TOWARDS AN ARTICULATORY MODEL OF TONE: A CROSS-LINGUISTIC INVESTIGATION

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Robin Park Karlin

December 2018

TOWARDS AN ARTICULATORY MODEL OF TONE: A CROSS-LINGUISTIC INVESTIGATION

Robin Park Karlin, Ph.D.

Cornell University 2018

This thesis examines the role of timing information in phonological representation, focusing on how tone aligns with segmental material. On the basis of three acoustic studies, I present a novel gestural model of tone representation, where tone gestures are durationally underspecified and receive their timing information from the constellation of segmental gestures they are coordinated with. I also argue that the distributional and temporal characteristics of tone are the direct result of gestural coordination: phonological association can be analyzed as the existence of coordinative relationships between a tone gesture and a constellation of segmental gestures, and the precise nature of that coordinative relationship produces the cross-linguistically variable acoustics.

Chapter 1 delineates two major approaches to tone representation: Autosegmental(-Metrical), which references point-like features and nominal time, and Articulatory Phonology, which uses gestures that unfold over time and space. I discuss the different ways in which each theory arrives at phonetic realization from underlying representation, as well as their differing notions of overlap.

Chapter 2 presents the results of an acoustic study on contour tones in Thai (Tai-Kadai), a tone language where the mora serves as a licensing unit for tone. Counter previous hypotheses, tonal extrema do not map to moraic edges: both moraic edges and tone extrema instead independently refer to the syllable as a unit of timing. Based on these results, I argue that the apparent acoustic mismatches can be straightforwardly derived from the application of different coordinative modes between tonal and segmental gestures, while maintaining the phonological licensing relationship between moras and tones. The data also suggests that the

F0 targets of tone gestures play a role in tone timing, indicating that there is an interaction between the tone gesture and segmental gestures when determining duration.

Chapter 3 presents the results of an acoustic study focused on the realization of the falling accent in the Belgrade and Valjevo dialects of Serbian (Indo-European—Slavic), a tone language where one syllable per word is specified for tone. I compare the alignment and duration of pitch excursions across varying phonetic and phonological properties of the syllable onset of the tone-bearing syllable. I show that the duration of pitch excursions increases with phonetically longer syllable onsets, which indicates that tone gestures are durationally underspecified and receive their timing information from the constellation of segmental gestures they are coordinated with. The two dialects also exhibit distinct patterns of both the duration and the alignment of the pitch excursion, and I argue that this is due to differences in the type of coordination used.

Chapter 4 focuses on the rising accents of the same dialects of Serbian, crucially examining the Valjevo dialect, which routinely retracts the pitch peak into the syllable preceding the H-bearing syllable. Despite this phonetic retraction, the patterns of alignment parallel those observed for the falling accent. This indicates that the tone gesture is still receiving timing information from the H-bearing syllable, and as such is still coordinated to it. Based on these results, I argue for the availability of gestural target coordination, in addition to gestural onset coordination.

In Chapter 5 I synthesize the findings from the experimental chapters to present a gestural model of tone representation, and discuss its implications for Articulatory Phonology and avenues for future research.

Biographical Sketch

Robin Karlin was born in Oshkosh, Wisconsin, on February 8th. In May 2012, she received Bachelor of Arts degrees in Linguistics, Psychology, and Spanish literature from the University of Wisconsin-Madison. She began as a graduate student at Cornell University in Fall 2012.

To those in academia who, nevertheless, persisted;

and to those who believed them.

Acknowledgements

As I have written an entire dissertation on the importance of timing, and of trajectories overlapping at specific times, it seems only fitting that I should thank the people who have overlapped with my life at key times. Chronologically first of course is my family: my sister, who I have always looked up to; and my parents, who bestowed upon me a genetic predisposition for loving languages, as well as a fondness for crafting arguments. Without them I would not be where—or who—I am today.

My academic trajectory was launched in a fieldwork class on Thai, where I crossed paths with Marlys Macken. Before this class, I had declared myself “afraid” of tone languages and, moreover, dedicated to syntax. It is now obvious how large an impact that fieldwork class had on my future trajectory: not only did I fully convert to phonology and phonetics, but I conquered my fear of tone languages and have actively sought them out ever since.

My quest for a greater understanding of tone languages led me to Cornell, for which I must acknowledge Tom Purnell, whose suggestions completely altered my pool of potential graduate schools and set Cornell at the top of my list. At Cornell, I have had the great fortune to work closely with four wonderful scholars, and I am extremely grateful for my experience with them. Draga Zec planted the initial seed of interest in Serbian, and somehow shows me how to reshape the most clunky of analyses and abstracts into something elegant; Abby Cohn has made sure that I consider the clarity and explicitness of my argument, and provides suggestions for all the literature I could ever need; Mats Rooth always asks thought-provoking questions that keep me on my toes; and Elizabeth Zsiga always gives thoughtful

feedback on my new ideas and keeps my analyses from running off into the brush. In addition to my committee, I would also like to thank Wayles Browne, Miloje Despić, Molly Diesing, Sarah Murray, Joe Pittayaporn, and John Whitman, all of whom have provided invaluable wisdom throughout my time here. I would also be remiss if I did not particularly thank Holly Boulia, who made sure I did not careen off an administrative cliff, as well as Gretchen Ryan and Jenny Tindall, who took over that unenviable task in my last year.

My dissertation—and my research more generally—would not exist without the multitudes of people who have agreed to be recorded, both at home in the P-lab and further afield. A full-hearted thank-you goes to all the people over the years who willingly shut themselves in a windowless room to have, at best, a microphone shoved in their face—or at worst, sensors glued inside their mouths and wires taped to their cheeks. An especial thank-you is due to my labmate Siree, who unflinchingly endured this last torture not just once but twice in one day due to my coding failures. I also owe an entire book of thank-yous to everyone in Belgrade who helped me conduct an incredible 80 total recording sessions over two visits to the Fakultet, including all the students who missed class to say nonsense words at me and their teachers for being so accommodating, and especially Biljana Čubrović for arranging it all and allowing me to stay in her home.

To Biljana I also owe thanks for introducing me to Andrej Bjelaković, who has been invaluable as both a friend and a colleague since my first trip to Belgrade. Andrej's first act when we met was to wrangle participants, to the tune of 58 participants in five days—surely a record-breaking performance. Since then he has patiently explained several aspects of Serbian dialectology to me, complied with countless requests for him to listen to such and such sound clip, and prodded me to stop procrastinating on my drafts. His kindness ultimately culminated in agreeing to conduct my second Serbian experiment on my behalf, for which I am eternally grateful. Baš si divan; puno ti hvala!

Closer to home, my trajectory intersected with a whole host of other graduate students, from Cornell and further afield. I would like to extend my heartfelt thanks to Emily, who saw to my non-linguistic education; Christina, who always pushes for a more thought-provoking analysis and a more delicious recipe; Chelsea, who is a constant friend, providing recipes, encouragement, and a voice of reason at every turn; Todd, who was the inspiration behind my last push to get my dissertation done; and Hao, who was always happy to discuss tones, gestures, statistics, and tennis. I would also like to thank Andrea, Anna, Carol-Rose, Chris, Dan, Eleanor, Emily C., Francesco, Geethika, Mary, Mia, Naomi, Okki, Rina, Ryan, Simone, Siree, Timo, Yang, Yu Yu, and Zac for years of discussions and support, and not a few meals. Grad school would have been much worse without all of you and I am extremely grateful for your friendship.

Finally, a network of incredible Ithacans made it possible for me to explore my non-academic interests, and to these friends I owe my mental and physical well-being. I am forever indebted to Andrew, who boosts my confidence and always has an idea for a new adventure; Meredith, who is always game for ice cream and a good ballet chat; Erika and Dan, who have made possible every horse-related field trip I could think of, and then some; Raina, who was among the first of the established company members to actively talk to me; Lauren, who is both an amazing mother and second mother to all of us at the studio; Cindy, who invited me to dance with the company and got me back en pointe after nine years; Hilary, with whom I have been a horse, a snowflake, and an authentically crying old lady; and Maria, who transfers her boundless energy into her friends and students. I also have to thank Alex, Alison, Anya, Ayla, Claire, Ellie, Finn, Jiayi, Keara, Kevin, Lauren S., Lucy, Lydia, Maddie, Miranda, Rachel, Sage, T. Rose, Vivian, and Xin for creating such a welcoming environment at the ballet studio—that can be a rarity and I am so grateful to have met you all.

Contents

Chapter 1

Introduction

This thesis examines the role of timing information in phonological representation, with a special focus on the alignment of tones and segmental material. I adopt the Articulatory Phonology perspective and treat tones as gestures that unfold over time, and investigate how the coordination of tone gestures with the segmental gestures reflects phonological relationships and phonetic properties.[1] I explore these issues through three acoustic experiments on two unrelated tone languages, Thai and Serbian, specifically probing inter- and intra-language variation in peak alignment.

1.1 Features, gestures, and tone timing

There are two main philosophies in generative theories of phonological representation, where the main distinguishing characteristic is how to include time in phonological representation. In frameworks that I will refer to as "featural", the use of time is limited. Features are abstracted away from time and treated as point values, as opposed to being grounded in time and space. Units such as segments are represented in linear order, as they are realized in time, and overlap is limited to units in different **tiers**. For example, tones and segments are regarded as being on separate tiers (Goldsmith 1976, 1990), and as such can

[1] Throughout this dissertation I will be using the term "phonological" to refer to qualitative, categorical patterns and relationships, and "phonetic" to refer to the quantitative, gradient properties of the physical realization.

occur simultaneously—i.e., a tone can be realized at the same time as a vowel. However, two tones cannot overlap each other, and neither can two segments.

In contrast, gestural models such as Articulatory Phonology (AP) are based on the idea that timing relationships between articulatory gestures are essential to phonological representation. **Articulatory gestures** are "characterizations of discrete, physically real events that unfold during the speech production process" (Browman & Goldstein 1992), and as such are anchored in both time and space. Individual gestures are combined via timing relationships into "constellations" (Browman & Goldstein 1989), which approximate variously sized units in the prosodic hierarchy. For example, a constellation involving lowering the velum and closing the lips forms the segment-like unit for /m/. Overlap between segmental gestures is assumed as a fact of motor control, and by extension, articulation. There is no tier-based restriction on gestural overlap; for example, a high degree of overlap is modeled for the CV syllable, where the C and V gestures are executed with in-phase coordination (i.e., the C and V gestures start at the same time).

One of the main objectives for any theory of representation is to separate the phonetic variability from systematic, meaningful differences. In tone, the phonetics-phonology interface is particularly entangled on the issue of timing. While current articulatory models do not predict as much variability in timing as is empirically present in lexical tone languages, featural models tend to delegate variation into the phonetics, thus putting very little restriction of the possible types of tone timing. Making this issue more complex is the concept of a "tone-bearing unit", a segmental unit that is frequently posited to address both phonological distribution and phonetic alignment. In this section, I provide a brief introduction to the major theories of tone and tone timing.

1.1.1 Featural models of tone

1.1.1.1 Autosegmental representation

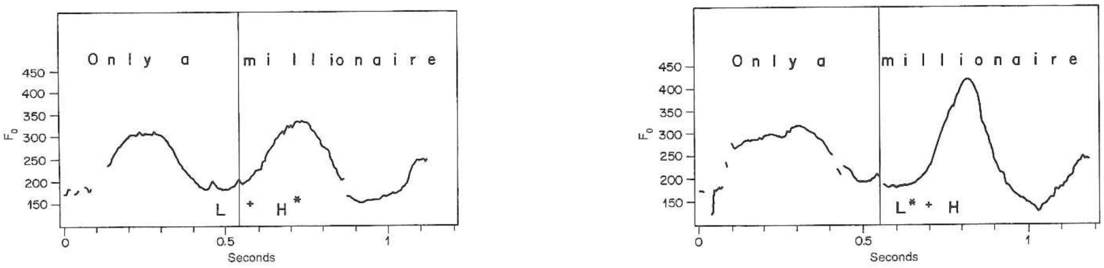
By far the most well-known member of the featural philosophy is Autosegmental(-Metrical) theory (AM). This theory builds on the concept of **suprasegmentals**, which are distinctive elements that occur over units longer than the segment, developed from a need to describe tones as independent from individual segments, rather than as part of the segmental featural bundle (Leben 1973). This separation was driven variously by patterns in English intonation (Liberman 1975), African tone languages (Goldsmith 1976; Leben 1973, inter alia), and Swedish pitch accent (Bruce 1977). Liberman describes the intellectual process of separating the "tune" of English intonation from the "text" as follows:

> "For example, I associated the tonal aspect of an utterance too closely with its "textual" aspect—one breakthrough came when I became willing to conceive of the "tune" as an entity which is in origin completely independent of the "text"; not an aspect of the features of the segmental string, not a set of suprasegmental diacritics, not even a separate string of segments, but a *completely independent structure.* Each step in this progression brought progress; each represented a more abstract conception of the nature of tonal phenomena in English." (Liberman 1975, p. 7a)

Autosegmental theory thus replaced features like [+high], [+low], and [+fall] with single tone targets, H(igh) and L(ow), or combinations thereof (i.e., a fall is composed of an H+L, not a single Fall tone). Typically, words are lexically specified for tone, but the phonetic realization and alignment of tones with the segmental material is determined through two stages. First, the **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. In this use of the TBU, phonological distribution plays a major role in determining the TBU for a particular language. For example, the four Mandarin tones can occur on any non-checked 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, contour tones in Thai (i.e., HL and LH tones) can only occur on words with two sonorant moras, while the simple tones H and L can occur on words with just one sonorant mora, which indicates that the mora is the TBU in Thai (Morén & Zsiga 2006).

Second, the phonetic realization of tones is determined by phonetic mapping rules that reference the association lines between tones and segmental structures. For example, the "right alignment" of tones in Thai (where tonal targets occur near the end, or right edge, of the moraic TBU) is included in the phonetic mapping rules, rather than represented in the phonology (Morén & Zsiga 2006). Some timing information may be incorporated into the tone targets themselves—for example, contrasts between late and early pitch targets are frequently included in the representation, using the **star convention**, which indicates which tone is aligned with the stressed syllable. For example, Pierrehumbert (1980) described two distinct rises in English, an early rise, denoted as L+H*, and a late rise, denoted as L*+H. 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 stress syllable, and the transition to H trails after (see Figure 1.1). Examples in lexical tone languages include Smiljanić's (2002) proposal of two pitch accents for Serbian, L+H* and L*+H, which represent early (falling accent) and late (rising accent) F0 rises, respectively, as well as Myrberg's (2010) account of Swedish word accent (in alignment with Bruce (1977), who did not specifically utilize the star convention), which uses H+L* for early (accent I) and H*+L for late (accent II) F0 peaks.

However, as noted by Atterer and Ladd (2004), the use of the star convention has expanded past denoting contrast, instead roughly capturing the previously mentioned early and late alignments in a purely phonetic sense. Arvaniti, Ladd, and Mennen (2000) also ar-



(a)

(b)

Figure 1.1: Figures showing the difference in alignment between L+H* (early rise) and L*+H (late rise) pitch accents in English, reproduced from Pierrehumbert and Steele 1989.

gue that the star convention is “ill-defined”, citing the empirically unsupported tendency to use phonetic alignment as the basis for positing phonological association and licensing. They argue that phonetic alignment does not in all cases correspond to phonological association, and vice versa, and provide data from Greek, where neither the L target nor the H target of a bitonal L+H accent phonetically aligns with the stressed syllable: the L target occurs before the stressed syllable, and the H target occurs after. That is, only the F0 transition, not the targets, occur during the syllable the accent is associated with.

It is not uncommon for pitch targets to occur outside their TBU; however, there is an asymmetry in when pitch targets occur relative to their TBU. Typically this “misalignment” is late—i.e., pitch targets more frequently occur after the TBU than before the TBU. Late peaks have been attributed to **peak delay**, a phenomenon where pitch targets are achieved after the syllable (or other unit) they are associated to (Xu 2001). Peak delay has been documented in several languages, including lexical tone languages (see Myers 1999 for Chichewa; Xu 2001 for Mandarin; Morén and Zsiga 2006 for Thai), as well as in intonation-only languages (see Silverman and Pierrehumbert 1990 for English; Arvaniti, Ladd, and 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). According to the PENTA model, pitch targets are "implemented in synchrony with the host [TBU]" (Xu and Wang 2001, p. 322), but landmarks such as pitch peaks and valleys are "delayed due to inertia" until after the end of the TBU (Xu 2004, p. 95). Thus, this delay is specifically not related to other pressures in the string, such as insufficient segmental material for full pitch target realization, or several preceding tones in the tonal string. This model predicts that all pitch turning points are realized after the TBU.

Despite the ubiquity of peak delay, phonetic alignment of pitch extremes is still one of the main sources of evidence for the association of tones to segmental structures. In some cases, this has led to phonological analyses of phonetic phenomena, such as tone "spreading" in Chichewa (Myers 1999). Chichewa was analyzed to have high tone spreading, where a "high tone spreads rightward onto the following syllable only if the high tone is not in the last three syllables of a phrase" (p. 215). This spreading was described as phonological, and the high tone "associated with two syllables" (ibid.). However, Myers showed that the phonetic overlap between the F0 peak and the rightward mora is simply a case of peak delay, not re-association. He also speculated that this was the case for several other Bantu languages with similar analyses.

There are comparatively fewer cases of an early peak—in fact, the rarity is such that there is no term generally used to describe this phenomenon. In the phenomenon **tonal crowding**, peaks occur early relative to the TBU; however, it 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 2006). This is distinct from early peaks that occur without additional time pressure, such as those described by 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 (emphasis mine):

> ACCENT I: HIGH in the **pre-stress syllable**, LOW in the stressed syllable.

ACCENT II: HIGH in the **stressed syllable**, LOW in the post-stress syllable.

The case of Swedish accent I is quite “extreme”, in a sense, as accent I not only occurs before the stressed syllable it is assigned to, but sometimes before the word:

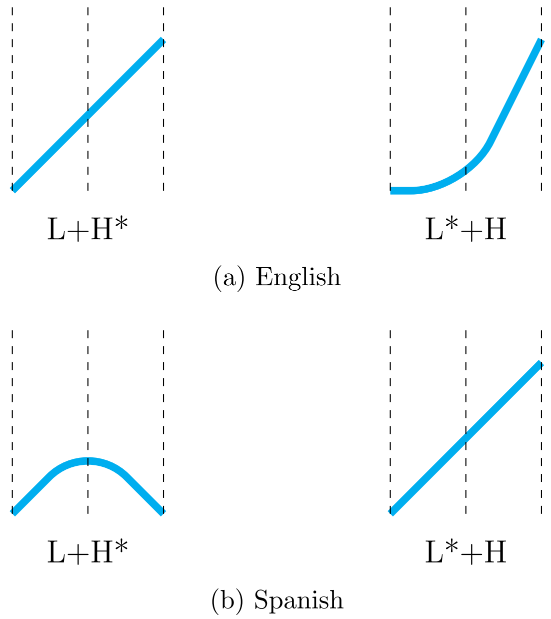
> **For the accent I-word the peak occurs as early as in the pre-stress syllable**, even if this syllable belongs to a preceding word... This is a first indication that the prosodically relevant F0-contributions do not necessarily relate to the domain of the word.

Myrberg (2010) (after Bruce 1977) addressed the alignment differences in the Swedish accentual system in the AM framework, using 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.

1.1.1.2 Phonetic mapping rules

In AM representations, phonological processes address the association of tunes to segmental units, and another set of rules is necessary for mapping from the phonology to the phonetics. One of the most widely investigated hypotheses in the AM framework is the **segmental anchoring hypothesis**, first formally proposed by Ladd, Faulkner, Faulkner, and Schepman (1999). Segmental anchoring hypothesizes an alignment of tunes to segmental material, which remains stable in the face of pressures such as speech rate. It specifically hypothesizes that both the start and the target of a pitch excursion are anchored to the segmental string—that is, for any rise or fall in F0, there is an anchoring point for the minimum as well as for the maximum (Prieto 2011).

One consequence of this type of anchoring is that hypotheses that include the concept of an invariant rise are not possible—and this is supported by empirical data. In general terms, as more segmental material comes between the anchoring points for the start and the end of a pitch movement, the longer the pitch movement is. For example, Dilley, Ladd,



(a) English

(b) Spanish

Figure 1.2: Schemata of L+H* and L*+H in English (a) and Spanish (b), adapted from Prieto 2011. Note that the Spanish L*+H and English L+H* correspond to the same F0 contour, while the same tone transcription L*+H in each language represents a different phonetic contour.

and Schepman (2005) found that pitch excursions in the bitonal L+H* accent in English increased in duration when the stressed (associated) syllable had a syllable onset compared to no onset.

These mapping rules are highly idiosyncratic to individual languages and thus it is difficult—if not impossible—to predict the phonetic form from underlying tones. This idiosyncracy also combines with the star convention to produce contours that are not cross-linguistically consistent, given the same underlying representation. One example of the phonetic consequences of this is provided by Prieto (2011), who notes that the same F0 contours can have different transcriptions. She gives the example of English L+H* and Spanish L*+H (illustrated in Figure 1.2), where a single label (for example, L*+H) does not correspond to a single contour, nor does a single contour correspond to a single pitch accent.

For tone languages, there has been less emphasis on segmental anchoring, instead utilizing the right and left edges of the distributional TBU for mapping phonetic alignment. For example, Morén and Zsiga (2006) argued that the distributional TBU in Thai (the mora) is responsible for phonetic alignment, in that the phonetic mapping rules specify that the tone associated to the TBU is exceptionlessly realized at the right edge of that TBU. Similarly, in Yolóxochitl Mixtec, which allows contours within a single mora, the issue is simply one of the mora as the TBU, which can have full tone melodies associated to it (DiCanio, Amith, & García 2014); the phonetic alignment of the tones then align to the moraic boundaries.

In some cases, it has been necessary to include some alignment information in the tones themselves, due to the existence of contrasts in alignment within a TBU—though the edges of TBUs are still referenced. Based on a lack of evidence for it, contrastive alignment within a syllable was hypothesized to not be possible (Hyman 1988, as cited in Remijsen and 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. Remijsen and Ayoker (2014) proposed a feature [±late-aligned] that would distinguish two HL contours, which refers to the relative timing of the first tone target in the contour; alternatively, they argue that an H tone that has a distinctive feature alignment to the right or left edge of the first mora (where the first vowel without its syllable onset is considered the first mora) could also produce the desired result.

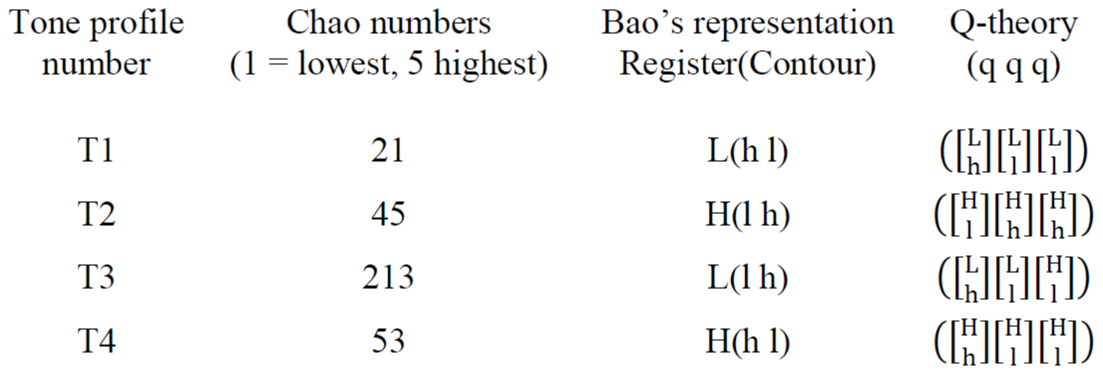
1.1.1.3 Q theory (ABC+Q for tone)

More recently, Shih and Inkelas (to appear, 2019) have proposed a representation of tone (with extension to other autosegmental phenomena) that includes a more granular linearization of time than is provided in typical featural representations, called Q theory. One of the arguments for this theory is in fact the existence of contours (both tonal and segmental in the Sagey 1986 sense) that contrast only by temporal alignment. In Q theory, each segment Q is composed of (maximally) three subsegments, q_1, q_2, q_e. Each subsegment can be specified with a different value: for example, an early fall (à la Shilluk) would be represented with Q{H L L}, while a late fall would be represented with Q{H H L}. In this representation, it is the subsegment, not the segment, that is the TBU.

It is unclear how the subsegments-as-TBU correspond to TBUs in the traditional autosegmental sense. Shilluk is cited as an example of a language that motivates the proposal for three subsegments; however, in Remijsen and Ayoker's (2014) analysis, the mora was posited as the TBU, and the entire rime (i.e., short vowel + sonorant coda) was recruited. If the "segment" Q is to correspond with the rime, this representation would work, though this does not seem likely given that a simple /t/ is represented as Q{t t t}. If, on the other hand, one Q (each with three subsegments) is posited for each segment, then there is a total of six subsegments in the rime, which overgenerates the number of contours demonstrated to be contrastive in tone languages (thus far).

Furthermore, although the number of subsegments stems from articulatory approaches, in that Shih and Inkelas draw a comparison to the gestural composition of onset—target—offset, it is unclear to what extent timing relationships are to be portrayed in this representation, and how much is left to the phonetics. For example, in the proposed representation of tones in Tianjin Mandarin, both tones that are phonetically described as having three distinct levels and tones described with just two are represented with three subsegments, with an apparent default of Q{H L L} and Q{L H H} as opposed to Q{H H L} or Q{L L H} (see Figure 1.3). Tone 3, evidently a contour with three targets, does have three distinct subsegments.

Moreover, in allowing multiple targets per *Q* (in that each subsegmental *q* can be distinctly represented), Q theory loses its relationship with Articulatory Phonology and its analogy to the onset, target, and release of a gesture. For example, a level high tone would be represented with an H at every subsegment; in accordance with the analogy with gestural phases, this high tone would have an onset, target, and release. However, the onset of a high tone as measured in an articulatory study would be from a low point in the F0 contour. Thus, although it posits a more detailed representation of linear time, quantal representation



Tone profile number	Chao numbers (1 = lowest, 5 highest)	Bao's representation Register(Contour)	Q-theory (q q q)
T1	21	L(h l)	$(\begin{bmatrix}L\\h\end{bmatrix}\begin{bmatrix}L\\l\end{bmatrix}\begin{bmatrix}L\\l\end{bmatrix})$
T2	45	H(l h)	$(\begin{bmatrix}H\\l\end{bmatrix}\begin{bmatrix}H\\h\end{bmatrix}\begin{bmatrix}H\\h\end{bmatrix})$
T3	213	L(l h)	$(\begin{bmatrix}L\\h\end{bmatrix}\begin{bmatrix}L\\l\end{bmatrix}\begin{bmatrix}H\\l\end{bmatrix})$
T4	53	H(h l)	$(\begin{bmatrix}H\\h\end{bmatrix}\begin{bmatrix}H\\l\end{bmatrix}\begin{bmatrix}H\\l\end{bmatrix})$

Figure 1.3: Q-theory representations of tone contours in Tianjin Mandarin, reproduced from Shih and Inkelas (to appear, 2019).

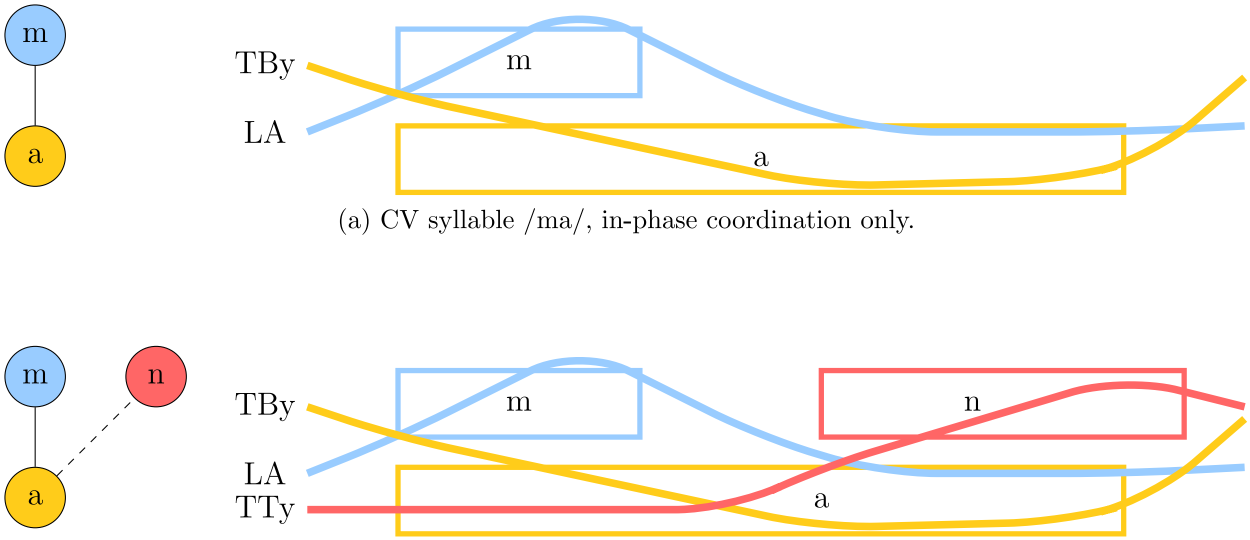
does not eliminate the need for mapping rules from underlying representation to phonetic alignment.

1.1.2 Gestural models of tone

1.1.2.1 Articulatory Phonology and the c-center hypothesis for tone

In contrast, phonetic alignment plays a central role in Articulatory Phonology (AP) concepts of tone, though the focus is less on acoustic simultaneity, in favor of articulatory coordination. AP makes use of the two major modes of motor coordination, which have differing degrees of overlap: in-phase, where the gestures begin simultaneously (as in the arms and legs in jumping jacks), and anti-phase, where the gestures begin 180° out of phase from each other (as in the legs while walking). These two basic modes of coordination are used to create speech units: for example, CV syllables consist of a C(onsonant) gesture that is in-phase with a V(owel) gesture, while VC syllables consist of a V gesture that is anti-phase coordinated with a C gesture (Browman & Goldstein 1988). A CVC syllable, then, is the first C gesture in-phase coordinated with the V gesture, while the second C gesture is anti-phase coordinated with the V gesture.

These coordinative modes are illustrated in Figure 1.4. In Figure 1.4a, where /ma/



(a) CV syllable /ma/, in-phase coordination only.

(b) CVC syllable /man/, in-phase and anti-phase coordination.

Figure 1.4: Schematic representations of coordinative diagrams (left) and corresponding gestural scores and trajectories (right). TBy: vertical position of tongue body sensor; LA: lip aperture (inverted, such that the high point is the closure for /m/); TTy: vertical position of the tongue tip.

is the example syllable, the /m/ (lip aperture; LA) and /a/ (TBy; vertical tongue body position) gestures are in-phase coordinated. In the coordinative diagram (on the left), this is represented by the solid line between the /m/ and /a/ nodes. This corresponds to the alignment of the left edge of the /m/ and /a/ rectangles in the gestural score (on the right), which in turn represent the onsets[2] of each gesture. In Figure 1.4b, where /man/ is the example syllable, the /m/ and /a/ gestures are in-phase coordinated with each other, while the /n/ gesture (TTy; vertical tongue tip position) is anti-phase coordinated with the /a/ gesture. In the coordinative diagram, the anti-phase relationship is represented by the dashed line between the /a/ and /n/ nodes. This corresponds to the alignment of the left edge of the rectangle for the /n/ gesture (i.e., the onset of the /n/ gesture) with the target of the /a/ gesture.

[2] As landmarked by the 20% velocity point of the trajectory.

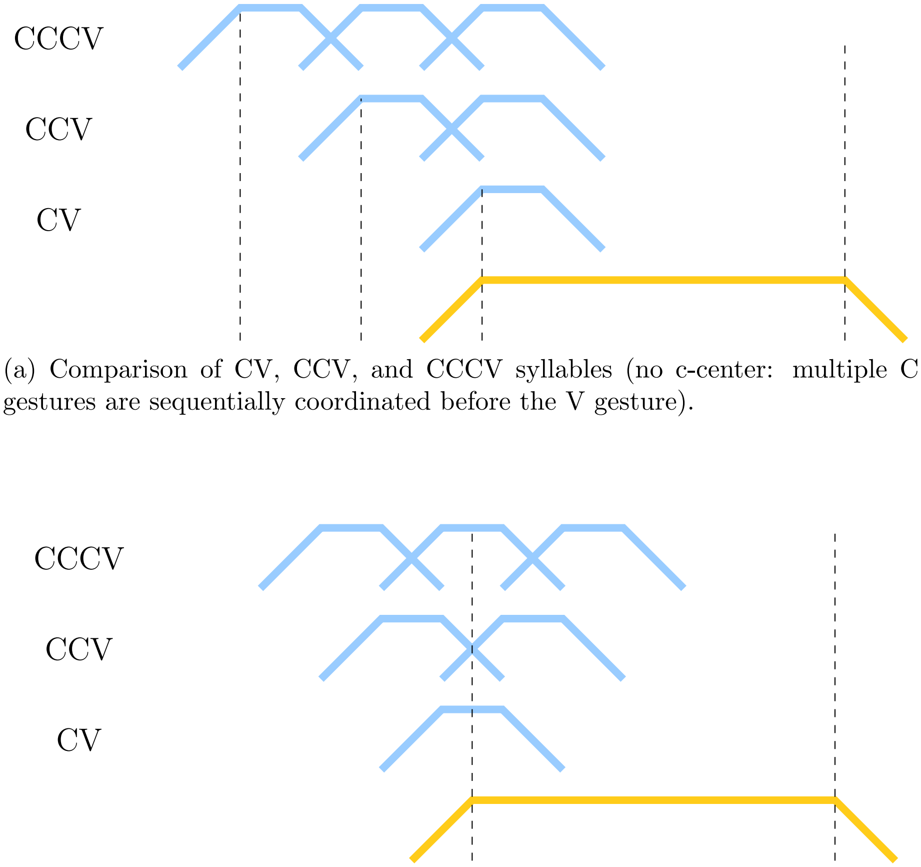
Gestures in AP are specifically viewed to play out over space and time. In the space dimension, gestures can contrast in **constriction location** (analogous to place of articulation) and **constriction degree** (analogous to manner of articulation); the **target** of a gesture is the combination of location and degree of closure. Thus, the gesture for a /t/ is distinct from the gesture for the /s/ in that the degree of constriction is lower for the /s/. The contrast in time is less well-defined, though gestures have been argued to have some sort of "intrinsic duration". The source of this is the stiffness parameter k, which is related to constriction degree and "specifies (roughly) the time required to get to a target" (Browman & Goldstein 1989). Thus, a gesture with high stiffness, such as a stop consonant, takes less time to be realized than a gesture with low stiffness, such as a vowel. The differential coordination of multiple gestures can also create contrast along the time dimension, though this of course involves contrasts between units larger than a single gesture.

The inclusion of tone as an articulatory gesture alongside segmental gestures is relatively new, first seriously considered by Ladd (2006). Gao (2008) was the first to investigate H and L specifications as tone gestures in a lexical tone language, and used Mandarin as a case study. She examined the timing relationships between changes in F0 (as a proxy for T(one) gestures), C, and V gestures, for all four Mandarin tones. She found that the onset of V was in the center of the interval between the onset of C and the onset of T—i.e., tones behaved as if they were the second member of a CC onset cluster with a **c-center** coordinative structure.

The c-center effect does not have its origins in tone research, but rather in research on the production of consonant clusters. It describes a timing relationship where the consonant gestures of an onset cluster are anti-phase coordinated with each other, and also in-phase timed with the vowel (nucleic) gesture (Browman and Goldstein 1988, 2000; Byrd 1995). The c-center structure is far from universal and there has been some speculation as to the reason why it occurs in some cases and not in others. Some have argued that c-center coordination is a more complicated type of coordination than local timing is, and as such is developed in a later stage of the acquisition of cluster production (Tilsen 2016). In this scenario, not all languages reach the stage of c-center coordination, though it is unclear whether this should be motivated by the individual language's phonology or not.

Some have proposed a connection between the phonological structure of apparent clusters and the mode of coordination. For example, Moroccan Arabic, which has been argued to not have true clusters, has repeatedly shown local timing, rather than the c-center (Gafos 2002; Shaw, Gafos, Hoole, & Zeroual 2011), and the same is true for Tashlhiyt Berber (Goldstein, Chitoran, & Selkirk 2007; Hermes, Ridouane, Mucke, & Grice 2011). In contrast, the c-center has been consistently found in English (Browman & Goldstein 2000; Marin & Pouplier 2010) and in Georgian (Chitoran, Goldstein, & Byrd 2002; Goldstein et al. 2007), which are both argued to have "true" phonological clusters. The c-center has also been found to occur with possible phonological exceptions within one language, for example in Italian (Hermes, Grice, Mücke, & Niemann 2008), where sC clusters do not show c-center coordination but CC clusters do, as well as the reverse in Romanian (Marin 2013), where sC clusters do show the c-center but CC clusters do not.

The method for analyzing the c-center in consonant clusters focuses on the timing of C gestures relative to some distinct articulatory "anchor", most frequently the target achievement of the nucleic gesture or the onset of the following syllable, and compares syllables with differing onset complexity—e.g., /ra/, /pra/, /spra/. If c-center coordination is being used, the rightmost gesture should be displaced towards the anchor as more consonant gestures are added, while the leftmost gesture should be displaced away from the anchor (see Figure 1.5b). Thus, the consonant gestures are displaced equally in both directions from the center point of the consonant gestures (i.e., the point halfway between the target achievement of the first consonant and the release of the last consonant)—thus the name "c-center". If the c-center is not used, the rightmost gesture will not move relative to the anchor gesture as more consonants are added, but the leftmost consonant should shift further leftward (see Figure 1.5a). This method does not measure the timing of the onset of the vowel gesture; however, it has been hypothesized that the c-center is coordinated with the onset of the



(a) Comparison of CV, CCV, and CCCV syllables (no c-center: multiple C gestures are sequentially coordinated before the V gesture).

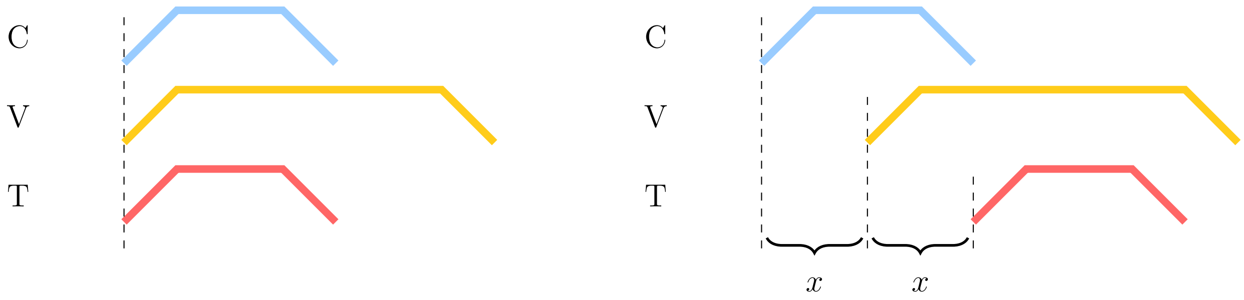
(b) Comparison of CV, CCV, and CCCV syllables (c-center: C gestures spread bidirectionally from the anchor).

Figure 1.5: Gestural score schematics (adapted from Shaw et al. 2011) for determining c-center using the anchor method, showing the relative timing of the center of C (blue) gestures and an anchor point in the V (yellow) gestures for CV, CCV, and CCCV syllables. Figure 1.5a shows no c-center; Figure 1.5b shows c-center.

vowel gesture.

This hypothesis is invoked in the second method of evaluating the c-center (Gao 2008), which directly measures the onset of the vowel gesture. A lack of both toneless syllables[3] and (appropriate) clusters makes it impossible to compare timing differences across syllables with

[3] Of the same phonological status as the words with lexical tone—in Mandarin, toneless syllables are exclusively "clitic-like suffixes or particles" and do not appear in phrase-initial position (Yip 1980). The status of the Thai mid tone is debatable Abramson (1962); Morén and Zsiga (2006).



(a) Onset method: no c-center.

(b) Onset method: c-center.

Figure 1.6: Gestural score schematics for determining c-center using gestural onset timing, showing the relative timing of C (blue), V (yellow), and T (red) gestures. Figure 1.6a shows no c-center (all in-phase); Figure 1.6b shows c-center.

simple and complex onsets—for tone languages like Mandarin, the limited cluster inventory is unsuitable, as it is not possible to track F0 through obstruents. In the method used by Gao (2008), rather than comparing the displacement of consonant gestures relative to an anchor, the timing of the vowel gesture is compared to the hypothesized c-center. In clusters that use the c-center, the onset of the vowel gesture should occur at the midpoint of the two consonant gestures (see Figure 1.6b); in clusters that use local timing, the onset of the vowel gesture should be coordinated with the onset of the last consonant (see Figure 1.6a). This method does not require a comparison between clusters of differing complexity, but it does require that the vowel gesture be visible—that is, the vowel must alternate in height from the pre-target syllable to the target syllable so there is active tongue movement from one nucleus to the next. Unfortunately, this method is also particularly prone to coarticulatory effects when consonants using the tongue are used.

Further studies on Mandarin tone have confirmed the c-center finding (Yi 2014), though with some variability introduced by the interaction of speed and bonding strength. To date, Thai is the only other tone language whose tones been examined from an AP standpoint. Karlin (2014) focused on the coordination of falling tones (analyzed as HL), taking into consideration the availability of the mora in Thai. Unlike Mandarin, Thai has been analyzed to use the mora as its TBU (Morén & Zsiga 2006), and so potentially has another unit available for coordination, rather than all tones being coordinated at the level of the syllable. Similarly to Mandarin tones 1-3, there was a c-center effect, where the onset of V occurred at the midpoint of the interval between the onset of C and the onset of T1 (the H of the HL contour); however, the Thai falling tone was different from the Mandarin falling tone (tone 4). The Mandarin falling tone behaved as if T2 (the L of HL) was a third consonant, thus pushing together the T1 and V onsets, but in Thai, the c-center effect was preserved, while T2 behaved more similarly to a syllable coda.

Karlin (2014) also examined the concept of the TBU in AP, exploring the use of mora-sized "co-selection sets" (Tilsen 2016). She found evidence that the mora is an active unit for coordinating the articulation of tone: T1 was coordinated with the first mora gestures (i.e., the gestures for syllable onset and V1 of diphthong sequences), while T2 was coordinated with the second mora gestures (i.e., V2 in diphthong sequences and non-moraic codas). However, there was also evidence that suggested that T1 and T2 in contour tones are timed to each other in addition to being timed with segmental gestures. Karlin looked at three word shapes with diphthong nuclei: /mua/ (no coda), /muan/ (non-moraic, sonorant coda), and /muat/ (non-moraic, obstruent coda). For all three word types, the lag between T1 and T2 was the same. However, the time lag between V1 and V2 depended on the presence/absence of a coda (regardless of sonority). Thus, the timing of the segmental gestures shifted without affecting the timing of the tone, which remained constant.

Previous articulatory studies on intonation have not found c-center structures (Mücke, Grice, Becker, & Hermes 2009; Prieto & Torreira 2007), instead proposing in-phase coordination between pitch and segmental gestures. Thus, it has been proposed that the c-center is unique to lexical tone languages due to the lexical nature of tone (Mücke, Nam, Hermes, & Goldstein 2011)—that is, lexical tones affect the coordinative structure of the syllable because they are integral contrastive units in the syllable, rather than phrase-based units that are overlaid on already-coordinated syllables. Relatedly, it has been hypothesized that all languages with lexical pitch contrast use the c-center structure to integrate tone gestures, though this hypothesis has not been extensively tested.

1.1.2.2 Articulatory anchoring

Articulatory anchoring (Ladd 2006) is a model of intonation/accentual timing that is the result of cross-pollination between AM theory and gestural theories. Rather than anchoring to acoustic landmarks in the segmental string, articulatory anchoring hypothesizes that F0 movements are anchored to articulatory landmarks—e.g. instead of the acoustic end of a vowel, perhaps the target achievement of the vowel gesture, or its gestural offset. Ladd (2006) argues that there is too much cross-linguistic variation in acoustic alignment to be accounted for by secondary association (as proposed in Prieto, D'imperio, and Fivela 2005, where pitch accents are secondarily associated to prosodic boundaries such as moras, syllables, and prosodic words, in addition to their association to the TBU). Furthermore, proponents of articulatory anchoring argue that the relative timing between intonational pitch gestures and segmental gestures is less variable than that between intonational pitch gestures and acoustic landmarks.

Articulatory anchoring permits a much wider set of coordinative relationships in tonal representation than the c-center hypothesis for tone. Work exploring the articulatory anchoring hypothesis has argued for anchor points beyond what is posited in most versions of AP; for example, D'Imperio et al. (2007) argue that the targets of H gestures in Neapolitan nuclear rises are more tightly coordinated with the maximum onset velocity of lip aperture (for syllables starting with /m/) than with any plausible acoustic landmark. In contrast, the strictest versions of AP only posit timing relationships between gestural onsets, while less restrictive models only additionally admit the coordination of gestural targets.

It is also unclear if articulatory anchoring is a gestural analog of segmental anchoring, where post-phonological processes address gestural anchor points rather than acoustic anchor points. This issue is further complicated by the fact that the hypothesis has largely been explored for intonation, where pitch gestures are not part of the lexical representation (as argued in Mücke et al. 2011). Thus, any coordinative relationships between tones (or intonational pitch accents) and the segmental material would not necessarily be directly with the segmental gestures in the lexical representation of the word, but potentially with units higher in the prosodic hierarchy.

1.2 Languages

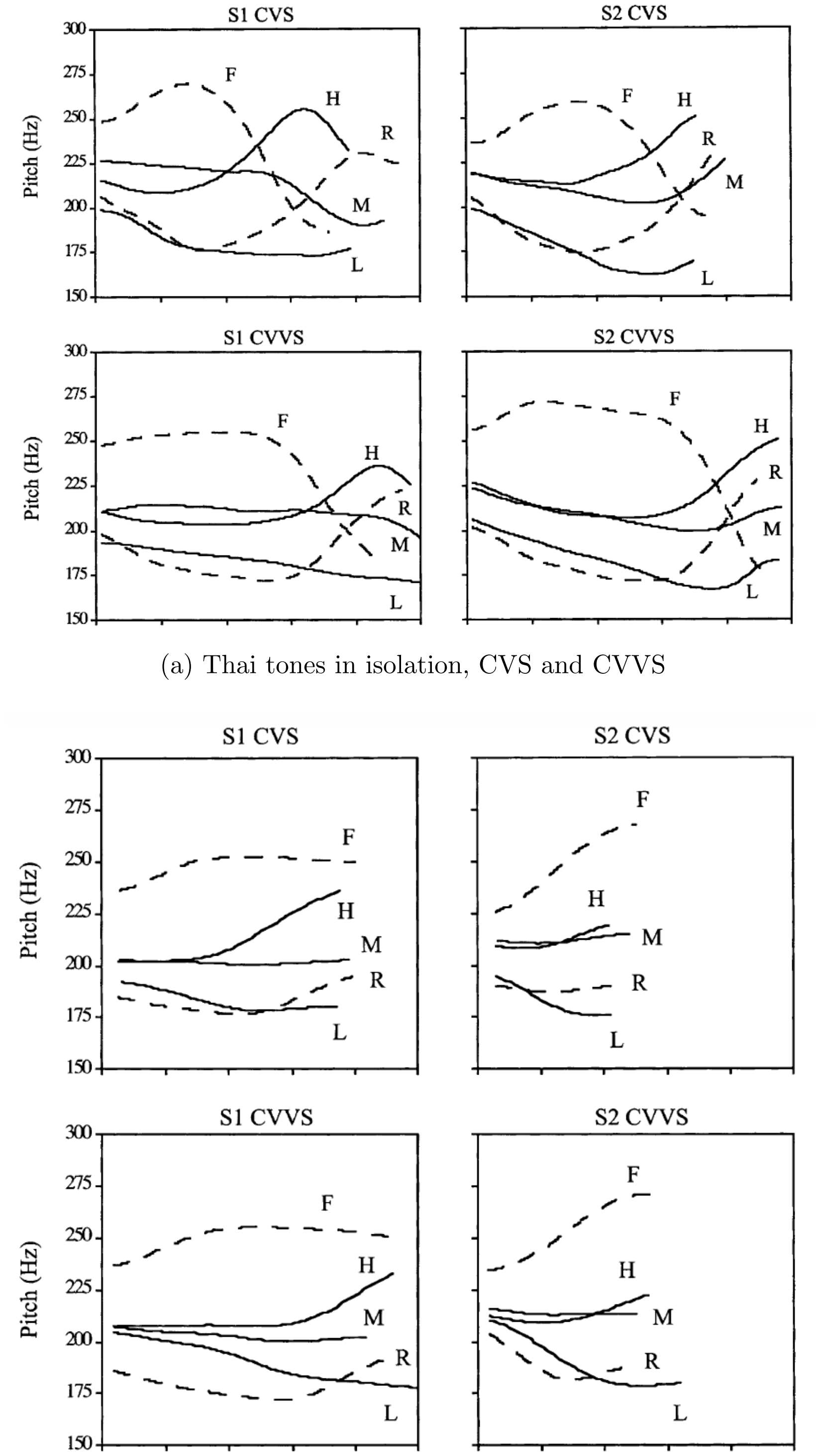
In this section, I provide a description of the contrastive systems of the two languages used in this study, Thai (Asian tone) and Serbian (lexical tone system traditionally called "pitch accent", where one syllable per morpheme is specified for tone).

1.2.1 Thai

There are five tones in Thai, three so-called "level" tones (High, Mid, and Low), and two contour tones (Falling and Rising). Figure 1.7 shows these pitch shapes, in isolation and in Mid-Target-Mid context. Despite their names, the level tones are not phonetically level: the high tone starts in the mid range, and rises near the end of the rime; the mid tone is fairly level, but is often realized as a shallow fall; and the low tone is characterized by a continuous fall. The contour tones are true contours, and thus are also slightly misnamed: the Falling tone first rises, and then falls, while the Rising tone first falls, and then rises. When reduced, as sometimes occurs in connected speech, the first pitch excursion is often preserved at the expense of the second, which results phonetically in a mostly rising Falling tone and a mostly falling Rising tone.

As every syllable in Thai is assigned one lexical tone, the TBU (tone-bearing unit) of Thai has been argued to be the syllable (Abramson 1978, 1979), where the tones were stored as their full shape. More recently, Morén and Zsiga (2006) argued that the TBU in Thai is the mora,[4] and that the lexical tones can be treated as sequences of H(igh), L(ow), and 0 (no

[4] It should be noted that Morén and Zsiga's (2006) argument for the mora as the TBU in Thai is not the same as DiCanio et al.'s (2014) argument for Yoloxóchitl Mixtec, which can assign complex tone melodies to two moras in one syllable.



(a) Thai tones in isolation, CVS and CVVS

(b) Thai tones in context, CVS and CVVS

Figure 1.7: Representative pitch tracks of each tone on CVS and CVVS words, produced by two participants (figures from Morén and Zsiga 2006). F = Falling, H = High, M = Mid, R = Rising, L = Low.

specification), with one tonal element associated to each TBU. This analysis is based on the restricted distribution of Falling and Rising tones, which only occur on syllables with two sonorant moras. Although all stand-alone syllables in Thai are minimally bimoraic,[5] Falling and Rising tones only appear on words with a long vowel, a diphthong, or a short vowel + sonorant (i.e., nasal or glide[6]) coda. Unlike Mandarin, Thai does have a long vs. short vowel contrast, and thus the mora does have an active phonological life independently of tone. Additionally, instead of defaulting to a long vowel, syllables that have only a short vowel are pronounced with a glottal stop final when alone (such as in พระ *phrá* [p^{h}ráʔ] ‘priest’, compared to in the compound พระเจ้า *phrá dʑâw* [p^{h}rá dʑâw] ‘God’). That is, contour tones in Thai do not appear on words with a short vowel + obstruent coda rime (see Table 1.1 for a full summary of tone distribution).

Shape	**Moras**	Low	Mid	High	Fall	Rise
		(0)L	00	(0)H	HL	LH
CV	1		▨		▨	▨
CVO	2*		▨		▨	▨
CVS	2					
CVV	2					
CVVO	2		▨			▨
CVVS	2					

Table 1.1: The distribution of tones in Thai. Greyed out cells are illegal. *CVO has two moras (satisfies minimality constraints), but only one sonorant mora.

As is evident from Table 1.1, there are a few caveats to the “two tone-bearing units, two tones” generalization of Thai tonal distribution. First, there is a gap at CVVO syllables, which have two sonorant moras, but do not permit Rising tones. This could be attributed to an accident of historical tonogenesis or tonal change, or possibly a straightforward pho-

[5] Not all syllables are bimoraic. Multisyllabic words can contain monomoraic syllables non-finally, such as in such as ประเทศ *prà.thêes* ([prà.thêet]) ‘country’ or ขนม *kh.nǒm* ([khà.nǒm]) ‘dessert’, which have just a short vowel nucleus and no coda.

[6] Liquids are also present in Thai, but syllable-finally are produced as [n], such as in the loanword แอปเปิ้ล *ɛ̀p.pɤ̂l* [ɛ̀p.pɤ̂n], ‘apple’

netic/perceptual constraint, as obstruent finals obscure the late final rise in Rising tones, even in citation form. Without the final rise, Rising tones are easily confusable with Low tones (Zsiga & Nitisaroj 2007). Both of these avenues, to my knowledge, remain unexplored, and are a topic for future investigation.

Second, Mid tones also exhibit the same restrictions as Rising tones, in that syllables with Mid tones must have two sonorant moras, and must not end with an obstruent. This is somewhat unusual behavior for a tone that is often described as unspecified for tone (as in Morén and Zsiga 2006)—if there are no tones to specify, there should be no restriction on the number of TBUs required. However, the Mid tone in Thai does not function in the same way as, for example, toneless syllables in Mandarin. Toneless syllables in Mandarin tend to receive tonal material from adjacent words (X. Liu 2014), which does not occur in Thai any more for the Mid tone than it does for other tones. Short syllables (i.e., monomoraic syllables) in Thai are not assigned a default Mid tone either, but rather are either High or Low. For example, in the word มะม่วง *mamuaŋ* 'mango', the first syllable has a spelled-out short /a/ and has a High tone. When the vowel is not explicitly in the orthography, the same happens, as in the "sesquisyllabic" words สบาย *sbaay* 'comfortable', pronounced [sàbaaj], and สมัย *smay* 'era, period', pronounced [sàmǎj]. Tonally unspecified syllables in Mandarin also often serve grammatical functions and are limited to the last syllable of a multisyllabic word (X. Liu 2014; Yip 1980). This is not the case for the Thai mid tone. In Thai, there are several mid tone content words, such as ตา *taa*, 'eye', and there is no restriction on where the Mid tone can occur, demonstrated by the word ความเป็นกลาง *khwaam.pen.klaaŋ*, 'neutrality' (composed of three Mid-tone morphemes, NOMINALIZER.to-be.middle). Thus, although describing Mid tones as a lack of tone is an elegant use of H and L in a five-tone system (thus creating the exhaustive 0, H, L, HL, LH), it remains an open question as to if its restricted distribution is indicative of being marked in representation, or if its distributional restrictions are a historical accident of tonogenesis and historical tone change. This, however, is beyond the scope of the current study and also remains as a topic of future investigation.[7]

In these studies, I am focusing on the two contour tones, Falling and Rising. These tones offer insight on the role of timing relationships in the representation of tone due to their contour shape. That is, they have two easily defined timing events, the first at the onset of the first tone specification (the H in HL, for example), and the second at the onset of the second tone specification (such as the L in HL). In order to examine the relationships between tones and segments, I am limiting the possible word shapes to those that have two different segments for each TBU, such as CVN (/man/) and CV_1V_2 (/mua/). Like contour tones, these word shapes also have two clearly demarcated timing events, one at the start of the first mora and the other at the start of the second mora.

1.2.2 Serbian

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

- Short falling (ȁ[8])
- Short rising (à)
- Long falling (â)
- Long rising (á)

All words in Serbian, excluding function words such as clitics and prepositions, have one primary stress, and thus one accent (Zec 2005). Some minimal pairs do exist, as in Tables 1.2 and 1.3; however, minimal accentual pairs are somewhat rare in Serbian, and except in dictionaries, Serbian orthography does not mark any aspect of the accentual system. The rarity of minimal pairs is due in large part to the near-complementary distribution of rising

[7] Gao (2008) suggested briefly at the end of her discussion of Mandarin tone that the Thai Mid tone may be composed of a simultaneous H and L tone gesture, as she described the non-extreme pitch levels in the rising tone of Mandarin. However, this seems as much an attempt to exhaust H and L in a five-tone system as using 0 does, with the added benefit only of the tone being fully specified and thus having a target.

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

and falling accents—falling accents can only occur on the initial syllable, while rising accents can occur on any non-final syllable (Browne & McCawley 1965).[9]

Table 1.2: An example of a three-way minimal accentual set in Serbian.

	Short	**Long**
Falling	*lȕka* 'onion.GEN'	*Lûka* 'Luka (name)'
Rising	—	*lúka* 'harbor'

Table 1.3: Examples of minimal accentual pairs in Serbian.

(a) Two minimal pairs that distinguish length only

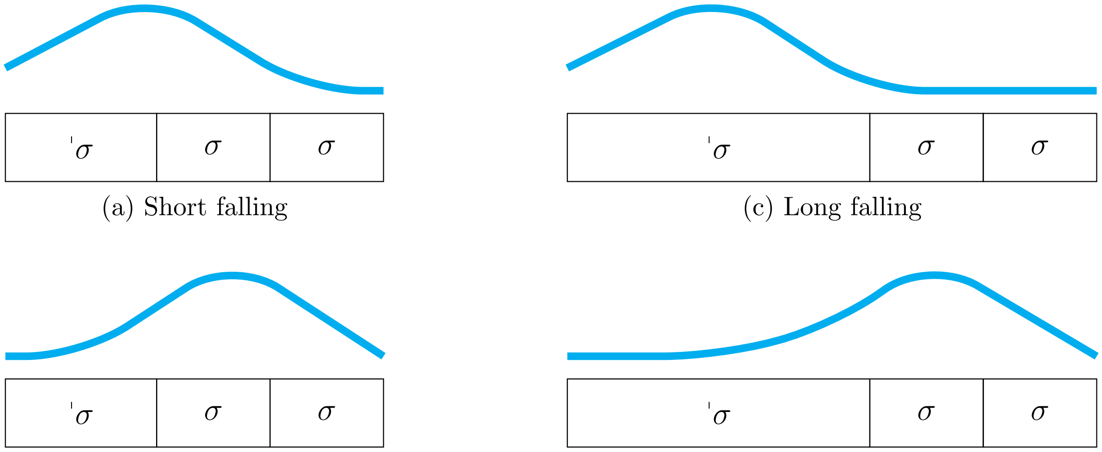
	Short	**Long**
Falling	*ȍran* 'plowed'	*ôran* 'disposed'
Rising	*sèdeti* 'to sit'	*sédeti* 'to go grey'

(b) Two minimal pairs that distinguish pitch contour only.

	Short	**Long**
Falling	*pȁra* 'steam'	*mlâda* 'bride'
Rising	*pàra* 'dime'	*mláda* 'young.FEM'

The names of the accents are indicative of the bundle of prosodic characteristics that have been included in the term "accent". The length descriptors refer to the phonological length of the vowel in the "accented" syllable: short accents have a short vowel, and long accents have a long vowel. The pitch descriptor refers, generally speaking, to the pitch contour of the "accented" syllable: falling accents start high and fall, while rising accents start low and rise. Only for long accents is the pitch contour realized within the stressed syllable; for short accents, it becomes necessary to view the contour from the stressed syllable to the following syllable. That is, the pitch contour is realized over two moras, which may be contained in the same syllable, or "stretched" over two syllables (Inkelas & Zec 1988). Schemata for the

[9] For further elaboration, see Chapter 4



(a) Short falling (c) Long falling

(b) Short rising (d) Long rising

Figure 1.8: Schemata of the F0 movements for the four accent types on trisyllabic words. Stressed vowels are longer than unstressed vowels; for the sake of simplicity, long vowels are represented as twice as long as short vowels.

four accent types are provided in Figure 1.8 (Lehiste and Ivić 1986, and references therein), as they would be produced in trisyllabic words with "accent" on the first syllable.[10]

In the 20th century, several studies analyzed the accentual system not as a bundle of stress, pitch, and length, but rather as separate features that converged to make a four-way distinction (Browne and McCawley 1965; Jakobson 1931, inter alia). In this chapter I assume the analysis presented by Inkelas and Zec (1988) in representing Serbian tone with H lexically assigned to some syllable, and stress one syllable to the left;[11] falling accents occur when the H is on the first syllable and there is no additional syllable to the left for the stress. However, although H placement and stress are determined by syllable, there is evidence that the mora is the TBU of Serbian. First, it has been reported that in several dialects, including that spoken in Belgrade, there is a degree of neutralization between short falling and short rising accents in disyllabic words (Magner & Matějka 1971). Zsiga and Zec (2013) showed that this

[10] Modeled after the Belgrade dialect, with fairly late peak alignment.

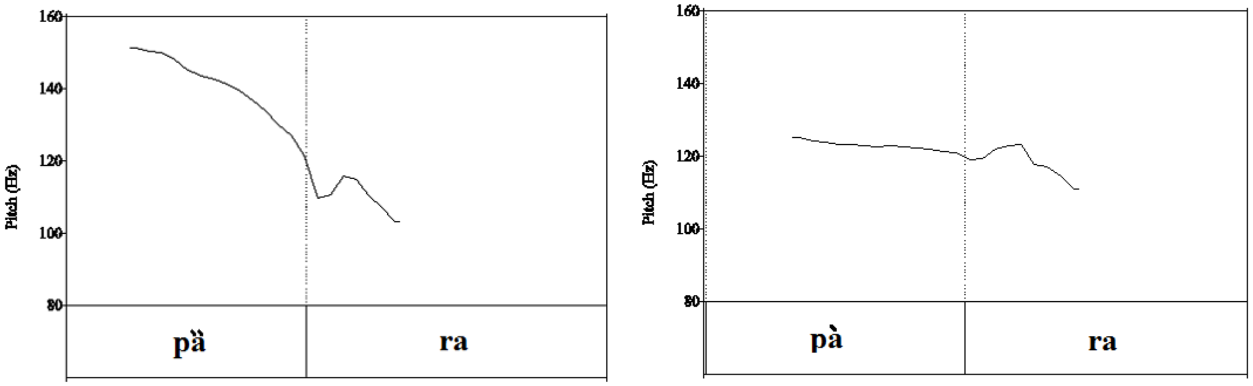
[11] See Chapter 4 for a more detailed examination of different proposals for the representation of Serbian tone.

neutralization is limited to when the lexical high is in utterance-final position, and argued that the neutralization stems from the presence of a L% boundary tone on the final mora of the utterance, which pushes the High in rising words to the previous mora. This shift only occurs when the final vowel is short—for example, it affects the word *ramèna* /raˈmena$_H$/ 'shoulder.GEN', but not the word *vòlan* /ˈvolaːn$_H$/ 'steering wheel'. The result of the shift is that the lexical H is realized on the stressed syllable; when the stressed syllable is short, the result is a surface short falling contour. As falling accents are restricted to the first syllable of a word, the fullest form of neutralization affects disyllabic words, as only then does the lexical high shift to the first syllable and form an underlyingly phonologically possible word.

In some cases, the shift of the High of a short rising accent to the stressed syllable may not result in a surface form that is identical to an underlying short falling accent. Some speakers or varieties preserve a pitch height distinction; in Bosnian Serbian, Wagner and Mandić (2005) found that short falling words have a sharper F0 fall, while short rising words are flatter. Anecdotally, I also found this effect as produced by a male native speaker of Belgrade Serbian with multiple close family members from the southern region of Bosnia and Herzegovina, illustrated in Figures 1.9a (short falling) and 1.9b (short rising) below. However, despite the failure to fully neutralize, there is still an effect of the boundary L% in this situation.

A moraic TBU further explains the shape of the long rising accent on disyllabic words in Belgrade Serbian. As previously mentioned, the contrast between rising and falling is only neutralized for short stressed vowels. The same utterance-final shift of H does not cause neutralization between long rising and long falling accents, as the lexical H shifts by just one mora, not one syllable. The result of the shift is a relatively late peak for long rising accents (near the end of the nucleus), while the long falling accent has a comparatively early peak (near the middle of the nucleus). These pitch contours are demonstrated in Figure 1.10 below.

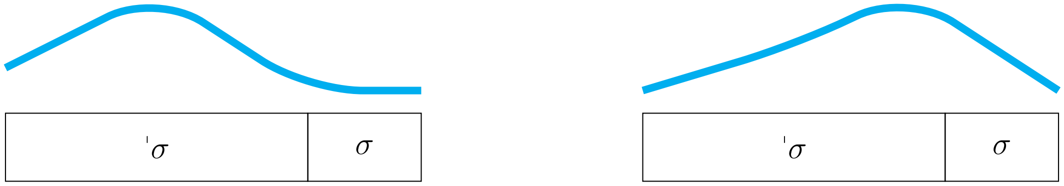
Thus, like Thai, Serbian tone demonstrates a mixed system that makes use of both the



(a) Short falling (*pȁra* ‘steam’)

(b) Short rising (*pàra* ‘cent, coin’)

Figure 1.9: Pitch contours from the minimal pair *pȁra*/*pàra*, where the falling accent has a higher F0 peak and sharper F0 fall than the rising accent (Andrej Bjelaković, personal correspondence).



(a) Long falling on a disyllable

(b) Long rising on a disyllable

Figure 1.10: Schemata of the F0 movements for the long falling and long rising accents on disyllables. The pitch peak in the long rising accent is shifted one mora back from its underlying position, and does not merge with the long falling accent.

syllable and the mora. Like Thai, tone is assigned lexically at the syllable level. However, unlike Thai, the mora unambiguously plays a role in the realization of pitch and interactions between tone and intonation. Serbian also provides a contrast with Thai in that it has tonally unspecified syllables (again, following Inkelas and Zec 1988).

1.2.2.1 The role of syllable onsets

For Serbian specifically, there has not been previous work that has focused on the effects of different syllable onsets on pitch landmark timing; previous studies have used exclusively real words and thus have been limited by lexical availability. Smiljanić (2002), for example, did in fact have a few complex onsets among her stimuli, such as *vrana* ‘crow’ and *mlada* ‘bride’, but onset complexity was not systematically manipulated in her study. In this study, I vary the syllable onset in two ways. First, I compare the timing of pitch peaks in words with complex onsets to those with simple onsets. Second, I examine the effect of syllable onsets with different intrinsic durations within phonological complexity—e.g. the relatively short /r/ vs. the relatively long /m/.

Serbian has a rich inventory of consonant clusters. The constraints on Serbian clusters can be stated as follows (Hodge 1946; Kreitman 2012; Tilsen et al. 2012):

- Obstruent-obstruent clusters must agree in voicing (e.g., *pdica)
- Sonority reversals are not permitted between the the major classes (that is, sonorant and obstruent, e.g., *lpica; reversals within major class are allowed, such as /sp/ (as in *spasti* ‘to save’) and /ps/ (*psovati* ‘to swear’).

Sonority plateaus are allowed, including sonorant-sonorant clusters such as /mn/, /ml/, and /mr/. These are ideal for tone research, as they effect minimal perturbation on F0, in contrast with clusters that contain obstruents.

Although there has been one previous articulatory study on the production of clusters in Serbian (Tilsen et al. 2012), it is still unclear if consonant clusters themselves use the c-center coordinative structure, as the experiment was limited to one participant, and the evidence was inconclusive. Tilsen et al. examined the production of {l, fl, sl, pl, spl, ml, sml, kl, skl}, {r, fr, sr, pr, spr, mr, smr, kr, skr}, {s, ps, ks}, {f, sf, kf}, {p, sp, kp}, and {k, sk, pk}; all clusters exist in Serbian except for {kf, kp, and pk}. The stimuli consisted of mostly nonwords of the form (C)(C)C/ata/, and the participant was instructed to produce the words with a short falling accent on the initial syllable (e.g., *rȁta* /ˈra$_H$ta/). In order to probe c-center organization, they used both indirect and direct measures: the indirect measure compared the timing of the right edge of the rightmost consonant gesture relative to the target achievement of the following /t/; the direct measure compared the timing of the consonant c-center (i.e., the midpoint between the leftmost target and the rightmost gestural offset) to the same anchor point.

The results of this study were mixed. Clusters where /l/ was the last member of a consonant cluster (i.e., /fl, sl, pl, spl, ml, sml, kl, skl/) position did not exhibit c-center like behavior, while clusters where /r/ was the last member of a consonant cluster did. This ran counter to the hypothesis that all complex onsets would be coordinated in a c-center structure. The researchers noted that one possible source of confusion is that it is unclear what measure is best for Serbian /l/, as it has a velar articulation that is timed distinctly from its apical closure. There was also evidence that even the three clusters that do not exist in Serbian ({kf, kp, pk}) were coordinated in a c-center structure, which ran contrary to the hypothesis that non-existing clusters would be produced with simplex timing (i.e., sequentially).

Although there was no manipulation of accent type in this study, this fact is unlikely to have affected the interpretation of the articulatory data. The methods used to determine c-center in this study required comparisons between simple and complex onsets with the same segments, such as {l, pl, spl}, rather than directly determining the timing of onsets of the consonant gestures and their relationships with the onset of the vowel gesture (as was the methodology in Yi 2014 and Karlin 2014). Thus, even if a tone gesture were participating in a c-center relationship with the consonant, the comparisons would still be valid—the only difference would be that instead of comparing an onset with one gesture (e.g., /l/) to onsets with two (/pl/) and three (/spl/) gestures, the comparison would in fact be between two- (/l/ + tone), three- (/pl/ + tone), and four-gesture (/spl/ + tone) onsets.

The precise coordinative patterns of onset consonants in Serbian is not the goal of the current study. Rather, I will examine differences in peak timing as they relate to the duration and phonological complexity of the syllable onset. These comparisons will serve to distinguish between acoustic and articulatory theories of pitch anchoring. I will furthermore examine differences in pitch onset timing and pitch excursion duration cross- and within-dialect (i.e., cross-onset), which serves to tease apart the predictions generated by different articulatory theories of tone alignment.

1.2.2.2 Zsiga and Zec 2013 and Smiljanić 2002: two analyses

Inkelas and Zec (1988) argue that Serbian accent is most fruitfully treated as a H(igh) pitch that determines the location of stress. Specifically, the location of the H is lexically specified, and stress is located one syllable to the left. This accounts for 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—since the stress cannot move one syllable to the left, H and stress occur on the same syllable, which results in a falling contour from the stressed syllable to the following syllable.

This approach accounts for the near-complementary distribution of rising and falling accents in Serbian. Although stress can occur on any non-final syllable, accents are not uniformly distributed. Browne and McCawley (1965) describe the distribution of accents as follows:[12]

(1) a. Rising accents can appear on any syllable other than final (i.e., they cannot appear on monosyllabic words).

b. Falling accents can appear only on the word-initial syllable (and on the only syllable of monosyllabic words).

These restrictions follow directly from Inkelas and Zec's (1988) account. First, the H itself can occur on any syllable; stress does not occur on word-final syllables (excluding monosyllables) because it is always to the left of the H. Second, falling accents only occur when there is no syllable left of the H—i.e., when H and stress are both initial. Thus, the four Serbian accents can be broken down and described as in Table 1.4.

[12] There are some exceptions to these rules, but they appear mostly in recent polysyllabic loan words, such as *asistȅnt* 'assistant' and *Austrâlija* 'Australia' (Inkelas & Zec 1988).

Table 1.4: 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.

Accent	**Vowel**	**Stress**	**Pitch**	**Phonology**	**Orthography**	Gloss
Short falling	short	initial	initial	/ˈje$_H$zero/	jȅzero	‘lake’
Short rising	short	initial	second	/ˈpapri$_H$ka/	pàprika	‘pepper’
Long falling	long	initial	initial	/ˈzaː$_H$stava/	zâstava	‘flag’
Long rising	long	initial	second	/ˈraːzli$_H$ka/	rázlika	‘difference’

This approach also simplifies accentual shifts caused by prefixing and procliticization. One example of an accentual shift is caused by *ne* proclitization, in which the negative particle *ne* procliticizes to the verb to form one phonological word. If the verb has a rising accent, the accent remains the same from affirmative to negative—i.e., neither the stress nor the H shifts. This results in non-alternations of accent, such as those in (2).

(2) a. *sèdīm* /ˈsediː$_H$m/ sit.1SG > *ne=sèdīm* /neˈsediː$_H$m/ NEG sit.1SG

b. *tȑčīm* /ˈtrʧiː$_H$m/ run.1SG > *ne=tȑčīm* /neˈtrʧiː$_H$m/ NEG run.1SG

c. *žívīm* /ˈʒiːviː$_H$m/ live.1SG > *ne=žívīm* /neˈʒiːviː$_H$m/ NEG live.1SG

d. *séčēm* /ˈseːʧeː$_H$m/ cut.1SG > *ne=séčēm* /neˈseːʧeː$_H$m/ NEG cut.1SG

On the other hand, if the verb has a falling accent, the stress shifts one syllable to the left (i.e., to *ne*), while the H (and length) remains on the original syllable. This results in falling > rising alternations such as those in (3). Under Inkelas and Zec's (1988) analysis, this "accent alternation" is simply an extension of the rule that places stress one syllable to the left of H. In the affirmative, there is no syllable to the left of the H for stress to be assigned to, but the negative proclitic expands the phonological word to the left, thus allowing the rule to apply.

(3)	a.	*vȉdīm*	/ˈvi$_H$diːm/	see.1SG	> *nè=vidīm*	/ˈnevi$_H$diːm/	NEG see.1SG
	b.	*mȍlīm*	/ˈmo$_H$liːm/	ask.1SG	> *nè=molīm*	/ˈnemo$_H$liːm/	NEG ask.1SG
	c.	*râdīm*	/ˈraː$_H$diːm/	do.1SG	> *nè=rādīm*	/ˈneraː$_H$diːm/	NEG do.1SG
	d.	*môrām*	/ˈmoː$_H$raːm/	must.1SG	> *nè=mōrām*	/ˈnemoː$_H$raːm/	NEG must.1SG

This could be described as a form of "non-culminativity", where pitch as a marker of prominence is not limited to the stressed syllable. Thus, the stressed syllable and the H tone syllable are distinct. However, recall that while this is true both phonologically and phonetically for rising accents in the Belgrade dialect, the peaks of rising accents in the Valjevo dialect frequently occur during the stressed syllable.

The main alternative analysis takes an Autosegmental-Metrical approach and focuses on phonetic alignment, but at the expense of explaining phonological processes. In her 2002 dissertation, Smiljanić analyzed Belgrade[13] 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 (in the Belgrade dialect), 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 (cf. Zsiga and Zec 2013). Furthermore, such a restriction would not account for the accentual shifts described in (3) above; the process would require first a stress shift, followed by a change of pitch melody, as well as a stipulation in the grammar that this process only occurs if the pitch melody on the non-cliticized form is L+H*.

Thus far, there has not yet been a study that can definitively rule out one proposed representation. The data presented in Chapter 3 suggests that the Inkelas and Zec (1988) representation is correct, as the timing of the H in the short rising accent was influenced

[13] As compared to Zagreb, which has a reduced set of contrasts.

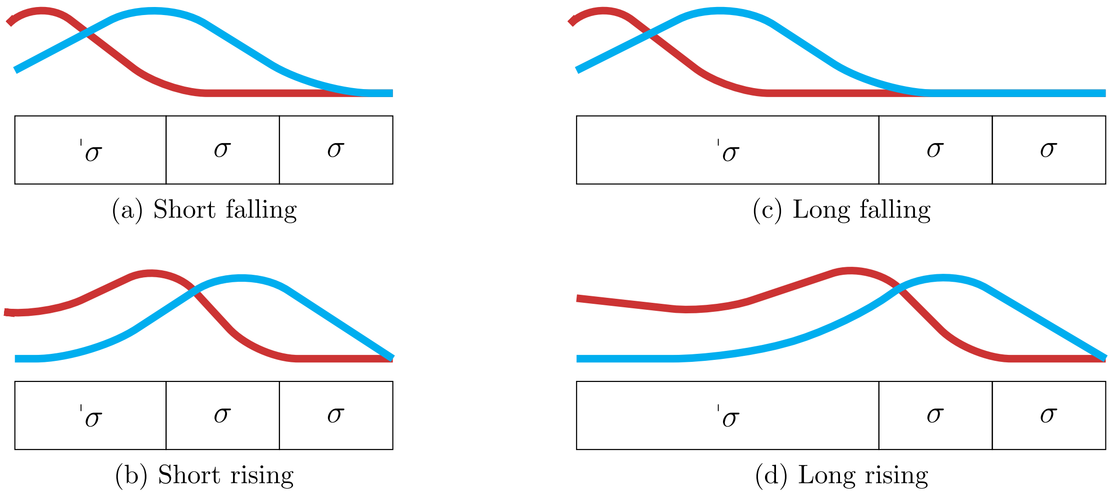
by the duration of the syllable onset of the post-tonic syllable. However, the patterns from Valjevo—in particular the earliness of the peak in rising accents—potentially indicate that the anchored representation proposed by Smiljanić (2002) is correct.

1.2.2.3 Dialectal differences

In these studies, I focus on Belgrade and Valjevo Serbian, which are both "Neo-Štokavian" dialects of Serbian. "Štokavian" refers to dialects of the BCS spectrum that use the word *što* for 'what', while the prefix "Neo" refers to dialects that underwent the diachronic shift that produced rising accents. As a general rule, Neo-Štokavian dialects all have the same four-way system of contrast described in the previous section (Lehiste & Ivić 1986); however, the realization of these accents and interaction with other parts of the prosodic system differ from dialect to dialect.

The pitch contours of the long accents are similar to their corresponding short versions. In the Belgrade dialect, the F0 peak of the long falling accent is realized near the middle of the long vowel (i.e., near the end of the first mora, as in the short falling accent), while in the Valjevo dialect the F0 peak occurs near the beginning of the nucleus (Figure 1.11c). Similarly, in long rising accents, the Belgrade F0 peak occurs in the second syllable (with the F0 rise beginning in the second mora of the first syllable), while in the Valjevo dialect the pitch peak occurs before the end of the first syllable (Figure 1.11d). In sum, the main difference between the two dialects is that Belgrade F0 peaks occur near the end of the mora the lexical H is associated to, while Valjevo F0 peaks occur near or even before the beginning of the mora the lexical H is associated to.

Zec and Zsiga (2016) describe "variable retraction" of the Valjevo rising accents in phrase-initial position: 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 first syllable. For the remaining speaker, the F0 peak consistently occurred on the stressed syllable, rather than on the post-stress syllable as is canonical for the Serbian rising accent. In phrase-final position, all speakers consistently retracted the H tone of a final syllable to the stressed syllable. It is



(a) Short falling

(c) Long falling

(b) Short rising

(d) Long rising

Figure 1.11: Schemata of the F0 movements for the four accent types on trisyllabic words, comparing Belgrade (blue) and Valjevo (red) dialects. Contours adapted from Zec and Zsiga 2016.

unclear if the non-retracted tokens in phrase-initial position were in a separate distribution from the retracted tokens, or if the timing of the F0 peak was simply distributed around the syllable boundary such that half of the tokens had a peak after the boundary, and the other half had a peak before—that is, it is unclear if the measurement of "retraction" is actually targeting a phonological process, or if it is simply a categorical description of phonetic variation. The contrasting behavior of phrase-initial vs. phrase-final rising accents also does not clarify the matter: all tokens appear as "retracted", but even in falling accents, the F0 peaks in phrase-final words occur significantly earlier than the F0 peaks in phrase-initial words. Thus, it could simply be the case that in phrase-final position, the distribution of F0 peaks shifted sufficiently leftward to place all F0 peaks before the syllable boundary, rather than categorically shifting to the left by one mora.

There is the issue of what Valjevo "retraction" creates in the acoustic realization, which is essentially a tone system that on the surface contrasts two falling contours with different timing. Although a full investigation of the consequences of such a system is not included in the scope of the current study, the studies presented in Chapter 3 and Chapter 4 provide some indirect evidence for different patterns of H association between the Belgrade and Valjevo dialects. These results will be considered as part of the larger discussion on the role of the phonological TBU in phonetic alignment.

1.3 Questions and structure of the dissertation

In this dissertation, I adopt the perspective that tone is represented as gestures, and I interpret acoustic data through a gestural lens. This potentially raises the issue of the tension between articulatory and acoustic notions of simultaneity. Much of the existing literature on tone has examined acoustic timing and evaluated simultaneity based on acoustic overlap (see Sagey 1986 for a discussion of simultaneity and association, for example). However, acoustic overlap and non-overlap do not map directly to specific articulatory coordinative structures. This mismatch is demonstrated clearly by CV syllables: as described above, the C and V gestures of a CV syllable are in-phase coordinated (that is, they start at the same time), but the acoustic consequence is the apparently linear sequence C-V. Similar mismatches can occur for tone, but the obfuscation can be minimized by including only sonorant segments, which do not obscure pitch trajectories, and examining the relative timing of gestural targets and onsets as would be done in an articulatory study.

The novel proposal in this dissertation is the concept of an **articulatory TBU**, which is a constellation of gestures that a tone gesture is coordinated with and receives timing information from. This model will address the nature of tone gestures and their relationships to segmental gestures, and how these relationships link the distributional and timing characteristics of tone. As speakers negotiate and produce both categorical and gradient phenomena (Zsiga 1993), it is important when forming an analysis to balance distributional and categorical phenomena with phonetic patterns and not give undue weight to one facet over the other.

The rest of this dissertation is organized as follows: in Chapter 2, I present the results of an acoustic study on Thai, which is designed to examine the proposed relationship between tones and a moraic TBU (Morén & Zsiga 2006), and show that different modes of coordination with the mora as an articulatory TBU can be the source of apparent variable timing with respect to a segmental boundary. In Chapter 3, I present the results of an acoustic study on two dialects of Serbian, which probes various aspects of the timing relationships between tone gestures and TBUs, and show that tone gestures get durational information from their TBU. In Chapter 4, I present the results of a second acoustic study on Serbian, which builds on the findings from Chapter 3 and targets a case of the mismatch between articulatory and acoustic timing. In Chapter 5 I summarize the findings of the acoustic studies and tie them together to present a gestural model of tone representation.

Chapter 2

Thai

In this chapter, I present the results of an acoustic study that provides a close examination of the relationship between tone gestures and segments, using Thai `(tha)` as the language of investigation. This study is novel in that it explicitly considers the timing of tone contours relative to moraic boundaries, as well as how timing is affected by tonal coarticulation.

I first investigate the effects that segments have on the realization of tone, as well as the effects that tones have on the realization of segments. To this end, I vary the segmental content of the carrier words, using word shapes with segmentally distinct moras (such as diphthongs) and different segmental makeups (such as a non-moraic codas vs. no coda). I pair these word shapes with the two Thai contour tones, F(alling) and R(ising), which are true contour tones (i.e., rise-fall and fall-rise) that make possible an examination of the timing of pitch excursions relative to the right edge of the first mora. I also investigate the effects of tone sequences both tone and segmental realization. In order to do this, I use sequences of two tones: F+F, F+R, R+F, and R+R.

The structure of this chapter is as follows: in Section 2.1, I provide the hypotheses for this chapter, and give details on the experimental design. In Section 2.2, I present the results of the acoustic study. Then, in Section 2.3, I give interim conclusions and discuss the implications of the findings.

2.1 Experimental design

2.1.1 Hypotheses

I present here the hypotheses in general terms; specific predictions referencing the variables included in the statistical analyses will be presented before each analysis. As briefly described in the introduction, there are three independent variables, the predictions for which I list here without null hypotheses:

> **Hypothesis A** (independent variable—segments): Different word shapes will have significantly different syllable durations and significantly different first mora durations.
>
> **Hypothesis B** (independent variable—tones): Different tones will have significantly different extremum timing—specifically, the Falling peak will occur earlier in the syllable than the Rising valley.
>
> **Hypothesis C** (independent variable—tone sequence): Anticipatory dissimilation (as described by Gandour, Potisuk, Dechongkit, and Ponglorpisit 1992; Potisuk, Gandour, and Harper 1997) affects both pitch height and elbow timing.

Hypothesis 1 examines the effect of tones on segments, where the segment that is the focus of these investigations corresponds to the first mora.

> **Hypothesis 1.0** (null hypothesis): There is no effect of tone identity on segmental realization.
>
> **Prediction 1.0**:
>
> - Tone identity will not be a significant predictor of word duration.
> - Tone identity will not be a significant predictor of the timing of the right edge of the mora (i.e., the fact that Falling tones have an early elbow does not mean that the first mora will be shorter in Falling tone words).

Hypothesis 1.1: Tone plays a significant role in the timing of the first mora.

Prediction 1.1: Tone identity will be a significant predictor of the timing of the right edge of the first mora, and elbow timing will parallel moraic timing: Falling tones (with early extrema) will have shorter first moras, and Rising tones (with late extrema) will have longer first moras.

Hypothesis 2 address the effects of segments on tone realization.

Hypothesis 2.0 (null hypothesis): There is no effect of segment duration on tone realization.

Prediction 2.0:

- Word duration will not be a significant predictor of extremum timing (i.e., tone gestures have absolute time specified, and need a specific amount of time to be realized, regardless of the accompanying segments).
- The right edge of the first mora will not be a significant predictor of extremum timing (i.e., differences in mora timing will not affect elbow timing within tone category).

Hypothesis 2.1: Syllables are a main driving force in Thai tone timing.

Prediction 2.1:

- Word shape (where each syllable shape has a different duration) *is* a significant predictor of elbow timing (in milliseconds);
- Word shape is *not* a significant predictor of time-normalized extremum timing—specifically, all word shapes will have the same extremum timing.

Hypothesis 2.2: Moras are a main driving force in Thai tone timing.

Prediction 2.2: The right edge of the first mora is a significant predictor of elbow timing.

I also consider the role of tone targets in tone gesture representation and timing. Hypotheses 3 - 8 address the relationship between the timing and excursion size of tone gestures, specifically focusing on the first tone gesture of the contour tone (i.e., the upward trajectory of Falling tones, and the downward trajectory of Rising tones). Hypothesis 3 addresses the effects of tone identity on the duration of the first tone gesture.

Hypothesis 3.0 (null hypothesis): The first excursion is identical in duration for Falling and Rising tones.
Predictions 3.0: Tone identity is not a significant predictor of the duration of the first pitch excursion.

Hypothesis 3.1: In line with the timing differences posited in Hypothesis B, Rising tones have a longer initial pitch excursion.
Predictions 3.1: Tone identity is a significant predictor of excursion duration—specifically, Rising tones have a longer initial excursion.

Hypothesis 4 investigates the effects of word shape on the duration of the first tone gesture.

Hypothesis 4.0 (null hypothesis): The first excursion is identical in duration across word shapes.
Predictions 4.0: Word shape is not a significant predictor of the duration of the first pitch excursion.

Hypothesis 4.1: Excursion duration is determined in part by word shape.
Predictions 4.1: Tone identity is a significant predictor of the duration of the first pitch excursion—specifically, longer words have longer initial excursions.

Hypothesis 5 investigates the effects of tone identity on the timing of the start of the first pitch excursion. Hypothesis B addresses the timing of the tonal extrema; Hypothesis 5 probes whether the difference in timing between Falling and Rising tones is due to the timing of the start of the first excursion, or the duration of the first excursion.

Hypothesis 5.0 (null hypothesis): The Falling and Rising tones use the same coordinative regime to coordinate the start of the tone gesture with the beginning of the word.

Prediction 5.0:

- There is no effect of tone identity on the timing of the start of the initial pitch excursion.
- Tone identity is a significant predictor of the duration of the first pitch excursion—specifically, Rising tones have longer excursions.

Hypothesis 5.1: Falling and Rising tones use different coordinative regimes to coordinate the start of the tone gesture with the beginning of the word.

Prediction 5.1:

- There is a significant effect of tone identity on the timing of the start of the first pitch excursion, where the first pitch excursion starts later for Rising tones than for Falling tones.
- Tone identity is not a significant predictor of excursion duration.

Similarly, Hypothesis 6 addresses the effects of word shape on the timing of the start of the first pitch excursion.

Hypothesis 6.0 (null hypothesis): All word shapes have the same coordinative relationship with the first tone gesture.

Prediction 6.0: There is no effect of word shape on the timing of the start of the first pitch excursion.

Hypothesis 6.1: The timing of the beginning of the first tone gesture depends on the shape of the word.

Prediction 6.1: Word shape will be a significant predictor of the timing of the start of the first pitch excursion.

Hypothesis 7 investigates the effects of tone identity on the timing of the start of the first pitch excursion, while Hypothesis 8 examines the effects of word shape on the timing of the start of the first pitch excursion.

Hypothesis 7.0 (null hypothesis): The first excursion is identical in magnitude for Falling and Rising tones.

Predictions 7.0: Tone identity is not a significant predictor of the magnitude of the first pitch excursion.

Hypothesis 7.1: In line with the timing differences posited in Hypothesis B, Rising tones have a more extreme initial pitch excursion.

Predictions 7.1: Tone identity is a significant predictor of excursion size—specifically, Rising tones have a greater excursion size.

Hypothesis 8.0 (null hypothesis): Tone targets are specified in the representation of the tone gesture.

Prediction 8.0: The duration of the initial pitch excursion (as related to word shape) is not a significant predictor of excursion size—i.e., word shapes that cause longer initial excursions do not also cause larger pitch excursions.

> **Hypothesis 8.1**: Pitch excursions and F0 extrema are the phonetic result of a tone gesture (upward for Falling tones; downward for Rising tones) that continues until the next gesture is activated.
>
> **Prediction 8.1**: Excursion duration (as related to word shape) is a significant predictor of excursion size.

Finally, in this study I also investigate tonal coarticulation and its effects on the relationship between tones and segments by varying sequences of two tones. Hypothesis 7 investigates the effects on segments found in sequences of tones, paralleling Hypothesis 1. In this study, I focus on the effects that the second word has on the first word. The null hypothesis is that tonal anticipatory dissimilation is phonetic in nature.

> **Hypothesis 9.0** (null hypothesis): There is no effect of tone sequence on the realization of the segments of the first word.
>
> **Prediction 9.0**:
>
> - Whether the two words in the sequence have the same tone or not (that is, whether the sequence is F+F/R+R or F+R/R+F) will not be a significant predictor of the duration of the first word.
> - Whether the two words in the sequence have the same tone or not will not be a significant predictor of `REMora`.
>
> **Hypothesis 9.1**: Tone sequence does affect the realization of the first word.
>
> **Prediction 9.1**:
>
> - Whether the two words in the sequence have the same tone or not will be a significant predictor of the duration of the first word.
> - Whether the two words in the sequence have the same tone or not will be a significant predictor of the duration of only the first mora in the first word.

2.1.2 Stimuli

2.1.2.1 Target words

This study uses the four possible dyads of contour tones (presented in Figure 2.1). These tones were combined with four different word shapes, making the words presented in Table 2.2 below. These target word shapes have been previously found to have different syllable duration, as well as different first mora duration Karlin (2014), and thus make it possible to compare hypotheses.

All target words are composed entirely of sonorants so that pitch can be tracked as accurately as possible. They also all have /m/ as their onset, which, as a bilabial consonant, minimizes mutual interference between the onset gesture and the vowel gestures. Due to these restrictions, over half of the words used are nonce words, but all are legal combinations. Due to happy coincidence, the syllable shapes with no real words with either Falling or Rising tones[1] (/mia/ and /mian/) are used as names.

Tone 2 Tone 1	Falling	Rising
Falling	F + F (HL + HL)	F + R (HL + LH)
Rising	R + F (LH + HL)	R + R (LH + LH)

Table 2.1: The target tone sequences.

In an ideal world, all syllable shapes would be identical in Target 1 and Target 2. However, there is some concern that simply repeating the same form for both targets will have a reduplicative-like effect, which could produce variation that is not the focus of this study. Reduplication in Thai is fairly common and results in an unstressed-stressed pair, where the unstressed item can be severely reduced. Thus, although /muan + muan/ would otherwise be a licit sequence, it is avoided for this reason.

[1] Mid tone เมีย *mia* means 'wife', but no other tones exist on this segment sequence. High tone เมี้ยน *mían* is the word for the Mien people, but no other tones exist on this segment sequence either.

Shape		Target 1	Target 2
CV_1V_2	Falling	เมี่ย mîa *nonce*	มั่ว mûa ‘guess wildly’
	Rising	เหมีย mǐa *nonce*	หมัว mǔa *nonce*
CV_1V_2N	Falling	เมี่ยน mîan *nonce*	ม่วน mûan ‘to have fun’
	Rising	เหมียน mǐan *nonce*	หมวน mǔan *nonce*
CVN	Falling	มั่น mân ‘to be sure’	มุ่น mûn ‘to be anxious’
	Rising	หมัน mǎn ‘to be sterile’	หมุน mǔn ‘to knot up’
CVVN	Falling	ม่าน mâan ‘screen, curtain’	มู่น mûun *nonce*
	Rising	หมาน mǎan *nonce*	หมูน mǔun *nonce*

Table 2.2: The words to be used in the study and their glosses. C: Consonant, V: Vowel, V_1V_2: diphthong; Target 1: first word in two-word sequence, Target 2: second word in two-word sequence.

2.1.2.2 Carrier phrases

The target words were embedded in a carrier phrase, either “ นาง *naaŋ* Target 1 + Target 2 + ดีๆ *diidii*” or “ คุณ khun Target 1 + Target 2 + ดีๆ *diidii*”, both of which mean “Ms. name verbs easily.” Thus, all Target 1 words are treated as names, and all Target 2 words are treated as verbs. Both of the words in the carrier phrase have a mid tone, which, while in Thai is not a default or “toneless” tone, is nevertheless near the middle of the F0 range and approximately level.

2.1.3 Task

Before the experiment began, the experimenter told the participant that the words were sometimes made up and didn’t make sense, and asked them to read the sentences as naturally as possible. Before the experiment, participants completed a practice block, where words were presented in isolation in order to let the participants become accustomed to the nonce words, followed by eight practice sentences. In each trial, the participant saw the entire sentence presented in Thai script on a screen set at a comfortable distance from them.

The experiment took place in a sound-attenuated booth in the phonetics lab of the Cornell University Linguistics Department. Stimuli were presented using PsychoPy (Peirce 2007), and audio data recorded using a Samson GoMic. Participants pressed the space bar to advance through the trials at their own pace. There was no manipulation of speech rate in this study.

The experiment was divided into blocks by tone sequence, and each sentence was presented twice in a block, for a total of 32 trials per block. All 16 sentences were presented before any repeated, and they were presented in a different order the second time. Blocks were presented in order 1. F+F, 2. F+R, 3. R+F, 4. R+R, and this tone sequence order was repeated for the second round. A full experiment consisted of 8 blocks, or 2 repetitions of each tone sequence. This leads to a total of 64 sentences per tone sequence type. Each unique combination of target words has 4 repetitions. Participants received 10 dollars for their participation in the study.

2.1.4 Participants

There were a total of six participants in this experiment: 3 male, and 3 female. Their ages ranged from 24 to 33 with a mean age of 27. All were native speakers of Central Thai from Thailand, though had been living in Ithaca, NY for one to five years.

2.1.5 Data labeling and analysis

2.1.5.1 Segmentation

Approximately half of the data was initially aligned using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger 2017), and the other half segmented by hand with no automatic alignment; the automatically aligned data was ultimately corrected by hand in Praat. As there were no statistical analyses that require accurate segment marking on the first or last word of the carrier phrase, only the word boundaries for these two words were corrected, which gives accurate phrase duration.

Marking segment boundaries was fairly straightforward, with two notable exceptions. First, the boundary between the two qualities of a diphthong were not marked by approximating the point of maximum velocity of F1 and F2 between the two qualities, but rather at a point in F2 (the more dynamic of the two formants) that demonstrated clear movement away from the first vowel quality. The reference to F2 approximates the use of vertical tongue position for the articulatory marking of vowels, while the point of clear movement away from the first vowel quality approximates the 20% maximum velocity threshold used to mark gestural onsets. Second, in some Word 1 + Word 2 sequences, there was a coda nasal followed by an onset nasal. In most cases, the boundary was quite visible in the spectrogram (example provided in Figure 2.1) and readily marked; the remaining tokens were marked at visible and reasonable points of F2 and F3 change.[2]

2.1.5.2 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 using a Matlab script. Pitch track landmarking was performed using a Matlab script that first found pitch extrema within the boundaries of each word—for Falling tones, the highest point within the boundaries of the word, and for Rising tones, the lowest point within the boundaries of the word. These values were then used to bound where further landmarks could be located. Two example landmarked trajectories from one of the participants are provided in Figure 2.2a and Figure 2.2b. Only landmarks that were used for analysis are included below.

- **Word 1 F0 elbow** (C): The most extreme F0 of target word 1 (in both figures, the maximum F0, as Word 1 is a Falling tone). The only boundaries specified for this landmark were the acoustic left and right edges of the word.

[2] Auditory assessment does not provide much insight in these cases, as in many cases the two points of comparison would be 10 ms or less apart—i.e., an /n/ with 10 ms vs. 0 ms of /m/ at the end.

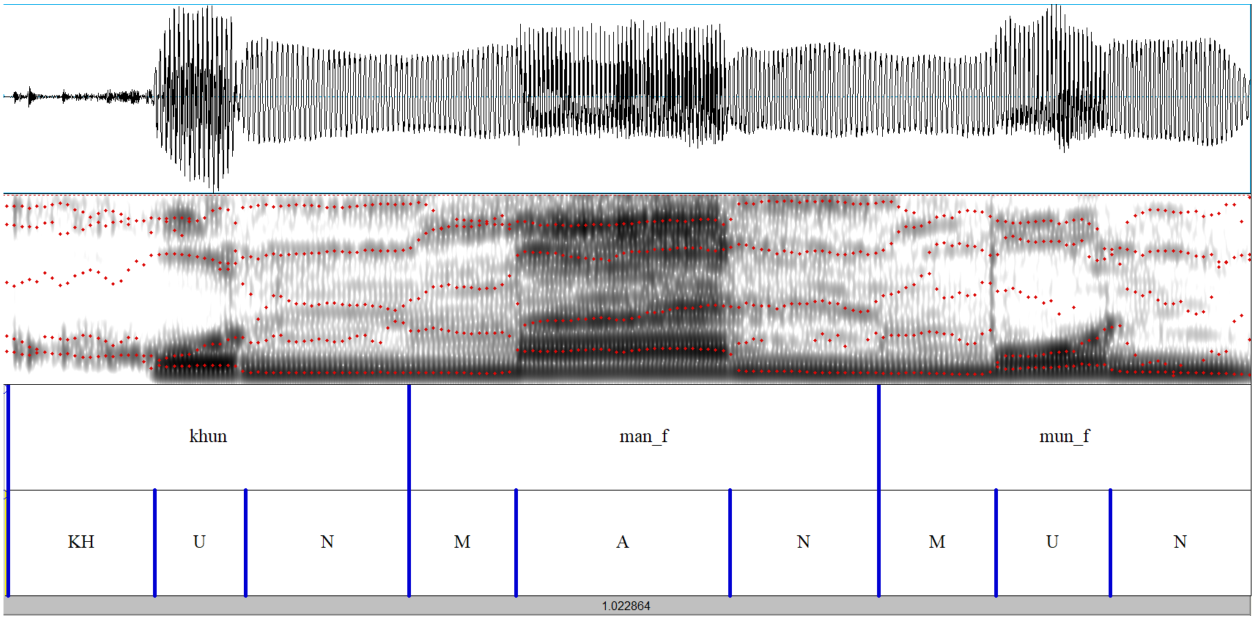
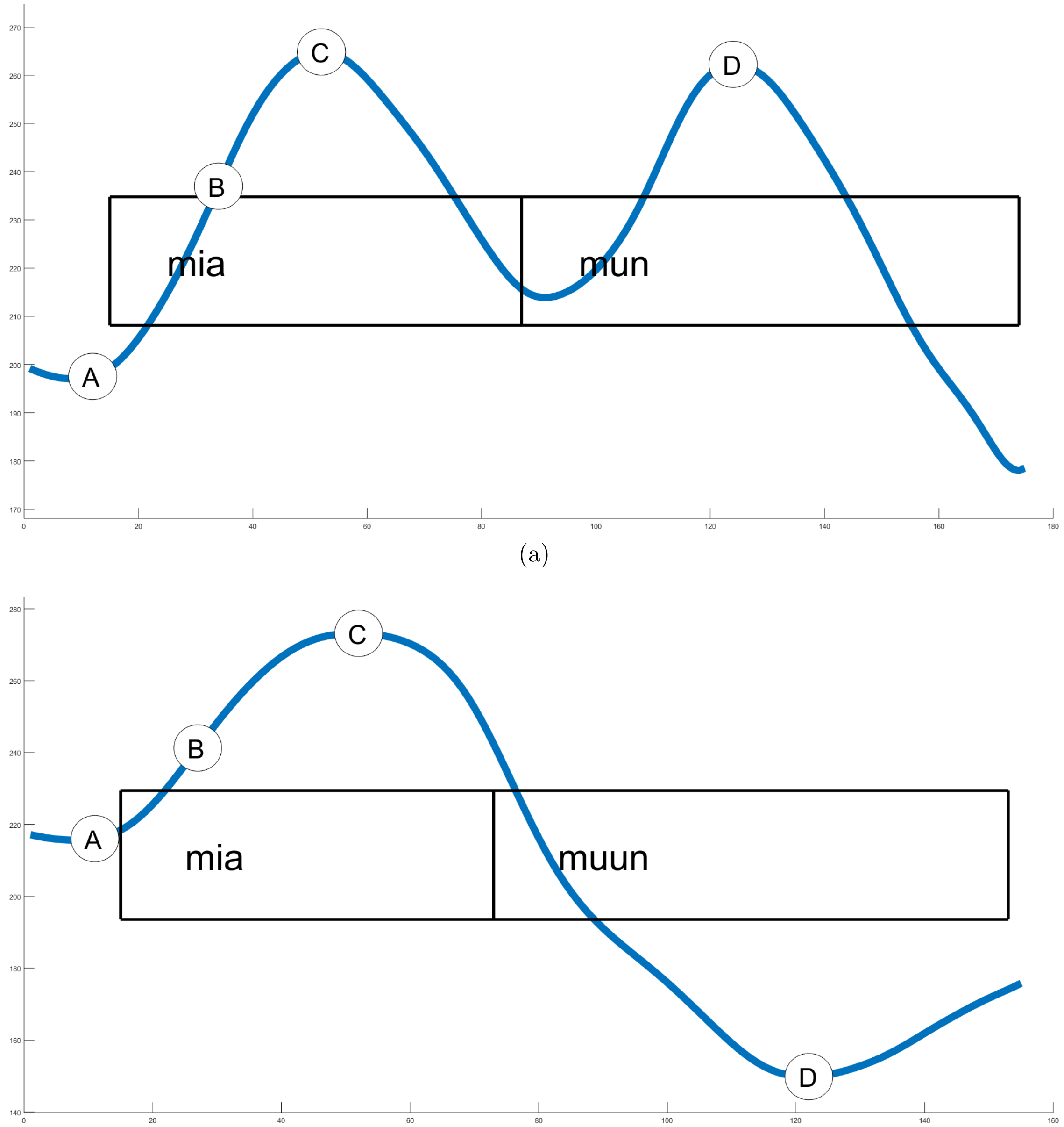


Figure 2.1: An example of two clearly defined boundaries between nasals, first between the coda /n/ of the carrier word and the onset /m/ of Word 1, and second between the coda /n/ of Word 1 and the onset /m/ of Word 2.

- **Word 2 F0 elbow** (D): The most extreme F0 of target word 2 (in Figure 2.2a, the maximum F0, as Word 2 is a Falling tone; in Figure 2.2b, the minimum F0, as Word 2 is a Rising tone). The only boundaries specified for this landmark were the acoustic left and right edges of the word.
- **Word 1 excursion onset** (A): The point at which the F0 trajectory reached 20% of Word 1 maximum onset speed (point marked at B). This point necessarily occurs before maximum onset speed is reached; the leftmost boundary was specified as $\frac{2}{3}$ of the way through the first carrier word (i.e., the kinship term).

Due to the lack of F0 plateaus in the vast majority of the pitch tracks in the Thai data, absolute minima and maxima were used as markers of elbow timing (compare to the experiments in Chapter 3 and Chapter 4, where peak offset was used as a more reliable marker of F0 timing due to the prevalence of F0 plateaus). However, due to the lack of



(a)

(b)

Figure 2.2: Two example landmarked trajectories (F+F and F+R). A = Initial pitch excursion onset; B = maximum onset speed; C = Word 1 F0 elbow; D = Word 2 F0 elbow.

consistent pitch extremum in the first carrier word, the beginning of the pitch excursion for Word 1 was defined at the 20% threshold, rather than some local low point before a Falling tone, or some local high point before a Rising tone. The use of this onset landmark both eliminates errors in trajectory marking due to small fluctuations in the relatively flat mid tone of the first word of the carrier phrase, and provides a consistent landmark for sentences with a Falling Word 1 and a Rising Word 1.

2.1.5.3 Time normalization

Due to wide variation in speech rate across speakers, as well as the need to analyze the role of the syllable vs. the mora in timing, analyses are performed both on raw and time-normalized data. Rather than referencing the duration of the entire carrier phrase, time normalization was performed within each target word only: time-normalized pitch and segmental landmarks are given as the proportion of the word at which they occur. For example, if a pitch elbow occurs 150 ms after the beginning of the word, and the word is 300 ms long, then the time-normalized pitch elbow landmark is 50%. This method of time normalization is used because it allows comparison across Word 1 and Word 2 despite phrasally-caused differences in the durations of Word 1 and Word 2—with this method, a pitch elbow occurring 150 ms after the beginning of Word 1 would have a smaller proportion (50% for a 300 ms word) than a pitch elbow occurring 150 ms after the beginning of Word 2 (e.g., 56% for a 270 ms word), but both would have the same proportion if compared to the duration of the whole phrase.

2.1.5.4 Statistical analyses

Throughout this chapter I will be using an α-level of 0.01; p-values below 0.05 but greater than 0.01 are considered “marginally significant” and the corresponding effects taken as a suggestion for further exploration. Statistical analyses were performed in R (R Core Team 2017), using the lme4 package (Bates, Maechler, Bolker, Walker, et al. 2014) for linear mixed effects models. Models were built and compared incrementally, starting with the null model, which includes just `Part` as a random effect. For all analyses, all predictors are first examined in single fixed-effect models (i.e., one fixed effect and `Part` as a random effect), and compared with the AIC (Akaike Information Criterion; lower values are better). Nested models were compared with likelihood ratio tests, using the `anova` function from the lmerTest package (Kuznetsova, Brockhoff, & Christensen 2015). Homoskedasticity and normality of the residuals were assessed graphically.

The analyses of this chapter fall into two categories: tones while collapsing across context (i.e., the effect of tone identity on timing—when examining the timing of Word 1, F+F and F+R are considered together, and R+F and R+R are considered together); and tones while considering context (i.e., the effect of context on timing—F+F and F+R are compared to each other).

The variables (presented in `monospace font`) used in the analyses of the pitch excursions are the following:

Random effects

- `Part` (participant): Random intercepts for participant are included in all linear models.

Order is not included as a random effect, as the target words and phrases were presented in random order in each block. Token (i.e., /muan/ vs. /mua/) is not included as a separate random factor because all aspects that distinguish tokens from each other are independent variables. For some analyses, random slopes are also included for `Part`, as there was significant variation that obscured results when all data was pooled together under the assumption of comparable slopes.

Fixed effects

- `Tone` (tone of the word being measured): categorical variable with two levels, `falling` or `rising`
- `Shape` (the shape of the rime that the word being measured has): categorical variable with four levels, `CVN`, `CVVN`, $\mathtt{CV_1V_2}$, or $\mathtt{CV_1V_2N}$

- WordDur (the duration of the word being measured): continuous variable, measured in seconds

- REMora (the timing of the right edge of the first mora, relative to the beginning of the word; time-normalized version is NormREMora): continuous variable, measured in seconds

- SameTone (if the two tones in the sequence are the same): categorical variable with two levels, same or different

Dependent variables For a schematic of these variables, see Figure 3.5.

- Elbow (timing of the F0 extremum—peaks for Falling tones, valleys for Rising tones; time-normalized version is NormElbow): The time interval between the acoustic beginning of the word and the F0 extremum (measured in seconds for raw data; percentage for time-normalized data)

- REMora (timing of the right edge of the first mora; time-normalized version is NormREMora): time interval between the beginning of the word and the end of the first mora (measured in seconds for raw data; percentage for time-normalized data)

- ExcurStart (start of the first pitch excursion): The time interval between the acoustic beginning of the word and the start of the upward F0 excursion (measured in seconds for raw data; percentage for time-normalized data)

- ExcurDur (excursion duration): The time interval between the peak offset and the start of the F0 excursion (measured in seconds for raw data; percentage for time-normalized data)

- ExcurSize (size of the first pitch excursion): The difference in Hz between the start of the excursion and the target of the excursion (measured in Hz)

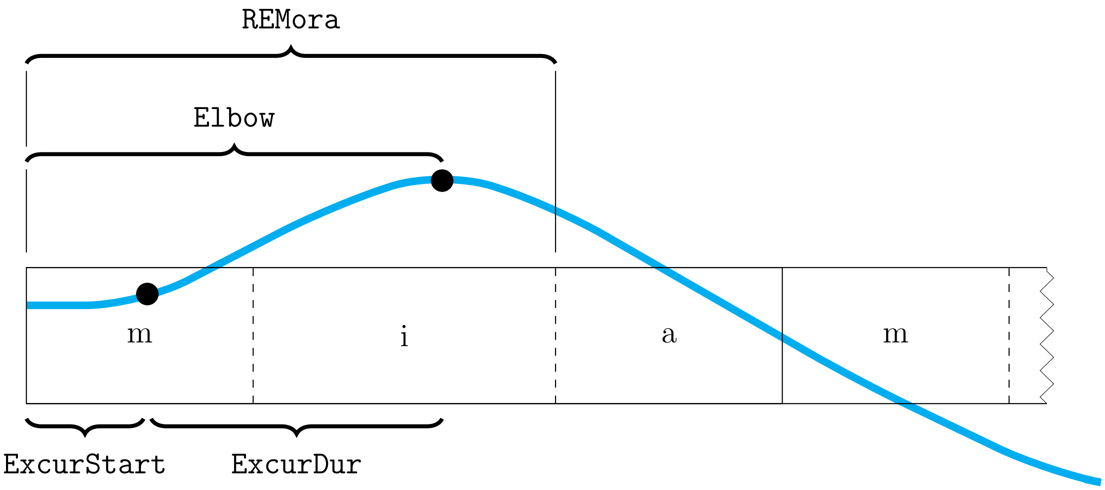


Figure 2.3: A schema of the dependent variables used in analysis, illustrated on Word 1 of a two-word sequence (schema is cut off before the end of Word 2). Segment boundaries are denoted by dashed lines; the word boundary is denoted by solid lines. The blue contour is a schematized Falling + Rising contour, with black dots to mark the start (leftmost) and peak (rightmost) of the first tone pitch excursion.

2.2 Results

The dataset used for analysis is a subset of the total trials collected. With six participants each doing 8 blocks of 32 sentences, there are a total of 1,536 possible trials. Trials with recording or production errors were excluded from analysis. Trials with too great a pause between Word 1 and Word 2 were also excluded from analysis, even if the pause was due to prosodic focus and not some error in reading or production, as the phrase break could have an effect on the timing of Word 1. A total of 19 trials (1.2%) were excluded due to pausing. This dataset of 1,517 trials is used for segment-only analyses (such as duration).

Of the remaining 1,517 trials, some were also marked for removal if the pitch tracking failed to produce a reasonable contour. In some cases (particularly in words with Rising tones), this was due to a short period of creakiness near the lowest part of the contour, which obscured the location of the F0 elbow. In other cases (particularly for Word 2), there were pitch perturbations that prevented reliable marking of F0 elbows. Only the word that was affected by the bad pitch tracking was removed from analysis; thus, if a trial had a good Word 1, but a bad Word 2, Word 1 is still included in Word 1 analyses, while Word 2 is excluded from Word 2 analyses. An additional 29 trials (1.9%, for a total of 1,488) were removed from Word 1 analyses, and 58 (3.8%, for a total of 1,459) were removed from Word 2 analyses. Of the trials removed due to pitch tracking errors, 7 trials were marked as bad for both Word 1 and Word 2.

Finally, for analyses of initial excursion duration and initial excursion size, only trials with clearly distinguishable excursion onsets (in addition to clear F0 elbows) were included. Excursion characteristics were only analyzed for Word 1, because Word 2 does not have an obvious excursion onset point in F+R and R+F sequences (where there is instead a smooth trajectory from one elbow to the next). From the dataset of 1,488 trials included in Word 1 pitch analyses, 38 additional trials were excluded (2.6% of tokens with correctly marked Word 1 elbows).

Table 2.3: A summary of the data used for pitch analyses.

	Word 1	Word 1 onset	Word 2
F + F	375	356	374
F + R	371	363	348
R + R	367	361	359
R + F	375	370	378
Total	1,488	1,450	1,459

2.2.1 Environment 1: Timing of tones collapsed across context

In this section, I discuss the characteristics of the pitch contours associated with the Falling and Rising tones, regardless of the tone context. For example, if discussing the timing of Falling tones on Word 1, both **F**+F and **F**+R trials are included in the dataset; similarly, if discussing the timing of Falling tones on Word 2, both F+**F** and R+**F** trials are included.

In this section, analyses that include only segmental data (i.e., word or segment durations) or categorical variables (i.e., the tone or segmental shape of the word) use the datasets that have no segmental errors, but may have pitch tracking errors. If any model being compared in the analysis includes a continuous variable related to F0 (such as the timing of the pitch elbow), the entire set of models is run using the dataset that does not have any pitch tracking errors. Furthermore, whenever discussing pitch trajectories of Word 1, I use the dataset that has no pitch tracking or segmental errors that affect Word 1, and when discussing Word 2, the dataset with no errors that affect Word 2.

2.2.1.1 Segmental characteristics

Effects of phrase position on word duration There is an effect of phrase position (Word 1 vs. Word 2) on the duration of the word. Overall, Word 1 is significantly longer than Word 2 (F(1,3032) = 240.4, $p < 0.0001$). The difference between the two word positions is approximately 30 ms (Word 1 M = 349.4 ms, SD = 53.0 ms; Word 2 M = 318.4 ms, SD = 57.2 ms). The effect size, while still small, also indicates a meaningful difference ($\eta^2 = 0.07$). There are two possible sources for this difference: first, Word 1 and Word 2 serve different syntactic functions and are in prosodically different positions; as mentioned, Word 1 serves as a name, while Word 2 is a verb. Second, the two words use different nuclei; in particular, the monophthongal nuclei are /a/ and /aa/ for Word 1, and /u/ and /uu/ for Word 2, which suggests that the difference between word positions is driven by vowel height differences in only a subset of the tokens.[3] However, there is no significant interaction between nucleus type and word position (F(2,3028) = 2.27, p = 0.10); for all nucleus types, Word 2 is significantly shorter than Word 1 (see Figure 2.4). Thus, it appears that prosodic position is the main contributor to Word 1 vs. Word 2 durational differences.

[3] Compare the diphthong tokens /mia(n)/ and /mua(n)/ for Word 1 and Word 2, respectively, which are both high vowels followed by low vowels.

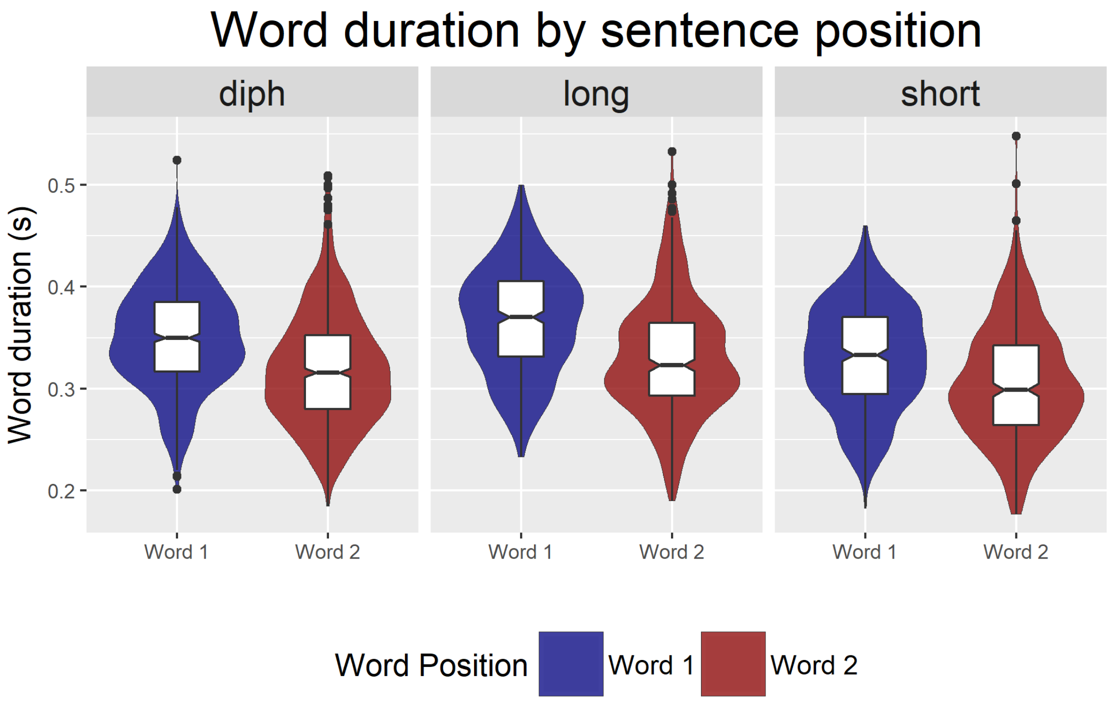


Figure 2.4: Violin plots comparing the durations of Word 1 and Word 2, divided by word shape.

Effects of segmental structures on duration As predicted by Hypothesis A, there are differences in duration between words of different shapes for both Word 1 and Word 2 ($\chi^2(3)$ = 215.24, $p < 0.0001$ compared to the null model for Word 1; $\chi^2(3) = 283.96$, $p < 0.0001$ compared to the null model for Word 2; see Table 2.4). For Word 1, CVN words are the shortest (M = 330.0 ms, SD = 51.5 ms), followed by CV_1V_2 words (M = 343.4 ms, SD = 53.6 ms), followed by CV_1V_2N words (M = 357.2 ms, SD = 46.5 ms), and CVVN words are the longest (M = 366.5 ms, SD = 53.0 ms); all shapes are significantly different from each other (using a least squares mean Tukey test, $p < 0.0001$ for all except CV_1V_2N and CVVN, which are $p = 0.002$). Word 2 behaves similarly, though the word shapes fall into two groups: CVN and CV_1V_2 pattern together as the shortest words (CVN M = 305.8 ms, SD = 56.9 ms; CV_1V_2 M = 301.7 ms, SD = 47.3 ms; $p = 0.41$) and CV_1V_2N and CVVN

Table 2.4: Comparison of single-factor linear mixed effects models for WordDur, for Word 1 and Word 2.

(a) Single predictor models for WordDur of Word 1.

Model for WordDur	AIC	χ^2	DegF	p[†]
Tone + (1|Part)	-5687.3	11.02	1	0.0009**
Shape + (1|Part)	-5897.5	215.24	3	< 0.0001**
[†]As compared to the null model, WordDur ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Single predictor models for WordDur of Word 2.

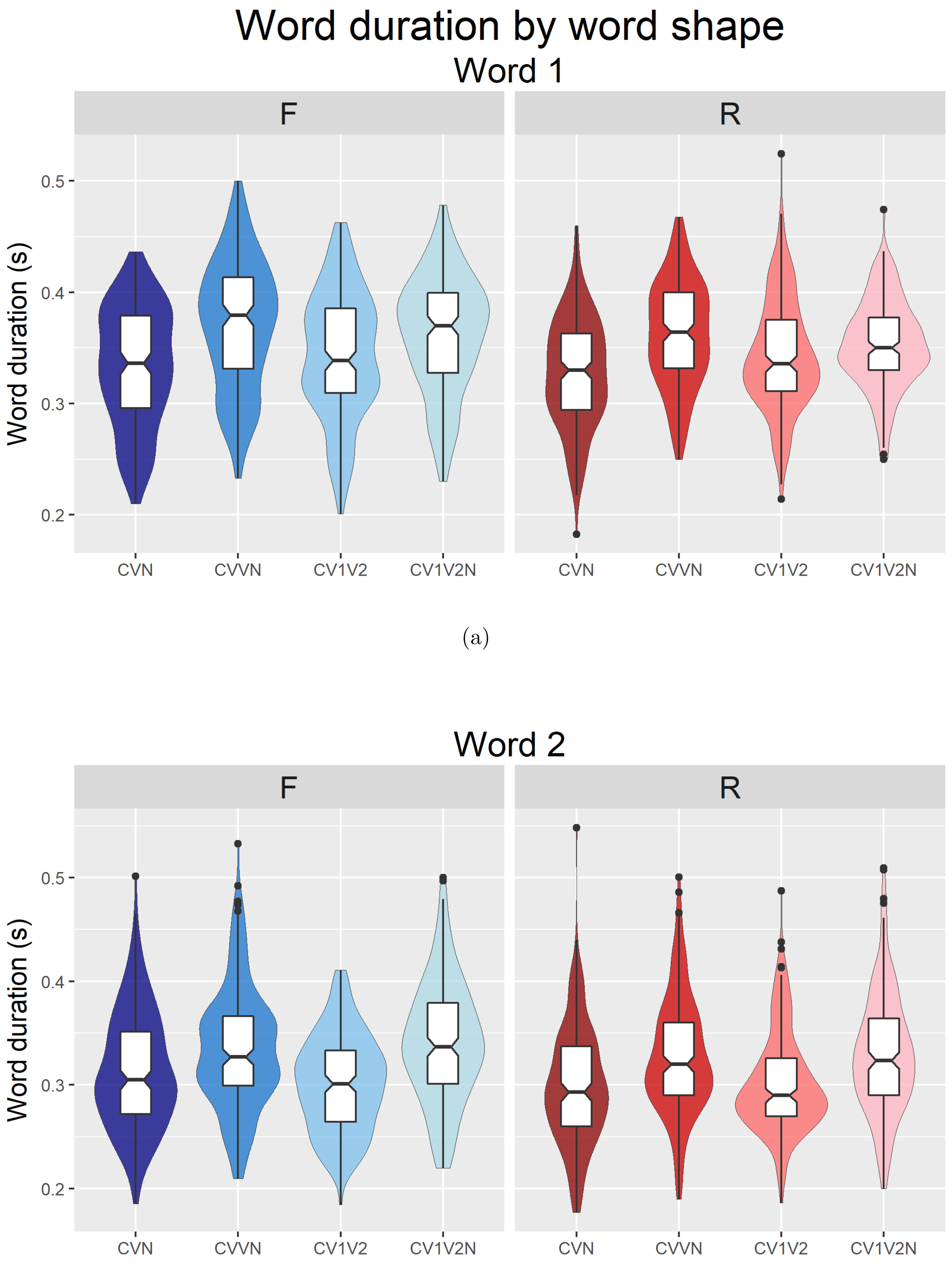
Model for WordDur	AIC	χ^2	DegF	p[†]
Tone + (1|Part)	-5846.4	21.13	1	< 0.0001**
Shape + (1|Part)	-6105.3	283.96	3	< 0.0001**
[†]As compared to the null model, WordDur ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

pattern together as the longest words (CV_1V_2N M = 335.2 ms, SD = 58.1 ms; CVVN M = 330.6 ms, SD = 58.0 ms; p = 0.31; $p < 0.0001$ for all other comparisons). Thus, the first prediction from Condition B is shown in the data. These patterns are illustrated in Figure 2.5.

There is also an effect of Shape on the duration of the first mora.[4] For Word 1, the addition of Shape as a single fixed effect significantly improves the model ($\chi^2(2) = 157.5$, $p < 0.0001$). CV_1V_2N words have the shortest first mora (M = 191.7 ms, SD = 30.2 ms), followed by CVN words (M = 204.7 ms, SD = 30.7 ms), and finally CV_1V_2 words have the longest first mora (M = 214.9 ms, SD = 35.8 ms, all $p < 0.0001$). Shape also significantly improves the fit of the model for Word 2 ($\chi^2(219.5) = 2$, $p < 0.0001$), and the patterns are similar to those in Word 1: CV_1V_2N words have the shortest first mora (M = 177.2 ms, SD = 27.5 ms), followed by CVN words (M = 182.0 ms, SD = 31.6 ms), with CV_1V_2 words again having the longest first mora (M = 201.7 ms, SD = 30.6 ms; p = 0.008 compared to CVN). Thus, Condition A is satisfied.

Note that the duration of the nucleus in the two word shapes with diphthongs (i.e., /mia/

[4] These analyses do not include CVVN words, as it is impossible to mark the end of the first mora.



(a)

(b)

Figure 2.5: Box plots the durations of words for each word shape for both Word 1 and Word 2, separated by tone.

vs. /mian/, and /mua/ vs. /muan/) is not the same. For both Word 1 and Word 2, the diphthong nucleus is shorter when there is a coda ($p < 0.0001$ for both): /mia/ M = 253.4 ms, SD = 47.4 ms vs. /mian/ M = 188.7 ms, SD = 33.6 ms; /mua/ M = 200.4 ms, SD = 43.8 ms vs. /muan/ M = 152.9 ms, SD = 35.6 ms. One might expect that the locus of this difference would rest entirely within the second element of the diphthong due to mora sharing;[5] however, this is not actually the case. Rather, both elements of the diphthong are affected ($p < 0.0001$ for all): both the [i] and [a] in /mia/ ([i] M = 125.0 ms, SD = 29.2 ms; [a] M = 128.4 ms, SD = 32.3 ms) are longer than the [i] and [a] in /mian/ ([i] M = 102.6 ms, SD = 24.2 ms; [a] M = 86.1 ms, SD = 19.7 ms). This is also true for Word 2: both the [u] and [a] in /mua/ ([u] M = 100.4 ms, SD = 23.4 ms; [a] M = 100.0 ms, SD = 23.3 ms) are longer than the [u] and [a] in /muan/ ([u] M = 80.4 ms, SD = 18.0 ms; [a] M = 72.5 ms, SD = 19.3 ms).

On the other hand, it is also the case that the magnitude of difference is greater for the second element (compare 23 ms difference to 42 ms difference for the /ia/ diphthong elements, and 20 ms difference to 28 ms difference for the /ua/ diphthong elements); the effect size is also greater for the second element in both cases (compare $\eta^2 = 0.15$ for /i/ to $\eta^2 = 0.39$ for /a/; $\eta^2 = 0.19$ for /u/ to $\eta^2 = 0.29$ for /a/). Thus, the presence of the coda does have a greater shortening effect on the element of the diphthong that it is sharing a mora with. However, the presence of the coda in fact shortens both elements of the nucleus.[6]

2.2.1.2 Pitch characteristics

Elbow timing As predicted by Hypothesis B, overall, F0 elbows consistently occur earlier in Falling tones than in Rising tones. This is true for both raw duration and time-normalized data. The difference in timing occurs both in Word 1 (Falling M = 176.6 ms, SD = 37.1 ms, or 49.4% through the word; Rising M = 220.4 ms, SD = 42.1 ms, or 63.2% through the word;

[5] Remember that syllables in Thai are maximally bimoraic.

[6] It is not the case that the entire rest of the word is affected by the presence of a coda: the duration of the syllable onset /m/ is the same for Word 1 ($p = 0.51$); for Word 2, the /m/ in /muan/ is shorter ($p = 0.005$), but the difference in means is approximately 5 ms (/mua/ M = 101.3 ms, SD = 24.6 ms; /muan/ M = 96.8 ms, SD = 20.0 ms) and the effect size is small ($\eta^2 = 0.01$), indicating again that the difference is significant but not meaningful.

$p < 0.0001$) and Word 2 (Falling M = 164.0 ms, SD = 37.1 ms, or 49.3% through the word; Rising M = 202.2 ms, SD = 46.6 ms, or 61.8% through the word; $p < 0.0001$). Similarly, there is a significant effect of `Tone` on `ExcurDur` ($\chi^2(1) = 211.15$, $p < 0.0001$ for Word 1; $\chi^2(1) = 115.43$, p = for Word 2), where the initial excursion of Rising tones is longer than the initial excursion of Falling tones (difference between estimates is 21.5 ms, SE = 1.4 ms for Word 1; 19.8 ms, SE = 1.8 ms for Word 2). Thus, the Falling and Rising accents are not simple mirror images of each other.

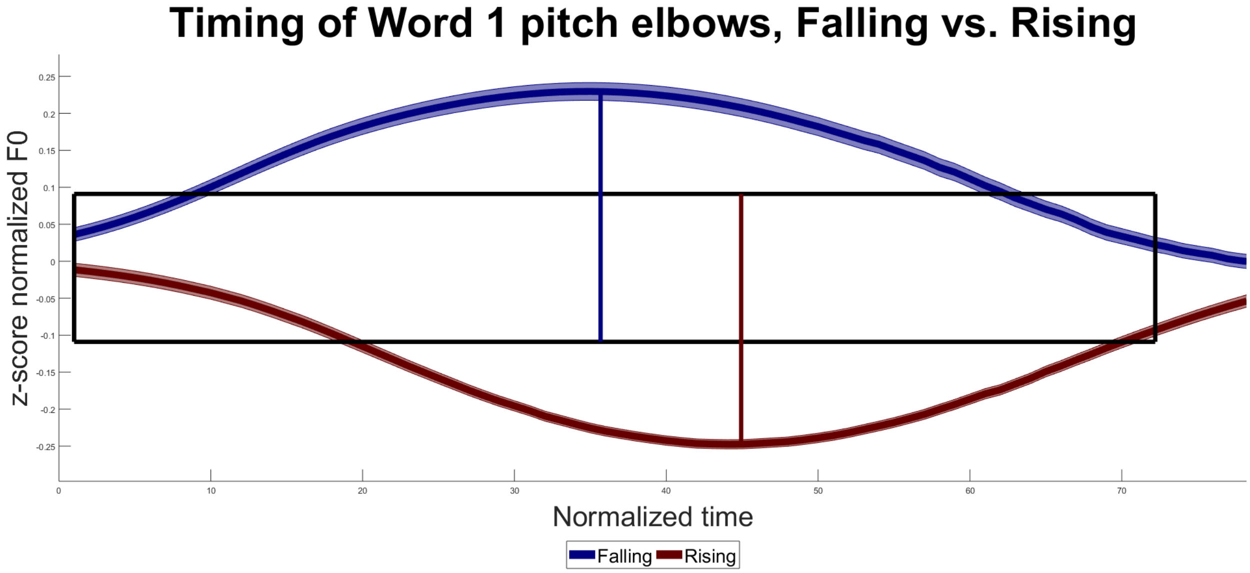
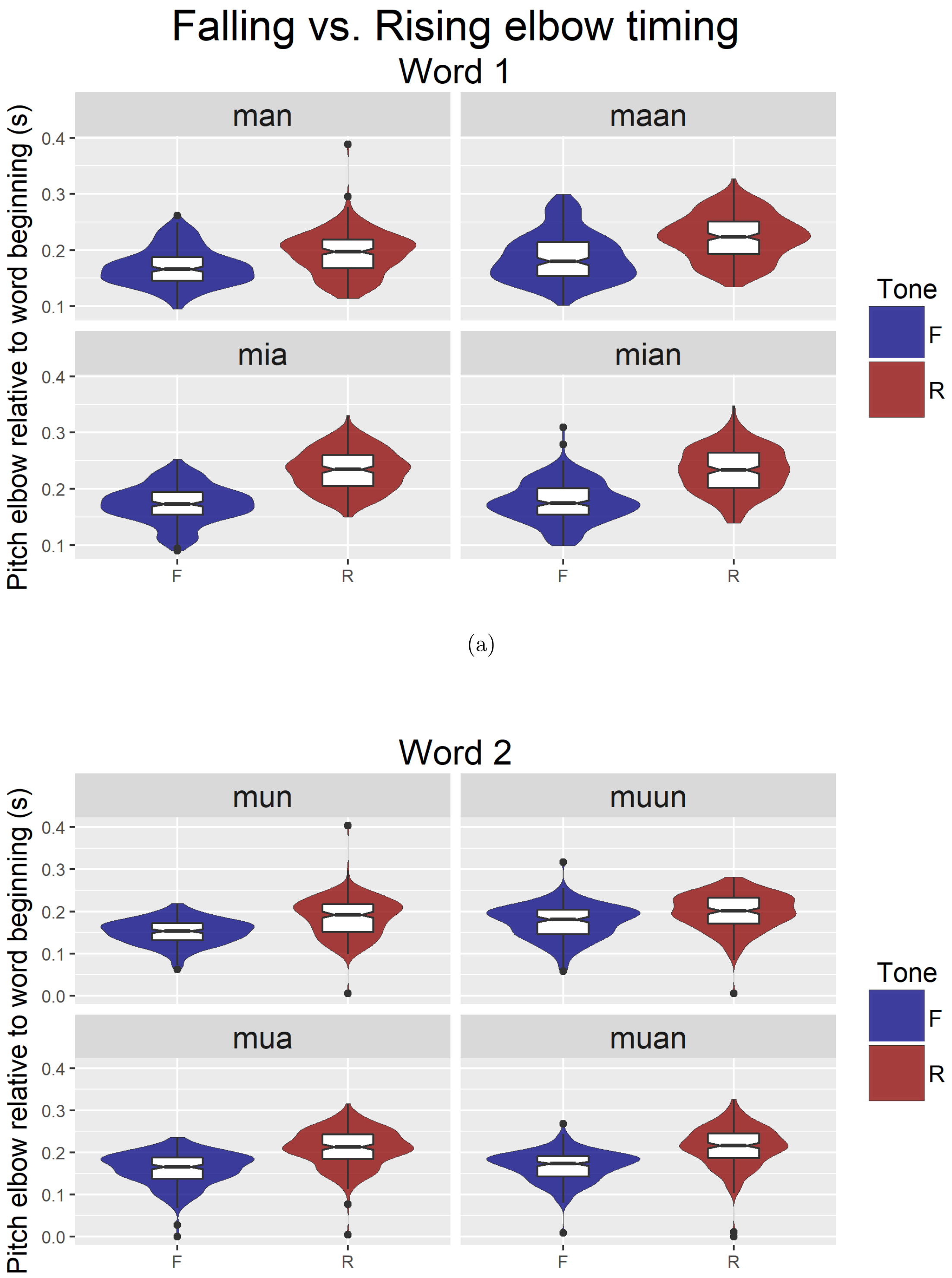


Figure 2.6: Time- and z-score normalized F0 trajectories of Falling (blue) and Rising (red) tones in Word 1, including all participants. The rectangle represents the acoustic start and end of the word; elbows are marked for each trajectory.

This difference in elbow timing also cannot be attributed to articulatory constraints: the pattern exhibited by Thai contour tones in this study is in fact the reverse of what is expected (and in fact runs contrary to the results found by Nitisaroj (2006), who found that the pitch elbow occurs earlier in Rising tones than in Falling tones). Xu and Sun (2002) found that, for equal duration of F0 change, *increases* in F0 are characterized by (1) a smaller excursion size, (2) lower average speed, and (3) a lower maximum velocity. Thus, the prediction is that increases in F0 would be allotted more time than decreases in a contour tone, or that words with F0 rises are longer than words with only F0 falls. However, the Thai Rising tone is precisely the opposite: the *falling* portion of the contour takes up 61.8% of the word, which generates the delayed elbow. In terms of raw duration, it is also not true that the remaining 38.1% of Rising tones is as long as the 49.3% of the word allotted to the F0 rise in Falling tones.

A comparison of linear mixed-effects models shows that the addition of `Tone` significantly improves the fit of the model compared to the null model for both Word 1 ($\chi^2(1) = 636.78$, $p < 0.0001$ for millisecond data; $\chi^2(1) = 936.25$, $p < 0.0001$ for normalized data) and Word 2 ($\chi^2(1) = 415.51$, $p < 0.0001$ for millisecond data; $\chi^2(1) = 586.19$, $p < 0.0001$ for normalized data). A comparison between a model with random intercepts for `Part` and a model with random slopes for `Part` indicates that the addition of random slopes significantly improves on a model with `Tone` as a fixed effect for both Word 1 ($\chi^2(2) = 35.55$, $p < 0.0001$ for millisecond data; $\chi^2(2) = 25.59$, $p < 0.0001$ for normalized data) and Word 2 ($\chi^2(2) = 111.36$, $p < 0.0001$ for millisecond data; $\chi^2(2) = 115.38$, $p < 0.0001$ for normalized data), indicating that some speakers have greater differences in elbow timing between Rising and Falling tones. (See Table 2.5 for model comparisons with millisecond data, and Table 2.6 for model comparisons with time-normalized data.)

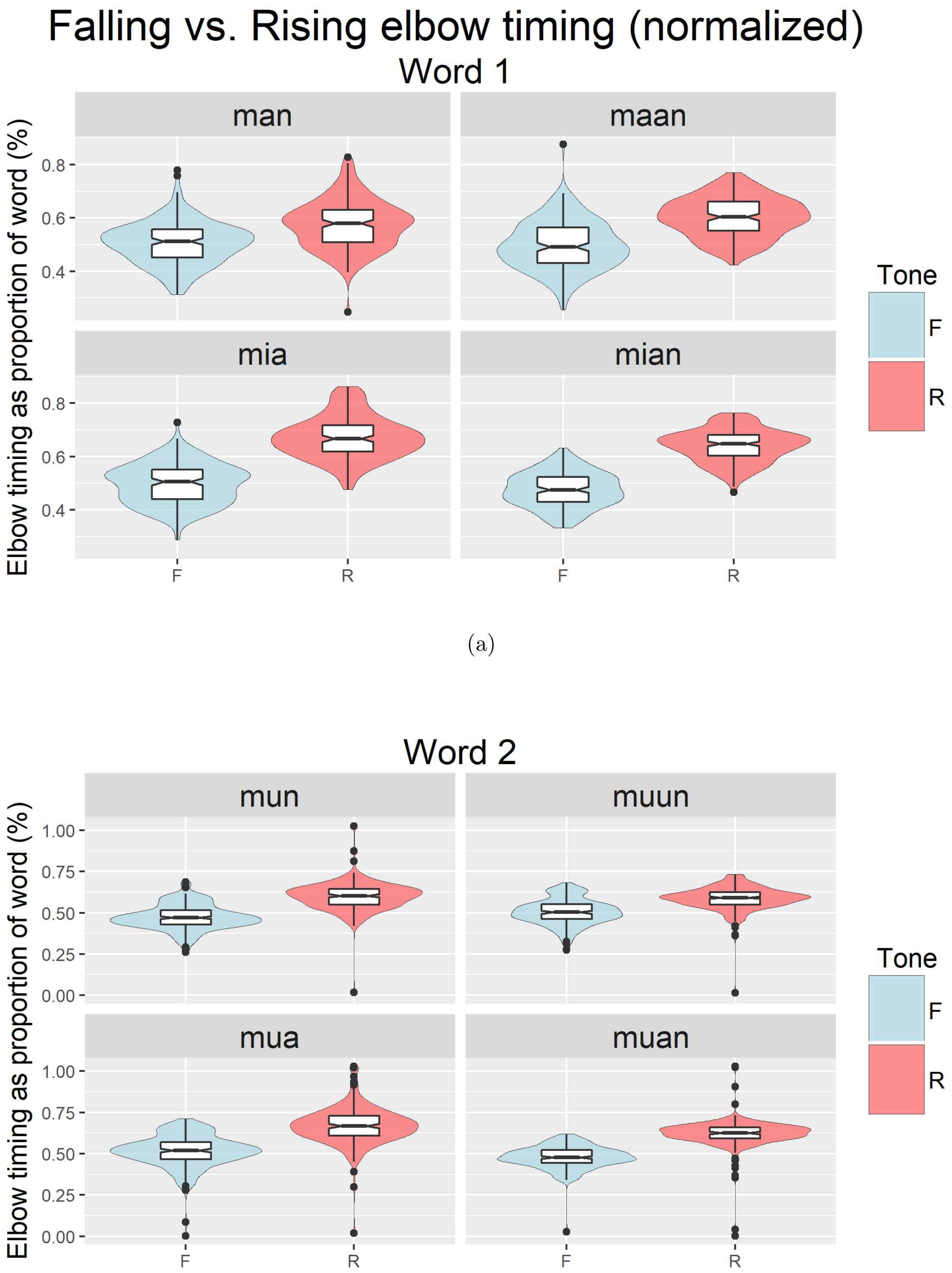
Despite individual differences in the duration of words with Falling vs. Rising words, the overall pattern where Falling elbows occur earlier than Rising elbow holds for all participants. Within this pattern there is a fair amount of inter-speaker variation, even when considering time-normalized data. For example, for participant BM01 the Falling tone elbow occurs at approximately 42% of the word, and the Rising tone at approximately 55%—which is where the Falling tone elbow for participants RH01 and SM01 occurs (54.5% and 54.6%, respectively; see Table 2.7a for all coefficients). It is also not the case that these differences are due to the combination of by-speaker variation in rate of speech and some target delay of x milliseconds from the onset of the word. For this to be true, participant BM01 and CP01



(a)

(b)

Figure 2.7: Box plots comparing the elbow timing of Falling (blue) and Rising (red) accents for each word position.



(a)

(b)

Figure 2.8: Box plots comparing the elbow timing of Falling (blue) and Rising (red) accents for each word position.

Table 2.5: Comparison of nested models for `Elbow` for Word 1 and Word 2 (millisecond data).

(a) Word 1.

Model for `Elbow`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`Tone + (1|Part)`	636.78	1	< 0.0001**
`Tone + (1+Tone|Part)`	35.55	2	< 0.0001**
`1 + (1+Tone|Part)`	—	—	—
`Tone + (1+Tone|Part)`	16.52	1	< 0.0001**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

(b) Word 2.

Model for `Elbow`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`Tone + (1|Part)`	415.51	1	< 0.0001**
`Tone + (1+Tone|Part)`	111.36	2	< 0.0001**
`1 + (1+Tone|Part)`	—	—	—
`Tone + (1+Tone|Part)`	9.30	1	0.002*
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

would have to have the slowest rate of speech, with words of greater duration (such that the same number of milliseconds would correspond to a smaller proportion of the word)—and this is not true, as illustrated in Figure 2.9. CP01 is, in fact, the speaker with the fastest rate of speech, while SM01 is the slowest; this is the opposite of what would be predicted. Therefore, there is in fact wide variation in tone timing.

A comparison of the random slopes model with the random intercepts model indicates that the addition of random slopes does improve the fit of the model, both for Word 1 ($\chi^2(2) = 25.59$, $p < 0.0001$) and Word 2 ($\chi^2(2) = 79.41$, $p < 0.0001$). At the same time, it is consistently true that the elbow occurs later in Rising tones than in Falling tones—i.e., it is not like in the case of word duration, where some slopes were negative and others were positive.

Table 2.6: Comparison of nested models for Elbow for Word 1 and Word 2 (time normalized data).

(a) Word 1.

Model for NormElbow	χ^2	DegF	p†
1 + (1|Part)	—	—	—
Tone + (1|Part)	936.25	1	< 0.0001**
Tone + (1+Tone|Part)	25.59	2	< 0.0001**
1 + (1+Tone|Part)	—	—	—
Tone + (1+Tone|Part)	20.57	1	< 0.0001**
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

(b) Word 2.

Model for NormElbow	χ^2	DegF	p†
1 + (1|Part)	—	—	—
Tone + (1|Part)	586.19	1	< 0.0001**
Tone + (1+Tone|Part)	115.38	2	< 0.0001**
1 + (1+Tone|Part)	—	—	—
Tone + (1+Tone|Part)	11.05	1	0.0009**
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

2.2.1.3 Effects of tones on segments

In this section, we turn to the effects of tones on segments. In determining predictions of the related hypotheses, there are two possibilities: either the mora is the TBU (and thus the unit to be potentially affected by tone), or the syllable is the TBU. The latter does not predict any effects, as there is no intrinsic reason for Falling and Rising tones to require (for example) different durations.

Hypothesis 1.0 (null hypothesis): There is no effect of tone identity on segmental realization.

Prediction 1.0:

- Tone will not be a significant predictor of WordDur.

Table 2.7: Estimates for proportion of word until the pitch elbow, by tone, where the intercept is the Falling tone and the direction and size of the effect is given under Rising.

(a) Word 1.

Partic.	**Falling**	**Rising**
BM01	42.0%	+13.3%
CP01	49.2%	+11.4%
TS01	51.2%	+16.0%
MA01	45.1%	+14.0%
RH01	54.5%	+13.3%
SM01	54.6%	+9.5%

(b) Word 2.

Partic.	**Intercept**	**Rising**
BM01	48.5%	+5.2%
CP01	46.9%	+13.2%
TS01	44.3%	+19.7%
MA01	50.5%	+16.4%
RH01	47.3%	+14.2%
SM01	60.0%	+6.3%

- `Tone` will not be a significant predictor of `REMora`.

Hypothesis 1.1: Tone plays a significant role in the timing of the first mora.

Prediction 1.1: `Tone` will be a significant predictor of `REMora`, and `Elbow` will parallel moraic timing: Falling tones (with early extrema) will have shorter first moras, and Rising tones (with late extrema) will have longer first moras.

Effects of tone on word duration There is an effect of `Tone` on the duration of words. In a linear mixed effects model with random intercepts only (where `Part` is the only random effect), `Tone` significantly improves the model ($\chi^2(1)$ = 11.02, p = 0.0009 compared to the null model). However, the difference between estimated means is just 6 ms (β = -6.2 ms, SE = 1.9 ms for the Rising tone).

In addition, there are large individual differences: for some speakers, Falling words are longer than Rising words, while the reverse is true for others (see Figure 2.9). In some cases, the difference seems meaningful, such as for MA01, where words with Rising tones are 20.5 ms longer, and for SM01, where words with Rising tones are 23.4 ms shorter. In other cases, the difference seems to be spurious—for example, it does not seem likely that the 5.2 ms difference for participant TS01 is meaningful, even if it is statistically significant.

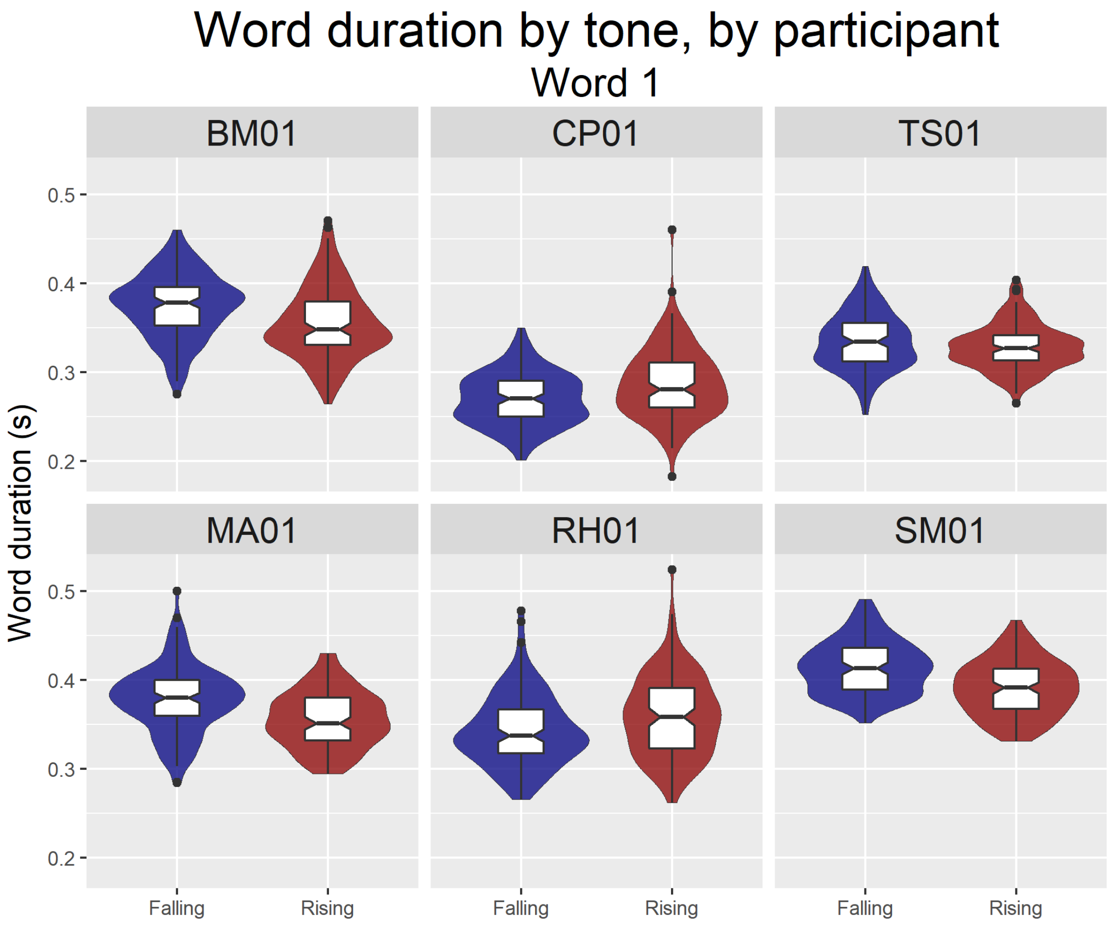


Figure 2.9: Violin plots comparing the durations of Falling and Rising tone words, Word 1.

In a linear mixed effects model specified for random slopes as well as intercepts (denoted as `(1+Tone|Part)`), `Tone` as a single fixed effect does not significantly improve the model ($\chi^2(1) = 0.74$, p = 0.39).

Word 2 patterns slightly differently, though with the same ultimate outcome, where `Tone` does not significantly (or meaningfully) affect word duration. In a linear mixed effects model with random intercepts for `Part` only, `Tone` significantly improves the model ($\chi^2(1) = 21.14$, $p < 0.0001$ compared to the null model). However, similarly to Word 1, the difference between Falling and Rising tones is very small, just 8 ms (β = -8.2 ms, SE = 1.8 ms for the

Rising tone).

In this case, the variation between participants is not as extreme: in almost all cases, Rising tone words are shorter than Falling tone words. In addition, for Word 2, the differences between Falling and Rising tone are even less smaller than for Word 1—participants TS01 and MA01 approach meaningful durational differences at 18 and 14 ms, respectively, but for everybody else, the difference remains in the ones digit of milliseconds. Thus, it seems unlikely that these differences are intentional and encoded in the phonology—even in the case of Word 1, the inconsistency of the patterning suggests some cause other than `Tone`, be it prosodic or purely phonetic. In a linear mixed effects model specified for random slopes as well as random intercepts, the effect of `Tone` is marginal at best ($\chi^2(1) = 3.77$, p = 0.05 compared to the null model).

Effects of tone on first mora duration Although there are robust differences in elbow timing between Falling and Rising tones, there is no meaningful corresponding difference in the tone-bearing segments. `Tone` is a significant predictor of `REMora` for both Word 1 ($\chi^2(1)$ = 16.85, $p < 0.0001$; see Table 2.9a) and Word 2 ($\chi^2(1) = 18.11$, $p < 0.0001$; see Table 2.10a). However, two aspects of these estimates indicate that the difference is not meaningful. First, in both cases, the difference in the duration of the first mora in Falling vs. Rising tones is miniscule, on the scale of 6 ms: for Word 1, Falling β = 207.2 ms, SE = 8.5 ms vs. Rising β = 200.6 ms, SE = 1.6 ms, and for Word 2, Falling β = 190.0 ms, SE = 7.7 ms vs. Rising β = 183.6 ms, SE = 1.5 ms. Second, the difference is actually in the opposite direction—that is, Rising tones are associated with a shorter first mora, even though the elbows of Rising tones are, on average, much later than the elbows of Falling tones.

Thus, there is no evidence that differences in the duration of the first pitch excursion cause parallel differences in the duration of the first TBU: even though the first pitch excursion is longer in the Rising tone than in the Falling tone, the first mora has the same duration in both cases (and, as there is no overall difference in word length, the second mora does

Table 2.8: Comparison of random intercept vs. random slope models for `WordDur` for Word 1 and Word 2.

(a) Comparison of nested models, Word 1.

Model for `WordDur`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`Tone + (1|Part)`	11.02	1	0.0009**
`Tone + (1+Tone|Part)`	69.87	2	< 0.0001**
`1 + (1+Tone|Part)`	—	—	—
`Tone + (1+Tone|Part)`	0.74	1	0.39
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of nested models, Word 2.

Model for `WordDur`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`Tone + (1|Part)`	21.14	1	< 0.0001**
`Tone + (1+Tone|Part)`	10.49	2	0.005*
`1 + (1+Tone|Part)`	—	—	—
`Tone + (1+Tone|Part)`	3.77	1	0.05
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

as well). There is also no evidence that changes in the duration of the first TBU cause changes in the duration of the first pitch excursion; the first mora in /mian/ is significantly shorter than in /mia/ or /man/—in terms of both milliseconds and proportion of the word—, but there is no corresponding decrease in the duration of the first pitch excursion. Thus, Hypothesis 1.0 is not rejected; tone identity does not determine the timing of the first mora.

2.2.1.4 Effects of segments on tone realization

In this section, I address the effect of segments on tone realization. Here too the tone-alignment capacity of the mora and the syllable can be differentiated. For these analyses, all comparisons will be carried out within tone category, as it has already been shown that the Falling and Rising tones have different extremum timing.

Table 2.9: Comparison of single-factor models for `REMora` (raw data) and `NormREMora` (time normalized data), excluding CVVN syllables, Word 1.

(a)

Model for `REMora`	AIC	χ^2	DegF	p[†]
`Tone + (1|Part)`	-4886.1	16.85	1	< 0.0001**
`Word + (1|Part)`	-5017.8	150.57	2	< 0.0001**
[†]As compared to the null model, `REMora` ~ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b)

Model for `NormREMora`	AIC	χ^2	DegF	p[†]
`Tone + (1|Part)`	-2836.6	6.87	1	0.009*
`Word + (1|Part)`	-3384.8	557.05	2	< 0.0001**
[†]As compared to the null model, `NormREMora` ~ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

Table 2.10: Comparison of single-factor models for `REMora` (raw data) and `NormREMora` (time normalized data), excluding CVVN syllables, Word 2.

(a)

Model for `REMora`	AIC	χ^2	DegF	p[†]
`Tone + (1|Part)`	-4907.1	18.11	1	< 0.0001**
`Word + (1|Part)`	-5095.7	208.67	2	< 0.0001**
[†]As compared to the null model, `REMora` ~ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b)

Model for `NormREMora`	AIC	χ^2	DegF	p[†]
`Tone + (1|Part)`	-2653.0	2.42	1	0.12
`Word + (1|Part)`	-3689.6	1041.1	2	< 0.0001**
[†]As compared to the null model, `NormREMora` ~ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

Hypothesis 2.0 (null hypothesis): There is no effect of segment duration on tone realization.

Prediction 2.0:

- `WordDur` will not be a significant predictor of `Elbow` (within tone category).
- `REMora` will not be a significant predictor of `Elbow` (within tone category).

Hypothesis 2.1: Syllables are a main driving force in Thai tone timing.

Prediction 2.1:

- `Shape` *is* a significant predictor of `Elbow` (in milliseconds, within tone category);
- `Shape` is *not* a significant predictor of `NormElbow` (in percentage, within tone category)—that is, all word shapes will have the same extremum timing.

Hypothesis 2.2: Moras are a main driving force in Thai tone timing.

Prediction 2.2: `REMora` is a significant predictor of `Elbow` (within category).

Effect of word shape The addition of `Shape` to a model that already has `Tone` (i.e., separated for Falling vs. Rising) significantly improves the model for both Word 1 ($\chi^2(3) = 174.57$, $p < 0.0001$; see Table 2.11a) and Word 2 ($\chi^2(3) = 102.28$, $p < 0.0001$; see Table 2.11b). Elbow timing by word patterns roughly in parallel with word duration—i.e., CVN and CV_1V_2 words have the shortest time interval between the beginning of the word and the elbow, and CVVN and CV_1V_2N words have the longest. This is true for both tones and both phrase positions (see Figure 2.10 for illustration).

There is also a significant interaction between `Tone` and `Shape` ($\chi^2(3) = 125.66$, $p < 0.0001$ for Word 1; $\chi^2(3) = 25.53$, $p < 0.0001$ for Word 2; see Table 2.11); that is, word shape affects elbow timing differently for each tone. For Word 1, only CVVN is significantly different from the other word shapes for Falling tones (using a least squares means Tukey test, $p < 0.0001$ compared to CV_1V_2 and CVN, $p = 0.002$ compared to CV_1V_2N), but for Rising tones, all are significantly different (at $p < 0.01$) except CV_1V_2 and CV_1V_2N ($p = 1.00$). For Word 2, Falling tone words show significant differences between CVN and CV_1V_2N ($p < 0.0001$), CVN and CVVN ($p < 0.0001$), and CV_1V_2 and CVVN ($p = 0.004$), while the other comparisons are not significant[7]; Rising tone words show significant differences between all

[7] Note that this follows word duration patterns, which were in the increasing order 1. CVN 2. CV_1V_2 3.

Table 2.11: Comparison of nested models for `Elbow` for Word 1 and Word 2 (millisecond data).

(a) Word 1.

Model for `Elbow`	χ^2	DegF	p[†]
`Tone + (1|Part)`	—	—	—
`Tone + Shape + (1|Part)`	174.57	3	< 0.0001**
`Tone + Shape + Tone:Shape + (1|Part)`	125.66	3	< 0.0001**

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

(b) Word 2.

Model for `Elbow`	χ^2	DegF	p[†]
`Tone + (1|Part)`	—	—	—
`Tone + Shape + (1|Part)`	102.28	3	< 0.0001**
`Shape + Tone + Tone:Shape + (1|Part)`	29.53	3	< 0.0001**

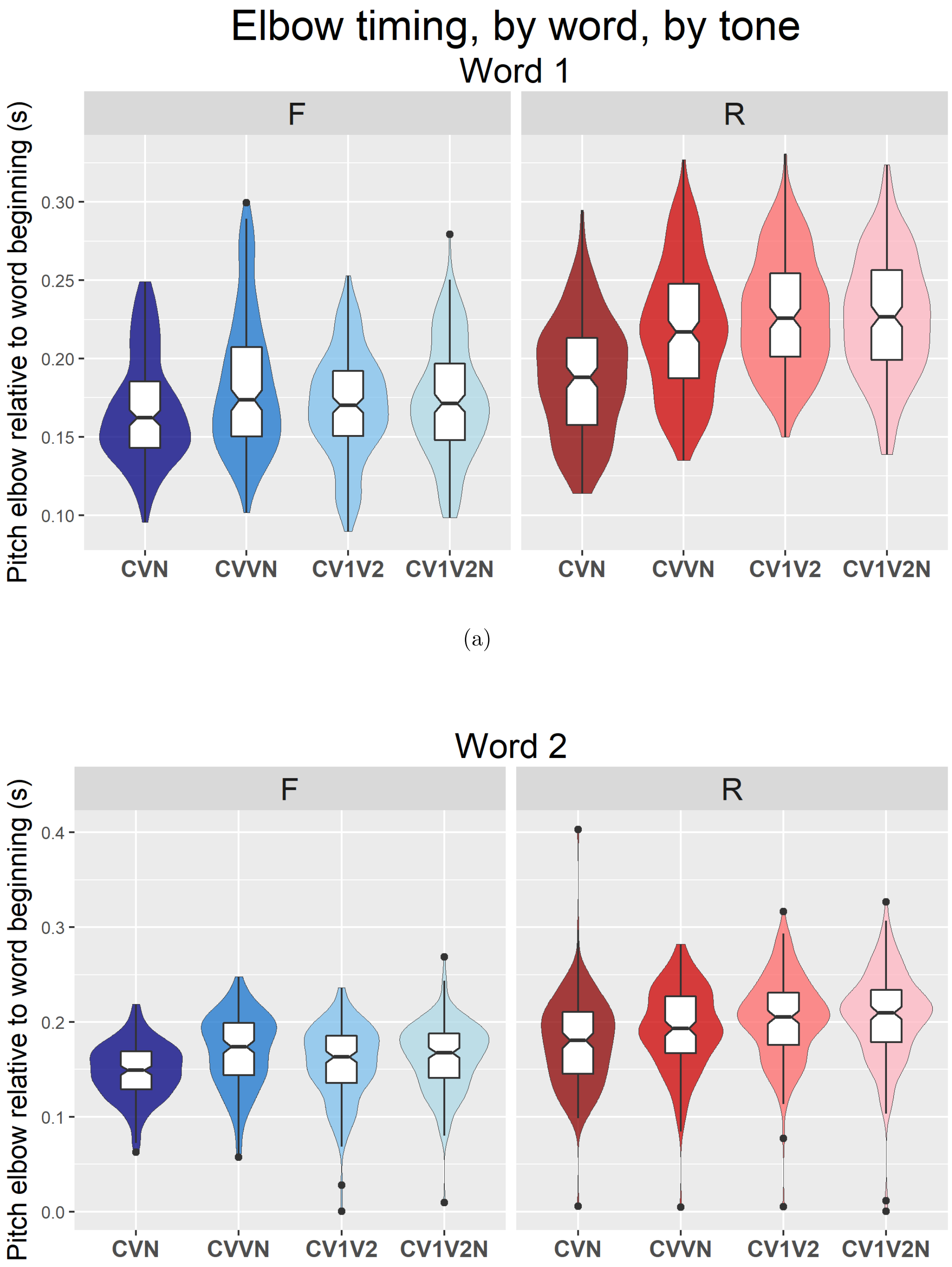
[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

word shapes (at $p < 0.01$) except between CV_1V_2 and CVVN ($p = 0.03$) and again, CV_1V_2 and CV_1V_2N ($p = 1.00$).

There is also an effect of `Shape` on `NormElbow` (the time-normalized equivalent of elbow timing—$\chi^2(3) = 74.58$, $p < 0.0001$ for Word 1; $\chi^2(3) = 110.85$, $p < 0.0001$ for Word 2; see Table 2.12). This indicates that there is not a set proportional target for tonal elbows. However, at least for Falling tones, the differences in means are quite small despite being statistically significantly different (for Word 1, the largest difference is 3.1%; for Word 2, 1.9%), which is suggestive of a more consistent "proportional" target for tonal elbows.

In contrast, there is quite a lot of variation in the Rising tone—in particular, for Word 1, all means are significantly different from each other ($p < 0.005$ for all comparisons using a Tukey HSD test). Although some of the differences in means are fairly small, such as the difference between CVN and CVVN (2.7%), others are fairly large, such as the difference between CV_1V_2 and CVN (9.5%; using mean duration of all Rising Word 1, on the order of

CV_1V_2N 4. CVVN; significant differences only exist between pairs that are two levels apart (i.e., 1/3, 1/4, 2/4).



(a)

(b)

Figure 2.10: Box plots comparing the elbow timing of Falling (blue) and Rising (red) tones for each word shape.

Table 2.12: Comparison of nested models for NormElbow for Word 1 and Word 2 (time normalized data).

(a) Word 1.

Model for NormElbow	χ^2	DegF	p[†]
Tone + (1|Part)	—	—	—
Tone + Shape + (1|Part)	74.58	3	< 0.0001**
Tone + Shape + Tone:Shape + (1|Part)	158.99	3	< 0.0001**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

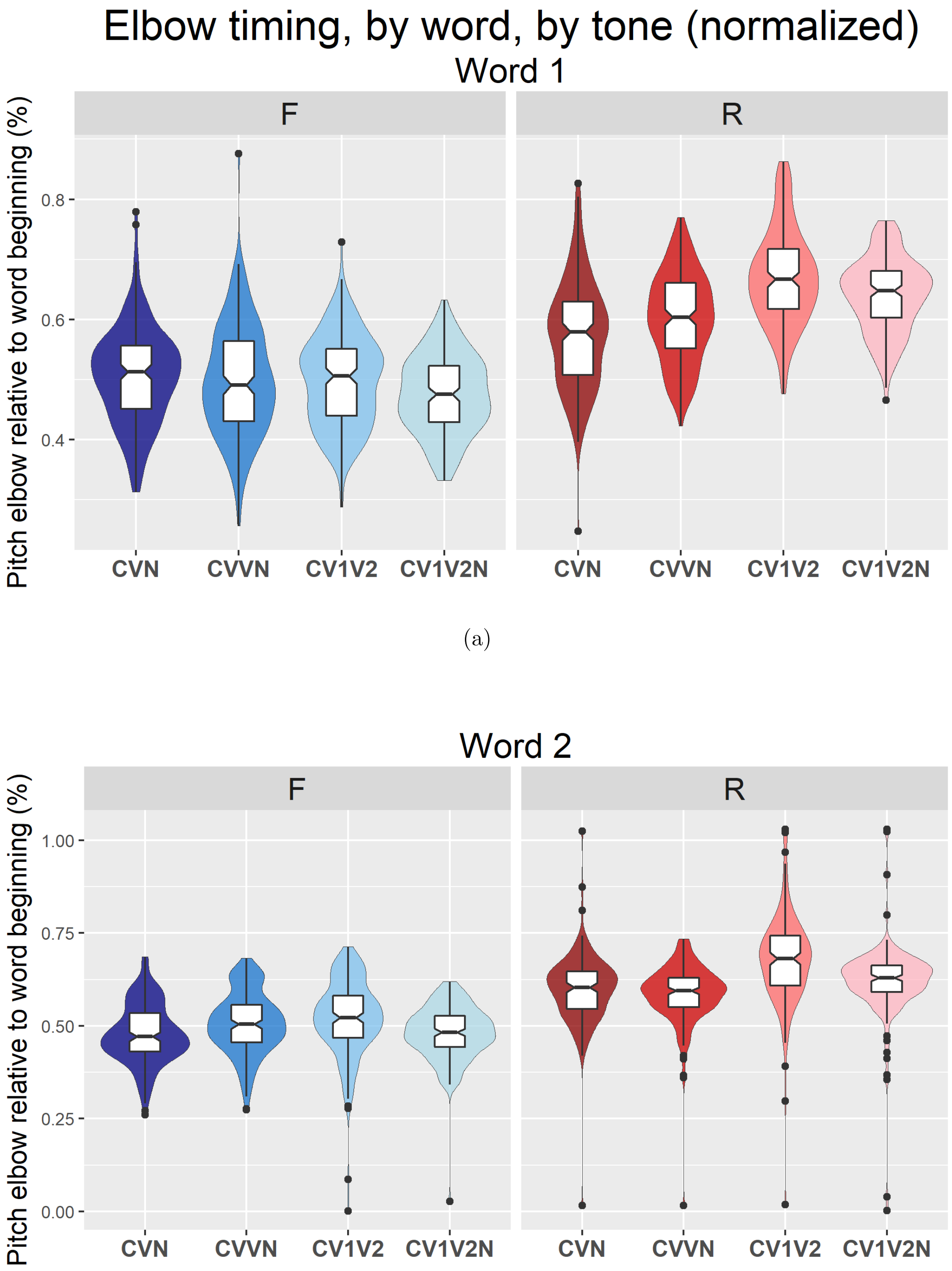
(b) Word 2.

Model for NormElbow	χ^2	DegF	p[†]
Tone + (1|Part)	—	—	—
Tone + Shape + (1|Part)	110.85	3	< 0.0001**
Tone + Shape + Tone:Shape + (1|Part)	39.14	3	< 0.0001**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

approximately 35 ms). There is slightly less variation in Word 2; only CV_1V_2 is significantly different from the rest of the word shapes ($p < 0.0001$ compared to CVN; $p = 0.03$ compared to CVVN; $p = 0.0004$ compared to CV_1V_2N), and the largest difference between means is 5.5% (using mean duration of all Rising Word 2, on the order of approximately 15 ms). These differences are illustrated in Figure 2.8. Thus, while for Falling tones, the null hypothesis is rejected in favor of Hypothesis 2.1, the results for the Rising tones satisfy none of the hypotheses.

Effect of first mora duration In the previous section, it was shown that Shape has a significant effect on Elbow for both tones. However, the effect is not straightforwardly based on the duration of the first mora: CV_1V_2N words have the shortest first mora, but they do not have the earliest elbows, indicating that the syllable is the active unit in tonal alignment. Again, CVVN words are excluded from this analysis due to the impossibility of marking the end of the first mora.

The addition of REMora to a model that already has Tone significantly improves the fit



(a)

(b)

Figure 2.11: Box plots comparing the elbow timing of Falling (blue) and Rising (red) tones for each word shape.

Table 2.13: Comparison of nested models for Elbow for Word 1 and Word 2 (millisecond data).

(a) Word 1.

Model for Elbow	χ^2	DegF	p[†]
Tone + (1|Part)	—	—	—
Tone + REMora + (1|Part)	247.44	1	$< 0.0001^{**}$
Tone + REMora + Tone:REMora + (1|Part)	15.95	1	$< 0.0001^{**}$
[†]As compared to model immediately above	$^{\circ} < 0.05$, $^{*} < 0.01$, $^{**} < 0.001$		

(b) Word 2.

Model for Elbow	χ^2	DegF	p[†]
Tone + (1|Part)	—	—	—
Tone + REMora + (1|Part)	196.08	1	$< 0.0001^{**}$
Tone + REMora + Tone:REMora + (1|Part)	6.07	1	0.01°
[†]As compared to model immediately above	$^{\circ} < 0.05$, $^{*} < 0.01$, $^{**} < 0.001$		

of the model both for Word 1 ($\chi^2(1) = 247.44$, $p < 0.0001$) and Word 2 ($\chi^2(1) = 196.08$, $p < 0.0001$; see Table 2.13). Increases in REMora correlate with later tonal elbows, though it is not a one-to-one relationship for either Word 1 ($\beta = 482.3$ ms, SE = 28.9 ms—that is, an increase of 1,000 ms in REMora corresponds to a 482.3 ms delay in elbow) or Word 2 ($\beta = 341.0$ ms, SE = 23.2 ms).

Like for the models with the categorical Shape predictor, the interaction between REMora and Tone is also significant for Word 1 ($\chi^2(1) = 15.95$, $p < 0.0001$) and marginally significant for Word 2 ($\chi^2(1) = 6.07$, $p = 0.01$). Increases in REMora have a smaller effect on Rising elbows than on Falling elbows in both Word 1 and Word 2. Estimates and standard errors are provided in Table 2.14.

The patterning is slightly different when considering time-normalized data, which eliminates speech rate variation and focuses on differences caused by word shape—in this case the interpretation is if the proportion of the word taken up by the first mora affects the proportion of the tone taken up by the first excursion. The addition of NormREMora to a

Table 2.14: Table of estimates and standard errors from the model `Elbow` $\sim$ `Tone + REMora + Tone:REMora + (1|Part)`. Intercept is Falling tone at `REMora` = 0 ms. All other estimates are with respect to the intercept. Units in ms.

(a) Word 1.

	β	Std. Error
Intercept	86.8	10.4
Rising, 0 ms	+10.5	9.4
Falling, 1,000 ms	+403.1	34.9
Rising, 1,000 ms	+183.3	45.7

(b) Word 2 (marginally significant interaction).

	β	Std. Error
Intercept	99.9	10.7
Rising, 0 ms	+19.5	8.1
Falling, 1,000 ms	+288.3	31.5
Rising, 1,000 ms	+98.0	39.7

Table 2.15: Comparison of nested models for `NormElbow` for Word 1 and Word 2 (time-normalized data).

(a) Word 1.

Model for `NormElbow`	χ^2	DegF	p[†]
`Tone + (1|Part)`	—	—	—
`Tone + NormREMora + (1|Part)`	62.93	1	< 0.0001**
`Tone + NormREMora + Tone:NormREMora + (1|Part)`	0.49	1	0.48
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

(b) Word 2.

Model for `NormElbow`	χ^2	DegF	p[†]
`Tone + (1|Part)`	—	—	—
`Tone + NormREMora + (1|Part)`	23.43	1	< 0.0001**
`Tone + NormREMora + Tone:NormREMora + (1|Part)`	11.35	1	0.0008**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

model that already includes `Tone` significantly improves the model for both Word 1 ($\chi^2(1)$ = 15.84, p < 0.0001) and Word 2 ($\chi^2(1)$ = 31.70, p < 0.0001; see Table 2.15). Overall, an increase in `NormREMora` corresponds with a delay in `Elbow`; as in the raw data, however, the effect is not one-to-one (β = 24.3%, SE = 3.0% for Word 1; β = 14.5%, SE = 3.0% for Word 2).

Table 2.16: Table of estimates and standard errors from the model `NormElbow` $\sim$ `Tone + NormREMora + Tone:NormREMora + (1|Part)`. Intercept is Falling tone at `REMora` = 0%. All other estimates are with respect to the intercept. Units in %.

(a) Word 1 (interaction NOT significant).

	β	Std. Error
Intercept	33.5	3.0
Rising, 0 ms	+16.4	3.5
Falling, 1,000 ms	+26.3	4.1
Rising, 1,000 ms	-4.2	5.9

(b) Word 2.

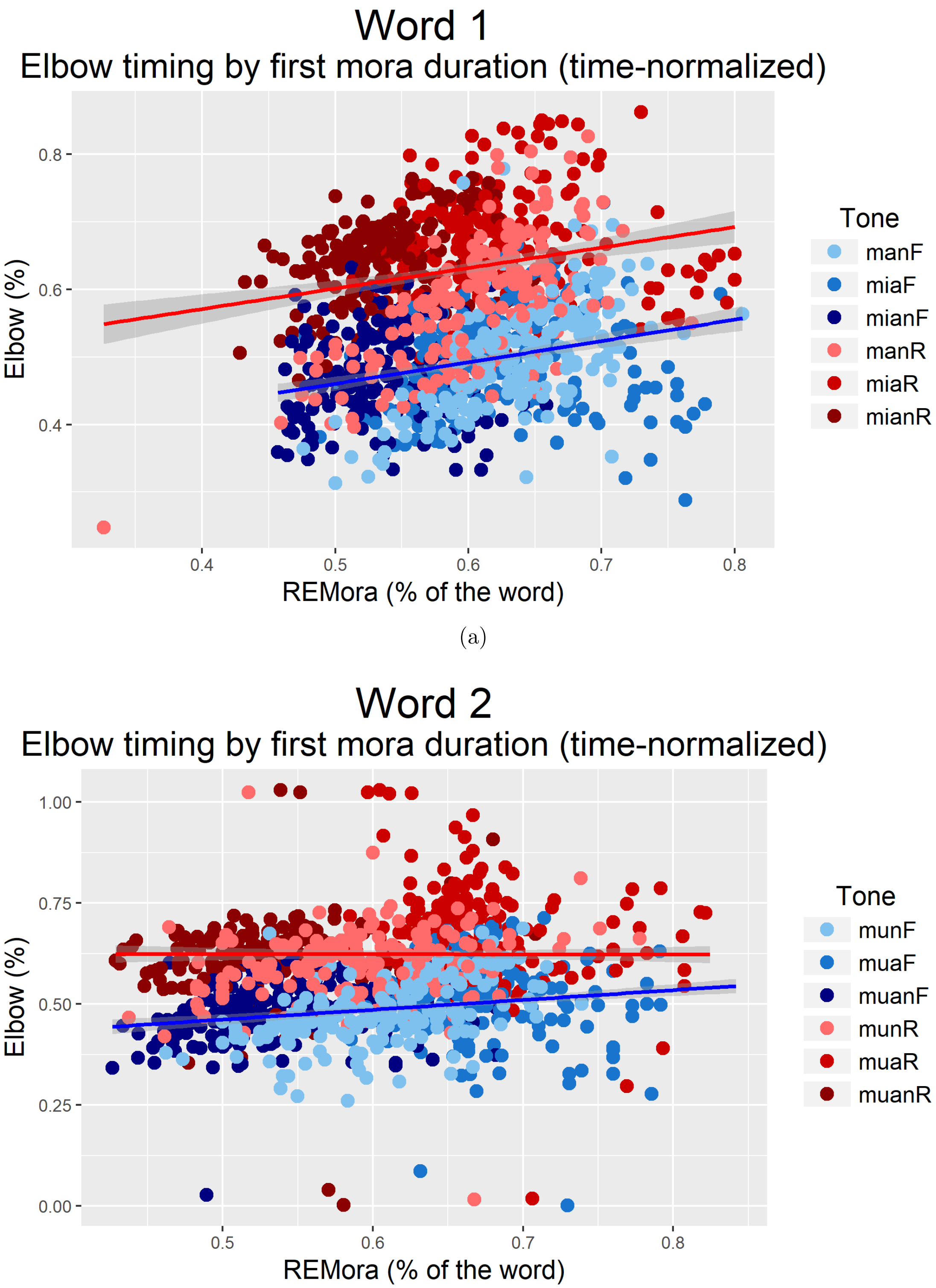
	β	Std. Error
Intercept	34.0	3.1
Rising, 0 ms	+25.3	3.7
Falling, 1,000 ms	+24.3	4.1
Rising, 1,000 ms	-19.5	5.8

The interaction term `Tone:NormREMora` does not significantly improve the model for Word 1 ($\chi^2(1) = 0.49$, p = 0.48), but does for Word 2 ($\chi^2(1) = 11.35$, p = 0.0008). As seen with the non-normalized data, increases in `NormREMora` have a smaller effect on Rising elbows than on Falling elbows. These patterns are illustrated in Figure 2.12.

Overall, Hypothesis 2.0 is rejected. The results are in favor of Hypothesis 2.1, where the syllable is the TBU for tone alignment, as `NormElbow` is fairly consistent across word shape, and does not parallel changes in `NormREMora`.

2.2.1.5 Excursion characteristics

The analysis of excursion size includes only the excursions of Word 1. This is due to the lack of clearly defined excursion onsets for half of the Word 2 trajectories—F+R and R+F sequences do not have an intervening extremum between the tones as F+F and R+R sequences do, but rather have one smooth pitch movement from elbow to elbow. The dataset used to analyze excursion size is also further pared down from the dataset used to analyze peak location, as some tokens had clearly defined peaks, but not clearly defined pitch onsets. A total of 1450 trials were included in this dataset (2.6% attrition from the peak timing dataset for Word 1): 356 F+F, 363 F+R, 361 R+R, and 370 R+F. The variables used in the following analyses are schematized in Figure 2.13.



(a)

(b)

Figure 2.12: Scatter plots (with fit lines by tone) showing the relationship between `NormREMora` and `NormElbow`. Falling tones are blue; Rising tones are red. Different target word shapes are indicated by different shades within colors.

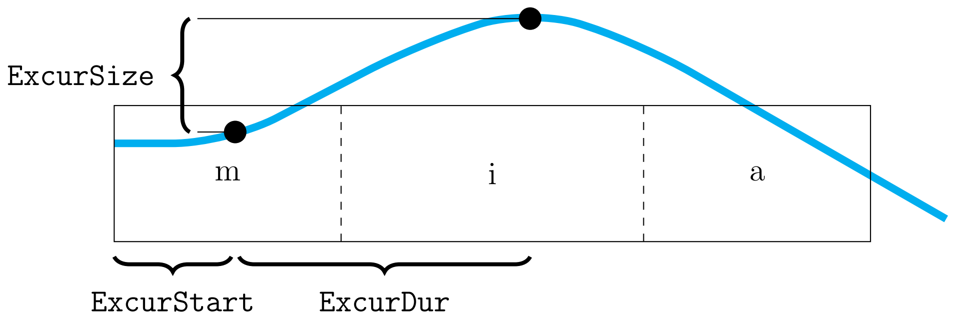


Figure 2.13: A schema of the dependent variables used for analysis of excursion characteristics, reproduced from Figure 2.3.

Excursion duration This section investigates the role of tone targets in tone timing, first by examining the duration of the initial first excursion. There are two potential effects: the effect of tone identity, and the effect of word shape (syllable duration).

> **Hypothesis 3.0** (null hypothesis): The first excursion is identical in duration for Falling and Rising tones.
>
> **Predictions 3.0**: `Tone` is not a significant predictor of `ExcurDur`.

> **Hypothesis 3.1**: In line with the timing differences posited in Hypothesis B, Rising tones have a longer initial pitch excursion.
>
> **Predictions 3.1**: `Tone` is a significant predictor of `ExcurDur`—specifically, Rising tones have a longer initial excursion.

> **Hypothesis 4.0** (null hypothesis): The first excursion is identical in duration across word shapes.
>
> **Predictions 4.0**: `Shape` is not a significant predictor of `ExcurDur`.

> **Hypothesis 4.1**: Excursion duration is determined in part by word shape.

Table 2.17

(a) Comparison of single-factor linear mixed effects models for ExcurDur, Word 1.

Model for ExcurDur	AIC	χ^2	DegF	p[†]
Tone + (1|Part)	-5690.8	22.89	1	< 0.0001**
Shape + (1|Part)	-5805.4	141.49	3	< 0.0001**
WordDur + (1|Part)	-5768.9	100.92	1	< 0.0001**
[†]As compared to the null model, ExcurDur ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of single-factor linear mixed effects models for NormExcurDur, Word 1.

Model for NormExcurDur	AIC	χ^2	DegF	p[†]
Tone + (1|Part)	-2520.4	37.12	1	< 0.0001**
Shape + (1|Part)	-2527.6	48.28	3	< 0.0001**
[†]As compared to the null model, ExcurDur ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

Predictions 4.1: Tone is a significant predictor of ExcurDur—specifically, longer words have longer initial excursions.

There is an effect of Tone on ExcurDur ($\chi^2(1) = 22.89$, $p < 0.0001$; see Table 2.17). That Tone has an effect indicates that tonal elbows are not later in Rising tones simply because the tone movement starts later, but rather that there are also durational differences; however, the difference between Falling and Rising tones is just 8.5 ms (SE = 1.8 ms), which does not indicate a meaningful (i.e., deliberate) difference in excursion duration. There is also quite a lot of individual variation: three speakers (CP01, TS01, MA01) exhibit very small differences between Falling and Rising excursion durations (in the ones digits of milliseconds), while two speakers (BM01 and SM01) exhibit medium-sized differences that are more likely to be meaningful—though in opposite directions from each other—, and the remaining speaker (RH01) shows a considerable difference between Falling and Rising excursions (see Table 2.18a). Tone also is a significant predictor of NormExcurDur ($\chi^2(1) = 37.12$, $p < 0.0001$; see Table 2.17b), where again Rising tones have longer excursions (3.2% longer; SE = 0.5%) than Falling tones. Thus, Hypothesis 3.0 is rejected.

There is also an effect of Shape on ExcurDur ($\chi^2(3) = 141.49$, $p < 0.0001$; see Table 2.17). The main difference is the word shape CVN, which has an excursion approximately 20 ms shorter than the other words ($p < 0.0001$ for all); the remaining word shapes have roughly equal excursion durations—all within 2 ms of each other ($p > 0.75$ for all). Similarly, there is also a significant effect of WordDur ($\chi^2(1) = 100.92$, $p < 0.0001$ compared to the null model), where increases in word duration are correlated with increases in excursion duration ($\beta = 237.7$ ms, SE = 22.8 ms). The AIC values indicate that the model with Shape alone provides a better fit than the model with WordDur alone. Thus, Hypothesis 4.0 is rejected.

Start of pitch excursion In this section, I examine the timing of the start of the initial pitch excursion. This encompasses two sets of analyses: the effect of tone identity, and the effect of word shape (syllable duration).

> **Hypothesis 5.0** (null hypothesis): The Falling and Rising tones use the same coordinative regime to coordinate the start of the tone gesture with the beginning of the word.
>
> **Prediction 5.0**:
>
> - There is no effect of Tone on ExcurStart.
> - Tone is a significant predictor of ExcurDur—specifically, Rising tones are longer.
>
> **Hypothesis 5.1**: Falling and Rising tones use different coordinative regimes to coordinate the start of the tone gesture with the beginning of the word.
>
> **Prediction 5.1**:
>
> - There is a significant effect of Tone on ExcurStart, where the first pitch excursion starts later for Rising tones than for Falling tones.
> - Tone is not a significant predictor of ExcurDur.

Hypothesis 6.0 (null hypothesis): All word shapes have the same coordinative relationship with the first tone gesture.

Prediction 6.0: There is no effect of `Shape` on `ExcurStart`.

Hypothesis 6.1: The timing of the beginning of the first tone gesture depends on the shape of the word.

Prediction 6.1: `Shape` will be a significant predictor of `ExcurStart`.

There is a significant effect of `Tone` on `ExcurStart` ($\chi^2(1) = 412.67$, $p < 0.0001$), where the initial excursion starts approximately 33.5 ms later (SE = 1.5 ms) for Rising tones than for Falling tones. Thus, Hypothesis 5.0 is rejected in favor of Hypothesis 5.1.

Table 2.18: Estimates for the Word 1 excursion characteristics of each participant for Falling and Rising tones (Rising given as difference from Falling).

(a) Excursion duration (units in ms).

Partic.	**Falling**	**Rising**
BM01	196.9	+19.1
CP01	185.3	+9.6
TS01	197.2	+9.6
MA01	197.3	+1.5
RH01	191.3	+31.1
SM01	245.2	-19.2

(b) Excursion size (units in Hz).

Partic.	**Falling**	**Rising**
BM01	37.4	+5.1
CP01	22.3	+1.3
TS01	21.0	+7.0
MA01	44.0	-13.6
RH01	53.4	+4.9
SM01	53.5	+23.5

However, there is no significant effect of `Shape`, either as a single fixed effect ($\chi^2(3) = 2.10$, $p = 0.55$) or when added as a second fixed effect alongside `Tone` ($\chi^2(3) = 1.97$, $p = 0.58$). Thus, Hypothesis 6.0 is rejected in favor of Hypothesis 6.1. This, alongside the results for `ExcurDur`, indicates that differences in peak timing between word shapes is due solely to differences in excursion duration, and pitch excursions start at the same time with respect to the beginning of the word.

Excursion size Finally, I investigate the size (in Hz) of the initial pitch excursions. As in the previous two analyses, I examine two effects: the effect of tone identity, and the effect of word shape (syllable duration).

> **Hypothesis 7.0** (null hypothesis): The first excursion is identical in magnitude for Falling and Rising tones.
>
> **Predictions 7.0**: `Tone` is not a significant predictor of `ExcurSize`.
>
> **Hypothesis 7.1**: In line with the timing differences posited in Hypothesis B, Rising tones have a more extreme initial pitch excursion.
>
> **Predictions 7.1**: `Tone` is a significant predictor of `ExcurSize`—specifically, Rising tones have a greater excursion size.
>
> **Hypothesis 8.0** (null hypothesis): Tone targets are specified in the representation of the tone gesture.
>
> **Prediction 8.0**: `ExcurDur` (as related to word shape) is not a significant predictor of `ExcurSize`—i.e., word shapes that cause longer initial excursions do not also cause larger pitch excursions.
>
> **Hypothesis 8.1**: Pitch excursions and F0 extrema are the phonetic result of a tone gesture (upward for Falling tones; downward for Rising tones) that continues until the next gesture is activated.
>
> **Prediction 8.1**: `ExcurDur` (as related to word shape) is a significant predictor of `ExcurSize`.

In contrast to the analyses of excursion duration and the timing of the start of the excursion, `ExcurSize` demonstrates an invariance that indicates that pitch targets have primacy in representation. `Tone` as a single fixed effect significantly improves the fit of the model

($\chi^2(1) = 71.82$, $p < 0.0001$ compared to the null model; see Table 2.19), where the initial excursion is 4.7 Hz larger (SE = 0.55 Hz) for Rising tones than for Falling tones.[8] Providing for random slopes for each participant significantly improves the model ($\chi^2(1) = 454.33$, $p < 0.0001$ compared to a model with just random intercepts). The estimates for each participant reveal a wide range of relationships between Falling and Rising excursion sizes; four of the six speakers have approximately equal excursion sizes for Falling and Rising tones (that is, single-digit Hz differences), but one speaker has a larger excursion for Falling tones, and the last speaker has a much larger excursions for Rising tones (see Table 2.18b). There is also a fairly wide range in excursion size between these six speakers: participants CP01 and TS01 have small overall mean excursions (22.9 and 25.0 Hz, respectively), while participants MA01 and BM01 have slightly larger excursions (37.1 and 39.9 Hz), and participants RH01 and SM01 have very large excursions (56.0 and 65.2 Hz). Overall, Hypothesis 7.0 is rejected, but the inconsistency of patterning between participants suggests that there is not a consistent effect.

However, there is not a significant effect of `REMora` ($\chi^2(1) = 0.56$, $p = 0.46$) `ExcurSize`. There is also no significant effect of `Shape`, either as a single factor ($\chi^2(3) = 2.73$, $p = 0.43$) or as a second fixed effect in a model that already accounts for `Tone` ($\chi^2(1) = 2.86$, $p = 0.41$); nor is there a significant effect of `WordDur`, either as a single factor ($\chi^2(1) = 0.09$, $p = 0.77$) or as a second fixed effect alongside `Tone` ($\chi^2(1) = 0.17$, $p = 0.68$). That is, although first mora duration, word shape, and word duration had a significant effect on the duration of the excursion, they do not directly have a significant effect on the size of the excursion. `ExcurDur` itself does significantly improve the model ($\chi^2(1) = 109.19$, $p < 0.0001$), where longer excursions are larger ($\beta = 84.81$ Hz, SE = 8.0 Hz; i.e., for every increase of 1,000 ms on excursion duration, the excursion is 84.81 Hz larger). Thus, the null hypothesis 8.0 is not rejected.

Two things are worthy of note at this juncture. The first is that speech rate (as approx-

[8] Note that excursion size is expressed in an absolute value, not positive for the initial upward trajectory of a Falling tone and negative for the initial downward trajectory of a Rising tone.

Table 2.19: Summary of linear mixed effects models for `ExcurSize`, Word 1.

(a) Comparison of single-factor linear mixed effects models for `ExcurSize`, Word 1.

Model for `ExcurSize`	AIC	χ^2	DegF	p[†]
`Tone + (1|Part)`	10,950	71.82	1	< 0.0001**
`Shape + (1|Part)`	11,023	2.73	3	0.43
`WordDur + (1|Part)`	11,022	0.09	1	0.77
`REMora + (1|Part)`	11,022	0.56	1	0.46
`ExcurDur + (1|Part)`	10,913	109.19	1	< 0.0001**

[†] As compared to the null model, `ExcurSize` $\sim$ 1 + `(1|Part)`

° < 0.05, * < 0.01, ** < 0.001

(b) Comparison of nested linear mixed effect models for `ExcurSize`, Word 1.

Model for `ExcurSize`	χ^2	DegF	p[†]
`Tone + (1|Part)`	—	—	—
`Tone + Shape + (1|Part)`	2.86	1	0.41
`Tone + (1|Part)`	—	—	—
`Tone + WordDur + (1|Part)`	0.17	1	0.68
`Tone + (1|Part)`	—	—	—
`Tone + WordDur + (1|Part)`	93.40	1	<0.0001**

[†] As compared to model immediately above

° < 0.05, * < 0.01, ** < 0.001

imated by `WordDur + Shape`, where including `Shape` addresses shape-related variation in word duration, leaving just trial-to-trial or participant-related variation within word shape) is not a strong predictor of excursion size ($\chi^2(4) = 2.93$, p = 0.57 compared to the null model). The range of combinations of excursion size and speech rate are demonstrated by the six participants: CP01, with small excursions (22.9 Hz) and a fast rate of speech (mean Word 1 duration 278.2 ms); SM01, with large excursions (65.2 Hz) and a slow rate of speech (403.2 ms); TS01, with small excursions (25.0 Hz) and a medium rate of speech (331 ms); and RH01, with large excursions (56.0 Hz) and a comparably medium rate of speech (350.3 ms). Although it may be true that deliberate changes in rate of speech within one person would affect the excursion size, each individual speaker appears to have a pitch excursion target that is fairly independent of their idiosyncratic rate of speech and that is not affected by word-to-word variation in duration.

2.2.2 Environment 2: Timing of tones, accounting for context

In this section, I will discuss the timing of tones while accounting for context: i.e., separating the timing of tones in Word 1 by the tone of Word 2, and vice versa. For analyses where only the timing of one word or the other is considered, the dataset with no errors for that word is used; for analyses that involve the timing of both words (e.g., timing of Word 1 elbow by the timing of Word 2 elbow), the dataset with no errors for either word is used. In this section, I will also refer to the variable `SameTone`, which describes the relationship between the tone of Word 1 and the tone of Word 2—i.e., if they are the same (`SameTone` = True, F+F and R+R sequences) or different (`SameTone` = False, F+R and R+F sequences).

2.2.2.1 Effects of tone sequence on elbow timing

Similarly to the investigations collapsed across context, there is a necessary condition to further investigate the effects of tone sequence:

> **Hypothesis C** (independent variable—tone sequence): Anticipatory dissimilation (as described by Gandour et al. 1992; Potisuk et al. 1997) affects both pitch height and elbow timing.

Since it has already been shown that `Tone` has an effect on the timing of the elbow, further testing will be compared to that model (i.e., `Elbow` $\sim$ `Tone + (1|Part)`) rather than to a fully null model. A comparison of linear mixed models (with random intercepts, but not random slopes) shows that the addition of `SameTone` significantly improves the model ($\chi^2(1) = 179.41$, $p < 0.0001$), where `SameTone` being false (i.e., F+R and R+F) delays the pitch elbow ($\beta = 19.5$ ms, SE = 1.4 ms). There is also a significant interaction between `SameTone` and `Tone` ($\chi^2(1) = 39.87$, $p < 0.0001$), where the effect is slightly greater if the first tone is Rising ($\beta = 10.6$ ms, SE = 2.0 ms for F+R compared to F+F; $\beta = 17.7$ ms, SE = 2.8 ms for R+F compared to R+R). The comparisons of raw elbow timing are presented in Figure 2.15, where each participant is shown in a separate panel.

Thus, when the two tones in the sequence are different, the elbow of the first tone occurs later than when the two tones are the same. That is, the first elbow in F+F sequences occurs earlier than the first elbow in F+R sequences, and the first elbow in R+R sequences occurs earlier than the first elbow in R+F sequences. Thus, Hypothesis C is upheld. This difference in timing is illustrated in Figure 2.14, where each trajectory has the first pitch elbow marked.

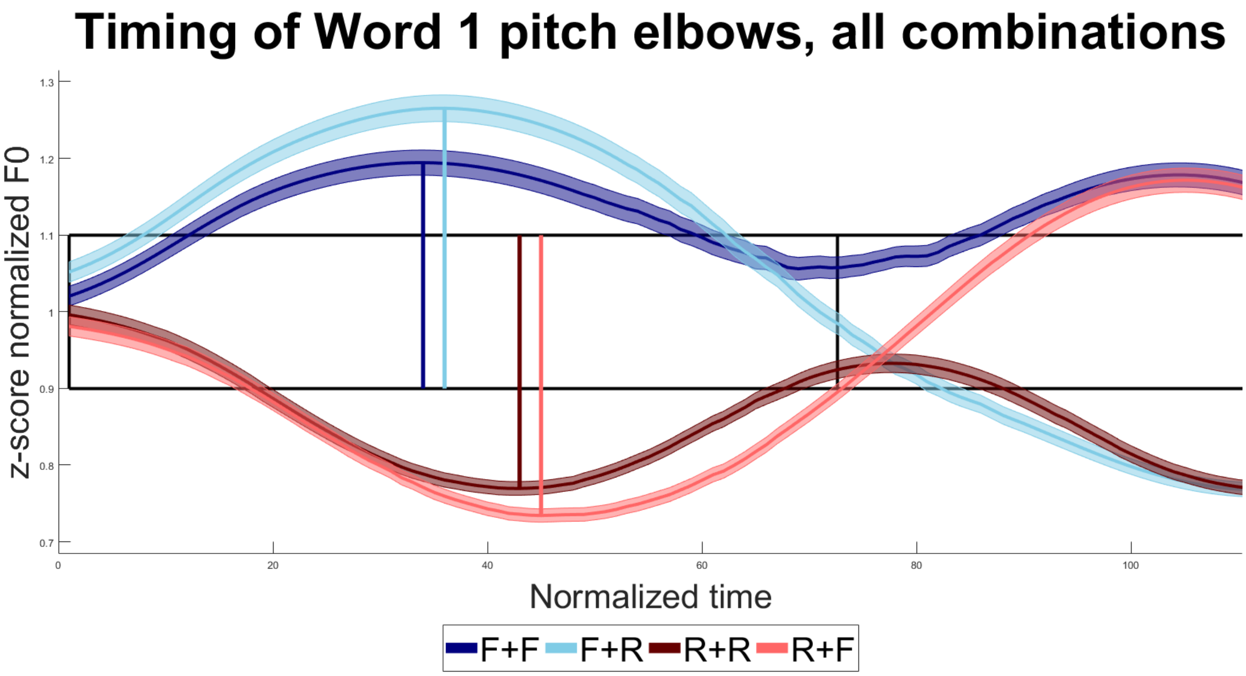


Figure 2.14: Z-scored and time-normalized F0 trajectories, including all participants. Dark blue is F+F, light blue is F+R, red is R+R, and pink is R+F. The rectangles indicate the timing of Word 1 and Word 2.

The same patterns hold for time-normalized data: `SameTone` significantly improves a model with just `Tone` ($\chi^2(1)$ = 143.35, p < 0.0001), and there is a significant interaction between `Tone` and `SameTone` ($\chi^2(1)$ = 16.55, p < 0.0001), and the difference again is quite small (β = 2.80 %, SE = 4.8 % for F+R compared to F+F; β = 2.76 %, SE = 6.8 % for R+F compared to R+R). Recall that for some speakers, the first word in R+F sequences was longer than the first word in R+R sequences—it is possible that the delayed elbow is obscured when considering the time normalized data, as the longer initial pitch excursion is then also being divided by a longer word. However, as the difference between delays for Falling vs. Rising tones is fairly small for the raw data as well, it is likely that this interaction is significant, but not meaningful.

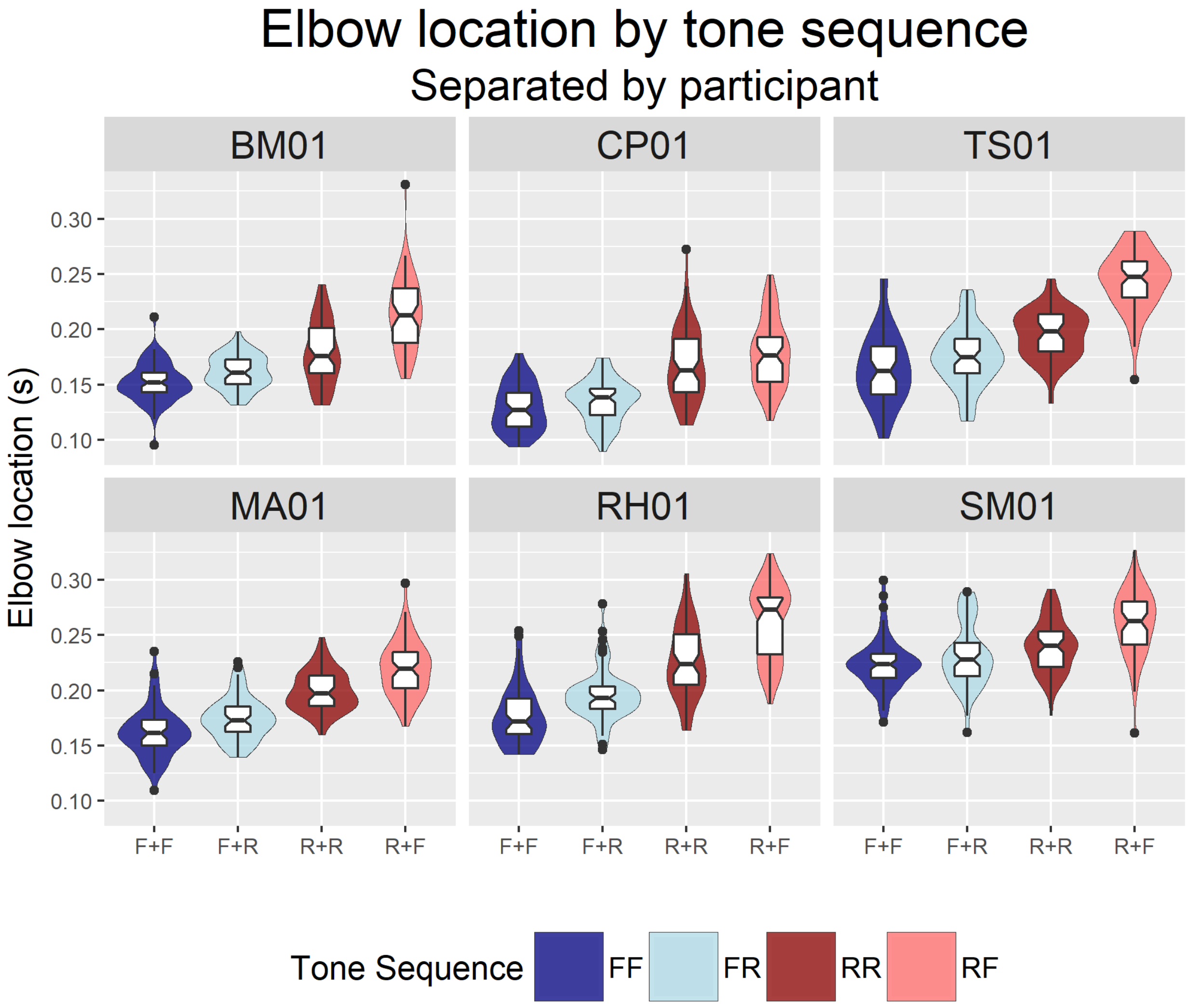


Figure 2.15: The timing of the first pitch elbow in each tone sequence, divided by participant.

One possible candidate for the effects of the tone of Word 2 on the realization of the tone on Word 1 is tonal crowding. However, this explanation does not explain the data well. In the case of F+R, tonal (de-)crowding could explain the rightward shift: the (phonetically) later elbow of the Rising tone allows the Falling tone to shift its own elbow later. In this case, the first elbow in F+F sequences occurs earlier because the following Falling tone has an earlier elbow, which puts pressure on the first elbow to occur earlier as well. However, by this logic, one would expect the first elbow in R+F sequences to be earlier than the first elbow in R+R sequences, as the following early peak in the Falling tone would put time pressure on the Rising tone's elbow and shift it earlier. As this is the opposite of the observed effect, tonal crowding does not offer a satisfactory explanation.

What is instead likely is that the timing changes due to anticipatory dissimilation. The contours in F+R and R+F sequences are both later and more extreme than F+F and R+R, with higher peaks and lower valleys. This is reminiscent of Xu's (1997) work that shows that the peak of falling tones are higher than the peak of level high tones in Mandarin, which he argues is due to dissimilation. The shift in timing is then likely due to the increase in pitch space traversed by the pitch excursion, rather than a deliberate change in timing.

2.2.2.2 Effects of tone sequence on word duration

For this section, there are two hypotheses that parallel those from the previous investigation of the effects of tone on segmental realization. In this case, however, the null hypothesis is that the already-demonstrated dissimilation is phonetic in nature:

> **Hypothesis 9.0** (null hypothesis): There is no effect of tone sequence on the realization of the segments of the first word.
>
> **Prediction 9.0**:
>
> - Whether the two words in the sequence have the same tone or not (that is, whether the sequence is F+F/R+R or F+R/R+F) will not be a significant predictor of the duration of the first word.
> - Whether the two words in the sequence have the same tone or not will not be a significant predictor of `REMora`.

Table 2.20: Comparison of nested linear mixed effects models for WordDur, Word 1.

(a) Comparison of nested models.

Model for WordDur	χ^2	DegF	p[†]
1 + (1|Part)	—	—	—
SameTone + (1|Part)	18.82	1	< 0.0001**
SameTone + Tone + (1|Part)	11.10	1	0.0009**
SameTone + Tone + SameTone:Tone + (1|Part)	13.78	1	0.0002**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

Hypothesis 9.1: Tone sequence does affect the realization of the first word.

Prediction 9.1:

- Whether the two words in the sequence have the same tone or not will be a significant predictor of the duration of the first word.
- Whether the two words in the sequence have the same tone or not will be a significant predictor of the duration of only the first mora in the first word.

When collapsing across participants, there is a significant effect of SameTone on the duration of Word 1 ($\chi^2(1) = 18.82$, p < 0.0001 compared to the null model). The difference once again is very small; the estimate for True (i.e., F+F and R+R sequences) is just 8.1 ms (SE = 1.9 ms) lower than for False (β = 353.5 ms, SE = 15.7 ms). The addition of Tone significantly improves the model ($\chi^2(1) = 11.10$, p = 0.0009), though again the difference is quite small—the Rising tone is 6.2 ms shorter than the Falling tone (SE = 1.9 ms). However, interaction between SameTone:Tone significantly improves the model ($\chi^2(1) = 13.78$, p = 0.0002), and here there is more separation of the duration of the first word in R+R sequences (β = -13.8 ms, SE = 3.7 ms compared to the intercept F+R).

When viewed altogether, it appears that the duration of the first word in F+F, F+R, and R+F sequences are all approximately equal, while R+R is the shortest (illustrated in Figure 2.16). This is a strange pattern if it is viewed as R+R being deliberately distinct, with the other three being the "default" duration. However, it is likely that this is not the case. As has been previously described, there are individual differences in the duration of Falling vs. Rising words—for some speakers, the words with Falling tones were longer, and for others, words with Rising tones were longer. Thus, it is necessary to consider separately the durational patterns of each Word 1 tone, as well as the durational patterns of each participant.

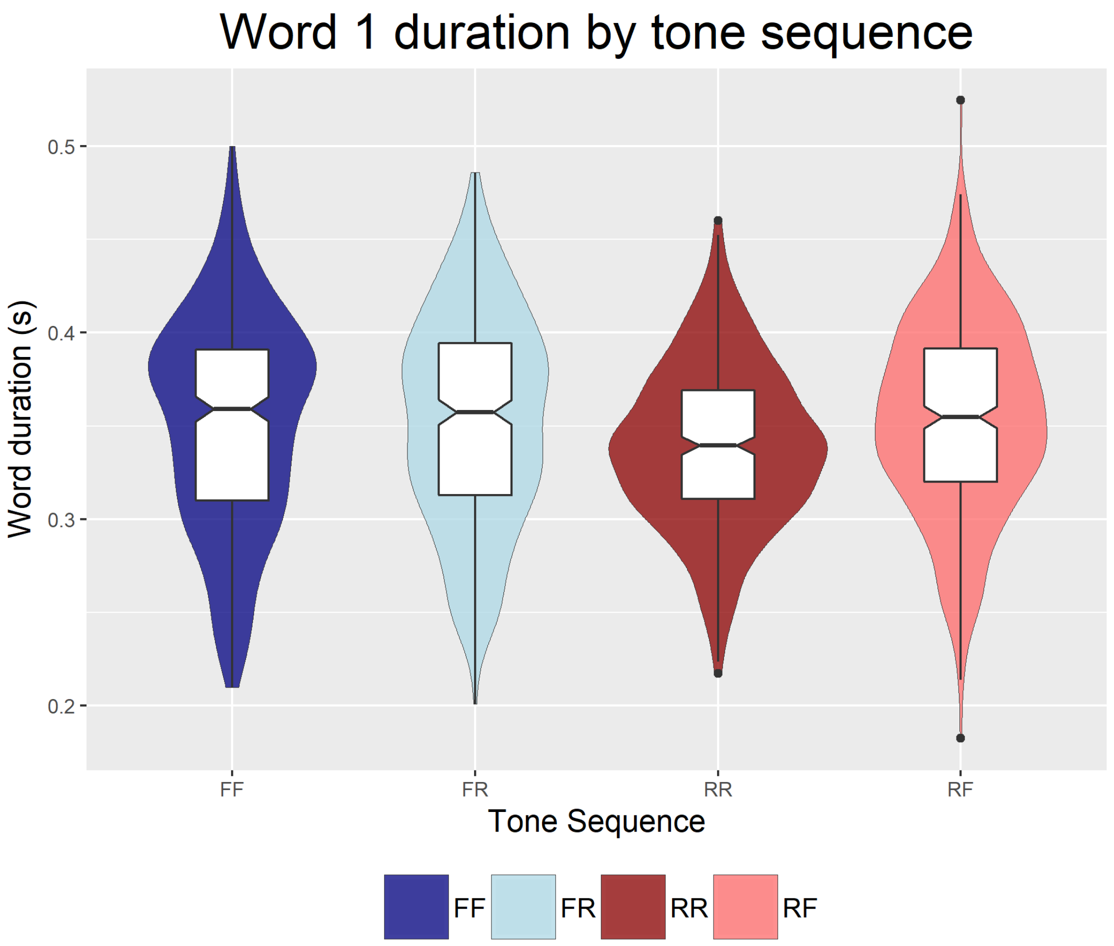


Figure 2.16: The durations of Word 1 for each tone sequence. Sequences with initial Falling tones are blue; sequences with initial Rising tones are red.

First, consider the patterns of F+F and F+R sequences (illustrated in Figure 2.17). As with the collapsed data, there is very little difference between F+F and F+R sequences;

participant SM01 appears to be verging on a statistically significant difference, but the separation is still small. A comparison of linear models confirms this impression: when considering just F+F and F+R sequences, SameTone does not improve the model ($\chi^2(1) = 0.27$, $p = 0.60$ compared to the null model). In contrast, there do appear to be significant (and meaningful) differences between R+R and R+F sequences (illustrated in Figure 2.18), though not for all participants: participants CP01, TS01, and MA01 appear to have no differences or possibly very small differences, while participants SM01 and RH01 in particular show fairly large differences. Again, a comparison of linear models confirms this impression: when considering just R+R and R+F sequences, SameTone significantly improves the model ($\chi^2(1) = 32.73$, $p < 0.0001$, R+F $\beta = +14.9$ ms, SE = 2.6 ms compared to R+R).

The addition of random slopes for participant (i.e., allowing for different magnitudes of effect of SameTone by participant) also significantly improves the model ($\chi^2(2) = 21.05$, $p < 0.0001$), and the participant-specific estimates also support the impressionistic conclusions (summary table provided in Table 2.21). Participant CP01 has a very small difference, and in fact the directionality is opposite to the other participants; participants TS01 and MA01 also show quite small differences, though they are both positive differences. In contrast, participants BM01, RH01, and SM01 all show differences that are more likely to be under some level of control, particularly in the last two cases, which approach a 30 ms difference.

Table 2.21: Estimates of the effect of SameTone in R+F vs. R+R sequences for each participant. Units in ms.

Partic.	R+R	R+F
BM01	345.7	+19.9
CP01	291.3	-7.0
TS01	323.2	+10.3
MA01	350.1	+10.6
RH01	345.6	+28.2
SM01	376.9	+27.9

Thus, with these differences in mind, and combined with the knowledge that for some

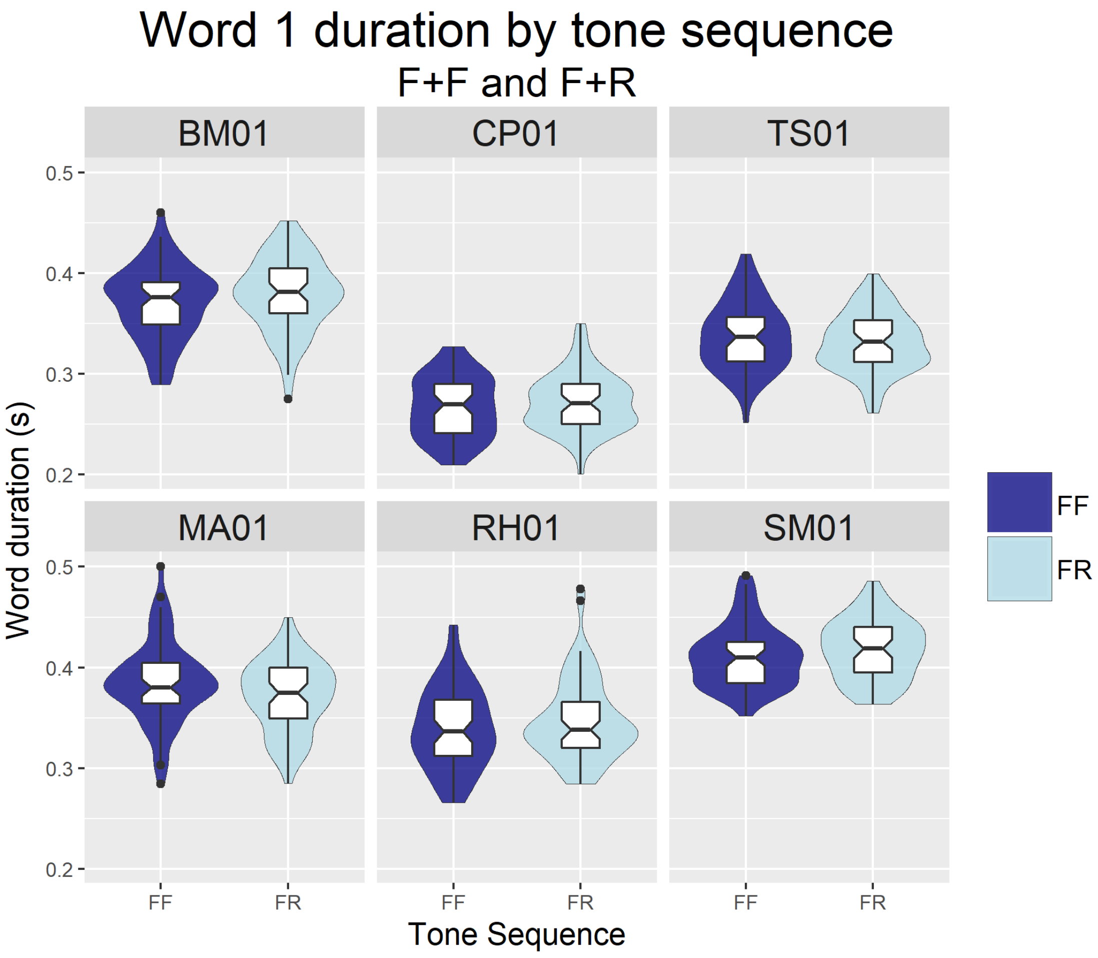


Figure 2.17: The durations of Word 1 for each tone sequence, F+F and F+R only, separated by participant.

speakers, words with Rising tones are overall shorter than words with Falling tones, one can conclude that R+R appears as shorter than the other three sequences likely due to the Rising baseline being lower, and R+F rising above that baseline in a lengthening process. This indicates that the null hypothesis 9.0 is not true. Further evidence for this process as lengthening of R+F rather than a shortening of R+R will be provided in the following section.

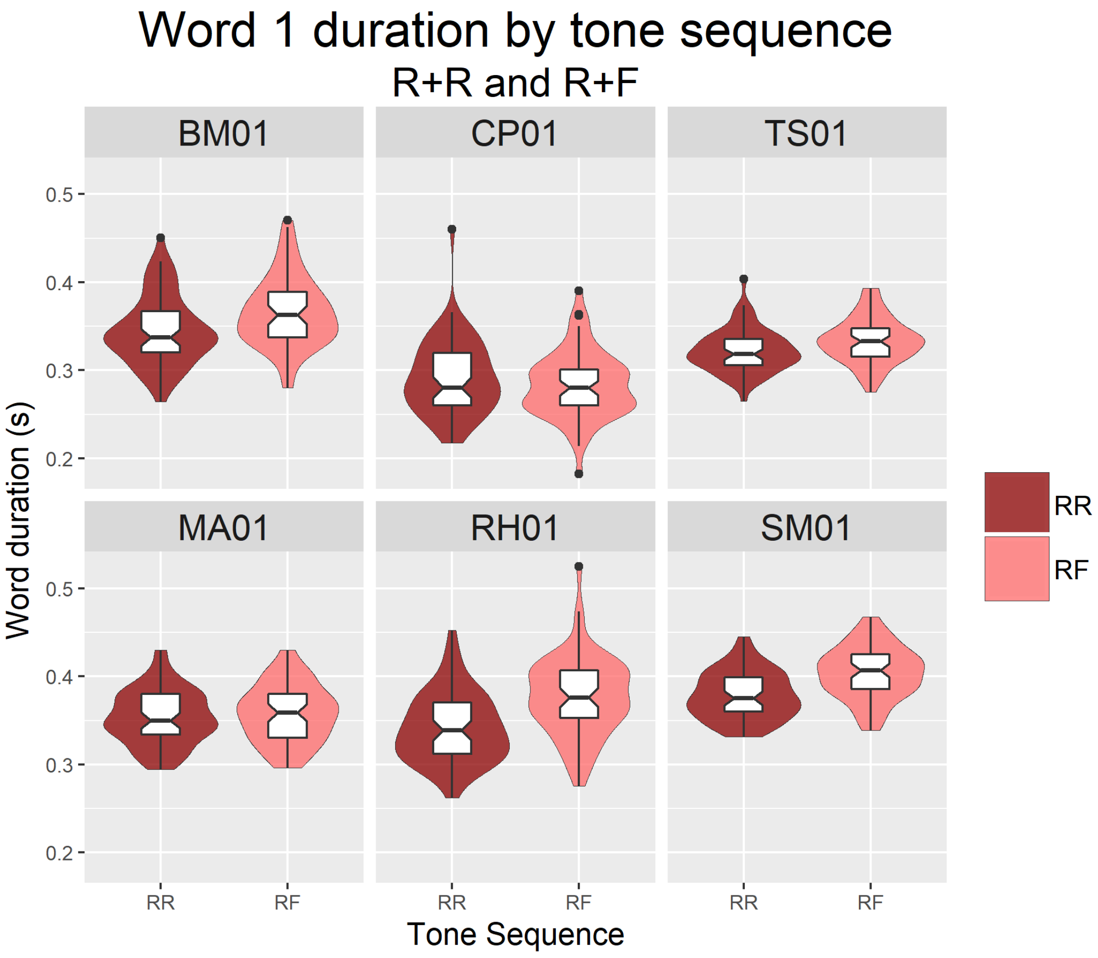


Figure 2.18: The durations of Word 1 for each tone sequence, R+R and R+F only, separated by participant.

2.2.2.3 Effect of tone sequence on first mora duration

All models for this comparison start from the model `REMora` $\sim$ `Shape + Tone + (1|Part)`, as it has been shown that `Shape` has a significant effect on `REMora` (e.g., the /mi/ in /mian/ is shorter than the /mi/ in /mia/), and also that `Tone` has a significant (but small) effect on `REMora`.

For speakers with a small pitch excursion, there is a significant but small (and in the opposite direction) effect of `Tone` on the timing of the right edge of the mora of Word 1 ($\chi^2(1)$

Table 2.22: Nested model comparisons for speakers with large pitch excursions and speakers with small pitch excursions.

(a) Speakers with small excursions (CP01, MA01, TS01).

Model for `REMora`	χ^2	DegF	p†
`Shape + (1|Part)`	—	—	—
`Shape + Tone + (1|Part)`	8.30	1	0.004*
`Shape + Tone + SameTone + (1|Part)`	0.25	1	0.62
`Shape + Tone + SameTone + Tone:SameTone + (1|Part)`	0.19	1	0.66
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

(b) Speakers with large excursions (BM01, RH01, SM01).

Model for `REMora`	χ^2	DegF	p†
`Shape + (1|Part)`	—	—	—
`Shape + Tone + (1|Part)`	10.09	1	0.001*
`Shape + Tone + SameTone + (1|Part)`	17.07	1	< 0.0001**
`Shape + Tone + SameTone + Tone:SameTone + (1|Part)`	11.21	1	0.0008**
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

= 8.30, p = 0.004; see Table 2.22a). There is no effect of `SameTone` on the duration of the first mora ($\chi^2(1) = 0.25$, p = 0.62)—i.e., the dissimilation-caused shift in tonal elbow timing is not accompanied by a shift in timing of the first TBU. This is as predicted in Hypothesis 69.0, given both the phonetic nature of the elbow shift; in addition, segments have not been shown to be reliably affected by elbow timing in previous analyses. The interaction between `Tone` and `SameTone` (i.e., comparing F+R to R+F) is also not significant ($\chi^2(1) = 0.19$, p = 0.66), indicating that there is truly no difference in the timing of the right edge of the mora, no matter the tonal sequence.

For speakers with a large pitch excursion, there is also a significant but small effect of `Tone` on the timing of the right edge of the first mora of Word 1 ($\chi^2(1) = 10.09$, p = 0.001; see Table 2.22b). However, in contrast with the small pitch excursion group, these speakers show an effect of `SameTone` on the duration of the first mora of Word 1 ($\chi^2(1) = 17.07$, $p < 0.0001$), where the right edge of the first mora occurs later in sequences with two different tones. Furthermore, there is a significant interaction between `Tone` and `SameTone` ($\chi^2(1) = 11.21$, $p = 0.0008$): there is a greater difference between the timing of the mora in R+R and R+F sequences (approximately 24 ms) than between F+F and F+R sequences (1.5 ms).

However, this difference does not appear to be a shift in *proportion* of mora 1 to mora 2, which would be the predicted patterning if the shift in elbow timing were causing a parallel shift in TBU duration. Rather, Word 1 in its entirety "stretches" in R+F sequences: the effect of `SameTone` on the time-normalized measure of `REMora` (i.e., the proportion of the word occupied by the first mora) is marginally significant ($\chi^2(1) = 6.31$, $p = 0.01$), but quite small and in the wrong direction (less than 1% difference, and earlier for sequences with two different tones, rather than later). The interaction `Tone:SameTone` also does not improve the model ($\chi^2(1) = 1.65$, $p = 0.20$), indicating that the larger difference between R+F and R+R sequences in raw terms is simply due the difference in overall duration of the word, not a particular shift of the right edge of the first mora. Thus, Hypothesis 9.1 is not supported. Furthermore, this result suggests that tone gestures can influence segmental timing when necessary.

2.2.2.4 Effects of tone sequence on excursion size

Initial pitch excursions in F+R and R+F sequences are also larger than those of F+F and R+R sequences. Again, as previously described, `Tone` has an effect on the size of the initial pitch excursion (where Falling tones generally have a smaller pitch excursion than Rising tones), thus further testing will be compared to a model that includes `Tone` (i.e., `ExcSize` $\sim$ `Tone + (1|Part)`) rather than the fully null model.

The addition of `SameTone` significantly improves the model ($\chi^2(1) = 89.63$, $p < 0.0001$; see Table 2.23), though the effect is fairly small; there is an increase of only 5.1 Hz when the tones are not the same (SE = 0.53 Hz). The inclusion of random slopes for `SameTone` additionally improves the model ($\chi^2(2) = 20.70$, $p < 0.0001$), but for all participants the

Table 2.23: Comparison of nested linear mixed effects models for ExcurSize, Word 1.

Model for ExcurSize	χ^2	DegF	p[†]
Tone + (1|Part)	—	—	—
Tone + SameTone + (1|Part)	89.63	1	< 0.0001**
Tone + SameTone + (1+SameTone|Part)	20.70	2	< 0.0001**
Tone + SameTone + ExcurDur + (1+SameTone|Part)	58.89	1	< 0.0001**
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

effect of having different tones in Word 1 and Word 2 is small and positive (coefficients range from 1.6 Hz difference to 8.9 Hz difference). Finally, ExcurDur still significantly improves the fit of the model for ExcurSize ($\chi^2(1) = 58.89$, $p < 0.0001$) even when possible confounding variables are already accounted for—i.e., Tone and SameTone.[9] Thus, it is likely that this effect is purely phonetic, which follows the first prediction of Hypothesis 9.0.

2.3 Conclusions

2.3.1 Summary: licensing vs. alignment

Although it is clear that the mora is used by Thai for distributional purposes, the results from this experiment indicate that the syllable, rather than the mora, is the TBU for alignment in Thai. Tonal elbows target proportions of the syllable, which do not correspond to moraic edges; nor do they track changes in mora duration that are caused by either intrinsic phonetic differences (e.g. [a] vs. [i]) or structural differences (e.g. mora sharing in /mian/ vs. /mia/). As shown in the analysis, the syllable as a TBU provides information to the tone gesture—specifically, the duration of its excursions. Importantly, however, this is not a one-way relationship: when under duress, tones can also force segments to accommodate them. This indicates that tone-segment timing is the result of multiple negotiations between segmental and tonal gestures.

[9] WordDur still does not, $\chi^2(1) = 0.39$, $p = 0.53$.

First, the null hypothesis 1.0 is upheld: despite differences in the timing of tonal extrema between Falling and Rising tones, there is no significant effect on the realization of the segments. This is particularly striking, because the directionality of difference was opposite to what would be predicted by phonetic effects, which suggests that the separation is encoded in the phonology.

> ✓**Hypothesis 1.0** (null hypothesis): There is no effect of tone identity on segmental realization.
>
> ✗**Hypothesis 1.1**: Tone plays a significant role in the timing of the first mora.

Furthermore, differences in first mora timing did not correspond to differences in tone realization (within tone identity). While `REMora` did have an effect on `Elbow`, the effects were not consistent: for both tones, differences in `REMora` did not effect changes of a similar magnitude in `Elbow`. Furthermore, there was an interaction between `Tone` and `REMora`, where the elbows of Rising tones were affected even less than the elbows of Falling tones. In contrast, `NormElbow` was fairly consistent across word shape, though again Rising tones did behave somewhat differently.

> ✗**Hypothesis 2.0** (null hypothesis): There is no effect of segment duration on tone realization.
>
> ✓**Hypothesis 2.1**: Syllables are a main driving force in Thai tone timing.
>
> ✗**Hypothesis 2.2**: Moras are a main driving force in Thai tone timing.

I then performed a set of analyses that examined the effects of tone identity and word shape on characteristics of the initial tone gesture. Both tone identity and word shape are significant predictors of the duration of the first pitch excursion, rejecting the null hypotheses 3.0 and 4.0.

> ✗**Hypothesis 3.0** (null hypothesis): The first excursion is identical in duration for Falling and Rising tones.

✓**Hypothesis 3.1**: In line with the timing differences posited in Hypothesis B, Rising tones have a longer initial pitch excursion.

✗**Hypothesis 4.0** (null hypothesis): The first excursion is identical in duration across word shapes.

✓**Hypothesis 4.1**: Excursion duration is determined in part by word shape.

There was a difference between Rising and Falling tones in the timing of the start of the first pitch excursion, which rejects Hypothesis 5.0. However, the magnitude of the difference was smaller (roughly half) than the magnitude of the difference between extremum timing, indicating that initial coordination differences are not the sole source of differences in elbow timing. In addition, timing of the start of the tone gesture relative to the beginning of the word was consistent across word shapes. Thus, differences in elbow timing between tone shape are not due to coordinative differences at the onset of the syllable. This suggests that although the coordinative regime may differ between tones (perhaps specifically between L and H tone gestures), there is a constancy across all words.

✗**Hypothesis 5.0** (null hypothesis): The Falling and Rising tones use the same coordinative regime to coordinate the start of the tone gesture with the beginning of the word.

✓**Hypothesis 5.1**: Falling and Rising tones use different coordinative regimes to coordinate the start of the tone gesture with the beginning of the word.

✓**Hypothesis 6.0** (null hypothesis): All word shapes have the same coordinative relationship with the first tone gesture.

✗**Hypothesis 6.1**: The timing of the beginning of the first tone gesture depends on the shape of the word.

Interestingly, although excursion duration directly influences the size of the excursion, no factors other than tone identity that were shown to affect `ExcurDur` significantly improved the models for `ExcurSize`. That is, although increases in word duration correlated with increases in the duration of the first excursion, there was not a corresponding increase in excursion size. There were some differences between the Falling and Rising tones, but different participants exhibited different directionalities despite all having early Falling peaks and late Rising valleys. Thus, the null hypothesis is not rejected for either Hypothesis 7 or 8. This suggests that the pitch target is included in the representation and has some sort of primacy when computing the precise trajectory of the first pitch excursion.

> ✓**Hypothesis 7.0** (null hypothesis): The first excursion is identical in magnitude for Falling and Rising tones.
>
> ✗**Hypothesis 7.1**: In line with the timing differences posited in Hypothesis B, Rising tones have a more extreme initial pitch excursion.

> ✓**Hypothesis 8.0** (null hypothesis): Tone targets are specified in the representation of the tone gesture.
>
> ✗**Hypothesis 8.1**: Pitch excursions and F0 extrema are the phonetic result of a tone gesture (upward for Falling tones; downward for Rising tones) that continues until the next gesture is activated.

Finally, Hypothesis 9.0 was rejected, as there were some effects of tone sequence on the realization of word 1—specifically, in R+F sequences, the first word was elongated compared to R+R sequences. Notably, this did not affect only the first mora (despite the delayed elbow), or just the second mora (which might indicate a smaller window of tone planning) but rather was a uniform stretching of the word. In addition, this difference was only exhibited by speakers with larger pitch excursions, which indicates that they are stretching the first word to accommodate the extra-late extremum in R+F sequences. This suggests that tone can play a role in timing segments, even when it is a purely phonetic dissimilatory effect.

✗**Hypothesis 9.0** (null hypothesis): There is no effect of tone sequence on the realization of the segments of the first word.

✓**Hypothesis 9.1**: Tone sequence does affect the realization of the first word.

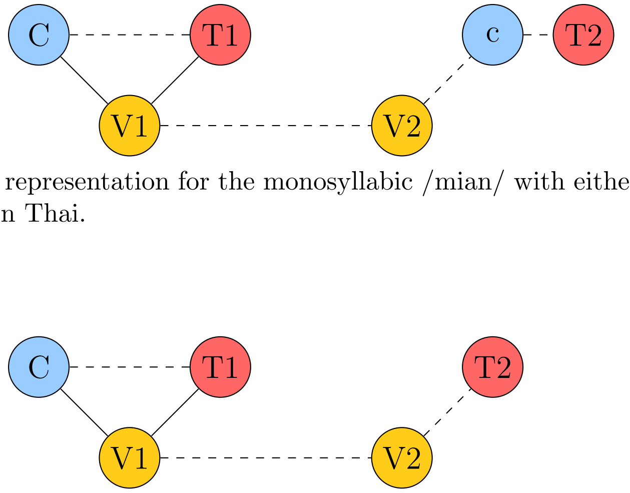
This effect is particularly interesting because it indicates the presence of a boundary between tone domains despite the continuous trajectory of the F0 contour. That is, in a LH+HL contour, there is a continuous rise from the valley of the LH contour to the peak of the HL contour. One could thus consider this to be one simple (merged) tone gesture/movement, where the H of the LH contour does not have a particular target it needs to reach until the next word. However, the stretching of the word indicates that the H gesture needs to fit into its syllabic segmental domain.

2.3.2 Gestural representation

This data indicates that the representation of tone gestures includes coordinative information, as well as information about the target, but gains specific duration information from the other gestures it is coordinated with. Although an additional articulatory study would be needed to verify the precise coordinative regime used for Thai tone,[10] the consistency in the lag between the start of the pitch excursion and the start of the word for all word shapes indicates that there is a consistent coordinative pattern for tone.

Thus, I propose the structure in Figure 2.19 for the representation of tone in Thai. This coordinative diagram shows the onset consonant (C), first vowel (V1), and the first tone gesture (T1) forming a c-center structure. The second vowel (V2) is anti-phase coordinated with V1 to form a syllable, but is not in-phase coordinated with the second tone gesture (T2). Instead, T2 is anti-phase coordinated with the last element in the syllable, be it V2 or a coda consonant (c). This provides both for the reference to syllable duration, and also generates the timing found in Karlin (2014).

[10] The lag in milliseconds between the start of the pitch excursion and the acoustic beginning of the syllable onset /m/ found in this study is similar to the lag reported by Karlin (2014), which suggests that this speaker too has a c-center relationship between tone and syllable onsets.



(a) Proposed representation for the monosyllabic /mian/ with either a Rising or Falling tone in Thai.

(b) Proposed representation for the monosyllabic /mia/ with either a Rising or Falling tone in Thai.

Figure 2.19: Model of tonal representation in Thai. T1 is coordinated in a c-center structure with the syllable onset and the first vowel; T2 is anti-phase coordinated to the last item in the second mora.

2.3.3 Discussion

There still remains the question of the differences in timing between Falling and Rising tones. The root cause of this variation is unclear. One possibility is that there is pressure from other contrasts in the tonal system, even if the link is not specifically between Falling and Rising tones. It would be worth considering the effects of the presence other tones in the system where timing is one of the main distinctions: the Falling and High tones both (historically) have a falling contour, where the High tone peak occurs later than the Falling tone peak (Pittayaporn to appear); Rising and Low tones have a similar contrast, except in this case the Low tone falls very rapidly and can get low very quickly (i.e., it is not a mirror of the high tone, which stays in the mid range for a while before rising towards the end). Thus, there would be pressure on the Falling tone to have an earlier peak (to maximize the contrast with the later High peak), and pressure on the Rising tone to have a later valley (to maximize the contrast with the rapid fall of the low tone). In this case, a syllabic window (rather than a strict moraic one) allows for greater separation of tonal contrast; contour tone timing can use the entire syllable to maximize separation rather than being constrained to the mora. Furthermore, the syllable itself is the unit on which contrast is allowed: individual tone levels may be licensed by the mora, but lexical tones are assigned to the syllable.

Finally, although the data shows that tones in Thai are aligned to the syllable, rather than to individual moraic TBUs, the dissimilation in F+R and R+F sequences indicates that the two pitch levels are still individually specified—i.e., tone in Thai is not stored as a single target contour (as suggested by Abramson (1979)). Similar dissimilation has been discussed within single tones in other languages, such as Mandarin (Xu & Wang 2001): the high point of a falling tone is higher than the high point of a high tone. In Thai, one might expect the same effect in Rising vs. Low tones (i.e., due to anticipatory dissimilation, the lowest point of an LH should be lower than the lowest point of an L). However, it is not clear that this is true; the peak of the Falling tone has been reported to be higher than the peak of the High tone in Thai, but the valley of the Rising tone has not been reported to be lower than the low point of the Low tone (Abramson 1962; Morén & Zsiga 2006; Zsiga & Nitisaroj 2007). For Falling tones, it is interesting that there is additional dissimilation when followed by a Rising tone. If the shifted extrema in R+F and F+R sequences is due to dissimilation, it is somehow an extension of the typical dissimilation already inherent in complex tones. That is, the tonal elbows are important in some way that makes Falling tones importantly high and Rising tones importantly low, even though they both have both tonal specifications.

Chapter 3

Serbian: Focus on falling accents

Serbian (srp) (previously known as Serbo-Croatian) is a South Slavic language that makes up part of the Bosnian-Croatian-Serbian (BCS) continuum. In this chapter I compare two dialects of Neo-Štokavian Serbian: the variety spoken in Belgrade (the capital city of Serbia), and the variety spoken in Valjevo (some 50 miles to the southwest of Belgrade). As a so-called "pitch accent" language, Serbian adds to the discussion of the representation of timing in tone languages in that only one syllable per word is phonologically specified for pitch.

Some of these differences have been previously described by Zec and Zsiga (2016). The main difference of interest to this study between the Belgrade and Valjevo dialects is the timing of pitch peaks, as was described in Chapter 1. For all accents, the Valjevo dialect has consistently earlier peaks than the Belgrade dialect. In the Belgrade dialect the F0 peak of a short falling accent occurs near the syllable boundary (i.e., near the end of the first mora), while for Valjevo the peak occurs near the beginning of the stressed vowel (see Figure 1.11a). Similarly, while the peak of short rising accents occurs in the post-tonic syllable in Belgrade Serbian, the corresponding peak in Valjevo Serbian occurs at or immediately after the boundary between the tonic and post-tonic syllables. The difference in timing is large enough that Valjevo rising accents in initial position are confusable with Belgrade falling accents.

In the current study, I expand on previous work done on dialectal differences in by considering the timing of the onset of H tone gesture, focusing on which aspects of the pitch excursion effect the major timing differences between the Belgrade and Valjevo dialects of Serbian. There are two main possibilities: first, the pitch excursions for both dialects start at the same time, but the Valjevo dialect specifies an excursion with a shorter duration; or second, the Valjevo pitch excursion initiates earlier than the Belgrade dialect, and the two dialects have excursions with the same duration. This targets the Articulatory Phonology arguments on the nature of tone gestures. It has been suggested that languages with lexical tone treat tone gestures as if they were the second gesture in a consonant cluster (the so-called c-center hypothesis for tone; Gao 2008; Karlin 2014; Yi 2014); however, this has not yet been studied in non-Asian tone languages. Second, the c-center hypothesis does not predict the differences in timing that have been documented between the Belgrade and Valjevo dialects, particularly when combined with the proposal that all gestures have an intrinsic duration (Browman & Goldstein 1990).

This study thus builds on previous work on Serbian and tonal alignment in general by systematically varying the syllable onsets of H syllables. I use /r, l, m, mr, ml/ as syllable onsets in order to probe both phonetic and phonological aspects of syllable onsets. Although there have been studies on the effect of the presence vs. absence of syllable onset on the timing of pitch landmarks (Ladd & Schepman 2003), there has not yet been work on the effect of simple vs. complex onsets. This combination of phonetic and phonological manipulations serves both to probe the acoustic anchoring hypothesis, which holds that pitch landmarks are aligned to acoustic segment edges, as well as to tease apart the predictions from various gestural anchoring proposals, which as a group hold that pitch should be treated as a gesture that is coordinated with segmental gestures.

3.1 Experiment design

3.1.1 Hypotheses

I present here the hypotheses in general terms; specific predictions referencing the variables included in the statistical analyses will be presented before each analysis. The broad purpose of this experiment is to examine the validity of the segmental anchoring hypothesis vs. the c-center hypothesis for tone, as manifested in the Belgrade and Valjevo dialects of Serbian. I am specifically focusing on the effects of syllable onset on tone timing. In order for there to be any meaningful investigation, however, varying the syllable onset has to have phonetic effects as well as phonological categories.

> **Hypothesis A**: Different syllable onsets have different durations.

Previous literature indicates that Belgrade and Valjevo Serbian have the same system of accentual contrast; this will be verified with this data in order to assess the different hypotheses.

> **Hypothesis B**: The Belgrade and Valjevo dialects of Serbian are two phonetic realizations of the same underlying accentual system (distribution and association), and not two distinct systems.

Previous literature has also indicated that the location of accent in the word preceding a target word influences the timing of the H gestures, which I will examine in this study as well.

> **Hypothesis 1.0** (null hypothesis): There is no effect of carrier verb on the timing of the accentual peak.
>
> **Hypothesis 1.1**: There is a significant effect of carrier verb on the timing of the

accentual peak.

Hypothesis 2.0 (null hypothesis): There is no effect of carrier verb on the timing of the start of the tone gesture.

Hypothesis 2.1: There is a significant effect of carrier verb on the timing of the start of the tone gesture.

The main questions of interest regard the interaction of segmental structures and tone—specifically, what the role of the syllable onset is in tone timing. All hypotheses will be tested for both the Belgrade and the Valjevo dialect. I present here hypotheses as they align with four theories of tone timing:

X.0 Null hypothesis, a sort of strawman hypothesis in which there is no relationship between a TBU and its tone—specifically, the target of a tone gesture is not anchored to any point, and the tone gesture simply starts at the beginning of the word;

X.1 Segmental anchoring, where both the start of a tone gesture and the end of a tone gesture are anchored to points in the acoustic segmental string;

X.2 C-center hypothesis, where only the start of a tone gesture is coordinated, and where the tone gesture acts as the last gesture of a complex syllable onset;

X.3 Articulatory anchoring, where both the start of a tone gesture and the end of a tone gesture are anchored to articulatory landmarks in the segmental structure.

I first examine the relationship between the properties of the syllable onset and the timing of the target of the H gesture.

Hypothesis 3.0 (null hypothesis): H targets are not anchored to any point in tone-bearing unit.

Prediction 3.0: There is no effect of syllable onset on the timing of the accentual peak.

Hypothesis 3.1 (segmental anchoring): H targets are acoustically anchored to some point in the nucleus.
Prediction 3.1: There is an effect of the phonetic duration of the syllable onset on the timing of the accentual peak (segmental anchoring hypothesis).

Hypothesis 3.2 (c-center): H targets are not articulatorily anchored, but are affected by the number of gestures in the tone-bearing unit onset.
Prediction 3.2: There is a significant effect of phonological complexity of the syllable onset on the timing of the accentual peak.

Hypothesis 3.3 (articulatory anchoring): H targets are anchored to some point in the nucleus, but precise timing also depends on the number of gestures in the tone-bearing unit onset.
Prediction 3.3: There is an effect of both the phonetic duration and the phonological complexity of the syllable onset on the timing of the accentual peak.

Second, I investigate the timing of the start of the H gesture. To distinguish between the null hypothesis and the segmental anchoring hypothesis, I set the start of the H gesture at 0 for the null hypothesis, while it can be at any other point for segmental anchoring.

Hypothesis 4.0 (null hypothesis): H gestures start at the same time as the word. **Prediction 4.0**: There is no effect of syllable onset on the timing of the start of the H tone gesture, and the lag between the start of the word and the start of the H gesture is 0.

Hypothesis 4.1 (segmental anchoring): The start of H gestures is anchored to some point in the tone-bearing unit.
Prediction 4.1: There is no effect of the syllable onset on the timing of the start of the H tone gesture, but the lag may not be 0.

Hypothesis 4.2 (c-center): H targets are coordinated as the second (in a simple onset) or third (in a complex onset) gesture with the syllable onset.
Prediction 4.2: There is a significant effect of the phonological complexity of the syllable onset (number of gestures) on the timing of the start of the H tone gesture.

Hypothesis 4.3 (articulatory anchoring): H targets are articulatory anchored to some point in the nucleus, and that point is influenced both by the duration of the other gestures in the onset, as well as the number.
Prediction 4.3: There is a significant effect of both the phonological complexity and the phonetic duration of the syllable onset on the timing of the start of the H tone gesture.

I also investigate the duration of the H gesture. Here, both the null hypothesis and the c-center hypothesis predict "ballistic" tone gestures; they would be distinguished by the results of Hypothesis 4.

Hypothesis 5.0 (null hypothesis): Tone gestures are ballistic in nature.
Prediction 5.0: There is no effect of syllable onset on the duration of the H

gesture.

Hypothesis 5.1 (segmental anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Prediction 5.1: There is a significant effect of the phonetic duration of the syllable onset on the duration of the H gesture.

Hypothesis 5.2 (c-center): Tone gestures are ballistic in nature.

Prediction 5.2: There is no effect of syllable onset on the duration of the H gesture.

Hypothesis 5.3 (articulatory anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Prediction 5.3: There is a significant effect of both the phonetic duration and phonological complexity of the syllable onset on the duration of the H gesture.

Finally, I compare excursion duration across dialects.

Hypothesis 6.0 (null hypothesis): Both dialects have H gestures with the same duration.

Prediction 6.0: There is no significant effect of dialect on the duration of the H excursion.

Hypothesis 6.1: Valjevo has shorter pitch excursions than Belgrade, in alignment with their earlier peaks.

Prediction 6.1: There is a significant effect of dialect on the duration of the H excursion; specifically, Valjevo excursions are shorter than Belgrade excursions.

In the following section, I describe the experiment conducted to investigate these hypotheses.

3.1.2 Stimuli

3.1.2.1 Target words

This study examines all four accents on either disyllabic or trisyllabic words. The short rising accent (*mravinjak*) is necessarily a trisyllabic word so as to prevent accent neutralization (as described in Chapter 1); the additional trisyllabic short falling word (*mramora*) was included to allow direct comparison if necessary.

A set of base words is set in combination with five syllable onsets, /r, l, m, mr, ml/, which vary on the first syllable onset of the word. Two different clusters were used in order to include a cluster that produces a real word as well as one that produces a nonce word, and to examine the effects of clusters with different phonetic durations, but the same phonological complexity. The singleton onsets were used as comparison, and included all the parts of the clusters; similarly to the two complex onsets, the singleton onsets are phonologically the same complexity but have different phonetic durations, which again allows a comparison of the effects of phonological complexity vs. phonetic duration. The words used in this study are provided in Table 3.1.

Since Serbian 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 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. 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 *mrämor* was a type of stone, *mlämor*, *mämor*, *lämor*, and *rämor* were other types of stone. In the case of the long rising accent, they were told that all were

Table 3.1: Target words used in Experiment 1, organized by accent type.

Length	**Orthography**	**Phonology**	**Gloss**
		Falling	
Long	mrave	/ˈmraː$_H$ve/	"ant.ACC.PL"
	mlave	/ˈmlaː$_H$ve/	*nonce*
	mave	/ˈmaː$_H$ve/	*nonce*
	lave	/ˈlaː$_H$ve/	*nonce*
	rave	/ˈraː$_H$ve/	*nonce*
Short	mramor	/ˈmra$_H$mor/	"marble"
	mlamor	/ˈmla$_H$mor/	*nonce*
	mamor	/ˈma$_H$mor/	*nonce*
	lamor	/ˈla$_H$mor/	*nonce*
	ramor	/ˈma$_H$mor/	*nonce*
	mramora	/ˈmra$_H$mora/	"marble.GEN.SG"
	mlamora	/ˈmla$_H$mora/	*nonce*
	mamora	/ˈma$_H$mora/	*nonce*
	lamora	/ˈla$_H$mora/	*nonce*
	ramora	/ˈma$_H$mora/	*nonce*
		Rising	
Long	Mronu	/ˈmroːnu$_H$/	*nonce*
	Mlonu	/ˈmloːnu$_H$/	*nonce*
	Monu	/ˈmoːnu$_H$/	"Mona.ACC.SG"
	Lonu	/ˈloːnu$_H$/	*nonce*
	Ronu	/ˈroːnu$_H$/	"Rhone.ACC.SG"
Short	mravinjak	/ˈmravi$_H$njak/	"anthill"
	mlavinjak	/ˈmlavi$_H$njak/	*nonce*
	mavinjak	/ˈmavi$_H$njak/	*nonce*
	lavinjak	/ˈlavi$_H$njak/	*nonce*
	ravinjak	/ˈmavi$_H$njak/	*nonce*

names of fashion brands, following *Mónu*, rather than *Rónu*[1]. Participants were allowed to reference this sheet through the study, though none had to.[2]

[1] Although *Ronu* (nominative form *Rona*) is a real river name, i.e. the river Rhône, it was found in the pilot study to be less commonly known than *Monu* (nominative form *Mona*), which is a popular fashion brand in Belgrade.

[2] There were two words that sometimes caused problems, due to sharing orthographic representation with existing words: first, the word *rave* with a long rising accent (rather than the long falling accent desired) is a slang term that means roughly 'ho.ACC.PL' (< whore); second, the word *lave* with a long rising accent (again, instead of the desired long falling accent) means 'lava.GEN'. Typically, participants were consistent

3.1.2.2 Carrier phrases

In order to probe the effects of tonal crowding (as reported in Zec and Zsiga 2016; Zsiga and Zec 2013), there were two carrier phrases in this study: one that has a pitch peak on the first of three syllables (*nêmamo X* /ˈneː$_H$mamo/ ‘we don’t have X’), and one that has a pitch peak on the second of three syllables (*ìmamo X* /ˈima(ː)$_H$mo/[3] ‘we have X’). These phrases and pragmatic contexts were chosen in order to encourage focus on the target word. There were 50 phrases total (**5** accent types x **5** syllable onsets x **2** carrier phrases).

3.1.2.3 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 ni na jednoj slici modnu kuću* ‘We don’t have a fashion brand in any picture’) or that they had all instances of a lexical category (e.g., *Imamo slike za sve modne kuće* ‘We have pictures for all of the fashion brands’). The participant then responded in disagreement with a written response that was presented on the screen. In order to prevent overlap, rushing, and list intonation, the written response only appeared on the screen after the context prompt ended. Two example exchanges are in Figures 3.1 (where the target word is in bold here for clarity, but was not marked in the experiment):

For this experiment, 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 using PsychoPy. Participants were recorded in a quiet

with their pronunciation; there were participants that almost always used the wrong accent, and participants that always remembered the correct and novel accent.

[3] For one and a half participants from Valjevo, this was produced as [iˈmáːmo], which is typical for the region. Only one participant had this naturally; the half is one participant that paused halfway through the experiment and noted that usually in Valjevo the stress is shifted rightward, though they did not typically use that pronunciation as their parents were from Belgrade (but then proceeded to complete the experiment with the Valjevo pronunciation). The other Valjevo participants had already shifted to the Belgrade stress pattern in the year they had been studying at the university.

Context:	Nemamo ni na jednoj slici modnu kuću.
	“We don’t have a fashion brand in any picture.”
Response:	Nije tačno! Imamo **Monu**.
	“That’s not true! We have Mona.”

(a)

Context:	Imamo sva mesta gde žive bube.
	“We have all the places where bugs live.”
Response:	Nije tačno! Nemamo **mravinjak**.
	“That’s not true! We don’t have an anthill.”

(b)

Figure 3.1: Example contexts and responses.

room with a Samson GoMic. The experiment was recorded independently from PsychoPy as one sound file.

3.1.3 Participants

Data was collected in late summer 2016 at the Faculty of Philology at the University of Belgrade in Belgrade, Serbia. In this chapter I am presenting the data from 5 native speakers of Belgrade Serbian (ages 19-39; 3 male, 2 female) and 4 native speakers of Valjevo Serbian (ages 19 - 22; 1 male, 3 female). Since living in Belgrade for an extended amount of time affects the realization of accent, the Valjevo speakers were all young university students that still had family ties in Valjevo and frequently visited home. The Belgrade speakers were all born and raised in Belgrade, though typically one or both parents were from elsewhere. Some speakers that participated in this experiment had also participated in a pilot study in March of the same year.

As all participants were fluent speakers of English, written consent was provided with a consent form in English.

3.1.4 Data labeling and analysis

3.1.4.1 Segmentation

Data was initially aligned with the Montreal Forced Aligner (McAuliffe et al. 2017), and then corrected by hand in Praat (Boersma & Weenink 2017). Due to wide variation in the production of the carriers "imamo" and "nemamo",[4] only the boundaries of the word were corrected; segments were not corrected (and segment edges are not used as landmarks in the analysis).

Marking segment boundaries was typically straightforward, with the exception of /r/ in word onsets. Although /r/ in Serbian is most often realized as a tap, I did not mark the onset at the beginning of the closure. This decision was made largely based on how /r/ presents in /mr/ clusters, which is as a brief, fully open period following the m, followed by the closure (see Figure 3.2). Although there is not always a clear demarcation of this small voiced period in r-initial words, the final vowel of the carrier word is consistently longer when preceding r-initial words than when preceding other words. Thus, I also marked the beginning of /r/ in r-initial words prior to the closure, and included a short vocalic interval at the end of the preceding vowel; in some cases, this was accompanied by either a leveling off or a drop in F3. Word-medial /r/ (as in *mramo**ra***) was not treated in the same way, but was simply labeled as the closure. This was due to extensive creak and devoicing near the word ends, which obscured the vowel movements. Additionally, these /r/'s are not used for landmarking in the analysis.

3.1.4.2 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 done using a Matlab script that first found pitch extrema located within certain boundaries—for example, no earlier than the acoustic beginning of

[4] Productions for *imamo* ranged from [imamo] to [imːo] to [imːː] to [iamo]; similarly, *nemamo* appeared as [neːmamo], [nemːo], [nemːː], and [neamo].

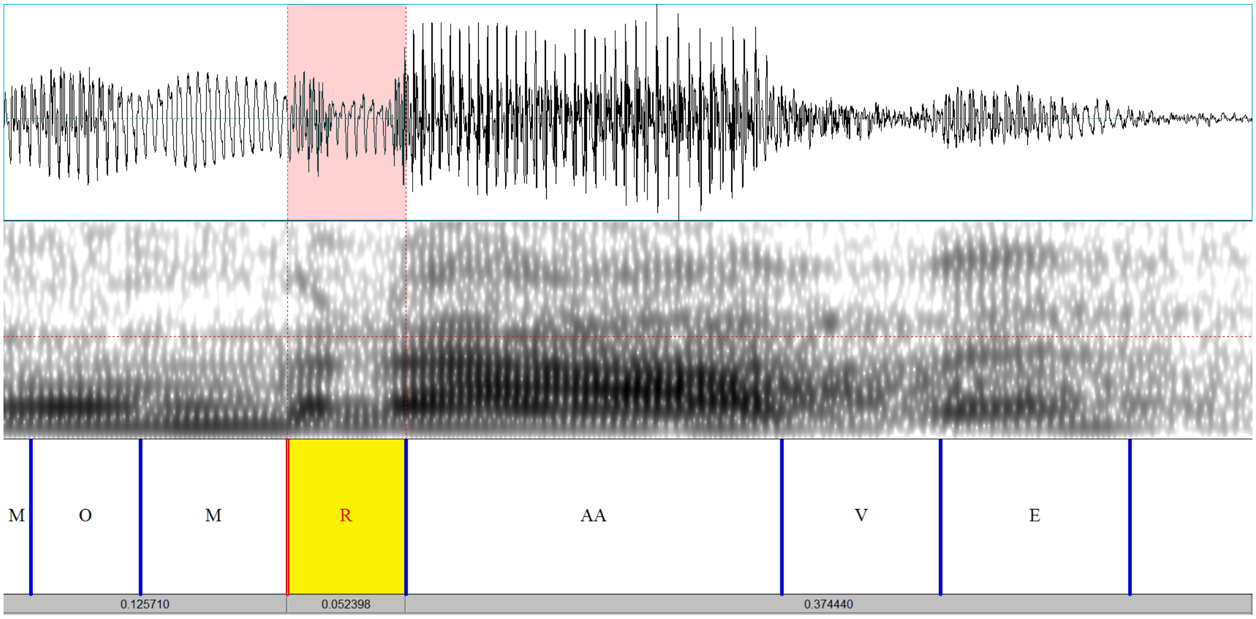


Figure 3.2: An example of /r/ segmentation in *mrave*. The open, voiced portion before the closure here is approximately 15 ms long.

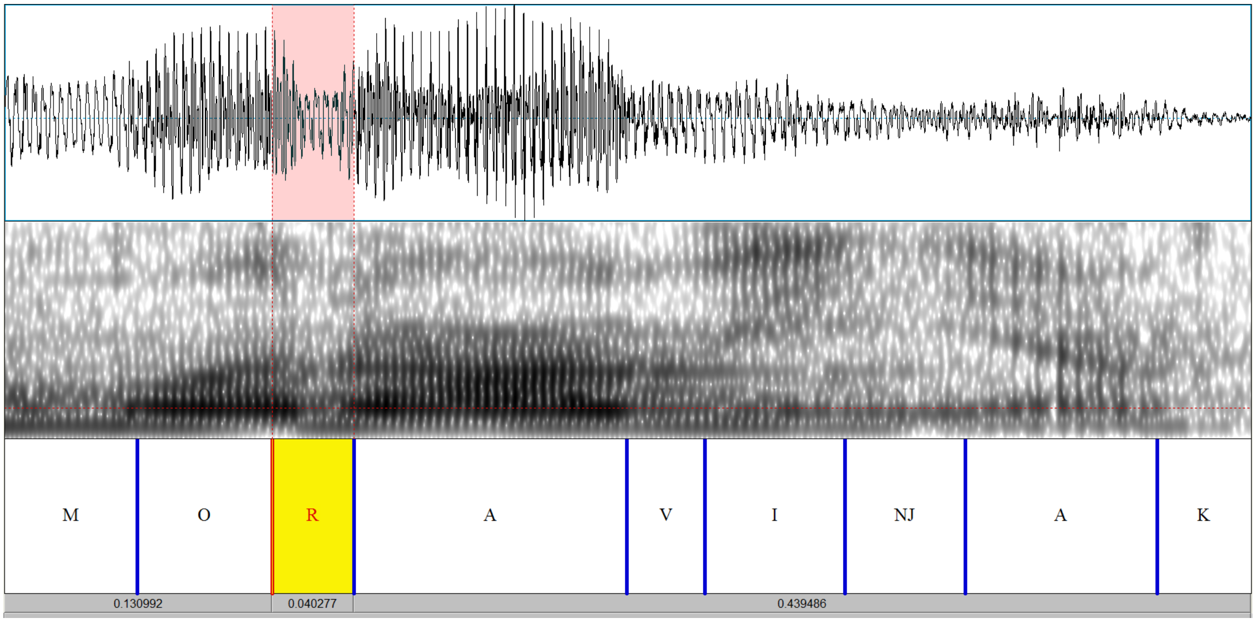


Figure 3.3: An example of /r/ segmentation in *ravinjak*.

the word, and no later than the second syllable nucleus for the F0 peak of a word with a falling accent.[5] These values were then used to bound where further landmarks could be located. An example of F0 marking on an actual token from Belgrade is given in Figure 3.4.

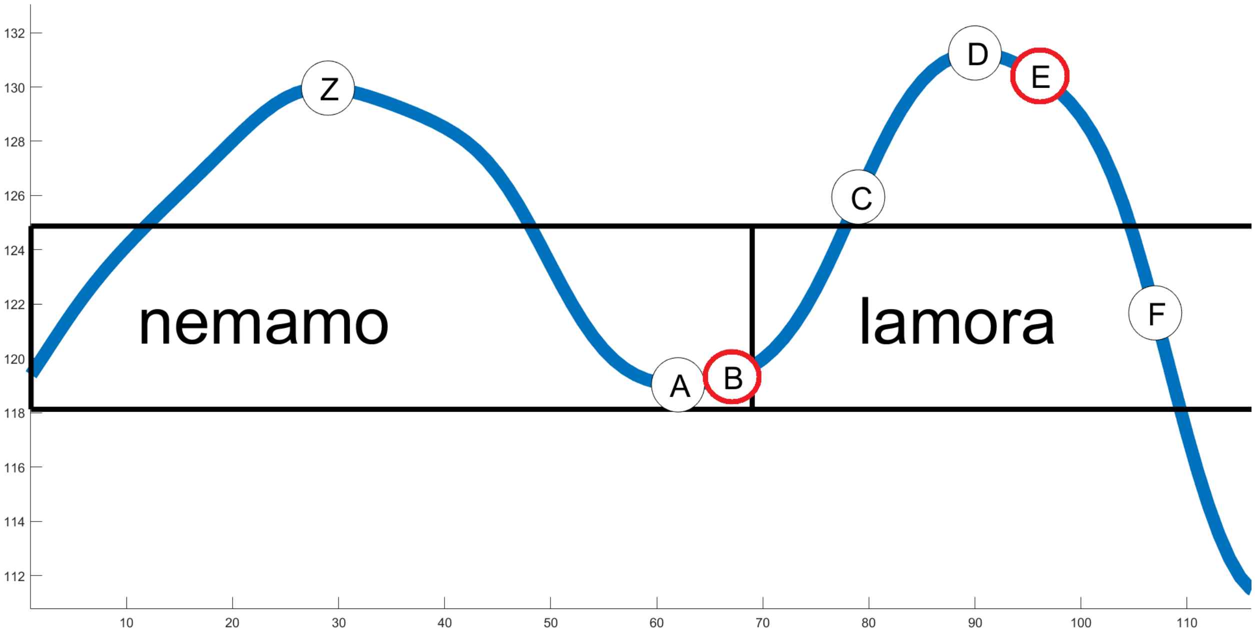


Figure 3.4: An example landmarked pitch trajectory. Z = carrier F0 peak; A = F0 valley; B = Excursion onset; C = maximum onset speed (Hz/s); D = target F0 peak; E = H release; F = maximum release speed.

As the absolute pitch peak is less stable and prone to small fluctuations, peak timing was measured using the gestural release rather than the actual target F0 peak; similarly, analyses that involve the start of upward F0 movement references the F0 onset, rather than the F0 valley:

- **H release** (E): The H gesture release was marked at the first point after the target word F0 peak (D) where F0 speed achieved 20% of the maximum release speed (this velocity point marked at F).
- **Excursion start** (B): The start of the F0 excursion was marked at the first point

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

after the F0 valley (A) where F0 speed achieved 20% of the maximum onset speed (this velocity point marked at C).

Except in cases of long plateaus (where any small fluctuations can result in a drastically displaced local maximum), the results do not wildly change in direction or significance when using F0 peak instead of H gesture relese; the 20% speed threshold ensures that only F0 changes of sufficient magnitude are labeled. These landmarks also make it possible to directly compare plateau-like peaks and true peaks, which is necessary due to the plateau-like nature of Valjevo pitch accents, particularly in rising accents. The H release marks a “shoulder” in the F0 trajectory, rather than a peaked “elbow”.

3.1.4.3 Statistical analyses

Throughout this chapter I will be using an α-level of 0.01; p-values below 0.05 but greater than 0.01 are considered “marginally significant” and the corresponding effects taken as a suggestion for further exploration. Statistical analyses were performed in R (R Core Team 2017), using the lme4 package (Bates et al. 2014) for linear mixed effects models. Models were built and compared incrementally, starting with the null model, which includes just `Part` as a random effect. For all analyses, all predictors are first examined in single fixed-effect models (i.e., one fixed effect and `Part` as a random effect), and compared with the AIC (Akaike Information Criterion; lower values are better). Nested models were compared with likelihood ratio tests, using the `anova` function from the lmerTest package (Kuznetsova et al. 2015). Homoskedasticity and normality of the residuals were assessed graphically.

Analyses were performed first within dialect, in order to examine each dialect’s timing patterns. The data from both dialects was then pooled for a cross-dialect analysis in order to probe interactions between timing and dialect. For all pitch-focused analyses, falling and rising accents are treated separately.

The variables (presented in `monospace font`) used in the analyses of the pitch excursions are the following:

Random effects

- `Part` (participant): Random intercepts for participant are included in all linear models.

Word group (e.g., -amora vs. -amor vs. -ave) is not included as a random effect, as there are too few groups to allow calculation of random intercepts. Order is also not included as a random effect, as the target words were presented in random order in each round. No models included random slopes.

Fixed effects

- `Complexity` (phonological complexity of the syllable onset): categorical variable with two levels, `simple` or `complex`
- `Identity` (identity of the syllable onset): categorical variable with five levels, /r/, /l/, /m/, /mr/, /ml/
- `OnsDur` (phonetic duration of the syllable onset): continuous variable, measured in seconds
- `Dialect` (dialect): categorical variable with two levels, `Belgrade` or `Valjevo`

Dependent variables All dependent variables were measured in seconds; for a schematic of these variables, see Figure 3.5.

- `PeakOffset` (peak timing relative to the beginning of the word): The time interval between the acoustic beginning of the word and the H gesture release (how much the H release is "offset" from the beginning of the word)
- `NucLag` (peak timing relative to the nucleus): The time interval between the acoustic beginning of the nucleus of the tone-bearing syllable and `PeakOffset`
- `ExcurStart` (start of the pitch excursion): The time interval between the acoustic beginning of the word and the start of the upward F0 excursion

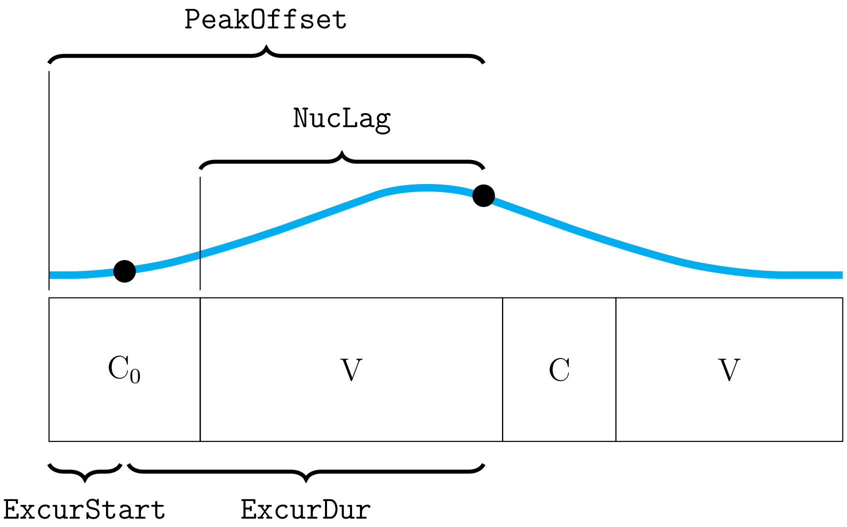


Figure 3.5: A schema of the dependent variables used in analysis. The blue line is a schematized short falling accent, with black dots to mark the start (leftmost) and peak offset (rightmost) of the pitch excursion.

- `ExcurDur` (excursion duration): The time interval between the peak offset and the start of the excursion

3.2 Results

The results section is structured as follows: first, a summary of segmental characteristics, followed by an analysis of the effects of carrier word, both of which present Belgrade and Valjevo results together as there are no major differences between dialects; second, a description of the phonetics of the Belgrade accentual system, including both falling and rising accents; third, a description of the phonetics of the Valjevo accentual system, including both falling and rising accents; and fourth, a comparison of the two dialects that focuses on just falling accents.

3.2.1 Segmental characteristics

3.2.1.1 Syllable onset duration

A comparison of linear mixed effects regressions shows that speakers of both dialects produced the hypothesized differences in duration for syllable onsets. When considering all tokens from the Belgrade dialect together, all syllable onsets have distinct durations: /r/ (M = 41.8 ms, SD = 8.1 ms) < /l/ (M = 66.1 ms, SD = 13.3 ms) < /m/ (M = 89.6 ms, SD = 15.8 ms) < /mr/ (M = 125.4 ms, SD = 18.9 ms) < /ml/ (M = 136.3 ms, SD = 21.4 ms), all p < 0.0001 (using least squares means Tukey test). When broken down by word type, the difference between /mr/ and /ml/ collapses for the *ȁmora* and *ónu* words (p = 0.33 and p = 0.35, respectively), though the means are numerically in the same relationship (i.e., /mr/ < /ml/). For the remaining word types, the difference between /mr/ and /ml/ is either significant (*ȁmor* p = 0.003; *âve* p = 0.001) or marginally significant (*àvinjak* p = 0.02). Syllable onset durations for each word type for the Belgrade dialect are illustrated in Figure 3.6.

The syllable onset duration patterns are similar in the Valjevo dialect, though with a few differences. There is a significant effect of syllable onset identity, though /mr/ and /ml/ are not significantly different from each other: /r/ (M = 51.7 ms, SD = 10.5 ms) < /l/ (M = 65.0 ms, SD = 11.9 ms) < /m/ (M = 98.4 ms, SD = 18.3 ms) < /mr/ (M = 137.0 ms, SD = 21.5 ms) = /ml/ (M = 141.1 ms, SD = 20.5 ms), all p < 0.0001 except /mr/ = /ml/ at p = 0.35. The distinction between /r/ and /l/ also collapses for *àvinjak* words (p = 0.11), and is only marginally significant for *ȁmora* words (p = 0.03).[6]). In addition, unlike for Belgrade, the /mr/ and /ml/ means numerically switch position *âve* words, which did not occur in the Belgrade dialect. These facts taken together suggest that in the Valjevo dialect /mr/ and /ml/ are even less distinct than in the Belgrade dialect, possibly due to /r/ and /l/ themselves being less distinct. Syllable onset durations for each word type are illustrated in

[6] It is possible that this is ultimately due to a smaller number of tokens (e.g., 14 for *ràvinjak* vs. 21 for *làvinjak*, compared to Belgrade's 28 *làvinjak* tokens and 30 *làvinjak* tokens.

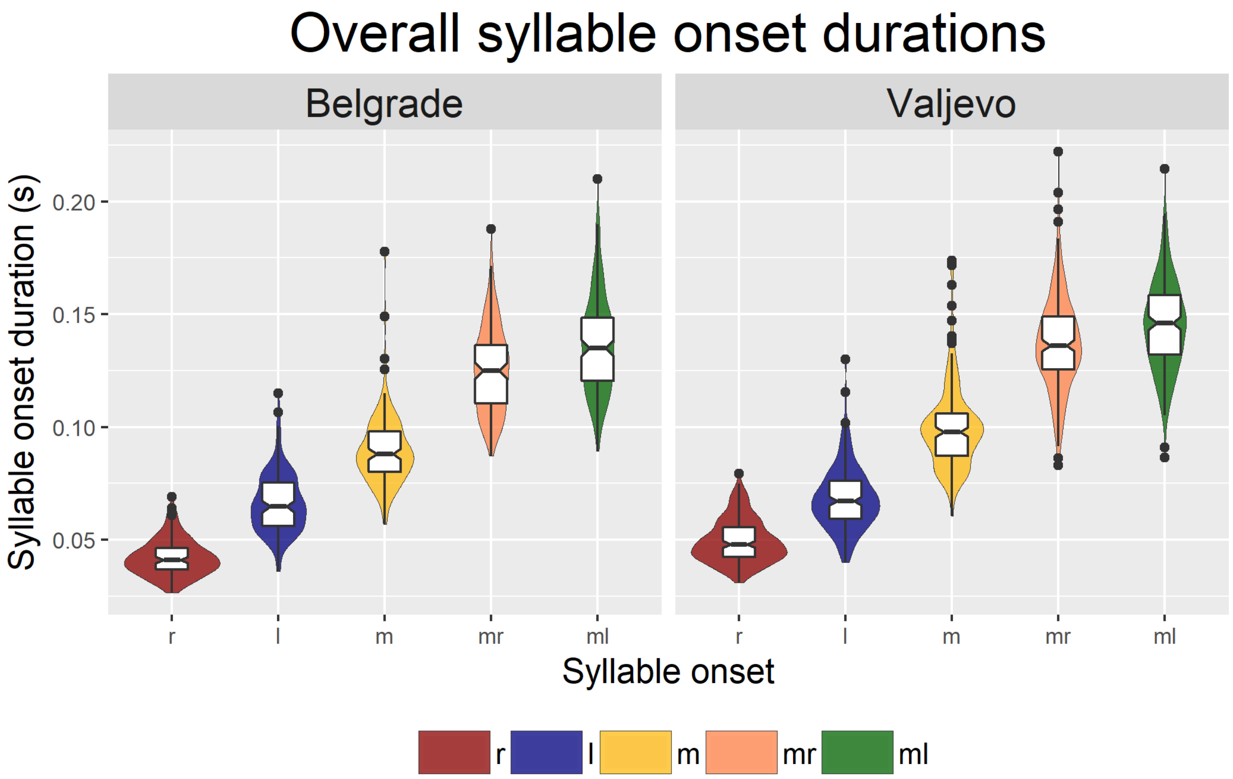


Figure 3.6: Syllable onset durations for the each dialect, all tokens combined.

Figure 3.6.

Finally, there is no effect of dialect on syllable onset duration, either as a single fixed effect ($\chi^2(1) = 1.74$, p $= 0.19$) or as a second main effect (in addition to syllable onset identity as the first main effect, $\chi^2(1) = 1.72$, p $= 0.19$). An illustration of the patterns of both dialects is provided in 3.6. Overall, Hypothesis A is upheld.

3.2.1.2 Duration of stressed nucleus

There is an effect of stressed vowel length (accent length) on the duration of the stressed nucleus ($\chi^2(1) = 2042.70$, p < 0.0001 for both dialects together; see Table 3.2). In the Belgrade dialect, the long accents (word types *ónu* and *âve*) are significantly longer than the short accents ($\chi^2(1) = 1114.00$, p < 0.0001). There are no statistically significant differences within short accents (*ȁmor*, *ȁmora*, and *àvinjak*), and there is only a marginally significant difference between the two long accents (p = 0.01 between *âve* and *ónu* according to a least squares means Tukey test). The difference between estimates here is quite small ($\beta = 6.2$ ms, SE = 1.9 ms), and is likely not meaningful.

There is also a significant effect of accent length in the Valjevo dialect; long accents are significantly longer than short accents ($\chi^2(1) = 1026.1$, $p < 0.0001$). However, in the Valjevo dialect there are statistically significant differences within lengths. There is no statistically significant difference between *ȁmor* and *àvinjak* (p = 0.75); however, nuclei are significantly longer in *ȁmor* words than in *ȁmora* words (p = 0.0002), and marginally shorter in *ȁmora* words than in *àvinjak* words (p = 0.04). For both statistically significant comparisons, the difference between estimates is fairly small ($\beta = 9.3$ ms, SE = 2.2 ms and $\beta = 6.5$ ms, SE = 2.3 ms, respectively). There is also a statistically significant difference between *ónu* and *âve* nuclei ($p < 0.0001$), and the difference between estimates is slightly larger ($\beta = 12.9$ ms, SE = 2.5 ms; *ónu* longer).

There is a marginally significant effect of `Dialect` on nucleus length, both when `Dialect` is a single fixed effect ($\chi^2(1) = 4.18$, p = 0.04), and when `Dialect` is a second fixed effect in a model that already includes `Length` ($\chi^2(1) = 4.26$, p = 0.04; see Table 3.2). In these comparisons, Valjevo nuclei are longer than Belgrade nuclei. Much of this effect seems to come from the long accents: there is a significant interaction between `Dialect` and `Length` ($\chi^2(1) = 134.67$, $p < 0.0001$). In this model, short accents are not significantly different between the two dialects (p = 0.35), while long accents are marginally different (p = 0.04). The differences between long and short accents are illustrated in Figure 3.7 for each dialect.

Finally, there is also a significant effect of `OnsDur` on nucleus duration ($\chi^2(1) = 120.28$, $p < 0.0001$; see Table 3.2). Longer syllable onsets are paired with shorter nuclei; however, as hypothesized, the effect is not sufficient to compensate for the duration of the syllable onset ($\beta = -133.0$ ms, SE = 11.8 ms; for every 1,000 ms increase in syllable onset duration, there is only a 133.0 ms decrease in nucleus duration). This effect is illustrated in Figure

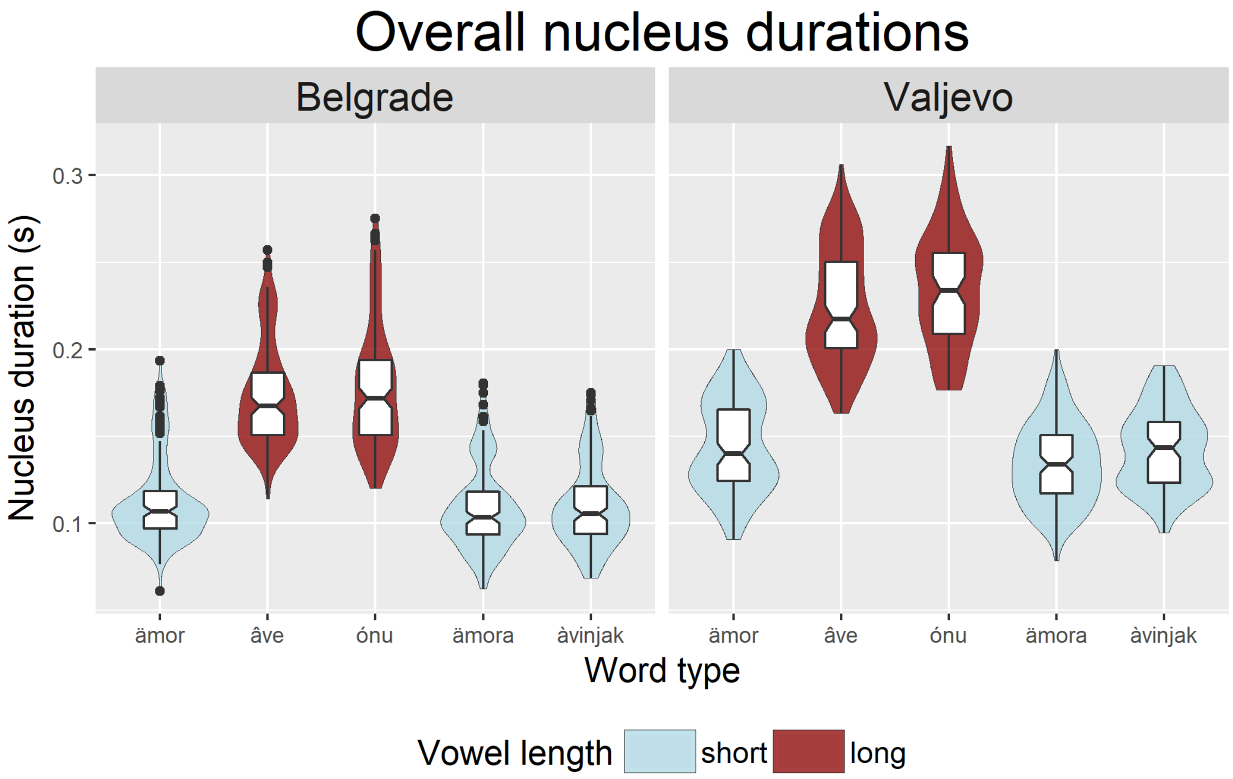


Figure 3.7: Nucleus durations, separated by word type; Belgrade on the left, Valjevo on the right.

Table 3.2: Comparison of linear mixed effects models for `NucDur` (stressed nucleus duration).

Model for `NucDur`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`Length + (1|Part)`	2042.70	1	$<$ 0.0001***
`Length + Dialect + (1|Part)`	4.26	1	0.04°
`Length + Dialect + Length:Dialect + (1|Part)`	134.67	1	$<$ 0.0001***
`Length + Dialect + Length:Dialect + OnsDur + (1|Part)`	120.28	1	$<$ 0.0001***
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

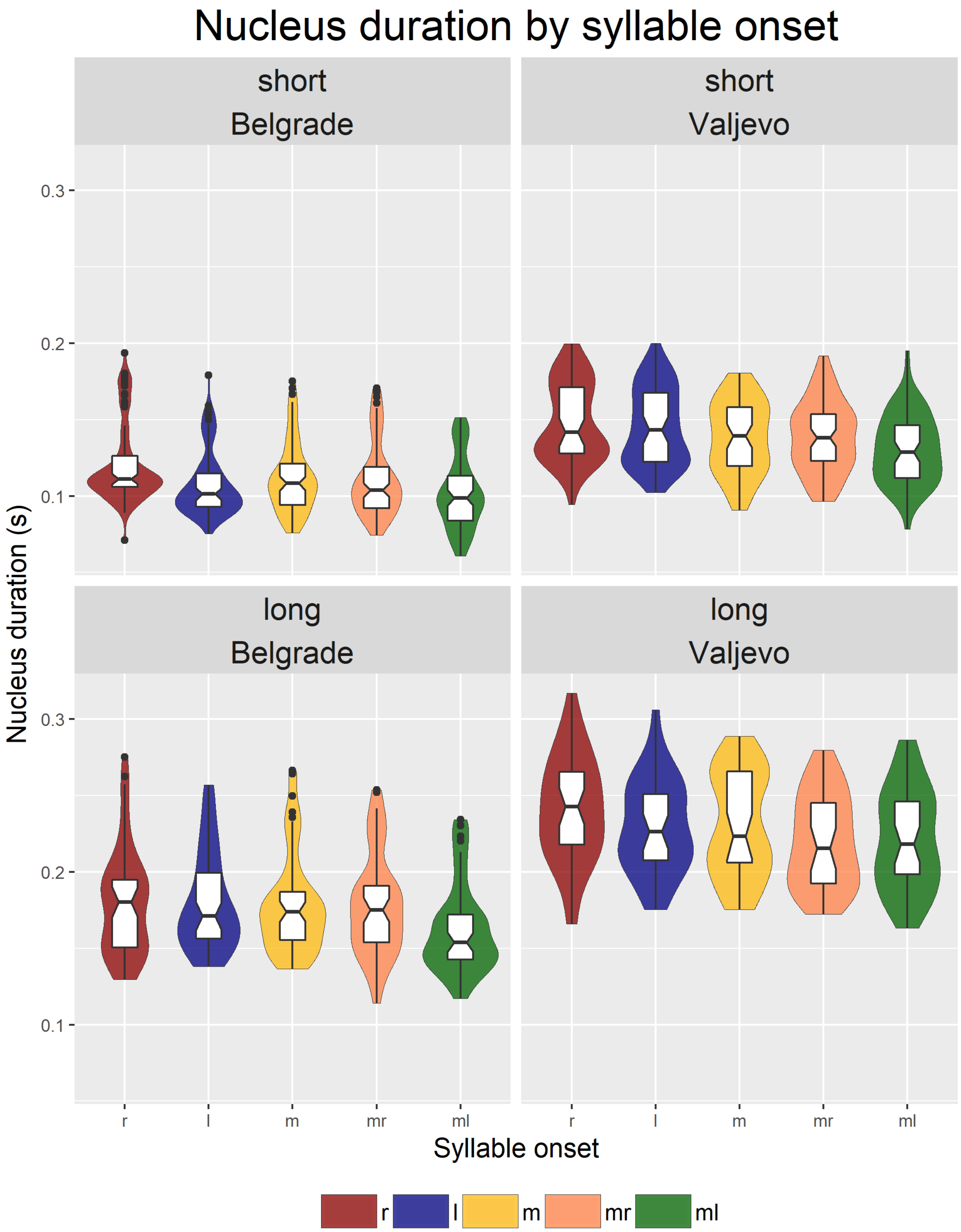


Figure 3.8: Nucleus durations, separated by syllable onset identity and vowel length; Belgrade on the left, Valjevo on the right.

3.8, grouped by syllable onset identity. Thus, differences predicted by Hypothesis A are not negated by compensatory shortening in the nucleus.

3.2.2 Effect of carrier

In this section I address Hypotheses 1 and 2:

> **Hypothesis 1.0** (null hypothesis): There is no effect of carrier verb on the timing of the accentual peak.
> **Prediction 1.0**: `Carrier` is not a significant predictor of `PeakOffset` (within accent type).

> **Hypothesis 1.1**: There is a significant effect of carrier verb on the timing of the accentual peak.
> **Prediction 1.1**: `Carrier` is a significant predictor of `PeakOffset` (within accent type).

> **Hypothesis 2.0** (null hypothesis): There is no effect of carrier verb on the timing of the start of the tone gesture.
> **Prediction 2.0**: `Carrier` is not a significant predictor of `ExcurStart` (falling accents evaluated only).

> **Hypothesis 2.1**: There is a significant effect of carrier verb on the timing of the start of the tone gesture.
> **Prediction 2.1**: `Carrier` is a significant predictor of `ExcurStart` (falling accents evaluated only).

There is not a significant effect of carrier (*nêmamo* or *ìmamo*) on either `PeakOffset` or `ExcurStart` in this study (see Table 3.3 for χ^2 and p values). This is true for both dialects

Table 3.3: A table of χ^2 values from comparing the null model $\sim$ `1 + (1|Part)` to $\sim$ `Carrier + (1|Part)` for `PeakOffset` and `ExcurStart`.

		Belgrade		Valjevo	
		$\chi^2(1)$	p	$\chi^2(1)$	p
`PeakOffset`	Falling	1.46	0.23	1.81	0.18
	Rising	0.90	0.34	0.17	0.68
`ExcurStart`	Falling	0.13	0.71	1.12	0.29
	Rising				

and both accents.[7] Thus, Hypotheses 1.0 and 2.0 are upheld. This contradicts the results reported by Zsiga and Zec (2013), who found that the location of the pitch peak in the carrier phrase had a significant effect on the timing of the valley before the target word peak (though not on the peak itself).

3.2.3 Pitch characteristics: Belgrade

3.2.3.1 Falling accents

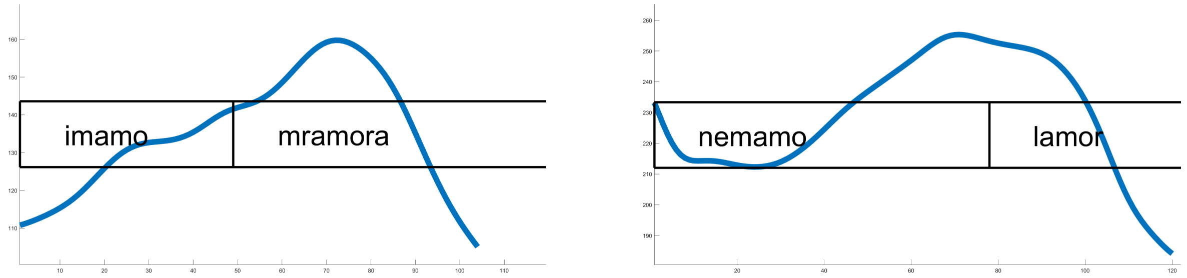
The data used to analyze the timing of the pitch excursions is a subset of all utterances produced by the participants. Tokens that were removed included tokens where an F0 peak (and thus peak offset) could not be found, segmental errors, accentual errors, and tokens with excessive pausing before the target word.[8] For analysis of H achievement (`PeakOffset` and `NucLag`), there were 429 falling tokens from Belgrade (out of a possible 450 tokens—4.5% removed).

This dataset was further cleaned for the analysis of the excursion characteristics (`ExcurStart` and `ExcurDur`), where 400 tokens from Belgrade were used (11% removed from total). Only F0 trajectories with clear and unambiguous valleys between the peaks of

[7] Note that `ExcurStart` is only measured for falling accents.

[8] The guideline for elimination was if F0 tracking was lost, as participants tended to have silent pauses; non-silent pauses were typically a false start on the target word without repeating the entire phrase.

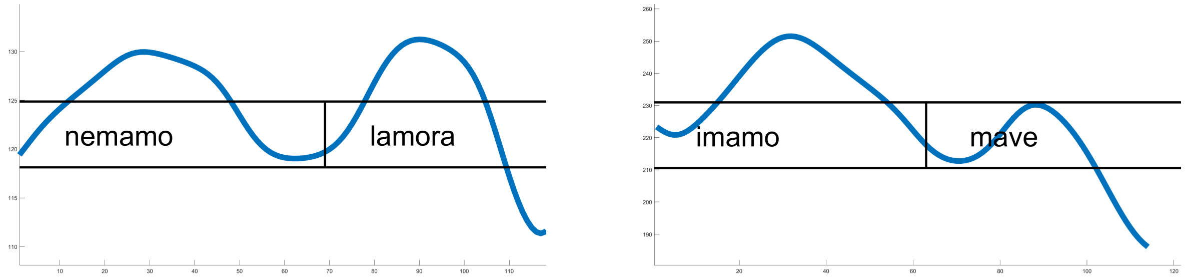
the carrier and target words can be used, as otherwise it is impossible to determine where the upward pitch excursion for the target word begins. Various intonational patterns completely obscured this minimum, including focus on the carrier word rather than on the target word. Two examples of such tokens from the Belgrade dialect are given in Figure 3.9; compare the tokens with clear minima in Figure 3.10.



(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 3.9: 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 3.10: Examples of F0 shapes that allow analysis of F0 onsets.

The breakdown of tokens used in the analysis is provided in Table 3.4.

H achievement (`PeakOffset` and `NucLag`) In this section, I address Hypothesis 3:

Hypothesis 3.0 (null hypothesis): H targets are not anchored to any point in

Table 3.4: Number of falling accent tokens with clear F0 peaks (H achievement) and clear minima (Excursion characteristics) for the Belgrade dialect.

Onset	**H achievement**			**Excursion char.**		
	ȁmor	ȁmora	âve	ȁmor	ȁmora	âve
ml	29	29	28	28	29	28
mr	31	28	28	30	26	24
m	30	27	28	29	26	27
l	28	28	29	25	28	26
r	29	29	28	26	24	24

tone-bearing unit.

Prediction 3.0:

- There is no effect of syllable onset on `PeakOffset`;
- `NucLag` will be increasingly negative with increased `OnsDur`.

Hypothesis 3.1 (segmental anchoring): H targets are acoustically anchored to some point in the nucleus.

Prediction 3.1:

- `OnsDur` positively correlates with `PeakOffset`; the changes in magnitude in `PeakOffset` parallel differences in `OnsDur`;
- `OnsDur` has no effect on `NucLag`.

Hypothesis 3.2 (c-center): H targets are not articulatorily anchored, but are affected by the number of gestures in the tone-bearing unit onset.

Prediction 3.2:

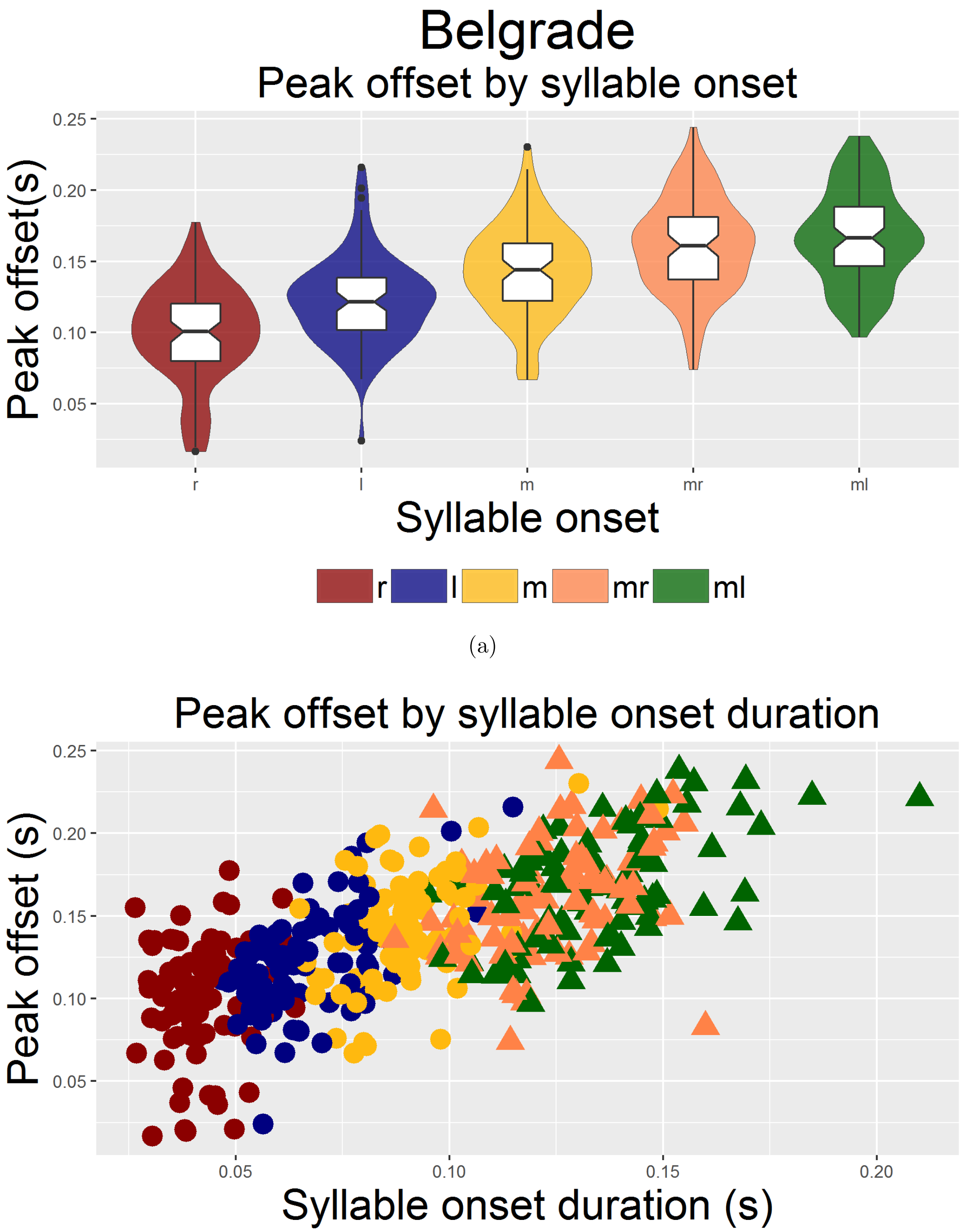
- There is a significant effect of `Complexity` on `PeakOffset`—complex onsets will have later peak offsets;
- There is a significant effect of `Complexity` on `NucLag`.

Hypothesis 3.3 (articulatory anchoring): H targets are anchored to some point in the nucleus, but precise timing also depends on the number of gestures in the tone-bearing unit onset.

Prediction 3.3: There is an effect of both `OnsDur` and `Complexity` on `PeakOffset` and `NucLag`.

There is an effect of syllable onset on the timing of the peak offset in the Belgrade dialect. A comparison of linear mixed effects regressions shows that both phonetic and phonological characteristics of the syllable onset affect the timing of the H achievement. `Complexity` as a single fixed effect provides a better fit than the null model ($\chi^2(1) = 141.42$, $p < 0.0001$, AIC = -1679.1), and `Identity` provides a slightly better fit ($\chi^2(4) = 220.25$, $p < 0.0001$ as compared to the null model; AIC = -1751.9). However, `OnsDur` as a single fixed effect provides a better fit than either categorical predictor ($\chi^2(1) = 293.67$, $p < 0.0001$ as compared to the null model; AIC = -1831.4; see Table 3.5a for a summary of the single predictor mixed models). The relationship between peak offset and syllable onset duration is illustrated in Figure 3.11. As syllable onset duration increases, the peak offset occurs later in the word: for every increase of 1000 ms in syllable onset time, the peak offset is delayed by 743.5 ms (SE = 36.1 ms).

Starting from a model that includes `OnsDur` as a fixed effect, the addition of `Complexity` as a second fixed effect does significantly improve the fit ($\chi^2(1) = 7.57$, $p = 0.006$). The further addition of the interaction term `Complexity:OnsDur` also significantly improves the fit ($\chi^2(1) = 9.79$, $p = 0.002$; see Table 3.5b). The interaction `Complexity:OnsDur` addresses differences in slope between complex and simple onsets, and as such is more meaningful than the main effect `Complexity` when added to `OnsDur`—`Complexity` predicts `OnsDur` in that complex onsets are longer than simple onsets, with very little overlap between the two types. However, the significant interaction indicates that an increase in `OnsDur` has a slightly larger effect in simple onsets ($\beta = 940.7$ ms, SE = 72.1 ms) than in complex onsets ($\beta = 809.0$ ms,



(a)

(b)

Figure 3.11: A boxplot (a) and scatter plot (b) showing the relationship between (intrinsic) syllable onset duration and F0 offset location in the Belgrade dialect. Red is /r/, blue /l/, yellow /m/, orange /mr/, green /ml/.

Table 3.5: Comparison of linear mixed effects models for `PeakOffset` (Belgrade dialect).

(a) Single predictor models, compared to the null model.

Model for `PeakOffset`	AIC	χ^2	DegF	p†
`Complexity + (1|Part)`	-1679.1	141.42	1	< 0.0001**
`Identity + (1|Part)`	-1751.9	220.25	4	< 0.0001**
`OnsDur + (1|Part)`	-1831.4	293.67	1	< 0.0001**

†As compared to the null model, `PeakOffset` $\sim$ 1 + `(1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `PeakOffset`	χ^2	DegF	p†
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Complexity + (1|Part)`	7.57	1	0.006*
`OnsDur + Complexity + Complexity:OnsDur + (1|Part)`	2.83	1	0.093
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Complexity:OnsDur + (1|Part)`	9.79	1	0.002*
`OnsDur + Complexity:OnsDur + Complexity + (1|Part)`	0.60	1	0.440

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

SE = 41.9 ms).

As the significance of the `Complexity:OnsDur` interaction suggests, there are differences between simple and complex onsets when comparing the lag between peak offset and the beginning of the nucleus (`NucLag`)—that is, while in simple onsets peak offset is delayed in a nearly one-to-one relationship with increase in `OnsDur`, the same is not true of complex onsets. As illustrated in Figure 3.12, peak offsets occur closer to the beginning of the nucleus in words with complex onsets than in words with simple onsets: /ml/ = /mr/ (p = 0.77); /mr/ < /m/ (p = 0.007) and /ml/ < /m/ (p < 0.0001); /m/ = /l/ (p = 0.99), /l/ = /r/ (p = 0.97), and /m/ = /r/ (p = 0.79).

In this case, the difference in `NucLag` is not driven by idiosyncratic variation in consonant duration, but rather by phonological complexity. Both `OnsDur` and `Complexity` significantly

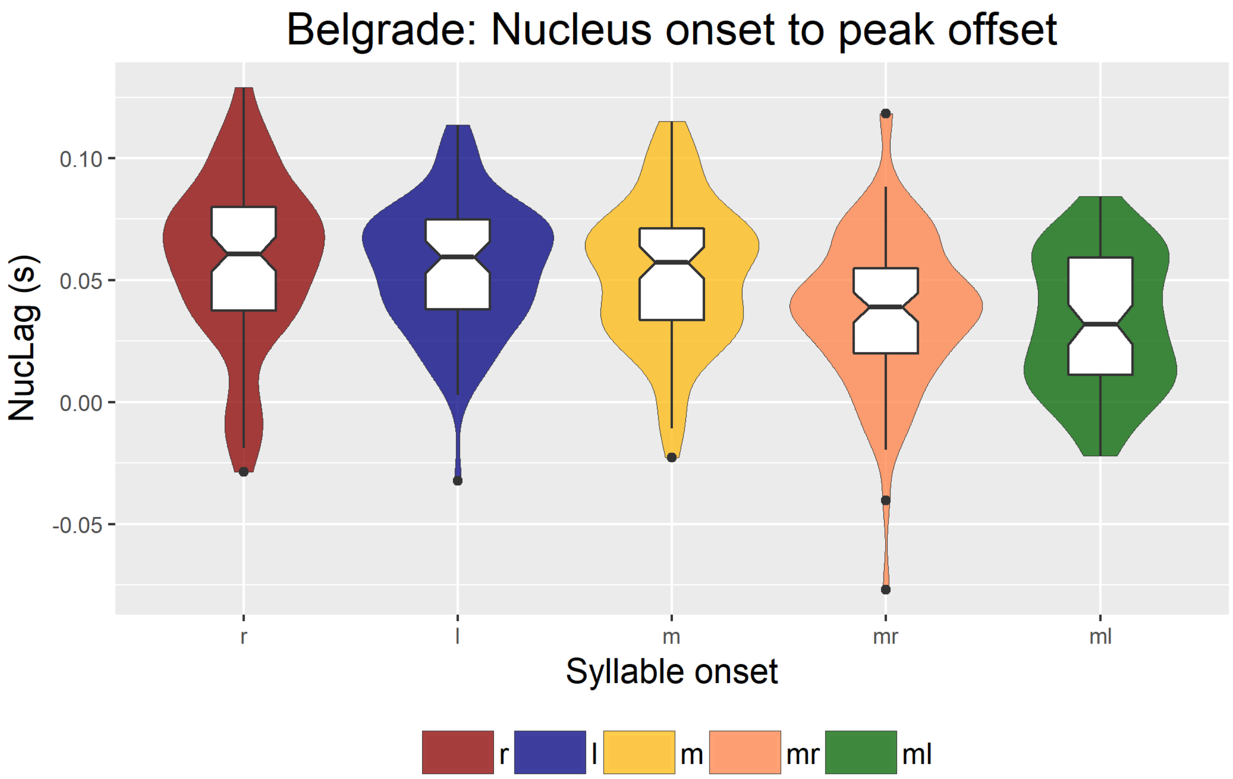


Figure 3.12: The time lag between the start of the nucleus and the peak offset, by onset identity (Belgrade dialect). Here 0 represents the start of the nucleus.

improve the model as compared to the null model (see Table 3.6a); as `OnsDur` is correlated with `Complexity`, it is not surprising that both predictors are significant. However, unlike the single predictor models for `PeakOffset`, the AIC does not suggest that `OnsDur` provides a better fit for `NucLag`. A comparison of three models further shows that `Complexity` is the best predictor of `NucLag` (see Table 3.6b). When `Complexity` is added as a second main effect alongside `OnsDur`, there is significant model improvement ($\chi^2(1) = 7.57$, p = 0.006); however, when `OnsDur` is added as a second main effect alongside `Complexity`, there is not significant model improvement ($\chi^2(1) = 2.96$, p = 0.085).

The same difference between simple and complex onsets is found when considering alignment of the peak offset relative to the end of the nucleus (`NucEndLag`, calculated as `time`

Table 3.6: Comparison of linear mixed effects models for NucLag (Belgrade dialect).

(a) Single predictor models, compared to the null model.

Model for NucLag	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-1836.0	52.16	1	< 0.0001**
OnsDur + (1|Part)	-1831.4	47.56	1	< 0.0001**

[†]As compared to the null model, NucLag ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for NucLag	χ^2	DegF	p[†]
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	7.57	1	0.006*
Complexity + (1|Part)	—	—	—
Complexity + OnsDur + (1|Part)	2.96	1	0.085

[†]As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

of peak offset - time of end of nucleus, thus peaks that occur before the end of the nucleus appear as negative values); given that there is no difference in nucleus duration between simple and complex onsets, this is not surprising. For short falling accents (i.e., word sets *ȁmor* and *ȁmora*), Complexity significantly improves the fit of the model ($\chi^2(1) = 18.78$, $p < 0.0001$), as does OnsDur ($\chi^2(1) = 10.34$, $p = 0.001$; see Table 3.7a); however, the AIC does not strongly suggest that one predictor provides a better fit than the other.

Upon comparing nested models, it becomes clear that for NucEndLag too, Complexity explains most of the variation: adding OnsDur to a model that already has Complexity does not significantly improve the fit ($\chi^2(1) = 0.48$, $p = 0.49$), but adding Complexity to a model that already has OnsDur does significantly improve the fit of the model ($\chi^2(1) = 8.92$, $p = 0.003$; see Table 3.7b). Thus, the peak offset overall occurs earlier relative to the nucleus in words with complex onsets than in words with simple onsets.

Thus, for the Belgrade dialect, Hypothesis 3.0 is rejected in favor of Hypothesis 3.3: peaks are not anchored to a single point in the nucleus; rather, both OnsDur and Complexity affect

Table 3.7: Comparison of linear mixed effects models for `NucEndLag` (Belgrade dialect, short falling accents only).

(a) Single predictor models, compared to the null model.

Model for `NucEndLag`	AIC	χ^2	DegF	p[†]
`Complexity + (1|Part)`	-1294.9	18.78	1	< 0.0001**
`OnsDur + (1|Part)`	-1286.5	10.34	1	0.001
[†]As compared to the null model, `NucEndLag` ∼ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b) Nested model comparisons.

Model for `NucEndLag`	χ^2	DegF	p[2]
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Complexity + (1|Part)`	8.92	1	0.003*
`Complexity + (1|Part)`	—	—	—
`Complexity + OnsDur + (1|Part)`	0.48	1	0.49
[2]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

the alignment of the peak.

Excursion characteristics (`ExcurStart` and `ExcurDur`) In this section, I investigate Hypothesis 4.

Hypothesis 4.0 (null hypothesis): H gestures start at the same time as the word.

Prediction 4.0:

- There is no effect of syllable onset on `ExcurStart`;
- `ExcurStart` is 0.

Hypothesis 4.1 (segmental anchoring): The start of H gestures is anchored to some point in the tone-bearing unit.

Prediction 4.1:

- There is no effect of the syllable onset on `ExcurStart`;

- ExcurStart may not be 0.

Hypothesis 4.2 (c-center): H targets are coordinated as the second (in a simple onset) or third (in a complex onset) gesture with the syllable onset.
Prediction 4.2: There is a significant effect of Complexity (number of gestures) on ExcurStart: ExcurStart is greater for complex onsets than for simple onsets.

Hypothesis 4.3 (articulatory anchoring): H targets are articulatory anchored to some point in the nucleus, and that point is influenced both by the duration of the other gestures in the onset, as well as the number.
Prediction 4.3: There is a significant effect of both Complexity and OnsDur on ExcurStart.

Both Complexity and OnsDur as single fixed effects significantly improve the fit of the model, compared to the null model; the AIC values of these two models suggest that OnsDur ($\chi^2(1) = 215.67$, $p < 0.0001$, AIC = -2040.4) provides a better fit than Complexity ($\chi^2(1) = 85.41$, $p < 0.0001$, AIC = -1910.1). Pitch excursions start later in words with longer syllable onsets (for the model with OnsDur as a single fixed effect, $\beta = 417.0$ ms, SE = 24.6 ms).

The addition of Complexity as a fixed factor to a model that already has OnsDur included significantly improves the model ($\chi^2(1) = 20.11$, $p < 0.0001$; see Table 3.8b). The difference between the two estimates is somewhat small (the pitch excursion starts approximately 14.6 ms earlier for complex onsets, SE = 3.2 ms), but on the same order of magnitude of the difference between estimates for the two complexity categories in NucLag (as discussed in Section **??**, 13.4 ms for a model that has both OnsDur and Complexity; 20.3 ms for a model that has only Complexity). In addition, there is not a significant interaction between OnsDur and Complexity on the start of the pitch excursion ($\chi^2(1) = 0.22$, $p = 0.64$; see Table 3.8b), which indicates that increases in syllable onset duration have the same delaying effect on the

Table 3.8: Comparison of linear mixed effects models for `ExcurStart`, Belgrade dialect.

(a) Single predictor models, compared to the null model.

Model for `ExcurStart`	AIC	χ^2	DegF	p[†]
`OnsDur + (1|Part)`	-2040.4	215.67	1	< 0.0001**
`Complexity + (1|Part)`	-1910.1	85.41	1	< 0.0001**

[†] As compared to the null model, `ExcurStart` $\sim$ 1 + `(1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Comparison of nested models.

Model for `ExcurStart`	χ^2	DegF	p[†]
`1 + (1|Part)`	—	—	—
`OnsDur + (1|Part)`	215.67	1	< 0.0001**
`OnsDur + Complexity + (1|Part)`	20.11	1	< 0.0001**
`OnsDur + Complexity + Complexity:OnsDur + (1|Part)`	0.22	1	0.64

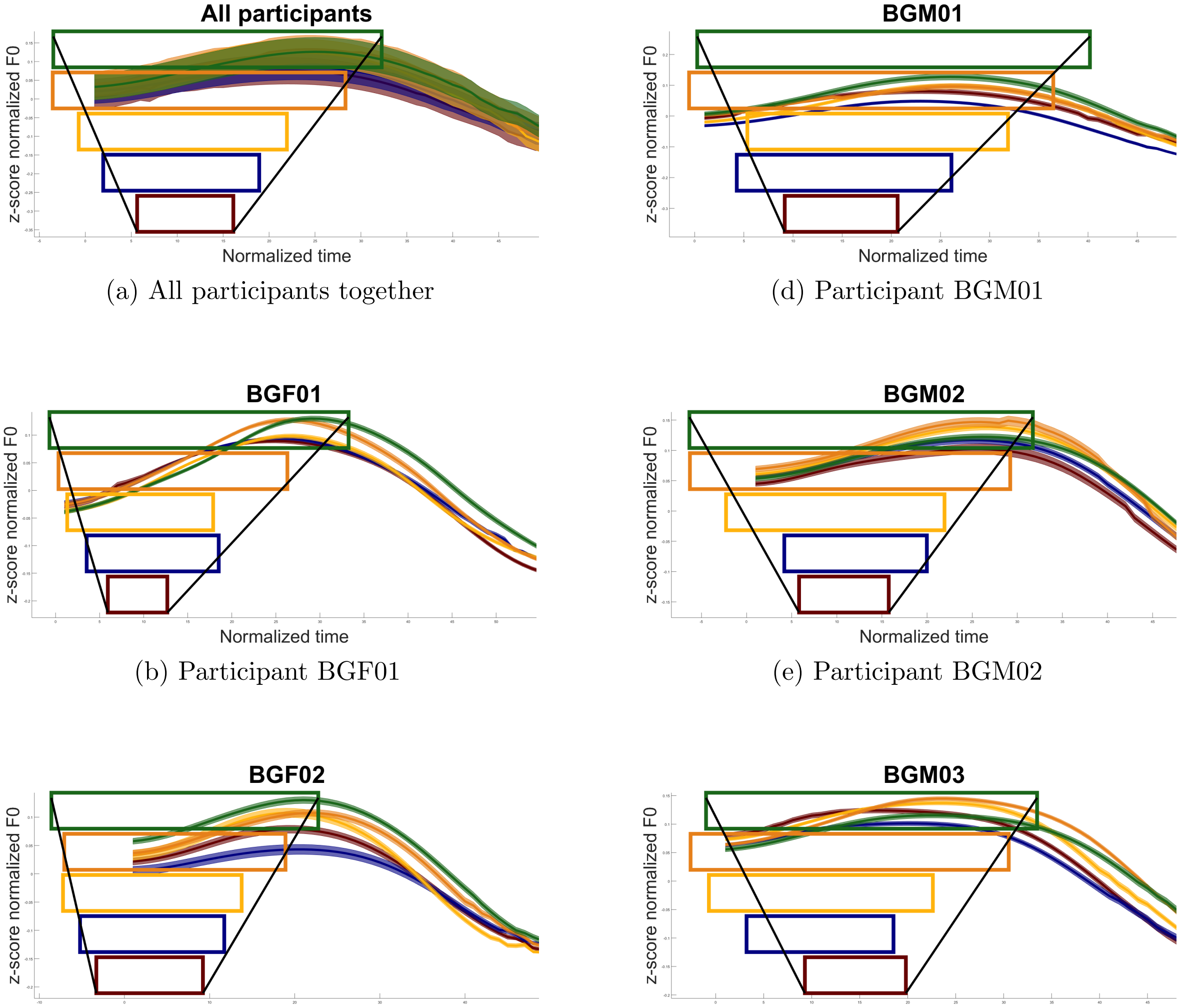
[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

start of the excursion in both complexity categories.

On average, pitch excursions start after the beginning of the word for complex onsets (/mr/ M = 14.3 ms, SD = 30.0 ms; /ml/ M = 13.3 ms, SD = 22.6 ms), and near the beginning of the word for both /m/ and /l/ (/m/ M = 3.2 ms, SD = 23.4 ms; /l/ M = -7.5 ms, SD = 22.2 ms). In contrast, pitch excursions start before the beginning of the word (i.e., in the carrier verb) with /r/ onsets (/r/ M = -21.8 ms, SD = 32.9 ms). Thus, the start of the pitch excursion is not anchored to the left[9] edge of the syllable onset. Rather, the acoustic edges of the syllable onset displace away from the start of the pitch excursion in both directions. This is illustrated in Figure 3.13, where 0 represents the beginning of the pitch excursion, not the beginning of the word. Figure 3.13a contains pitch trajectories and acoustic syllable onset edges for all participants together, while the remaining figures illustrate each participant's patterns separately. For each participant, the edges of the syllable onset displace in both directions from the start of the pitch excursion (highlighted by the black line from /r/ to

[9] Or right: significant effect of `OnsDur` on interval between the start of the pitch excursion and the beginning of the nucleus, $\chi^2(1) = 140.26$, $p < 0.0001$

/ml/).



(a) All participants together

(d) Participant BGM01

(b) Participant BGF01

(e) Participant BGM02

(c) Participant BGF02

(f) Participant BGM03

Figure 3.13: Z-score and time-normalized F0 trajectories (with standard error shading), Belgrade dialect. The acoustic edges of the syllable onset marked by boxes. Zero on the x axis represents the start of the pitch excursion. Both syllable onsets and F0 trajectories are color-coded in the same color scheme as used elsewhere; syllable onsets are arranged in order from /r/ at the bottom to /ml/ at the top.

These patterns suggest that it may be the midpoint of the syllable onset that F0 gestures are timed to—i.e., if the consonant edges are spreading in away from the start of the pitch excursion in both directions, the constant point may be the midpoint. However, there is still a significant effect of both OnsDur and Complexity on ExcurStarttoMid (the time interval between the start of the pitch excursion and the midpoint of the syllable onset); these patterns are illustrated in Figure 3.14. A comparison of linear models indicates that OnsDur significantly improves the fit of the model ($\chi^2(1) = 11.24$, p = 0.0008; see Table 3.9a); longer syllable onsets have slightly earlier F0 excursion starts, but the effect is quite small (β = -83.1 ms, SE = 24.6 ms, i.e. 83.1 ms earlier for every 1000 ms increase in syllable onset duration). Complexity appears to have a greater effect on the location of the start of the excursion: Complexity as a single predictor also significantly improves the fit of the model ($\chi^2(1) = 28.09$, $p < 0.0001$; see Table 3.9a), and the AIC suggests a slight advantage over the model with just OnsDur. However, as for the model with OnsDur as a single predictor, the difference in estimates is very small (β = -9.8 ms, SE = 1.8 ms).

Furthermore, the addition of Complexity to a model that includes OnsDur significantly improves the fit of the model ($\chi^2(1) = 20.11$, $p < 0.0001$; see Table 3.9b), but the converse is not true ($\chi^2(1) = 3.25$, p = 0.07). With OnsDur included, the effect of Complexity is slightly larger (β = -14.6 ms, SE = 3.2 ms), now closer to the realm of a meaningful difference, and is in fact the same difference reported when comparing the start of the pitch excursion to the left edge of the word. This suggests once again that the timing of the start of the F0 excursion is explained more by phonological characteristics (Complexity) than by phonetic characteristics (OnsDur). The interaction Complexity:OnsDur also does not significantly improve the model ($\chi^2(1) = 0.22$, p = 0.64). Thus, in the Belgrade dialect, Hypothesis 4.0 is rejected in favor of Hypothesis 4.4: there is a significant effect of both Complexity and OnsDur.

The following section concerns Hypothesis 5:

Hypothesis 5.0 (null hypothesis): Tone gestures are ballistic in nature.

Prediction 5.0: There is no effect of syllable onset on ExcurDur.

Hypothesis 5.1 (segmental anchoring): Tone gestures stretch with more seg-

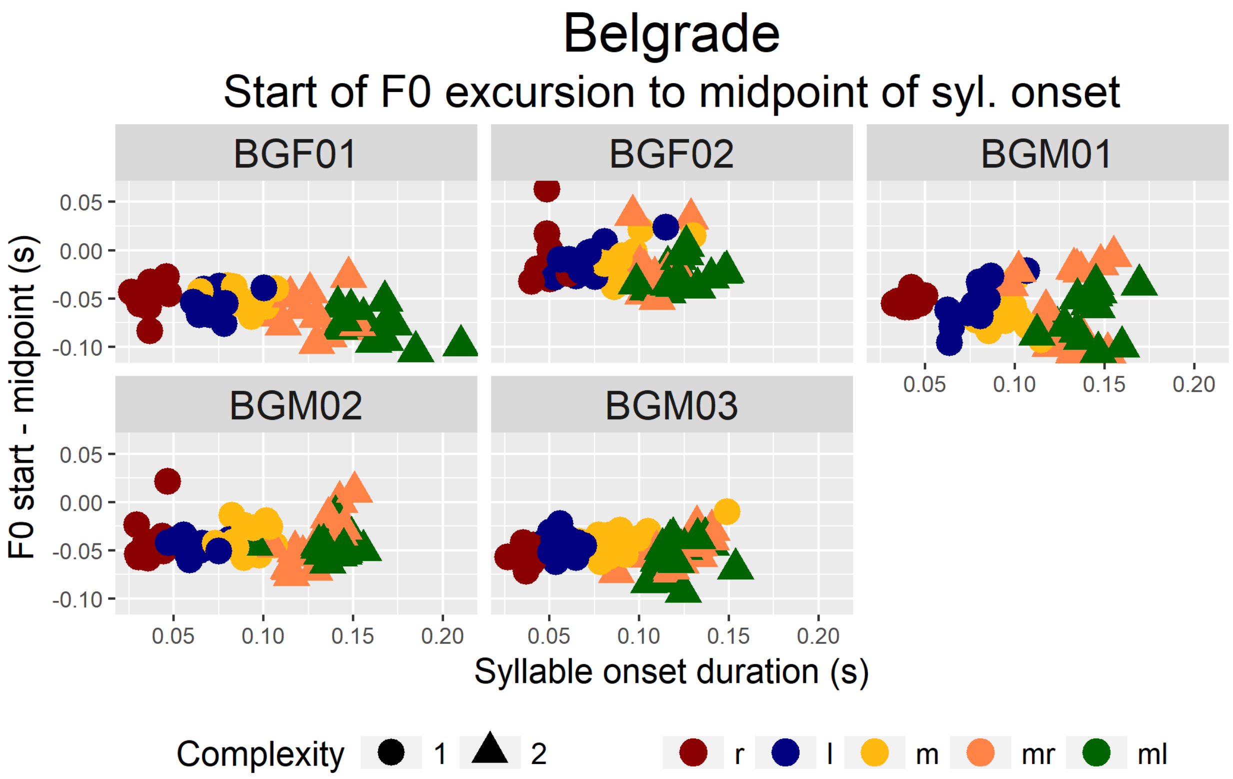


Figure 3.14: Distance from the midpoint of the syllable onset to the start of the F0 excursion (negative values indicate that the start of the excursion occurs prior to the midpoint), separated by participant (Belgrade dialect).

mental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Prediction 5.1: There is a significant effect of `OnsDur` on `ExcurDur`; changes in `OnsDur` parallel changes in `ExcurDur`.

Hypothesis 5.2 (c-center): Tone gestures are ballistic in nature.

Prediction 5.2: There is no effect of syllable onset on `ExcurDur`.

Hypothesis 5.3 (articulatory anchoring): Tone gestures stretch with more seg-

Table 3.9: Comparison of linear mixed effects models for F0StarttoMid, Belgrade dialect.

(a) Single predictor models, compared to the null model.

Model for F0StarttoMid	AIC	χ^2	DegF	p†
OnsDur + (1|Part)	-2024.4	11.24	1	0.0008**
Complexity + (1|Part)	-2057.2	28.09	1	< 0.0001**
†As compared to the null model, F0StarttoMid ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of nested models.

Model for F0StarttoMid	χ^2	DegF	p†
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	20.11	1	< 0.0001**
Complexity + (1|Part)	—	—	—
Complexity + OnsDur + (1|Part)	3.25	1	0.07
Complexity + OnsDur + Complexity:OnsDur + (1|Part)	0.22	1	0.64
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

mental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Prediction 5.3: There is a significant effect of both Complexity and OnsDur on ExcurDur; increases in Copmlexity and OnsDur correlate with increases in ExcurDur.

Syllable onsets affect the duration of the pitch excursion. Both Complexity ($\chi^2(1) = 49.56$, $p < 0.0001$) and OnsDur ($\chi^2(1) = 73.40$, $p < 0.0001$) significantly improve the model as single fixed effects; the AIC indicates that OnsDur provides a better fit than Complexity (see Table 3.10a). Increases in syllable onset duration elongate the upward pitch excursion, though the estimate is somewhat smaller than the estimates for both PeakOffset and ExcurStart ($\beta = 293.0$ ms, SE = 32.6 ms). Excursions are the shortest in words with /r/ and /l/ (/r/ M = 125.9 ms, SD = 29.6 ms; /l/ M = 132.5 ms, SD = 27.5 ms), and the

Table 3.10: Comparison of single-factor linear mixed effects models for ExcurDur, Belgrade dialect.

(a) Single predictor models, compared to the null model.

Model for ExcurDur	AIC	χ^2	DegF	p[†]
OnsDur + (1|Part)	-1814.4	73.40	1	< 0.0001**
Complexity + (1|Part)	-1790.6	49.56	1	< 0.0001**
[†]As compared to the null model, ExcurDur $\sim$ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of nested models.

Model for ExcurDur	χ^2	DegF	p[†]
1 + (1|Part)	—	—	—
OnsDur + (1|Part)	73.40	1	<0.0001**
OnsDur + Complexity + (1|Part)	0.01	1	0.91
OnsDur + Complexity + Complexity:OnsDur + (1|Part)	1.72	1	0.19
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

longest in words with complex onsets (/mr/ M = 148.1 ms, SD = 29.1 ms; /ml/ M = 154.4 ms, SD = 36.0 ms); /m/ once again patterns between the two extreme groups, but is only significantly different from /r/ (M = 142.5 ms, SD = 31.5 ms).

Neither Complexity as a simple fixed effect nor the interaction Complexity:OnsDur significantly improve the fit of a model that already has OnsDur as a fixed effect ($\chi^2(1) = 0.01$, p = 0.91 and $\chi^2(1) = 1.72$, p = 0.19 respectively; see Table 3.10b). This indicates a steady, shallow increase in excursion duration as the phonetic duration of the syllable onset increases. Thus, Hypothesis 5.0 is rejected in favor of Hypothesis 5.1.

The relationships between OnsDur, PeakOffset, ExcurDur, and ExcurStart are illustrated in Figure 3.15, as well as density plots (divided by syllable onset identity) for each individual variable. In the Belgrade dialect, both *when* the pitch excursion starts, as well as how long the pitch excursion is, contribute to the timing of the peak offset. However, only ExcurStart exhibits the same interaction between syllable onset duration and syllable onset complexity that PeakOffset has.

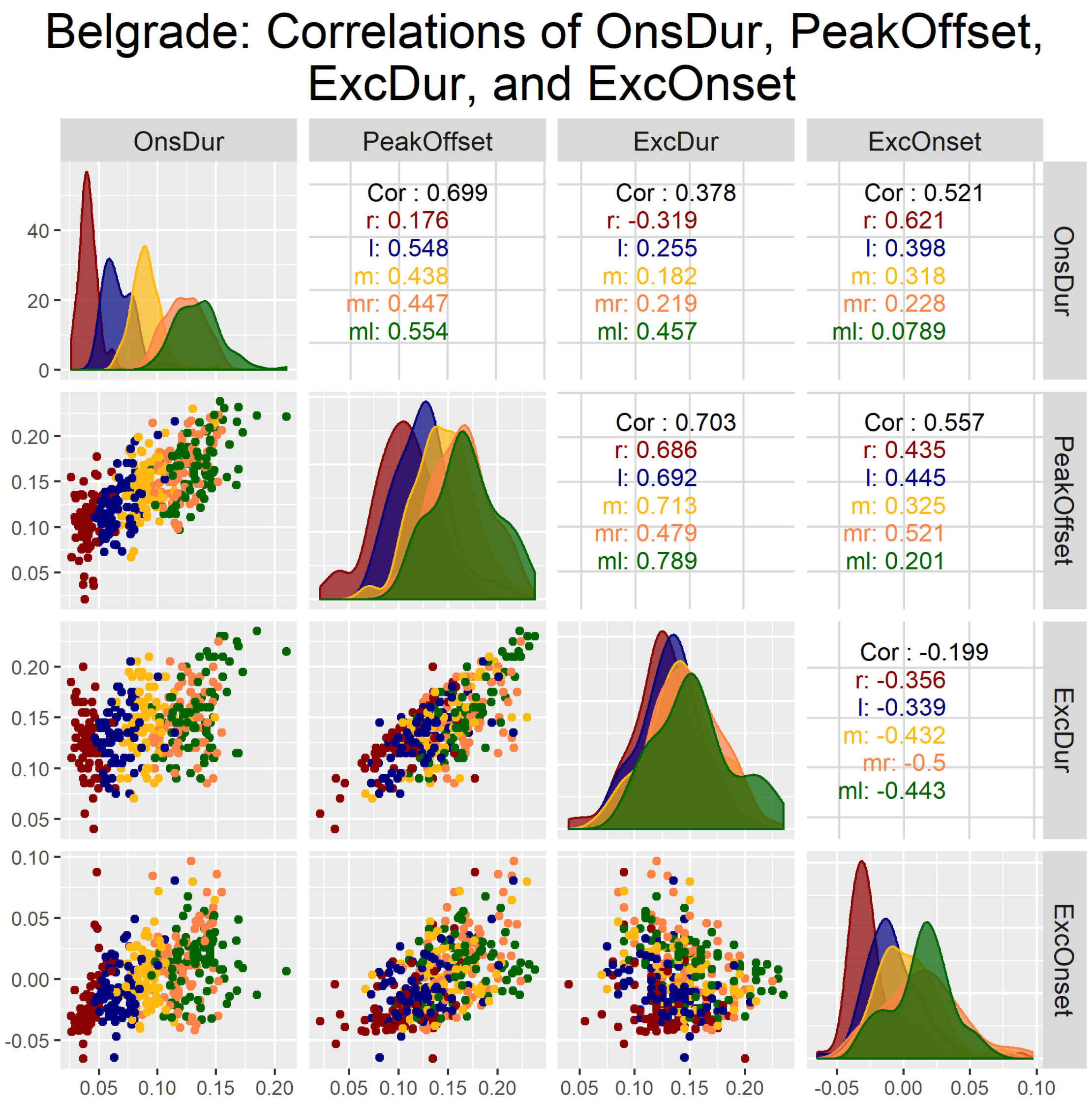


Figure 3.15: Correlation matrix for the Belgrade dialect that shows the relationships between syllable onset duration (`OnsDur`), peak offset location relative to the beginning of the word (`PeakOffset`), excursion duration (`ExcurDur`), and time lag between the start of the excursion and the beginning of the word (`ExcurStart`).

3.2.3.2 Rising accents

This section will be a fairly brief summary of the behavior of the rising accents. Due to extensive variation in the timing of the start of the rising pitch accents,[10] only the timing of the H achievement will be analyzed. After processing the data from the Belgrade dialect, 275 tokens out of the possible 300 remained in the Belgrade dialect (25 removed; 8.3% attrition). These tokens are summarized in Table 3.11.

Table 3.11: Number of rising accent tokens for each syllable onset with clear F0 peaks for the Belgrade dialect.

Onset	**Belgrade**	
	àvinjak	ónu
ml	29	28
mr	29	26
m	24	29
l	30	28
r	28	24

Short rising Short rising and falling accents are distinct in the Belgrade dialect: there is a significant effect of accent type (rising vs. falling) on `PeakOffset` ($\chi^2(1) = 403.66$, $p < 0.0001$). The difference between estimated means for the two accent types is 108.2 ms (SE = 4.2 ms), where rising accents occur later relative to the beginning of the word (here, the same as the beginning of the stressed syllable) than falling accents. On average, short rising accent peaks occur 60.5 ms *after* the acoustic beginning of the second syllable (SD = 26.5 ms), while short falling accent peaks occur 57.3 ms (SD = 29.1 ms) *before* the beginning of the second syllable. Thus, the contrast inherent in Hypothesis B is upheld for short accents in Belgrade.

Syllable onset has a significant effect on `PeakOffset`, where `PeakOffset` still reflects the location of the peak offset relative to the beginning of the word, not the beginning of

[10] As was documented by Inkelas and Zec (1988), rising accents can plateau over two moras or be confined to the tone-bearing mora.

Table 3.12: Comparison of nested linear mixed effects models for `PeakOffset`, short rising accents (Belgrade dialect).

(a) Single predictor models, compared to the null model.

Model for `PeakOffset`	AIC	χ^2	DegF	p†
`OnsDur + (1|Part)`	-670.5	155.75	1	< 0.0001**
`SylDur + (1|Part)`	-746.1	231.38	1	< 0.0001**
†As compared to the null model, `PeakOffset` $\sim$ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b) Nested model comparisons.

Model for `PeakOffset`	χ^2	DegF	p†
`OnsDur + (1|Part)`	—	—	—
`OnsDur + SylDur + (1|Part)`	81.02	1	< 0.0001**
`SylDur + (1|Part)`	—	—	—
`SylDur + OnsDur + (1|Part)`	5.39	1	0.02°
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

the H syllable. `OnsDur` as a single fixed effect significantly improves the fit of the model ($\chi^2(1) = 155.75$, $p < 0.0001$; see Table 3.12a); as the syllable onset increases in duration, `PeakOffset` also increases ($\beta = 802.6$ ms, SE = 47.0 ms). However, the duration of the first syllable (`SylDur`) also significantly improves the fit of the model ($\chi^2(1) = 231.38$, $p < 0.0001$), and the AIC value (-746.1) suggests that it provides a better fit than `OnsDur` (AIC = -670.5). In addition, the estimate is even closer to a one-to-one relationship with `SylDur`: for every 1,000 ms increase in syllable duration, there is a 927.1 ms delay in `PeakOffset` (SE = 38.1 ms). This indicates that the effect of `OnsDur` is due to the fact that syllable duration increases with longer syllable onsets—and the longer the first syllable is, the later the second (tone-bearing) syllable starts.[11]

This is confirmed when comparing models with both `OnsDur` and `SylDur`. When `SylDur` is added as a second fixed effect to a model that already has `OnsDur`, there is a significant

[11] `SylDur` also significantly improves the model for falling accents ($\chi^2(1) = 83.9$, $p < 0.0001$), but the AIC value (-1621.6) indicates that it this model provides a worse fit than the model with just `OnsDur` (AIC = -1831.4).

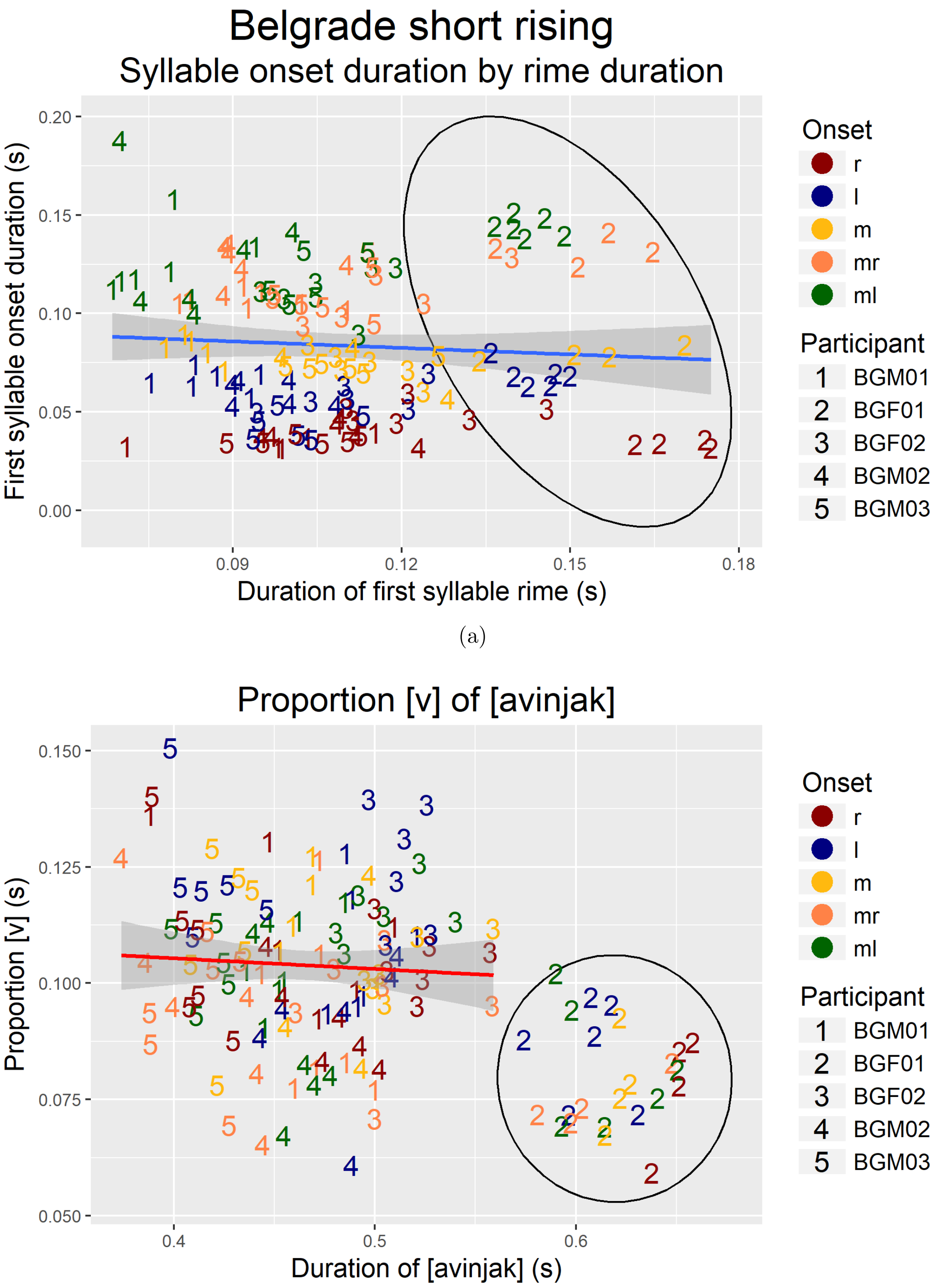
improvement in model fit ($\chi^2(1) = 81.02$, $p < 0.0001$; see Table 3.12b). In contrast, in the opposite order, there is only a marginal improvement when OnsDur is added ($\chi^2(1) = 5.39$, $p = 0.02$).[12] Thus, the duration of the whole syllable, rather than the duration of just the syllable onset, is more predictive of the timing of the H for rising accents.

An additional measure to consider for rising accents is the time interval between the acoustic beginning of the second syllable and the peak offset, henceforth PeakOffset2. This is the phonological parallel to PeakOffset for falling accents, i.e. the time lag between the pitch peak offset and the acoustic start of the syllable that underlyingly has lexical H. A comparison of linear models shows that OnsDur2 (the duration of the onset of the second syllable) significantly improves the model ($\chi^2(1) = 15.08$, $p = 0.0001$); as OnsDur2 increases, so does PeakOffset2 ($\beta = 315.4$ ms, SE = 138.6 ms).

In this case, unlike the variation found between syllable onsets in OnsDur, variation in OnsDur2 is entirely due to trial-to-trial variation or subject-to-subject variation, since all onsets are /v/. It is worth noting that there appears to be individual variation in C-to-V ratio, with one participant having particularly long vowels without a corresponding increase in syllable onset duration. This is illustrated in Figure 3.16. In Figure 3.16a, the duration of the first syllable onset (i.e., the one manipulated in this study) is plotted as a function of the duration of the first nucleus. As demonstrated by the horizontal separation, participant BGF01's nuclei are longer (significantly so; $p < 0.0001$ compared to all participants in a Tukey HSD test); however, the syllable onsets do not consistently exhibit a difference (as visualized by vertical separation; when accounting for intrinsic differences due to syllable onset identity, $p = 0.03$ compared to BGM01, $p = 0.20$ compared to BGM02, $p < 0.0001$ compared to BGM03, and $p = 0.0006$ compared to BGF02). Thus, BGF01 appears to have a different C-to-V ratio than the other speakers.

The effect of this C-to-V ratio on the duration of [v] is illustrated in Figure 3.16b, where the proportion of [avinjak] that is taken up by the duration of [v] is plotted against

[12] For falling accents, the reverse is true: OnsDur provides a significantly better fit when added to a model with SylDur ($\chi^2(1) = 218.94$, $p < 0.0001$).



(a)

(b)

Figure 3.16: Two figures showing the anomalous C-to-V ratio of participant BGF01 (marked with a "2"); ellipse marks a 0.95 confidence interval of the BGF01 tokens. Red is /r/, blue /l/, yellow /m/, orange /mr/, green /ml/; numbers mark each participant's tokens.

the duration of [avinjak]. The other participants seem to exhibit "uniform stretching": participant BGF02 (marked by "3") has a generally longer [avinjak] duration, but the [v] takes up the same proportion of the longer [avinjak] as for faster-speaking participants like BGM03 (marked by "5"); that is, speech rate does not appear to affect vowels more than consonants (or vice versa). This consistent proportion is supported by the flat fit line, which excludes participant BGF01. In contrast, BGF01's tokens are detached entirely from the other participants, both in duration of [avinjak] and in proportion taken up by [v]; the proportion of [v] is lower than other participants (in a Tukey HSD test, $p < 0.0001$ compared to all participants except BGM02, who is marked by "4", and where $p = 0.04$).

Finally, there is the question of `NucLag2`, which is the time interval between the peak offset and the beginning of the nucleus of the second syllable. It is not predicted that `OnsDur2` will have an effect on `NucLag2`; as shown in the section on falling accents, `NucLag` was consistent within phonological complexity, despite differences in intrinsic duration. Working under the thus-far supported assumption that the H in rising accents is in fact associated to the second syllable, not the first, `OnsDur` and `SylDur` are not predicted to have an effect on `NucLag2` either. These predictions are borne out: neither `OnsDur2` ($\chi^2(1) = 3.29$, $p = 0.07$), nor `OnsDur` ($\chi^2(1) = 1.24$, $p = 0.27$), nor `SylDur` ($\chi^2(1) = 0.03$, $p = 0.85$) significantly improve the null model.

Long rising Long rising and falling accents are also distinct in the Belgrade dialect: there is a significant effect of accent type (rising vs. falling) on `PeakOffset` ($\chi^2(1) = 225.07$, $p < 0.0001$). The difference between estimated means for the two accent types is 92.9 ms (SE = 5.0 ms), where rising accents occur later relative to the beginning of the word than falling accents. The long rising accent was in a phrase-final disyllabic word in this experiment, and as such, the peak is realized on the second mora of the first syllable, retracted from its typical position on the second syllable. On average, long falling peaks occur 138.9 ms (SD = 33.5 ms) before the start of the second syllable, while long rising peaks occur 64.7 ms (SD

Table 3.13: Comparison of linear mixed effects models for PeakOffset (Belgrade long rising accent).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-550.0	90.33	1	< 0.0001**
OnsDur + (1|Part)	-629.7	169.94	1	< 0.0001**

[†] As compared to the null model, PeakOffset ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p[2]
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	0.25	1	0.62
OnsDur + Complexity + Complexity:OnsDur + (1|Part)	0.15	1	0.70

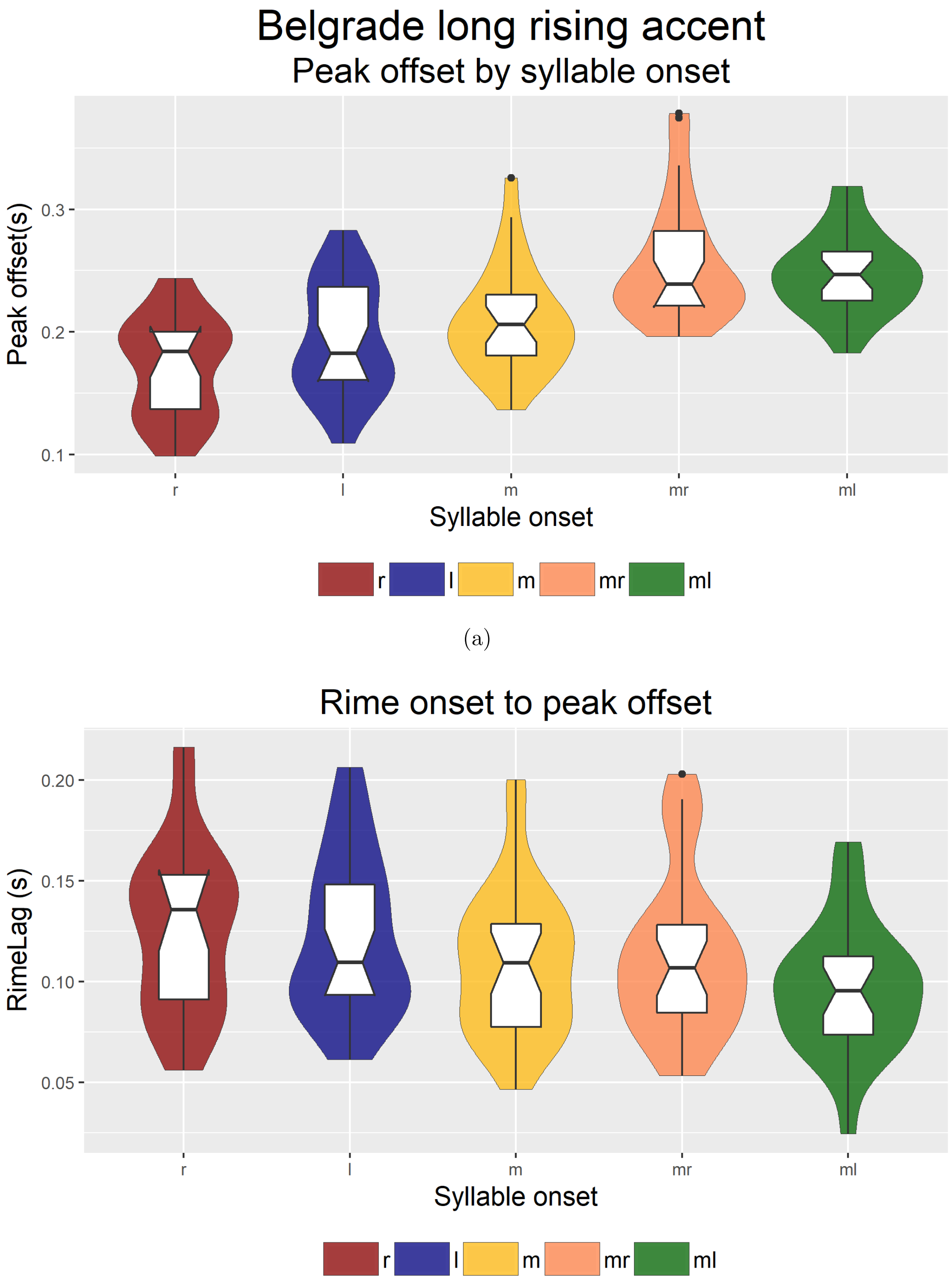
[2] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

= 20.4 ms) before the start of the second syllable. Thus, Hypothesis B is upheld for rising vs. falling accents of both lengths in the Belgrade dialect.[13]

A comparison of linear mixed effects models indicates that both Complexity ($\chi^2(1) = 90.33$, $p < 0.0001$; see Table 3.13a) and OnsDur ($\chi^2(1) = 169.94$, $p < 0.0001$) significantly improve the fit of the model as single fixed effects. The AIC values of each model indicate that OnsDur provides a better fit than Complexity. Unlike for the falling accents, adding Complexity to a model that already has OnsDur does not significantly improve the fit, either as a second fixed effect ($\chi^2(1) = 0.25$, $p = 0.62$) or in the interaction term Complexity:OnsDur ($\chi^2(1) = 0.15$, $p = 0.70$; see Table 3.13b). The timing of the peak offsets in long rising accents is illustrated in Figure 3.17a.

The lack of interaction indicates that the relationship between syllable onset duration and peak offset in long rising accents is similar to the relationship demonstrated by the

[13] Recall that the length distinction within rising or within falling accents is based on the phonological length of the stressed vowel; refer back to Section ?? for a description of the realization of the length distinction in Serbian.



(a)

(b)

Figure 3.17: Violin plots showing the location of the peak offset relative to the left edge of the word (a) and the beginning of the nucleus (b) in Belgrade long rising accents.

short rising accent: rather than being directly coordinated or aligned with the first mora (syllable onset included), the H in long rising accents is coordinated with the second mora. The relationship between `PeakOffset` and `OnsDur` is thus not direct, but rather due to the relationship between `OnsDur` and the beginning of the second mora, which is delayed by increases in syllable onset duration. However, there is also an effect of `OnsDur` on `NucLag` ($\chi^2(1)$ = 22.93, p < 0.0001; timing illustrated in Figure 3.17b): longer syllable onsets have a shorter interval between the beginning of the nucleus and the peak offset (β = -211.7 ms, SE = 42.2 ms). This effect moves in the same direction as in falling accents (i.e., for falling accents, peak offsets in words with complex onsets are closer to the beginning of the nucleus); however, there is not a clear step between simple and complex onsets. As there is no way in this data to locate the beginning of the second mora, it is impossible to determine if the H is in fact timed to the second mora, rather than the first syllable, and is a question left for future work.

3.2.4 Pitch characteristics: Valjevo

3.2.4.1 Falling accents

The Valjevo dialect proved additionally problematic, as the pitch range was smaller overall and also had more plateaus that covered the whole utterance. For analysis of H achievement (`PeakOffset` and `NucLag`), there were 351 falling tokens from Valjevo (out of a possible 360 tokens—2.5% removed). For analysis of the excursion characteristics (`ExcurStart` and `ExcurDur`), there were 200 tokens (44% removed from total). Details of these tokens is provided in Table 3.14 .

H achievement (`PeakOffset` and `NucLag`) In this section, I address Hypothesis 3.

> **Hypothesis 3.0** (null hypothesis): H targets are not anchored to any point in tone-bearing unit.
>
> **Prediction 3.0**:

Table 3.14: Number of falling accent tokens for each syllable onset with clear F0 peaks (H achievement) and clear minima (excursion characteristics) for the Valjevo dialect.

Onset	**H achievement**			**Excursion char.**		
	ȁmor	ȁmora	âve	ȁmor	ȁmora	âve
ml	24	24	23	16	15	11
mr	26	24	23	17	18	13
m	21	24	23	15	15	11
l	24	23	21	12	10	10
r	24	23	24	17	13	7

- There is no effect of syllable onset on `PeakOffset`;
- `NucLag` will be increasingly negative with increased `OnsDur`.

Hypothesis 3.1 (segmental anchoring): H targets are acoustically anchored to some point in the nucleus.

Prediction 3.1:

- `OnsDur` positively correlates with `PeakOffset`; the changes in magnitude in `PeakOffset` parallel differences in `OnsDur`;
- `OnsDur` has no effect on `NucLag`.

Hypothesis 3.2 (c-center): H targets are not articulatorily anchored, but are affected by the number of gestures in the tone-bearing unit onset.

Prediction 3.2:

- There is a significant effect of `Complexity` on `PeakOffset`—complex onsets will have later peak offsets;
- There is a significant effect of `Complexity` on `NucLag`.

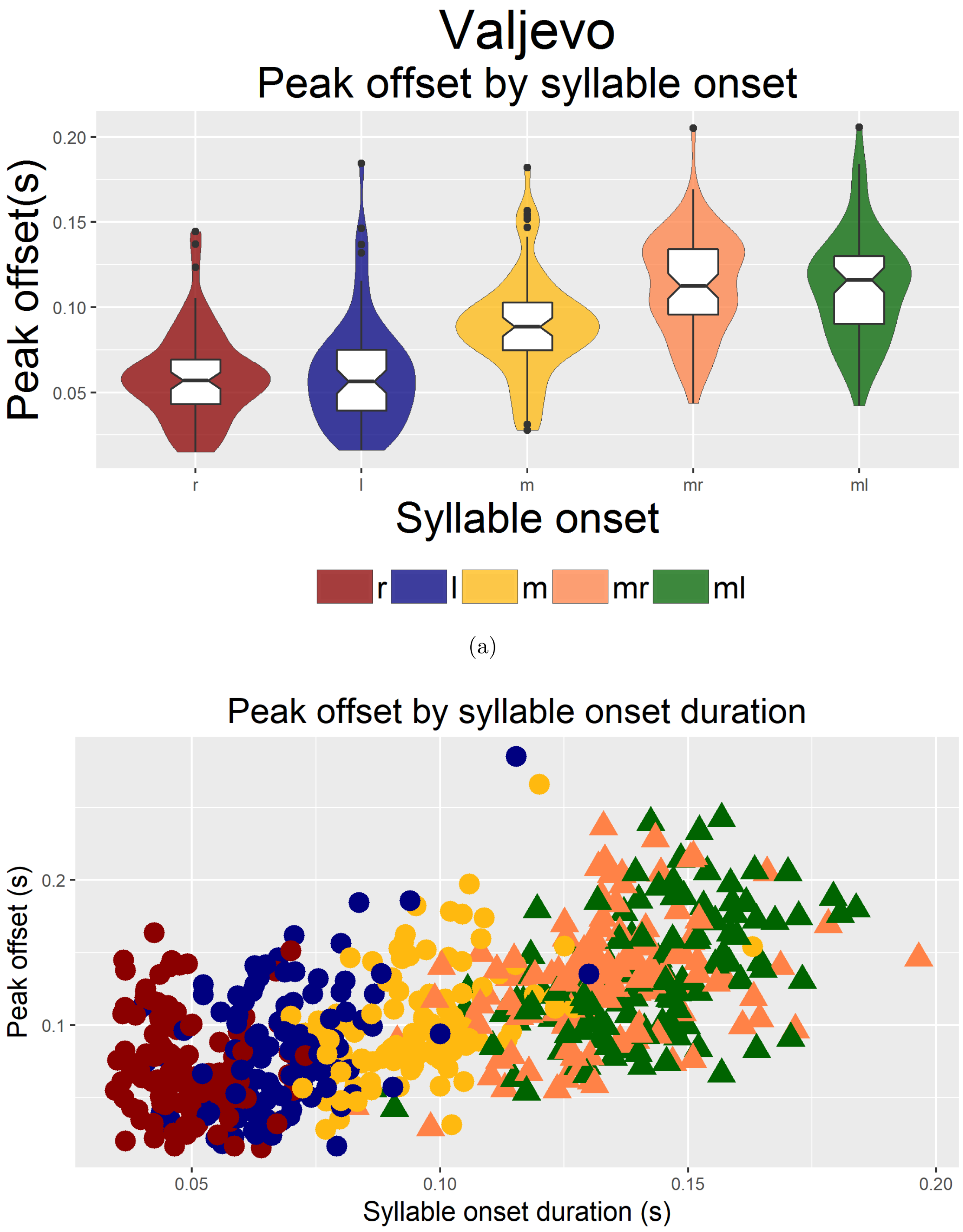
Hypothesis 3.3 (articulatory anchoring): H targets are anchored to some point in the nucleus, but precise timing also depends on the number of gestures in the tone-bearing unit onset.

Prediction 3.3: There is an effect of both OnsDur and Complexity on PeakOffset and NucLag.

There is also an effect of syllable onset on H achievement in the Valjevo dialect. Complexity as a single fixed effect improves the fit over the null model ($\chi^2(1) = 143.6$, $p < 0.0001$, AIC = -1451.5), and Identity as a single fixed effect also significantly improves the fit of the model ($\chi^2(4) = 188.24$, $p < 0.0001$ as compared to the null model). The AICs of these two models suggest that Identity (AIC = -1490.2) provides a better fit than Complexity. In the Valjevo dialect, there is no significant difference in peak offset timing between either /r/ and /l/ onsets (p = 0.998 using a least squares mean Tukey test) or /mr/ and /ml/ onsets (p = 0.997); all other comparisons, however, are significant at $p < 0.0001$ (see Figure 3.18a). This is somewhat reflective of the durations of syllable onsets in Valjevo Serbian; recall that /mr/ and /ml/ are never distinct, and the distributions for /r/ and /l/ are not distinct for some word groups.

Despite the collapse of both /r/ with /l/ and /mr/ with /ml/, the patterning of /m/ indicates that pitch timing is not determined by complexity. OnsDur predicts peak offset location better than either categorical predictor ($\chi^2(1) = 241.91$, $p < 0.0001$ as compared to the null model; AIC = -1549.8; see Table 3.15a for a summary of the single predictor mixed models). As syllable onset duration increases, the peak offset occurs later in the word: for every increase of 1,000 ms in syllable onset time, the peak offset is delayed by 689.4 ms (SE = 36.8 ms). These patterns are illustrated in Figure 3.11b.

Unlike in the Belgrade dialect, the addition of Complexity to a model that already has OnsDur does not significantly improve the fit of the model ($\chi^2(1) = 2.94$, p = 0.09). The interaction term Complexity:OnsDur also does not provide a better fit ($\chi^2(1) = 0.10$, p = 0.75), indicating that increases in syllable onset duration affect peak timing equally for simple onsets as for complex onsets. Similarly, neither Identity ($\chi^2(4) = 8.61$, p = 0.07)



(a)

(b)

Figure 3.18: A boxplot (a) and scatter plot (b) showing the relationship between (intrinsic) syllable onset duration and F0 offset location in the Valjevo dialect. Red is /r/, blue /l/, yellow /m/, orange /mr/, green /ml/.

Table 3.15: Comparison of linear mixed effects models for PeakOffset (Valjevo dialect).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-1451.5	143.60	1	< 0.0001**
Identity + (1|Part)	-1490.2	188.24	4	< 0.0001**
OnsDur + (1|Part)	-1549.8	241.91	1	< 0.0001**

[†] As compared to the null model, PeakOffset ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p[†]
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	2.94	1	0.09
OnsDur + Complexity + Complexity:OnsDur + (1|Part)	0.10	1	0.75
OnsDur + (1|Part)	—	—	—
OnsDur + Identity + (1|Part)	8.61	4	0.07
OnsDur + Identity + Identity:Duration + (1|Part)	2.64	4	0.62

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

nor the interaction Identity:OnsDur ($\chi^2(4) = 2.64$, p = 0.62) significantly improve the fit of the model.

In order to re-examine the effect of Complexity without the confound of an abnormal /r/, another series of linear models was compared, this time leaving out all /r/ tokens. In these models, the interaction Identity:Duration moves out of marginal improvement on the model (p = 0.05; previously p = 0.04), while the interaction Complexity:OnsDur moves from no improvement on the model to a marginal improvement (p = 0.01, previously p = 0.08). As in the Belgrade dialect, an increase in OnsDur has a larger effect in simple onsets ($\beta = 978.1$ ms, SE = 106.1 ms) than in complex onsets ($\beta = 848.5$ ms, SE = 50.9 ms).

Using this same reduced data set, there is also an effect of Complexity on NucLag (the time between peak offset and the beginning of the nucleus). As illustrated in Figure 3.19, F0

Table 3.16: Table of estimates and standard errors for each syllable onset in the Valjevo dialect, from the model `PeakOffset` $\sim$ `Duration + Identity:Duration + (1|Part)`.

	β	Std. Error
Intercept (`OnsDur` = 0 ms)	-2.7	12.3
`OnsDur`: /r/	1207.3	210.8
/l/	1003.2	84.3
/m/	972.6	113.1
/mr/	866.2	139.0
/ml/	835.6	140.0

Table 3.17: Comparison of linear mixed effects models for `PeakOffset`, excluding /r/ onsets (Valjevo dialect).

Model for `PeakOffset`	χ^2	DegF	$p^\dagger$
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Complexity:OnsDur + (1|Part)`	6.40	1	0.01°
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Identity:Duration + (1|Part)`	7.99	1	0.05
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

offsets occur earlier in words with complex onsets than in words with simple onsets: /ml/ = /mr/ (p = 0.80); /mr/ < /m/ (p = 0.01) and /ml/ < /m/ (p = 0.0005); /m/ = /l/ (p = 0.95).[14] Both `Complexity` ($\chi^2(1) = 34.48$, p < 0.0001) and `OnsDur` are significant predictors by themselves (see Table 3.18a), and the AIC does not suggest that either predictor provides a better fit.

A comparison of three models further shows that `Complexity` is the best predictor of `NucLag` (see Table 3.18b). When `Complexity` is added as a second main effect alongside `OnsDur`, there is marginal model improvement ($\chi^2(1) = 5.20$, p = 0.02); however, when `OnsDur` is added as a second main effect alongside `Complexity`, there is no improvement on

[14] In an ANOVA that includes /r/: /ml/ = /mr/ (p = 0.89); /mr/ < /m/ (p = 0.01) and /ml/ < /m/ (p = 0.0006); /m/ = /l/ (p = 0.98), /l/ = /r/ (p = 0.16), /m/ < /r/ (p = 0.04).

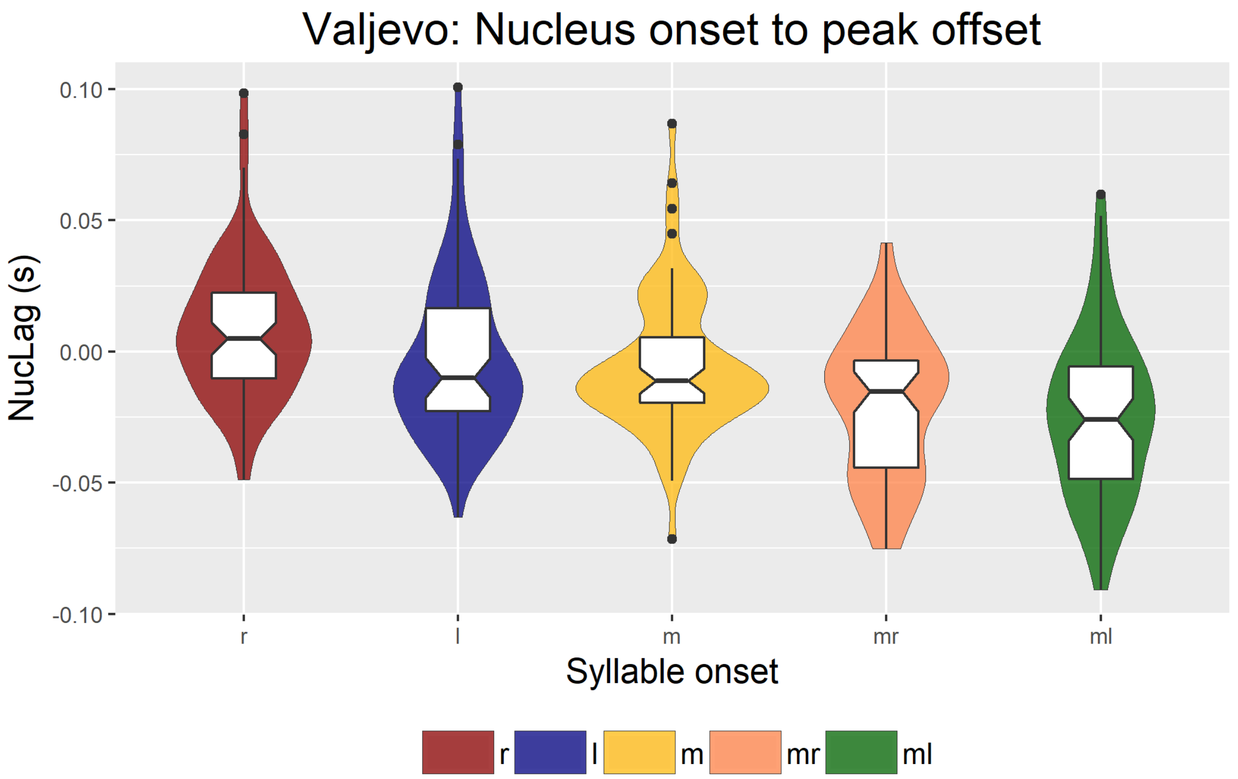


Figure 3.19: The time lag between the start of the nucleus and the F0 offset, by onset identity (Valjevo dialect). Here 0 represents the start of the nucleus.

the model ($\chi^2(1) = 0.94$, p = 0.33). Thus, `Complexity` addresses most of the variation in `NucLag`; as `OnsDur` is correlated with `Complexity`, it is not surprising that `OnsDur` is still significant as a single predictor.

Overall, Hypothesis 3.0 is rejected, largely in favor of Hypothesis 3.1: `OnsDur` predicts both `PeakOffset` and `NucLag`, with the caveat that the addition of `Complexity` to a model with `OnsDur` marginally improves the model fit for `NucLag`.

Excursion characteristics (`ExcurStart` and `ExcurDur`) In this section, I address Hypotheses 4 and 5. First, I investigate the patterns of the timing of the start of the H gesture.

Hypothesis 4.0 (null hypothesis): H gestures start at the same time as the

Table 3.18: Comparison of linear mixed effects models for NucLag, excluding /r/ onsets (Valjevo dialect).

(a) Single predictor models, compared to the null model.

Model for NucLag	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-1235.2	34.48	1	< 0.0001**
OnsDur + (1|Part)	-1230.9	30.22	1	< 0.0001**
[†]As compared to the null model, NucLag ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Nested model comparisons.

Model for NucLag	χ^2	DegF	p[†]
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	5.20	1	0.02°
Complexity + (1|Part)	—	—	—
Complexity + OnsDur + (1|Part)	0.94	1	0.33
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

word.

Prediction 4.0:

- There is no effect of syllable onset on ExcurStart;
- ExcurStart is 0.

Hypothesis 4.1 (segmental anchoring): The start of H gestures is anchored to some point in the tone-bearing unit.

Prediction 4.1:

- There is no effect of the syllable onset on ExcurStart;
- ExcurStart may not be 0.

Hypothesis 4.2 (c-center): H targets are coordinated as the second (in a simple onset) or third (in a complex onset) gesture with the syllable onset.

Prediction 4.2: There is a significant effect of `Complexity` (number of gestures) on `ExcurStart`: `ExcurStart` is greater for complex onsets than for simple onsets.

Hypothesis 4.3 (articulatory anchoring): H targets are articulatory anchored to some point in the nucleus, and that point is influenced both by the duration of the other gestures in the onset, as well as the number.

Prediction 4.3: There is a significant effect of both `Complexity` and `OnsDur` on `ExcurStart`.

In the Valjevo dialect, there is not a significant effect of syllable onset on the timing of the start of the pitch excursion. A comparison of linear mixed effects models shows that neither `Complexity` ($\chi^2(1) = 2.31$, p = 0.13) nor `OnsDur` ($\chi^2(1) = 3.03$, p = 0.08; see Table 3.19a) is a significant predictor of `ExcurStart`; pitch excursions start at the same time regardless of the duration of the syllable onset. Overall, pitch excursions start slightly before the beginning of the word (M = -12.6 ms, SD = 27.4 ms); this is true for all syllable onsets individually as well (/ml/ M = -6.5 ms, SD = 26.6 ms; /mr/ M = -11.2 ms, SD = 30.0 ms; /m/ M = -20.8 ms, SD = 25.0 ms; /l/ M = -10.1 ms, SD = 16.9 ms; /r/ M = -14.2 ms, SD = 33.2 ms). As the pitch excursions reliably begin before the acoustic beginning of the word, Hypothesis 4.0 is rejected in favor of Hypothesis 4.1.

Second, I address the duration of the H gesture:

Hypothesis 5.0 (null hypothesis): Tone gestures are ballistic in nature.

Prediction 5.0: There is no effect of syllable onset on `ExcurDur`.

Hypothesis 5.1 (segmental anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Table 3.19: Comparison of nested linear mixed effects models for `ExcurStart`, Valjevo dialect.

(a) Single predictor models, compared to the null model.

Model for `ExcurStart`	AIC	χ^2	DegF	p[†]
`OnsDur + (1|Part)`	-876.8	3.03	1	0.08
`Complexity + (1|Part)`	-876.1	2.31	1	0.13
[†]As compared to the null model, `ExcurStart` $\sim$ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

Prediction 5.1: There is a significant effect of `OnsDur` on `ExcurDur`; changes in `OnsDur` parallel changes in `ExcurDur`.

Hypothesis 5.2 (c-center): Tone gestures are ballistic in nature.

Prediction 5.2: There is no effect of syllable onset on `ExcurDur`.

Hypothesis 5.3 (articulatory anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Prediction 5.3: There is a significant effect of both `Complexity` and `OnsDur` on `ExcurDur`; increases in `Copmlexity` and `OnsDur` correlate with increases in `ExcurDur`.

The variation in `PeakOffset` instead comes from `ExcurDur`, which increases as syllable onset duration increases ($\chi^2(1) = 57.73$, $p < 0.0001$; see Table 3.20a). The two liquid onsets have the shortest excursion, and are not distinct from each other (/r/ M = 83.6 ms, SD = 32.0 ms; /l/ M = 81.2 ms, SD = 36.7 ms). The remaining three onsets are significantly longer than the liquid onsets, and not significantly different from each other (/m/ M = 111.0 ms, SD = 35.5 ms; /ml/ M = 119.0 ms, SD = 25.4 ms; /mr/ M = 122.4 ms, SD =

Table 3.20: Comparison of nested linear mixed effects models for ExcurDur, Valjevo dialect.

(a) Single predictor models, compared to the null model.

Model for ExcurDur	AIC	χ^2	DegF	p[†]
OnsDur + (1|Part)	-812.1	57.73	1	< 0.0001**
Complexity + (1|Part)	-788.4	33.99	1	< 0.0001**
[†]As compared to the null model, ExcurDur ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of nested models.

Model for ExcurDur	χ^2	DegF	p[†]
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	1.06	1	0.30
OnsDur + Complexity:OnsDur + (1|Part)	0.17	1	0.68
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

32.5 ms). There is no additional effect of Complexity ($\chi^2(1) = 1.06$, p = 0.30). Although the interaction Complexity:OnsDur was significant for PeakOffset (when excluding /r/), it is not significant for ExcurDur either when including /r/ ($\chi^2(1) = 0.17$, p = 0.68; see Table 3.20b) or when excluding /r/ ($\chi^2(1) = 2.50$, p = 0.11). Thus, in the Valjevo dialect, Hypothesis 5.0 is rejected in favor of Hypothesis 5.1; the start of the excursion does not move with added complexity, and all changes in peak location are attributable to the phonetic duration of the excursion.

The relationships between OnsDur, PeakOffset, ExcurDur, and ExcurStart are illustrated in Figure 3.20, as well as density plots of each measure. In contrast with the Belgrade correlations, the plot of excursion onset as predicted by syllable onset duration (first column, fourth row) is flat, indicating a uniform timing of the onset of the pitch excursion.

3.2.4.2 Rising accents

Again due to a tendency for plateaus, only 174 out of a possible 240 remained (66 removed; 27.5% attrition) for analysis of rising accent peak timing in the Valjevo dialect. Excessive variation in the timing of the start of the excursion also prevented analysis of excursion

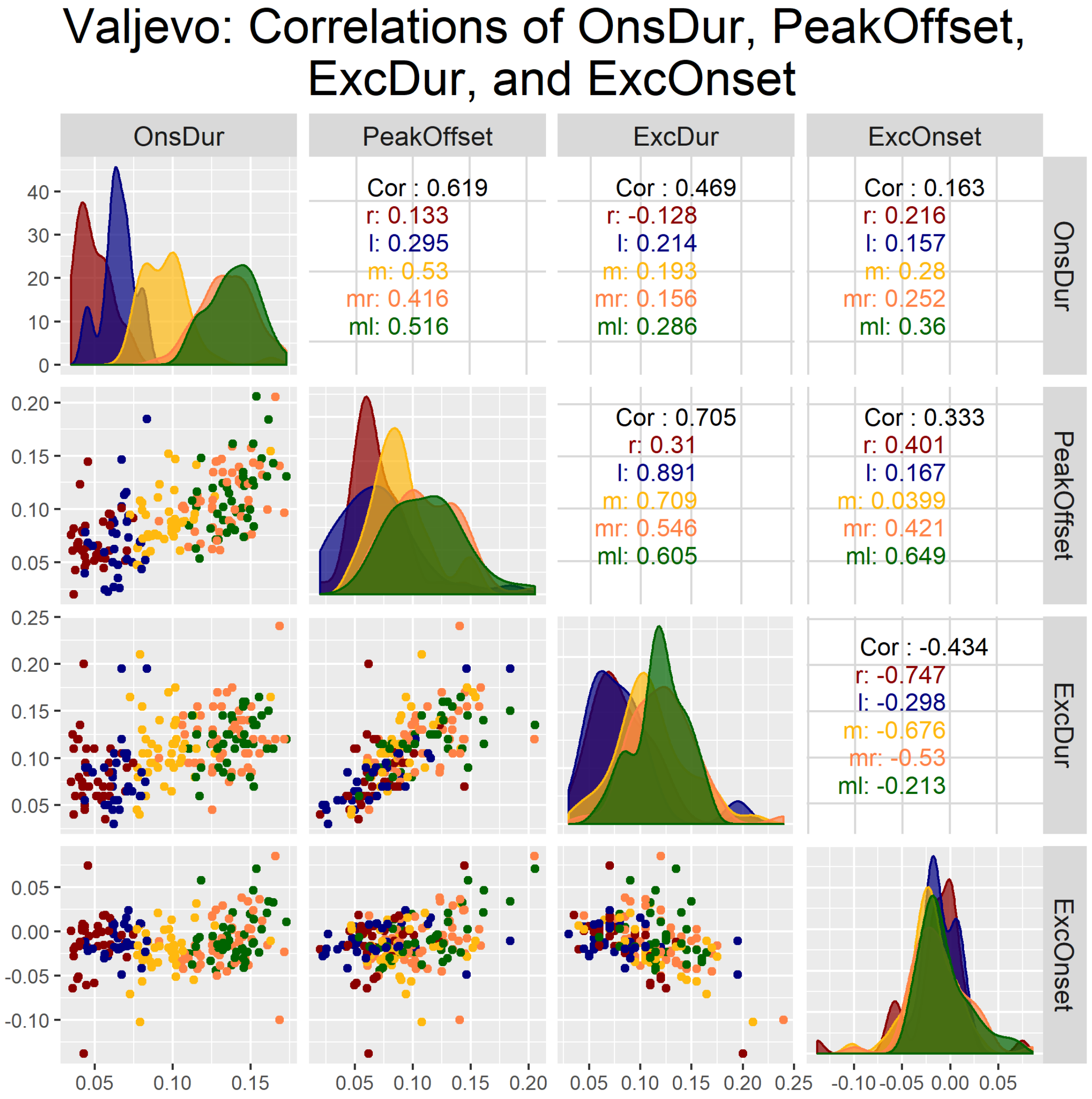


Figure 3.20: Correlation matrix for the Valjevo dialect that shows the relationships between syllable onset duration (`OnsDur`), peak offset location relative to the beginning of the word (`PeakOffset`), excursion duration (`ExcurDur`), and time lag between the start of the excursion and the beginning of the word (`ExcurStart`).

characteristics. Counts of each token for each syllable onset are provided in Table 3.21.

Table 3.21: Number of rising accent tokens for each syllable onset with clear F0 peaks for the Valjevo dialect.

Onset	**Valjevo**	
	àvinjak	ónu
ml	20	16
mr	20	14
m	19	17
l	21	16
r	14	17

Short rising Short rising and falling accents are also distinct in the Valjevo dialect: there is a significant effect of accent type (rising vs. falling) on `PeakOffset` ($\chi^2(1) = 330.50$, p < 0.0001). The difference between estimated means for the two accent types is 125.7 ms (SE = 5.3 ms), where rising accents occur later relative to the beginning of the word than falling accents. However, unlike in the Belgrade dialect, the peaks for both falling and rising accents occur before the second syllable, though there is quite a lot of variation in the rising accents. On average, short rising accent peaks occur 25.6 ms before the acoustic beginning of the second syllable (SD = 54.1 ms), while short falling accent peaks occur 145.7 ms (SD = 30.2 ms) before the beginning of the second syllable. Thus, Hypothesis B is upheld for short accents in the Valjevo dialect—there is still distinct timing for falling vs. rising accents, even though both peaks tend to occur during the stressed syllable.

Syllable onset has a significant effect on `PeakOffset`. `OnsDur` as a single fixed effect significantly improves the fit of the model ($\chi^2(1) = 14.95$, p = 0.0001; see Table 3.22a); as the syllable onset increases in duration, `PeakOffset` also increases ($\beta = 563.2$ ms, SE = 139.1 ms). `SylDur` also significantly improves the fit of the model ($\chi^2(1) = 20.65$, p $<$ 0.0001), though the AIC value (-279.8) does not robustly suggest that it provides a better fit than `OnsDur` (AIC = -274.1). However, the estimate for `SylDur` closer to a one-to-one

Table 3.22: Comparison of nested linear mixed effects models for PeakOffset, short rising accents (Valjevo dialect).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p†
OnsDur + (1|Part)	-274.1	14.95	1	0.0001**
SylDur + (1|Part)	-279.8	20.65	1	< 0.0001**

†As compared to the null model, PeakOffset ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p†
OnsDur + (1|Part)	—	—	—
OnsDur + SylDur + (1|Part)	5.97	1	0.01°
SylDur + (1|Part)	—	—	—
SylDur + OnsDur + (1|Part)	0.27	1	0.60

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

relationship: for every 1,000 ms increase in syllable duration, there is a 717.5 ms delay in PeakOffset (SE = 123.0 ms). Thus, as in the Belgrade dialect, this indicates that the effect of OnsDur is due to longer syllable onsets on the first syllable delaying the beginning of the second syllable.

This is supported by comparing models with both OnsDur and SylDur. When SylDur is added as a second fixed effect to a model that already has OnsDur, there is a marginal improvement in model fit ($\chi^2(1) = 5.97$, p = 0.01; see Table 3.22b). In contrast, in the opposite order, there is no significant improvement on the model ($\chi^2(1) = 0.27$, p = 0.60). Unlike in the Belgrade dialect, however, OnsDur2 does not significantly improve the model for PeakOffset2 ($\chi^2(1) = 0.29$, p = 0.59).

Long rising Long rising and falling accents are also distinct in the Valjevo dialect: there is a significant effect of accent type (rising vs. falling) on PeakOffset ($\chi^2(1) = 173.18$, p < 0.0001). The difference between estimated means for the two accent types is 123.8 ms (SE

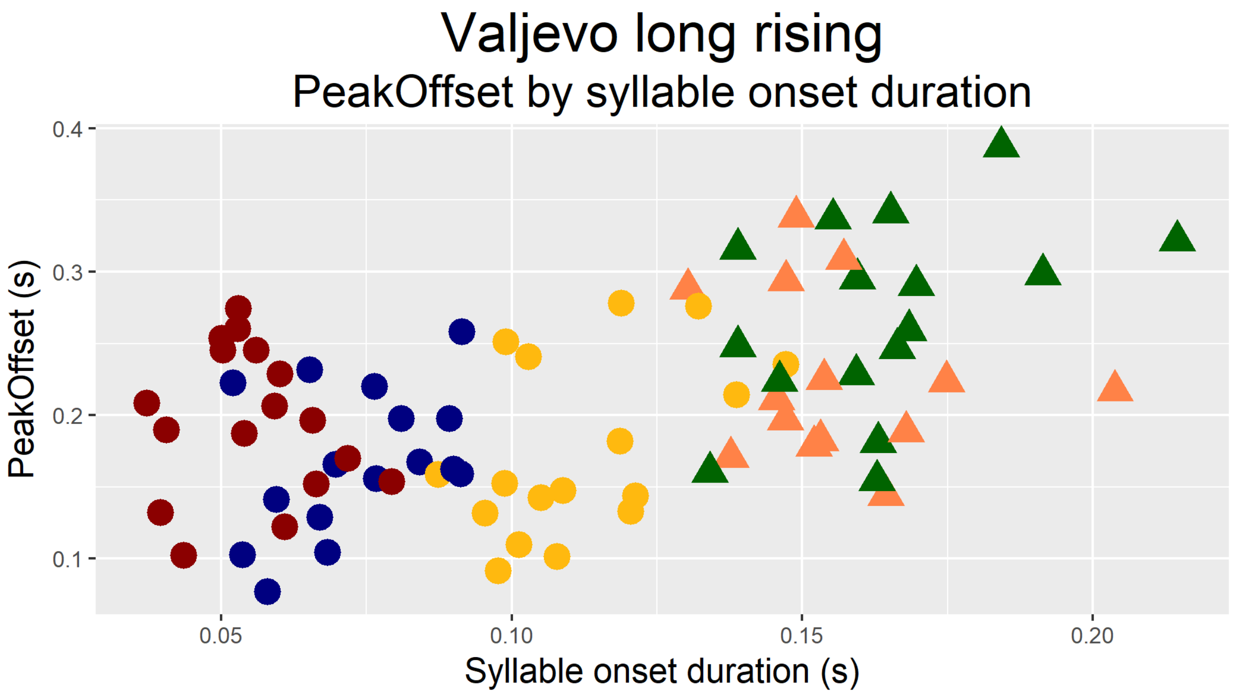


Figure 3.21: A scatter plot of peak offset times for long rising accents, Valjevo dialect.

= 7.4 ms), where rising accents occur later relative to the beginning of the word than falling accents. On average, long falling peaks occur 238.8 ms (SD = 40.1 ms) before the start of the second syllable, while long rising peaks occur 139.2 ms (SD = 50.8 ms) before the start of the second syllable. Thus, Hypothesis B is upheld for both sets of rising vs. falling accents in Valjevo. Furthermore, Hypothesis B is upheld in its entirety—both Belgrade and Valjevo exhibit a clear distinction between rising and falling accents.

There is an effect of syllable onset on `PeakOffset` on long rising accents; however, unlike in the Belgrade dialect, phonological complexity seems to be the determining factor, rather than phonetic duration. Both `OnsDur` ($\chi^2(1) = 19.19$, $p < 0.0001$) and `Complexity` ($\chi^2(1) = 24.85$, $p < 0.0001$; see Table 3.23a) significantly improve the fit of the model as compared to the null model. The AIC values of each model suggest that `Complexity` as a single fixed

Table 3.23: Comparison of linear mixed effects models for PeakOffset (Valjevo long rising accent).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-234.9	24.85	1	< 0.0001**
OnsDur + (1|Part)	-219.7	19.19	1	< 0.0001**
[†]As compared to the null model, PeakOffset $\sim$ 1 + (1|Part)	° < 0.05, * < 0.01, ** < 0.001			

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p[2]
Complexity + (1|Part)	—	—	—
Complexity + OnsDur + (1|Part)	0.14	1	0.71
OnsDur + (1|Part)	—	—	—
OnsDur + Complexity + (1|Part)	5.80	1	0.02°
OnsDur + Complexity + Complexity:OnsDur + (1|Part)	1.09	1	0.30
[2]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

effect provides a better model than OnsDur.

Comparing models with both Complexity and OnsDur further suggests that Complexity provides the most predictive power for Valjevo long rising accents. The addition of OnsDur to a model that already has Complexity does not significantly improve the model ($\chi^2(1) = 0.14$, p = 0.71; see Table 3.23b), but the addition of Complexity to a model that already has OnsDur marginally improves the fit of the model ($\chi^2(1) = 5.80$, p = 0.02). These patterns are illustrated in Figure 3.21.

3.2.5 Pitch characteristics: Dialect comparison

3.2.5.1 Falling accents

H achievement (PeakOffset and NucLag) When considered separately, the behavior of peak offset timing in the Belgrade and Valjevo dialects is similar. In both dialects, the peak occurs later relative to the acoustic beginning of the word if the syllable onset is longer. This is true even within simple onsets: /r/ has the shortest intrinsic duration, and peaks correspondingly occur the earliest in words with /r/ onsets. However, the relationship between peak location and syllable onset duration is not entirely straightforward in either dialect; rather, there is an interaction between syllable onset duration and complexity, where increases in syllable onset duration have a smaller effect on peak offset delay when the syllable onset is complex (i.e., /mr, ml/) than when the syllable onset is simple (i.e., /r, l, m/).

As expected, both `Complexity` ($\chi^2(1) = 282.5$, $p < 0.0001$) and `OnsDur` ($\chi^2(1) = 532.99$, $p < 0.0001$) also significantly improve the fit as compared to the null model (see Table 3.24a). `Dialect` is also a significant predictor of `PeakOffset` when considering the entire set of falling accents ($\chi^2(1) = 15.8$, $p < 0.0001$; see Table 3.24a for single factor model comparisons). As has been found in previous literature (Zec & Zsiga 2016), peaks occur earlier in the Valjevo dialect than in the Belgrade dialect.

As there were no major differences found in syllable onset duration between dialects (or phonological complexity, of course, as it was a directly manipulated categorical variable),[15] I took the model `PeakOffset` $\sim$ `Dialect + (1|Part)` as a starting point for the nested model comparisons. Both `OnsDur` and `Complexity` significantly improve the model when added as a second fixed main effect ($\chi^2(1) = 534.35$, $p < 0.0001$ and $\chi^2(1) = 282.20$, $p < 0.0001$, respectively); however, `Dialect` does not interact with either `Complexity` ($\chi^2(1) = 0.02$, $p = 0.90$) or `OnsDur` ($\chi^2(1) = 1.09$, $p = 0.30$; see Table 3.24b). Thus, increases in syllable onset duration and changes in phonological complexity have the same effect in both dialects. The relationship between the Belgrade and Valjevo dialects is illustrated in Figure 3.22. In this figure, the fit lines for the two dialects are roughly parallel, and the Valjevo line appears as simply shifted down from the Belgrade line.

Finally, as with the by-dialect analyses, the interaction `Duration:Complexity` significantly improves the fit of the model ($\chi^2(1) = 12.90$, $p = 0.0003$);[16] also as with the by-

[15] Refer back to **??** for more details.

[16] As previously discussed, `Complexity` as a main effect alongside `OnsDur` as a main effect is not being considered, as the two are strongly correlated.

Table 3.24: Comparison of nested linear mixed effects models for PeakOffset, both dialects (and all onsets) included.

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p†
Dialect + (1|Part)	-2851.6	15.80	1	< 0.0001**
Complexity + (1|Part)	-3118.3	282.50	1	< 0.0001**
OnsDur + (1|Part)	-3368.8	532.99	1	< 0.0001**

†As compared to the null model, NucLag $\sim$ 1 + (1|Part)

° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons, two-way interations with dialect.

Model for PeakOffset	χ^2	DegF	p†
Dialect + (1|Part)	—	—	—
Dialect + Complexity + (1|Part)	282.20	1	< 0.0001**
Dialect + Complexity + Dialect:Complexity + (1|Part)	0.02	1	0.90
Dialect + (1|Part)	—	—	—
Dialect + OnsDur + (1|Part)	534.35	1	< 0.0001**
Dialect + OnsDur + Dialect:OnsDur + (1|Part)	1.09	1	0.30
Dialect + OnsDur + (1|Part)	—	—	
Dialect + OnsDur + OnsDur:Complexity + (1|Part)	12.90	1	0.0003**
Dialect + OnsDur + OnsDur:Complexity + OnsDur:Complexity:Dialect + (1|Part)	1.33	2	0.51

†As compared to model immediately above

° < 0.05, * < 0.01, ** < 0.001

dialect analyses, increases in syllable onset duration have a lesser effect on peak offset delay for complex onsets than for simple onsets. This interaction is also the same across dialects: the interaction Duration:Complexity:Dialect does not significantly improve the model ($\chi^2(2) = 1.33$, $p = 0.51$, see Table 3.24b). All of these results hold when /r/ is entirely eliminated from analysis (see Table 3.25 for the full summary).

The patterns for NucLag are similar and follow from the patterns observed in PeakOffset. First considered are the analyses including all syllable onsets. As expected, Dialect is a significant predictor of NucLag ($\chi^2(1) = 17.00$, $p < 0.0001$), where peak offsets in the Valjevo

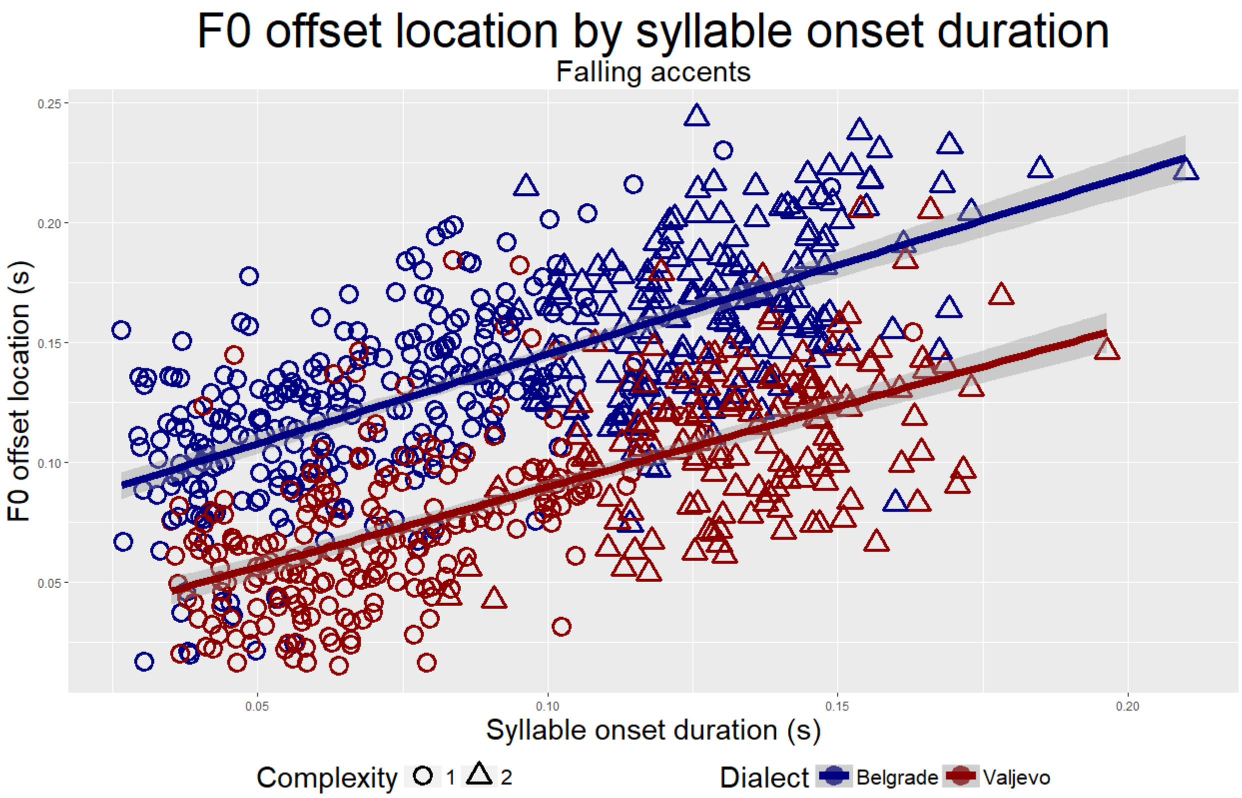


Figure 3.22: A scatter plot with fit lines comparing the Belgrade and Valjevo dialects with syllable onset duration on the x axis and F0 offset location on the y axis. Circles are simple onsets; triangles are complex onsets. Blue is Belgrade, red is Valjevo.

dialect occur near or before the beginning of the nucleus, while in the Belgrade dialect they occur well after the beginning of the nucleus. `Complexity` and `OnsDur` are also significant predictors as single factors, compared to the null model ($\chi^2(1) = 111.04$, $p < 0.0001$ and $\chi^2(1) = 109.49$, $p < 0.0001$, respectively); the AIC does not indicate that either of these models provide a better fit than the other (see Table 3.26a).

Further model comparisons were performed once again taking a model with `Dialect` included as a basic fixed effect. `OnsDur` as a second fixed effect significantly improves model fit ($\chi^2(1) = 109.65$, $p < 0.0001$), as does `Complexity` ($\chi^2(1) = 111.06$, $p < 0.0001$, see Table 3.26b). As in the `PeakOffset` analyses, there is no interaction between `Dialect` and

Table 3.25: Comparison of nested linear mixed effects models for `PeakOffset`, both dialects included, /r/ excluded from both dialects.

(a) Single predictor models, compared to the null model.

Model for `PeakOffset`	AIC	χ^2	DegF	p[†]
`Dialect + (1|Part)`	-2381.2	15.38	1	< 0.0001**
`Complexity + (1|Part)`	-2539.6	173.77	1	< 0.0001**
`OnsDur + (1|Part)`	-2720.9	355.09	1	< 0.0001**
[†]As compared to the null model, `PeakOffset` $\sim$ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b) Comparison of nested models.

Model for `PeakOffset`	χ^2	DegF	p[†]
`Dialect + (1|Part)`	—	—	—
`Dialect + Complexity + (1|Part)`	173.45	1	< 0.0001**
`Dialect + Complexity + Dialect:Complexity + (1|Part)`	1.66	1	0.20
`Dialect + (1|Part)`	—	—	—
`Dialect + OnsDur + (1|Part)`	356.26	1	< 0.0001**
`Dialect + OnsDur + Dialect:OnsDur + (1|Part)`	0.05	1	0.83
`Dialect + OnsDur + (1|Part)`	—	—	
`Dialect + OnsDur + OnsDur:Complexity + (1|Part)`	19.49	1	< 0.0001**
`Dialect + OnsDur + OnsDur:Complexity + OnsDur:Complexity:Dialect + (1|Part)`	0.40	2	0.82
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

`OnsDur` ($\chi^2(1) = 1.09$, p = 0.30) or `Dialect` and `Complexity` ($\chi^2(1) = 0.42$, p = 0.52), again supporting the conclusion that increases in syllable onset duration and changes in complexity have the same effect on peak offset location in both dialects.

Interestingly, the addition of `OnsDur` as a fixed effect significantly improves model fit when `Complexity` is already included ($\chi^2(1) = 9.27$, p = 0.002); and, in turn, the addition of `Complexity` as a fixed effect significantly improves model fit when `OnsDur` is already included ($\chi^2(1) = 10.69$, p = 0.001). This was not true in the Belgrade-only analyses, nor in the Valjevo-only analysis that excluded /r/—in those analyses, `OnsDur` did not improve

Table 3.26: Comparison of linear mixed effects models for NucLag, both dialects (and all onsets) included.

(a) Single predictor models, compared to the null model.

Model for NucLag	AIC	χ^2	DegF	p†
Dialect + (1|Part)	-3276.3	17.00	1	< 0.0001**
Complexity + (1|Part)	-3370.3	111.04	1	< 0.0001**
OnsDur + (1|Part)	-3368.8	109.49	1	< 0.0001**
†As compared to the null model, NucLag ∼ 1 + (1|Part)		° < 0.05, * < 0.01, ** < 0.001		

(b) Nested model comparisons.

Model for NucLag	χ^2	DegF	p†
Dialect + (1|Part)	—	—	—
Dialect + OnsDur + (1|Part)	109.65	1	< 0.0001**
Dialect + OnsDur + Dialect:OnsDur + (1|Part)	1.09	1	0.30
Dialect + (1|Part)	—	—	—
Dialect + Complexity (1|Part)	111.06	1	< 0.0001**
Dialect + Complexity + Dialect:Complexity + (1|Part)	0.42	1	0.52
Dialect + Complexity + (1|Part)	—	—	—
Dialect + Complexity + OnsDur + (1|Part)	9.27	1	0.002*
Dialect + Complexity + OnsDur + Complexity:OnsDur + (1|Part)	2.42	1	0.12
Dialect + Duration + (1|Part)	—	—	—
Dialect + OnsDur + Complexity + (1|Part)	10.69	1	0.001*
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

on a model that already included Complexity.

In analyses that exclude /r/ entirely, the results are much more similar to previous findings. Once again, Dialect, Complexity, and OnsDur are significant predictors of NucLag when considered as single factors compared to the null model (see Table 3.27); there are also no significant interactions between Dialect and OnsDur or Complexity. In contrast with the fully inclusive models, OnsDur does not significantly improve the fit of a model when Complexity is already included ($\chi^2(1) = 1.62$, p = 0.20), but adding Complexity to a model

Table 3.27: Comparison of nested linear mixed effects models for `NucLag`, both dialects included, /r/ excluded from both dialects.

(a) Single predictor models, compared to the null model.

Model for `NucLag`	AIC	χ^2	DegF	p[†]
`Dialect + (1|Part)`	-2673.1	16.27	1	< 0.0001**
`Complexity + (1|Part)`	-2734.1	77.29	1	< 0.0001**
`OnsDur + (1|Part)`	-2720.9	64.09	1	< 0.0001**
[†]As compared to the null model, `NucLag` $\sim$ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

(b) Nested model comparisons.

Model for `NucLag`	χ^2	DegF	p[†]
`Dialect + (1|Part)`	—	—	—
`Dialect + OnsDur + (1|Part)`	64.37	1	< 0.0001**
`Dialect + OnsDur + Dialect:Duration + (1|Part)`	0.05	1	0.83
`Dialect + (1|Part)`	—	—	—
`Dialect + Complexity (1|Part)`	77.312	1	< 0.0001**
`Dialect + Complexity + Dialect:Complexity + (1|Part)`	0.01	1	0.93
`Dialect + OnsDur + (1|Part)`	—	—	—
`Dialect + OnsDur + Complexity + (1|Part)`	14.56	1	0.0001**
`Dialect + Complexity + (1|Part)`	—	—	—
`Dialect + Complexity + OnsDur + (1|Part)`	1.62	1	0.20
`Dialect + Complexity + OnsDur + Complexity:OnsDur + (1|Part)`	6.43	1	0.01°
`Dialect + Complexity + OnsDur + Complexity:OnsDur + Complexity:OnsDur:Dialect + (1|Part)`	0.68	2	0.71
[†]As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

that includes `OnsDur` does significantly improve the model ($\chi^2(1) = 14.56$, p = 0.0001). This supports the conclusion that `NucLag` is largely determined by complexity, rather than idiosyncratic variation in syllable onset duration.

Excursion characteristics (ExcurStart and ExcurDur) As was established in Section **??**, increases in syllable onset duration affect the timing of the peak offset in the same way in both dialects. However, the strategy used by each dialect to accomplish the delay in peak offset is different: the Belgrade dialect employs a combination of delaying the onset of the pitch excursion and increasing the duration of the pitch excursion, while in the Valjevo dialect, the onset of the pitch excursion remains constant and the pitch excursion lengthens in parallel with the syllable onset duration. However, Valjevo pitch accents are aligned earlier in the syllable than Belgrade pitch accents.

As demonstrated in the preceding two sections, a comparison of linear mixed effects models shows that OnsDur as a single fixed effect significantly improves the fit of the model when both dialects are considered together ($\chi^2(1) = 140.26$, $p < 0.0001$; see Table 3.28a). In contrast, Dialect as a single fixed effect does not ($\chi^2(1) = 2.05$, $p = 0.15$); Dialect also does not significantly improve the fit of the model when OnsDur is already included ($\chi^2(1) = 2.71$, $p = 0.10$; see Table 3.28b). However, there is a significant interaction between Dialect and OnsDur ($\chi^2(1) = 43.72$, $p < 0.0001$; see Table 3.28b). Given the results of the individual dialects, this interaction is expected. For the Belgrade dialect, OnsDur has an estimate of 416.4 ms (SE = 28.5 ms), while for Valjevo the estimate is 81.4 ms (SE = 49.7), which reflects the non-effect of OnsDur on ExcurStart in the Valjevo dialect.

Finally, we come to Hypothesis 6, which addresses the duration of the H gesture across dialects:

> **Hypothesis 6.0** (null hypothesis): Both dialects have H gestures with the same duration.
>
> **Prediction 6.0**: Dialect is not a significant predictor of ExcurDur.
>
> **Hypothesis 6.1**: Valjevo has shorter pitch excursions than Belgrade, in alignment with their earlier peaks.
>
> **Prediction 6.1**: Dialect is a significant predictor of ExcurDur; Valjevo dialects

Table 3.28: Comparison of nested linear mixed effects models for `ExcurStart`, both dialects included.

(a) Single predictor models, compared to the null model.

Model for `ExcurStart`	AIC	χ^2	DegF	p†
`Dialect + (1|Part)`	-2703.4	2.05	1	0.15
`OnsDur + (1|Part)`	-2841.6	140.26	1	< 0.0001**

†As compared to the null model, `ExcurStart` ∼ 1 + `(1|Part)`

° < 0.05, * < 0.01, ** < 0.001

(b) Comparison of nested models.

Model for `ExcurStart`	χ^2	DegF	p†
`OnsDur + (1|Part)`	—	—	—
`OnsDur + Dialect + (1|Part)`	2.71	1	0.10
`Dialect + OnsDur + Dialect:OnsDur + (1|Part)`	43.72	1	< 0.0001**

†As compared to model immediately above

° < 0.05, * < 0.01, ** < 0.001

have shorter pitch excursions.

`Dialect` does significantly improve the model for `ExcurDur` ($\chi^2(1) = 7.22$, p = 0.007): excursions are overall shorter in the Valjevo dialect than in the Belgrade dialect. Thus, Hypothesis 6.0 is rejected in favor of Hypothesis 6.1. `OnsDur` also significantly improves the fit of the model, both as a single fixed effect ($\chi^2(1) = 130.09$, p < 0.0001) and when added as a second main effect alongside `Dialect` ($\chi^2(1) = 131.21$, p < 0.0001; see Table 3.29). The interaction `Dialect:OnsDur` also significantly improves the model ($\chi^2(1) = 9.58$, p = 0.002; see Table 3.29b). Although excursion duration increases with increased syllable onset duration in both dialects, the effect is greater in the Valjevo dialect ($\beta = 487.3$ ms, SE = 62.2 ms) than in the Belgrade dialect ($\beta = 293.7$ ms, SE = 35.7 ms).

The relative excursion onset times and excursion durations are illustrated in Figure 3.23. In Figure 3.23a, the Belgrade dialect clearly shows an effect of syllable onset, with the mean excursion onset time of /r/ occurring before the beginning of the word and the two complex onsets occurring well after the beginning of the word, but the overall range of excursion onset times largely contains the small range of times for Valjevo. Although on average, the two

Table 3.29: Comparison of nested linear mixed effects models for `ExcurDur`, both dialects included.

(a) Single predictor models, compared to the null model.

Model for `ExcurDur`	AIC	χ^2	DegF	p†
`Dialect + (1|Part)`	-2476.8	7.22	1	0.007*
`OnsDur + (1|Part)`	-2599.7	130.09	1	< 0.0001**
†As compared to the null model, `ExcurDur` $\sim$ 1 + `(1|Part)`		° < 0.05, * < 0.01, ** < 0.001		

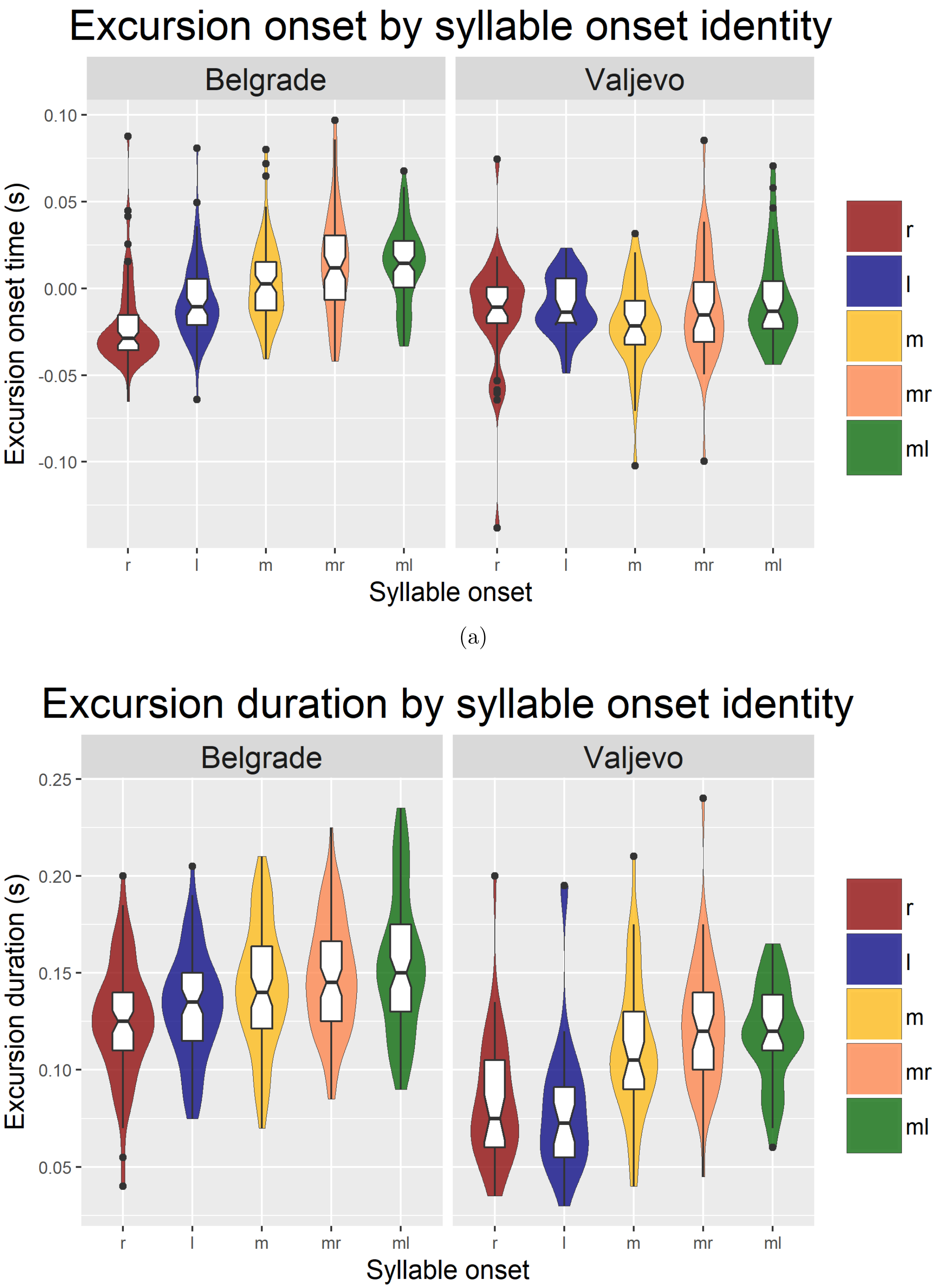
(b) Comparison of nested models.

Model for `ExcurDur`	χ^2	DegF	p†
`Dialect + (1|Part)`	—	—	—
`Dialect + OnsDur + (1|Part)`	131.21	1	< 0.0001**
`Dialect + OnsDur + Dialect:OnsDur + (1|Part)`	9.58	1	0.002*
†As compared to model immediately above	° < 0.05, * < 0.01, ** < 0.001		

dialects have the same excursion timing, the start of pitch excursions are affected by syllable onset in the Belgrade dialect, but not the Valjevo dialect. The two dialects also exhibit fairly distinct ranges for excursion duration, though there is quite a bit of overlap due to Valjevo exhibiting a larger effect of syllable onset duration. Finally, Valjevo excursions are shorter than Belgrade pitch excursions overall, but the effect of increased syllable onset duration is much greater in the Valjevo dialect than in the Belgrade dialect.

3.3 Discussion

This experiment addresses the question of what exactly a TBU is and what role it serves. For this study, I accepted the argument put forward by Zec and Zsiga (2016), who argued that the TBU in Serbian is the mora, based on the behavior of utterance-final rising accents. The data from this experiment shows that these are two dialects (or languages) with the same phonological system of contrast, but with very different phonetic realizations. Differences in timing exist both at the start and the target of the pitch excursions. This indicates that a



(a)

(b)

Figure 3.23: Box plots of excursion onset time and excursion duration for each dialect, by onset type.

TBU serves as the phonological unit that supplies information to tone realization, even if it does not necessarily bound its timing.

3.3.1 Summary

The data from this study upholds Hypothesis B: both Belgrade and Valjevo Serbian have a four-way system of accentual contrast, where rising accents occur later relative to the stressed syllable than falling accents. Overall, pitch accent peaks occur earlier in the Valjevo dialect than in the Belgrade dialect (confirming the findings reported in Zec and Zsiga 2016). The difference in alignment between the two dialects is especially pronounced for rising accents; in the Belgrade dialect, rising peaks reliably occur in the post-stressed syllable, but in the Valjevo dialect, rising peaks frequently occur in the stressed syllable. That is, in the Valjevo dialect, both falling and rising accents have peaks in the stressed syllable, and the contrast (phonetically) is one of timing alone.

First, the location of the H syllable in the carrier verb did not have an effect on peak offset timing or on the timing of the start of the H gesture in either dialect. Thus, the null hypotheses 1.0 and 2.0 are upheld.

✓✓**Hypothesis 1.0** (null hypothesis): There is no effect of carrier verb on the timing of the accentual peak.

✗✗**Hypothesis 1.1**: There is a significant effect of carrier verb on the timing of the accentual peak.

✓✓**Hypothesis 2.0** (null hypothesis): There is no effect of carrier verb on the timing of the start of the tone gesture.

✗✗**Hypothesis 2.1**: There is a significant effect of carrier verb on the timing of the start of the tone gesture.

Overall, neither the segmental anchoring hypothesis nor the c-center hypothesis for tone fully predict the data presented in this study. First, in the Belgrade dialect, there was a significant effect of both syllable onset duration and syllable onset complexity on the timing of the peak offset. Notably, complexity alone provided the best fit when considering peak offset timing relative to the beginning of the nucleus, with no additional information from syllable onset duration. This suggests that articulatory anchoring is a more likely model for Belgrade timing. H achievement in the Valjevo dialect, on the other hand, mostly exhibited effects of syllable onset duration, though with the caveat that the /r/ presented with some abnormal behavior. Thus, for both dialects, Hypothesis 3.0 is rejected, in favor of Hypothesis 3.3 for Belgrade and in favor of Hypothesis 3.1 for Valjevo.

> ✗✗**Hypothesis 3.0** (null hypothesis): H targets are not anchored to any point in tone-bearing unit.
>
> ✗✓**Hypothesis 3.1** (segmental anchoring): H targets are acoustically anchored to some point in the nucleus.
>
> ✗✗**Hypothesis 3.2** (c-center): H targets are not articulatorily anchored, but are affected by the number of gestures in the tone-bearing unit onset.
>
> ✓✗**Hypothesis 3.3** (articulatory anchoring): H targets are anchored to some point in the nucleus, but precise timing also depends on the number of gestures in the tone-bearing unit onset.

Despite similar behavior of the peak offset, the two dialects exhibited very different methods for achieving H target timing. In the Belgrade dialect, there was an effect of both syllable onset duration and syllable onset complexity on the timing of the start of the H gesture: pitch movements started later with longer syllable onsets. However, in the Valjevo dialect, there was no effect of syllable onset on the timing of the start of the H gesture: pitch movements started before the beginning of the word in all cases. Thus, Hypothesis 4.0 is rejected in favor of Hypothesis 4.3 in the Belgrade dialect, but in favor of Hypothesis 4.1 in the Valjevo dialect.

> ✗✗**Hypothesis 4.0** (null hypothesis): H gestures start at the same time as the

word.

✗✓**Hypothesis 4.1** (segmental anchoring): The start of H gestures is anchored to some point in the tone-bearing unit.

✗✗**Hypothesis 4.2** (c-center): H targets are coordinated as the second (in a simple onset) or third (in a complex onset) gesture with the syllable onset.

✓✗**Hypothesis 4.3** (articulatory anchoring): H targets are articulatory anchored to some point in the nucleus, and that point is influenced both by the duration of the other gestures in the onset, as well as the number.

There was also an effect of syllable onset duration of the upward pitch excursion in both dialects, thus rejecting the null Hypothesis 5.0 for both—in favor for either of the anchoring hypotheses. When looking at excursion duration alone, the two types of anchoring hypotheses do not make distinct predictions. However, while Belgrade peaks were timed using a combination of shifts in the start of the H gesture and duration of the H gesture, Valjevo solely utilized changes in gesture duration. In combination with the other hypotheses, Valjevo appears to use segmental anchoring, while Belgrade appears to use some sort of articulatory anchoring. However, it is only possible to positively rule out articulatory anchoring (without the c-center) for Valjevo with a careful articulatory study.

✗✗**Hypothesis 5.0** (null hypothesis): Tone gestures are ballistic in nature.

✓✓**Hypothesis 5.1** (segmental anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

✗✗**Hypothesis 5.2** (c-center): Tone gestures are ballistic in nature.

✓✓**Hypothesis 5.3** (articulatory anchoring): Tone gestures stretch with more segmental material in between the anchoring point for the start of the H gesture and the anchoring point for the end of the H gesture.

Finally, there was also a cross-dialect difference in duration of the H gesture: Belgrade pitch excursions are (on average) longer than Valjevo pitch excursions. Thus, the null Hypothesis 6.0 is rejected.

> ✗**Hypothesis 6.0** (null hypothesis): Both dialects have H gestures with the same duration.
>
> ✓**Hypothesis 6.1**: Valjevo has shorter pitch excursions than Belgrade, in alignment with their earlier peaks.

3.3.2 The contribution of syllable onsets

3.3.2.1 A note on /r/ in Valjevo

The individual patterning of /r/ in the Valjevo dialect is worth some discussion. One possible source of the differences is segmentation. As described previously, the segmentation of /r/ included the schwa-like interval before the closure of the tongue tip, but did not include any open interval after the closure. It is possible that another open interval should be included post-closure, which would increase the duration of the /r/ and decrease the duration of the nucleus. However, the abnormal patterning could not have been caused by the possibly rogue /r/ in the *ónu* word type, as *ónu* has a long rising accent and these analyses only include falling accents. It is also unlikely that the difference between /r/ and /ml/ for the *ȁmora* word type is generating the entire pattern, particularly because that same exception exists in the Belgrade dialect, but the patterns of peak timing are not affected.

Another possible source of the different patterning is that F0 peaks in the Valjevo dialect are quite close to the boundary between the syllable onset and nucleus. As shown in 3.19, the mean `NucLag` for all syllable onsets (except /r/) is *before* the onset of the nucleus. Unlike the other onsets, /r/ has a full closure; in the method of segmentation that I chose, this means that the last approximately 30 ms of the syllable onset is the closure. It is possible that this perturbed the timing of the peak and pushed it later, producing the abnormally high β estimates for `PeakOffset` $\sim$ `Identity:Duration` and `NucLag` $\sim$ `Identity:Duration`.

The closure-related perturbation seems the more likely source of the abnormal patterning, as the same segmentation method was used for Belgrade and Valjevo, but only Valjevo had an errant /r/. As Belgrade's F0 peaks occur well after the onset of the nucleus (grand mean 47.3 ms after the beginning of the nucleus, compared to Valjevo's 9.2 ms before), the closure of the /r/ is not in a position to perturb the timing.

3.3.2.2 Duration and complexity

Some questions still remain, which are complicated by the nature of simple and complex onsets. The data suggested that increases in duration did not have the same effect at all points in the scale—that is, increases in the duration of /m/ did not have the same magnitude of effect as increases in the duration of /l/, where /m/ overall is longer than /l/. Since complex onsets are naturally longer than simple onsets, it is possible that the interaction between syllable onset duration and complexity is simply due to complex onsets being on the long end of the duration scale—perhaps it is not complexity that causes this apparent category split, but some threshold of duration. However, the means of /l/ and /m/ are not separated by a drastically different amount than what separates /m/ and /mr/—e.g. in the Belgrade dialect, the mean durations for /l/ and /m/ are approximately 23 ms apart, and the mean durations for /m/ and /mr/ are approximately 35 ms apart. Even more convincing is that the distance between /r/ and /m/ is approximately 45 ms, and yet /r/ and /m/ pattern the same way with respect to distance between peak offset and nucleus beginning (and end).

Nevertheless, a worthy future study would examine several more syllable onsets, ranging from the shortest possible complex onsets (perhaps /vl/ or /vr/) to the longest possible complex onsets (perhaps /mn/), and from the shortest possible simple onsets (/r/) to the longest possible simple onset (perhaps /m/, or even a fricative). The goal of this study would be to get two onsets of differing complexity that in fact overlap in duration (i.e., the longest simple onset and the shortest complex onset) and compare the timing of the pitch peaks, as well as the effects of increased duration within phonological complexity categories.

3.3.2.3 Tone gestures and tone-bearing units

It should be noted that the failure to support the c-center as a possible coordinative structure for Belgrade is largely due to the strictness of this particular conceptualization: the c-center in its most strict form requires zero input from the duration of other gestures, as well as a ballistic tone gesture (i.e., a tone gesture that is fully determined by a stiffness parameter). Under a less strict interpretation of the c-center hypothesis, however, it is possible that the Belgrade dialect does use the c-center structure to coordinate tone. Neither the right nor the left edge of the syllable onset serves as an anchoring point for F0 excursions; rather, the left and right edges of the onset consonants spread bidirectionally away from the start of the pitch excursion. The results for the midpoint of the syllable onset show significant, but possibly not meaningful differences between complexity categories. This "center-like" alignment is reminiscent of the motivation for the name of the c-center effect—that is, the center, rather than the edges, of the consonant gestures (here, using acoustic landmarks as proxy for gestures) remains constantly timed to some reference landmark. However, in the original c-center, the anchoring gesture (that is, the gesture that the consonants were bidirectionally displaced from) is a vowel gesture, not a tone gesture (Browman & Goldstein 1988); moreover, the c-center literature on Mandarin and Thai tone argue that tone gestures behave as consonants, not vowels. Without an actual articulatory study, it is impossible to determine whether Belgrade uses the c-center structure for its pitch accent. Moreover, the comparison with Valjevo data indicates that the c-center is not the only possible coordinative regime for lexical tone.

Finally, there is the question of how gestures receive their timing. This is particularly interesting for the Valjevo dialect, where the duration of the H tone gesture varies directly with the duration of the syllable onset. For the H gesture, then, the proposal of "intrinsic timing" (related to gestural stiffness; Browman and Goldstein 1990) is not viable. Instead, the duration of the gesture is unspecified, and only receives its timing information when coordinated with some TBU.

One way to implement this in Articulatory Phonology is to allow for both the onsets and targets of gestures to be coordinated, rather than just the onsets. This would allow for gestural "stretching," as occurs in Valjevo Serbian, as well as in other pitch systems (English, for example; Pierrehumbert 1980); this resembles the articulatory anchoring hypothesis. However, it is unclear if target timing is necessary for any other type of gesture, and if not, why pitch gestures are special.

Another way to implement this would be to specify coordinative structures and pitch targets, which would require some amount of time to be accomplished. A qualitative assessment of the data suggests that Valjevo pitch excursions are smaller in both duration and in magnitude than Belgrade pitch excursions. It is unclear how these two aspects of the pitch excursion are related—specifically, it is unclear if there is a causal relationship between the two, and if so, in which direction it runs. In addition, in cleaning the data, multiple trials had to be eliminated due to a focus intonation on the carrier word obscuring the timing of falling accent peaks. in this type of focus, the pitch contour would rise from the carrier word until the very beginning of the target word, at which point the F0 would fall sharply.[17] This could be indicative of the target of an H gesture playing a large role in timing: as long as the peak is achieved, the particular timing of the peak relative to a point in the nucleus is less important.

[17] Interestingly, focus on the carrier word did not eliminate pitch contours in rising accents.

Chapter 4

Serbian: Focus on rising accents

In this chapter, I build on the study presented in Chapter 3 and focus on the realization of short rising accents in Belgrade and Valjevo Serbian. In Chapter 3, I showed that varying the syllable onset of the H syllable affects the timing of the peak; however, only the syllable onset of the stressed syllable was varied. As described in Chapter 1, in falling accents the stressed syllable is the same as the H syllable; however, in rising accents, the stressed syllable and the H syllable are distinct. Thus, the goal of this study is to determine the effects of the H syllable onset vs. the stressed syllable onset on peak timing in rising accents. Using the same set of syllable onsets as in Chapter 3 (/r, l, m, mr, ml/), I vary syllable onsets over three loci: Locus 1, a falling accent where stress and H are on the same syllable; Locus 2, a rising accent where only the H syllable onset is varied; and Locus 3, a rising accent where only the stressed syllable onset is varied.

The structure of this chapter is as follows: In Sections 4.1 and 4.2, I present the second acoustic study on Serbian, followed finally by a discussion of the results and conclusion of the chapter in Section 4.3.

4.1 Experiment design

4.1.1 Hypotheses

The purpose of this experiment is to probe the association of the H tone in rising accents in Belgrade and Valjevo Serbian. In order to do this, I examine the effects of syllable onset on the coordination and timing of the rising accent in two rising accent manipulations: first, when only the stressed syllable varies in onset complexity, and second, when only the syllable with the lexical H varies in onset complexity. These two situations are then compared to the effect of syllable onset on falling accents, where the lexical H is on the stressed syllable.

I take as a starting point the results from the study in Chapter 3, where both phonetic and phonological characteristics of the syllable onset affected H target timing in falling accents.

> **Hypothesis A**: Differences in syllable onsets affect peak timing for falling accents (Locus 1) similarly to the results from Chapter 3.

The hypotheses of interest for this study concern the phonological relationship between the H and the post-stress syllable in Serbian.

> **Hypothesis 1.0** (null hypothesis): The H of rising accents is not influenced by either the stressed syllable or the H syllable onset.
>
> **Prediction 1.0**: Peak timing will not be affected by variation in syllable onset in either Locus 2 or Locus 3 words.

> **Hypothesis 1.1**: H is associated to the stressed syllable in rising accents.
>
> **Prediction 1.1**: Peak timing for Locus 3 words (varying onset of stressed syllable) will pattern like peak timing for Locus 1 words (single stressed/H syllable).

> **Hypothesis 1.2**: H is associated to the post-stress syllable in rising accents.
>
> **Predictions 1.2**: Peak timing for Locus 2 words (varying onset of H syllable)

will pattern like peak timing for Locus 1 words (single stressed/H syllable).

4.1.2 Stimuli

4.1.2.1 Target words

The target words for Experiment 2 were very similar to those used in Experiment 1, but focus solely on short accents. 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****ȁmora* 'marble.GEN', ***mr****àvinjak* 'anthill', and *ò****ml****adinu* 'youth.ACC', where the syllable onsets in boldface are the varied onsets. The word *òmladinu* is an ideal word for an analysis of clusters, as the syllabification is unambiguous: first, the principle of maximal syllable onset would encourage a syllabification of /o.mla.di.nu/ rather than /om.la.di.nu/, as /ml/ clusters are permitted in Serbian; second, the inclusion of the root *mlâd* "young" is transparent, which further encourages a syllabification of /o.mla.di.nu/.

4.1.2.2 Carrier phrases

As no significant differences in peak timing were found in Experiment 1 between the two carrier phrases, Experiment 2 did not have a manipulation of distance between the pitch peaks of the carrier and the target words. 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?*[1] 'Do you want X?' and *Da li ste rekli X?* 'Did you say X?'. These two stimuli elicited a slightly different initial response (*Neću*! 'I don't want that!' vs. *Nisam*! 'I didn't [say that]!'), but the actual carrier sentence, *Daj mi Y* 'Give me Y', was the same for both stimulus contexts. This context ensured that focus would be put on the target word.

[1] The pronoun *Vi* '2SG.FORMAL' was chosen in order to be able to use the same stimulus for both men and women; with *ti* '2SG.FAMILIAR' gender agreement would be required in the 'Did you say...' sentences.

Table 4.1: Target words used in Experiment 2, organized by accent type, and indicating which syllable onset was manipulated.

Orthography	**Phonology**	**Gloss**
Falling: stress and pitch		
mlamora	/ˈmra$_H$mora/	*nonce*
mramora	/ˈmla$_H$mora/	‘marble.GEN.SG’
mamora	/ˈma$_H$mora/	*nonce*
lamora	/ˈla$_H$mora/	*nonce*
ramora	/ˈma$_H$mora/	*nonce*
Rising: just stress		
mlavinjak	/ˈmlavi$_H$njak/	*nonce*
mravinjak	/ˈmravi$_H$njak/	‘anthill’
mavinjak	/ˈmavi$_H$njak/	*nonce*
lavinjak	/ˈlavi$_H$njak/	*nonce*
ravinjak	/ˈmavi$_H$njak/	*nonce*
Rising: just pitch		
omladinu	/ˈomla$_H$dinu/	‘youth.ACC.SG’
omradinu	/ˈomra$_H$dinu/	*nonce*
omadinu	/ˈoma$_H$dinu/	*nonce*
oladinu	/ˈola$_H$dinu/	*nonce*
oradinu	/ˈora$_H$dinu/	*nonce*

4.1.2.3 Task

As in Experiment 1, participants first heard a context prompt (recorded in advance by a native speaker of Valjevo Serbian that has been living in Belgrade for 20 years) and then read their response. 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[2] replacement for the target word, and had the same accent and syllable number as the target words. The six possible questions and sample responses are presented in Figure 4.1 (where the target word was presented in upper case for the experiment as well, in order to encourage a focused reading; accents marked in the target and replacement words for clarity, but not marked in presentation to the participants):

[2] Insofar as a replacement for demanding youthful people or an anthill is semantically plausible.

Context: Da li želite dȑveta?
“Do you want [pieces of] wood?”
Response: Neću! Daj mi MRÄMORA, molim te.
“I don’t! Give me [pieces of] MARBLE, please.

Context: Jeste li rekli ‘dȑveta’?
“Did you say ‘[pieces of] wood’?”
Response: Nisam! Daj mi MRÄMORA, molim te.
“I didn’t! Give me [pieces of] MARBLE, please.

(a) Trisyllabic, short falling (varied onset of the stressed and H syllable)

Context: Da li želite pčèlinjak?
“Do you want a beehive?”
Response: Neću! Daj mi MRÀVINJAK, molim te.
“I don’t! Give me an ANTHILL, please.

Context: Jeste li rekli ‘pčèlinjak’?
“Did you say ‘a beehive’?”
Response: Nisam! Daj mi MRÀVINJAK, molim te.
“I didn’t! Give me an ANTHILL, please.

(b) Trisyllabic, short rising (varied onset of stressed syllable only)

Context: Da li želite òčevinu?
“Do you want the inheritance?”
Response: Neću! Daj mi ÒMLADINU, molim te.
“I don’t! Give me YOUNG PEOPLE, please.

Context: Jeste li rekli ‘òčevinu’?
“Did you say ‘inheritance’?”
Response: Nisam! Daj mi ÒMLADINU, molim te.
“I didn’t! Give me YOUNG PEOPLE, please.

(c) 4-syllabic, short rising (varied onset of lexical H syllable only)

Figure 4.1: Contexts and example responses.

Participants received a list of words that would be used in the study, which marked the accents and grouped them together to make clear what accents they had. Participants were allowed to reference this sheet through the study, though most did not have to. In order to prevent overlap, rushing, and list intonation, the written response only appeared on the screen after the context prompt ended. It was not possible to fully anticipate what the response was, as there are always five possibilities for one given prompt.

There were 15 target phrases total (**3** accent types x **5** 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 5 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 (if recorded by the graduate student) or a Sennheiser noise-canceling headset (if recorded by the professor).

4.1.3 Participants

Data for this experiment was collected in summer and fall 2017 at the Faculty of Philology at the University of Belgrade in Belgrade, Serbia. In this chapter I am presenting the data from 4 native speakers of Belgrade Serbian (ages 30-38; 1 male, 3 female) and 5 native speakers of Valjevo Serbian (ages 19 - 22; 1 male, 4 female). Three of the Belgrade speakers had previously participated in Experiment 1. Again, since living in Belgrade for an extended amount of time affects the realization of accent, the Valjevo speakers were all young university students that still had family ties in Valjevo and frequently visited home. The Belgrade speakers were all born and raised in Belgrade, though typically one or both parents were from elsewhere.

The experimenter did not personally collect this data. At the University of Belgrade, a graduate student in English phonetics that had participated in Experiment 1 ran the experiment on themselves and three other participants. The remainder of the data was collected by a professor of phonetics in the English department. All participants that were fluent speakers of English, and written consent was provided with a consent form in English.

4.1.4 Data labeling and analysis

4.1.4.1 Segmentation

As for Experiment 1, data was initially aligned with the Montreal Forced Aligner, and then corrected by hand in Praat. Only the boundaries of the carrier word were corrected; segments in the carrier word were not corrected (and these segment edges are not used as landmarks in the analysis); the segments of the target word were fully corrected. 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.

Marking segment boundaries proceeded as for Experiment 1. The boundary between the /i/ in *daj mi* and the initial /o/ of the *omladinu* set was marked at approximately the point of maximum velocity of F2, which proved reliable and consistent across speakers. The same landmarks were used for pitch landmarking as in Experiment 1. Again, as the absolute pitch peak is less stable and prone to small fluctuations, peak timing was compared using the H gesture release rather than the actual F0 peak; analyses that involve the start of upward F0 movement also reference the excursion onset, rather than the F0 valley.

4.1.4.2 Statistical analyses

Statistical analyses were performed using the same methodology as in Experiment 1. The variables (presented in `monospace font`) used in the analyses of the pitch excursions are the following:

Random effects

- `Part` (participant): Random intercepts for participant are included in all linear models.

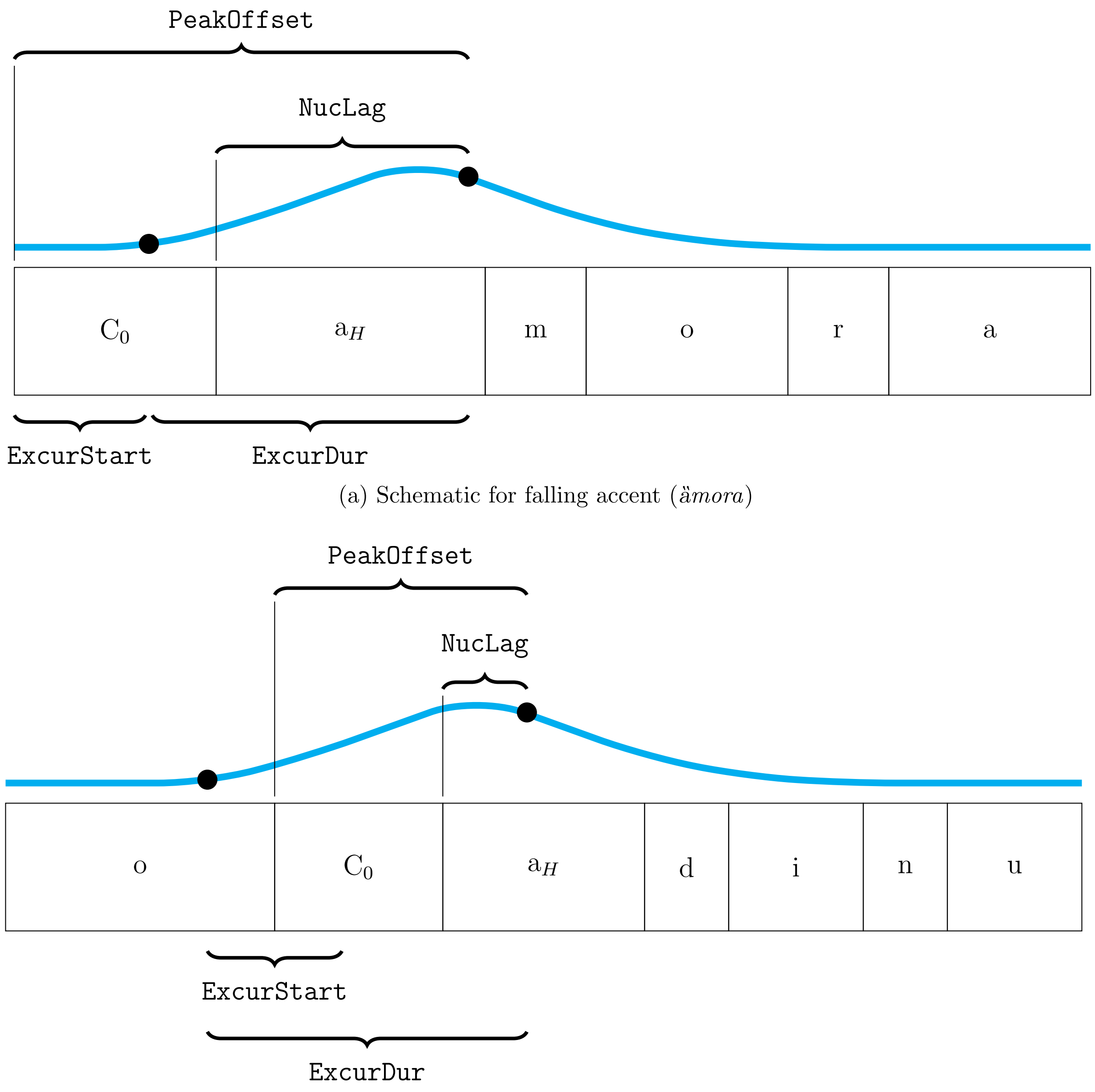
Word group (i.e., -amora vs. -adinu vs. -avinjak) is not included as a random effect, as there are too few groups to allow calculation of random intercepts. Order is also not included as a random effect, as the target words were presented in random order in each round. No models included random slopes.

Fixed effects

- `Complexity` (phonological complexity of the varied syllable onset): categorical variable with two levels, `simple` or `complex`
- `VarOnsDur` (phonetic duration of the *varied* syllable onset): continuous variable, measured in seconds
- `HsylOnsDur` (phonetic duration of the syllable onset of the phonologically H-bearing syllable): continuous variable, measured in seconds
- `Locus` (which locus group the target word belongs to): categorical variable with three levels, `Locus1` (*ȁmora* words), `Locus2` (*adinu* words), or `Locus3` (*àvinjak* words)
- `Dialect` (dialect): categorical variable with two levels, `Belgrade` or `Valjevo`

Dependent variables All dependent variables were measured in seconds; for a schematic of these variables, see Figure 4.2. In this chapter, all F0 measurements are taken relative to the phonologically H-bearing syllable, not the beginning of the word.

- `PeakOffset` (peak timing relative to the beginning of the word): the time interval between the acoustic beginning of phonologically H syllable and the H gesture release (how much the peak is "offset" from the beginning of the syllable)
- `NucLag` (peak timing relative to the nucleus): the time interval between the acoustic beginning of the nucleus of the tone-bearing syllable and the F0 peak offset
- `ExcurStart` (start of the pitch excursion): the time interval between the acoustic beginning of the H-bearing syllable and the start of the upward F0 excursion
- `ExcurDur` (excursion duration): the time interval between the peak offset and the start of the F0 excursion



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

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

Figure 4.2: 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 peak offset (rightmost) of the pitch excursion.

4.2 Results

4.2.1 Segmental characteristics

4.2.1.1 Syllable onsets

Belgrade There is a significant effect of syllable onset on syllable onset duration ($\chi^2(4) = 726.34$, $p < 0.0001$): /r/ < /l/ < /m/ < /mr/ < /ml/ (all $p < 0.0001$ using least squares means Tukey test, except between /l/ and /m/, where $p = 0.01$.). There is also a main effect of word set on syllable onset duration ($\chi^2(2) = 60.39$, $p < 0.0001$); Locus 2 words overall have shorter syllable onsets ($p < 0.0001$, using least squares means Tukey test) than Locus 1 and Locus 3 words, while the latter two are not significantly different from each other ($p = 0.08$). This was predicted, as in the *adinu* word set, the syllable onset being measured is part of the unstressed syllable, and one of the phonetic correlates of stress in Serbian is duration. When considering the *adinu* word set alone, /mr/ and /ml/ are not distinct ($p = 0.51$), but all other onsets remain distinct ($p < 0.0001$ except between /l/ and /m/, where $p = 0.001$).These differences are illustrated in Figure 4.3.

In the Valjevo dialect there is also a significant effect of syllable onset on syllable onset duration ($\chi^2(4) = 1073.90$, $p < 0.0001$): /r/ < /l/ < /m/ < /mr/ < /ml/ ($p < 0.0001$ for all comparisons using least squares means Tukey test). There is also a significant effect of `Group` on syllable onset duration ($\chi^2(2) = 54.08$, $p < 0.0001$); Locus 2 words overall have shorter syllable onsets ($p< 0.0001$ for both comparisons), while Locus 1 and Locus 3 words are not significantly different from each other ($p = 0.23$). When considering the Locus 2 word set alone, /mr/ and /ml/ are not significantly different ($p = 0.72$), but all other onsets remain distinct ($p < 0.0001$ for all). These results are illustrated in Figure 4.3.

Hypothesis 1 is upheld in large part; the one caveat is that the range of phonetic duration is reduced when the varied syllable onset is on the unstressed syllable. However, this only affects the statistical separation between the complex onsets.

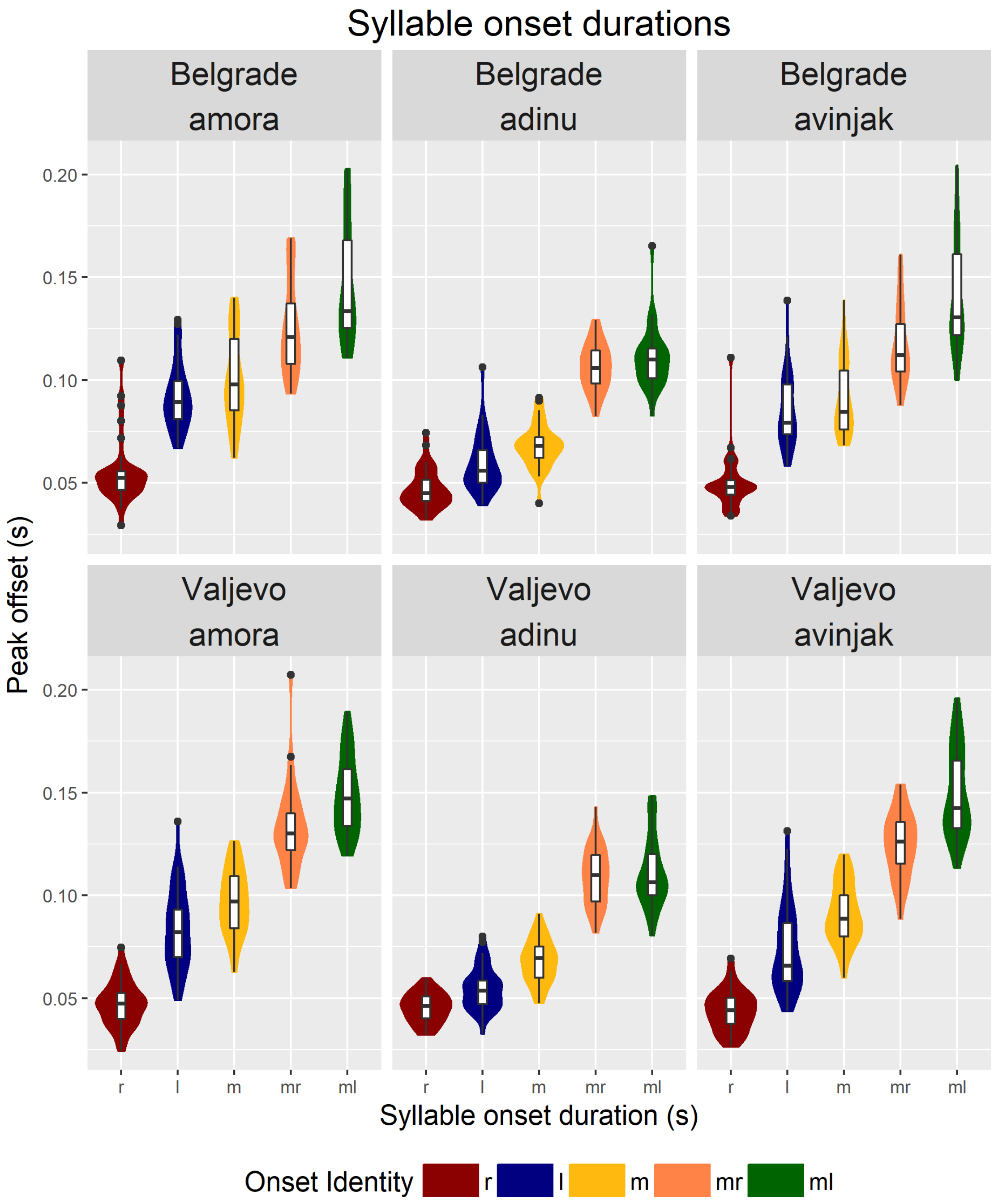


Figure 4.3: Violin plots of syllable onset duration for each word group, both dialects (Belgrade on top; Valjevo on the bottom).

4.2.1.2 Nucleus durations

For all comparisons reported, there is no statistically significant difference between dialects (see individual tables for model comparisons); thus, the dialects will be reported together.

Stressed (first) nucleus For this comparison, the stressed (first) nucleus is compared—i.e., mr[a]mora, mr[a]vinjak, and [o]mladinu—though for the majority of comparisons, Locus 2 words will not be included.

Table 4.2: Comparison of linear mixed effects models for `StressNucDur` (both dialects combined; only *ȁmora* and *àvinjak* sets).

(a) Single predictor models, compared to the null model.

Model for `StressNucDur`	AIC	χ^2	DegF	p[†]
`Dialect + (1|Part)`	-4815.4	0.05	1	0.83
`Group + (1|Part)`	-4879.9	64.47	1	< 0.0001**
`OnsDur + (1|Part)`	-5028.4	213.00	1	< 0.0001**

[†] As compared to the null model, `StressNucDur ∼ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `StressNucDur`	χ^2	DegF	p[†]
`Group + (1|Part)`	—	—	—
`Group + OnsDur + (1|Part)`	209.31	1	< 0.0001**
`Group + OnsDur + Group:OnsDur + (1|Part)`	0.002	1	0.97

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

There is an effect of `Group` on the duration of the stressed nucleus ($\chi^2(1) = 64.47$, $p < 0.0001$ when comparing Locus 1 and Locus 3 groups; see Table 4.2a), though the differences in estimates are quite small: the nucleus of the stressed syllable in Locus 1 words (e.g. mr[**a**]mora]) is 7.2 ms shorter (SE = 0.9 ms) than the nucleus of *àvinjak* words (e.g. mr[**a**]viɲak). Similarly, in a model that includes all three word groups, `Group` still significantly improves the fit of the model ($\chi^2(2) = 24.31$, $p < 0.0001$), though again with differences of a small magnitude; the nucleus of Locus 1 words is significantly different from both Locus 2 ($\beta = -5.3$ ms, SE = 1.4 ms, p = 0.0005) and Locus 3 ($\beta = -6.6$ ms, SE = 1.4

ms, $p < 0.0001$) words, where Locus 1 words have a shorter nucleus. The difference between Locus 3 and Locus 2 in this model is not significant ($p = 0.60$).[3] The durations of the nuclei for each word set are illustrated in Figure 4.4.

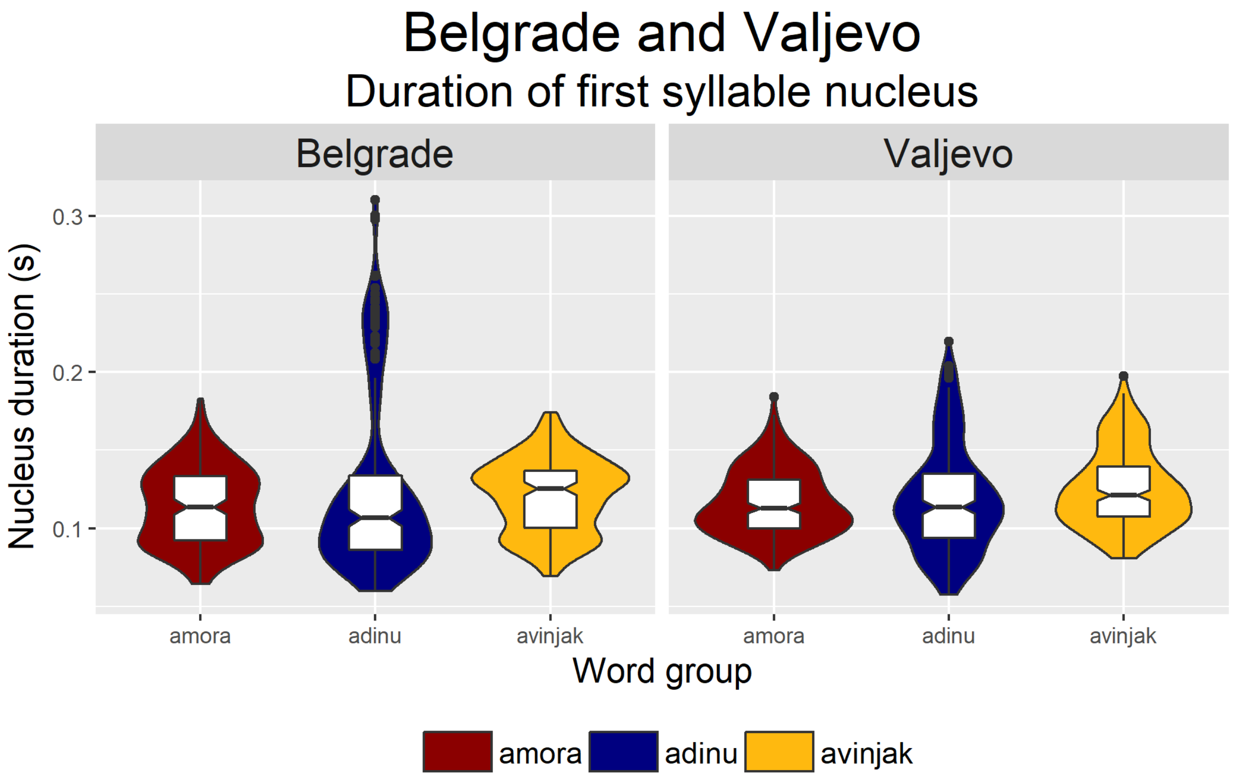


Figure 4.4: Duration of the nucleus of the stressed (first) syllable for each word group, separated by dialect.

There is also an effect of syllable onset duration ($\chi^2(1) = 213.00$, $p < 0.0001$; see Table 4.2a), which parallels the differences found in Experiment 1. For this effect, only Locus 1 and Locus 3 words will be considered, as they compare the same effect (i.e., tautosyllabic onset and nucleus, rather than nucleus and following onset). There is a small decrease in nucleus duration as the syllable onset duration increases (β = -169.7 ms, SE = 10.9 ms, i.e., for every 1,000 ms increase in syllable onset duration, there is a 169.7 ms decrease in

[3] In a conversation with Draga Zec, the possibility of the nucleus of the stressed vowel of rising accents being longer than a comparable falling accent was raised, based on anecdotal observation of a young person's pronunciation several years previously; this trend appears to exist in this data, though to a small magnitude.

nucleus duration); the decrease does not fully compensate for the increase in syllable onset duration—i.e., syllables with long syllable onsets are longer overall than syllables with short syllable onsets.

The effect of syllable onset duration remains when added to a model that already has `Group` ($\chi^2(1)$ = 209.31, p < 0.0001; see Table 4.2b). Again, there is a slight decrease in nucleus duration with increased syllable onset duration (β = -156.8 ms, SE = 17.2 ms). There is no significant interaction between `StressOnsDur` and `Group`; the nuclei in each set of words are affected in the same way by increases in syllable onset duration.

Nucleus of varied syllable For this comparison, the nucleus of the syllable with the varied onset is compared—i.e., mr[a]mora, mr[a]vinjak, and oml[a]dinu. There is an effect of `Group` on the duration of the nucleus ($\chi^2(1)$ = 1343.90, p < 0.0001; see Table 4.3a); as in the previous comparison, Locus 1 and Locus 3 have very similar estimates of duration (Locus 3 is 7.6 ms longer), while Locus 2 is meaningfully shorter (β = -35.4 ms, SE = 0.9 ms). This is tantamount to observing that stress has an effect on vowel duration—note that for this comparison, all nuclei have the same quality (i.e., a short /a/).

With `OnsDur` as a single predictor, there is not a significant improvement on the model ($\chi^2(1)$ = 0.00, p = 0.9988; see Table 4.3a). However, this is likely due to the additional variability introduced by not differentiating between word groups. When `Group` is already included in the model, the addition of `OnsDur` significantly improves the fit ($\chi^2(1)$ = 206.29, p < 0.0001; see Table 4.3b). The interaction between `Group` and `OnsDur` is marginally significant ($\chi^2(1)$ = 7.36, p = 0.03; see Table 4.3b); however, the differences are of magnitude, not direction, and are quite small (as illustrated in Figure 4.5). As for the stressed nucleus, for all word groups, there is a slight negative (but not compensatory) effect of `OnsDur`.

Nucleus of syllable with lexical H For this comparison, the nucleus of the syllable that the H is phonologically associated to is compared—i.e., mr[a]mora, mrav[i]njak, and oml[a]dinu. As for the other comparisons, there is an effect of `Group` ($\chi^2(1)$ = 1793.00, p

Table 4.3: Comparison of linear mixed effects models for `VarNucDur` (both dialects combined; all word sets).

(a) Single predictor models, compared to the null model.

Model for `VarNucDur`	AIC	χ^2	DegF	p†
`Dialect + (1|Part)`	-5925.2	0.32	1	0.57
`Group + (1|Part)`	-7266.7	1343.90	1	$< 0.0001^{**}$
`OnsDur + (1|Part)`	-5924.9	0.00	1	1.00

†As compared to the null model, `VarNucDur ~ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `VarNucDur`	χ^2	DegF	p†
`Group + (1|Part)`	—	—	—
`Group + OnsDur + (1|Part)`	206.29	1	$< 0.0001^{**}$
`Group + OnsDur + Group:OnsDur + (1|Part)`	7.36	1	0.03°

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

< 0.0001; see Table 4.4a). In this case, there are meaningful differences between each word group (p $<$ 0.0001 for all comparisons, using a least squares means Tukey test): Locus 1 words have the longest nucleus (β = 115.7 ms, SE = 3.6 ms), followed by Locus 2 (β = -35.3 ms, SE = 0.9 ms), and then by Locus 3 (β = -57.8 ms, SE = 0.9 ms). As previously noted, duration is a major correlate of stress in Serbian; thus, it was predicted that Locus 1 will have the longest duration. Though not predicted, it is also not surprising that the unstressed [i] in Locus 3 words is shorter than the unstressed [a] in Locus 2 words due to intrinsic vowel height differences.

As in the previous comparisons, there is also an effect of syllable onset duration (χ^2(1) = 210.81, p $<$ 0.0001; see Table 4.4a), though in a single factor model the effect is in the opposite direction from the previous comparisons: as the syllable onset gets longer, so does the nucleus (β = 308.0 ms, SE = 20.3 ms). This is likely due largely to the effect of stress, where stress affects not just nucleus durations, but also syllable onset durations (as described in Section 4.2.1.1). Thus, Locus 1 words would have a longer nucleus as well as a longer

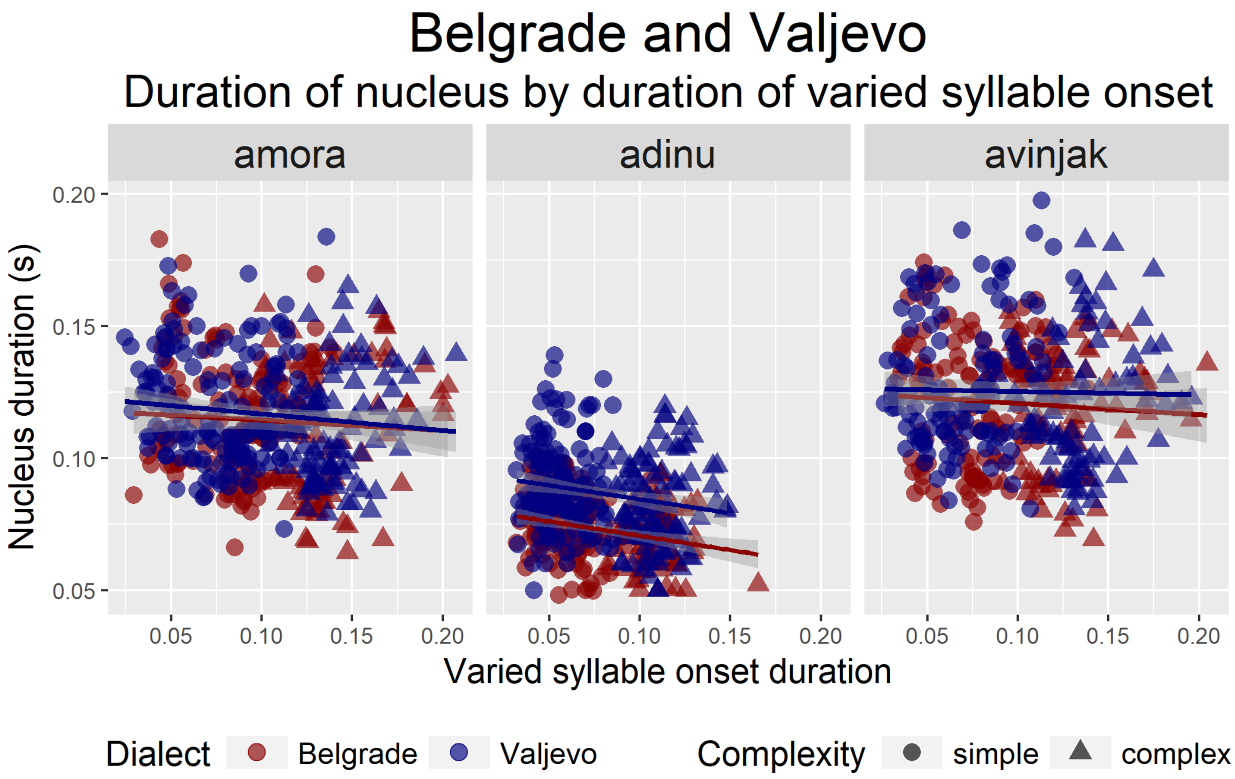


Figure 4.5: Duration of the nucleus of the varied syllable for each word group, colored by dialect.

syllable onset, while Locus 2 words would have a shorter nucleus as well as a shorter syllable onset. The [v] of the H syllable in Locus 3 words is also quite short, which further correlates with the short unstressed [i]. This tendency is illustrated in Figure 4.6, where H syllable nucleus durations, colored by word group, are presented with the fit line for HsylOnsDur.

When Group is included as a fixed effect, the addition of HsylOnsDur improves the fit of the model ($\chi^2(1) = 121.56$, $p < 0.0001$), but now the direction of the effect is reversed—i.e., with increased syllable onset duration, the duration of the nucleus decreases slightly. There is also a significant interaction between Group and HOnsDur ($\chi^2(1) = 32.80$, $p < 0.0001$; see Table 4.4b), though the direction is still negative for all groups; there is simply an even larger effect in Locus 3 words (see Table 4.5).

Table 4.4: Comparison of linear mixed effects models for HNucDur (both dialects combined; all word sets).

(a) Single predictor models, compared to the null model.

Model for HNucDur	AIC	χ^2	DegF	p†
Dialect + (1|Part)	-5546.6	0.80	1	0.37
Group + (1|Part)	-7336.8	1793.00	1	< 0.0001**
HsylOnsDur + (1|Part)	-5924.9	210.81	1	< 0.0001**

†As compared to the null model, HNucDur ~ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for HNucDur	χ^2	DegF	p†
Group + (1|Part)	—	—	—
Group + HsylOnsDur + (1|Part)	121.56	1	< 0.0001**
Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)	32.80	1	< 0.0001**

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

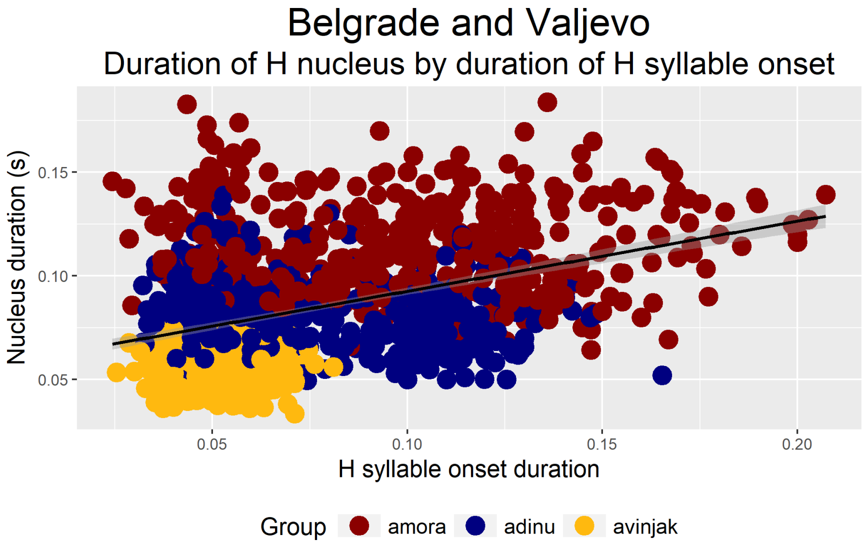


Figure 4.6: Duration of the nucleus of the H syllable, colored by word group (Locus 1 red, Locus 2 blue, Locus 3 yellow), with the fit line for HsylOnsDur. Note that Locus 3, with the unvaried [v], is clustered in the bottom left with the short [i].

Table 4.5: Estimates for the model `HNucDur` $\sim$ `Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)`, both dialects combined. The first line of each group is the intercept (here, *ȁmora*); following values indicate distance from the intercept. Values in ms.

		β	SE
Intercept	(*ȁmora*)	127.8	4.3
	adinu	-34.0	2.4
	àvinjak	-42.8	4.0
`HsylOnsDur`	:*ȁmora*	-118.9	16.0
	:*adinu*	-52.8	26.0
	:*àvinjak*	-407.9	72.5

These differences in H syllable nucleus durations make certain predictions for the behavior of the F0 trajectory. There are two possibilities for the timing of the peak. One possibility is that the F0 trajectory has some intrinsic duration that does not differ based on variation in syllable duration: under this hypothesis, there would be some consistent interval measured in ms across all word types. A second possibility is that the timing of the F0 peak relies on some proportion of the H syllable, in which case there would be some consistent interval measured in percentage of the duration of the H syllable. There still would remain the question of phonologically-based variation (i.e., differences between falling and rising accents, which may or may not be independent from differences in duration between stressed and unstressed syllables) vs. phonetically-based variation (i.e., the intrinsic durational difference between [a] and [i]).

Although in this experiment I am only examining words with phonologically short vowels (compared to the previous experiment, which had both short and long vowels), it is still not preferable to use duration of the H syllable as basis for time normalization, at least not in the Valjevo dialect. In the Belgrade dialect, the peaks of rising accents occur in the post-tonic syllable (i.e., the syllable the H is associated to) almost exceptionlessly, which makes them parallel falling accents; however, in the Valjevo dialect, these peaks are often retracted into the preceding syllable. Thus, it is unclear exactly time normalization using the duration of the phonologically H syllable would be achieved, and what exactly such normalization would mean for each dialect.

4.2.2 Pitch characteristics: Belgrade

Some additional data cleaning was necessary for the analysis of the pitch excursions. For the analysis of peak offset timing, 587 tokens out of a possible 600 had a clear maximum and were retained (2.2% attrition). For the analysis of excursion characteristics, 531 showed clear minima and were retained (9.5% attrition from total). A summary of these tokens is presented in Table 4.6.

Table 4.6: Number of falling accent tokens for each syllable onset with clear F0 landmarks in the Belgrade dialect.

Onset	**H achievement**			**Excursion char.**		
	ȁmora	adinu	àvinjak	ȁmora	adinu	àvinjak
ml	37	40	40	36	35	37
mr	37	39	41	35	37	39
m	38	40	39	37	33	36
l	38	39	40	35	33	40
r	39	40	40	33	34	31

Again, in this chapter `PeakOffset` will always refer to the interval of time between the peak offset and the beginning of the syllable that is phonologically associated with the H. That is, for the *ȁmora* set, it is the time lag between the beginning of the first syllable and the peak offset, but for the *àvinjak* and *adinu* sets, it is the time lag between the beginning of the *second* syllable and the peak offset. Similarly, `VarOnsDur` refers to the duration of the varied syllable onset, and `HsylOnsDur` refers to the duration of the syllable onset of the syllable with the lexical H.

The analyses below serve to address Hypothesis 1:

Hypothesis 1.0 (null hypothesis): The H of rising accents is not influenced by either the stressed syllable or the H syllable onset.

Prediction 1.0: `PeakOffset` will not be affected by variation in syllable onset in either Locus 2 or Locus 3 words.

Hypothesis 1.1: H is associated to the stressed syllable in rising accents.

Prediction 1.1: `PeakOffset` for Locus 3 words (varying onset of stressed syllable) will pattern according to syllable onset like `PeakOffset` for Locus 1 words (single stressed/H syllable).

Hypothesis 1.2: H is associated to the post-stress syllable in rising accents.

Predictions 1.2: `PeakOffset` for Locus 2 words (varying onset of H syllable) will pattern according to syllable onset like `PeakOffset` for Locus 1 words (single stressed/H syllable).

4.2.2.1 H achievement (`PeakOffset` and `NucLag`)

A comparison of linear mixed models shows that the duration of the syllable onset has a positive effect on the timing of the peak relative to the beginning of the word, but only for words where the onset of the phonologically H syllable is varied. Thus, Hypothesis 2 is upheld for the Belgrade dialect. When considering the dataset as a whole, `VarOnsDur` as a single predictor significantly improves the fit of the model ($\chi^2(1) = 94.5$, $p < 0.0001$). However, `HsylOnsDur` as a single predictor also significantly improves the model ($\chi^2(1) = 538.87$, $p < 0.0001$). The AIC values of each model indicate that `HsylOnsDur` provides a better fit (compare -1834.0 for `VarOnsDur` to -2278.4 for `HsylOnsDur`)—and the AIC for `HSylOnsDur` suggests an even better fit than `Group` as a single factor (also significant at $\chi^2(2) = 416.97$, $p < 0.0001$, AIC = -2154.5; see Table 4.7a).

Table 4.7: Comparison of linear mixed effects models for PeakOffset (Belgrade, all word types).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p†
VarOnsDur + (1|Part)	-1834.0	94.50	1	< 0.0001**
HsylOnsDur + (1|Part)	-2278.4	538.87	1	< 0.0001**
Group + (1|Part)	-2154.5	416.97	2	< 0.0001**

†As compared to the null model, PeakOffset ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p†
VarOnsDur + (1|Part)	—	—	—
VarOnsDur + Group + (1|Part)	440.17	2	< 0.0001**
Group + (1|Part)	—	—	—
Group + VarOnsDur + (1|Part)	117.69	1	< 0.0001**
Group + VarOnsDur + Group:VarOnsDur + (1|Part)	204.87	2	< 0.0001**
HsylOnsDur + (1|Part)	—	—	—
HsylOnsDur + Group + (1|Part)	165.23	2	< 0.0001**
Group + (1|Part)	—	—	—
Group + HsylOnsDur + (1|Part)	287.13	1	< 0.0001**
Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)	28.79	2	< 0.0001**

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

The addition of Group to a model with just VarOnsDur significantly improves the fit ($\chi^2(2) = 440.17$, p < 0.0001; see Table 4.7b). There is also a significant interaction between VarOnsDur and Group ($\chi^2(2) = 204.87$, p < 0.0001); varied syllable onsets with longer durations correlate with longer intervals between the start of the H syllable and the peak offset for the *ȁmora* and *adinu* groups, but with (slightly) shorter intervals for the *àvinjak* group (see Table 4.8a). These differences are illustrated in Figure 4.7a, where the duration of the varied syllable onset is plotted against peak offset delay. Although the slope of the

Table 4.8: Table of estimates and standard errors for each group in the Belgrade dialect, comparing `VarOnsDur` and `HsylOnsDur`. Units in ms.

(a) `PeakOffset ∼ Group + VarOnsDur + VarOnsDur:Group + (1|subj)`.

		β	SE
Intercept	(*ämora*)	55.9	13.7
	adinu	-0.4	8.6
	àvinjak	33.7	8.4
`VarOnsDur`	:*ämora*	1057.5	56.4
	:*adinu*	-460.5	91.3
	:*àvinjak*	-1236.9	79.4

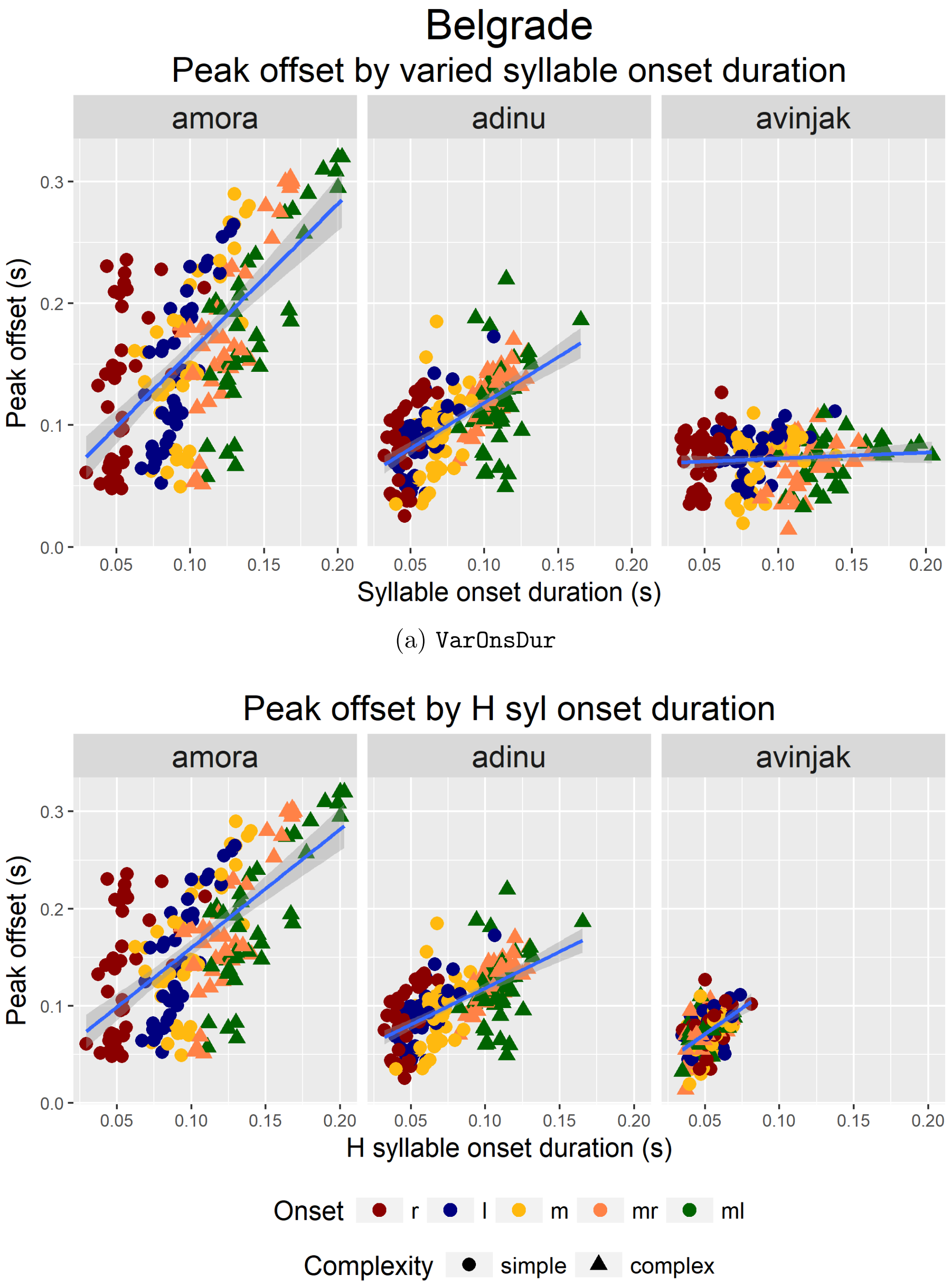
(b) `PeakOffset ∼ Group + HsylOnsDur + HsylOnsDur:Group + (1|subj)`.

		β	SE
Intercept	(*ämora*)	54.7	13.2
	adinu	0.2	8.6
	àvinjak	-3.2	13.9
`VarOnsDur`	:*ämora*	1069.4	56.7
	:*adinu*	-463.7	91.8
	:*àvinjak*	-661.9	250.1

fit line in the *adinu* panel is not as steeply positive as the slope in the *ämora* panel, the increase in peak offset delay is evident in comparison to the *àvinjak* panel. Thus, Hypothesis 3 is upheld for the Belgrade dialect; `VarOnsDur` affects the timing of the peak offset in the expected way only when the syllable with the lexical H has the varied syllable onset.

`Group` still significantly improves the model when added to a model that already has `HsylOnsDur` ($\chi^2(2) = 165.23$, p < 0.0001; see Table 4.7b); however, in this case, there is a positive relationship between `HsylOnsDur` and `PeakOffset` for all groups (see Table 4.8b). In this case, any differences in `HsylOnsDur` for the *àvinjak* group are related to either speaker-specific differences or token-to-token variation in speech rate or production. However, the relationship between `HsylOnsDur` and peak offset is positive, rather than negative (as was the case for `VarOnsDur`). This is illustrated in Figure 4.7b.

Although `HsylOnsDur` is the major predictor of `PeakOffset`, as in the previous experiment, `Complexity` appears to be the major predictor of the interval between the acoustic left edge of the nucleus and the peak offset (`NucLag`). In these models, only *ämora* and *adinu* groups were considered, as the previous discussion showed that only the onset of the syllable with lexical pitch systematically influences the timing of the peak, and `Complexity` was not



(a) `VarOnsDur`

(b) `HsylOnsDur`

Figure 4.7: Three scatter plots comparing the relationships between syllable onset durations and and `PeakOffset` in *ȁmora*, *àvinjak*, and *adinu* words (Belgrade dialect).

manipulated for the *àvinjak* group of words. First, `Group` as a single predictor significantly improves the model ($\chi^2(1)$ = 111.00, p < 0.0001; see Table 4.9a); the peak offset of the rising accent (*adinu*) occurs 38.0 ms (SE = 3.4 ms) earlier relative to its nucleus edge than the peak offset of the falling accent (*ȁmora*; β = 62.2 ms, SE = 14.8 ms). There is not an effect of `OnsDur` ($\chi^2(1)$ = 1.4, p = 0.24);[4] however, there is a significant effect of `Complexity` ($\chi^2(1)$ = 15.45, p < 0.0001; see Table 4.9a).

When `OnsDur` is added as a second fixed effect to a model with `Group`, there is a significant improvement on model fit ($\chi^2(1)$ = 9.48, p = 0.002); the addition of `Complexity` as a third fixed effect further improves the model ($\chi^2(1)$ = 11.10, p = 0.0009; see Table 4.9b). Much like in Experiment 1, the converse is not true—i.e., `Complexity` as the second fixed effect significantly improves the model ($\chi^2(1)$ = 19.93, p < 0.0001), but `OnsDur` as a third fixed effect does not improve the fit ($\chi^2(1)$ = 0.64, p = 0.42; see Table 4.9b). Finally, there is no interaction between `Group` and `Complexity`; i.e., for both the falling and the rising accent, peaks occur closer to the left edge of the nucleus in words with complex syllable onsets than in words with simple syllable onsets. This again indicates that `Complexity` as a phonological characteristic has more of an effect on the time interval between the peak and the left edge of the nucleus. Overall, Hypothesis 1.0 is rejected in favor of Hypothesis 1.2 in the Belgrade dialect.

The magnitude of the effect of `Complexity` is also similar to that reported in Chapter 3 (~ 15 ms). That there is no interaction between `Group` and `Complexity`, however, is perhaps somewhat surprising, given the previously described differences in syllable onset duration and peak offset timing in the falling (stressed) and rising (unstressed) syllables.[5] The results from Experiment 1 suggested that the origin of this difference is in the timing of the start of the pitch excursion, as it was significantly affected by both phonetic duration

[4] Here I am collapsing the two syllable onset distinctions as for both word groups considered in these analyses, the varied and H syllable onset are one and the same

[5] In the model that includes the interaction, `NucLag ~ Group + Complexity + Group:Complexity + (1|Part)`, the difference in estimates between *ȁmora* and *adinu* words with complex onsets is just 3.0 ms. This difference is far smaller than that between nucleus duration between the same two groups.

Table 4.9: Comparison of linear mixed effects models for NucLag (Belgrade, only *ȁmora* and *adinu* types).

(a) Single predictor models, compared to the null model.

Model for NucLag	AIC	χ^2	DegF	p[†]
Complexity + (1|Part)	-1422.5	15.45	1	< 0.0001**
OnsDur + (1|Part)	-1408.5	1.40	1	0.24
Group + (1|Part)	-1518.1	111.00	1	< 0.0001**

[†] As compared to the null model, NucLag ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for NucLag	χ^2	DegF	p[†]
Group + (1|Part)	—	—	—
Group + OnsDur + (1|Part)	9.48	1	0.002*
Group + OnsDur + Complexity + (1|Part)	11.10	1	0.0009**
Group + (1|Part)	—	—	—
Group + Complexity + (1|Part)	19.93	1	< 0.0001**
Group + Complexity + OnsDur + (1|Part)	0.64	1	0.42
Group + Complexity + (1|Part)	—	—	—
Group + Complexity + Group:Complexity + (1|Part)	0.20	1	0.65

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

and complexity. If the difference in timing due to complexity is ultimately due to differences in gestural coordination, as was speculated, differences in gestural duration (as reflected in different acoustic durations) should produce parallel differences in peak timing relative to the left edge of the nucleus.

4.2.2.2 Excursion characteristics (ExcurStart and ExcurDur)

Belgrade Overall, the results of Experiment 2 parallel the results of Experiment 1. A comparison of mixed effect models shows that there is an effect of HsylOnsDur on ExcurStartHsyl ($\chi^2(1) = 225.67$, $p < 0.0001$); for every 1,000 ms increase in syllable onset duration, there is a 740.8 ms delay (SE = 69.1 ms) delay in the start of the excursion. There

Table 4.10: Estimates for `NucLag` in the model `NucLag` $\sim$ `Group + Complexity + (1|Part)`. Units in ms.

		β	SE
`Group`	*ȁmora*	68.1	14.8
	adinu	-37.9	3.3
`Complexity`	(simple)	—	—
	complex	-15.1	3.3

is only a marginally significant effect of `Complexity` ($\chi^2(1)$ = 6.16, p = 0.01); this lack of significance is likely due to the conflation of the two word groups (that is, stress-based durational differences that are obscured by a categorical separation). There is also a significant effect of `Group` ($\chi^2(1)$ = 225.67, p < 0.0001; see Table 4.11a), where pitch excursions start 70.8 ms earlier (SE = 4.0 ms) in Locus 2 words than in Locus 1 words.

Compared to a base model with `Group` as a fixed effect, the addition of `HsylOnsdur` also significantly improves the fit of the model ($\chi^2(1)$ = 48.63, p < 0.0001), but now the estimate is smaller; for every 1,000 ms increase in `HsylOnsDur`, there is just a 416.6 ms delay (SE = 57.7 ms) in the start of the excursion (compare 740.8 ms for the model with just `HsylOnsdur`). There is also a significant interaction between `Group` and `HsylOnsDur` ($\chi^2(1)$ = 15.58, p < 0.0001); increases in `HsylOnsDur` have less of an effect on the start of Locus 2 excursions than on the start of Locus 1 excursions.

The addition of `Complexity` additionally improves the fit of the model ($\chi^2(1)$ = 8.94, p = 0.003; see Table 4.11b); somewhat unexpectedly, the intercept for complex onsets is 18.6 ms earlier (SE = 6.2 ms) than for simple onsets. Furthermore, unlike in Experiment 1, the interaction `Complexity:HsylOnsDur` significantly improves the fit of the model ($\chi^2(1)$ = 16.56, p < 0.0001; see Table 4.11b. However, participant BGF02-ii is exerting quite a lot of influence on the model (Cook's D = 5.21): they had a slower rate of speech, as well as

Table 4.11: Comparison of linear mixed effects models for `ExcurStartHsyl` (Belgrade, only *ȁmora* and *adinu* words).

(a) Single predictor models, compared to the null model.

Model for `ExcurStartHsyl`	AIC	χ^2	DegF	p[†]
`Complexity + (1|Part)`	-1062.7	6.16	1	0.01°
`HsylOnsDur + (1|Part)`	-1155.8	99.27	1	< 0.0001**
`Group + (1|Part)`	-1282.2	225.67	1	< 0.0001**

[†] As compared to the null model, `ExcurStartHsyl ~ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `ExcurStartHsyl`	χ^2	DegF	p[†]
`Group + (1|Part)`	—	—	—
`Group + HsylOnsDur + (1|Part)`	48.63	1	< 0.0001**
`Group + HsylOnsDur + Complexity + (1|Part)`	8.94	1	0.003*
`Group + HsylOnsDur + Complexity + Complexity:HsylOnsDur + (1|Part)`	16.56	1	< 0.0001**
`Group + HsylOnsDur + (1|Part)`	—	—	—
`Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)`	15.58	1	< 0.0001**

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

late excursion starts and late peaks in their falling accents.[6] Thus, they have both complex onsets with the longest phonetic durations, as well as extremely late pitch excursions. This exceptional behavior is illustrated in Figure 4.8.

There is also an effect of syllable onset duration on the duration of the pitch excursion (`ExcurDur`), ($\chi^2(1) = 30.16$, $p < 0.0001$; see Table 4.12a); for every 1,000 ms increase in syllable onset duration, there is a 305.0 ms increase (SE = 54.3 ms) in excursion duration. There is also a significant effect of `Complexity` ($\chi^2(1) = 24.06$, $p < 0.0001$); however, it is likely that this is simply due to the correlation between the complexity and duration of an onset, as `Complexity` does not contribute to a model that already has `HsylOnsDur` and

[6] Proportionally late—that is, their peaks occurred later in the syllable as well as later in absolute millisecond values than would be predicted by the slower rate of speech.

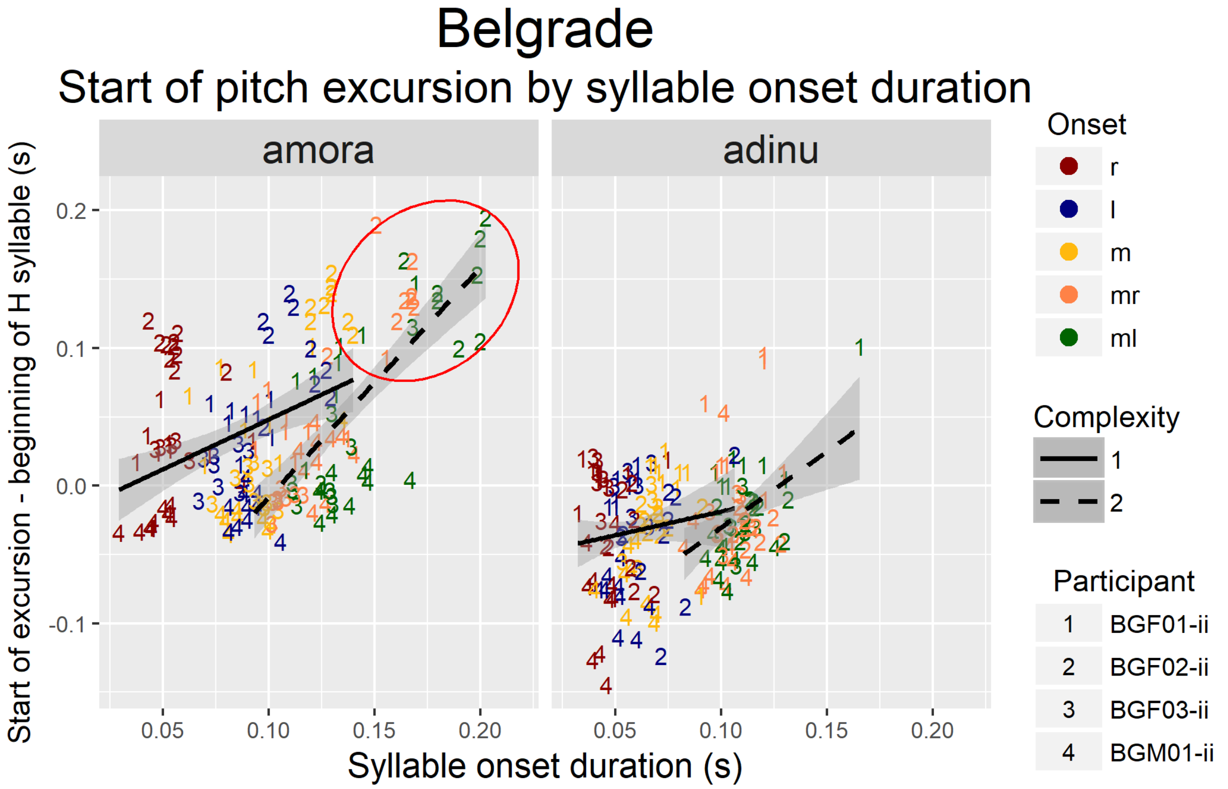


Figure 4.8: Scatter plots showing the relationships between syllable onset duration and the start of the pitch excursion, separated by word type, for the Belgrade dialect. Datapoints are marked according to participant (number) and syllable onset identity (color); fit lines are a simple linear model line for simple and complex onsets. The ellipse marks a 0.95 confidence interval of the complex onset tokens of BGF02-ii.

`Group` ($\chi^2(1) = 0.17$, p = 0.68; see Table 4.12b).[7]

`Group` as a single fixed effect does not significantly improve the fit of the model ($\chi^2(1) = 1.68$, p = 0.19; see Table 4.12a). However, it does significantly improve on a model that already has `HsylOnsDur` as a fixed effect ($\chi^2(1) = 13.86$, p = 0.0002; see Table 4.12b); the interaction between `Group` and `HsylOnsDur` is not significant ($\chi^2(1) = 0.30$, p = 0.58). The estimated difference between groups is middlingly large (14.9 ms; see Table 4.13), and it is Locus 2 words that have longer excursions.

However, recall that pitch excursions start much earlier for Locus 2 words than for Locus

[7] Or, compared to a model with just `HsylOnsDur` (and not `Group`), $\chi^2(1) = 1.77$, p = 0.18.

Table 4.12: Comparison of linear mixed effects models for `ExcurDur` (Belgrade, only *ȁmora* and *adinu* words).

(a) Single predictor models, compared to the null model.

Model for `ExcurDur`	AIC	χ^2	DegF	p[†]
`Complexity + (1|Part)`	-1319.5	24.06	1	< 0.0001**
`HsylOnsDur + (1|Part)`	-1325.6	30.16	1	< 0.0001**
`Group + (1|Part)`	-1297.1	1.68	1	0.19

[†] As compared to the null model, `ExcurDur ~ 1 + (1|Part)`

° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `ExcurDur`	χ^2	DegF	p[†]
`HsylOnsDur + (1|Part)`	—	—	—
`HsylOnsDur + Group + (1|Part)`	13.86	1	0.0002**
`HsylOnsDur + Group + Complexity + (1|Part)`	0.17	1	0.68
`HSylOnsDur + (1|Part)`	—	—	—
`HsylOnsDur + Group + (1|Part)`	13.86	1	0.0002**
`HsylOnsDur + Group + HsylOnsDur:Group + (1|Part)`	0.30	1	0.58

[†] As compared to model immediately above

° < 0.05, * < 0.01, ** < 0.001

Table 4.13: Table of estimates and standard errors for each group in the Belgrade dialect, from the model `ExcurDur ~ HsylOnsDur + Group + (1|subj)`. Units in ms.

		β	SE
Intercept	(*ȁmora*)	83.9	9.7
	adinu	+14.9	4.0
`HsylOnsDur`	+1,000 ms	384.6	57.3

1 words (β = 70.8 ms, SE = 4.0 ms); thus, even in the case that the approximately 15 ms difference is meaningful as well as significant, its contribution to the timing of the peak is comparatively small. Furthermore, there appears to be some between-speaker variation: participants BGF02-ii and BGF03-ii appear to have more separation between their word groups, and with longer excursions for Locus 2, while BGF01-ii has more overlap but with the Locus 2 excursions generally shorter than Locus 1 excursions. These relationships are illustrated in Figure 4.9.

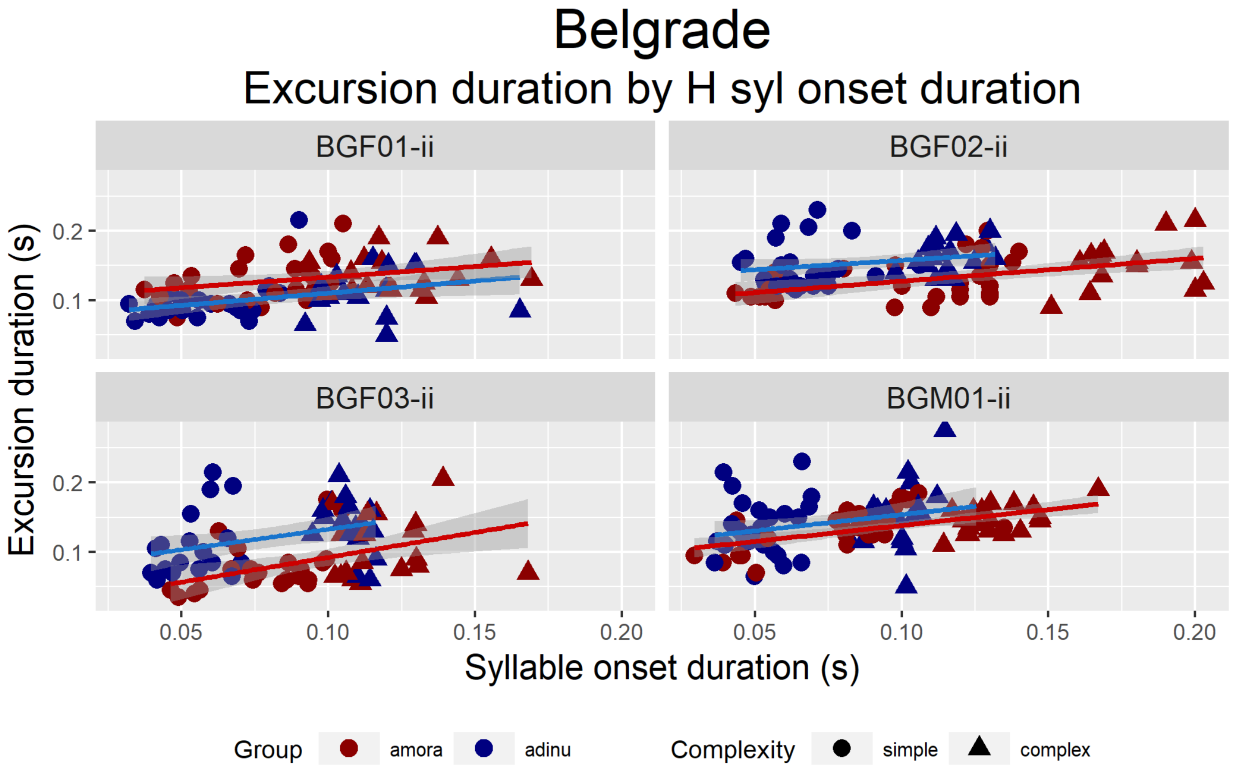


Figure 4.9: Scatter plots showing the relationships between syllable onset duration and excursion duration, separated by participant, for the Belgrade dialect; only Locus 1 (red) and Locus 2 (blue) groups included.

Thus, as found for the Belgrade dialect in Experiment 1, changes in the timing of the peak offset relative to the acoustic beginning of the word are caused by changes both in the timing of the start of the pitch excursion, as well as the duration of the excursion. The relationship between excursion duration and syllable onset is fairly straightforward in the

Belgrade dialect, where increases in syllable onset duration are correlated with increases in excursion duration, while the relationship between the start of the excursion and the syllable onset is additionally affected by phonological complexity. The relationships between characteristics of the pitch excursion and syllable onset duration for the dialect are illustrated below in Figure 4.14a.

4.2.3 Pitch characteristics: Valjevo

Out of the total 750 tokens for the Valjevo dialect, 690 were retained for analysis of peak offset timing (8% attrition). For the analysis of the excursion characteristics, only 369 tokens have clear minima (46.5% attrition). The Locus 2 and Locus 3 words (the rising accents) were most severely affected, with 56.8% and 51.5% attrition, respectively (compare 32.0% attrition for Locus 1 words). The number of tokens available for each word is given in Table 4.14.

Table 4.14: Number of falling accent tokens for each syllable onset with clear F0 landmarks for the Valjevo dialect.

Onset	**H achievement**			**Excursion char.**		
	ȁmora	adinu	àvinjak	ȁmora	adinu	àvinjak
ml	47	49	40	32	20	19
mr	46	49	40	31	22	21
m	49	48	38	35	22	14
l	50	49	46	33	24	24
r	49	48	42	33	17	22

These analyses also serve to address Hypothesis 1:

Hypothesis 1.0 (null hypothesis): The H of rising accents is not influenced by either the stressed syllable or the H syllable onset.

Prediction 1.0: `PeakOffset` will not be affected by variation in syllable onset in either Locus 2 or Locus 3 words.

Hypothesis 1.1: H is associated to the stressed syllable in rising accents.

Prediction 1.1: `PeakOffset` for Locus 3 words (varying onset of stressed syllable) will pattern according to syllable onset like `PeakOffset` for Locus 1 words (single stressed/H syllable).

Hypothesis 1.2: H is associated to the post-stress syllable in rising accents.

Predictions 1.2: `PeakOffset` for Locus 2 words (varying onset of H syllable) will pattern according to syllable onset like `PeakOffset` for Locus 1 words (single stressed/H syllable).

4.2.3.1 H achievement (`PeakOffset` and `NucLag`)

In the Valjevo dialect, the duration of the syllable onset has a positive effect on the timing of the peak relative to the beginning of the word, but only for words where the onset of the phonologically H syllable is varied. Thus, Hypothesis 2 is upheld for the Valjevo dialect. When considering the dataset as a whole, `VarOnsDur` significantly improves the fit of the model ($\chi^2(1) = 56.47$, $p < 0.0001$), though going by the AIC, a model with `Group` as a single predictor provides a better fit ($\chi^2(1) = 782.60$, $p < 0.0001$; see Table 4.15a). This is likely for two reasons: first, as discussed for Belgrade, even though the measure of `PeakOffset` has been unified for falling and rising accents (i.e., taking as the leftmost point the beginning of the phonologically H syllable, not the stressed syllable), the phonologically H syllable in rising accents is much shorter than in falling accents, since stress elongates syllables in Serbian. Second, which is unique to Valjevo, the timing of the peak relative to the phonologically H syllable is often different for falling and rising accents. As observed by Zec and Zsiga (2018), the peak of Valjevo rising accents often occurs before the beginning of the phonologically H syllable. For falling accents this is impossible, as all falling accents have the H associated to the first syllable.

Due to the large effect of `Group`, the nested model comparisons start from a model with

Table 4.15: Comparison of linear mixed effects models for PeakOffset (Valjevo, all word groups).

(a) Single predictor models, compared to the null model.

Model for PeakOffset	AIC	χ^2	DegF	p[†]
VarOnsDur + (1|Part)	-1775.2	56.47	1	< 0.0001**
HsylOnsDur + (1|Part)	-2097.4	378.65	1	< 0.0001**
Group + (1|Part)	-2499.4	782.61	2	< 0.0001**

[†] As compared to the null model, PeakOffset ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for PeakOffset	χ^2	DegF	p[†]
Group + (1|Part)	—	—	—
Group + VarOnsDur + (1|Part)	51.13	1	< 0.0001**
Group + VarOnsDur + Group:VarOnsDur + (1|Part)	183.06	2	< 0.0001**
Group + (1|Part)	—	—	—
Group + HsylOnsDur + (1|Part)	173.50	1	< 0.0001**
Group + HsylOnsDur + Group:HSylOnsDur + (1|Part)	29.97	2	< 0.0001**

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

Group as a fixed effect. The addition of VarOnsDur still significantly improves the model ($\chi^2(1) = 51.13$, $p < 0.0001$). In addition, there is a significant interaction between VarOnsDur and Group ($\chi^2(1) = 183.06$, $p < 0.0001$; see 4.15b). For the *ȁmora* and *adinu* groups, there is a positive relationship between VarOnsDur and PeakOffset, while for the *àvinjak* group there is a (shallowly) negative relationship (see Table 4.16). That is, for *ȁmora* and *adinu* words, increases in syllable onset duration correlate with later peaks, but for *àvinjak* words, increases in syllable onset duration correlate with earlier peaks. Thus, Hypothesis 3 is also upheld for the Valjevo dialect; the two groups where the syllable with the varied onset is the phonologically H syllable exhibit the same relationship between syllable onset duration and peak timing as found in Experiment 1, while the group where the syllable with the varied onset is just the stressed syllable (with no lexical H) exhibits the opposite relationship.

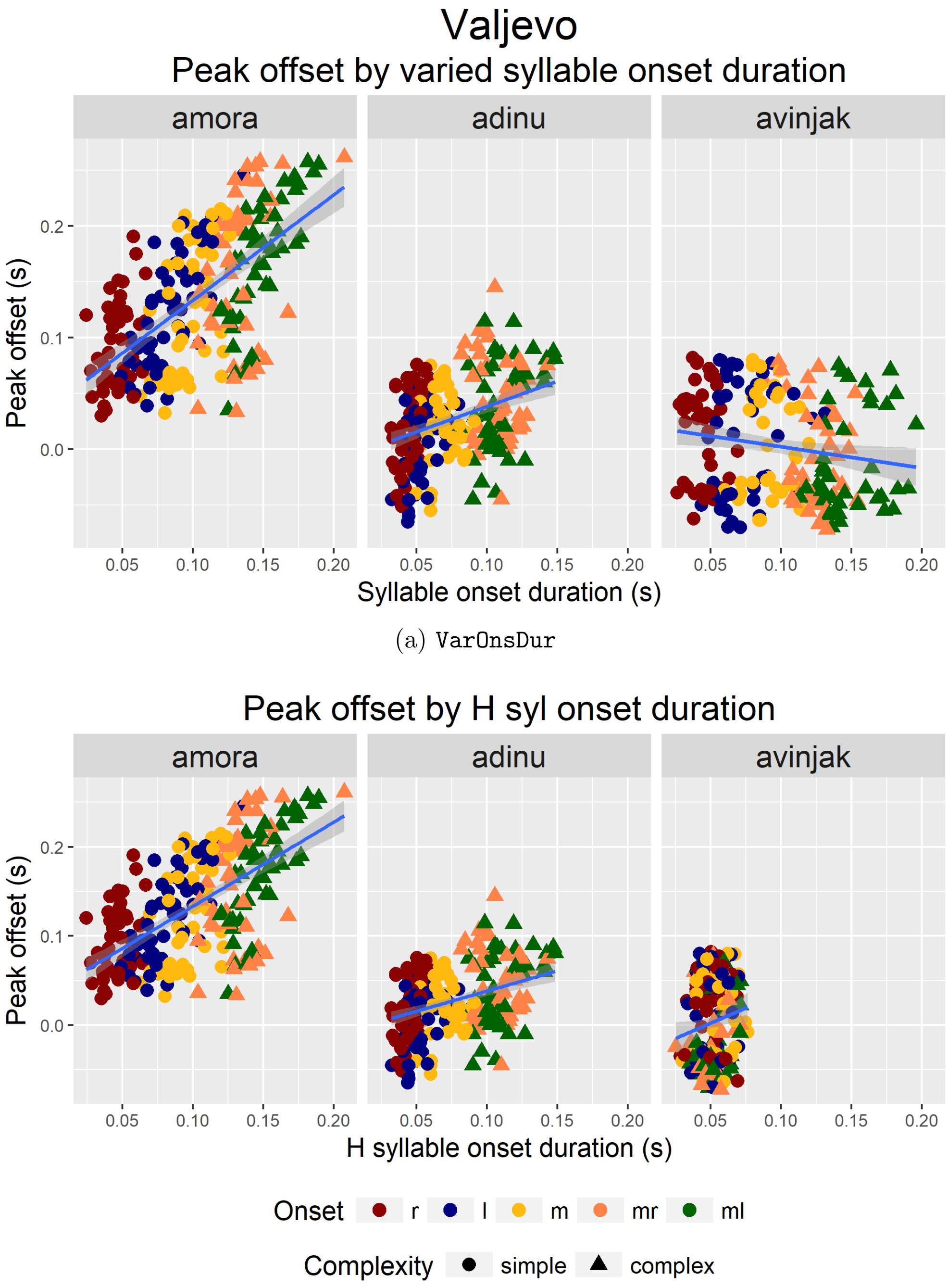
Table 4.16: Table of estimates and standard errors for each group in the Valjevo dialect, from the model `PeakOffset` $\sim$ `Group + VarOnsDur + VarOnsDur:Group + (1|subj)`. Units in ms.

		β	SE
Intercept	(*ȁmora*)	54.3	14.1
	adinu	-55.8	8.2
	àvinjak	-18.4	8.2
`VarOnsDur`	:*ȁmora*	790.7	53.8
	:*adinu*	-414.2	88.2
	:*àvinjak*	-1109.7	77.0

However, they do not behave in precisely the same way. Even though both relationships are positive, the slope is shallower for *adinu* words than for *ȁmora* words. In a model that includes just these two word groups, the interaction `Group:VarOnsDur` still significantly improves the fit of the model ($\chi^2(1) = 27.22$, p $<$ 0.0001). The difference in slopes is illustrated in Figure 4.10a.

As alluded to previously, using `HsylOnsDur` (rather than `VarOnsDur`) as the predictive factor shows a more consistent relationship across word groups. `HsylOnsdur` as a single fixed effect significantly improves the fit of the model ($\chi^2(1) = 378.65$, p $<$ 0.0001), and the AIC suggests that this is a greater improvement than using just `VarOnsDur` (compare -2097.4 for `HsylOnsDur` and -1775.2 for `VarOnsDur`; see Table 4.15a). There is still a significant interaction between `Group` and `HsylOnsDur` ($\chi^2(2) = 29.97$, p $<$ 0.0001; see Table 4.15b), which indicates that increases in `HsylOnsDur` does not have the same effect across groups. However, unlike for `VarOnsDur`, the slopes are all positive for all groups—i.e., increases in `HsylOnsDur` correlate with later peaks for all groups (see Table 4.17 for estimates and Figure 4.10b for a scatter plot with fit lines).[8]

[8] It should be noted that increases in `HsylOnsDur` for the *àvinjak* group stem from either speech rate differences or trial-to-trial differences, not deliberately varied durational differences, and thus the durations



(a) `VarOnsDur`

(b) `HsylOnsDur`

Figure 4.10: Three scatter plots comparing the relationships between syllable onset durations and and `PeakOffset` in *ȁmora*, *adinu*, and *àvinjak* words (Valjevo dialect).

Table 4.17: Table of estimates and standard errors for each group in the Valjevo dialect, from the model `PeakOffset` ∼ `Group + HsylOnsDur + HsylOnsDur:Group + (1|subj)`. Units in ms.

		β	SE
Intercept	(*ȁmora*)	53.6	13.9
	adinu	-55.6	8.4
	àvinjak	-40.0	14.3
`HsylOnsDur`	:*ȁmora*	797.2	55.1
	:*adinu*	-413.6	90.2
	:*àvinjak*	-955.3	255.5

There is a significant effect of `Group` on the timing of the peak offset relative to the start of the nucleus ($\chi^2(1) = 445.92$, $p < 0.0001$; see Table 4.18a); peaks occur 83.2 ms earlier (SE = 3.1 ms) relative to the nucleus for Locus 2 words than for Locus 1 words. However, unlike in Experiment 1, there is no apparent effect of `Complexity` on `NucLag` except a model that includes only `Complexity` as a fixed effect ($\chi^2(1) = 23.10$, $p < 0.0001$; see Table 4.18a). `HsylOnsDur` as a single fixed factor does not significantly improve the model ($\chi^2(1) = 0.25$, $p = 0.62$); however, this is likely due to the two major effects of `Group`, the first on the duration of `HsylOnsDur` (where Locus 1 words have longer syllable onsets due to stress), and the second on the timing of the peak (where peaks in Locus 1 words occur later, as they are never retracted to the preceding syllable).

When `Group` is already included as a fixed effect, however, the addition of `HsylOnsDur` does significantly improve the model ($\chi^2(1) = 70.37$, $p < 0.0001$; see Table 4.18b). It also significantly improves the fit of a model that has both `Group` and `Complexity` as fixed effects ($\chi^2(1) = 26.86$, $p < 0.0001$), which indicates that, unlike in Experiment 1, differences in peak timing relative to the nucleus are not largely determined by phonological characteristics of the

are fairly clustered and as such do not illustrate a strong relationship.

Table 4.18: Comparison of linear mixed effects models for `NucLag` (Valjevo, only *ȁmora* and *adinu* types).

(a) Single predictor models, compared to the null model.

Model for `NucLag`	AIC	χ^2	DegF	p[†]
`Complexity + (1|Part)`	-1456.0	23.10	1	< 0.0001**
`HsylOnsDur + (1|Part)`	-1433.1	0.25	1	0.62
`Group + (1|Part)`	-1878.8	445.92	1	< 0.0001**

[†] As compared to the null model, `NucLag ~ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `NucLag`	χ^2	DegF	p[†]
`Group + (1|Part)`	—	—	—
`Group + HsylOnsDur + (1|Part)`	79.37	1	< 0.0001**
`Group + HsylOnsDur + Complexity + (1|Part)`	0.22	1	0.67
`Group + (1|Part)`	—	—	—
`Group + Complexity + (1|Part)`	52.72	1	< 0.0001**
`Group + Complexity + HsylOnsDur + (1|Part)`	26.86	1	< 0.0001**
`Group + Complexity + (1|Part)`	—	—	—
`Group + Complexity + Group:Complexity + (1|Part)`	0.20	1	0.65
`Group + HsylOnsDur + (1|Part)`	—	—	—
`Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)`	27.22	1	< 0.0001**

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

syllable onset. The effects are in the same direction as in Experiment 1: peaks of words with longer syllable onsets occur further to the left relative to the beginning of the nucleus (see the negative estimates under `HsylOnsDur` in Table 4.19). This effect is more pronounced in Locus 2 words, which is also expected, given the flatter patterning of `PeakOffset` described previously.

Table 4.19: Table of estimates and standard errors for each group in the Valjevo dialect, from the model `NucLag` $\sim$ `Group + HsylOnsDur + Group:HsylOnsDur + (1|subj)`, only Locus 1 and Locus 2 words included. Units in ms.

		β	SE
Intercept	(*ȁmora*)	56.5	14.2
	adinu	-55.0	7.6
`HsylOnsDur`	:*ȁmora*	-230.4	50.1
	:*adinu*	-432.5	81.7

4.2.3.2 Excursion characteristics (`ExcurStart` and `ExcurDur`)

For the Valjevo dialect, the results from Experiment 2 are also very similar to those from Experiment 1; the main driving force behind the syllable onset-based is the duration of the excursion, rather than the timing of the start of the excursion. A comparison of linear mixed effects models shows that there is a significant effect of `Group` on the timing of the start of the F0 excursion ($\chi^2(1)$ = 352.41, p < 0.0001); the pitch excursions for Locus 2 words start 102.9 ms earlier (SE = 3.8 ms) relative to the beginning of the phonologically H syllable than the pitch excursions for Locus 1 words. This is the only logical choice, as the peaks of Valjevo rising accents often occur before the beginning of the H syllable; thus, the entirety of the H excursion is outside its TBU. There is also a significant effect of `HsylOnsdur` ($\chi^2(1)$ = 28.58, p < 0.0001) on `ExcurStartHsyl`, but not of `Complexity` ($\chi^2(1)$ = 0.46, p = 0.50; see Table 4.20a), which suggests that some of the effect of `HsylOnsDur` may be due to group-(stress-)based differences.

When added to model that already has `Group` as a fixed effect, `HsylOnsDur` somewhat improves the model ($\chi^2(1)$ = 6.79, p = 0.009). The effect is fairly small; for a 1,000 ms increase in the duration of the syllable onset, there is a 137.4 ms delay (SE = 52.3 ms) in the start of the excursion. The interaction `Group:HsylOnsDur` is not significant ($\chi^2(1)$ =

Table 4.20: Comparison of linear mixed effects models for `ExcurStartHsyl` (Valjevo, only *ămora* and *adinu* words).

(a) Single predictor models, compared to the null model.

Model for `ExcurStartHsyl`	AIC	χ^2	DegF	p[†]
`Complexity + (1|Part)`	-764.2	0.46	1	0.50
`HsylOnsDur + (1|Part)`	-792.3	28.58	1	< 0.0001**
`Group + (1|Part)`	-1116.1	352.41	1	< 0.0001**

[†] As compared to the null model, `ExcurStartHsyl ~ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `ExcurStartHsyl`	χ^2	DegF	p[†]
`Group + (1|Part)`	—	—	—
`Group + HsylOnsDur + (1|Part)`	6.79	1	0.009*
`Group + HsylOnsDur + Group:HsylOnsDur + (1|Part)`	0.36	1	0.55

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

0.36, p = 0.55; see Table 4.20b), indicating that this effect is the same for both word groups. These relationships are illustrated in Figure 4.11.

In contrast, characteristics of the syllable onset do have an effect on the duration of the pitch excursion. Both `Complexity` ($\chi^2(1) = 48.66$, p < 0.0001) and `HsylOnsDur` ($\chi^2(1) = 54.32$, p < 0.0001) significantly improve the model as single fixed effects (see Table 4.12a); for every 1,000 ms increase in syllable onset duration, there is a 455.9 ms (SE = 58.8 ms) increase in excursion duration. This increase is smaller than the one observed for `PeakOffset` (β = 1,034.0 ms, SE = 68.1 ms for a model with all good maxima included; β = 965.0 ms, SE = 84.1 ms for a model with the subset of data used to examine the start of the excursion). However, there is not a significant effect of `Group` alone ($\chi^2(1) = 1.13$, p = 0.29),

In this comparison too, `Group` as a single fixed effect does not significantly improve the model ($\chi^2(1) = 1.13$, p = 0.29), which suggests that the bulk of the difference in peak timing between between falling (*ămora*) and rising (*adinu*) accents stems from differences in the

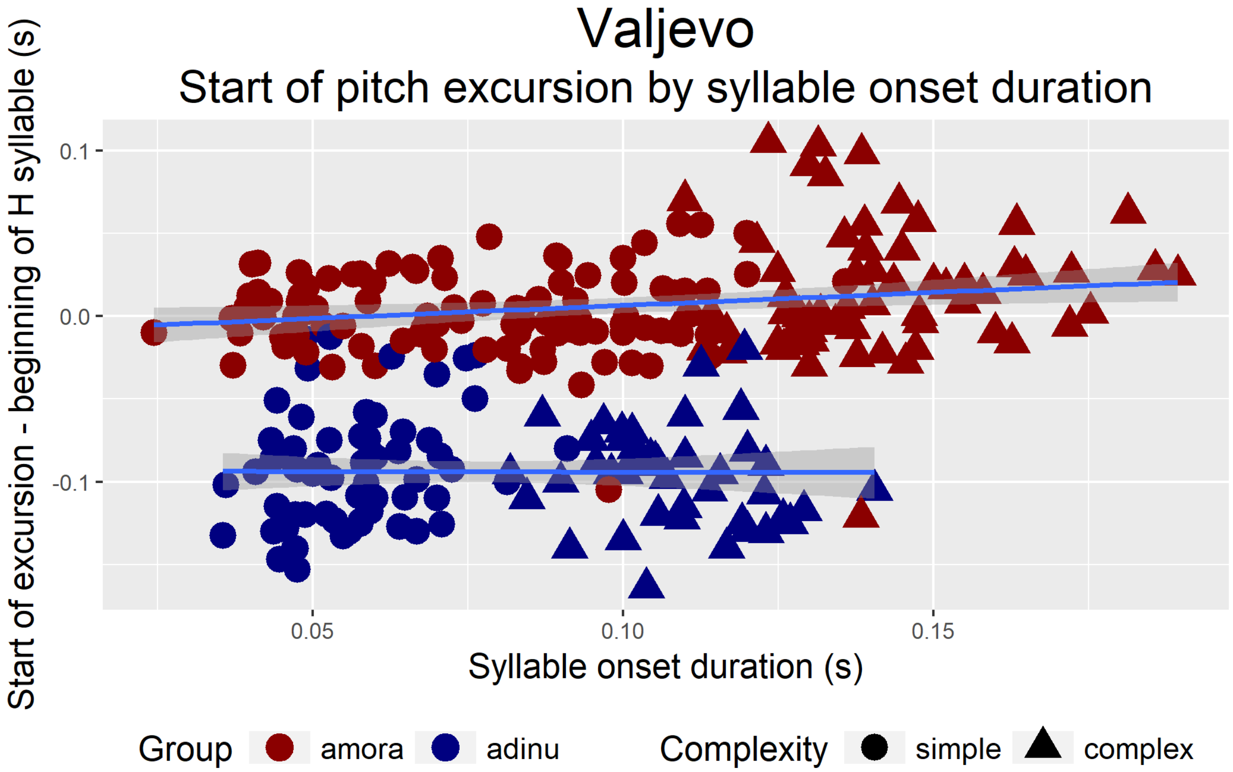


Figure 4.11: A scatter plot showing the flat relationship between the timing of the start of the pitch excursion and the syllable onset duration (only Locus 1 and Locus 2 groups included).

timing of the beginning of the excursion—i.e., rising accent peaks occur early relative to the syllable they are phonologically associated to because their excursions start early. `Group` does significantly improve on a model that already has `HsylOnsDur` as a fixed effect ($\chi^2(1) = 11.76$, p = 0.0006); there is also a marginally significant interaction between `Group` and `HsylOnsDur` ($\chi^2(1) = 6.71$, p = 0.01; see Table 4.21b).

The estimates for this model are given in Table 4.22. Increases in `HsylOnsDur` have a greater effect on the length of the excursion for Locus 1 words than for Locus 2 words, as shown by the very negative value for Locus 2 under `HsylOnsDur`; rather than an approximately 530 ms increase in excursion duration for every 1,000 ms increase in syllable onset duration, there is only an approximately 210 ms increase. It is unclear, however, if this is

Table 4.21: Comparison of linear mixed effects models for ExcurDur (Valjevo, only *ȁmora* and *adinu* words).

(a) Single predictor models, compared to the null model.

Model for ExcurDur	AIC	χ^2	DegF	p†
Complexity + (1|Part)	-1029.8	48.66	1	< 0.0001**
HsylOnsDur + (1|Part)	-1035.5	54.32	1	< 0.0001**
Group + (1|Part)	-982.3	1.13	1	0.29

†As compared to the null model, ExcurDur ∼ 1 + (1|Part) | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for ExcurDur	χ^2	DegF	p†
HsylOnsDur + (1|Part)	—	—	—
HsylOnsDur + Complexity + (1|Part)	3.88	1	0.05
HsylOnsDur + (1|Part)	—	—	—
HsylOnsDur + Group + (1|Part)	11.76	1	0.0006**
HsylOnsDur + Group + HsylOnsDur:Group + (1|Part)	6.71	1	0.01°

†As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

Table 4.22: Table of estimates and standard errors for each group in the Valjevo dialect, from the model ExcurDur ∼ HsylOnsDur + Group + HsylOnsDur:Group + (1|subj). Units in ms.

		β	SE
Intercept	(*ȁmora*)	65.9	15.6
	adinu	44.4	12.0
HsylOnsDur	:*ȁmora*	597.7	67.6
	:*adinu*	-345.1	132.3

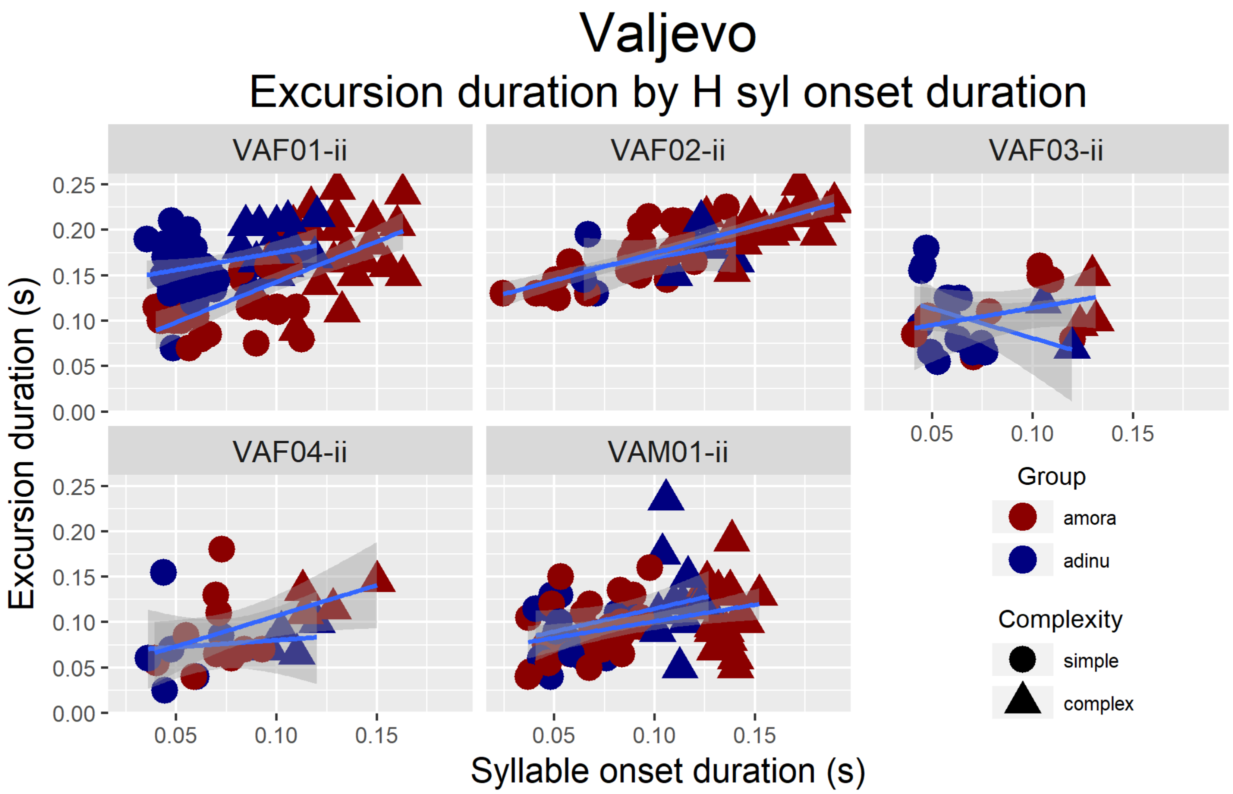


Figure 4.12: Scatter plots, separated by participant, showing the relationship between syllable onset duration (HsylOnsDur) and excursion duration (ExcurDur), with groups differentiated by color (red for *ȁmora*; blue for *adinu*) and individual fit lines. Valjevo dialect.

simply due to the comparatively high attrition rate for Locus 2 words, particularly given the marginal p-value and the high standard error for this estimate. Some participants were also affected more than others. A scatter plot illustrating the relationship between syllable onset duration and excursion duration is provided in Figure 4.12, separated by participant. Note that, for example, participants VAF03-ii and VAF04-ii have particularly few tokens in general, but especially of Locus 2 words, and the resulting large standard error intervals. For participants with more tokens, there is a more consistent relationship between HsylOnsDur and ExcurDur for both word types.

4.2.4 Pitch characteristics: Dialect comparison

4.2.4.1 H achievement (PeakOffset)

Interestingly, despite the robust findings in both previous studies Zec and Zsiga (2018) and Experiment 1, the Belgrade and Valjevo dialects do not exhibit as much separation in the timing of their pitch accents in this experiment. This is particularly true of the falling accents, which in this experiment have nearly the same distribution in both dialects (see Figure 4.13 for an illustration). When considering models with single fixed effects, Dialect only marginally improves the model ($\chi^2(1) = 4.04$, $p = 0.04$; see Table 4.23a). Not surprisingly, both Group ($\chi^2(1) = 657.57$, $p < 0.0001$) and HsylOnsDur ($\chi^2(1) = 412.18$, $p < 0.0001$) significantly improve the model.

Even with Group already included in the model, Dialect as a simple fixed effect does not significantly improve the model ($\chi^2(1) = 4.02$, $p = 0.05$; see Table 4.23b); however, the interaction between Dialect and Group does ($\chi^2(1) = 110.44$, $p < 0.0001$). This is due to the dialect separation in just the Locus 2 group of words, illustrated in Figure 4.13—Valjevo rising accents occur much earlier, frequently retracting into the preceding syllable. The interaction between Dialect and HsylOnsDur does not significantly improve the model ($\chi^2(1) = 0.13$, $p = 0.72$), which indicates that the influence of the duration of the syllable onset on peak offset timing is the same in both dialects.

As in the analyses of each dialect separately, there is a significant interaction between Group and HsylOnsDur ($\chi^2(1) = 44.80$, $p < 0.0001$); increases in syllable onset duration have a smaller effect on peak timing in Locus 2 words. The three-way interaction between Group, Dialect, and HsylOnsDur is not significant ($\chi^2(1) = 0.00$, $p = 1.00$). Thus, although Valjevo rising accent peaks retract into the preceding syllable, the duration of the onset of the phonologically H syllable has the same effect as in Belgrade.

Table 4.23: Comparison of linear mixed effects models for `PeakOffset` (both dialects, only *ȁmora* and *adinu* types).

(a) Single predictor models, compared to the null model.

Model for `PeakOffset`	AIC	χ^2	DegF	p[†]
`Dialect + (1|Part)`	-2387.1	4.04	1	0.04°
`Group + (1|Part)`	-3040.8	657.67	1	< 0.0001**
`HsylOnsDur + (1|Part)`	-2795.3	412.18	1	< 0.0001**

[†] As compared to the null model, `PeakOffset ~ 1 + (1|Part)` | ° < 0.05, * < 0.01, ** < 0.001

(b) Nested model comparisons.

Model for `PeakOffset`	χ^2	DegF	p[†]
`Group + (1|Part)`	—	—	—
`Group + Dialect + (1|Part)`	4.02	1	0.05
`Group + Dialect + HsylOnsDur + (1|Part)`	337.36	1	< 0.0001**
`Group + Dialect + HsylOnsDur + Dialect:HsylOnsDur + (1|Part)`	0.13	1	0.72
`Group + Dialect + HsylOnsDur + Dialect:HsylOnsDur + Dialect:Group + (1|Part)`	110.44	1	< 0.0001**
`Group + Dialect + HsylOnsDur + Dialect:HsylOnsDur + Dialect:Group + Group:HsylOnsDur + (1|Part)`	44.80	1	< 0.0001**
`Group + Dialect + HsylOnsDur + Dialect:HsylOnsDur + Dialect:Group + Group:HsylOnsDur + Dialect:Group:HsylOnsDur + (1|Part)`	0.00	1	1.00

[†] As compared to model immediately above | ° < 0.05, * < 0.01, ** < 0.001

4.2.4.2 Excursion characteristics (`ExcurStart` and `ExcurDur`)

The data from this experiment parallels the data from Experiment 1, suggesting more strongly that there are multiple ways to achieve the same phonetic effect and, crucially, that the Belgrade and Valjevo dialects use different strategies: Belgrade achieves later peaks by starting excursions later and making the excursions longer, while Valjevo excursions only stretch. Individual participants within each dialect also tend to exhibit patterns similar to the overall pattern of the dialect, which suggests that these alignment strategies are dialect-

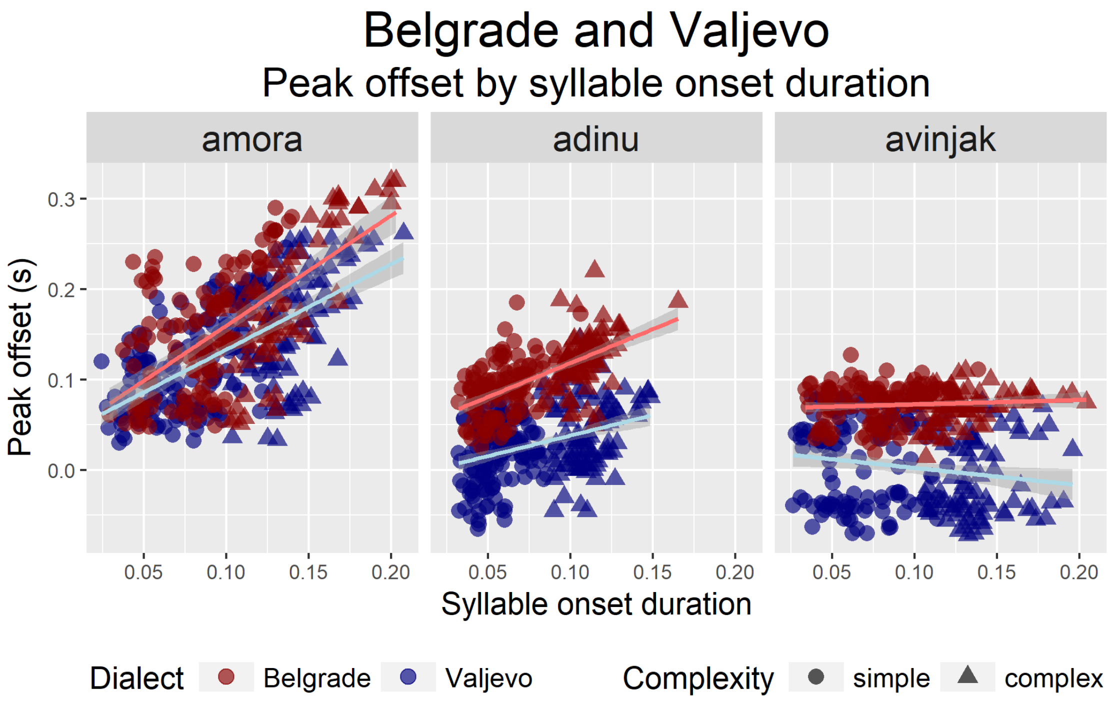


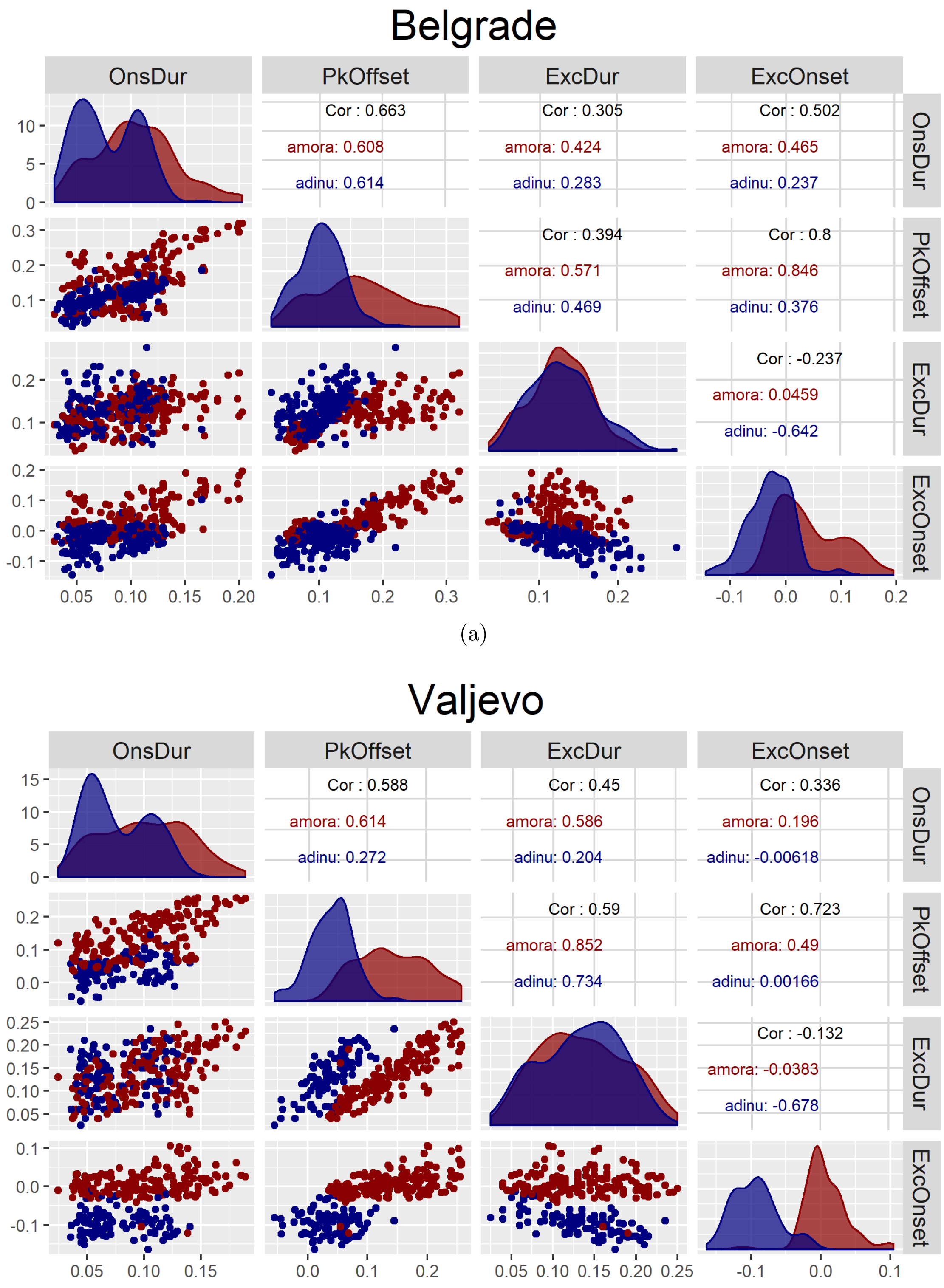
Figure 4.13: Scatter showing the relationship between syllable onset duration and peak offset timing, separated by word group and color-coded according to dialect (Belgrade is red; Valjevo blue).

specific and not random variation by speaker. The patterns for each dialect are illustrated in Table 4.14.

4.3 Conclusions

4.3.1 Summary

This study confirms the findings from Chapter 3, and also shows that rising accents show similar effects when the onset of the H syllable is varied. The results support the Inkelas and Zec (1988) account of rising accents, rather than Smiljanić's (2002) proposal. When the post-tonic syllable in a rising accent word has a longer syllable onset, the peak offset of that rising accent is delayed relative to the beginning of the second syllable. This is not predicted



(a)

(b)

Figure 4.14: Scatter plots showing the relationships between syllable onset duration, peak offset timing, excursion duration, and excursion start time; only Locus 1 (red) and Locus 2 (blue) groups included. Belgrade in (a); Valjevo in (b).

by a L*+H representation of the rising accent, which instead predicts that the peak offset will occur after a set interval of time from the stressed nucleus, regardless of the duration of the second syllable onset.

✗✗**Hypothesis 1.0** (null hypothesis): The H of rising accents is not influenced by either the stressed syllable or the H syllable onset.

✗✗**Hypothesis 1.1**: H is associated to the stressed syllable in rising accents.

✓✓**Hypothesis 1.2**: H is associated to the post-stress syllable in rising accents.

4.3.2 Discussion

The variation between speakers raises some issues. Despite an effort to recruit young Valjevo speakers that had not been in Belgrade for very long, some speakers still showed signs of assimilation to Belgrade timing. Some speakers are more similar to Belgrade speakers in that their peaks occur within the post-tonic syllable (i.e., the H syllable); others exhibit peak retractions where the peak now occurs within the tonic syllable (i.e., the syllable prior to the lexical H), which aligns with the retractions reported by Zec and Zsiga (2018). However, among those that did appear to be converging to Belgrade, the shift did not appear to uniformly affect their accentual system. Specifically, there was much greater variation in the realization of rising accents—for example, there was one speaker that did the first half of the experiment using Belgrade timing, and the second half using Valjevo timing, with no apparent trigger in the environment to encourage one style or the other. Notably, this speaker did *not* also change the timing of their falling accents; they remained uniformly Valjevo-like throughout. This again calls into question the possibility of a single H gesture with timing fully specified.

The classic Valjevo retraction puts pressure on the current conceptualization of the TBU. In autosegmental theory, there is some assumption of simultaneity (Sagey 1986); in Valjevo, this is not always the case. It is true that segmental anchoring does not have such strict requirements for phonetic alignment; however, it is very unlikely that a pitch accent would be posited to be associated to the post-tonic syllable if it is phonetically realized entirely in the tonic syllable (i.e., both the beginning and the end of the pitch movement are in the tonic syllable). Similarly, in Articulatory Phonology, tone gestures should be coordinated with the segmental gestures that make up the relevant prosodic unit. This restriction, in combination with the idea that only gestural onsets can be coordinated with each other, would rule out the retraction exhibited by Valjevo speakers (though, of course, without direct articulatory evidence, the precise coordination of the H gesture is still an open question): a (relatively slow) tone gesture that started even at the same time as the first consonant gesture of the post-tonic syllable should have a target that at least overlaps with the acoustic onset of that syllable.

Additionally, the extreme difference in timing between Valjevo falling and rising accents is troublesome for a gestural account with a single H gesture in the inventory. Unlike for rising accents, falling accents always occur in the syllable they are associated to;[9] this suggests that there is a different coordinative structure for falling and for rising accents. In a gestural model of representation, this would rule out a single H gesture, as timing is part and parcel of a gesture and thus of contrast. If the two accents have different timing relationships with the segmental material, they are by definition distinct gestures. However, given the rising-falling alternations in Serbian, having two completely distinct gestures for the two accents does not seem sufficiently parsimonious.

Also troubling for a uniform H gesture is the result that, in Belgrade, Locus 2 excursions were longer than Locus 1 excursions. This is unexpected under a theory that assumes a uniform H gesture for both falling and rising accents: either excursion duration should be the same for both accents (i.e., word groups) or, as much of the previous data has suggested, dependent on the duration of the syllable onset (or possibly nucleus, or syllable duration in general), which for Locus 2 words is shorter. One possible solution for this is that duration

[9] One can compare Stockholm Swedish (Bruce 1977), which allows the pitch contours to escape the word entirely. The failure of Valjevo Serbian to do this may be some domain restriction—but this remains yet another open question.

is not fully specified in the tone gesture, and it receives its durational information from a combination of target specification and the duration of the segmental gestures in the tone-bearing unit. In addition, there may be a different coordinative schema for rising accents in Valjevo: as has been discussed previously, it is not likely that Valjevo uses the c-center to coordinate tone to its TBU. Other options are thus available—for example, coordinating tone to the onset of the vowel gesture (which could potentially be affected by gestures in the syllable onset).

Interestingly, the Locus 1 and Locus 2 word sets did not exhibit identical behavior. It is possible that this is simply due to the fact that the syllable onsets in the Locus 2 words were on an unstressed syllable. Alternatively, there could be more distinguishing falling and rising accents in Serbian than the H tone's position relative to the stressed syllable.

Chapter 5

Discussion and conclusion

5.1 An articulatory model of tone representation

I have shown in this dissertation that a gestural representation of tone, where time is present in the form of coordinative relationships between gestures, is necessary in order to capture the relationship between the distribution and the realization of a tone. Segmental anchoring (as developed in the AM approach, and described in Section 1.1.1.2) does not predict all the data produced by these studies. Three cases in particular stand out: first, pitch excursions in the Belgrade dialect of Serbian showed an effect of syllable onset structure in addition to the effects of syllable onset duration (Section 3.2.3.1); the segmental anchoring hypothesis only predicts effects of phonetic duration. Second, the onset of pitch excursions in both Thai (Section 2.2.1.5) and Belgrade Serbian (Section 3.2.3.1) occurred after the acoustic beginning of the word, but still within the syllable onset; as argued by D'Imperio et al. (2007), the beginning and end of segmentally anchored tones should approach 0 ms lags with their anchors. Third, and finally, the pitch excursion in Valjevo rising accents lengthens even when the bulk of the segmental material causing the stretch is after the end of the excursion (Section 4.2.3.2); segmental anchoring only predicts tone gesture "stretching" in cases where the segmental material causing the stretch is between the start and endpoint of the gesture.

Thus, I propose a gestural model of tone, where the representation of tone includes specifications for the following three facets:

1. **Tone gestures**;
2. **Segmental gestures**, which by virtue of coordination are grouped into prosodic units, such as the mora and the syllable;
3. The **coordinative relationships** between **(1)** and **(2)**

In what follows, I discuss numbers **(1)** and **(2)**, and their relations **(3)**.

5.1.1 Tone gestures and segmental gestures

In Articulatory Phonology (AP), the gesture is the fundamental unit of contrast (Browman & Goldstein 1989). A tone gesture is an articulatory gesture that is specified with a target, given in terms of F0. Gestures in AP are always specified with a target; for segmental gestures, this is expressed in constriction degree and constriction location. For example, a /t/ is specified with "closed" constriction degree (full contact) and an "alveolar" constriction location with the tongue tip articulator set (Browman & Goldstein 1989). For tone gestures, the target is given as a window of F0 values (after Keating 1990). Strictly speaking, F0 is not an articulatory variable; however, limitations in theoretical and methodological arenas limit the current ability to model F0 like oral gestures (Yi 2017). Furthermore, the linguistic goal of tone gestures is to produce an F0 contour that changes over time, and as such the use of F0 as a tract variable has theoretical grounding (McGowan & Saltzman 1995).

A crucial aspect of this model is that the tone gesture is underspecified for duration, which differs from previous articulatory models. There are several findings in this dissertation that show that tone gestures do not have intrinsic duration. For example, in contour tones in Thai, where each tonal component has a specific gesture (i.e., an H and an L gesture for an HL tone) the duration of the first tone gesture is dependent on the length of the syllable it is realized with; CVN syllables have shorter first tone excursions than CVVN syllables, but the size (in Hz) of the excursions remained constant (Section 2.2.1.5). In Serbian, the

H gesture in both dialects showed "stretching" that roughly paralleled the duration of the syllable onset of the H-bearing syllable (see Section 3.2.3.1 and Section 3.2.4.1). The Valjevo dialect in particular provides a textbook example of a "stretching" tone gesture: the onset of the tone gesture consistently occurs at a point that precedes the beginning of the word, and the target of the tone gesture consistently occurs at some point in the nucleus; as more segmental material is added in between those two points, the gesture is elongated.

The notion of tone as a gesture is relatively new, in comparison with segments, and as such it has not enjoyed much discussion in the AP literature regarding intrinsic duration; however, the assumption has been that stiffness and duration in tone gestures should function as in other articulatory gestures. As discussed in Chapter 1, the crucial difference between AP and featural theories of phonology is the presence of time as a sequence of points in the representation: gestures are viewed as unfolding over time, with specified duration. Thus, gestures with different durations are considered distinct gestures. However, also as discussed, the origin of gestural duration has not been fully explored in AP. One suggestion is the notion of gestural stiffness (Browman & Goldstein 1989; Sorensen & Gafos 2016), where gestures are viewed as simple oscillators with a stiffness parameter like a spring—the stiffer a gesture, the shorter it is. For example, a vowel has a lower stiffness value than a stop consonant, and thus lasts longer.

Durational underspecification of tone gestures contributes to the simplicity of the analysis. With full durational specification, there would have to be a distinct H gesture for every possible syllable onset, or every possible syllable shape; the language would then have to select the corresponding gesture for each tone-bearing syllable. However, this is not a viable analysis: not only would this be highly unparsimonious, one might also expect a certain difficulty with nonce words, or possibly speech errors with mismatched tones and TBUs. Thus, a tone gesture is specified with a target, but not with an absolute duration.

Although tone provides a good example of gestures that are able to stretch, consonants often appear to be ballistic—particularly in cases like a tapped /r/. However, it may not be the case that tone gestures are special; there is some evidence that segmental gestures are also underspecified for duration. In Thai, the first word in a R+F sequence was longer than the first word in an R+R sequence (Section 2.2.2.2); I argue that this is caused by the late extremum in Word 1 caused by dissimilation with Word 2. Thus, segment duration can be affected by planning effects in the tonal domain. Work in Mandarin has also shown that word duration is a major correlate of tone (S. Liu & Samuel 2004), where rising tones are the longest, indicating that segmental gestures can stretch to accommodate tone contours even without pressure from other words in a phrase. Thus, gestural stiffness may have an effect on how much a gesture is able to stretch, but does not itself determine timing.

5.1.2 The articulatory TBU

In my model, I introduce the concept of an articulatory TBU, which is the gestural "constellation" forming a prosodic unit (such as the mora or the syllable) that a specific tone gesture (or series of tone gestures, as proposed for Mandarin by Gao (2008)) is coordinated with. With a constellation of gestures forming a prosodic unit I am invoking the "co-selection sets" proposed by Tilsen (2016). The tone gesture gets durational information from the other gestures in its articulatory TBU.

When claiming that a tone is lexically associated with a unit—as demonstrated by either licensing, as in Thai (described in Section 1.2.1), or phonological processes that categorically target tones, as in Serbian (described in Section 1.2.2)—then the tone gesture is coordinated in some way to the segmental gestures that themselves are coordinated to make up that unit. Thus, the categorical patterns exhibited by tones make reference to the binary existence vs. non-existence of a coordinative relationship between a tone gesture and an articulatory TBU.

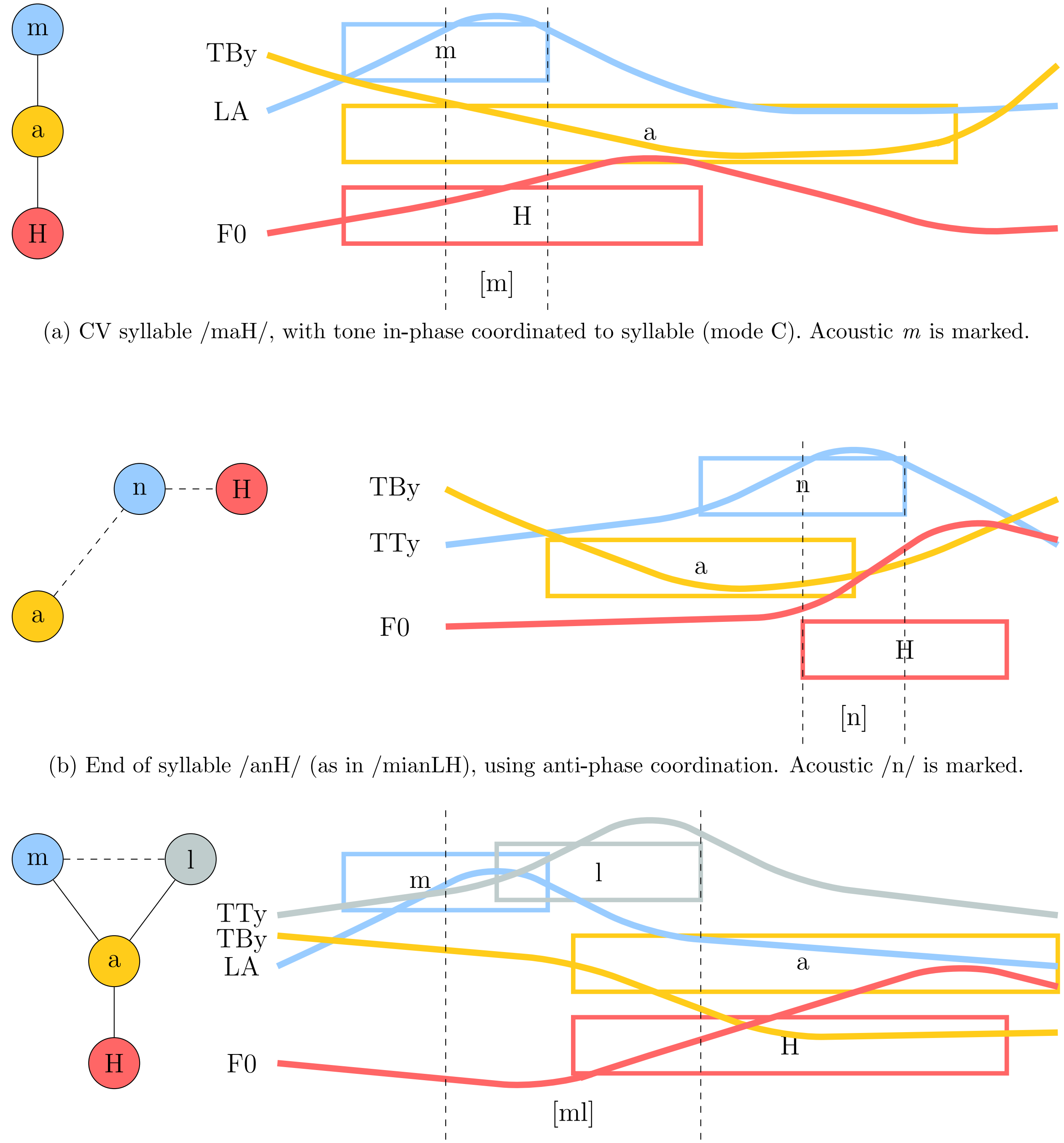
The specific realization of these tones is also derived from the coordinative relationship: different modes of coordination between a tone gesture and its articulatory TBU result in different timing patterns. In this dissertation, I have shown that tone gestures can be coordinated to segmental gestures in all the same ways that segmental gestures can be coordinated to each other, contrary to the c-center hypothesis for tone (refer to Section 1.1.2.1

Table 5.1: A list of the coordinative modes for lexical tone that reference gestural onsets displayed in this dissertation, and the acoustic consequences.

	Coordinative mode	**Acoustic consequence**	**Example**
A.	In-phase	Tone gesture starts at or slightly before the left edge of the articulatory TBU	Valjevo falling
B.	Anti-phase	Tone gesture starts at or slightly after left edge of the articulatory TBU	Thai (T2)
C.	C-center: consonant-like	Tone gesture starts slightly after (∼15 ms) the left edge of the articulatory TBU if only one C; increasingly later- as more C gestures are added to syllable onset	Thai (T1)
D.	C-center: vowel-like	Syllable onset gestures displace in both directions away from tone gesture onset	Belgrade

for details). The c-center hypothesis predicts that tone gestures can only be coordinated with segmental gestures as an additional consonant-like gesture; however, I found that tone gestures can be coordinated in-phase (as in the Valjevo falling accent), anti-phase (as in the second tone gesture of a contour tone in Thai), and even as a vowel-like gesture in a c-center structure (as in the Belgrade accents; see Table 5.1). Rather than assuming that the coordination of tone gestures is more restricted than that of segmental gestures, I assume that both types of gestures can be coordinated in the same way. The coordinative modes proposed in this model exhaust the list of possible modes generally posited in AP.

Some schematized examples of coordinative relationships from Table 5.1 and their acoustic consequences are illustrated in Figure 5.1 for sample timing differences. On the left of each figure is a coordinative diagram, which shows the coordinative relationships between the gestures: solid lines between nodes indicates in-phase relationships, and dashed lines indicate anti-phase relationships. To the right is a gestural score with corresponding gestural trajectories: each gesture has a rectangle which indicates its active period, with gestural onset on the left edge and gestural release on the right edge. The gestural trajectories line



(a) CV syllable /maH/, with tone in-phase coordinated to syllable (mode C). Acoustic *m* is marked.

(b) End of syllable /anH/ (as in /mianLH), using anti-phase coordination. Acoustic /n/ is marked.

(c) Syllable /mlaH/, coordinating vowel and H gesture to the c-center (as defined in Browman and Goldstein 1988). Acoustic *ml* is marked.

Figure 5.1: Schematics for coordinative modes A, B, and D. In the coordinative diagrams, solid lines represent in-phase coordination; dashed represent anti-phase. In the gestural scores, the left edge of the rectangle indicates gestural onset, and the right edge indicates gestural release. Acoustic boundaries for the consonants are marked (corresponds to the interval between gestural target and release).

up with these edges. In addition, the acoustic boundaries of the consonants are marked with dashed vertical lines: here, the acoustic beginning corresponds to the target achievement of a gesture (i.e., the achievement of closure), and the acoustic end corresponds to the gestural release. Thus, when viewing the F0 trajectory, it is possible to see how the onsets of F0 gestures, though in-phase coordinated with consonantal gestures, appear to start earlier than the consonant (in Figure 5.1a); meanwhile, the onset of the F0 gesture in anti-phase coordination with a coda consonant approximately lines up with the acoustic edge of the consonant (see Figure 5.1b).

In addition to the typically assumed set of coordinative modes that refer to gestural onsets, I also propose gestural *release* coordination—that is, the deliberate timing of a tone gesture release (i.e., the gestural "shoulder") to the onset or release of a second gesture, or to the c-center of a constellation of gestures. These modes are listed in Table 5.2, along with their acoustic consequences. The patterns of the Valjevo rising accent suggest that tone can be coordinated in this way: as shown in Section 4.2.3.2, H gesture excursions were affected by the duration of the syllable onset of the H-bearing syllable (longer syllable onsets correlated with longer pitch excursions), even though the pitch excursion started well before that H-bearing syllable, and ended early on in that syllable's onset. This indicates that the H gesture is still receiving timing information from the H-bearing syllable, and as such, is coordinated in some way to that syllable. The early alignment can be achieved by coordinating the release of the H gesture with the H-bearing syllable.

Although I propose that release timing is available for tone gestures, at this juncture I am not arguing that all gestures are coordinated at both the onset and the release. Some durational information can come from the target of the gesture itself; for tone gestures, more extreme F0 targets take longer. This relationship between gestural target and duration can be seen when comparing the timing of Belgrade and Valjevo dialects. As described in Section 1.2.2, Valjevo peaks occur earlier than Belgrade peaks; the pitch excursions also start earlier and are shorter in Valjevo than in Belgrade (see Section 3.2.5.1). Importantly, the excursions

Table 5.2: A list of the coordinative modes that refer to gestural releases for lexical tone displayed in this dissertation, and the acoustic consequences.

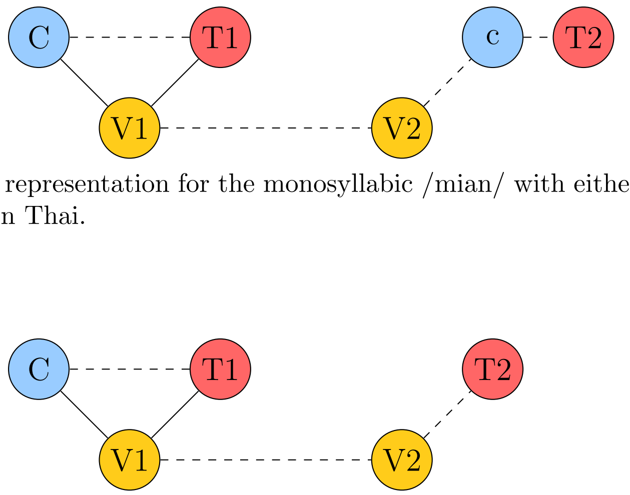
	Coordinative mode	**Acoustic consequence**	**Example**
E.	Release to onset	(Can be conceptualized as a restatement of the anti-phase relationship: 180° phase to 0° phase)	(not in this dissertation)
F.	Release to release	Tone gesture target achieved at acoustic beginning of second gesture; gestures may not have started at the same time	(not in this dissertation)
G.	Release to c-center	Tone gesture starts well before the left edge of the articulatory TBU; the consonants that form the c-center it is coordinated with displace in both directions from the target of the tone gesture	Valjevo rising

in Valjevo are also smaller in Hz space, and, impressionistically, the overall pitch trajectories of Valjevo Serbian are markedly flatter than those in Belgrade Serbian. Thus, it is possible to generate the earlier peaks of the Valjevo dialect from the less extreme target F0 specification, without additionally coordinating the release of the H gesture to a point in the syllable.

5.1.3 Language-specific examples

5.1.3.1 Thai

The models for Thai are illustrated in Figure 5.2. As detailed in Section 1.2.1, contour tones (tones with two tone gestures) are limited to words with two sonorant moras. In the gestural model, a sonorant must be present in a moraic constellation of gestures to license coordination to a tone. Furthermore, only one tone gesture is permitted per moraic unit. This produces the distributional patterns described by Morén and Zsiga (2006).



(a) Proposed representation for the monosyllabic /mian/ with either a Rising or Falling tone in Thai.

(b) Proposed representation for the monosyllabic /mia/ with either a Rising or Falling tone in Thai.

Figure 5.2: Model of tonal representation in Thai.

The timing of the tone gestures relative to the segments is reflected in the specific coordinative modes recruited to coordinate the tone gestures. In Thai, the timing of the first tone gesture of a contour tone (T1) relative to the acoustic start of the word is comparable to the timing reported in Karlin 2014, where T1 acts as the second consonant gesture in a c-center structure (Section 2.2.1.5). This configuration is illustrated in Figure 5.2 (reproduced from Chapter 2). T1 is coordinated in a c-center structure with the syllable onset and the first vowel (coordinative mode C), which generates the lag from the left edge of the syllable to the onset of the tone gesture; T2 is anti-phase coordinated to the last item in the second mora (coordinative mode B), which generates a tonal extremum near the acoustic beginning of the second mora (i.e., near the target of the initial gesture of the second mora).

This model can also address the timing differences between CV_1V_2 and CV_1V_2N syllables. Recall that both vowels of the nucleus in a diphthong are shortened, and that T2 does not move earlier to match the shifted moraic edge. This is captured in the model by the fact that T2 is coordinated to the last member of the second mora unit, which generates a delay with additional material included in the second mora (such as a non-moraic coda). This also addresses the patterns found in Karlin 2014, where it was reported that the time lag between the onsets of a coda *n* gesture and the T2 gesture varied depending on the position of the /n/: when the /n/ was moraic (i.e., in /man/ syllables), there was a larger onset-to-onset lag between the gestures. This parallels the differences found between the /mia/ and /mian/ syllables, specifically the overall shortening of both vowels in the nucleus—as more gestures are added to the overall syllable constellation, the lags between them decrease. Thus, at this juncture it is not necessary to posit a distinct relationship between T2 and the last member of the second mora constellation when there is a non-moraic coda in order to address the compressed time lags.

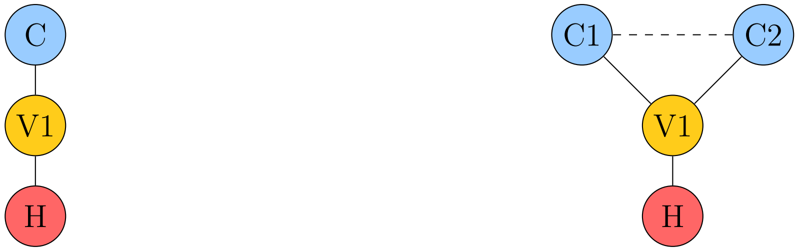
5.1.3.2 Serbian

In Chapter 3, I show that the H tone gesture for falling accents in the Belgrade dialect of Serbian displays timing patterns that are most simply generated by the c-center structure, though with the tone gesture itself acting as a vowel-like gesture (coordinative mode D; see Table 5.1), rather than a consonant-like gesture. With tone as a vowel-like gesture, the syllable onset consonants would displace away from the onset of the H gesture in both directions, as was shown in Chapter 3. 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 a two-dimensional format Figure 5.3.[1]

In contrast, the falling accent in the Valjevo dialect displays timing patterns that are most simply generated by an in-phase coordinative relationship between the H gesture and the first consonant gesture of the syllable onset (coordinative mode A; see Table 5.1). This representation is illustrated in Figure 5.4.[2] Here it is important to note that in my model, the entire gestural "constellation" that makes up the distributional TBU informs the timing of the tone gesture, not just the gesture(s) that the tone gesture is immediately coupled to.

[1] Perhaps this would be more clearly represented in three dimensions, to better depict the relationship between the H gesture and the consonant gestures. However, if both the H and V gestures are used as anchors, they are definitionally in-phase with each other and thus equally affected by the consonant gestures.

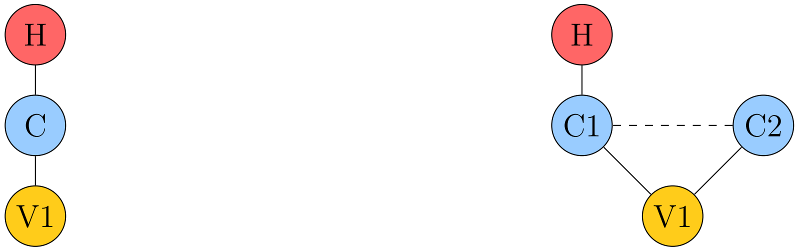
[2] Note that the vertical positions of gestures in these coordinative diagrams does not signify anything about their relationships with the other gestures—vertical divisions are formed largely on the basis of aesthetic readability, such as avoiding crossing lines and crowded nodes.



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

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

Figure 5.3: Model of tonal representation for Belgrade Serbian falling (and rising) accent. This model produces the bidirectional displacement of the consonant gestures away from the tone gesture.



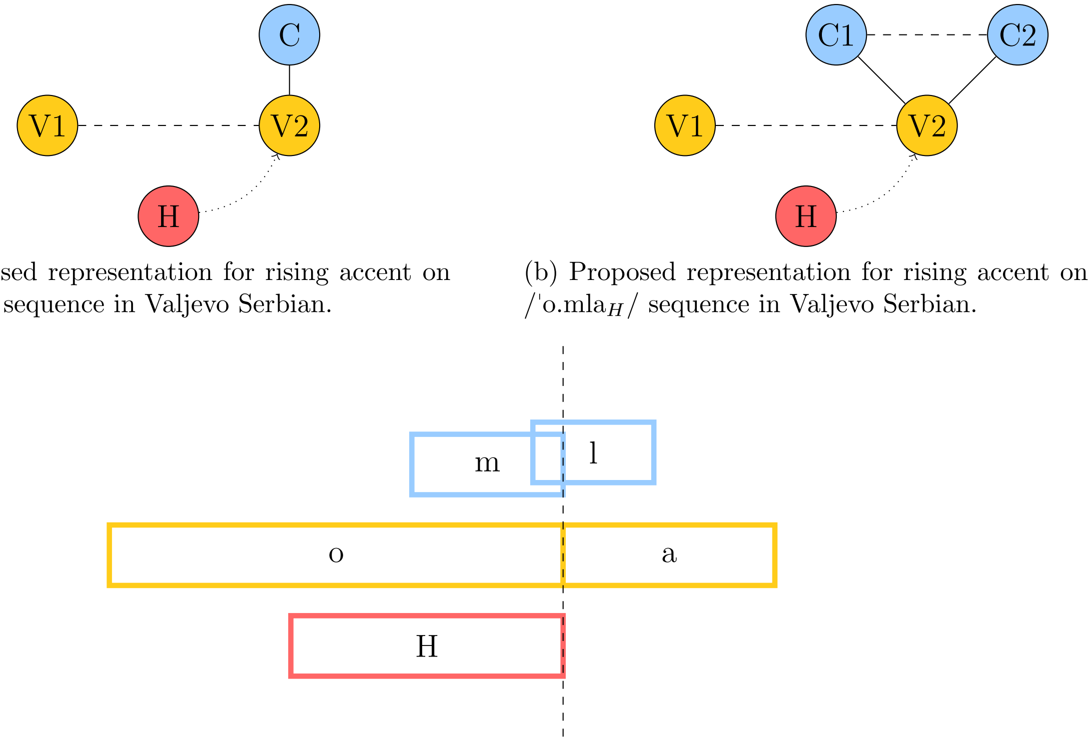
(a) Proposed representation for /ma/ syllable with a falling accent in Valjevo Serbian.

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

Figure 5.4: Model of tonal representation for Valjevo Serbian falling accent. Here the H gesture is only in-phase coordinated to the onset consonant. This produces the consistent timing of the onset of the H gesture, regardless of the complexity of the syllable onset.

The data from the Valjevo dialect provides the most compelling evidence: the H gesture is in-phase coordinated with the first consonant gesture of the syllable onset, but still increases in duration when moving from /m/ to /mr/ (Section 3.2.4.1), which indicates that the /r/ gesture provides timing information to the tone gesture despite not being directly coordinated with it.

Finally, an additional model configuration is necessary for the rising accent in Valjevo Serbian. As mentioned above in Section 5.1.2, the H gesture in Valjevo rising accents receives timing information from the post-stress syllable, which indicates that it is coordinated in



(a) Proposed representation for rising accent on /ˈo.ma$_{H}$/ sequence in Valjevo Serbian.

(b) Proposed representation for rising accent on /ˈo.mla$_{H}$/ sequence in Valjevo Serbian.

(c) Gestural score that corresponds to (b): 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 5.5: Model of tonal representation for Valjevo Serbian rising accent, showing both the stressed and post-stress (H-bearing) syllable. The bending dotted line represents release timing (release of the H gesture, here timed to the c-center of the following syllable).

some way to the segmental gestures of the second syllable. A coordinative relationship between the H gesture and the c-center of the post-stress syllable produces the phonological association, as well as the early alignment. The most plausible perspective is that the *release*, not the onset, of the H gesture is coordinated with the gestures of the post-stress syllable—possibly to the c-center of that syllable (coordinative mode G; see Table 5.2). The use of the c-center as a reference point would generate the slightly flatter relationship between syllable onset duration and peak offset in rising accents as compared to falling accents: the c-center moves later as the syllable onset gets longer, but by a smaller amount. This structure is modeled in Figure 5.5.

5.1.4 Differences from previous gestural models

This model predicts a wider range of alignment possibilities than the c-center hypothesis for tone, and argues that tone gestures use the full set of coordinative modes available to segmental gestures. This addition is necessary in order to predict the data produced by these studies. As discussed in Chapter 1, *articluatory* simultaneity (i.e., direct coordination between gestures) does not necessarily correlate with *acoustic* simultaneity. The example provided was that of CV syllables: CV syllables consist of a consonant gesture in-phase coupled with a vowel gesture and are thus articulatorily simultaneous; however, the acoustic signal suggests that they are sequential (Browman & Goldstein 1988).

The availability of different coordinative regimes—particularly in combination with different targets—addresses these different degrees of acoustic overlap. For example, the c-center structure as described by Karlin (2014) results in an apparent "delay" of pitch gesture onset, where there is a delay between the acoustic start of the syllable onset and the start of the first pitch gesture. Similarly, in F+F sequences and R+R sequences, the intervening midpoint (i.e., where the first tone gesture of the second word starts) occurs after the acoustic beginning of the second word. It thus may be tempting to expand the TBU "window" of timing past the syllable itself. However, in articulatory terms, the simultaneity is still there—the entire constellation for the syllable has started, but the implementation simply has not yet gotten to the next tone gesture. This can be compared to the segmental lengthening effect found in Chapter 2 for R+F sequences: the segment is in fact lengthening to accommodate the second tone gesture; the tone gesture is not allowed to simply bleed into the next TBU (despite the fact that it is, phonetically, one long upward excursion).

Though not addressed in this dissertation, the number of tone gestures per licensing unit is an additional variable to consider in the future. For example, as mentioned in Chapter 1, a mora can license multiple tone targets in Yoloxóchitl Mixtec; each whole tone contour is aligned to the mora (DiCanio et al. 2014), but each tone target in a contour must have some other reference unit of timing. A similar case was presented by Gao (2008), where Tone 4 is presented as an HL contour, with the L coordinated only to the H, and only the H coordinated to the segmental gestures of the syllable. These two cases suggest that tone gestures can be coordinated to each other (anti-phase coordination), and do not have to be singly linked to segmental gestures.

5.1.5 Beyond lexical representation

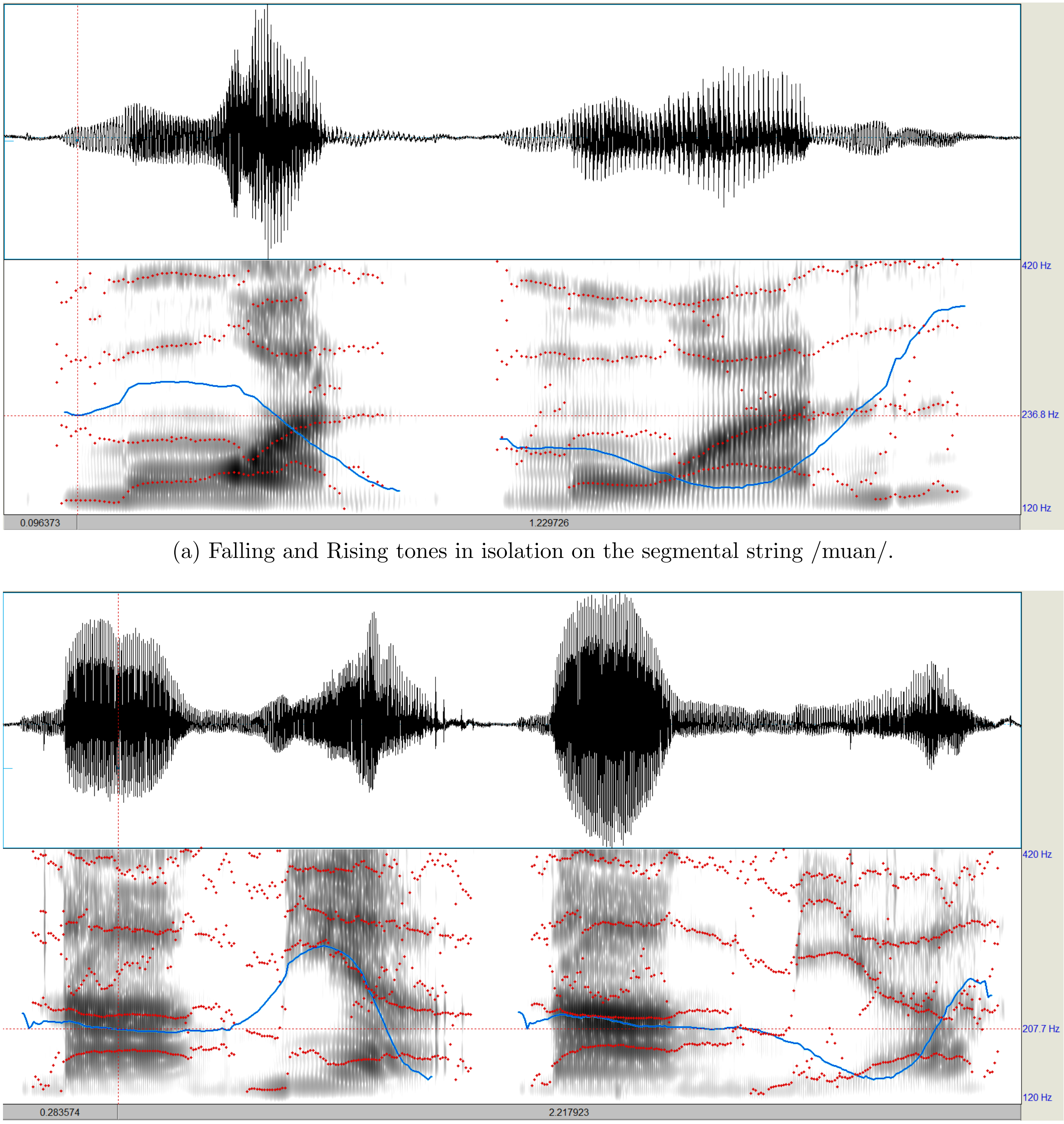
As described in the introduction and examined in-depth in Chapter 2, the units that license and align tone are not necessarily the same. In Thai, tone targets are licensed by the mora, but the precise alignment of extrema is likely influenced by the timing of extrema in other tones in the system, where tones are lexically assigned to the syllable, not the mora. The Serbian study did not specifically focus on what unit drives pitch timing, but previous work has shown effects of the mora on peak retraction (Zsiga & Zec 2013), while syllables are the domain of lexical stress and pitch assignment. However, the results presented in Chapters 3 and 4 show that the syllable onset must be included in the TBU for alignment. Another more extreme example from beyond this dissertation is the licensing of question word intonation in Tashlhiyt Berber, where an H target is licensed by a question word but not consistently aligned with any particular landmark (Bruggeman, Roettger, & Grice 2017).

In addition to these two tonal domains, there must also be some broader domain that is related to the planning of tone. As demonstrated in Chapter 2, tones in adjacent words can have an effect on the realization of a tone—e.g., in a two-word sequence in Thai, the tone of Word 2 affects both the timing and the height of the tone in Word 1. This indicates that there is a planning domain that can encompass multiple alignment domains. Again, although this dissertation did not specifically examine this domain in Serbian, peak retraction as caused by boundary tones also suggests the existence of a planning domain for tone (Zsiga & Zec 2013). More generally, such a domain is well-supported by the phenomenon of tonal crowding (Arvaniti et al. 1998, 2006; Prieto & Torreira 2007), where the precise timing of accentual peaks is influenced by the proximity of later boundary accents.

This model is similar to that proposed by Keating (1990), where targets are specified not as single ideal points, but "windows" of realization, as well as the target-interpolation models in AM (Pierrehumbert 1980). However, an underspecification of timing with specific timing information provided by the phonetic realization should specifically not be seen as a return to universal phonetics. That is, the existence of a gestural target does not predict the precise realization for each language. Interactions of this sort for nasality are discussed at length by Cohn (1990); additionally, one might predict more variation cross-linguistically in segments that have more conflicting pressures—for example, contour tones in a language with length contrast in vowels (where pressure to achieve multiple tonal targets comes to a head with preserving length distinctions, cf. Remijsen and Ladd 2008 for Dinka), or voiced obstruents (where pressure to achieve a laryngeal target for voicing comes to a head with pressure to achieve a high degree of closure, cf. Bjorndahl 2018 for voiced fricatives).

Although this dissertation does not directly provide evidence for target-related timing, some anecdotal evidence from data not included in the analysis is highly suggestive of a balance between timing and target realization (illustrated in Figure 5.6). First, Thai provides evidence for a primacy of timing and an active T2 gesture (rather than a simple release of T1): when produced in isolation, words with contour tones do not consistently start from the middle of the F0 range (as they do in the experiment, where the target words are flanked by mid tone words); rather, they start closer to their targets. However, the F0 contour simply follows a relatively shallow trajectory to the first target before changing direction at roughly the same point as words in context. Thus, even though the F0 target could very easily be reached early in the word, there is some notion of gestural timing (perhaps contrast timing) that prevents this from happening. Note the difference in "peakiness" of the tones in context in Figure 5.6, compared to the tones in isolation.

Just as an underspecified tone gesture accounts for stretching phenomena, an underspecified segmental gesture accounts for phenomena such as tone-related segmental lengthening, or even in the creation of segmental material for tone (Grice, Röttger, Ridouane, & Fougeron



(a) Falling and Rising tones in isolation on the segmental string /muan/.

(b) String /mia/ with a Rising tone and a Falling tone in context, following mid-tone /naang/.

Figure 5.6: Examples of Thai contours in isolation (a) and in context (b), provided by the same speaker in the acclimation period before the experiment.

2011; Grice, Savino, Caffò, & Roettger 2015). What is in the phonetics then is not specific alignment rules, but language-specific rules for "optimizing" output: restrictions on what tones can be truncated or flattened; what segments can be truncated or lenited; the extent to which overlap is allowed; what contrasts need to be preserved and how they are indicated; etc. Some cross-linguistic differences in this respect are summarized by Gibson (2013) for tone languages, and by Ladd et al. (1999) for issues of intonation. Thus, rather than phonetic "mapping" rules providing idiosyncratic timing differences with no reference to the system from whence they come, the phonological system as a whole is referenced in creating the acoustic output.

5.2 Further implications for the representation of tone

The results from this investigation point in a number of future directions. The first direction is to test the predictions of the given models for the languages investigated in this dissertation. Although acoustic data provides many insights on how tone is represented, some timing relationships are obscured. For example, the onset of the first tone gesture of Rising tones in Thai occurs later than the first tone gesture of Falling tones. This does not mean that the two tone gestures are coordinated differently, however, as the location c-center as described by Browman and Goldstein (1988) depends on the duration of the gesture: a longer second consonant gesture would correspond with a later c-center. As T1 in Rising tones is longer than T1 in Falling tones, probing the exact timing of the vowel gesture would be informative. Preliminary analysis of a slightly modified articulatory study on Thai does suggest that the c-center is the mode of coordination for the first tone gesture in Thai contour tones, confirming the findings from Karlin 2014, as well as the predictions of the model presented here.

It would also be informative to examine the issue of timing pressure provided by other tones in the phonological space. For example, I proposed that Thai tone timing is driven in part by gestural coordination, but with an additional overlay of systemic pressures. That is, the target of the first tone gesture in Thai contour tones approximates some kind of proportional target in the syllable, the precise position of which is influenced by the shape of other tones in the Thai tonal space. Given that the individual speakers showed a reasonable amount of variation in their realization of the Falling and Rising tones, a careful investigation of each speaker's tonal space—in terms of perception as well as production—would provide insight on the role of contrast in realization.

There are also further studies to do regarding the c-center patterning of some tone gestures. We may speculate that the coordinative timing of lexical tone is related to its historical origin. Tonogenesis in Asian languages is frequently tied to voicing differences in the initial consonant (Thurgood 2002); it is possible that the laryngeal coordination used to create voicing distinctions was shifted to tone in the reanalysis. Other tonal features also arise from codas—for example, the tones in Burmese have been analyzed as coming from glottal finals (Gruber 2011); the coordination of T2 in Thai is reminiscent of the coordination of coda consonants (as described in Marin and Pouplier 2010). The preservation of these patterns across centuries of tonal shift would be remarkable; possibly a large-scale typological study of languages with known consonantal effects on F0 could shed some light on this issue. However, such an explanation does not address the distinction between the Belgrade and Valjevo dialects. Future work on other languages may also investigate the possibility of a c-center structure with the tone gesture behaving in diverse ways:

1. Similarly to the V gesture (as suggested by Belgrade Serbian)
2. Similarly to a second C gesture (as suggested by Thai, Mandarin)
3. Taking the place of the first C gesture (as possibly suggested by Valjevo Serbian—not likely, but possible)

That the c-center is used at all for coordinating tone also provides additional avenues of future research. It has been suggested that the existence of tone in Mandarin syllables is what drives the out-of-phase timing of the onset C and V gesture (Mücke et al. 2011), but this hypothesis has not yet been thoroughly investigated. Though Gao (2008) did not examine toneless syllables, the prediction is that the absence of tone would be accompanied by a phase shift back to in-phase for the CV in toneless syllables. Serbian (and other languages that are sparsely specified for lexical tone) would be another fertile ground for such an investigation, as the majority of syllables are toneless. One could thus, for example, compare the consonant-to-vowel timing in the stressed syllable of rising accents (where there is no associated tone gesture) to the stressed syllable of falling accents (where there is an associated tone gesture).

References

Abramson, A. S. (1962). The vowels and tones of Standard Thai: Acoustical measurements and experiments. *International journal of American linguistics*, *28*(2), x–146.

Abramson, A. S. (1978). Static and dynamic acoustic cues in distinctive tones. *Language and speech*, *21*(4), 319–325.

Abramson, A. S. (1979). The coarticulation of tones: An acoustic study of Thai. *Studies in Tai and Mon-Khmer phonetics and phonology in honour of Eugenie JA Henderson*, 1–9.

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.

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.

Arvaniti, A., Ladd, D. R., & Mennen, I. (2006). Phonetic effects of focus and "tonal crowding" in intonation: Evidence from Greek polar questions. *Speech Communication*, *48*(6), 667–696.

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.

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.

Bjorndahl, C. (2018). *(manuscript) A story of /v/: voiced spirants in the obstruent-sonorant divide* (Unpublished doctoral dissertation). Cornell University.

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.

Browman, C. P., & Goldstein, L. (1989). Articulatory gestures as phonological units. *Phonology*, *6*(02), 201–251.

Browman, C. P., & Goldstein, L. (1990). Gestural specification using dynamically-defined articulatory structures. *Status report on speech research*, 95.

Browman, C. P., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, *49*(3-4), 155–180.

Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Les Cahiers de l'ICP. Bulletin de la communication parlée*(5), 25–34.

Browne, E. W., & McCawley, J. D. (1965). Srpskohrvatski akcenat. *Zbornik za filologiju i lingvistiku*, *8*, 147–151.

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

Bruggeman, A., Roettger, T. B., & Grice, M. (2017). Question word intonation in tashlhiyt berber: Is 'high'good enough? *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, *8*(1).

Byrd, D. (1995). C-centers revisited. *Phonetica*, *52*(4), 285–306.

Chitoran, I., Goldstein, L., & Byrd, D. (2002). Gestural overlap and recoverability: Articulatory evidence from Georgian. *Laboratory phonology*, *7*, 419–447.

Cohn, A. C. (1990). *Phonetic and phonological rules of nasalization* (Unpublished doctoral dissertation). University of California Los Angeles.

DiCanio, C., Amith, J., & García, R. C. (2014). The phonetics of moraic alignment in Yoloxóchitl Mixtec.

Dilley, L. C., Ladd, D. R., & Schepman, A. (2005). Alignment of L and H in bitonal pitch accents: testing two hypotheses. *Journal of Phonetics*, *33*(1), 115–119.

D'Imperio, M., Espesser, R., Loevenbruck, H., Menezes, C., Nguyen, N., & Welby, P. (2007). Are tones aligned with articulatory events? evidence from Italian and French. *Papers in Laboratory Phonology 9*, 577–608.

Gafos, A. I. (2002). A grammar of gestural coordination. *Natural Language & Linguistic Theory*, *20*(2), 269–337.

Gandour, J., Potisuk, S., Dechongkit, S., & Ponglorpisit, S. (1992). Anticipatory tonal coarticulation in Thai noun compounds. *Linguistics of the Tibeto-Burman Area*, *15*(111-124).

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

Gibson, M. (2013). Lexical tone, intonation, and their interaction: a scopal theory of tune association.

Goldsmith, J. A. (1976). *Autosegmental phonology* (Unpublished doctoral dissertation). Indiana University (Bloomington).

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).

Grice, M., Röttger, T., Ridouane, R., & Fougeron, C. (2011). The association of tones in Tashlhiyt Berber. In *Proc. 17th int. conf. phon. sci* (pp. 775–778).

Grice, M., Savino, M., Caffò, A., & Roettger, T. B. (2015). The tune drives the text–schwa in consonant-final loan words in Italian. *The Scottish Consortium for ICPhS*.

Gruber, J. F. (2011). *An articulatory, acoustic, and auditory study of Burmese tone* (Unpublished doctoral dissertation). Georgetown University.

Hermes, A., Grice, M., Mücke, D., & Niemann, H. (2008). Articulatory indicators of syllable affiliation in word initial consonant clusters in italian. *Proceedings of the International*

Seminar on Speech Production, *8*(1), 433-436.

Hermes, A., Ridouane, R., Mucke, D., & Grice, M. (2011). Kinematics of syllable structure in Tashlhiyt Berber: The case of vocalic and consonantal nuclei. In *9th international seminar on speech production.* (pp. 1–6).

Hodge, C. T. (1946). Serbo-Croatian phonemes. *Language*, *22*(2), 112–120.

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.

Jakobson, R. (1931). *Die Betonung und ihre Rolle in der Wort-und Syntagmaphonologie...* Státní tiskárna.

Karlin, R. (2014). The articulatory TBU: Gestural coordination of lexical tone in Thai. *Cornell Working Papers in Phonetics and Phonology*.

Keating, P. A. (1990). The window model of coarticulation: articulatory evidence. *Papers in laboratory phonology I*, *26*, 451–470.

Kreitman, R. (2012). On the relations between [sonorant] and [voice]. In P. Hoole, L. Bombien, M. Pouplier, C. Mooshammer, & B. Kühnert (Eds.), *Consonant clusters and structural complexity* (chap. 2). Walter de Gruyter.

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., 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.

Ladd, D. R., & Schepman, A. (2003). "Sagging transitions" between high pitch accents in English: experimental evidence. *Journal of phonetics*, *31*(1), 81–112.

Leben, W. R. (1973). *Suprasegmental phonology.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

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

Liberman, M. Y. (1975). *The intonational system of english.* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Liu, S., & Samuel, A. G. (2004). Perception of mandarin lexical tones when f0 information is neutralized. *Language and speech*, *47*(2), 109–138.

Liu, X. (2014). Mandarin neutral tone—does it change target. *International Journal of Language and Linguistics*, *2*(1), 5–18.

Magner, T. F., & Matějka, L. (1971). *Word accent in modern Serbo-Croatian.* Penn State University Press.

Marin, S. (2013). The temporal organization of complex onsets and codas in Romanian: A gestural approach. *Journal of Phonetics*, *41*(3-4), 211–227.

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).

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

McGowan, R. S., & Saltzman, E. L. (1995). Incorporating aerodynamic and laryngeal components into task dynamics. *Journal of Phonetics*, *23*(1-2), 255–269.

Morén, B., & Zsiga, E. (2006). The lexical and post-lexical phonology of Thai tones. *Natural Language & Linguistic Theory*, *24*(1), 113–178.

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.

Mücke, D., Nam, H., Hermes, A., & Goldstein, L. (2011). Coupling of tone and constriction gestures in pitch accents. *Consonant Clusters and Structural Complexity*.

Myers, S. (1999). Tone association and f0 timing in Chichewa. *Studies in African Linguistics*, *28*(2).

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

Nitisaroj, R. (2006). Thai tonal contrast under changes in speech rate and stress. *Speech Prosody 2006, Dresden, Germany, May 2-5 2006*.

Peirce, J. W. (2007). Psychopy—psychophysics software in Python. *Journal of neuroscience methods*, *162*(1), 8–13.

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.

Pittayaporn, P. (to appear). Phonetic and systemic biases in tonal contour changes in Bangkok Thai.

Potisuk, S., Gandour, J., & Harper, M. P. (1997). Contextual variations in trisyllabic sequences of Thai tones. *Phonetica*, *54*(1), 22–42.

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.

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.

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.

R Core Team. (2017). 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.

Remijsen, B., & Ayoker, O. G. (2014). Contrastive tonal alignment in falling contours in Shilluk. *Phonology*, *31*(03), 435–462.

Remijsen, B., & Ladd, D. R. (2008). *The tone system of the luanyjang dialect of dinka.* Walter de Gruyter GmbH & Co. KG.

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

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

Shih, S., & Inkelas, S. (2014). A subsegmental correspondence approach to contour tone (dis) harmony patterns. In *Proceedings of the annual meetings on phonology* (Vol. 1).

Shih, S., & Inkelas, S. (to appear, 2019). Autosegmental aims in surface optimizing phonology. *Linguistic Inquiry*.

Silverman, K., & Pierrehumbert, J. (1990). The timing of prenuclear high accents in English. *Papers in laboratory phonology I*, 72–106.

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.

Sorensen, T., & Gafos, A. (2016). The gesture as an autonomous nonlinear dynamical system. *Ecological Psychology*, *28*(4), 188–215.

Thurgood, G. (2002). Vietnamese and tonogenesis: Revising the model and the analysis. *Diachronica*, *19*(2), 333–363.

Tilsen, S. (2016). Selection and coordination: The articulatory basis for the emergence of phonological structure. *Journal of Phonetics*, *55*, 53–77.

Tilsen, S., Zec, D., Bjorndahl, C., Butler, B., L'Esperance, M.-J., Fisher, A., … Sanker, C. (2012). A cross-linguistic investigation of articulatory coordination in word-initial consonant clusters. *Cornell Working Papers in Phonetics and Phonology*.

Wagner, P., & Mandić, J. (2005). Are pitch contour and quantity independent distinctive features in Bosnian Serbian? *IKP Arbeitsberichte NF/IKP Working Papers, New Series*, *14*.

Xu, Y. (1997). Contextual tonal variations in Mandarin. *Journal of phonetics*, *25*(1), 61–83.

Xu, Y. (2001). Fundamental frequency peak delay in Mandarin. *Phonetica*, *58*(1-2), 26–52.

Xu, Y. (2004). The PENTA model of speech melody: Transmitting multiple communicative functions in parallel. *Proceedings of from sound to sense*, *50*, 91–96.

Xu, Y. (2005). Speech melody as articulatorily implemented communicative functions. *Speech communication*, *46*(3), 220–251.

Xu, Y., & Sun, X. (2002). Maximum speed of pitch change and how it may relate to speech. *The Journal of the Acoustical Society of America*, *111*(3), 1399–1413.

Xu, Y., & Wang, Q. E. (2001). Pitch targets and their realization: Evidence from mandarin chinese. *Speech communication*, *33*(4), 319–337.

Yi, H. (2014). A gestural account of Mandarin tone sandhi. *The Journal of the Acoustical Society of America*, *136*(4), 2144–2144.

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

Yip, M. (1980). *The tonal phonology of Chinese* (Unpublished doctoral dissertation). Massachusetts Institute of Technology.

Yip, M. (1989). Contour tones. *Phonology*, *6*(01), 149–174.

Zec, D. (2005). Prosodic differences among function words. *Phonology*, *22*(1), 77–112.

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

Zec, D., & Zsiga, E. (2018). Phonological consistency and phonetic variation: Tonal alignment in three Neo-Štokavian dialects. Ms.

Zsiga, E. (1993). *Features, gestures, and the temporal aspects of phonological organization* (Unpublished doctoral dissertation). Yale University.

Zsiga, E., & Nitisaroj, R. (2007). Tone features, tone perception, and peak alignment in Thai. *Language and Speech*, *50*(3), 343–383.

Zsiga, E., & Zec, D. (2013). Contextual evidence for the representation of pitch accents in standard Serbian. *Language and speech*, 0023830912440792.