Contextual Evidence for the Representation of Pitch Accents
in Standard Serbian

Elizabeth Zsiga and Draga Zec

Elizabeth Zsiga
Dept. of Linguistics
Georgetown University
Washington, D.C. 20057
zsigae@georgetown.edu
(Corresponding author)

Draga Zec
Dept. of Linguistics
Cornell University
dz17@cornell.edu
Contextual Evidence for the Representation of Pitch Accents
in Standard Serbian

ABSTRACT

This paper reports the results of an experiment that elicits contextual effects on Rising and Falling accents in Standard Serbian, with the goal of determining their acoustic correlates and their phonological representation. Materials systematically vary the distance between pitch accents, inducing "tone crowding," in order to identify the phonetic dimensions that consistently distinguish the two pitch accent types, to examine the association between accents and the segmental string, as well as the timing relationship between accent minima and maxima, and to investigate the interaction between lexical accents and boundary tones. On the basis of the phonetic findings, a unified analysis of the phonological distribution and phonetic realization of Falling and Rising accents in Standard Serbian is proposed. It is proposed that both Rising and Falling accents consist of a single lexical High. The restricted distribution of the two accents, emerges from the interaction of stress and tone: Falling accents are monosyllabic, such that stress and pitch prominence coincide; Rising accents are bisyllabic, such that the stressed syllable precedes the pitch-accented syllable. The phonetic differences between the Falling and Rising accent follow from the place of lexically designated H, the location of stress, and the effects of boundary tones.

The larger issue we address concerns the phonological characterization of tone/stress interactions. Given the two general types of interactions, one in which the place of stress is predictable from the place of tone, and the other with the reversed direction of influence, we analyze Standard Serbian as belonging to the former type. While both types can be characterized
in systems of tonal phonology, which allow free interaction of tone and stress, the type exemplified by Standard Serbian, with contrastive tonal specifications governing the distribution of stress, cannot be captured in an Autosegmental-Metrical (AM) framework, in which stress serves as anchor for tonal melodies.
INTRODUCTION

This paper investigates the representation, distribution, and realization of Rising and Falling accents in Standard Serbian. Previous accounts agree in finding that each accent is associated with a different pitch pattern over the stressed syllable, but there is no agreement about the phonological characterization of the two pitch accent types. Beyond accounting for the language-particular facts, investigating the Standard Serbian system is of particular interest because it provides an opportunity to explore the nature of interactions between stress, tone, and intonation, taking into account both phonetic and phonological aspects.

As has been noted in a number of studies that focus on the interactions of lexical tone and word stress (de Lacy 2002, Hyman 2006, Zec 1999), it is possible, in the typology of these systems, either for the position of tone to determine the distribution of stress, or for the position of stress to determine the distribution of tone, including also the trivial type of no interactions. By contrast, intonational systems are characterized by a one-way attraction between stress prominence and tonal melodies. In autosegmental-metrical (AM) approaches to intonational theory, it is rightly assumed that tonal melodies, that is, pitch accents or starred tones, associate with metrically prominent stressed syllables (Pierrehumbert 1980, Prieto 2011 and references therein). Because AM intonational models work so well in capturing the correlation between metrical prominence and tonal melody at the phrasal level, it may seem theoretically desirable to extend them to systems such as Standard Serbian in which tone and stress correlate at the lexical level (Ladd 2008:156, and the references therein). However, if an AM model of intonational
theory is employed in an analysis of how lexical tone interacts with stress, this formal choice restricts the range of possible interactions. Only one type of interaction is captured: the case of stress governing the distribution of tone. Most approaches to lexical tone and stress, however, are not constrained in this way, and thus make no specific predictions regarding the direction of influence between the two subsystems (e.g. de Lacy 2002, Hyman 2006, Zec 1999).

Various approaches have been used in previous research on Standard Serbian. We begin with the classic study of Lehiste and Ivić (1986), who characterize both pitch accents in terms of F0 maxima with different syllable affiliations. They find that in Falling accents, the F0 maximum is on the stressed (or tonic) syllable, while in Rising accents, it is on the post-tonic syllable. Under this view that goes back to Masing (1876), Rising accents are disyllabic, with stress and pitch maxima residing on different syllables, while Falling accents are monosyllabic, with stress and pitch maxima coinciding. These results suggest that both pitch accents are represented as a single H tone, as phonologically captured in Browne & McCawley (1965) and Inkelas & Zec (1988).

Two recent proposals are couched in the formalism of AM intonational theory, specifically, bitonal pitch accents that associate to stressed syllables. Godjevac (2000) argues that the Falling accent should be represented as H*L, while the Rising accent is L*H, based on qualitative examination of the accent shape. In contrast, Smiljanić (2002) argues that both pitch accents are LH, with early alignment, that is, LH*, for the Falling accent and late alignment, that is, L*H, for the Rising accent. Smiljanić argues for bitonal LH* and L*H pitch accents based on three findings from an acoustic study. First she finds that not only the peak (corresponding to H) but also the beginning of the rise to the peak (corresponding to L) is systematically later in Rising accents than in Falling accents. She further finds that alignment of
H and L are correlated over variation in vowel length and pragmatic condition. The location of the L following the peak, however, is found not to correlate with peak location, but to vary according to phrasal influences. Smiljanić thus argues that both Falling and Rising accents are lexically specified as LH, but the following L is not lexically specified for either accent type.

In addition to characterizing accent shape and peak alignment, an account of the Standard Serbian prosodic system must address the interaction of tone and stress in determining the distribution of Falling and Rising accents. The distribution of the two pitch accents, shown in Table 1, is not free.

**Table 1. Distribution of Pitch Accents**

<table>
<thead>
<tr>
<th></th>
<th>Monosyllables</th>
<th>Polysyllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>Initial σ</td>
</tr>
<tr>
<td>Falling</td>
<td>✓</td>
<td>✓</td>
</tr>
<tr>
<td>Rising</td>
<td></td>
<td>✓</td>
</tr>
</tbody>
</table>

The Falling accent occurs on monosyllables and the initial syllable of polysyllables, while the Rising accent occurs only on polysyllables, on any syllable other than the final. Any phonological analysis of the pitch accents should include an account of their asymmetric distribution. Analyses cast in terms of an AM model extended to lexical interactions are formally committed to the assumption that the place of stress determines the distribution of tone, but models that assume that Standard Serbian pitch accent is a tonal contrast can have access to a wider range of interactions. It will be argued here that these patterns are better accounted for if it is assumed that tone governs the place of stress.

A further issue involves the interaction of lexical and phrasal tones, specifically with respect to "tonal crowding" (Arvaniti, Ladd & Mennen 1998, 2000, 2006; Grice 1995; Grice, Ladd & Arvaniti 2000; Silverman & Pierrehumbert 1990). There is a limit on the number of pitch targets than can be realized within a given domain. When this limit is pushed, as can
happen when both lexical and intonational tones associate to the same word or syllable, adjustments, which might include deletion, reduction, or realignment, have to be made. It is often assumed that, due to a principle of contrast preservation, post-lexical intonational adjustments will not override lexical specifications. Smiljanić (2002) found, for example, that when lexical pitch accents occur adjacent to phrasal boundary tones, the lexical tones shift somewhat to avoid having multiple pitch targets associated to a single syllable, but contrast is maintained. She states (2002:182-183) that while "both prosodic and pragmatic factors significantly change pitch alignment. . .[i]n narrow focus in Belgrade the overriding consideration is to maintain the accentual distinctions and this precludes alignment changes" that would result in neutralization. Smiljanić's data, however (p. 207-208), did not include utterance-final words with stressed short vowels, where the pressure of tonal crowding would be greater than for the long vowels that Smiljanić tested. The possibility of neutralization between Rising and Falling accents on words with stressed short vowels in utterance-final position in Serbian is explicitly tested here. It is of particular interest to test for the possibility of neutralization under narrow focus, because of the prevailing assumption that focus necessarily enhances lexical distinctions (Lindblom 1990).

The descriptive goal of the present research is to contribute new data toward resolving the representational debate on Standard Serbian accent. This paper reports the results of an experiment that elicits contextual effects on Rising and Falling accents in Serbian, by varying their distance from the preceding peak (cf. Ladd & Schepman 2003, Meyers 2003, Pierrehumbert & Beckman 1988), as well as vowel length and number of syllables in the word. Materials systematically vary the distance between pitch targets, in order to determine the effect of tone crowding on the realization of both lexical accents and boundary tones. Acoustic measurements
are used to identify the phonetic dimensions that consistently distinguish the two pitch accent types across different contexts, with the larger goal of discerning the phonological distinctions between the two pitch accent types.

The present study thus builds on the results of previous phonetic studies. Lehiste and Ivić’s (1986) study, which also focused on Standard Serbian, investigated disyllabic and trisyllabic words produced in a frame sentence, varying accent type, accent position and vowel length. Smiljanić’s (2002) study examined the production of disyllabic and trisyllabic words with penultimate accent in Standard Serbian, varying vowel length, accent type, focus and sentence position (see also Smiljanić & Hualde 2000). Previous studies did not systematically vary the preceding context, which is an important feature of the present experimental design.

Beyond the phonetic and phonological description of Serbian, the larger theoretical goal of this study is to investigate several general questions of tonal and intonational representation and realization: possible interactions of tone and stress, the interaction of intonational and lexical tones, differences between simple and bitonal accents, and the role of prosodic units and their edges in the alignment of tones to the segmental string.

The next sections of this paper describe the design of the present experiment, and the results of the phonetic analysis. Phonological analysis, accounting for accent distribution, accent shape, and accent neutralization, follows.

EXPERIMENTAL DESIGN

Hypotheses

In order to better discern the phonological distinctions between Rising and Falling accents and the principles governing their alignment to the segmental string, an experiment was
designed to elucidate contextual effects on the two pitch accent types. There are three specific aims: to identify the phonetic dimensions that consistently distinguish the two pitch accent types under word-internal and word-external contextual variation, to examine the association between accents and the segmental string, and to examine the timing relationship between accent minima and maxima, if any. Following Pierrehumbert and Beckman (1988), materials were designed in which the distance from the preceding peak to the accent under study was systematically varied. Word and syllable length were also varied. The effect of these manipulations on both maxima and minima were examined.

All previous research agrees on one point, stated here as Hypothesis 1:

*Hypothesis 1:* Alignment of maxima will differ by accent type: As reported in previous literature, it is predicted that the peak of Rising accents will be consistently later in the target word than the peak of Falling accents, regardless of syllable affiliation.

Previous research differs, however, in accounts of the syllable affiliation of the pitch peak, particularly for the Rising accent. Here, following Lehiste and Ivić (1986) and Smiljanić (2002), it is hypothesized that Falling accents reach their peak on the tonic syllable and Rising accents in general reach their peak on the post-tonic syllable.

*Hypothesis 2:* The peak of the Falling accent will occur on the tonic syllable.

*Hypothesis 3:* The peak of the Rising accent will occur on the post-tonic syllable.

Different phonological representations for the accents make different predictions concerning the correlation between the location of F0 maxima and minima, and the effects of phrasal and word-internal variation. We follow Pierrehumbert (1980), Pierrehumbert and Beckman (1988), and Smiljanić (2002), among others, in assuming that the location of H and L
in bitonal accents will be correlated. If the two parts of the accent are a phonological unit (like a diphthong or affricate, as suggested by Gussenhoven 2004, see also the references therein), then it is predicted the two parts will move in tandem under any phrasal effects, maintaining a "constant duration of the interval" (Frota 2002: 402) between the two tones. We assume here the strong version of this hypothesis, according to which two tones of a bitonal accent will be correlated in this fashion whether the starred tone is initial or final in the bitonal group. In the weaker version of this hypothesis, proposed by Grice (1995), a bitonal pitch accent whose starred tone is initial forms a tighter unit than a bitonal accent with a final starred tone. Frota (2002) found this asymmetry in the timing relations between intonational H*L and HL* pitch accents in European Portuguese: only the former maintained an interval of constant duration. We opt for the stronger hypothesis for two reasons. First, it may be that lexical bitonal accents form tighter units than their intonational counterparts, because they have a more restricted range of possible segmental alignments. Thus it makes sense to test for correlations for all lexical accents. Second, we note that Smiljanić (2002) found correlations for both Rising (which she analyzed as L*H) and Falling (which she analyzed as LH*) accents in Standard Serbian, and this result is further tested here. Thus our Hypotheses 4 and 5 are stated with maximum generality, predicting a correlation between the two pitch targets of any bitonal accent. In the weaker version of the hypothesis, correlations would be predicted only for bitonal accents with trailing tones (i.e., L*H), and no prediction would be made for bitonal accents with leading tones (i.e., LH*).

**Hypothesis 4:** If a pitch accent consists only of H*, as would follow from Lehiste and Ivić’s (1986) characterization of both Rising and Falling accents only in terms of maxima, then:
4a: There may be only High tones specified in the tonal string and any minima will be the result of phonetic effects. In that case, the likelihood of finding a minimum between peaks will increase as the peaks on the two words are further apart.

4b: If there are phonologically specified L tones, they will be either boundary tones or post-lexically inserted, rather than lexical. Minima will be consistently found at the prosodic boundary. However, the timing of the L to the segmental string and the amount of F0 fall from maximum to minimum will vary depending on syllable structure and distance from the preceding peak. Conversely, the timing of the H will be lexically determined and relatively unaffected by word-external variation (distance from the preceding peak).

4c: The timing of the pitch minimum corresponding to the L and the pitch maximum corresponding to the H will not show a significant correlation.

Hypothesis 5: If a pitch accent consists of a bitonal L+H, as proposed by Smiljanić (2002) for both Rising and Falling accents, and by Godjevac (2000) for Rising accents only, then:

5a: One or more minima will be consistently found, at least one of which is lexical.

5b: Both minimum and maximum will be relatively unaffected by phrasal context, particularly distance from the preceding peak.

5c: The timing of the pitch minimum corresponding to the lexical L and the pitch maximum corresponding to the lexical H will show a significant correlation.
Because all target words are phrase-final in the present study, no hypotheses concerning alignment of the final L are tested. Following Smiljanić (2002), who showed conclusively that the final L is phrasally conditioned, it is assumed that the final L is intonational. The final hypothesis concerns the interaction of lexical and phrasal specifications:

*Hypothesis 6:* Tone realignment or retraction due to variation in the distance between tonal specifications (tonal crowding) will not result in neutralization between the two accent types.

**Materials**

In order to test these hypotheses, 24 two-word phrases of Serbian were constructed. In each case, the first word was a trisyllabic verb, with a pitch peak occurring either early in the word (first syllable), in the middle of the word (second syllable), or at the end of the word (third syllable). The second word consisted of a noun with initial stress. No words with non-initial stress were included because Rising and Falling accents are only contrastive on words with initial stress. The second word was varied according to whether the lexical accent was Rising or Falling, whether the stressed vowel was long or short, and whether the word was disyllabic or trisyllabic. These variables are schematized in Figure 1.
In order to control for sentence stress and focus, each of the test phrases was paired with a question to which the test phrase was the answer. (For example: “Koga volite? Volimo Nemanju.” Whom do you love? We love Nemanja.) The question contained or implied the verb, thereby ensuring that sentence stress and narrow focus would fall on the noun, the answer to the question. In so far as possible, target phrases consisted only of sonorants. The full set of target phrases is given in Table 2.

Figure 1. Schematic of experimental variables.
Table 2. Target phrases. In the first word of the phrase, the syllable realizing the pitch peak is underlined. In the second word of the phrase, the accented syllable is in bold.

<table>
<thead>
<tr>
<th>Target word shape</th>
<th>Preceding peak</th>
<th>Falling accent on target word</th>
<th>Rising accent on target word</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trisyllabic Long V</td>
<td>Early σ σ σ</td>
<td>/krijemo na:mere/</td>
<td>/volimo ne:manju/</td>
</tr>
<tr>
<td></td>
<td>Middle σ σ σ</td>
<td>/znamo im na:mere/</td>
<td>/imamo ne:manju/</td>
</tr>
<tr>
<td></td>
<td>Late σ σ σ</td>
<td>/razume na:mere/</td>
<td>/razume ne:manju/</td>
</tr>
<tr>
<td>Disyllabic Long V</td>
<td>Early σ σ σ</td>
<td>/volimo mo:re/</td>
<td>/volimo ne:nu/</td>
</tr>
<tr>
<td></td>
<td>Middle σ σ σ</td>
<td>/imamo mo:re/</td>
<td>/imamo ne:nu/</td>
</tr>
<tr>
<td></td>
<td>Late σ σ σ</td>
<td>/ma plovi mo:rem/</td>
<td>/razume ne:nu/</td>
</tr>
<tr>
<td>Disyllabic Short V</td>
<td>Early σ σ σ</td>
<td>/volimo mamu/</td>
<td>/volimo neven/</td>
</tr>
<tr>
<td></td>
<td>Middle σ σ σ</td>
<td>/znamo im mamu/</td>
<td>/beremo neven/</td>
</tr>
<tr>
<td></td>
<td>Late σ σ σ</td>
<td>/razume mamu/</td>
<td>/ma bere neven/</td>
</tr>
</tbody>
</table>

Verbs with an early pitch peak (krijemo, volimo, hranimo) all had stress on the initial syllable with Falling accent, verbs with a medial pitch peak (znamo (im), imamo, beremo) all had initial stress with Rising accent, and verbs with a late pitch peak (razume, ma love, ma plovi, ma bere) all had penultimate stress with Rising accent. These patterns were hypothesized to result in the peak being reached on the first, second, and third syllables respectively, and this alignment was in fact experimentally confirmed. (Pitch peaks are allowed on word-final syllables except in utterance-final position.) However, exact alignment of this initial peak was not crucial to the design. As noted above, it is uncontroversial that the peaks of Rising accents are later than the peaks of Falling accents, and that peaks are later still in words with non-initial
stress. This three-way distinction was all that was necessary to conform to the Early, Middle, Late design.

**Participants and recording procedures**

Participants were three female native speakers of Standard Serbian. All three were born and grew up in Belgrade. Speakers 1 and 2 were recorded in Belgrade, in a quiet room. Speaker 3 (Zec) was recorded in the phonetics lab at Cornell University.

Each question and answer pair was written on an index card, using normal Serbian spelling (Roman alphabet). The speaker read aloud both question and answer, going through the set of index cards three times. The order of the cards was randomized for each speaker and each repetition. There were thus a total of 216 tokens (24 phrases x 3 speakers x 3 repetitions).

Productions were recorded directly onto a laptop computer, digitized at 22050 Hz.

**Analysis**

**Labeling.** Pitch tracks were obtained via the autocorrelation algorithm in Praat (Boersma & Weenink 2008). Each target sentence was labeled to mark the time at the left edge of the target noun (referred to below as simply "word boundary"), the time at all syllable boundaries, and the time (ms) and F0 value (Hz) at four pitch extrema:

1. Peak1: peak on the first word
2. Peak2: peak on the second word
3. F_End: end of the fall from the first peak
4. R_Beg: beginning of the rise to the second peak
Whenever possible, locations of minima and maxima were found using Praat’s peak-picking algorithms.

Typical examples of labeled tokens are shown in Figure 2 (Falling accent) and Figure 3 (Rising accent). With a few exceptions, discussed below, peaks were clear and the highest point was unambiguous. Regarding minima, in many tokens, there was a smooth curve between the two peaks, as in Figure 2. In that case, the same point -- the pitch minimum at the bottom of the curve -- was marked as both F_End and R_Beg. In other tokens, however, there was a relatively flat low stretch between the two peaks, as in Figure 3. We will refer to such stretches as plateaus. For purposes of labeling, a plateau was defined as an extent of the pitch track over which F0 changed less than 10 Hz per 100 ms, and that was marked by an abrupt change in slope at either end. In that case, F_End was marked at the left edge of the plateau and R_Beg was marked at the right edge of the plateau.
Figure 2. Labels for a typical token with Falling accent.

Figure 3. Labels for a typical token with Rising accent.
In some cases (10 tokens with Rising accent and 10 with Falling accent, 9% of the data), there was no measurable peak on the second word. We attribute this to narrowing of the pitch range within an intonational phrase. As noted by Godjevac (2005, p.157), within an intonational phrase, "the pitch range is downstepped for each subsequent phonological word." While contrast is generally maintained within the narrowed range, in these few tokens the narrowing was extreme. These 20 tokens are excluded from the statistical analysis.

Statistical analysis. In order to investigate the factors that affect accent timing and shape, a series of mixed-model analyses (with Subject as a random variable) was conducted. The independent variables were Accent Type (Rising vs. Falling), Preceding Context (Early, Middle, or Late peak on the preceding word), Syllable Count (disyllabic vs. trisyllabic target word) and Vowel Length (long vs. short initial syllable). Dependent variables were computed from the F0 and time labels, and are defined in Table 3. The locations of pitch minima and maxima were defined relative to the beginning of the target word as a stable anchor point. Later segmental landmarks, such as beginning of the first vowel or end of the first syllable, vary according to segment quality or vowel length, and are thus too variable to allow valid comparison across contexts. For each dependent variable, three questions are addressed:

1. Does this variable reliably distinguish Rising and Falling accents (effects of Accent Type)?

2. How is this variable affected by word-internal variation in syllable and word structure (effects of Vowel Length and Syllable Count)?
3. How is the variable affected by word-external variation (effects of Preceding Context, that is, distance from the preceding peak)?

In addition to the mixed model analyses, a set of correlations was computed to test the alignment of H and L points to each other and to the segmental string. For this analysis, the locations (relative to the left edge of the target word) of R_Beg (L) and Peak2 (H) were correlated with each other, and with the location (again relative to the left edge of the target word) of the boundary between the first and second syllables of the target word.

### Table 3. Dependent variables.

<table>
<thead>
<tr>
<th>Dependent variable name</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>Peak Location</td>
<td>Peak2(ms) – WordBoundary(ms)</td>
</tr>
<tr>
<td>F_End Location</td>
<td>F_End(ms) – WordBoundary(ms)</td>
</tr>
<tr>
<td>R_Beg Location</td>
<td>R_Beg(ms) – WordBoundary(ms)</td>
</tr>
<tr>
<td>Minima Separation</td>
<td>(R_Beg(ms) – F_End(ms)) / (Peak2(ms) – Peak1(ms))</td>
</tr>
<tr>
<td>Peak F0</td>
<td>Peak2(Hz)</td>
</tr>
<tr>
<td>Valley F0</td>
<td>Mean(F_End(Hz), R_Beg(Hz))</td>
</tr>
</tbody>
</table>

Measures of alignment and F0 value are considered separately, rather than through a composite variable such as slope, so that contextual effects on timing and F0 height can be separately evaluated.

**RESULTS**

**Mixed Model Analysis**

Peak Location. Significant effects for Peak Location are listed in Table 4. There were significant effects of Accent Type, Syllable Count, Vowel Length, and all their two-way and three-way interactions, but no main effect of Preceding Context, and no significant interaction that includes Preceding Context. This result is consistent with Hypothesis 1 and with both Hypotheses 4 and 5: peak alignment is predicted to be lexically or phonologically determined...
for both simple and bitonal accents, and neither hypothesis predicts an effect of preceding context.

**Table 4. Significant effects on Peak 2 Location.**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent Type</td>
<td>1,170</td>
<td>397.0</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Syllable Count</td>
<td>1,170</td>
<td>29.42</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Vowel Length</td>
<td>1,170</td>
<td>41.85</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * SylC</td>
<td>1,170</td>
<td>93.03</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * VL</td>
<td>1,170</td>
<td>145.0</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>SylC * VL</td>
<td>1,170</td>
<td>10.41</td>
<td>.002</td>
</tr>
<tr>
<td>Acc * SylC * VL</td>
<td>1,170</td>
<td>15.25</td>
<td>&lt; .001</td>
</tr>
</tbody>
</table>

The many significant effects and interactions show that accent realization is affected by the phonological shape of the target word, and that Falling and Rising accents are affected differently. Figure 4 graphs the interaction of Accent Type and Vowel Length. Trisyllables are shown in Figure 4A, disyllables in Figure 4B. The dark bar in each graph indicates the mean location of the end of the stressed syllable of the target word in that context. These syllable boundary markers are provided here for reference; syllable affiliations for each token are graphed in Figures 10 – 13.
Figure 4. The effect of Accent Type, Vowel Length, and Syllable Count on Peak Location.

A. Trisyllables.  B. Disyllables. Values show Peak Location relative to the beginning of the target word. Error bars show +/- 1 standard deviation. The black bar shows the mean location of the end of the tonic syllable.
Peak Location consistently distinguishes Rising vs. Falling accents in three of four contexts. The peak of the Rising accent is later than the peak of the Falling accent in all cases except disyllables with short vowels.

Differential effects of Syllable Count and Vowel Length are also seen. For the Falling accent, the peak always occurs within the first syllable (somewhat later in short vowels than in long vowels). For the Rising accent, the peak occurs in the second syllable on trisyllabic words. Note that as the syllable boundary moves later in words with a long vowel in the first syllable, the peak of the Rising accent also moves, maintaining a consistent distance from the syllable boundary. (This alignment will be measured directly in the correlations discussed below.)

For Rising accents on disyllabic words, the peak is realized on the first syllable. It will be argued below that this is due to an interaction with phrasal boundary tones that link to phrase-final syllables, forcing the lexical accent back one syllable. On disyllabic words with a long first vowel, the Rising and Falling accent remain distinct, with the peak of the Rising accent occurring late in the syllable and the peak of the Falling accent occurring earlier in the syllable. On disyllabic words with a short first vowel, peak location for Rising and Falling accents does not differ.

F_End Location. The mixed model analysis found only two significant effects on the location of F_End (end of the fall from the first peak): Preceding Context (F(2,170) = 27.17, p < .001) and the interaction of Accent Type and Syllable Count (F(2,170) =5.40 , p = .021). The highly significant effect of Preceding Context shows that the location of F_End was largely governed by the location of the pitch peak in the first word, regardless of the identity of the second word. As
shown in Table 5, the end of the fall from the first peak consistently occurred just after the word boundary, with the minimum occurring later as the peak occurred later.

**Table 5. Effect of Preceding Context on F_End Location**

<table>
<thead>
<tr>
<th>Preceding Word Accent:</th>
<th>Distance from word boundary to F_End</th>
</tr>
</thead>
<tbody>
<tr>
<td>Early</td>
<td>10 ms</td>
</tr>
<tr>
<td>Middle</td>
<td>20 ms</td>
</tr>
<tr>
<td>Late</td>
<td>62 ms</td>
</tr>
</tbody>
</table>

The interaction between Accent Type and Syllable Count is harder to interpret, but it is likely a small segmental effect: for Falling accents, F_End is slightly (14 ms) later in disyllables than trisyllables. It happens that the Falling disyllables begin with [m] ([mamu], [mo:re]), while the other lexical items began with [n] or [l], and this difference in place of articulation may be sufficient to cause a difference in precise F0 alignment.

**R_Beg location.** We now turn to results for R_Beg Location (the point in the target word at which the rise to the peak begins), keeping in mind the point made by Figures 2 and 3 that the beginning of the rise may or may not begin immediately at the end of the fall. (The degree of separation between these two points is tested directly below.) Results of the mixed model analysis on R_Beg Location are shown in Table 6. Main effects of Accent Type, Syllable Count, and Vowel Length, as well as their two-way and three-way interactions, are highly significant (p < .001). The effect of Preceding Word Accent, and the interaction of Vowel Length with Preceding Word Accent, are also significant.
Table 6. Significant effects on R_Beg Location, all cases included.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent Type</td>
<td>1,170</td>
<td>381.33</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Syllable Count</td>
<td>1,170</td>
<td>57.01</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Vowel Length</td>
<td>1,170</td>
<td>31.00</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * SylC</td>
<td>1,170</td>
<td>74.97</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * VL</td>
<td>1,170</td>
<td>25.29</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>SylC * VL</td>
<td>1,170</td>
<td>16.9</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * SylC * VL</td>
<td>1,170</td>
<td>13.26</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Preceding Word Accent</td>
<td>2,170</td>
<td>10.22</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>PWA * VL</td>
<td>1,170</td>
<td>3.76</td>
<td>.025</td>
</tr>
</tbody>
</table>
Figure 5. The effect of Accent Type, Vowel Length, and Syllable Count on R_Beg (beginning of the rise to the peak on the target word).  A.  Trisyllables.  B. Disyllables.  Values show R_Beg relative to the beginning of the target word.  Error bars show +/- 1 standard deviation.  The black bar shows the mean location of the end of the tonic syllable.
The interactions of Accent type, Syllable Count, and Vowel Length are graphed in Figure 5. These effects on the minimum location are very similar to those on the peak location (Figure 4). For three of four contexts, minima for the Rising accents are later than the minima for the Falling accents (consistent with the findings of Smiljanić 2002). There is no difference between Rising and Falling accents on short disyllables.

Across all contexts, minima occur in the first syllable of the target word (or in the case of Rising accents on trisyllables, within 10 ms of the syllable boundary). For the Falling accent, differences in Vowel Length and Syllable Count have little effect on minimum location: across all contexts, the minimum occurs early in the first syllable, on average 57 ms into the target word. For Rising accents, however, both Vowel Length and Syllable Count have a strong effect. In trisyllables, the minimum for Rising accents occurs late in the first syllable, at or very close to the syllable boundary. In disyllables with a long initial vowel, the minimum occurs in the middle of the vowel, later than the minimum for the Falling accent. In disyllables with a short initial vowel, however, the mean minimum location for Rising accents is 69 ms into the word, not significantly different from the minimum location for Falling accents. As with Peak Location, these interactions with syllable structure will be explored further below, when correlations with the syllable boundary location are discussed.

There was also a significant effect of Preceding Accent. However, in interpreting this result, the fact must be kept in mind that in many tokens F_End and R_Beg refer to the same point, and a significant effect of context on F_End was previously found. Table 7 shows the results of the mixed model analysis when only those cases for which R_Beg was separate from F_End are included. If those cases are excluded, all of the word-internal effects remain significant except for the three-way interaction. The main effect of Preceding Accent also
remains significant, although the size of the effect is smaller at \( p = .025 \). As shown in Table 8, the direction of the effect is the same as that for F_End: minima are later when the preceding accent is closer.

Table 7. Significant effects on R_Beg Location, including only cases where R_Beg \( \neq \) F_End.

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent Type</td>
<td>1,95</td>
<td>257.43</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Syllable Count</td>
<td>1,95</td>
<td>36.70</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Vowel Length</td>
<td>1,95</td>
<td>25.83</td>
<td>&lt; .001</td>
</tr>
<tr>
<td>Acc * SylC</td>
<td>1,95</td>
<td>11.70</td>
<td>.001</td>
</tr>
<tr>
<td>Acc * VL</td>
<td>1,95</td>
<td>4.394</td>
<td>.04</td>
</tr>
<tr>
<td>SylC * VL</td>
<td>1,95</td>
<td>5.064</td>
<td>.027</td>
</tr>
<tr>
<td>Preceding Word Accent</td>
<td>2,95</td>
<td>3.83</td>
<td>.025</td>
</tr>
</tbody>
</table>

Table 8. Effect of Preceding Context on R_Beg Location.

<table>
<thead>
<tr>
<th></th>
<th>Distance from word boundary to R_Beg</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>All cases</td>
</tr>
<tr>
<td></td>
<td>n = 196</td>
</tr>
<tr>
<td>Preceding Accent:</td>
<td></td>
</tr>
<tr>
<td>Early</td>
<td>109 ms</td>
</tr>
<tr>
<td>Middle</td>
<td>123 ms</td>
</tr>
<tr>
<td>Late</td>
<td>156 ms</td>
</tr>
</tbody>
</table>

Minima Separation. This section now explicitly turns to the separation of fall and rise in these tokens. The degree to which F_End and R_Beg are separated by a low pitch plateau (Figure 3) is encoded in the variable Min_Separation. This variable is defined (Table 3) as the distance in ms between F_End and R_Beg, as a percentage of the distance in ms between Peak1 and Peak2. For
tokens where $F_{\text{End}}$ and $R_{\text{Beg}}$ refer to the same point in time (Figure 2), $\text{Min} \_ \text{Separation} = 0$. In tokens where $\text{Min} \_ \text{Separation}$ is greater than 0, the accent on the target word consists of a plateau followed by a rise, rather than a rise of uniform slope.

Table 9 shows that tokens realized with a separation between the fall and rise were disproportionately Rising accents. More than 90% of Rising accents (excluding those on disyllables with short vowels) were realized with a plateau. (The different total token counts in each cell arise because of the excluded tokens that had no peak on the target word at all.)

<table>
<thead>
<tr>
<th></th>
<th>Rising Accent</th>
<th></th>
<th>Falling Accent</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Disyllabic</td>
<td>Trisyllabic</td>
<td>Disyllabic</td>
<td>Trisyllabic</td>
</tr>
<tr>
<td><strong>Short Vowel</strong></td>
<td>13/21</td>
<td>26/27</td>
<td>2/21</td>
<td>8/25</td>
</tr>
<tr>
<td></td>
<td>62%</td>
<td>96%</td>
<td>10%</td>
<td>32%</td>
</tr>
<tr>
<td><strong>Long Vowel</strong></td>
<td>22/23</td>
<td>25/27</td>
<td>12/26</td>
<td>10/26</td>
</tr>
<tr>
<td></td>
<td>96%</td>
<td>93%</td>
<td>46%</td>
<td>38%</td>
</tr>
</tbody>
</table>

The higher number of plateaus for Rising accents can not be attributed solely to the fact that the Rising accent Peak is later in the word. Even when overall peak-to-peak duration is controlled for, the presence of a plateau between minima is far more common for Rising accents. Figure 6 (modified from Figure 1) shows the count of syllables that intervene between peaks in different contexts, taking into account the syllable on which the accent of the first word occurs, and the syllable on which the accent of the second word occurs: first (tonic) for Falling and second (post-tonic) for Rising. While cases of peaks on immediately adjacent syllables occur only for Falling accents and cases with three intervening syllables occur only for Rising accents, cases with 1 or 2 intervening syllables occur for both.
Figure 6. Diagram of number of intervening syllables between peaks for Rising and Falling accents, based on Figure 1. The horizontal lines indicate the number of syllables that intervene between pitch peaks in different contexts, depending on whether the accent on Word B is Falling or Rising, and whether the accent is on the first, second, or third syllable of Word A.

Figure 7 then graphs the percentage of tokens realized with a plateau between the fall and rise, as a function of the number of syllables between peaks. The figure includes trisyllabic target words only, because of the accent retraction found for Rising accents in utterance-final disyllables. It is evident that even when peak-to-peak duration is controlled for, Rising accents are far more likely to be realized with a plateau.
Figure 7. Percent of tokens realized with a plateau for Rising and Falling accents, grouped by number of syllables intervening between peaks. Trisyllables only.

Stretches of flat pitch do occur on some 30% of tokens with a Falling accent, but when they occur these plateaus are generally much shorter than those that occur with Rising accents. Differences in plateau duration are shown in Table 10, and Figures 8 and 9. Table 10 breaks down the data by syllable count and vowel length, and includes only cases where plateau duration does not equal zero.

Table 10. Plateau duration, in ms, excluding cases where F_End = R_Beg.

<table>
<thead>
<tr>
<th></th>
<th>Disyllabic</th>
<th>Trisyllabic</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Short vowel</td>
<td>Long vowel</td>
</tr>
<tr>
<td>Falling accent</td>
<td>112 ms</td>
<td>73 ms</td>
</tr>
<tr>
<td>Rising accent</td>
<td>102 ms</td>
<td>152 ms</td>
</tr>
</tbody>
</table>
Note that in three of four contexts, including long disyllables, the plateaus for Rising accents are more than twice as long as those for Falling accents. For short disyllables, Rising plateaus are not longer than those for Falling.

Figures 8 and 9 represent plateau duration, again controlling for number of intervening syllables. Figure 8 presents the mean duration of occurring plateaus (that is, only cases where F_End ≠ R_Beg). Figure 9 presents the mean plateau duration over all tokens, including the cases where plateau duration = 0. Figure 9 thus represents the overall extent to which the accent type is characterized by separation between F_End and R_Beg. It is clear from this figure that Rising accents are systematically realized with separation between Rise and Fall, while Falling accents are not.

![Mean Plateau Duration Excl 0s](chart)

**Figure 8.** Mean plateau duration, including only trisyllables, and only those cases where F_End ≠ R_Beg.
Figure 9. Mean plateau duration, including only trisyllables, and but not excluding cases where F_End = R_Beg.

Statistical analysis further confirms the significance of the differences shown in Tables 9 and 10 and Figures 7 – 9. A mixed model analysis with dependent variable Minima Separation was run twice, once with cases where F_End = R_Beg included, once with cases where F_End = R_Beg excluded.

Results for the analysis when all cases are included are shown in Table 11. There are highly significant main effects of Accent Type, Vowel Length, and Syllable Count, reflecting the facts that plateaus are more likely to occur with Rising Accents, on long vowels, and on trisyllables. The significant interaction of accent and syllable count is due to the fact that syllable count has a bigger effect on Rising accents than on Falling, particularly due to the short disyllables. As was the case with peak location, Rising Accents in this context are indistinguishable from Falling accents on the basis of minima separation. Rising accents on
short disyllables are more likely to be realized with a plateau than are Falling accents (62% of tokens vs. 10% of tokens), but plateau duration does not differ in this context.

Table 11. Significant effects on Minima Separation, all tokens

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>df</th>
<th>F</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accent Type</td>
<td>1,170</td>
<td>188.96</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Syllable Count</td>
<td>1,170</td>
<td>22.85</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Vowel Length</td>
<td>1,170</td>
<td>9.48</td>
<td>.002</td>
</tr>
<tr>
<td>Acc * SylC</td>
<td>1,170</td>
<td>12.10</td>
<td>.001</td>
</tr>
</tbody>
</table>

When cases with no plateau are excluded, there is only one significant effect: Plateaus for Rising accents are a significantly longer percentage of peak-to-peak duration than those that occur with Falling accents (F(1,94) = 46.85, p < .001). Taken together, the results for Minima Separation show that accent shape is significant factor in distinguishing accent type. (This is also noted by Godjevac (2005:159): "The shape of the F0 contour provides an easy visual differentiation between the two types of accents.") The consistent presence of a plateau for Rising accents in 3 of 4 contexts suggests the presence of an additional phonological target in Rising-accent words. We return to the phonetic and phonological analysis of this difference in accent shape in the discussion section below.

Peak F0. The statistical analysis found only three significant effects on the F0 height of the pitch peak: Vowel Length (F(1,170) = 5.83, p = .017), Preceding Word Accent (F(2,170) = 13.94, p < .001), and the three-way interaction between Accent Type, Vowel Length, and Syllable Count (F(1,170) = 12.78, p < .001). The interaction is presumably a lexical effect: due to segmental or
idiosyncratic factors, some words were pronounced with a higher pitch than others. The difference between cell means in this interaction is in no case greater than 10 Hz.

The effects of Preceding Word Accent and Vowel Length are probably temporal. Peaks are slightly higher on long vowels (230 Hz) than on short vowels (224 Hz). F0 values are slightly lower when the preceding peak is Early in the preceding word than when the preceding peak is Middle or Late: 221.3 Hz for Early preceding peak, 233.0 Hz for Middle, 233.6 Hz for Late. Note that this is different from the finding for Peak location: there is an effect of Preceding Context on F0 height, but not F0 location.

Peak F0 does not consistently distinguish the two accent types: there is no main effect of Accent Type. Mean Peak F0 for the Rising Accent is 228 Hz, and Mean Peak F0 for the Falling Accents is 230 Hz.

Valley F0. The statistical results for Valley F0 mirror those for Peak F0. There is again a small, presumably lexical, three-way interaction of Accent Type, Vowel, Length and Syllable Count (F(1,170) = 5.96, p = .016). The effect of Preceding Word Accent is highly significant (F(2,170) = 29.22, p < .001). The closer the two peaks are to one another, the higher the level of the valley between them: 187 Hz for Early preceding peak, 198 Hz for Middle, 208 Hz for Late.

Summary of Results of the Mixed Model Analyses.

1. Which variables reliably distinguish Rising and Falling accents?

Three variables were found to reliably distinguish Rising and Falling accents: Peak Location, R_Beg Location, and Minima Separation. Consistent with the results of Smiljanić (2002), both Peak Location and R_Beg Location were consistently later for Rising than for...
Falling accents, except on disyllables with a short initial vowel. In that context, the two accents were not distinguished by the timing of minima and maxima. Related to the finding of delayed rises for Rising accents, the two accents were found to have distinct shapes: Rising accents were more likely to be composed of a plateau followed by an upturn (Figure 3), while Falling accents were more likely to have a smooth rise (Figure 2). While some Falling accents had regions of flat pitch, these were shorter than plateaus for Rising accents with comparable peak-to-peak durations.

Also consistent with the results of Smiljanić (2002), Rising and Falling accents did not consistently differ in F0 height of either peaks or valleys. Nor did they differ in the location of the end of the fall from the peak on the first word (F_End).

2. How are these variables affected by word-internal variation in syllable and word structure?

Significant effects of Syllable Count and Vowel Length were found for both Peak Location and R_Beg Location, but these effects were different for the two accent types (Figures 4 and 5). The peak of the Falling accent was realized on the first syllable in all contexts, but was slightly later on short vowels than long vowels. R_Beg location for the Falling accents did not vary according to syllable count or vowel length.

For the Rising accents, syllable affiliation of the peak differed by syllable count. In trisyllables the peak was realized on the second syllable of the word, but in disyllables the peak was realized on the first syllable of the word, and the preceding minima also moved accordingly. (These parallel changes in syllable affiliation, we would argue, account for the significant correlations in L and H location found by Smiljanić 2002). In disyllables with long vowels, accent location for Rising and Falling accents remained distinct, with the Rising peak reached
late in the syllable and the Falling peak reached early. As noted above, however, in disyllables with short vowels the two accent types were not distinguished by the location of either minima or maxima.

There were also small effects of word-internal variation on Minima Separation and Peak F0. On long vowels and trisyllables, plateaus were more likely and F0 peaks were slightly higher. These and other variables (F_End and Valley F0) were also found to exhibit two- and three-way interactions that can most likely be explained by segmental or lexical differences.

3. How are these variables affected by word-external variation, specifically distance from the preceding accent?

Distance from the preceding peak had no effect on Peak Location in the target word for either Rising or Falling Accents. On the other hand, it did have a significant effect on F0 height of both peaks and valleys: both became slightly higher as the peaks on the first and second word moved closer. Preceding Word Accent had no significant effect on Minima Separation.

In contrast to Peak Location, the location of F_End was strongly affected by Preceding Word Accent: F_End moved later in the target word when the peak was on the final syllable of the preceding word. There was also a significant effect of Preceding Word Accent on R_Beg, in addition to the significant word-internal effects on this variable described above.

To summarize, Hypotheses 1 – 3 were partially supported. Phrase-final trisyllabic words followed the prediction: the peak of the Rising accent occurred later than the peak of the Falling accent, with the peak of Falling accent occurring on the tonic syllable and the peak of the Rising accent occurring on the post-tonic (Figure 4a). In phrase-final disyllables, however, the peak of the Rising accent was retracted to the first syllable of the word (the tonic), contra Hypothesis 3.
If the stressed syllable contained a long vowel, the two accent types remained distinct, with the peak of the Falling accent realized early in the initial vowel and the peak of the Rising accent realized late. If the initial syllable of a disyllabic word contained a short vowel, however, the difference between Falling and Rising accents was neutralized (Figure 4b). (The only difference found is that a plateau may be more likely for Rising accents (Table 9), suggesting that neutralization is common but isn't obligatory.) Thus Hypothesis 6 is not supported: tone retraction leads to neutralization in phrase-final disyllables with a short initial vowel. The lack of neutralization in the long-vowel context is consistent with the findings of Smiljanić (2002:182-183) who tested only forms with stressed long vowels. She did not, however, test forms with short stressed vowels, and thus missed the neutralization that takes place in this context.

Hypothesis 4a is not supported. A pitch minimum is consistently found between the two peaks. The crucial question, however, is whether or not this pitch minimum is the L of a bitonal LH pitch accent, as claimed by Smiljanić (2002) for both Rising and Falling accents and by Godjevac (2000) for Rising accents only.

The data analyzed thus far provide no evidence that Falling accents are bitonal. For Falling accents, locations of the maxima and minima did not move in parallel and were affected by different variables: Peak Location was not affected by Preceding Context, while R_Beg Location was. This pattern is consistent with a simple H lexically associated to the first syllable of the word. Thus Hypothesis 4b is supported for the Falling accent.

Results for Rising accents are more complex. As predicted, the peak of Rising accent was consistently later in the target word than the peak of the Falling accent (disyllables with short vowels excluded). However, an additional difference in shape was also found: in the three
contexts where there was no neutralization, over 90% of Rising accents were realized with a plateau followed by an upturn, rather than a smooth rise. The finding of a plateau in these accents argues for a second specification, but the evidence thus far does not indicate that this specification is lexical.

In particular, Preceding Context had no significant effect on Peak Location, but it did have a significant effect on R_Beg. The fact that minima but not maxima are affected by Preceding Context suggests that they are not part of a lexical unit.

Peak location did change categorically, from second syllable to first, depending on syllable count, and R_Beg location moved correspondingly. Locations of Peak2 and R_Beg also varied based on vowel length in disyllables: both were late in the syllable on long vowels, and early in the syllable (indistinguishable from Falling accents) on short vowels. This effect, it will be argued below, is phonological, due to interactions with tones associated to the intonational phrase boundary, and would apply whether the Rising accent was represented as simple or bitonal.

The next section of the analysis now turns to specific tests of whether the locations of R_Beg and Peak2 correlate with each other across word-internal and word-external variation.

**Correlations**

Preceding analysis has shown that Peak Location and R_Beg Location differentiate Rising and Falling accents, and both are affected by differences in syllable count and vowel length. We turn now to what determines the location of these points. With what anchors are they correlated? If an accent is bitonal, the relationship between L and H is expected to be stable: that is, the locations of the pitch minimum and maximum will be correlated with each
other.

To test these hypotheses, the correlations of Peak Location and R_Beg Location to each other and to the boundary between the first and second syllable of the target word are tested. Each vowel-length/syllable-count context is considered separately, as justified by the significant Accent*VowelLength*SyllableCount interaction on both R_Beg and Peak2. Then the correlation shows how the minima and maxima move (or don't move) in the same way due to different preceding contexts, different subjects, and token-to-token variation.

Table 12 gives the correlations for the Falling accents. Each cell gives the Pearson Correlation (R) for the column and row variables.

**Table 12. Pearson Correlation (R) for minima, maxima, and syllable boundary locations (all computed in ms from the word boundary) for Falling accents.** * = p < .05 ** = p < .01

<table>
<thead>
<tr>
<th>Context</th>
<th>R_Beg</th>
<th>Peak2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long trisyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.054</td>
<td>.752**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.142</td>
<td></td>
</tr>
<tr>
<td>Short trisyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.284</td>
<td>.358 (p = .079)</td>
</tr>
<tr>
<td>Peak2</td>
<td>.219</td>
<td></td>
</tr>
<tr>
<td>Long disyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.236</td>
<td>.485*</td>
</tr>
<tr>
<td>Peak2</td>
<td>.064</td>
<td></td>
</tr>
<tr>
<td>Short disyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>-.024</td>
<td>.945**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.014</td>
<td></td>
</tr>
</tbody>
</table>

Table 12 shows that for the Falling accent, Peak Location tends to be correlated with the location of the syllable boundary. That is, it is located at a consistent percentage of the syllable length in each context. The strength of the correlation varies – for short trisyllables, the correlation only approaches significance, while for short disyllables and long trisyllables, it is highly significant. There is, however, no evidence of a correlation between the locations of the
beginning of the rise to the peak and the peak on the target word, nor a correlation between the beginning of the rise and the syllable boundary.

The conclusion follows that for Falling accents, the minimum at the beginning of the rise does not form a phonological unit with the peak, and it is not aligned with any particular point in the accented syllable. Its occurrence is governed by the location of the preceding peak. The F0 minimum might be the result of a phonological L that is part of the preceding accent, or an L associated to the word boundary; or it could be the phonetic result of the absence of a phonological pitch target. The preceding findings, however, in showing that the minima and maxima of the Falling accent move separately in response to contextual changes, do not support an analysis in which the Falling accent is a bitonal L+H.

Table 13 gives the results for the correlations with the Rising accents.

**Table 13. Pearson Correlation (R) for minima, maxima, and syllable boundary locations (all computed in ms from the word boundary) for Rising accents. * = p < .05 ** = p < .01**

<table>
<thead>
<tr>
<th></th>
<th>R_Beg</th>
<th>Peak2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Long trisyllable</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.564**</td>
<td>.803**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.499**</td>
<td></td>
</tr>
<tr>
<td>Short trisyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.771**</td>
<td>.868**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.713**</td>
<td></td>
</tr>
<tr>
<td>Long disyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.261</td>
<td>.759**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.277</td>
<td></td>
</tr>
<tr>
<td>Short disyllables</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Syllable boundary</td>
<td>.156</td>
<td>.669**</td>
</tr>
<tr>
<td>Peak2</td>
<td>.377</td>
<td></td>
</tr>
</tbody>
</table>

Table 13 shows both similarities and differences compared to Table 12. As was the case with the Falling accents, the location of Peak2 is correlated with the location of the syllable boundary: more strongly so for the Rising than for the Falling accents, in 3 of 4 cases. The
difference is that highly significant correlations are also found with R_Beg, both to Peak2 and to the syllable boundary, but only for trisyllabic words.

In order to elucidate these similarities and differences, Figures 10 – 13 graph the correlations between Peak2 and the syllable boundary for each context. In each case, the diagonal indicates the line where the location of Peak2 equals the location of the syllable boundary.

![Figure 10. Correlation between Peak2 and Syllable boundary, long trisyllables.](image)

**Trisyllables, Long Vowel**
- O Falling accents
- ▲ Rising accents

Figure 10. Correlation between Peak2 and Syllable boundary, long trisyllables.
Figure 11. Correlation between Peak2 and Syllable boundary, short trisyllables.

Figure 12. Correlation between Peak2 and Syllable boundary, long disyllables.
Figure 13. Correlation between Peak2 and Syllable boundary, short disyllables.

In these graphs, the significant correlations between Peak2 and the syllable boundary are clear, and the locations relative to the syllable boundary that were graphed in Figure 4 are confirmed. Also clear is that the correlation is stronger for the Rising accents than for Falling accents. Figure 13 shows that peaks for the two accent types are not distinct on short disyllables.

Figures 14 – 17 graph the correlations between Peak2 and R_Beg in each context. In each case, a strong correlation is expected for a bitonal accent. All points are above the diagonal, since the peak is always later than the minimum.
Figure 14. Correlation between Peak2 and R_Beg, long trisyllables.

Figure 15. Correlation between Peak2 and R_Beg, short trisyllables.
Figure 16. Correlation between Peak2 and R_Beg, long disyllables.

Figure 17. Correlation between Peak2 and R_Beg, short disyllables.
For the trisyllables (Figures 14 and 15), the correlation of R\_Beg and Peak2 for the Rising accents, and the difference between the two accent types, is clear. In Figure 14, there are three outliers among the Rising accents, with unexpectedly early R\_Beg. Inspection of these tokens shows them to be cases of Rising accents with no plateau (F\_End = R\_Beg). They have a long, gradual, irregular rise from a single minimum to a late maximum, but no plateau that met the labeling criteria for separation of F\_End and R\_Beg. Two similar outliers are seen in Figure 16.

For the disyllables (Figures 16 and 17), the lack of correlation between minimum and maximum for both accent types is evident. While there is good separation between the accent types on the long disyllables, again the lack of a difference between Rising and Falling accents is seen on short disyllables.

Why the difference between the trisyllables and disyllables? One crucial difference is that in the trisyllables, the syllable boundary intervenes between minimum and maximum for the Rising accents. The beginning of the rise occurs at the end of the first syllable (Figure 5), and the peak near the beginning of the second syllable (Figure 4). The correlations (Table 13) show that R\_Beg and Peak2 are correlated with each other only when both are also correlated with the syllable boundary: it's both or neither. It is argued below that R\_Beg and Peak2 are independently aligned with the syllable boundary, not directly to each other.

Thus the long disyllables (Figure 16) prove to be the crucial case for examining whether the Rising accents are bitonal. The two accent types remain distinct in this context, and 96% of Rising accents are realized with a plateau, that is, separation of F\_End and R\_Beg, but both R\_Beg and Peak2 occur within the same syllable. It is exactly in this context that a high
correlation between R_Beg and Peak2 would be expected, if the accent was bitonal. But no significant correlation is found. (Removing the two outliers only makes the correlation worse, not better.)

General conclusions concerning the correlations are as follows.

1. For both Rising and Falling accents, location of Peak2 is correlated with the location of the syllable boundary. This consistent alignment, based on word-internal factors, is what is predicted for a lexically-specified tone. The correlation is stronger for Rising accents than for Falling.

2. For Rising accents on trisyllables, there is a correlation between R_Beg and both Peak2 and syllable boundary. This can be seen as independent alignment of both minima and maxima to the boundary, not correlation of R_Beg and Peak2 to each other. When separated out by syllable count and vowel length, which determine phonological association, the locations of R_Beg and Peak2 are not correlated with each other for Falling accents in any context. For Rising accents the two points are not correlated on disyllables, the crucial case where no syllable boundary intervenes. Thus we conclude that neither accent is a bitonal L+H: Hypothesis 5 is not supported.

3. The contrast between Rising and Falling accents is usually neutralized on short disyllables.

While the evidence from the correlations does not support an analysis of either accent as bitonal, there is a difference in pitch shape between Rising and Falling accents that cannot be accounted for as a simple difference in peak alignment. The evidence shows that Rising and Falling accents differ not just in the location of their peaks, but in the location of the minima
preceding the peak. Rising accents are consistently realized with a plateau on the tonic syllable.

It will be argued below that Rising accents do have an additional L pitch target, which results in two separate minima with a plateau between them. The correlations indicate, however, that the additional specification is not the first part of a bitonal accent.

The next section now turns to the phonological analysis.

PHONOLOGICAL ANALYSIS

The phonological analysis proposed in this section is based on the experimental findings reported above. The evidence shows that neither of the two pitch accents is lexically bitonal, and thus it is proposed that, phonologically, both the Falling and the Rising accent are represented as simple H. The H location is specified lexically and can fall on any syllable within a word. If it falls on the initial syllable, as in (1), the accent is Falling. If it falls on any non-initial syllable, the accent is Rising, as in (2).

(1) Falling accent, trisyllabic word with short initial vowel (e.g., nenada)

\[
\begin{array}{l}
\text{H} \\
\sigma \quad \sigma \quad \sigma \\
\text{ne} \quad \text{na} \quad \text{da}
\end{array}
\]

(2) Rising accent, trisyllabic word with short initial vowel (e.g., lavove)

\[
\begin{array}{l}
\text{H} \\
\sigma \quad \sigma \quad \sigma \\
\text{la} \quad \text{vo} \quad \text{ve}
\end{array}
\]
The place of stress is predictable from the place of tone: in Falling accents, stress is on the initial syllable, and coincides with the location of H. In Rising accents, stress is on the syllable immediately preceding the one bearing H. Falling accents are thus ‘monosyllabic’, with both H and stress residing on the same syllable, while Rising accents are ‘bisyllabic’, with the place of H separated from the place of stress. This analysis, which echoes the traditional view (Browne & McCawley 1965, Garde 1976, Ivić 1976, Inkelas & Zec 1988) captures in a straightforward fashion the distribution of pitch accents presented in Table 1.

As shown in Table 14, the tonal representation of the two pitch accents as single lexical H fully captures the asymmetries in the distribution of the Falling and Rising accents. (In this and the following table, "c" stands for any consonant, "a" for any vowel, and the acute accent indicates stress.)

<table>
<thead>
<tr>
<th></th>
<th>Monosyllables</th>
<th>Polysyllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial σ</td>
<td>Medial σ</td>
</tr>
<tr>
<td>‘Falling’</td>
<td>c á c</td>
<td>c á c a (c a)</td>
</tr>
<tr>
<td>‘Rising’</td>
<td></td>
<td>c á c a (c a)</td>
</tr>
</tbody>
</table>

As already noted, given this tonal representation, stress falls on the syllable that immediately precedes the one specified for tone, if there is one, otherwise on the initial syllable. Thus, the distribution of stress is predictable from the place of H, but the place of H is not predictable from stress.

The gaps in Table 14 are systematic because the occurring forms exhaust the possible associations between syllables and H tones. In polysyllables, there is an apparent contrast between Rising and Falling accents only on words with stress on the initial syllable: H.
associated to the first syllable results in initial stress with a Falling accent, H associated to the second syllable results in initial stress with a Rising accent, and H associated to any subsequent syllable results in non-initial stress with Rising accent. Thus stress cannot fall on the final syllable of polysyllables. And, it follows straightforwardly that monosyllables can only bear a Falling pitch accent. (See Zec & Zsiga 2010 for an Optimality Theoretic analysis of the interactions of tone and stress in Serbian, with potentially conflicting constraints on the stressed syllable to align with the left edge, and to coincide with the syllable associated with H which insures its tonal prominence.)

Although we argue on phonetic grounds against a bitonal characterization of the two pitch accents, it is worth showing how such analyses fare in capturing the overall distribution. For example, Smiljanić’s (2002) characterization of Rising and Falling accents as L*H and LH* respectively yields the distributions shown in Table 15.

**Table 15. Distribution of Pitch Accents represented as L*H and LH***

<table>
<thead>
<tr>
<th></th>
<th>Monosyllables</th>
<th>Polysyllables</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Initial σ</td>
<td>Medial σ</td>
</tr>
<tr>
<td>‘Falling’</td>
<td>c á c</td>
<td>c á c a (c a)</td>
</tr>
<tr>
<td></td>
<td>L_H*</td>
<td>LH*</td>
</tr>
<tr>
<td>‘Rising’</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>c á c a (c a)</td>
<td>c a c á c a</td>
</tr>
<tr>
<td></td>
<td>L* H</td>
<td>L* H</td>
</tr>
</tbody>
</table>

The shaded cells in Table 15 indicate gaps that, under this analysis, are accidental. There is no explanation of why the two melodies do not associate freely: specifically why L*H can not associate to monosyllables, why LH* can not associate to non-initial syllables, and why neither can associate to final syllables. Moreover, in this type of analysis, not only the LH melodies but also the place of stress needs to be lexically determined, prior to the alignment of melodies to metrically strong positions. As stated by Smiljanić and Hualde (2000: 473), dialects such as
Standard Serbian “have both lexical stress and, in the case of words with initial stress, a lexical contrast in the type of tonal contour that the stressed syllable may receive.” That is, both stress and tonal melody are unpredictable and therefore lexically specified. This is so because analyses cast in terms of an AM model extended to lexical interactions are formally committed to the assumption that the place of stress determines the distribution of tone. But, if both Falling and Rising pitch accents are represented as single lexical High tones, as we propose, the overall system is much simpler as the place of stress is predictable from the place of H tone.

We now turn to characterizing the phonetic differences between the Falling and Rising pitch accent, which will follow from the place of lexically designated H, the location of stress, and the effects of boundary tones. Maxima will be accounted for in a simple way: each will correspond to a lexical H. In order to account for the minima found in the present experiment, it is proposed that both words and phrases are delimited by a L boundary tone (as argued by both Godjevac 2000 and Similjanić 2002). There is thus a word-boundary L (symbolized $L_ω$) between the two peaks in all of the phrases studied in the present experiment, accounting for the consistent finding of a pitch minimum at or near the boundary between the two words. There is also a phrase-final boundary L (symbolized $L_Φ$), which must associate to the final mora of each target word. The proposed phrasal tones are diagrammed in (3). The example word would be a trisyllabic word with a short initial vowel, e.g. *lavove* and *nenada*. Following Hayes and Lahiri (1991), we indicate the boundary tones as linked to a boundary symbol.
The difference between Falling *nenada* and Rising *lavove* is due to the position of $H$. A pitch accent is realized as Falling if $H$ is specified on the word-initial syllable, and as Rising in all other cases.

The difference in pitch accent shape comes about as an interaction between tone and stress placement. The Falling accent is simple $H$ on the initial syllable, specifically, on its first mora, and the $H$ is aligned to the mora's right edge. For these accents the present experiment found that there was generally a smooth curve between the peaks of the two words. The valley between the peaks comes from the word initial boundary $L$, as well as from the distance from the preceding peak, which shapes the valley between the two peaks. If there is an area of flat pitch in Falling accents, it is short (often just on the initial sonorant) and arises from phonetic effects in the absence of phonological specification.

The proposed phonological representation of Falling accents in all four word types is shown in (4).

(4a) Falling accent, trisyllabic word with long initial vowel (e.g., *na:mere*)
(4b) Falling accent, disyllabic word with long initial vowel (e.g., mo:re)

\[
\begin{array}{ccc}
L_\omega & H & L_\phi \\
\# & \mu & \mu & \mu \\
\sigma & \sigma & \sigma \\
\text{mo:} & \text{re}
\end{array}
\]

(4c) Falling accent, trisyllabic word with short initial vowel (e.g., nenada)

\[
\begin{array}{ccc}
L_\omega & H & L_\phi \\
\# & \mu & \mu & \mu \\
\sigma & \sigma & \sigma \\
\text{ne} & \text{na} & \text{da}
\end{array}
\]

(4d) Falling accent, disyllabic word with short initial vowel (e.g., mamu)

\[
\begin{array}{ccc}
L_\omega & H & L_\phi \\
\# & \mu & \mu \\
\sigma & \sigma \\
\text{ma} & \text{mu}
\end{array}
\]

In Falling accents, the lexically-specified pitch prominence and left-aligned metrical prominence both fall on the initial syllable, thus insuring the tonal prominence of the stressed syllable. Further, there is no conflict between the H associated to the first mora and the boundary L associated to the final mora.
Forms with Rising accents differ lexically only in that the H is associated to a non-initial syllable, here, the second syllable. Another difference is that Rising accents are ‘bisyllabic’: stress falls on the syllable immediately preceding the H bearer. As is usually stated, the pitch peak of the Rising accent is realized on the post-tonic syllable. Thus in (most) forms with Rising accents, the stressed syllable does not coincide with the syllable that realizes the H as pitch maximum.

There is, however, an additional L target in Rising accents that is absent from Falling accents. That is, there is an L specification present on the tonic syllable of Rising accents. It is argued here, however, that the L is not lexically specified as part of a bitonal pitch accent, but is inserted post-lexically, due to the interaction with stress. Unlike the H target (Peak 2), which was stable with respect to the preceding context, this L target (R_Beg) was not; and as further shown, the locations of the two points were not correlated. It is argued that a L is inserted on the stressed syllable of a Rising accent in order for this syllable (in fact, every mora of this syllable) to be associated with some pitch prominence. (Repetti & Kim 2010 give a similar analysis of post-lexical stress shift in Sardinian.) A high tone is generally preferred for marking prominence (deLacy 2002), but H insertion is ruled out in Serbian by high ranking of OCP-H. No more than one H is allowed per word. The stress-driven insertion of L on the stressed syllable of words with Rising accent results in the plateau across the stressed syllable that was consistently found for these words.

The examples in (5) illustrate Rising accents on trisyllabic words.
(5a) Rising accent, trisyllabic word with long initial vowel (e.g., ne:manju)

(5b) Rising accent, trisyllabic word with short initial vowel

Thus, in trisyllables with Rising accents, the peak is realized early in the second syllable, and a plateau occurs across the first syllable.
In disyllabic words with Rising accents, yet another interaction occurs. This is attributed to the utterance-final phrasal L. In these words, the lexical H is associated to the last mora of the word. A constraint on realizing the phrasal L takes precedence over the H alignment, pushing the H back one mora, onto the first syllable in the words considered here. We first consider the case of disyllabic words with a Rising accent and a long initial vowel. In this case, the H is realized on the second mora of the long vowel, as close as possible to its lexical alignment. As with the trisyllables, a constraint requiring a prominence associated with every mora of the stressed syllable causes insertion of L on the first mora, resulting in the characteristic Rising accent shape. This is illustrated in (6)

(6) Rising accent, disyllabic word with long initial vowel (e.g., ne:nu)

Thus words like *ne:nu* are realized with a short plateau followed by a peak late in the first syllable.

If the first syllable has a short vowel, the contrast between Rising and Falling accent is lost on such words when they occur in phrase-final position (as shown by the phonetic results). This neutralization follows from the representations and interactions already proposed. The phrase-boundary L pushes the H back to the first syllable, where the H associates to the only mora available. This association then satisfies the constraint requiring an accent on the only mora of the stressed syllable, and no L is inserted. Thus, the contrast between Rising and Falling
accents is neutralized: the representation in (7) is the same as that in (4d): H on the first syllable and L on the second.

(7) Rising accent, disyllabic word with short initial vowel

\[
L_\omega \quad H \quad L_\Phi
\]

Note that the neutralization of contrast only follows if the mora is the tone-bearing unit in Serbian (see Myers 2003 for a similar conclusion about Kinyarwanda). Also note that if the Rising accent were represented as L*H and the Falling accent as LH*, this neutralization could only be captured as an arbitrary change from one accent type to the other. In contrast, representing both accent types as H allows the neutralization to be stated as a simple and well-motivated reassociation.

To summarize, the proposed phonological analysis accounts for the phonetic facts confirmed in the experiment:

1. The pitch minimum between the two peaks (F_end) is due to a word-boundary tone, \(L_\omega\) in (3). Its location is influenced by the proximity of the peak on the first word.
2. The contrast between Falling and Rising accents is neutralized on phrase-final disyllables with a short stressed vowel (7), due to tone retraction. In all other contexts:
3. The peak of the Rising accent is later than the peak of the Falling accent. The peak of the Falling accent consistently occurs near the middle of the tonic syllable (4a - d).
The peak of the Rising accent occurs at the beginning of the post-tonic syllable in trisyllables (5) and at the end of the tonic syllable in disyllables (6).

4. For both Rising and Falling accents, peak location is correlated with the location of the boundary between the first two syllables of the accented word. In all cases (4 – 7), no tones intervene between H and the syllable boundary.

5. Rising accents are realized with a plateau preceding the rise to the peak, due to the inserted L on the tonic (5a, 5b, 6).

6. In general, the location of pitch minima and maxima are not correlated. For Falling accents, the L is a word boundary tone and the H is a lexical tone. For Rising accents, the H is lexical and the L is inserted due to interactions with stress.

7. In trisyllabic words with Rising accents (5a and 5b), but no other cases, the (inserted) L and the (lexical) H are both adjacent to the boundary between the first and second syllables of the target word. They are both correlated with the location of the boundary, and thus also seem to be correlated with each other. In no other case is the L adjacent to this boundary.

CONCLUSION

The present paper has proposed a unified analysis of the phonological distribution and phonetic realization of Falling and Rising accents in Standard Serbian.

This distribution has a straightforward phonological account: both Rising and Falling accents consist of a single lexical H, and their distribution emerges from the interaction of stress and tone. If the High-toned syllable is initial, stress and tone coincide, yielding a Falling accent. Otherwise, the stressed syllable immediately precedes the High-toned one, yielding a Rising
accent. Thus the Falling and Rising pitch accents appear to contrast only on the initial syllables of polysyllabic words. If Rising and Falling accents consist of contrastive tonal melodies, then the absence of Rising accents on monosyllables and the absence of Falling accents on non-initial syllables is difficult to explain.

The current proposal thus improves on the analyses of Godjevac (2000) and Smiljanić (2002). In these studies the representations emerge directly from their phonetic findings, but no account of phonological distributions is proposed. The present study generally confirms the phonetic findings, particularly the instrumental acoustic analyses of Smiljanić, but offers new data from varied contexts that leads to a different phonological representation.

The current proposal is also an improvement over previous phonological accounts (Browne & McCawley 1965, Garde 1976, Ivić 1976, Inkelas & Zec 1988) that proposed lexical representations of pitch accents as a diacritic that can occur on any syllable in a word; diacritics that fall on the initial syllable will coincide with stress; and if a diacritic falls on a non-initial syllable, the immediately preceding syllable will be stressed. Rather than relying on abstract diacritic markings that have no phonetic correlates, the present analysis proposes that these "diacritics" are H melodies with real phonetic correlates. Lexically-specified H tones are realized phonetically as pitch maxima.

Lehiste and Ivić (1986) build into their proposal both the phonetic and phonological findings about the two pitch accents. The current proposal follows Lehiste and Ivić in proposing that the Falling and Rising pitch accents are both represented as pitch maxima, in our terms H, with different lexical associations.

Moreover, the interaction of tone and stress gives rise to the distinctive pitch shapes of the Rising and Falling accents. We have argued on several grounds that, within the typology of
tone/stress interactions, Standard Serbian exemplifies one of the three types of interactions: the place of lexical tone is a predictor of the place of stress. Further types include the place of stress predicting the place of tonal melody, and the trivial case of no interactions (de Lacy 2002, Hyman 2006, Zec 1999). Earlier approaches cast in the AM model (Godjevac 2000, Smiljanić 2002) analyze the Standard Serbian prosodic system as the type of interaction with stress predicting the place of tonal melody. These analyses, committed to the restricted set of tone/stress interactions accommodated within an AM model, do not account for the regularities in the distribution of Standard Serbian pitch accents. And finally, neutralization due to tone retraction provides a case where prosodic requirements override lexical specification which, as we argue, is due to tone crowding, enriching the cross-linguistic typology in yet another respect. To conclude, Serbian pitch accent provides an interesting test case of the interaction of lexical tone, stress, and prosody.

REFERENCES


