UDC 811.163.41'34
UDC 811.111'34
https://doi.org/10.18485/ms_zmsfil.2025.68.1.8
ORCID 0000-0003-3046-1257
Изворни научни рад

Bojana Jakovljević

F0 PROMINENCE AND SPEECH RHYTHM: THE INITIAL EXPERIMENTS BASED ON ENGLISH AND SERBIAN

The paper examines the relation between F0 prominence and speech rhythm. The research is based on English and Serbian as they differ in the realization of pitch accents in prosodic words. The research included two experiments. The production experiment involved measuring F0 prominence in English and Serbian initially-stressed di- and trisyllabic feet with falling F0 slope, realized as prosodic words within IPs. Prior to the perception experiment, F0 of Serbian IPs was manipulated according to the English measurements and the IPs were resynthesized. The perception experiment involved evaluating the rhythmicity of the original and modified Serbian IPs with uniform and non-uniform interstress intervals on a 5-point Likert scale by 30 Serbian L1 speakers. The production analysis points to a greater F0 prominence in English feet. The perception experiment shows that greater F0 prominence results in higher rhythmicity judgments.
Key words: speech rhythm, F0, foot, speech production, speech perception.

У раду се испитује природа везе између проминентности основног тона (f0) и говорног ритма. Истраживање је спроведено у енглеском и српском, будући да се дати језици разликују у реализацији акцента на нивоу прозодијске речи. Истраживање чине два експеримента. Експеримент продукције заснивао се на мерењу проминентности f0 у енглеским и српским иницијално наглашеним двосложним и тросложним стопама са силазним тонским кретањем, које су биле реализоване као прозодијске речи унутар интонацијских фраза. Пре експеримента перцепције, контура f0 српских интонацијских фраза модификована је према мерењима у енглеском, а затим су интонацијске фразе ресинтетизоване. Експеримент перцепције заснивао се на процени ритмичности српских оригиналних и модификованих интонацијских фраза које су садржавале или само двосложне стопе, или комбинацију двосложних и тросложних стопа. Процену је вршило 30 изворних говорника српског језика на петостепеној Ликертовој скали. Анализа продукције указала је на већу проминентност f0 у енглеским стопама. Експеримент перцепције показао је да већа проминентност f0 доприноси вишем вредновању ритмичности говора.
Кључне речи: говорни ритам, основни тон (f0), стопа, продукција говора, перцепција говора

1. Introduction. Speech rhythm is a universal prosodic property of languages, which plays a significant role in both speech production and speech perception. The first and still rather influential theoretical models of speech rhythm focus on temporal cues or, more precisely, on the concept of isochrony, which is viewed as an absolute rhythmic feature (Pike 1945; Abercrombie 1967). Despite the absence of any solid empirical evidence, the models in question established the first rhythmic classification of languages into either stress-timed or syllable-timed.

The empirical evidence for the traditional language typology was to some degree provided in cross-linguistic production studies which were based on temporal measurements (Ramus et al. 1999; Grabe – Low 2002). Although these studies did not reject the traditional classification of languages, they showed that isochrony is a gradual rather than absolute feature, which resulted in placing

languages along a continuum from maximum stress-timing to maximum syllable-timing. However, as in the models by Pike (Pike 1945) and Abercrombie (Abercrombie 1967), the classification is primarily based on temporal cues.

On the other hand, the first perception studies of speech rhythm reduced the significance of temporal cues by pointing to the mismatch between isochrony in speech production and speech perception, thus suggesting that speech rhythm is primarily a perceptual phenomenon (Lehiste 1977; 1979). The concept of 'perceived rhythmicity', i.e., the impression of rhythmic strength of speech (Barry et al. 2009), has proved to be rather significant as it contributed to the development of new lines of research in the studies of speech rhythm. Instead of focusing on duration as the only significant factor in the realization of speech rhythm across languages, these studies also considered the importance of other components of prosodic prominence, such as pitch (F0) (Dauer 1983; 1987; Barry et al. 2009; Arvaniti – Rodriquez 2013), intensity (loudness) (Barry et al. 2009), as well as the reduction of vowel quality in prosodically non-prominent syllables (Dauer 1983; 1987; Barry et al. 2009).

1.1. Traditional approaches to isochrony and rhythm typology. The first theoretical approach to speech rhythm was developed by Pike (Pike 1945), and later adopted by Abercrombie (Abercrombie 1967). The approach is essentially based on the notion of 'isochrony', i.e., the reoccurrence of stressed syllables at regular temporal intervals, which results in equal or approximately equal duration of interstress intervals. Consequently, the vowels in unstressed syllables are reduced in both quantity and quality. The concept of isochrony was the basis of a well-known rhythm typology according to which languages are classified as either 'stress-timed' or 'syllable-timed' (Pike 1945: 34–35; Abercrombie 1967: 97). Stress-timed languages such as English, Russian and Arabic are characterized by the presence of isochrony, i.e., the interstress intervals are of (approximately) equal duration and the vowels in unstressed syllables are both durationally and qualitatively reduced. On the other hand, languages such as French, Spanish, Telugu and Yoruba are described as syllable-timed, i.e., due to the absence of isochrony, the interstress intervals become longer as the number of unstressed syllables increases and the vowels in unstressed syllables are not reduced in either duration or quality. Equally important, these approaches view isochrony as well as stress-timing vs. syllable-timing as absolute rather than gradual phenomena. In other words, isochrony is either present or absent in a given language and, accordingly, a language can be either stress-timed or syllable-timed.

The classification of languages into stress-timed and syllable-timed is still rather influential despite the fact that its founders do not provide any empirical data to support it. Also, it is rather disputable whether this typology is based on speech production, speech perception, or both. Abercrombie (Abercrombie 1967) relies on the Pulse (or Motor) Theory (Stetson 1951) to account for the two classes of languages, which points to the production basis of the typology. According to the Pulse Theory, syllables correspond to 'chest pulses', i.e., the bursts of the activity of the rib cage muscles, which result in the increase in air pressure. As opposed to syllable-timed languages which are characterized by a periodic repetition of chest pulses coinciding with each syllable, stress-timed languages are

characterized by the reinforced periodic chest pulses which coincide with stressed syllables, thus causing the reduction of the unstressed ones (Abercrombie 1967: 97). However, Abercrombie (Abercrombie 1967) often mentions 'hearing' as well as 'feeling' the rhythm of languages with different timing properties, which suggests the perceptual basis of the typology. Therefore, the traditional stress-timing vs. syllable-timing language typology is to a certain extent impressionistic, lacking clearly defined theory and solid evidence.[1]

The dichotomy between stress-timed and syllable-timed languages, originally proposed by Pike (Pike 1945) was later modified by introducing the class of so-called 'mora-timed languages', such as Japanese (Bloch 1950; Han 1962; Ladefoged 1982). Essentially, the underlying mechanism of mora-timed languages is rather similar to the one of syllable-timed languages, i.e., syllable-timed languages are characterized by (approximately) the same duration of each syllable, whereas mora-timed languages are characterized by (approximately) the same duration of each mora.

1.2. Empirical evidence for isochrony and rhythm typology. The production studies which are based on the measurements of speech rhythm indicate that the models proposed by Pike (Pike 1945) and Abercrombie (Abercrombie 1967) are not empirically grounded. The studies in question show that absolute isochrony in stress-timed languages does not exist (Lea 1974) as well as that all human languages show a tendency towards isochrony and stress-timing (Dauer 1983). Lea (Lea 1974: 34–42) measures the duration of interstress intervals in English and finds that it is directly proportional to the number of unstressed syllables.[2] Dauer (Dauer 1983) measures the duration of interstress intervals in English, Thai, Spanish, Italian and Greek and shows that the tendency towards isochrony and stress-timing is present in all the analyzed languages. Also, the measurements indicate that the duration of interstress intervals containing the same number of syllables is not statistically different in English and Spanish despite the fact that these languages are traditionally classified as stress-timed and syllable-timed respectively. Accordingly, instead of relying only on duration cues, Dauer (Dauer 1983) proposes the set of phonetic and phonological factors which characterize the languages traditionally classified as stress-timed or syllable-timed. In contrast to syllable-timed languages, stress-timed languages are characterized by the reduction of vowel quality in unstressed syllables, which become more centralized. Another important phonetic factor is the realization of lexical stress which is rather complex in stress-timed languages as it includes many components of prominence such as pitch, loudness, vowel quantity and vowel quality. On the other hand, the realization of stress (or accent) in syllable-timed languages is most often reduced to the variations in pitch. As a result, the difference in overall prosodic prominence

[1] As suggested by Dauer (Dauer 1983), the absence of a clearly defined theoretical framework as well as the lack of any empirical evidence to support absolute isochrony and either-or language typology, makes the theoretical nature of the models by Pike (Pike 1945) and Abercrombie (Abercrombie 1967) rather disputable.

[2] The same tendency regarding the realization of rhythm in English had been proposed by Classe (Classe 1939) before the postulation of the models by Pike (Pike 1945) and Abercrombie (Abercrombie 1967).

between stressed and unstressed syllables is higher in stress-timed than in syllable-timed languages. The most significant phonological difference between stress-timed and syllable-timed languages concerns the syllabic structure. As opposed to syllable-timed languages which have rather simple CV syllables, stress-timed languages are characterized by the presence of both simple and complex syllables, i.e., syllables containing complex consonant clusters, which influences vowel duration as well.

The approach to isochrony as a gradual rather than absolute feature resulted in the development of more advanced measurement procedures which placed languages along a continuum from the highest degree of stress-timing to the highest degree of syllable-timing (Ramus et al. 1999; Grabe – Low 2002). The procedure by Ramus et al. (Ramus et al. 1999) is based on measuring the duration of the whole sentence as well as the duration of vocalic and consonantal intervals in a sentence and then calculating the proportion of vocalic intervals (%V) and the standard deviation of consonantal intervals (*Δ*C). The procedure obviously draws from Dauer's (Dauer 1983) observation concerning the differences in the syllabic structure in stress-timed vs. syllable-timed languages. As mentioned, stress-timed languages have complex consonant clusters whose duration varies considerably. At the same time, complex consonant clusters reduce the percentage of speech occupied by vowels or vocalic intervals. Accordingly, stress-timed languages are characterized by lower %V and higher *Δ*C than syllable-timed languages. Mora-timed languages have the highest %V and the lowest *Δ*C. The results of the measurements show that Polish, English and Dutch are (to a greater or lesser degree) stress-timed. Syllable-timed languages include French, Spanish, Italian and Catalan, whereas Japanese is classified as mora-timed.

Another measurement procedure (Grabe – Low 2002) is based on so-called 'pairwise variability indices' ('PVIs'). The procedure involves calculating the mean durational differences between vocalic and consonantal (intervocalic) intervals in the pairs of successive syllables ('raw PVI', 'rPVI'), and then dividing these differences by the mean duration of successive pairs ('normalized PVI', 'nPVI'). While rPVI is used for consonantal intervals, nPVI is used for vocalic intervals as vowels tend to be more susceptible to speaking rate than consonants. Regarding that stress-timed languages exhibit greater durational differences in successive syllables, they are characterized by higher PVI than syllable-timed languages. In the research based on PVIs, there were 18 languages examined (Grabe – Low 2002). The results show that stress-timed languages include British English, German, Dutch and Thai, whereas Spanish, French, Luxembourgish and Mandarin are classified as syllable-timed. Japanese was described as a mora-timed language although it exhibits the same properties as syllable-timed languages. The rest of the languages show either the properties of both stress-timing and syllable-timing (Polish, Estonian, and Catalan) or they remain unclassified (Greek, Romanian, Singapore English, Welsh, Malay, and Tamil).

The measurement procedures discussed above unquestionably provide a useful tool for measuring the production of rhythm cross-linguistically. At the same time, they make it possible to examine the subtle differences in the realization of rhythm in the languages traditionally classified as either stress-timed or syllable-timed. Instead, the languages are described as more or less stress-timed vs. more

or less syllable-timed (Ramus et al. 1999), or as mixed (Grabe – Low 2002). However, as Arvaniti (Arvaniti 2009) points out, the fact that both measurement procedures analyze rhythm only from the perspective of speech production as well as the fact that they rely only on temporal cues, thus neglecting the importance of other factors in the realization of speech rhythm seems to be their major disadvantage. Also, the reliance on temporal cues only could be the reason why many languages remain unclassified.

1.3. Isochrony and speech rhythm reconsidered. The view of isochrony and stress-timing vs. syllable-timing as gradual phenomena as well as the observation that all human languages tend towards isochrony and stress-timing considerably changed the initial attitudes towards speech rhythm. However, the overall approach to speech rhythm was crucially influenced by the results of perception studies. The first perception studies (Lehiste 1977; 1979) point to the discrepancy between isochrony in speech production and speech perception, eventually suggesting that rhythm is primarily a perceptual phenomenon. In a series of perception experiments, Lehiste (Lehiste 1977; 1979) tests the perception of synthetized utterances containing the intervals of both identical and different duration by English native speakers. Her main conclusion is that people hear what they expect to hear. In the utterances which contained the intervals of different duration, the subjects did not perceive the difference between them as they expected to hear isochronous utterances. At the same time, in the utterances which contained the intervals of identical duration, the last interval was consistently perceived as the shortest one because the subjects expected it to be the longest due to the process of final lengthening.

The concepts of perceived isochrony resulted in considering other factors which contribute to the perception of speech rhythm. One of these factors involves the differences in pitch prominence between stressed and unstressed syllables. As previously mentioned, the realization of lexical stress in stress-timed languages is rather complex as it involves manifold components of prominence such as pitch, loudness, vowel quantity and vowel quality, thus creating "a discernible beat" (Dauer 1983: 58). First, the use of the term 'beat' clearly points to the significance of perceived rhythmicity. And second, in view of the experiments which show that pitch has the most significant role in the perception of lexical stress in some languages such as English (Fry 1958) and German (Kohler 2009), it can be hypothesized that pitch prominence also affects the perception of speech rhythm which in these languages results from stress/unstress alternations.

This hypothesis was confirmed in the subsequent perception studies (Barry et al. 2009; Arvaniti – Rodriquez 2013). In the perception experiment by Barry et al. (Barry et al. 2009), the components of prominence were varied in trochaic logatomes (nonsensical *dada* sequences) and the rhythmicity of the utterances containing the modified logatomes was evaluated by Bulgarian, German and English native speakers. The results of the research showed that duration and F0 have the major role in the perception of rhythmicity, whereas intensity and vowel quality are less significant.[3] In addition, the research revealed that the ranking of

[3] Almost identical ranking was observed in the perception of English lexical stress (Fry 1958; 1965), which suggests that the components of rhythmic prominence in English follow a similar pattern as the components of word-stress prominence.

the components of prominence depends on the listeners' L1. For the native speakers of German and English, which are stress-timed languages, duration was slightly more informative than F0, whereas for the native speakers of Bulgarian, which is considered to be syllable-timed, the two components of prominence were equally informative.

The study by Arvaniti and Rodriquez (Arvaniti – Rodriquez 2013) examines the influence of language-inherent speaking rate and F0 prominence on the perceptual discrimination of English from rhythmically similar (stress-timed) Polish and Danish as well as from rhythmically different (syllable-timed) Spanish, Greek and Korean by English native speakers. The initial hypothesis of the research was that each language has its own inherent speaking rate which does not necessarily correspond to stress-timing vs. syllable-timing typology, but which together with F0 prominence contributes to the perception of rhythm. The utterances from the above-mentioned languages were converted into *sasasa* logatomes and the information about speaking rate and F0 was either retained (original stimuli) or removed (modified stimuli). The results showed that both language-inherent speaking rate and F0 prominence significantly contribute to the perceptual discrimination of English from other languages.[4] Although the effect of speaking rate was more consistent, the discrimination based on F0 prominence only was also confirmed.

Therefore, it can be concluded that speech rhythm represents a rather complex prosodic feature which plays an important role in both speech production and speech perception and whose realization does not depend only on temporal cues, as it was initially claimed, but on a variety of different factors, including F0 prominence.

2. Rhythm in English and Serbian. Both English and Serbian exhibit the properties of stress-timed languages. English is labelled as a stress-timed language in the traditional models of speech rhythm, which are based on either-or approach to isochrony (Pike 1945; Abercrombie 1967: 97) as well as in the empirically grounded and more advanced classifications which view isochrony as a gradual feature (Ramus et al. 1999; Grabe – Low 2002). Also, English exhibits all the properties which Dauer (Dauer 1983) relates to stress-timed languages. Therefore, it is characterized by the presence of both simple and complex syllabic structures, the reduction of vowel quantity and quality in unstressed syllables, i.e., vowels are centralized and most often reduced to [ə] or syllabic consonants, as well as by a rather complex realization of lexical stress which involves pitch, intensity, vowel quantity and vowel quality. In addition, English is prone to the process of compensatory shortening, i.e., the shortening of stressed syllables (or the nuclei of stressed syllables) as the number of unstressed syllables in a foot increases, which is another significant feature of stress-timed languages (Bertinetto 1989).

The analysis of Serbian rhythm also points to the presence of compensatory shortening, which is especially significant in the production of long vowels (Marković – Milićev 2011). Also, the vowels in unstressed syllables undergo the

[4] It is also important to mention that language non-inherent, i.e., speaker-dependent, speaking rate can influence the perception of rhythm, i.e., faster speech tends to be perceived as more rhythmical (Ordin – Polyanskaya 2015).

reduction in both quantity (Sovilj-Nikić 2007: 55) and quality (Marković 2016), i.e., the vowels become more centralized and rather similar to the English vowel [ə], as the number of unstressed syllables in a foot increases. The measurements of formant frequencies show that the reduction of vowel quality takes place in unstressed syllables both preceding and following the stressed syllable in a foot, although the reduction tends to be greater in the former case (Marković 2016).

However, English and Serbian differ with respect to the degree of stress-timing, which was observed in the studies of L2 English acquisition by Serbian native speakers (Marković 2011a). The measurements of the production by Serbian learners of L2 English show that the shortening of unstressed vowels is directly proportional to the number of unstressed syllables in a foot, which further supports the tendency towards stress-timing in Serbian. However, as the degree of shortening is higher in the production by L1 English speakers, Marković (Marković 2011a) concludes that both English and Serbian exhibit the properties of stress-timed languages, although the degree of stress-timing tends to be higher in English than in Serbian.

Shifting focus to the role of F0 in the perception of speech rhythm, it is expected that the changes in F0 prominence in a foot will influence the rhythmicity judgments in both English and Serbian. The relation between F0 prominence and perceived rhythmicity in English has already been proved (Barry at al. 2009; Arvaniti – Rodriquez 2013). Similarly, we predict that greater F0 prominence will contribute to higher evaluations of rhythmicity in Serbian, but only on the condition that the highest F0 peaks coincide with stressed syllables, which is not always the case in the Serbian accentual system (see Section 3).

3. Word accent in English and Serbian. English word accent (also referred to as 'word-stress' or 'lexical stress') is dynamic, i.e., stressed syllables are produced with considerable expiratory force, which results in greater intensity and therefore loudness (Ohala 1977). However, in addition to intensity, English word-stress is characterized by a complex set of other components of prominence, including pitch, or rather pitch movement, vowel quality, and vowel quantity (Roach 1991: 86; Collins – Mees 2003: 109–110; Cruttenden 2008: 236–237).[5]

On the other hand, Serbian is a pitch accent language in which stress is combined with tones, i.e., the variations in pitch (Inkelas – Zec 1988; Godjevac 2005; Zec – Zsiga 2010). Serbian has 4 pitch accents which are referred to as 'long falling', 'long rising', 'short falling' and 'short rising' and which can be used to distinguish among word meanings (Subotić et al. 2012: 98). Long and short accents are primarily differentiated on the basis of distinctive length of the vowel in a stressed syllable.[6] The difference between falling and rising accents primarily depends on F0 of the vowel in the syllable which immediately follows the stressed one, i.e., lower F0 of the vowel in the following syllable is the property of falling accents,

[5] These components of prosodic prominence refer only to the realization of English primary stress.

[6] Still, it should be mentioned that distinctive quantitative differences, at least in the production of some stressed vowels, can be accompanied by qualitative differences, i.e., the differences in the formant structure, which was observed in the variant of Standard Serbian spoken in Novi Sad (Marković – Sredojević 2021).

whereas higher F0 characterizes rising accents (Лехисте – Ивић 1996: 53–69; Godjevac 2005; Петровић – Гудурић 2010: 118–122; Zec – Zsiga 2010; Средојевић 2011: 32–33, 42–43; Sredojević 2017: 79–83).

Another difference between the accentual systems of English and Serbian concerns the correlates of stress. As mentioned, English stressed syllables are, among other factors, consistently characterized by greater intensity. As for Serbian, the most significant correlate of stress is duration, i.e., stressed syllables are longer than the unstressed ones (Ивић – Лехисте 2002: 17, 24, 32, 36; Јокановић-Михајлов 2007: 91–93; Zec – Zsiga 2010; Sredojević 2017: 223). The intensity, on the other hand, is not fully reliable correlate of stress in Serbian as stressed syllables are not necessarily produced with greater intensity, which is especially the case with the words in IP-initial position (Sredojević 2017: 204). According to Ivić and Lehiste, rising accents, unlike the falling ones, can exhibit the ascending intensity relation between the stressed and the following syllable (Ивић – Лехисте 2002: 41–45), whereas Jokanović-Mihajlov finds that the stressed and the following syllable are characterized by either equal or ascending intensity in the case of both falling and rising accents (Јокановић-Михајлов 2007: 71–96).

Finally, English and Serbian accentual systems differ with respect to the co-occurrence of a stressed syllable and F0 maximum of a word. English stressed syllables always coincide with F0 maximums, whereas Serbian exhibits this tendency only in the words with falling accents (Zec – Zsiga 2010). To put it another way, while the co-occurrence of stress and F0 maximum in English words is consistent, Serbian lacks this constancy due to a complex interplay of stress and tone. Therefore, we expect a more significant role of F0 in the realization of stressed syllables in English than in Serbian. Specifically, we predict that English stressed vowels will exhibit greater F0 prominence at the lexical level than the comparable Serbian vowels, i.e., Serbian stressed vowels in the words with falling accents.

4. The design of the research. Based on the predictions about F0 prominence in English and Serbian lexically stressed syllables, i.e., the syllables which are perceived as the most salient in lexical (or more generally prosodic) words, as well as the role of F0 prominence in the perception of speech rhythm, the paper aims at testing the following hypotheses:

> Hypothesis 1 (H1): If F0 maximum coincides with a stressed syllable, F0 prominence in a foot will be greater in English than in Serbian;
> Hypothesis 2 (H2): If F0 maximum coincides with a stressed syllable, F0 manipulation in Serbian feet in accordance with English production measurements will contribute to higher rhythmicity judgments in Serbian.

The research consisted of production and perception parts. In the production experiment, we measured the difference in the maximum and mean F0 values between the vowel in a stressed syllable (V1) and the vowel in the immediately following unstressed syllable (V2) in English and Serbian metrical feet. In both languages, the measurements were conducted in the initially-stressed disyllabic and trisyllabic feet which corresponded to the initially-stressed disyllabic and trisyllabic prosodic words (DSWs and TSWs). The difference in F0 maximum and

F0 mean between V1 and V2 will from this point on be referred to as the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) respectively.

Both English and Serbian DSWs and TSWs had CV syllabic structure. Also, both analyzed vowels were preceded by a voiced consonant, as voiceless consonants can cause rather abrupt and inconsistent pitch increase at the onset of the following vowel (Maddieson 1996: 166–167), which results in the inability to determine the true F0 maximum.[7] The research was limited to English words with short V1 and short V2 and Serbian words with short falling accent. The reason for choosing short falling rather than long falling accent was practical, as the words with short falling accent, CV syllabic structure and a voiced consonant in syllable onset were more frequent. In order to overcome the differences between male and female speakers, the values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) were expressed in semitones (STs), i.e., the units of perceived pitch differences.

The prosodic words (feet) analyzed in the production experiment were extracted from IP non-final (non-focalized) and IP-final (focalized) position, ensuring that the number of words in the initial and medial IP-position is as similar as possible. The IPs were selected from the corpus of read speech produced by 5 English and 5 Serbian professional newsreaders (3 female and 2 male speakers per language). English and Serbian IPs were comparable regarding the recording quality, i.e., the sampling rate was 44.1 kHz, as well as the articulation rate (AR), i.e., according to the t-test, the difference in AR between English and Serbian IPs is not statistically significant (Table 1). Also, the results of One-Way ANOVA did not point to a statistically significant difference in AR among 10 subsets of IPs produced by individual newsreaders ($F(9,773)=1.628$, $p=0.103$). All the selected IPs were realized as matrix clauses with final information focus followed by a silent pause and they had falling intonation. The tokens analyzed in both languages included 200 DSWs and 200 TSWs, half of which were in IP non-final position, whereas the other half were in IP-final position. The measurements were conducted in *Praat* (version 6.0.29, Boersma – Weenink 2018) and the results of the measurements were statistically analyzed in SPSS 20.

Table 1. Mean AR (syll/sec) of the IPs produced by 5 English and 5 Serbian newsreaders.

N English	Mean English	SD English	N Serbian	Mean Serbian	SD Serbian	Mean Diff.	t	p (2-tailed)
388	5.51	0.607	395	5.56	0.703	-0.05	-1.159	0.274

The most frequent DSWs in the English corpus were (synchronically) simple adjectives and nouns ending in *-y* (e.g. *busy*, *ready*, *body*, *money*, etc.) as well as the agent nouns derived from monosyllabic verbs by adding the suffix *-er/-or* (e.g.,

[7] Although voiced consonants can cause the opposite process, i.e., the lowering of F0 at the onset of the following vowel, the influence of voiceless consonants on vowel pitch is regarded as considerably greater (Maddieson 1996: 166). The choice of voiced consonants in syllable onset is additionally motivated by a rather significant F0 increase after voiceless aspirated stops (Ladefoged 2003: 86–87) which are typical realizations of English voiceless stops at the beginning of both stressed and unstressed syllables, although the degree of aspiration tends to be higher in the former case (Cruttenden 2008: 236–237).

bomber, *robber*, etc.). The set of TSWs primarily included adverbs with the suffix *-ly* (e.g., *busily*, *readily*, etc.) and agent nouns derived from disyllabic verbs (e.g., *visitor*, *manager*, etc.). Both DSWs and TSWs included frequent personal names (e.g., *Johnny*, *Danny*, *Jennifer*, etc.). Other common words in the English corpus were *many*, *bigger*, *manner*, *bother*, *register* (noun and verb), *legacy*, *remedy* (noun and verb), *misery* and *malady*. As for Serbian, the most frequent DSWs in the corpus were the comparative forms of the adjectives *mali* (small), *dobar* (good), *dug* (long), *nizak* (low) and *dalek* (far) with different gender, number and case features (e.g. *manji*, *manja*, *manje*, *manju*, *bolji*, *bolja*, *bolje*, *bolju*, *duži*, *duža*, *duže*, *dužu*, *niži*, *niža*, *niže*, *nižu*, *dalji*, *dalja*, *dalje*, *dalju*).[8] Other frequent DSWs were the 3rd person singular forms of the verbs *žaliti* (to regret), *delati* (to act), *nuditi* (to offer) and *voditi* (to lead) in the present tense (e.g., *žali*, *dela*, *nudi*, *vodi*). Regarding TSWs, the most frequent ones were some infinitival and 1st person plural forms of the abovementioned verbs in the present tense (e.g., *žaliti*, *delati*, *nuditi*, *žalimo*, *delamo*, *nudimo*, *vodimo*) as well as the deverbal nouns *žaljenje* (regret), *delanje* (acting, action) and *viđenje* (view, opinion). Other highly recurrent words in the Serbian corpus were the nouns *vođa* (leader) and *godina* (year), which had different forms with respect to case and number features (e.g., *vođa*, *vođi*, *vođu*, *vođe*, *vođama*, *godina*, *godini*, *godinu*, *godine*). The examples of some IPs selected from the English and Serbian corpus, which contain the above listed words in both IP-positions are given in (1) and (2) respectively.[9]

(1) a. [Two months later,]$_{IP}$ [the Boston Marathon *bomber* was sentenced to death.]$_{IP}$
b. [Nonetheless,]$_{IP}$ [federal government is pretending to look *busy*.]$_{IP}$
c. [In 2020,]$_{IP}$ [*Jennifer* received the Nobel Prize in Chemistry.]$_{IP}$
d. [In cases like this,]$_{IP}$ [the employees should speak to the *manager*.]$_{IP}$

(2) a. [Međutim,]$_{IP}$ [njihove *vođe* zagovaraju nasilje.]$_{IP}$
however their leaders promote violence
'However, their leaders promote violence.'
b. [Kako je već najavljeno,]$_{IP}$ [prosečne neto zarade će biti *manje*.]$_{IP}$
as was already announced average net earnings will be smaller
'As already announced, the average net earnings will be lower.'
c. [A trenutno,]$_{IP}$ [*vodimo* razgovore sa evropskim zvaničnicima.]$_{IP}$
and currently lead.Pres.1pl conversations with European officials
'And currently, we are leading conversations with European officials.'
d. [Prema njegovim rečima,]$_{IP}$ [trenutna situacija zahteva *delanje*]$_{IP}$
according-to his words current situation requires action
'According to his words, the current situation requires (taking) action.'

The perception part of the research involved compiling the corpus of Serbian read speech which included isolated sentences realized as IPs with final information focus as well as estimating their rhythmicity on a 5-point Likert scale by 30 native speakers of Serbian.

The corpus of isolated sentences consisted of 20 original IPs (O-IPs), 20 modified IPs (M-IPs) and 20 distractor sentences (DSs), which resulted in the

[8] Also, it should be mentioned that the neuter singular forms of comparatives in the nominative case can be used as adverbs, which were also rather frequent DSWs in the Serbian corpus when used unemphatically.

[9] English and Serbian DSWs and TSWs analyzed in the production experiment are italicized.

total of 60 IPs. All O-IPs were SVO clauses which consisted of 3 initially-stressed words realized as initially-stressed metrical feet. As in the production experiment, the words were disyllabic (DSWs) and trisyllabic (TSWs) with short falling accent, CV syllabic structure and a voiced consonant preceding V1 and V2. The set of DSWs included rather frequent common nouns such as *mama* (mother), *baba* (grandmother) and *deda* (grandfather), feminine personal names such as *Nina* and *Dunja* as well as the 3rd person singular forms of commonly used transitive verbs such as *voleti* (to love) and *voziti* (to drive sb somewhere/to give sb a ride) in the present tense (e.g., *voli*, *vozi*, etc.). The set of TSWs included the euphemistic/diminutive forms of the above listed nominals, some of them being idiomatized, which contain the suffix *-ica* (e.g., *mamica*, *babica*, *dedica*, *Dunjica*, *Ninica*, etc.). Regarding the number of syllables in a foot, out of 20 O-IPs, 10 IPs had uniform feet (O-U-IPs), i.e., they consisted of 3 DSWs, whereas 10 IPs contained non-uniform feet (O-NU-IPs), i.e., they consisted of 1 TSW followed by 2 DSWs, as illustrated in (3)[10]. The set of 20 O-IPs and 20 DSs were recorded as they were read by a Professor at the Faculty of Philosophy in Novi Sad, whose expertise includes Serbian accents, intonation and diction. Prior to the recording, the speaker was asked to control the production of O-IPs for AR.[11] As in the production study, the sampling rate was 44.1 kHz.

(3) a. Dunja voli mamu.
Dunja.Nom love.Pres.3sg mother.Acc
'Dunja loves her mother.'
b. Dunjica vozi dedu.
Dunjica.Nom give-a-ride.Pres.3sg grandfather.Acc
'Dunjica is giving her grandfather a ride.'

The set of 20 M-IPs was created by carefully manipulating F0 in 20 O-IPs with respect to the measurements of F0 in English words, which were previously obtained in the production experiment.[12] The pitch prominence of Serbian initially-stressed DSWs and TSWs was changed according to the pitch prominence of the corresponding English words by considering the parameters D ($F0max_{V1} - F0max_{V2}$) and D ($F0mean_{V1} - F0mean_{V2}$). The position of each pitch point of V1 in the originally produced Serbian DSWs and TSWs was changed manually, ensuring that the shape of the F0 contour observed in the realizations of short falling accent of the original words is preserved to the greatest extent possible. The F0 manipulation in O-IPs and resynthesis of 20 M-IPs was done in *Praat* (version 6.0.29, Boersma – Weenink 2018).

[10] Other examples of O-U-IPs and O-NU-IPs used in the perception task followed the same pattern.

[11] The mean AR was 5.26 syll/sec and it was almost identical in all O-IPs. Also, according to the Mann-Whitney U test, the difference in AR between O-U-IPs and O-NU-IPs was not statistically significant ($U(n1=n2=10)=44.5$, $Z=-0.417$, $p=0.677$), which eliminated the influence of AR on the perception of rhythmicity of IPs consisting of uniform and non-uniform feet.

[12] Although it was experimentally shown that the short falling accent has different realizations in continuously read speech and isolated sentences, the differences primarily concern the tonal slope on the stressed vowel rather than the entire word (Sredojević 2017: 58–60, 84–86, 99–101, 122–126). Thus, the manipulation of F0 in the words in isolated IPs according to the measurements of continuously read speech is not expected to influence the relevance of the results of this research.

The main difficulty in creating M-IPs was the realization of F0 in English and Serbian prosodic words in IP-final position. While English marks final information focus by considerable F0 prominence (Lambrecht 2000; Godjevac 2000: 13–44; Marković 2011b), Serbian makes use of preboundary lengthening, thus relying on durational cues only (Jakovljević – Marković 2020; Jakovljević 2021). As a result, the F0 contour of English IPs with final information focus is characterized by a considerable increase in F0 on a final focalized word, whereas Serbian IPs exhibit a rather smooth declining F0 contour which does not display F0 jumps. In order to avoid the influence of English intonation and focus marking strategies on the perception of rhythm of Serbian M-IPs, the values of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) of English DSWs and TSWs in IP non-final position were applied to the corresponding Serbian words in both IP non-final and IP-final position. Accordingly, a smooth and rather consistent declination was preserved in all resynthesized Serbian M-IPs. This was also confirmed by the auditory impression of resynthesized M-IPs which were described as having either flat or slightly falling intonation without unnaturally high or low prominence of the words in IP-final position.[13]

When the corpus of 60 IPs was compiled, a random sample of 30 second year students of English at the Faculty of Philosophy in Novi Sad were asked to listen carefully to each IP and estimate its rhythmicity on a 5-point Likert scale. The notion of rhythmicity or the strength of speech rhythm was defined as a temporally regular beating pattern, which was done for two reasons. First, the students were not familiar with the notion of perceived rhythmicity, but rather isochrony; and second, this approach made it possible to analyze the interaction between temporal cues and F0 prominence in the perception of speech rhythm in Serbian. The IPs were randomized and the students could hear each sentence twice if uncertain about the evaluation. After the evaluation process, the results of the rhythmicity judgments of O-IPs and M-IPs with both uniform and non-uniform feet were statistically analyzed (SPSS 20).

5. The results of the production experiment. In this section we present the results of the production experiment. We begin by giving the overview of the results regarding F0 prominence of English initially-stressed DSWs and TSWs containing short vowels, realized in IP non-final and IP-final position. The following subsection focuses on F0 prominence of the corresponding Serbian words, i.e., the initially-stressed DSWs and TSWs with short falling accent. In the last part we compare the measurements of F0 prominence of English and Serbian words.

5.1. English. The results of the production experiment show that F0 prominence of the initially-stressed DSWs (Table 2) and TSWs (Table 3) is greater in IP-final than in IP non-final position. The difference between the mean values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) measured in IP non-final and IP-final position is statistically significant for both DSWs and TSWs ($p<0.001$). Although the results point to somewhat higher means of

[13] The evaluators of M-IPs were 4 high school English teachers (Novi Sad, Serbia), 3 high school Serbian teachers (Novi Sad, Serbia), and 3 students at the Faculty of Music (Belgrade, Serbia). All of them are Serbian native speakers.

D ($F0max_{V1}$ – $F0max_{V2}$) compared to D ($F0mean_{V1}$ – $F0mean_{V2}$), the difference between the means is not statistically significant in any realization of the analyzed words, i.e., IP non-finally in DSWs (t(198)=1.023, p=0.306) and TSWs (t(198)=0.755, p=0.451) as well as IP-finally in DSWs (t(198)=0.586, p=0.559) and TSWs (t(198)=0.686, p=0.493).

Table 2. The comparison of D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) (ST) between English initially-stressed DSWs in IP non-final (IP-NF) and IP-final (IP-F) position, produced by 5 native speakers of English.

Parameter	Mean IP-NF	SD IP-NF	Mean IP-F	SD IP-F	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1}$ – $F0max_{V2}$)	2.71	1.102	4.73	1.818	-2.02	-9.511	100	0.000
D ($F0mean_{V1}$ – $F0mean_{V2}$)	2.55	1.117	4.58	1.812	-2.03	-9.538	100	0.000

Table 3. The comparison of D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) (ST) between English initially-stressed TSWs in IP non-final (IP-NF) and IP-final (IP-F) position, produced by 5 native speakers of English.

Parameter	Mean IP-NF	SD IP-NF	Mean IP-F	SD IP-F	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1}$ – $F0max_{V2}$)	2.29	1.307	3.90	1.726	-1.61	-7.436	100	0.000
D ($F0mean_{V1}$ – $F0mean_{V2}$)	2.15	1.315	3.73	1.781	-1.58	-7.137	100	0.000

The results of the measurements also point to a greater F0 prominence of DSWs compared to TSWs realized both IP non-finally (Table 4) and IP-finally (Table 5). The difference in the mean values of D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) between DSWs and TSWs is statistically significant in both IP-positions ($p<0.05$).

Table 4. The comparison of D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) (ST) between English initially-stressed DSWs and TSWs in IP non-final position, produced by 5 native speakers of English.

Parameter	Mean DSW	SD DSW	Mean TSW	SD TSW	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1}$ – $F0max_{V2}$)	2.71	1.102	2.29	1.307	0.42	2.451	100	0.012
D ($F0mean_{V1}$ – $F0mean_{V2}$)	2.55	1.117	2.15	1.315	0.40	2.316	100	0.022

Table 5. The comparison of D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) (ST) between English initially-stressed DSWs and TSWs in IP-final position, produced by 5 native speakers of English.

Parameter	Mean DSW	SD DSW	Mean TSW	SD TSW	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1}$ – $F0max_{V2}$)	4.73	1.818	3.90	1.726	0.83	3.313	100	0.002
D ($F0mean_{V1}$ – $F0mean_{V2}$)	4.58	1.812	3.73	1.781	0.85	3.348	100	0.002

5.2. Serbian. The analysis of F0 prominence in Serbian words points to rather similar tendencies as in the corresponding English words. Thus, the mean values of the parameters D ($F0max_{V1}$ – $F0max_{V2}$) and D ($F0mean_{V1}$ – $F0mean_{V2}$) are higher in IP-final than in IP non-final realizations of Serbian DSWs (Table 6)

and TSWs (Table 7). The difference between the means of the parameters measured IP non-finally and IP-finally is statistically significant for both DSWs ($p<0.001$) and TSWs ($p<0.05$). As in English, the mean values of D ($F0max_{V1} – F0max_{V2}$) are higher than the mean values of D ($F0mean_{V1} – F0mean_{V2}$), but the difference between the means of the parameters in question is not statistically significant in any realization of the analyzed words, i.e., in IP non-final DSWs ($t(198)=0.336$, $p=0.739$) and TSWs ($t(198)=0.341$, $p=0.734$) as well as in IP-final DSWs ($t(198)=0.439$, $p=0.661$) and TSWs ($t(198)=0.349$, $p=0.727$).

Table 6. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between Serbian initially-stressed DSWs in IP non-final (IP-NF) and IP-final (IP-F) position, produced by 5 native speakers of Serbian.

Parameter	Mean IP-NF	SD IP-NF	Mean IP-F	SD IP-F	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1} – F0max_{V2}$)	1.37	0.831	2.01	1.122	-0.64	-4.586	100	0.000
D ($F0mean_{V1} – F0mean_{V2}$)	1.33	0.845	1.94	1.141	-0.61	-4.298	100	0.000

Table 7. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between Serbian initially-stressed TSWs in IP non-final (IP-NF) and IP-final (IP-F) position, produced by 5 native speakers of Serbian.

Parameter	Mean IP-NF	SD IP-NF	Mean IP-F	SD IP-F	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1} – F0max_{V2}$)	0.82	0.626	1.15	0.809	-0.33	-3.226	100	0.002
D ($F0mean_{V1} – F0mean_{V2}$)	0.79	0.621	1.11	0.815	-0.32	-3.123	100	0.003

English and Serbian also exhibit rather similar tendencies regarding the F0 prominence in DSWs and TSWs. The mean values of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) are higher in Serbian DSWs than in TSWs both IP non-finally (Table 8) and IP-finally (Table 9). The difference in the means of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) between DSWs and TSWs is statistically significant in both IP-positions ($p<0.001$).

Table 8. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between Serbian initially-stressed DSWs and TSWs in IP non-final position, produced by 5 native speakers of Serbian.

Parameter	Mean DSW	SD DSW	Mean TSW	SD TSW	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1} – F0max_{V2}$)	1.37	0.831	0.82	0.626	0.55	5.286	100	0.000
D ($F0mean_{V1} – F0mean_{V2}$)	1.33	0.845	0.79	0.621	0.54	5.149	100	0.000

Table 9. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between Serbian initially-stressed DSWs and TSWs in IP-final position, produced by 5 native speakers of Serbian.

Parameter	Mean DSW	SD DSW	Mean TSW	SD TSW	Mean Diff.	t	n1=n2	p (2-tailed)
D ($F0max_{V1} – F0max_{V2}$)	2.01	1.122	1.15	0.809	0.86	6.217	100	0.000
D ($F0mean_{V1} – F0mean_{V2}$)	1.94	1.141	1.11	0.815	0.83	5.919	100	0.000

5.3. The comparison of English and Serbian. The comparison of F0 measurements in English and Serbian points to a few important common tendencies. However, language-specific tendencies regarding F0 prominence in English and Serbian initially-stressed words realized as metrical feet were also observed.

The main similarities between English and Serbian concern the prominence of F0 with respect to IP-position and the number of syllables in a word (foot). The mean values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) in both DSWs and TSWs are higher in IP-final than in IP non-final position (cf. Table 10 and Table 11). Furthermore, the parameters in question have higher mean values in DSWs than in TSWs in both IP-positions. The measurements also show that the difference between the means of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) is not statistically significant in any realization of the analyzed words in both English and Serbian.

On the other hand, the main difference between English and Serbian relates to the degree of F0 prominence in all realizations of the analyzed words (feet). The mean values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) are higher in English than in Serbian DSWs and TSWs, especially in IP-final position (cf. Table 10 and Table 11). The difference between the mean values of the given parameters measured in English and Serbian words is statistically significant ($p<0.001$) both IP non-finally (Table 10) and IP-finally (Table 11).

Table 10. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between English and Serbian initially-stressed DSWs and TSWs in IP non-final position.

Parameter	Mean English	SD English	Mean Serbian	SD Serbian	Mean Diff.	t	n1= n2	p (2-tailed)
DSW D ($F0max_{V1} – F0max_{V2}$)	2.71	1.102	1.37	0.831	1.34	9.712	100	0.000
DSW D ($F0mean_{V1} – F0mean_{V2}$)	2.55	1.117	1.33	0.845	1.22	8.719	100	0.000
TSW D ($F0max_{V1} – F0max_{V2}$)	2.29	1.307	0.82	0.626	1.47	10.144	100	0.000
TSW D ($F0mean_{V1} – F0mean_{V2}$)	2.15	1.315	0.79	0.621	1.36	9.352	100	0.000

Table 11. The comparison of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) between English and Serbian initially-stressed DSWs and TSWs in IP-final position.

Parameter	Mean English	SD English	Mean Serbian	SD Serbian	Mean Diff.	t	n1=n2	p (2-tailed)
DSW D ($F0max_{V1} – F0max_{V2}$)	4.73	1.818	2.01	1.122	2.72	12.733	100	0.000
DSW D ($F0mean_{V1} – F0mean_{V2}$)	4.58	1.812	1.94	1.141	2.64	12.329	100	0.000
TSW D ($F0max_{V1} – F0max_{V2}$)	3.90	1.726	1.15	0.809	2.75	14.427	100	0.000
TSW D ($F0mean_{V1} – F0mean_{V2}$)	3.73	1.781	1.11	0.815	2.62	13.378	100	0.000

6. The results of the perception experiment. The results of the perception experiment indicate that greater F0 prominence in the initially-stressed feet (the initially-stressed DSWs and TSWs with short falling accent) contributes to higher evaluations of rhythmicity. Thus, M-IPs were evaluated as more rhythmical than O-IPs (Table 12) and the difference between the mean rhythmicity judgments is statistically significant ($p<0.001$).

Table 12. The evaluation of rhythmicity of Serbian original IPs (O-IPs) and modified IPs (M-IPs) by 30 native speakers of Serbian.

Mean O-IPs	SD O-IPs	Mean M-IPs	SD M-IPs	Mean Diff.	t	n1=n2	p (2-tailed)
3.07	0.887	3.59	0.844	-0.52	-10.536	600	0.000

Furthermore, the results point to higher rhythmicity judgments of M-IPs with both uniform (Table 13) and non-uniform feet (Table 14). In both cases, the difference between the means is statistically significant ($p<0.001$). However, it can also be observed that F0 prominence has a greater influence on the rhythmicity evaluations of U-IPs compared to NU-IPs, which indicates that the uniformity of metrical feet with respect to the number of syllables has an impact on the evaluation of rhythmicity of Serbian IPs (cf. Table 13 and Table 14).

Table 13. The evaluation of rhythmicity of Serbian original uniform IPs (O-U-IPs) and modified uniform IPs (M-U-IPs) by 30 native speakers of Serbian.

Mean O-U-IPs	SD O-U-IPs	Mean M-U-IPs	SD M-U-IPs	Mean Diff.	t	n1=n2	p (2-tailed)
3.26	0.881	3.93	0.733	-0.67	-10.124	300	0.000

Table 14. The evaluation of rhythmicity of Serbian original non-uniform IPs (O-NU-IPs) and modified non-uniform IPs (M-NU-IPs) by 30 native speakers of Serbian.

Mean O-NU-IPs	SD O-NU-IPs	Mean M-NU-IPs	SD M-NU-IPs	Mean Diff.	t	n1=n2	p (2-tailed)
2.87	0.849	3.25	0.811	-0.38	-5.652	300	0.000

This observation is confirmed by the comparison of the rhythmicity judgments of U-IPs and NU-IPs with both original (Table 15) and modified F0 contour (Table 16). The mean rhythmicity judgments are higher for both O-IPs and M-IPs provided that they contain uniform feet ($p<0.001$), although the means are greater in the latter case (cf. Table 15 and Table 16).

Table 15. The evaluation of rhythmicity of Serbian original uniform IPs (O-U-IPs) and original non-uniform IPs (O-NU-IPs) by 30 native speakers of Serbian.

Mean O-U-IPs	SD O-U-IPs	Mean O-NU-IPs	SD O-NU-IPs	Mean Diff.	t	n1=n2	p (2-tailed)
3.26	0.881	2.87	0.849	0.39	5.565	300	0.000

Table 16. The evaluation of rhythmicity of Serbian modified uniform IPs (M-U-IPs) and modified non-uniform IPs (M-NU-IPs) by 30 native speakers of Serbian.

Mean M-U-IPs	SD M-U-IPs	Mean M-NU-IPs	SD M-NU-IPs	Mean Diff.	t	n1=n2	p (2-tailed)
3.93	0.733	3.25	0.811	0.68	10.773	300	0.000

In order to test whether the uniformity of metrical feet in terms of the number of syllables is reflected in temporal features, we additionally examined the duration of unstressed intervals in 10 O-U-IPs and 10 O-NU-IPs used in the perception experiment. The results showed that the unstressed intervals are longer in O-NU-IPs than in O-U-IPs, as well as that the difference between them is statistically significant ($t(58)=2.542$, $p=0.014$), which indicates that the duration of unstressed intervals in metrical feet increases with the number of unstressed syllables.[14]

Therefore, the results of the perception experiment show that the duration of unstressed intervals is sufficient to differentiate between metrically uniform and non-uniform O-IPs. The unstressed intervals are shorter in the initially-stressed disyllabic feet (DSWs) than in the initially-stressed trisyllabic feet (TSWs) and therefore the O-IPs which contain uniform feet, i.e., 3 disyllabic feet, are perceived as more rhythmical than the O-IPs which contain non-unform feet, i.e., 1 trisyllabic foot followed by 2 disyllabic feet. However, the results also indicate that the increase in F0 prominence in metrical feet contributes to higher rhythmicity judgments of Serbian IPs, i.e., M-IPs with both uniform and non-uniform feet are evaluated as more rhythmical than the corresponding O-IPs although the evaluators were instructed to view rhythmicity as a temporally regular beating pattern, which excludes other components of rhythmic strength.

7. Conclusions. The results of the research verify both pre-specified hypotheses:

Hypothesis 1 (H1): If F0 maximum coincides with a stressed syllable, F0 prominence in a foot will be greater in English than in Serbian;

Hypothesis 2 (H2): If F0 maximum coincides with a stressed syllable, F0 manipulation in Serbian feet in accordance with English production measurements will contribute to higher rhythmicity judgments in Serbian.

Regarding H1, the results of the production experiment show that English initially-stressed feet realized as the initially-stressed DSWs and TSWs with short vowels exhibit greater F0 prominence than the corresponding Serbian feet realized as the initially-stressed DSWs and TSWs with short falling accent. The measurements point to higher mean values of the parameters D ($F0max_{V1} - F0max_{V2}$) and D ($F0mean_{V1} - F0mean_{V2}$) in English than in Serbian DSWs and TSWs both IP non-finally (Figure 1) and IP-finally (Figure 2). The difference in the mean values between English and Serbian words is especially large in IP-final position (cf. Figure 1 and Figure 2) as the final information focus in English IPs is marked by F0 prominence, whereas Serbian makes use of preboundary lengthening.

[14] In order to eliminate the influence of inherent consonant duration, we measured only the duration of vowels in unstressed intervals.

The results of the production task also point to the similarities between English and Serbian. The first similarity concerns the relation between F0 prominence and IP-position as higher mean values of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) were observed in IP-final realizations of both DSWs and TSWs. The second similarity refers to the relation between F0 prominence and the number of syllables in a word considering that the mean values of D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) are higher in DSWs than in TSWs in both IP-positions.

The results of the perception experiment verify H2 as M-IPs with both uniform and non-uniform feet were evaluated as more rhythmical than the corresponding O-IPs (Figure 3). Therefore, we conclude that that greater F0 prominence in metrical feet contributes to higher evaluations of rhythmicity.

Figure 1. The mean values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) in English and Serbian DSWs and TSWs in IP non-final position.

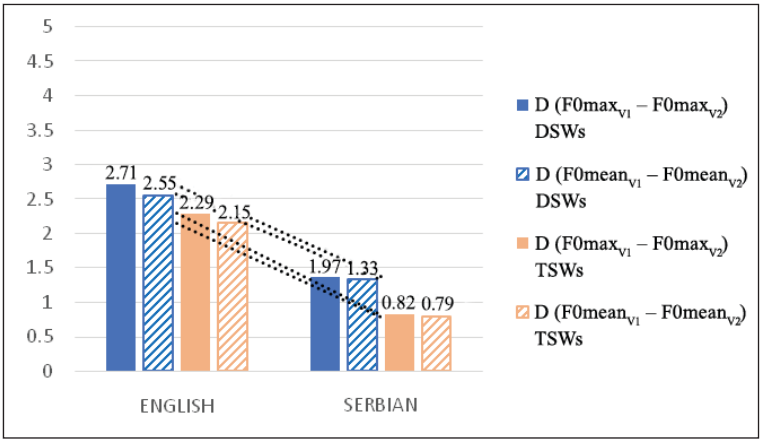
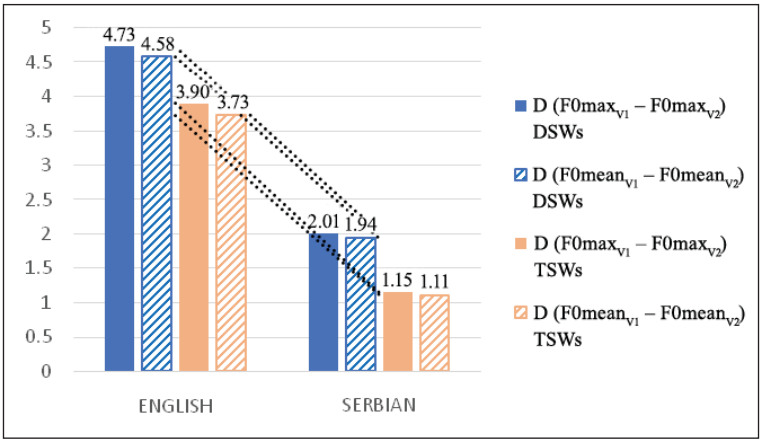
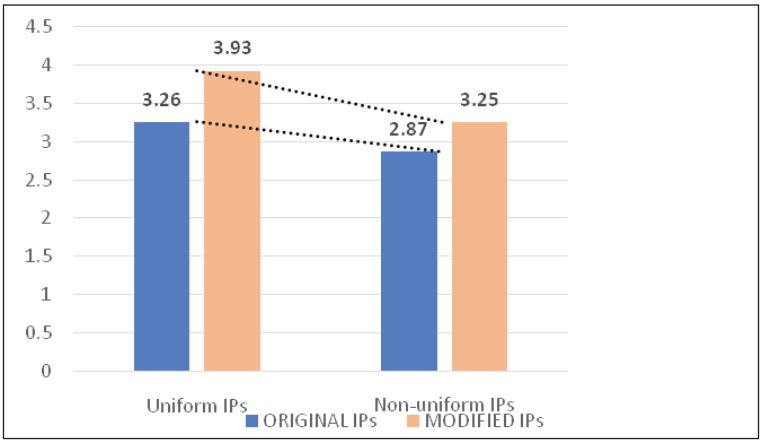


Figure 2. The mean values of the parameters D ($F0max_{V1} – F0max_{V2}$) and D ($F0mean_{V1} – F0mean_{V2}$) (ST) in English and Serbian DSWs and TSWs in IP-final position.



However, the results also indicate that the duration of unstressed intervals is sufficient to differentiate between metrically uniform and metrically non-uniform IPs. The duration of unstressed intervals increases with the number of unstressed syllables and therefore both O-IPs and M-IPs containing uniform feet, i.e., the feet which have the same number of unstressed syllables, are perceived as more rhythmical than the corresponding IPs containing non-unform feet, i.e., the feet which differ in the number of unstressed syllables. Considering that M-IPs with both uniform and non-uniform feet are evaluated as more rhythmical than the corresponding O-IPs, we conclude that F0 prominence additionally contributes the perception of rhythmicity in Serbian.

Figure 3. The rhythmicity judgments of Serbian O-IPs and M-IPs by 30 native speakers of Serbian.



In future research, we are to analyze whether the F0 manipulation of voiced consonants, primarily sonorants, in the onset of the initially-stressed syllables of Serbian words with short falling accent contributes to the perception of rhythmicity of Serbian IPs. Also, we are to examine how the temporal uniformity of metrical feet and the decrease in F0 prominence according to the Serbian production measurements presented in the paper interact in the perception of rhythmicity of English IPs by English native speakers. Finally, the future research involves analyzing the role of F0 prominence in the perception of rhythmicity of Serbian IPs which consist of initially-stressed prosodic words with rising accents.

REFERENCES

Abercrombie, David. *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.
Arvaniti, Amalia. Rhythm, Timing and the Timing of Rhythm. *Phonetica* 66 (2009): 46–63.
Arvaniti, Amalia, Tara Rodriquez. The Role of Rhythm Class, Speaking Rate and F0 in Language Discrimination. *Laboratory Phonology* 4(1) (2013): 7–38.
Barry, William, Bistra Andreeva, Jacques Koreman. Do Rhythm Measures Reflect Perceived Rhythm? *Phonetica* 66 (2009): 78–94.

Bertinetto, Pier Marco. Reflections of the Dichotomy "Stress" vs. "Syllable-timing". *Revue de Phonétique Appliquée* 91(93) (1989): 99–130.
Bloch, Bernard. Studies in Colloquial Japanese IV: Phonemics. *Language* 26 (1950): 86–125.
Boersma, Paul, David Weenink. *Praat: Doing Phonetics by Computer* (Version 6.0.29) [Computer Program] <http://www.praat.org> <25th June 2018>.
Classe, André. *The Rhythm of English Prose*. Oxford: Basil Blackwell, 1939.
Collins, Beverley, Inger Mees. *Practical Phonetics and Phonology*. London: Routledge, 2003.
Cruttenden, Alan. *Gimson's Pronunciation of English*. First published 1962. London: Hodder Education, 2008.
Dauer, Rebecca. Stress-timing and Syllable-timing Reanalyzed. *Journal of Phonetics* 11(1) (1983): 51–62.
Dauer, Rebecca. Phonetic and Phonological Components of Language Rhythm. *Proceedings of the Eleventh International Congress of Phonetic Sciences*. Tillinn: Academy of Science of the Estonian SSR., 1987, 447–450.
Fry, Dennis Butler. Experiments in the Perception of Stress. *Language and Speech* 1 (1958): 126–152.
Fry, Dennis Butler. The Dependence of Stress Judgments on Vowel Formant Structure. Eberhard Zwirner, Wolfgange Bethge (eds.). *Proceedings of the 5th International Congress of Phonetic Sciences*. Basel and New York: Karger, 1965, 306–311.
Godjevac, Svetlana. *Intonation, Word Order and Focus Projection in Serbo-Croatian*. PhD Thesis. Ohio: The Ohio State University, 2000.
Godjevac, Svetlana. Transcribing Serbo-Croatian Intonation. Sun-Ah Jun (ed.). *Prosodic Typology: The Phonology of Intonation and Phrasing*. Oxford: Oxford University Press, 2005, 146–171.
Grabe, Esther, Ee Ling Low. Durational Variability in Speech and the Rhythm Class Hypothesis. Carlos Gussenhoven, Natasha Warner (eds). *Papers in Laboratory Phonology*. Berlin: Mouton de Gruyter, 2002, 515–546.
Han, Mieko. The Feature of Duration in Japanese. *Onsei no Kenkyuu* 10 (1962): 65–80.
Inkelas, Sharon, Draga Zec. Serbo-Croatian Pitch Accent: The Interaction of Tone, Stress, and Intonation. *Language* 64/2 (1988): 227–248.
Jakovljević, Bojana. *Temporalne odlike graničnih segmenata prozodijskih jedinica u engleskom i srpskom jeziku: korpusno istraživanje*. Doktorska disertacija. Novi Sad: Filozofski fakultet, 2021.
Jakovljević, Bojana, Maja Marković. Properties of I-boundary Lengthening of Vowels in English and Serbian. *Godišnjak Filozofskog fakulteta u Novom Sadu* XLV-5 (2020): 95–111.
Kohler, Klaus. The Perception of Prominence Patterns. *Phonetica* 65 (2009): 257–269.
Ladefoged, Peter. *A Course in Phonetics*. First published 1975. New York: Harcourt Brace Jovanovich, 1982.
Ladefoged, Peter. *Phonetic Data Analysis: An Introduction to Fieldwork and Instrumental Techniques*. Oxford: Blackwell Publishing, 2003.
Lambrecht, Knud. When Subjects Behave Like Objects: An Analysis of the Merging of S and O in Sentence-Focus Constructions Across Languages. *Studies in Language* 24:3 (2000): 611–682.
Lea, Wayne. *Prosodic Aids to Speech Recognition IV. A General Strategy for Prosodically-Guided Speech Understanding*. Univac Report PX10791. St Paul, Minnesota: Sperry Univac, 1974.
Lehiste, Ilse. Isochrony Reconsidered. *Journal of Phonetics* 5 (1977): 253–263.
Lehiste, Ilse. The Perception of Duration within Sequences of Four Intervals. *Journal of Phonetics* 7 (1979): 313–316.
Maddieson, Ian. Phonetic Universals. *UCLA Working Papers in Phonetics* 92 (1996): 160–178.
Marković, Maja. Prozodijski transfer u učenju stranog jezika: temporalne karakteristike. *Зборник Матице српске за филологију и лингвистику* LIV/2 (2011a): 135 –146.
Marković, Maja. Acquiring second language prosody: fundamental frequency. Ivana Đurić Paunović, Maja Marković (eds.). *English Language and Anglophone literatures Today (ELALT)*. Novi Sad: Filozofski fakultet, 2011b, 238–249.
Marković, Maja. Kvalitet posttoničnih vokala u govoru Novog Sada. Jasmina Dražić, Isidora Bjelaković, Dejan Sredojević (ur.). *Teme jezikoslovne u srbistici kroz dijahroniju i sinhroniju: zbornik u čast Ljiljani Subotić*. Novi Sad: Filozofski fakultet, 2016, 261–277.
Marković, Maja, Tanja Milićev. The Effect of Rhythm Unit Length on the Duration of Vowels in Serbian. Eliza Kitis, Nikolaos Lavidas, Nina Topintzi, Anastasios Tsangalidis (eds.). *Selected Papers from the 19th International Symposium of Theoretical and Applied Linguistics (19th ISDAL)*. Thessaloniki: Monochromia, 2011, 305–313.

Marković, Maja, Dejan Sredojević. Spectral Characteristics of Stressed Vowels in Standard Serbian: Quantity and Quality Relations. *Зборник Матице српске за филологију и лингвистику* LXIV/2 (2021): 53–76.
Ohala, John. The Physiology of Stress. Larry Hyman (ed.). *Studies in Stress and Accent (Southern California Occasional Papers in Linguistics 4)*. Los Angeles: University of Southern California, 1977, 145–168.
Ordin, Mikhail, Leona Polyanskaya. Perception of Speech Rhythm in Second Language: The Case of Rhythmically Similar L1 and L2. *Frontiers in Psychology* 6 (2015): 1–15.
Pike, Kenneth Lee. *Intonation of American English*. Ann Arbor: University of Michigan Press, 1945.
Ramus, Franck, Marina Nespor, Jacques Mehler. Correlates of Linguistic Rhythm in the Speech Signal. *Cognition* 73 (1999): 265–292.
Roach, Peter. (1991). *English Phonetics and Phonology: A practical Course*. First published 1983. Cambridge: Cambridge University Press, 1991.
Sovilj-Nikić, Sandra. *Trajanje vokala kao jedan od prozodijskih elemenata u sintezi govora na srpskom jeziku*. Nepublikovana magistraska teza. Fakultet tehničkih nauka, Univerzitet u Novom Sadu, 2007.
Sredojević, Dejan. *Fonetsko-fonološki opis akcenata u standardnom srpskom jeziku: od specifičnog ka opštem*. Novi Sad: Filozofski fakultet, 2017.
Stetson, Raymond Herbert. *Motor Phonetics: A Study of Speech Movements in Action*. Originally published 1928. Amsterdam: North-Holland, 1951.
Subotić, Ljiljana, Dejan Sredojević, Isidora Bjelaković. *Fonetika i fonologija: ortoepska i ortografska norma standardnog srpskog jezika*. Novi Sad: Filozofski fakultet, 2012.
Zec, Draga, Elizabeth Zsiga. Interaction of Stress and Tone in Standard Serbian. *Formal Approaches to Slavic Linguistics* 18 (2010): 535–555.

*

Ивић, Павле, Илсе Лехисте. *О српскохрватским акцентима*. Сремски Карловци – Нови Сад: Издавачка књижарница Зорана Стојановића, 2002.
Јокановић-Михајлов, Јелица. *Акценат и интонација говора на радију и телевизији*. Београд: Друштво за српски језик и књижевност Србије, 2007.
Лехисте, Илсе, Павле Ивић. *Прозодија речи и реченице у српскохрватском језику.* Сремски Карловци – Нови Сад: Издавачка књижарница Зорана Стојановића, 1996.
Петровић, Драгољуб, Снежана Гудурић. *Фонологија српског језика.* Београд – Нови Сад: САНУ – Институт за српски језик – Београдска књига – Матица српска, 2010.
Средојевић, Дејан. *Експериментално-фонетско испитивање прозодијских карактеристика новосадског говора*. Докторска дисертација. Нови Сад: Филозофски факултет, 2011.

Бојана Јаковљевић

ПРОМИНЕНТНОСТ ОСНОВНОГ ТОНА И ГОВОРНИ РИТАМ: ИНИЦИЈАЛНИ ЕКСПЕРИМЕНТИ ЗАСНОВАНИ НА ЕНГЛЕСКОМ И СРПСКОМ ЈЕЗИКУ

Р е з и м е

У раду се испитује природа везе између проминентности основног тона и говорног ритма. Истраживање је спроведено у енглеском и српском језику, будући да се дати језици разликују у реализацији акцента на нивоу прозодијске речи. Истраживање чине два експеримента. Експериментална анализа продукције заснивала се на мерењу проминентности основног тона у енглеским и српским двосложним и тросложним стопама, које су биле реализоване као прозодијске речи са акцентом на првом слогу и које је одликовало силазно тонско кретање. У том смислу, мерене су разлике у максималним и средњим вредностима основног тона између вокала у иницијалном наглашеном слогу и вокала у слогу који непосредно следи. Мерења су вршена у прозодијским речима у нефиналној и финалној позицији интонацијских фраза које је читало пет изворних говорника енглеског, односно пет изворних говорника српског језика. Резултати анализе продукције указали су на већу проминентност основног тона у енглеским него у српским прозодијским речима, односно стопама. Пре експеримента перцепције, основни

тон српских интонацијских фраза модификован је у складу са резултатима мерења у енглеском језику, а затим су интонацијске фразе ресинтетизоване. Експеримент перцепције заснивао се на процени ритмичности српских оригиналних и модификованих интонацијских фраза, како униформних, тако и неуниформних. Униформне интонацијске фразе садржавале су три двосложне прозодијске речи (стопе), док су неуниформне интонацијске фразе садржавале једну тросложну и две двосложне прозодијске речи (стопе). У оба случаја, прозодијске речи је одликавао акценат на првом слогу и силазно тонско кретање. Ритмичност је процењивало 30 изворних говорника српског језика на Ликертовој скали од 5 ступњева. Перцептивни експеримент показао је да већа проминентност основног тона доприноси вишим проценама ритмичности. Резултати експеримента такође указују да је трајање ненаглашених интервала довољан предуслов за перцепцију разлика између метрички униформних и метрички неуниформних интонацијских фраза, те да темпоралне одлике говора неспорно доприносе опажању ритмичности у српском језику. Из свега наведеног проистиче да је говорни ритам изузетно сложен продуктивно-перцептивни феномен на чије одлике у различитим језицима не утиче само темпорална организација говора, већ и чиниоци прозодијске проминентности, што је у овом истраживању показано на примеру основног тона.

Универзитет у Новом Саду
Филозофски факултет
Одсек за англистику
Др Зорана Ђинђића 2, 21000 Нови Сад, Србија
bojana.jakovljevic@ff.uns.ac.rs

(Примљено: 20. фебруара 2025;
прихваћено: 15. маја 2025)