Experiment 1. Single trials were excluded if there were significant disfluencies or segmental or accentual errors (participants who were entirely excluded tended to have at least 30 errors in the experiment; participants who had occasional trials excluded did not exceed 10 errorful trials). Some additional data cleaning was necessary for



(a) Schematic for falling accent (*ǎmora*)

(b) Schematic for rising accent (*adinu*)

Figure 11: A schema of the dependent variables used in analysis, as marked on falling **(a)** and rising **(b)** accents. The blue line is a schematized accent, with black dots to mark the start (leftmost) and H offset (rightmost) of the pitch excursion.

the analysis of the pitch excursions. Out of a possible 1,200 Belgrade tokens, 1,163 had a clear maximum and were retained for an analysis of H achievement. Of those, 1,073 also had clear minima and were retained. Out of a possible 900 tokens in the Valjevo dialect, 882 tokens had a clear maximum, and 846 tokens additionally had clear minima.

Data was analyzed in R using the same procedures as Experiment 1. Models had fixed effects of TBU syllable onset duration (duration of the onset of the syllable proposed to be the TBU in the Inkelas and Zec 1988 analysis) or duration of the varied syllable onset, and prosodic template, referring to which syllable has the varied onset (H + Stress *amora* versus H-only *adinu* versus Stress-only *avinjak*). Dialects are first examined separately, and then together to examine the effect of dialect. Random intercepts were included for participant. Models were built incrementally, starting with syllable onset duration, prosodic template, and the interaction between them; for models with both dialects, dialect was then added, as well as its interactions with the other factors. Likelihood ratio tests were used to compare models. Post-hoc tests were done using least means squared tests with a Tukey adjustment using the emmeans package.

The dependent variables are the same as in Experiment 1: H offset, timing of F0 excursion onset, duration of F0 excursion. However, for the rising accents the syllable that the timing of the pitch excursion is compared to is the proposed TBU (i.e., the second syllable), rather than the first syllable. This is illustrated in **Figure 11**.

3.2. Syllable onset duration in unstressed syllables

The same set of onsets was used in Experiment 2 as in Experiment 1; however, as Serbian uses duration as a correlate of stress, it is also necessary to verify the durational differences between onsets in unstressed syllables. As for Experiment 1, the addition of dialect as a fixed effect does not significantly improve the model ($\chi^2(1) = 0.60, p = 0.81$ for syllable onset duration, $\chi^2(1) = 0.77$, $p = 0.38$ for rime duration), so the results are pooled and reported together. There is a significant

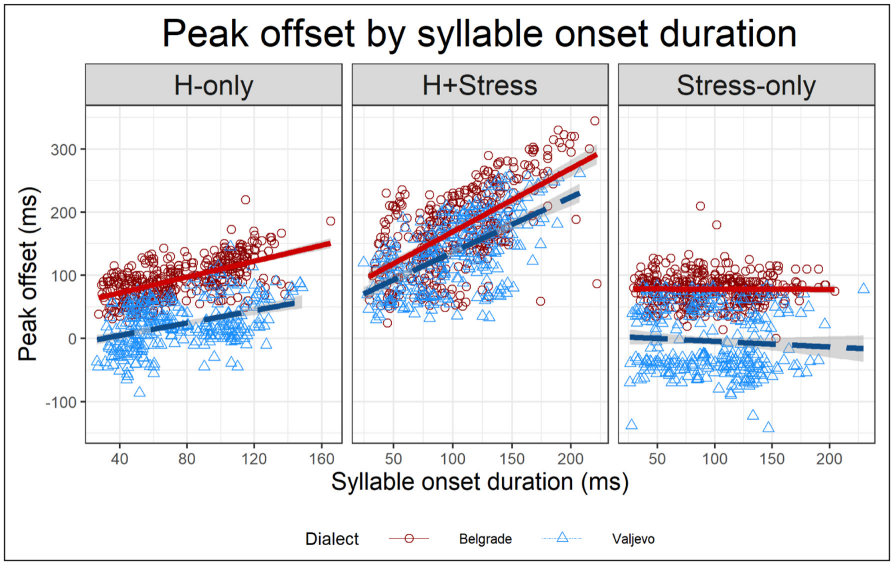


Figure 12: Scatter plots comparing the relationships between varied syllable onset durations and H offset in H + Stress, Stress-only, and H-only words (both dialects).

effect of syllable onset identity on syllable onset duration ($\chi^2(4) = 2{,}833.9$, $p < 0.0001$), /r/ < /l/ < /m/ < /mr/ < /ml/ (all $p < 0.0001$). Adding prosodic template as a second fixed effect significantly improves the model, where H-only syllable onsets are the shortest (76.0 ± 2.3 ms), followed by Stress-only syllable onsets (96.7 ± 2.6 ms), with H + Stress syllable onsets the longest (102.3 ± 2.6 ms, all $p < 0.0001$). As the syllable onsets in H-only words belong to an unstressed syllable, it is expected that they would be the shortest; the syllable onsets in the other two loci belong to stressed syllables and are accordingly longer (with estimated means separated by 5.6 ms).

There is a similar pattern for the duration of the nucleus of the syllable with the varied syllable onset. Prosodic template as a fixed effect significantly improves the fit of the model ($\chi^2(2) = 2{,}021.4$, $p < 0.0001$); as expected, H-only has the shortest nucleus duration (79.3 ± 4.4 ms), followed by H + Stress (113.5 ± 4.4 ms), and Stress-only words with the longest nuclei (120.2 ± 4.4 ms, all $p < 0.0001$). The addition of syllable onset duration significantly improves the model ($\chi^2(1) = 340.3$, $p < 0.0001$), where nuclei are slightly shorter when there is a phonetically longer syllable onset ($\beta = -152.9$ ms, $SE = 7.9$ ms for an increase of 1 s in syllable onset duration).

3.3. Results

3.3.1. Target achievement (H offset)

The patterns of H offset timing are illustrated for both dialects in **Figure 12**. As in Experiment 1, H offsets in falling accents (H + Stress words) occur later in syllables with longer syllable onsets; for rising accents, this relationship holds in H-only words but not Stress-only words, indicating that the post-stress syllable is the TBU in both dialects.

In the Belgrade dialect, the duration of the varied syllable onset as a single predictor significantly improves the fit of the model; accentual peaks occur later when the syllable onset is longer (β = 0.58 ms, SE = 0.04 ms, $p < 0.0001$). Adding prosodic template significantly improves the model fit, as does the interaction between prosodic template and syllable onset duration. The positive relationship between varied syllable onset duration and H offset exists only for words in the H + Stress and H-only prosodic templates, i.e., words where the onset of the phonologically H syllable is varied; words in the Stress-only prosodic template have a flat-to-negative relationship.

In contrast, when using the duration of the TBU syllable onset, all templates have a positive relationship between syllable onset duration and H offset. The duration of TBU syllable onset duration as a single fixed effect significantly improves the fit of the model. This model (AIC = 11,676.4) provides a better fit than the model with the varied syllable onset duration (AIC = 12,441.6). Adding prosodic template as a second fixed effect also significantly improves model fit, as does the interaction between TBU syllable onset duration and prosodic template. The effect of syllable onset duration is not different for the H-only and Stress-only templates (β = –0.02 ms, SE = 0.15 ms, p = 0.91), which are the two templates with rising accents. Note that in these models the changes in syllable onset duration for Stress-only are simply trial-to-trial variability in the duration of /v/, rather than differences due to different segments.

Accentual peaks pattern similarly in the Valjevo dialect. The duration of the varied syllable onset as a single predictor significantly improves the fit of the model; accentual peaks occur later when the syllable onset is longer (β = 0.51 ms, SE = 0.06 ms, $p < 0.0001$). Adding prosodic template significantly improves model fit, as does the interaction between template and syllable onset duration. The difference between the effect of syllable onset duration in H-only and H + Stress words is statistically significant (β = 0.34 ms, SE = 0.09 ms, $p < 0.0001$), but the relationship between syllable onset duration and H offset is still positive for both prosodic template. In contrast, there is a significant difference between H-only and Stress-only words (β = –0.61 ms, SE = 0.08 ms, $p < 0.0001$), where Stress-only words have a negative relationship between syllable onset duration and H offset. This is the same as in the Belgrade dialect.

As in the Belgrade dialect, when using the duration of the TBU syllable onset, all prosodic templates have a positive relationship between syllable onset duration and H offset. The duration of TBU syllable onset duration significantly improves the fit of the model, and this model (AIC = 9,594.6) provides a better fit than the model with the varied syllable onset duration (AIC = 10,038.6). Adding prosodic template significantly improves model fit, as does the interaction between TBU syllable onset duration and prosodic template. The effect of syllable onset duration is not different for the H-only and Stress-only templates (β = 0.03 ms, 0.24 ms, p = 0.89), as in the Belgrade dialect. However, in this case, the accentual peaks in H-only words occur approximately at the same time as the start of the post-tonic syllable (intercept β = –9.18 ms, SE = 12.23 ms, p = 0.47); compare the Belgrade dialect where accentual peaks in H-only

words occur after the start of the post-tonic syllable (intercept β = 54.08 ms, SE = 8.40 ms, $p < 0.0001$). In **Figure 12** it is evident that a good number of tokens have accentual peaks preceding the start of the post-tonic syllable, mainly tokens with short onsets (/r/ and /l/) but also including some with the longer syllable onsets. This indicates that in the Valjevo dialect, the accentual contour is getting timing information from the onset of the post-tonic syllable even when it has been acoustically retracted from that syllable.

Overall, the Belgrade and Valjevo dialects look similar to each other. In a model including both dialects, adding dialect as a fixed effect significantly improves the fit of the model; Valjevo peaks occur earlier relative to the start of the syllable (β = –62.92 ms, SE = 12.76 ms, p = 0.0002), which aligns with the data from Experiment 1, as well as previous descriptions of the Valjevo dialect. Adding the interaction between dialect and syllable onset duration also significantly improves the model; syllable onset duration has a great effect on H offset in the Valjevo dialect (β = 0.25 ms, SE = 0.04 ms, $p < 0.0001$). The interaction between dialect and prosodic template also significantly improves the model. Although in both dialects the peaks in falling accents occur later than peaks in rising accents (due to the increased duration of stressed syllables, which falling accentual peaks occur on), in the Valjevo dialect there is a greater difference between H-only and H + Stress words (β = 42.34 ms, SE = 9.40 ms, $p < 0.0001$). This is likely due to the fact that rising accents are sometimes fully retracted off the H-bearing syllable, but falling accentual peaks still occur within the word, meaning that they cannot be retracted to before the first syllable of the word. The three-way interaction between onset duration, prosodic template, and dialect does not improve model fit (for full model summary, see **Table 5**).

	β	***SE***	***t* value**	***p* value**
(Intercept)	53.79	9.16	5.87	< 0.0001***
OnsetDuration	0.54	0.05	11.79	< 0.0001***
H + Stress	27.47	4.94	5.56	< 0.0001***
Stress-only	–3.39	7.57	–0.45	0.65
Valjevo	–62.54	13.46	–4.65	0.0002***
OnsetDuration:H + Stress	0.34	0.05	6.83	< 0.0001***
OnsetDuration:Stress-only	0.002	0.13	0.01	0.99
OnsetDuration:Valjevo	–0.13	0.05	–2.66	0.008**
H + Stress:Valjevo	42.34	3.62	11.688	< 0.0001***
Stress-only:Valjevo	–13.23	3.60	–3.67	0.0002***

Table 5: Summary of the LME model for H offset, using the duration of the TBU syllable onset. Model: `HOffset ~ onsetDuration + template + onsetDuration:template + dialect + dialect:onsetDuration + dialect:template + (1|Participant)`. *Reference Levels: template = H-only, dialect = Belgrade.*

3.3.2. Excursion characteristics

For the excursion characteristics, we will focus on the duration of the TBU syllable onset, as it provides a better prediction of the data than the varied syllable onset.

Excursion start As in Experiment 1, syllable onset duration affects the timing of the onset of the H gesture for falling accents in the Belgrade dialect, but not in the Valjevo dialect. In neither dialect does syllable onset duration affect the timing of the H gesture onset in rising accents. This timing of the start of the pitch excursion in both dialects is illustrated in **Figure 13**.

In the Belgrade dialect, the addition of prosodic template significantly improves the fit of the model. Pitch excursions in words in the H-only template start the earliest (–35.3 ± 9.70 ms) and are similar but significantly different from excursions in words in the Stress-only template (–22.3 ± 9.70 ms, $p = 0.0004$). Pitch excursions in the H+Stress template start the latest (42.6 ± 9.7 ms, $p < 0.0001$ for both comparisons). The addition of syllable onset duration also significantly improves the fit of the model; pitch excursions start later in words with longer syllable onsets ($\beta = 0.43$ ms, $SE = 0.05$ ms). There is also a significant interaction between prosodic template and syllable onset duration. Rising accent excursions do not change when they start based on syllable onset duration (onset duration $\beta = 0.12$ ms, $SE = 0.08$ ms, $p = 0.14$; no significant difference in effect of onset duration between H-only and Stress-only words, $\beta = 0.47$ ms, $SE = 0.27$ ms, $p = 0.07$). In contrast, falling pitch accents start later in words with longer syllable onsets ($\beta = 0.45$ ms, $SE = 0.10$ ms, $p < 0.0001$).

The patterns differ in the Valjevo dialect. The addition of prosodic template significantly improves the fit of the model. Pitch excursions start the earliest in rising words; for the H-only template, pitch excursions start 100.81 ± 9.89 ms before the start of the TBU syllable, and in the

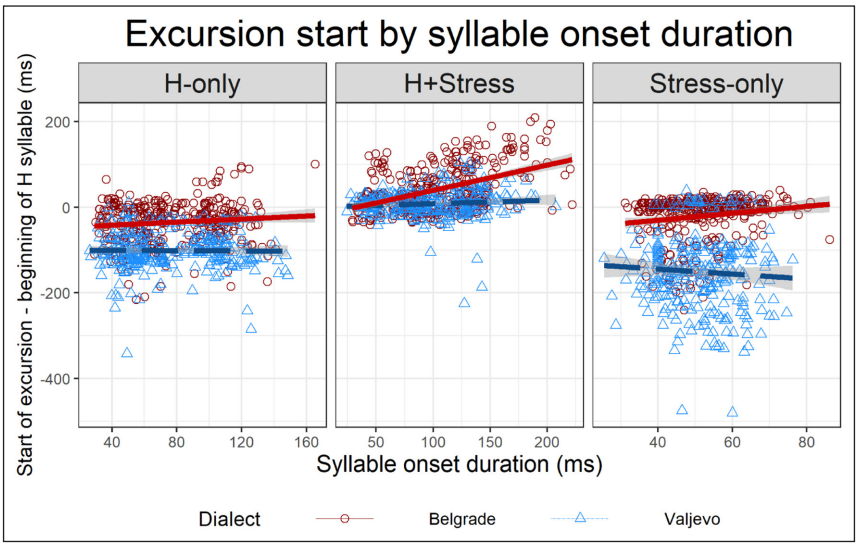


Figure 13: Scatter plots comparing the relationships between varied syllable onset durations and H offset in H+Stress, Stress-only, and H-only words (both dialects).

Stress-only template 150.39 ± 9.90 ms before the start of the TBU syllable. Pitch excursions in the H + Stress template start approximately concurrently with the start of the TBU (7.24 ± 9.95 ms). All templates are significantly different from each other ($p < 0.0001$ for all comparisons). Unlike in the Belgrade dialect, the addition of onset duration does not significantly improve the fit of the model ($\chi^2(1) = 2.88, p = 0.09$), nor does the interaction between prosodic template and syllable onset duration ($\chi^2(2) = 1.27, p = 0.53$). Adding complexity instead of phonetic duration does significantly improve the model ($\chi^2(1) = 6.76, p = 0.009$). The timing of the start of the pitch excursion is not dependent on the duration of the syllable onset in the Valjevo dialect.

For models considering both dialects, the addition of dialect significantly improves the fit of the model. The interaction between dialect and syllable onset duration also significantly improves the model fit, as does the interaction between dialect and prosodic template. Although overall pitch excursions start earlier in the Valjevo dialect than in the Belgrade dialect, the effect of dialect is larger for rising accent words (where accentual peaks are retracted from the second syllable in the Valjevo dialect) than for falling accent words (where accentual peaks still occur in the stressed syllable in both dialect). The three-way interaction between dialect, prosodic template, and syllable onset duration also significantly improves model fit. As suggested by the individual dialect models, syllable onset duration affects the timing of the start of the excursion in H + Stress words but not the two rising accent words in the Belgrade dialect; this differs from the Valjevo dialect where syllable onset duration does not affect the start of the pitch excursion in any template. These results parallel those from Experiment 1. The full model summary is available in **Table 6**.

	β	***SE***	***t* value**	***p* value**
(Intercept)	–44.70	11.15	–4.00	0.0002***
OnsetDuration	0.12	0.09	1.29	0.20
H + Stress	28.51	10.48	2.72	0.007**
Stress-only	–7.89	16.43	–0.48	0.63
Valjevo	–62.77	16.64	–3.77	0.0005***
OnsetDuration:H + Stress	0.45	0.11	3.96	< 0.0001***
OnsetDuration:Stress-only	0.47	0.29	1.59	0.11
OnsetDuration:Valjevo	–0.03	0.14	–0.25	0.81
H + Stress:Valjevo	70.42	15.61	4.51	< 0.0001***
Stress-only:Valjevo	–22.60	24.73	–0.91	0.36
OnsetDuration:H + Stress:Valjevo	–0.38	0.17	–2.21	0.03*
OnsetDuration:Stress-only:Valjevo	–0.80	0.45	–1.78	0.808

Table 6: Full LME model for the start of pitch excursion, with the duration of the TBU syllable onset. Model: `excursionStart ~ onsetDuration*template*dialect + (1|Participant)`. *Reference Levels: template = H-only, dialect = Belgrade.*

Excursion duration The patterns of excursion duration for both dialects are illustrated in **Figure 14**. In the Belgrade dialect, the addition of prosodic template significantly improves the fit of the model. Words in the H-only (131.0 ± 9.17 ms) and the H + Stress (130.0 ± 9.16 ms) templates have the longest pitch excursions (no significant difference, $p = 0.94$); words in the Stress-only template have the shortest excursions (100.0 ± 9.16 ms). The addition of syllable onset duration also significantly improves the fit of the model; words with a longer syllable onset have longer pitch excursions ($\beta = 0.32$ ms, $SE = 0.04$ ms). The interaction between prosodic template and onset duration is also significant; syllable onset duration affects excursion duration equally in H-only and H + Stress words ($p = 0.24$), but differently in Stress-only words ($\beta = -0.64$ ms, $SE = 0.23$ ms, $p = 0.006$). It is unclear why Stress-only excursions are different than H-only and H + Stress excursions, but one possibility is that differences in the duration of the TBU syllable onset are simply from trial-to-trial differences in the production of /v/, rather than a gross change from /r/ to /ml/.

The Valjevo dialect patterns slightly differently than Belgrade in the duration of their pitch excursions. The addition of prosodic template significantly improves the fit of the model. Unlike in the Belgrade dialect, however, words in the H-only (123.0 ± 17.50 ms) and the H + Stress (126.0 ± 17.50 ms) templates have the shortest pitch excursions (no significant difference between the two, $p = 0.80$), while Stress-only words have the longest pitch excursions (146.0 ± 17.50 ms). The addition of syllable onset duration significantly improves the fit of the model; words with longer syllable onsets have longer pitch excursions ($\beta = 0.47$ ms, $SE = 0.07$ ms). The interaction between prosodic template and syllable onset duration is not significant ($\chi^2(2) = 2.98, p = 0.23$), indicating that both accents have this effect.

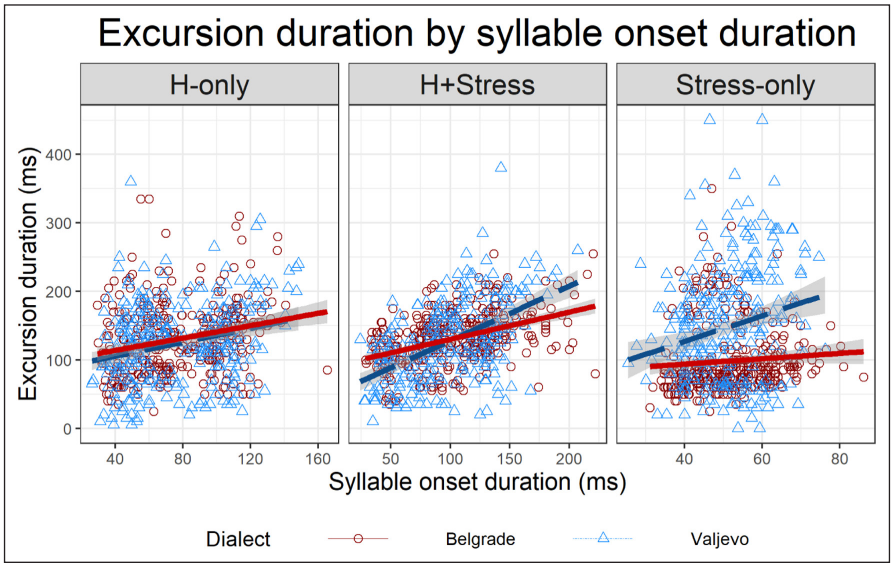


Figure 14: Scatter plots comparing the relationships between varied syllable onset durations and H offset in H + Stress, Stress-only, and H-only words (both dialects).

In a model that considers both dialects together, dialect as a fixed effect does not significantly improve the fit of the model ($\chi^2(1) = 0.59$, $p = 0.44$), suggesting that overall pitch excursions are the same duration in both dialects. However, there is a significant interaction between dialect and syllable onset duration. Adding the interaction between dialect and prosodic template also significantly improves the fit of the model, as does the three-way interaction between syllable onset duration, prosodic template, and dialect. This comes mainly from an increased effect of syllable onset duration in H + Stress ($\beta = 0.34$ ms, $SE = 0.16$ ms, $p = 0.03$) and Stress-only ($\beta = 0.92$ ms, $SE = 0.42$ ms, $p = 0.03$) words in the Valjevo dialect; there is no difference in the effect of syllable onset duration in H-only words between the Valjevo and Belgrade dialects ($\beta = –0.08$ ms, $SE = 0.13$ ms, $p = 0.52$). Rising accents in both the Belgrade and the Valjevo dialect receive timing information from the TBU, even though in the Valjevo dialect the peaks occur much earlier. A summary of the full model is provided in **Table 7**.

Taken together, these results suggest that Valjevo accentual peaks occur earlier not via shorter pitch excursions, but rather through starting the pitch excursions earlier relative to the TBU. However, in both dialects, the pitch accents receive timing information (durational in the case of Valjevo; both duration and initial timing in Belgrade) from the TBU proposed in Inkelas and Zec (1988). This indicates that Valjevo accentual peaks are merely retracted to the tonic syllable, and that rising accents have not undergone a shift to be associated to the tonic syllable.

	β	***SE***	***t* value**	***p* value**
(Intercept)	100.22	12.505	8.02	< 0.0001***
OnsetDuration	0.41	0.09	4.60	< 0.0001***
H + Stress	–0.79	9.87	–0.08	0.94
Stress-only	11.70	15.49	0.76	0.45
Valjevo	–1.48	18.78	–0.08	0.94
OnsetDuration:H + Stress	–0.11	0.11	–1.00	0.32
OnsetDuration:Stress-only	–0.64	0.28	–2.30	0.02*
OnsetDuration:Valjevo	–0.08	0.13	–0.64	0.52
H + Stress:Valjevo	–26.00	14.71	–1.77	0.08
Stress-only:Valjevo	5.21	23.31	0.22	0.82
OnsetDuration:H + Stress:Valjevo	0.34	0.16	2.14	0.03*
OnsetDuration:Stress-only:Valjevo	0.92	0.42	2.17	0.03*

Table 7: Full LME model for the duration of the pitch excursion, with the duration of the TBU syllable onset. Model: `excursionDuration ~ onsetDuration*template*dialect + (1|Participant)` *Reference Levels: template = H-only, dialect = Belgrade.*

4. Discussion

4.1. Summary

The results from these two experiments are both consistent with each other and the Inkelas and Zec (1988) account of pitch accents for both the Belgrade and Valjevo dialects. In Experiment 1, we showed that the timing of accentual peaks is influenced by the phonetic characteristics of the TBU in both dialects; in this case, the duration of the syllable onset was positively correlated with the timing of the H offset. Although accentual peaks occur earlier in the Valjevo dialect, both dialects had this positive relationship. However, the two dialects achieved this timing in different ways. In the Belgrade dialect, the pitch excursion both shifted rightward and increased in duration as syllable onset duration increased; in the Valjevo dialect, the start of pitch excursions were far more consistent across syllable onset duration, and the duration of the excursion changed more. Thus, although the segmental characteristics of the TBU provide timing information for pitch excursions in both dialects, the precise implementation of that timing information differs across dialects.

In Experiment 2, we compared the behavior of rising accents to falling accents in both dialects. The falling accents confirmed the findings from Experiment 1; peak timing is influenced by TBU characteristics in both dialects, but that timing is achieved through different timing strategies in each dialect. In both dialects, the interactions between tones and segments were parallel in rising accents when considering the segments of the post-tonic syllable, rather than the tonic syllable, indicating that the post-tonic syllable is the TBU for the lexical H in rising accents in both dialects. This supports Hypothesis 1, which is that both dialects follow the Inkelas and Zec (1988) proposal of Serbian rising accents. In this proposal, stress is one syllable to the left of a lexical H in rising accent words, and on the same syllable as the lexical H in falling accents.

Critically, the effect of the post-tonic syllable onset was found in both dialects, even though rising accent peaks in the Valjevo dialect were retracted, in some cases fully into the tonic syllable. Note that it is likely that 'true' peaks in fact occurred earlier than what was documented in this study, as H offset timing used the 20% offset velocity threshold for consistency, which places H offset timing slightly later than either gestural target achievement or gestural peak. Thus, the Belgrade and Valjevo dialects share the abstract aspects of their phonology, but do not share phonetic realization.

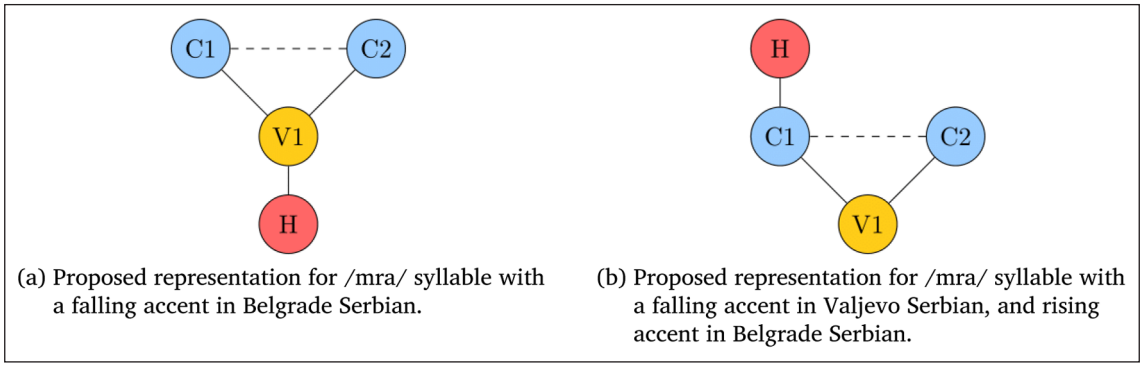
The c-center hypothesis for tone fails to predict two aspects of the data presented above. First, it is impossible for the same c-center coordinative structure to produce the two alignment strategies exhibited by the two dialects: The Belgrade dialect achieves peak alignment by changing both the timing of the start of the pitch excursion as well as the duration of the pitch excursion; the Valjevo dialect mainly utilizes the duration of the pitch excursion, keeping the onset of the gesture relatively constant. The behavior of Valjevo H gestures more closely resembles in-phase coordination between the intonational tone gesture and the syllable onset, which previously has only been documented in the small set of intonation systems examined articulatorily (Mücke et al., 2011; Prieto et al., 2007).

Second, the c-center hypothesis for tone does not predict the temporal retraction of Valjevo rising pitch accents, nor does a more general onsets-only theory of gestural coordination, as in the coupled oscillator model (Goldstein & Pouplier, 2014; Nam et al., 2009). However, it is still possible to maintain the idea that TBUs and tones interact in the temporal sphere by allowing for coordination between additional gestural landmarks—critically for Valjevo, the coordination of the H gesture target. As previously described, the Articulatory Phonology model gives priority to articulatory relationships, and articulatory coordination does not necessarily result in acoustic simultaneity. For example, complex codas are anti-phase coordinated with each other (Marin & Pouplier, 2010), but the acoustic output is sequential. Similarly, coordination between a tone gesture and segmental gestures need not necessarily result in acoustic simultaneity; in order to predict Valjevo rising accents, it is only necessary to expand the AP model of lexical tone from exclusively c-center to all modes of coordination allowed to segmental gestures under a model such as the one presented by Gafos (2002).

4.2. Modeling Serbian tones

We can now build coordinative models for both the Belgrade and the Valjevo dialects. In the Belgrade falling accent, the H gesture must be coordinated such that the start of the pitch excursion shifts rightward as the syllable onset grows longer. This is simply generated by coordinating the onset of the tone gesture with the onset of the vowel as a centering gesture, while the onset consonants are coordinated essentially in a c-center structure (onset-to-onset coordination with the vowel, and onset-to-target coordination between multiple C gestures). This differs from the c-center models proposed for Mandarin and Thai in that the tone is coordinated with the 'centering' gesture, rather than behaving as a consonant-like gesture that displaces away from the c-center. This generates the rightward shift of the tone gesture onset that is only approximately 50% of the duration of the syllable onset, as the syllable onset consonants would displace away from the onset of the H gesture in both directions, producing only a partial shift to the right of the acoustic start of the word. The timing of the gestural onset of the H gesture is affected by the c-center structure, just as the vowel gesture would be affected. This configuration is illustrated in **Figure 15**.

In contrast, the onset of the H gesture in the Valjevo falling accent was largely unaffected by syllable onset duration, a timing pattern most simply generated by an onset-to-onset coordinative relationship between the H gesture and the first consonant gesture of the syllable onset, illustrated in **Figure 15b**. This coordinative structure may also be used for the Belgrade rising accent, as the duration of the syllable onset in H-only words did not predict the timing of the onset of the H gesture, much like in the Valjevo dialect. In these studies, the timing of the F0 contours in the Valjevo falling accent and Belgrade rising accent are indeed quite similar, though it is difficult to directly compare them as stress has a major effect on syllable duration. The H gesture in Valjevo falling accents and Belgrade rising accents stretches with the duration of the syllable onset,



(a) Proposed representation for /mra/ syllable with a falling accent in Belgrade Serbian.

(b) Proposed representation for /mra/ syllable with a falling accent in Valjevo Serbian, and rising accent in Belgrade Serbian.

Figure 15: Coordinative topology models of tonal representation for Belgrade and Valjevo falling accents. The altered c-center structure in the Belgrade dialect produces the bidirectional displacement of the consonant gestures away from the tone gesture. In contrast, the Valjevo tone gesture is coordinated with the first onset consonant only and thus the onset of the tone gesture does not shift with additional consonants.

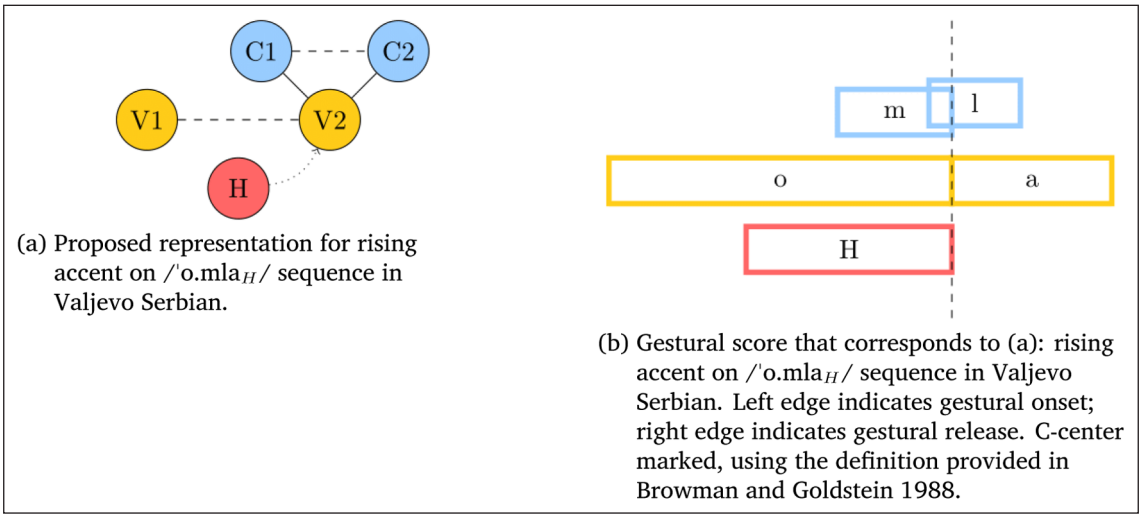
even though it is not directly coordinated with them, suggesting that, similar to the syllable- or mora-based TBUs proposed in autosegmental theories of tone, the articulatory TBU refers to the coordination of a tone gesture with a larger constellation of gestures, not the single point of gestural coordination (for more discussion, see section 4.3 below).

To capture target timing (H offset in this study), there are two possibilities. First, tone gestures can have some sort of 'base' stiffness which provides a degree of timing, but is then influenced by the gestures it is coordinated with. Alternatively, we can specify that gestures have target coordination as well as onset coordination, similarly to proposals of gestural anchoring (Ladd, 2006). The latter proposal is more appealing, as there are specific predictions: Tone targets must be timed to some other articulatory landmark (onset, target, c-center, release), and the degree of stretching then falls out naturally from the two coordinative points. In the former proposal, tone gestures would need some independent motivation to become longer when they are coordinated with more or longer gestures, which would not be shared more generally by gestures (cf. compensatory shortening in syllables with more segments, e.g., Fowler, 1981; Munhall, Fowler, Hawkins, & Saltzman, 1992). In either case, there must be some language-specific optimization of balancing the temporal demands between pitch and segmental movements. In some languages, pitch takes precedence and forces segments to be longer; in others, segmental gestures take precedence and tones will truncate or flatten. These studies do not provide definitive evidence for either of these proposals and as such the question is left for future work.

Finally, an additional model configuration is necessary for the rising accent in Valjevo Serbian, which takes advantage of two additional points of coordination as offered by Gafos (2002). The H gesture in Valjevo rising accents receives timing information from the post-tonic syllable: F0 excursions lengthen as syllable onset lengthens, and H offsets occur later with longer

syllable onsets. These patterns indicate that the H gesture is coordinated in some way to the segmental gestures of the second syllable. However, unlike falling accents in either dialect, the tone gesture must start well in advance of the post-tonic syllable, meaning that its onset cannot be coordinated to a gesture from the post-tonic syllable. A coordinative relationship between the *target* of the H gesture and the vowel gesture (which would be coordinated as the in-phase gesture of a c-center coordinative constellation) of the post-tonic syllable produces the desired phonological association and early alignment. This also produces the later peak achievement as syllable onsets get longer, similarly to how Belgrade tone gestures in falling accents start later with longer onset consonants: The c-center moves later as the syllable onset gets longer, but not by the entire duration of the onset consonants. This structure is modeled in **Figure 16**.

The relatively stable timing of the H gesture onset regardless of syllable onset duration indicates that the onset of the H gesture may be coordinated to some other gesture, but it is unclear which gesture that may be. Although an onset-to-onset relationship between the H gesture and the first consonant gesture of the syllable would produce a symmetric analysis of falling and rising accents in the Valjevo dialect, such coordination seems unlikely: Compare a mean H gesture onset – acoustic syllable onset lag of –100.9 ms (SD = 45.0 ms) in H-only rising accent words to the same in falling accents, which is only 8.5 ms (SD = 34.7 ms). One candidate may be coordination of the tone onset with the release or peak achievement of the vowel of the preceding syllable, similar to a consonant coda (anti-phase in coupled oscillator models) and a recent proposal for the second tone in Thai contour tones (Karlin, 2018). This 'dual loyalty' of the H gesture, if confirmed, would open the door



(a) Proposed representation for rising accent on /ˈo.mla$_H$/ sequence in Valjevo Serbian.

(b) Gestural score that corresponds to (a): rising accent on /ˈo.mla$_H$/ sequence in Valjevo Serbian. Left edge indicates gestural onset; right edge indicates gestural release. C-center marked, using the definition provided in Browman and Goldstein 1988.

Figure 16: Coordinative topology model of tonal representation and resulting gestural score for Valjevo Serbian rising accent, showing both the stressed and post-stress (H-bearing) syllable. The bending dotted line denotes that the gestural target of the H gesture is coordinated with the vowel gesture/c-center of the next syllable.

to three possibilities. First, it may indicate that target coordination is more important than onset coordination for abstract phonological relationships, as suggested by Turk and Shattuck-Hufnagel (2020), at least in the Valjevo dialect, though this may vary by language (cf. discussion of peak delay in Section 1.1). Second, it could be an indicator of an unstable phonological system, where such extreme retraction is liable to being reinterpreted with alternative association (consider the relative rarity of systems with very early tone targets). And finally, it may simply be an artefact of the fact that coordinating gestural onsets is more motorically stable and requires less computation than coordinating only a gestural target (Turvey, 1990). Such questions are left for future work.

4.3. Implications: An articulatory TBU

Based on these findings, we propose that an articulatory TBU corresponds to a constellation of segmental gestures that a tone gesture is included in. In this model, the existence of a coordinative relationship plays the role of phonological association and the precise parameters determine the output phonetics. To be clear, the articulatory TBU is not only the segmental gesture that the tone gesture is directly coordinated with, but rather the entire constellation in which the tone gesture participates. In the case of Serbian, the tone gesture is included in a syllable-sized constellation of gestures. Thus, a tone gesture may be directly coordinated with just the syllable onset gesture, or just the vowel gesture, but through that coordination is incorporated with the rest of the syllable as well.

Framing the articulatory TBU as the whole constellation, rather than only the gestures with direct coordinative relationships, has two advantages. First, it captures the exchange of timing information between gestures that are not directly coordinated with each other. For example, as shown in these two studies, pitch gestures shift and stretch in accordance with the duration of the entire segment onset, even in the case of complex syllable onsets, where not all of the segmental gestures have a direct coordinative relationship with the tone gesture. This conceptualization also suggests a more equitable role between segments and tone in terms of speech planning and control, rather than tones simply being 'carried' by or even 'aligned to' segments. There is mounting evidence of this prediction, in that tones can also influence temporal aspects of segments—for example, differences in duration between contour versus level tones (e.g., Mandarin) or between rising versus falling tones (Yu, 2010), segmental effects of tonal anticipatory coarticulation (Karlin, 2018), or even the insertion of segmental material in Italian (Grice, Savino, & Roettger, 2018) or Berber (Ridouane & Cooper-Leavitt, 2019).

Second, this formulation preserves the insights on phonological patterning from both traditional tonal phonology and more recent articulatory work. For example, Gao's (2008) model of Mandarin tones does not include a direct relationship between a consonant coda and the tone gestures, but does not suggest that the TBU of Mandarin should exclude the coda. Furthermore, a constellation view of the articulatory TBU does not predict that different

coordinative relationships would generate wildly different higher-level phonological patterns. That is, a Mandarin-like tone gesture that is coordinated to an onset consonant gesture and a vowel gesture is still similar in the abstract to some hypothetical language where the tone gesture is only coordinated to a vowel gesture. Given the paucity of cross-linguistic data, at this point in time there is not compelling evidence that languages with different coordinative structures for tone should have fundamentally different TBU behavior (e.g., onset consonant behaving as a TBU, with implications for phenomena such as tonal crowding or mora-based distributions). On the contrary, this paper provides evidence against that very proposal, in that Belgrade and Valjevo Serbian have the same TBU-based processes of tonal crowding despite their differences in coordinative structure (Zec & Zsiga, to appear).

5. Conclusion

Using gestural models provides tonal phonology with useful tools for expressing relationships between tones and their TBUs. The timing patterns of lexical pitch accents in the Belgrade and Valjevo dialects of Serbian indicate that an articulatory conceptualization of the TBU with potential coordination with gestural targets as well as gestural onsets better predicts existing tone data than more restrictive models that only allow the coordination of gestural onsets. Allowing coordination with the same gestural landmarks as proposed for segmental gestures both unifies the treatment of segmental versus pitch gestures and predicts a greater range of tone patterns that may be found across languages, including those observed in Valjevo Serbian. This highlights the need for more cross-linguistic work in tone under an Articulatory Phonology lens, as tones do vary in how they coordinate with segmental gestures, as demonstrated in this paper.

Acknowledgements

I would like to thank Draga Zec and Elizabeth Zsiga for their valuable insight on Serbian pitch accent, as well as Andrej Bjelaković and Biljana Čubrović for their help recruiting participants and collecting data. I am also extremely grateful to all of the speakers who agreed to participate in this research.

Competing Interests

The author has no competing interests to declare.

References

Arvaniti, A., Ladd, D. R., & Mennen, I. (1998). Stability of tonal alignment: the case of Greek prenuclear accents. *Journal of Phonetics*, *26*(1), 3–25. DOI: https://doi.org/10.1006/jpho.1997.0063

Arvaniti, A., Ladd, D. R., & Mennen, I. (2000). What is a starred tone? Evidence from Greek. *Papers in laboratory phonology V: Acquisition and the lexicon*, 119–131. DOI: https://doi.org/10.1017/S0142716401222073

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006a). Phonetic effects of focus and "tonal crowding"in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*(6), 667–696. DOI: https://doi.org/10.1016/j.specom.2005.09.012

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006b). Tonal association and tonal alignment: evidence from Greek polar questions and contrastive statements. *Language and speech*, *49*(4), 421–450. DOI: https://doi.org/10.1177/00238309060490040101

Atterer, M., & Ladd, D. R. (2004). On the phonetics and phonology of "segmental anchoring"of F0: evidence from German. *Journal of Phonetics*, *32*(2), 177–197. DOI: https://doi.org/10.1121/1.427151

Bates, D., Maechler, M., Bolker, B., Walker, S., et al. (2014). lme4: Linear mixed-effects models using Eigen and s4. *R package version*, *1*(7), 1–23. DOI: https://doi.org/10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2017). *Praat: doing phonetics by computer*. http://www.fon.hum.uva.nl/praat/

Browman, C. P., & Goldstein, L. (1988). Some notes on syllable structure in articulatory phonology. *Phonetica*, *45*(2–4), 140–155. DOI: https://doi.org/10.1159/000261823

Browman, C. P., & Goldstein, L. M. (1986). Towards an articulatory phonology. *Phonology*, *3*, 219–252. DOI: https://doi.org/10.1017/S0952675700000658

Bruce, G. (1977). *Swedish word accents in sentence perspective* (Vol. 12). Lund University.

DiCanio, C., Amith, J., & García, R. C. (2014). The phonetics of moraic alignment in Yoloxóchitl Mixtec. In *Proceedings of the 4th interntional symposium on tonal aspects of languages* (pp. 203–210).

Fowler, C. A. (1981). A relationship between coarticulation and compensatory shortening. *Phonetica*, *38*(1–3), 35–50. DOI: https://doi.org/10.1159/000260013

Fuchs, S., Perrier, P., & Hartinger, M. (2011). A critical evaluation of gestural stiffness estimations in speech production based on a linear second-order model. *Journal of Speech, Language, and Hearing Research*, *54*(4), 1067–1076. DOI: https://doi.org/10.1044/1092-4388(2010/10-0131)

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269–337. DOI: https://doi.org/10.1023/A:1014942312445

Gao, M. (2008). *Mandarin tones: An articulatory phonology account* (Unpublished doctoral dissertation). Yale University.

Geissler, C. (2019). Tonal and laryngeal contrasts in Diaspora Tibetan. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the XVIIIth International Congress of Phonetic Sciences, Melbourne, Australia* (pp. 2421–2424). Canberra, Australia.

Gibson, M., Sotiropoulou, S., Tobin, S., & Gafos, A. (2017). On some temporal properties of Spanish consonant-liquid and consonant-rhotic clusters. In *Proceedings of the conference on phonetics and phonology in German-speaking countries, berlin, germany* (pp. 73–76).

Goldsmith, J. A. (1990). *Autosegmental and metrical phonology* (Vol. 1). Basil Blackwell.

Goldstein, L., Chitoran, I., & Selkirk, E. (2007). Syllable structure as coupled oscillator modes: evidence from Georgian vs. Tashlhiyt Berber. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 241–244).

Goldstein, L., & Pouplier, M. (2014). The temporal organization of speech. In M. Goldrick, V. Ferreira, & M. Miozzo (Eds.), *The oxford handbook of language production* (pp. 210–227). Oxford University Press.

Grice, M., Savino, M., & Roettger, T. B. (2018). Word final schwa is driven by intonation—the case of Bari Italian. *The Journal of the Acoustical Society of America*, *143*(4), 2474–2486. DOI: https://doi.org/10.31234/osf.io/e9cug

Hoole, P., Mooshammer, C., & Tillmann, H. G. (1994). Kinematic analysis of vowel production in German. In *Third international conference on spoken language processing.*

Hyman, L. M. (1988). Syllable structure constraints on tonal contours. *Linguistique Africaine*, *1*, 49–60.

Inkelas, S., & Zec, D. (1988). Serbo-Croatian pitch accent: the interaction of tone, stress, and intonation. *Language*, 227–248. DOI: https://doi.org/10.2307/415433

Kanerva, J. M. (1990). *Focus and phrasing in Chichewa phonology* (Unpublished doctoral dissertation). Stanford University.

Karlin, R. (2014). *The articulatory TBU: Gestural coordination of lexical tone in Thai*. (Cornell Working Papers in Phonetics and Phonology). DOI: https://doi.org/10.1121/2.0000089

Karlin, R. (2018). *Towards an articulatory model of tone: a cross-linguistic investigation* (Unpublished doctoral dissertation). Cornell University.

Katsika, A., Krivokapić, J., Mooshammer, C., Tiede, M., & Goldstein, L. (2014). The coordination of boundary tones and its interaction with prominence. *Journal of Phonetics*, *44*, 62–82. DOI: https://doi.org/10.1016/j.wocn.2014.03.003

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2015). Package 'lmerTest'. *R package version*, *2*(0).

Ladd, D. R. (2006). Segmental anchoring of pitch movements: Autosegmental association or gestural coordination? *Italian Journal of Linguistics*, *18*(1), 19.

Ladd, D. R. (2008). *Intonational phonology*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511808814

Ladd, D. R., Faulkner, D., Faulkner, H., & Schepman, A. (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. DOI: https://doi.org/10.1121/1.427151

Laniran, Y. O. (1992). *Intonation in tone languages: the phonetic implementation of tones in Yoruba* (Unpublished doctoral dissertation). Cornell University.

Lehiste, I., & Ivić, P. (1986). *Word and sentence prosody in Serbocroatian*. MIT Press.

Lenth, R. (2019). emmeans: Estimated marginal means, aka least-squares means [Computer software manual]. Retrieved from https://CRAN.R-project.org/package=emmeans (R package version 1.3.5.1)

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*(3–4), 211–227. DOI: https://doi.org/10.1016/j.wocn.2013.02.001

Marin, S., & Pouplier, M. (2010). Temporal organization of complex onsets and codas in American English: testing the predictions of a gestural coupling model. *Motor Control*, *14*(3). DOI: https://doi.org/10.1123/mcj.14.3.380

Martinuzzi, C., & Schertz, J. (2021). Sorry, not sorry: The independent role of multiple phonetic cues in signaling the difference between two word meanings. *Language and Speech*, 0023830921988975. DOI: https://doi.org/10.1177/0023830921988975

McAuliffe, M., Socolof, M., Mihuc, S., Wagner, M., & Sonderegger, M. (2017). *Montreal forced aligner*. http://montrealcorpustools.github.io/Montreal-Forced-Aligner/

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory*, *24*(1), 113–178. DOI: https://doi.org/10.1007/s11049-004-5454-y

Mücke, D., Grice, M., Becker, J., & Hermes, A. (2009). Sources of variation in tonal alignment: evidence from acoustic and kinematic data. *Journal of Phonetics*, *37*(3), 321–338. DOI: https://doi.org/10.1016/j.wocn.2009.03.005

Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2011). Coupling of tone and constriction gestures in pitch accents. *Consonant Clusters and Structural Complexity*, *26*. DOI: https://doi.org/10.1515/9781614510772.205

Munhall, K., Fowler, C. H., Hawkins, S., & Saltzman, E. (1992). "Compensatory shortening"in monosyllables of spoken English. *Journal of Phonetics*, *20*, 225–239. DOI: https://doi.org/10.1016/S0095-4470(19)30624-2

Myers, S. (1999). Tone association and f0 timing in Chichewa. *Studies in African Linguistics*, *28*(2). DOI: https://doi.org/10.32473/sal.v28i2.107375

Myrberg, S. (2010). *The intonational phonology of Stockholm Swedish* (Unpublished doctoral dissertation). Acta Universitatis Stockholmiensis.

Nam, H., Goldstein, L., & Saltzman, E. (2009). Self-organization of syllable structure: A coupled oscillator model. In *Approaches to phonological complexity* (pp. 297–328). De Gruyter Mouton. DOI: https://doi.org/10.1515/9783110223958.297

Odden, D. (1995). Tone: African languages. In J. A. Goldsmith (Ed.), (Vol. 1, pp. 444–475). Oxford: Blackwell.

Pierrehumbert, J. B. (1980). *The phonology and phonetics of English intonation* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Pierrehumbert, J. B., & Steele, S. A. (1989). Categories of tonal alignment in English. *Phonetica*, *46*(4), 181–196. DOI: https://doi.org/10.1159/000261842

Prieto, P. (2011). Tonal alignment. In M. van Oostendorp, C. J. Ewen, E. V. Hume, & K. Rice (Eds.), *The Blackwell companion to phonology* (pp. 1185–1203). Blackwell Publishing. DOI: https://doi.org/10.1002/9781444335262.wbctp0050

Prieto, P., D'Imperio, M., & Fivela, B. G. (2005). Pitch accent alignment in Romance: primary and secondary associations with metrical structure. *Language and Speech*, *48*(4), 359–396. DOI: https://doi.org/10.1177/00238309050480040301

Prieto, P., Mücke, D., Becker, J., & Grice, M. (2007). Coordination patterns between pitch movements and oral gestures in Catalan. In *Proceedings of the XVIth International Congress of Phonetic Sciences* (pp. 989–992). Pirrot GmbH: Dudweiler.

Prieto, P., & Torreira, F. (2007). The segmental anchoring hypothesis revisited: Syllable structure and speech rate effects on peak timing in Spanish. *Journal of Phonetics*, *35*(4), 473–500. DOI: https://doi.org/10.1016/j.wocn.2007.01.001

R Core Team. (2019). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from https://www.R-project.org/

Remijsen, B. (2013). Tonal alignment is contrastive in falling contours in Dinka. *Language*, *89*(2), 297–327. DOI: https://doi.org/10.1353/lan.2013.0023

Remijsen, B., & Ayoker, O. G. (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology*, *31*(3), 435–462. DOI: https://doi.org/10.1017/S0952675714000219

Ridouane, R., & Cooper-Leavitt, J. (2019). A story of two schwas: a production study from Tashlhiyt. *Phonology*, *36*(3), 433–456. DOI: https://doi.org/10.1017/S0952675719000216

Roettger, T. (2017). *Tonal placement in Tashlhiyt: How an intonation system accommodates to adverse phonological environments*. Language Science Press. DOI: https://doi.org/10.5281/zenodo.814472

Ruthan, M., Durvasula, K., & Lin, Y.-H. (2018). Temporal coordination and sonority of Jazani Arabic word-initial clusters. In K. Hout, A. Mai, A. McCollum, S. Rose, & M. Zaslansky (Eds.), *Proceedings of the 2018 Annual Meeting on Phonology*. DOI: https://doi.org/10.3765/amp.v7i0.4485

Sagey, E. (1986). *The representation of features and relations in autosegmental phonology* (Unpublished doctoral dissertation). MIT.

Schweitzer, K., Walsh, M., Calhoun, S., Schütze, H., Möbius, B., Schweitzer, A., & Dogil, G. (2015). Exploring the relationship between intonation and the lexicon: Evidence for lexicalised

storage of intonation. *Speech Communication*, *66*, 65–81. DOI: https://doi.org/10.1016/j.specom.2014.09.006

Selkirk, E., & Durvasula, K. (2013). *Acoustic correlates of consonant gesture timing in English*. Acoustical Society of America. DOI: https://doi.org/10.1121/1.4831423

Shaw, J. A., & Chen, W.-r. (2019). Spatially-conditioned speech timing: evidence and implications. *Frontiers in Psychology*, *10*, 2726. DOI: https://doi.org/10.3389/fpsyg.2019.02726

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. *Papers in laboratory phonology I*, 72–106. DOI: https://doi.org/10.1017/CBO9780511627736.005

Smiljanić, R. (2002). *Lexical, pragmatic and positional effects on prosody in two dialects of Croatian and Serbian: An acoustic study* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign.

Turk, A., & Shattuck-Hufnagel, S. (2020). Timing evidence for symbolic phonological representations and phonology-extrinsic timing in speech production. *Frontiers in Psychology*, *10*(2952), 1–20. DOI: https://doi.org/10.3389/fpsyg.2019.02952

Turvey, M. T. (1990). Coordination. *American psychologist*, *45*(8), 938–953. DOI: https://doi.org/10.1037/0003-066X.45.8.938

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*(1–2), 26–52. DOI: https://doi.org/10.1159/000028487

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech communication*, *46*(3), 220–251. DOI: https://doi.org/10.1016/j.specom.2005.02.014

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from Mandarin Chinese. *Speech communication*, *33*(4), 319–337. DOI: https://doi.org/10.1016/S0167-6393(00)00063-7

Yi, H. (2014). A gestural account of Mandarin tone sandhi. *The Journal of the Acoustical Society of America*, *136*(4), 2144–2144. DOI: https://doi.org/10.1121/1.4899737

Yi, H. (2017). *Lexical tone gestures* (Unpublished doctoral dissertation). Cornell University.

Yip, M. (1989). Contour tones. *Phonology*, *6*(1), 149–174. DOI: https://doi.org/10.1017/S0952675700000097X

Yip, M. (2002). *Tone*. Cambridge University Press. DOI: https://doi.org/10.1017/CBO9781139164559

Yu, A. C. (2010). Tonal effects on perceived vowel duration. In C. Fougeron, B. Kuehnert, M. D'Imperio, & N. Vallee (Eds.), *Laboratory phonology 10* (pp. 151–168). Mouton de Gruyter. DOI: https://doi.org/10.1515/9783110224917.2.151

Zec, D. (2005). Prosodic differences among function words. *Phonology*, *22*(1), 77–112. DOI: https://doi.org/10.1017/S0952675705000448

Zec, D., & Zsiga, E. (2016). *(talk) A new typology of tone and stress interactions*. Manchester Phonology Meeting 24.

Zec, D., & Zsiga, E. (to appear). Tone and stress of agents of cross-dialectal variation: the case of Serbian. In H. Kubozono, J. Ito, & A. Mester (Eds.), *Prosody and prosodic interfaces*. Oxford University Press.

Zhang, M., Geissler, C., & Shaw, J. (2019). Gestural representations of tone in Mandarin: evidence from timing alternations. In S. Calhoun, P. Escudero, M. Tabain, & P. Warren (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 1803–1807).