Effects of F0 acoustic parameters on the perception of Serbian lexical pitch accents

Dušan Nikolić*
University of Calgary, Linguistics Division

Abstract

The study examines the discrimination of Serbian lexical pitch accent types by English, Mandarin, and Persian listeners. These groups of speakers were selected because of different word-prosodic systems of each language. For example, Serbian is a lexical pitch accent language, English is a stress-accented language, Mandarin is a tone language, and Persian could likely belong either to stress languages or to lexical pitch accent languages. All the participants were naïve listeners, that is, they had never been exposed to Serbian prior to the task. The participants were required to carry out an AX task on short Serbian sentences in which 1) a target item was not manipulated at all, 2) a target item's F0 alignment was modified, 3) a target item's F0 height was modified, and 4) a target item's F0 alignment and height were modulated in tandem. The results showed that both the F0 alignment and F0 height had an impact on the perception of Serbian lexical pitch accents for Mandarin and English speakers, but not for Persian speakers, which suggested that the perception of lexical pitch accents was contingent on the speakers' L1 prosodic system. *(received: 10 May 2021; accepted: 16 June 2021)*

АНАЛИ
ФИЛОЛОШКОГ ФАКУЛТЕТА

1. F0 acoustic parameters of pitch accents

F0 is often considered to be the most relevant acoustic correlate of pitch accents (in contrast to acoustic correlates of stress, which can be F0, duration, and intensity; Beckman, 2012; Levi, 2005). In particular, previous research has found that the perception of pitch accents is predicated on F0 alignment and F0 height, inter alia (Dilley/Heffner, 2013; Kügler/Gollrad, 2015; Grice, et al., 2017). Although these pitch accent F0 parameters are necessary for signalling *post-lexical* pitch accents, it has remained unclear whether the same F0 parameters would facilitate the perception of *lexical* pitch accents.[1] In this study, I aimed to fill this gap in research by exploring whether F0 alignment and F0 height affected the perception of Serbian lexical pitch accents.

In Section 2, I review the most relevant research papers exploring the perception of F0 acoustic parameters of pitch accents. In Section 3, I illustrate the prosodic systems of the languages used in the study. In Section 4, I provide a detailed description of the present study, while in Section 5, I report the results of the study. Discussion of the results and conclusion are given in Sections 6 and 7, respectively.

2. Research review

The previous research on acoustic parameters of pitch accents has mainly focused on post-lexical pitch accents (Smiljanić/Hualde, 2000; Dilley/Heffner, 2013; Kügler/Gollrad, 2015; Grice, et al. 2017). Grice and colleagues (2017) explored how F0 parameters, peak alignment, target height, and the tonal onglide, were used to signal different focus types – broad, narrow, and contrastive focus. In their study, the authors created three nonce words, 'Bieber' [ˈbiːbɐ], 'Bahber' [ˈbaːbɐ], and 'Bohber' [ˈboːbɐ]. The three target items appeared in three different focus types: broad, narrow, and contrastive. Participants were instructed to read a dialogue which consisted of questions and answers, whereby answers contained the target, nonce words. The read material was annotated by two trained transcribers who labelled the material according to German ToBI guidelines (Grice/Benzmüller, 1995). In addition to labelling, the F0 acoustic parameters, F0 alignment, height, and onglide, were analyzed by setting up specific measures for each acoustic parameter. The F0 peak alignment was measured by calculating the time difference between the onset of the stressed vowel and the F0 peak. Positive F0 alignment indicated that the peak occurred within or after the accented syllable, whereas negative values suggested that the peak occurred before the accented syllable. F0 target height was "the difference in semitones between the tonal target on the accented syllable, corresponding to the starred tone, and a following low reference point towards the end of the phrase, corresponding to a low boundary tone" (1995: 93). The authors placed this point towards the boundary tone at 210ms from the target word. The target height was therefore the difference between the F0 peak and this point. Lastly,

---

1 Based on their functions, pitch accents can be lexical or post lexical. For example, speakers of German use post-lexical pitch accents to indicate different types of focus, whereas speakers of Swedish use lexical pitch accent to differentiate segmentally identical lexical items (Jun, 2005).

the tonal onglide was the measure which indicated whether the F0 was rising or falling, and it was calculated by measuring the difference in semitones between the starred tone (where labelers placed a star) and the point 30 ms before the start of the syllable. This point, as the authors suggested, provided a consistent F0 point of measurement for a more optimal approximation of the pitch movement. A positive onglide value meant that the tone was rising up to the peak, whereas the negative value indicated a fall. The results of the study showed a lot of variation across speakers as each of the three F0 parameters, alignment, height and onglide, were used differently on broad focus and on narrow and contrastive focus. In general, participants exhibited "a later peak alignment, a greater tonal onglide and a higher target for contrastive focus than for narrow focus, and likewise for narrow focus in relation to broad focus" (1995: 102). Therefore, the authors concluded that the given F0 acoustic parameters were reliable cues to different pitch accent types.

In a series of four experiments, Dilley and Heffner (2013) tested whether the changes in an F0 peak and valley created a perceptual shift from H* to H+L* and from L* to L+H* pitch accents in American English. The authors used four different tasks: AX, AXB, imitation, and the judgement of the stimuli. Results of their study revealed that the shifts in F0 peak and valley alignment significantly influenced the perception of pitch accent categories. The authors found that "some timing differences have little or no impact on phonological representations for intonation, while other timing differences of comparable magnitude have a quite significant impact on representations" (2013: 49). More importantly, though, the experiments revealed that the adjustment in the alignment of the F0 peak could influence the perception of the distinction between the H*+L and H*.

Kügler and Gollrad (2015) investigated phonetic realizations of the nuclear rise-fall contour in German. The authors explored "how the phonetic realization of the rise-fall contour interacts with contexts that require a contrastive or broad focus interpretation in the answer" (2015: 12). In a series of production and perception experiments, the authors revealed that phonetic parameter height/alignment (they equated these two) of the rise-fall contour changed as the contrastive focus shifted. They found that the parameter height/alignment was sufficient to generate the contrast of the contour as the "speakers realized significantly higher and later F0 peaks in contrastive contexts" (2015: 12). Therefore, alignment/height again proved a very robust F0 parameter of pitch accents.

Smiljanić and Hualde (2000) explored accentual peak alignment in two Serbo-Croatian dialects (Belgrade and Zagreb) in order to develop a better understanding of whether lexical or pragmatic specification constraints could play a role in the accentual peak placement. The authors hypothesized, given the nature of the two dialects, that the accentual peak would be contrastive for Belgrade speakers, but that Zagreb speakers' peak alignment would be influenced mainly by pragmatic conditions such as narrow and broad focus. The authors therefore developed four different sentences in which the target lexical item was always the initial word because the pitch movement was maximized at the beginning of a sentence.

Four sentences were produced in both narrow and broad focus context, whereby whole sentences were produced in broad focus, while only the target word was produced under the narrow focus. The results showed that the contrastive nature of the accentual peak for Belgrade speakers was preserved under each focus type, while Zagreb speakers tended to shift peaks towards the middle of the stressed syllable under the narrow context resembling the lexical distinction preserved with Belgrade speakers. The study therefore showed that even for the speakers of the same language peak alignment was quite an important property in signalling lexical or pragmatic differences.

Levi (2005), in her study, asked whether Turkish belonged to stress or lexical pitch accent languages. The author assumed that Turkish could belong to lexical pitch accent languages (e.g. Serbian, Swedish, Japanese, Basque), and was one of the few who investigated the acoustic parameters of lexical pitch accents. To that end, the author analyzed F0, duration, and intensity of the two sets of minimal pairs of nouns and verbs. Each target item was embedded in a carrier phrase "Ahmet said X", whereby, due to the Turkish word order, the target item was in the middle of the sentence. Although the results showed that all the three cues were quite reliable indicators of lexical pitch accents in Turkish, F0 was found to be the most robust acoustic parameter of Turkish lexical pitch accents, whereas duration and intensity were less prominent. The study showed that lexical pitch accents, albeit signaled primarily by F0, could be indicated by duration and intensity, too.

Developing on the procedure used by Grice and colleagues (2017) (explained below), in the present study, I investigated whether F0 alignment and F0 height, which were taken to be important acoustic parameters of pitch accents influenced the perception of pitch accents. However, the present study deviated from the previous ones in many respects. First, the present study investigated whether F0 alignment and F0 height affected the perception of *lexical* pitch accents, instead of post-lexical pitch accents. Second, the lexical pitch accents are an integral property of Serbian prosodic system, which is why the stimuli that participants heard were in Serbian.[2] Third, although Serbian lexical pitch accent categories depend on the vowel duration, too, in the present study, I maintained the duration constant in order to be able to explore which F0 acoustic parameters influenced the perception of these categories. Lastly, I explored whether naïve speakers, i.e. speakers who had never been exposed to Serbian, could differentiate between lexical pitch accents. The naïve speakers were the native speakers of English, Mandarin, and Persian. I selected these three languages because they were typologically different in their prosodic systems (Hyman, 2006), which further allowed me to make assumptions on the ways participants would perceive lexical pitch accents. In the next section, I briefly describe Serbian, English, Mandarin, and Persian word-prosodic systems.

---

2   The author of the study, who is a native speaker of Serbian, recorded the stimuli.

3. Word-prosodic systems

Standard Grammars of the Serbian language maintain that Serbian uses four lexical accent types (Stanojčić/Popović, 1992). These are divided into *short falling, short rising, long falling,* and *long rising.* Accents in Serbian are contrastive, which means that two words with the same segmental configuration such as *râd* ("work") and *ràd* ("willing to do something"), or *dúga* ("rainbow") and *dȕga* ("long.NOM.FEM") have different meanings. Although the context in which these lexical items are found is fundamental, the accentual distinction cannot be overlooked as it indicates the meaning difference, too. Svetlana Godjevac (2000, 2005) argues that the four accents can be represented phonologically as two bitonal pitch accents: L*+H and H*+L. Godjevac (2005) bases her description on Lehiste and Ivić (1986) who hold that rising and falling accents differ mainly in the F0 patterns of the postaccented syllable. For example, Lehiste and Ivić (1986) found that the F0 of the "syllable following a falling accent was considerably lower than the fundamental frequency of the syllable following a rising accent. The peak F0 value of the postaccentual syllable following a rising accent was comparable to or higher than the peak F0 value of the syllable carrying the rising accent" (Lehiste/Ivić, 1986: 45). Godjevac (2005) retains this difference in stipulating the sequence of two tones: L and H, whereby the starred tone is anchored to the nuclear syllable, while the tone that follows, the trailing tone, is prominent on the second syllable. The distinction between long and short pitch accents is maintained, too, that is, short accents are aligned to a stressed monomoraic nucleus, whereas long pitch accents are associated with a bimoraic nucleus (Zec, 1999).

In the present study, I adhered to Godjevac's (2005) description of pitch accents with a distinction that I did not make a difference between long and short lexical pitch accents in order to be able to ascertain whether *F0 parameters* were affecting the perception rather than duration. Godjevac's (2005) classification of Serbian lexical pitch accents is presented in Table 1 below.

Duration	Pitch movement	
	Falling	Rising
Short	H*+L	L*+H
Long	H*+L	L*+H

Table 1. Serbian lexical pitch accents according to Godjevac (2005)

With respect to English, Beckman and colleagues (2005), and Ladefoged and Johnson (2014) list out five pitch accents in English: H*, L*, L*+H, L+H*, and H+!H*. Apart from L* and H*, high tones can be preceded by low pitch, as in L+H*, whereby the listener hears a "sharply rising pitch. Similarly, L* can be followed by a closely attached high pitch, L*+H, so that the listener hears a scoop upward in pitch after the low pitch at the beginning of the stressed syllable" (Ladefoged/Johnson, 2014: 135). A 'high plus downstepped high', H+!H*, indicates a slight step-down in the pitch height that is relative to the initial high tone. This means that every subsequent pitch accent from the starting high-pitched tone is only slightly lower than the

first tone, but it is not low enough to be an L* (2014: 135). The main distinction between pitch accents in English and Serbian is that English pitch accents are used to express meaning at a post-lexical level, while Serbian pitch accents are inherent to lexical items (Ladd, 1996; Jun, 2005). This means that English speakers do not rely on pitch accent to discriminate between the two lexical items, but they use them to indicate post-lexical information structure types.

Mandarin is a tone language. It includes four tones: high level, high rising, low falling rising, and high falling (Ladefoged/Johnson, 2014). Even though the phonetic reality is usually not as indicative as the annotation of Mandarin tones, the categorization can represent all the syllables in Mandarin (Ladefoged/Johnson, 2014). Mandarin speakers are, therefore, attuned to perceiving tonal contrast in a more categorical manner than the speakers of non-tonal languages (Dupoux, et al., 2008). In addition, it has been shown that Mandarin speakers are more sensitive to F0 directionality in the perception of tones rather than F0 height (Gandour, 1983).

It is still unclear what kind of word prosody Persian has (Eslami/Bijankhan, 2002; Sadat-Tehrani, 2009; Sadeghi, 2011; Abolhasanizadeh, et al., 2012). Persian phono-logical inventory accepts post-lexical pitch accents. These pitch accents are associated with the word-final syllables (Eslami/Bijankhan, 2002). The number of pitch accents is a matter of debate, though. Eslami and Bijankhan (2002) posits four pitch accents: H*, L*, L*+H, and L*+H. Sadat-Tehrani (2009) claims that there is only one pitch accent in Persian – L+H*. This pitch accent "has two morpheme alternants, L+H* in polysyllabic accentual phrases and H* in monosyllabic ones" (Abolhasanizadeh, et al., 2012: 1382). According to Abolhasanizadeh et al. (2012), Persian is different from English in that its prosodic system does not incorporate a contrast between stressed and unstressed syl-lables independently of the pitch accent presence. Sadeghi (2011), on the other hand, claims that stress is the main suprasegmental feature of Persian words, independently of pitch accents, and that duration is its most robust acoustic correlate.

4. Present study

In the present study, I asked whether the F0 alignment and F0 height were robust acoustic cues of Serbian lexical pitch accents, and whether naïve listeners could distinguish between Serbian lexical pitch accent types based on these F0 parameters.

In the following three sections, I elaborate on the methods of the study, I list the most relevant results of the study, and I discuss the findings in line with the research questions.

4.1. Methodology

4.1.1. Stimuli

The stimuli were recorded by a female and a male native speaker of Serbian (mean age = 27.5). The recordings were made in a soundproof booth, at the Phonetics laboratory, at the University of Calgary. All the stimuli were recorded in Praat at a sampling rate 44.1 kHz (16bit) by using a condenser microphone and a high-quality amplifier (24bit, 192kHz).

The stimuli consisted of four sentences produced by Serbian speakers. Each sentence contained one target lexical item bearing a pitch accent. The lexical target items were the following: línija ("line"), màlina ("raspberry"), jâvan ("public"), jälov ("of poor quality"). The target items were placed in a sentence-medial position, that is, in a position where the influence of the boundary tones would be minimal. The target items bore the primary but not emphatic or sentence stress, and each target item consisted of a sonorant so that F0 contours would not be interrupted by an obstruent.[3] The following sentences were used as the stimuli:

1.   a.   Ovo je   **línija**   crvene   boje.
            This be.3rd line._NOM_   red._GEN_   colour._GEN_
            'This is a red line.'

       b.   Ovo je   **màlina**   crvene   boje.
            This be.3rd raspberry._NOM_   red._GEN_   colour._GEN_
            'This is a red raspberry.'

       c.   Ovo je   **jâvan**   rad.
            This be.3rd public._NOM_   work._NOM_
            'This is public work.'

       d.   Ovo je   **jälov**   rad.
            This be.3rd poor quality._NOM_   work._NOM_
            'This is the work of poor quality.'

I manipulated each target item to receive three additional versions. The versions were marked in the following way. (1) original, (2) original + alignment, (3) original + height, (4) original + alignment + height. Original versions (1) were the originally produced sentences, that is, the sentences without any modifications of the lexical target item. The pitch contour of the original version of a target item produced by a female Serbian speaker is illustrated in Figure 1.
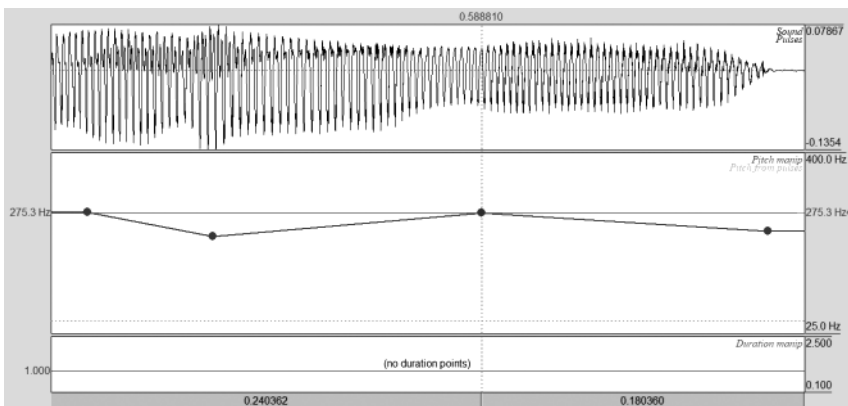

Figure 1. Version (1) – the originally produced pitch contour

---

3   [v] in Serbian is a labio-dental approximant.

In the present study, I developed an original + alignment version (2) in which the peak of the pitch accent was modified in order to align to either the beginning of the syllable or the end of the syllable depending on the pitch accent. L*+H pitch accent peaks were aligned to the left as the peak was found at the end of the syllable, whereas H*+L pitch accents were aligned to the right as the peak was positioned at the beginning of the syllable. Thus, the timing of the peak was changed, while the syllable duration and the pitch height were unaltered (Figure 2).
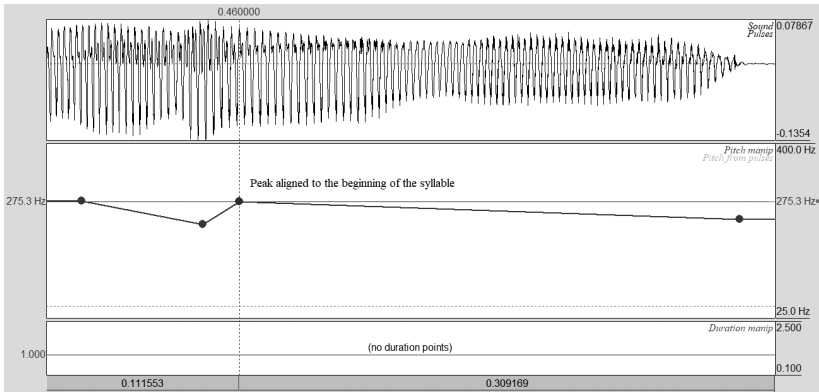

Figure 2. Version (2) – the pitch contour of the modified alignment

The version (3), in the present study, included the height modification. The pitch peak was increased by 50 Hz, as this was thought to be a range that could induce a perceptually significant difference from the original sound (Figure 3).[4] Lastly, the fourth version (4) witnessed the alteration of both the alignment and height. On a duration scale, the peak was aligned either to the left or to the right of its initial position, and the height was increased for 50 Hz (Figure 4).
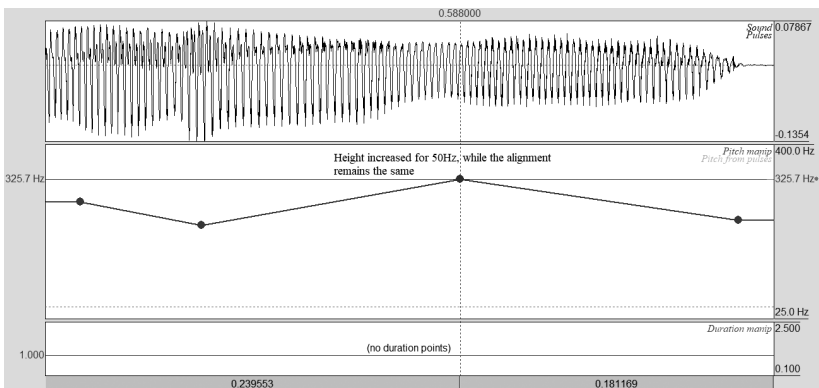

Figure 3. Version (3) – pitch height increased on the lexical pitch peak

---

4   Previous research on F0 peak increase or decrease did not *precisely* indicate the 'value' that should be used so that listeners could detect any difference in the signals (cf. Kohler 2008; Niebuhr/Winkler, 2017). Since the listeners had never been exposed to the language they listened to, and the stimuli were given as full sentences, the cognitive load was rather high. Therefore, to render the alteration of the F0 height perceivable, I decided to take a value of 50 Hz as an increase.
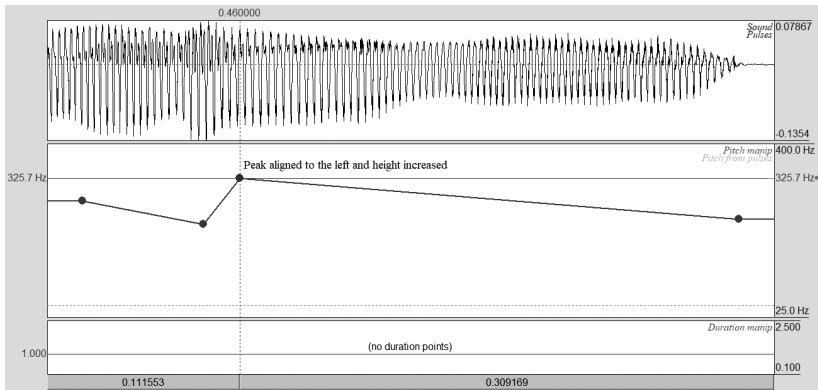
Figure 4. Version (4) – modified peak alignment and height

4.1.2. Participants

There were four groups of listeners: English, Mandarin, Persian, and Serbian speakers (Table 3). There were 5 English listeners, two male, and three female speakers (mean age = 30). There were two Mandarin, two Persian, and two Serbian speakers, one female and one male speaker per each language group. The mean age of the groups was 25, 35, and 24, respectively. All the listeners were the students at the University of Calgary.

4.1.3. Procedure

Participants were required to carry out an AX discrimination task that was administered in PsychoPy3 (Pierce, et. al, 2007). An AX discrimination task was used to determine whether the stimulus A was the same as the stimulus X. I paired each version with another version, and thus I received 10 combinations per target item. There were 80 trials per participant altogether (10 combinations *x* 2 speakers *x* 4 target items), and the stimuli were presented in pairs with a 1000ms inter-stimulus interval (ISI). Participants were told that all the paired recordings were the same segmentally, but that they differed in some suprasegmental aspect. After hearing each pair of sentences, participants judged whether the two sentences were the same or different by pressing the key [A] for "same", and [L] for "different". Once the response was collected, the next pair was played following a 500ms pause.

4.1.4. Data analysis

Although there were 10 combinations for each target item in total, for the purposes of the present study, I counted the responses on those pairs that included the original version and any of the three additional versions. Participants' accuracy scores and reaction times were measured in PsychoPy3, and these were statistically analyzed and processed in RStudio (RStudio Team, 2019) by using the Generalized Estimating Equations (GEE) approach (Liang/Zeger, 1986), and the Signal Detection Theory (SDT) (Macmillan/ Creelman, 2004).

5. Results

Using the Generalized Estimating Equation (GEE), I developed and compared several models wherein *Accuracy* and *Reaction Times* were dependent variables, and *F0 Parameters* was an independent variable. Models revealed that F0 alignment and F0 height, significantly influenced the accuracy scores (Table 2). The pairwise comparisons of estimated marginal means between the parameters confirmed this finding, too, as all the stimuli pairs significantly differed from the original version (Table 3). That is, F0 alignment, F0 height, and the combination of F0 alignment and height were significant predictors of accuracy.

F0 parameters	B	SE	z-score	p-value
Alignment	- 2.656	0.51	15.8	< 0.01
Height	- 1.946	0.49	27.2	< 0.01
Alignment+Height	- 1.236	0.54	5.25	0.022

Table 2. GEE model summary on *Accuracy* as the dependent variable and F0 parameters as the independent variables

Stimuli pairs	odds ratio	SE	z-score	p-value
Original - Alignment	14.24	7.25	5.22	< 0.01
Original - Height	7.00	3.43	3.97	< 0.01
Original - Alignment+Height	3.44	1.86	2.29	0.022

Table 3. Bonferroni pairwise comparisons of estimated marginal means

Different results were obtained on reaction times. While F0 alignment was not a reliable predictor of reaction times, F0 height was (Table 4). The estimated marginal means showed no difference between the stimuli pairs.

F0 parameters	B	SE	z-score	p-value
Alignment	- 0.051	0.147	0.12	0.732
Height	- 0.246	0.115	4.58	0.032
Alignment+Height	- 0.175	0.117	2.26	0.133

Table 4. GEE model summary on *Reaction Time* as the dependent variable and F0 parameters as the independent variables

The results obtained from the GEE models suggested that both F0 alignment and height were robust acoustic parameters of lexical pitch accents. Because the reaction times were affected by F0 height, this acoustic parameter could be taken as more robust than the alignment. Figure 5 shows that listeners indeed needed less time to respond to the pairs in which the original stimuli were coupled with the stimuli with altered height. Estimated marginal means indicate this, too (Alignment – 1.08, Height – 0.833, Alignment+Height – 0.95).
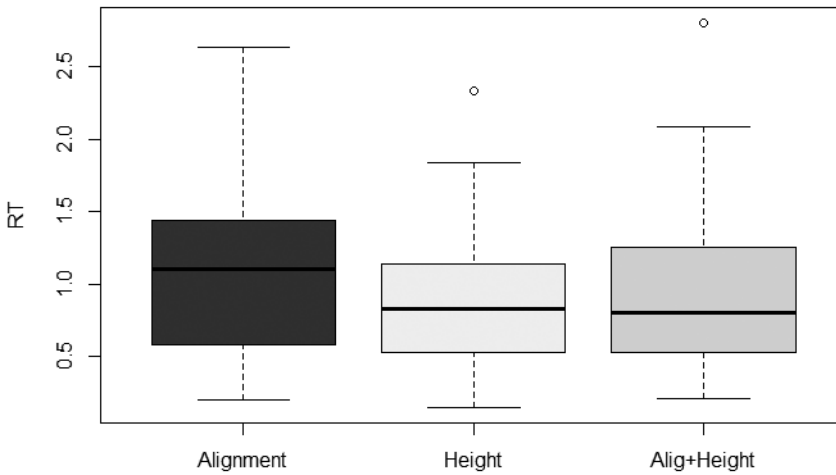
Figure 5. Mean reaction time values per each F0 parameter

The Signal Detection Theory (SDT) was used to investigate differences in performances between groups. Table 5 below presents the raw number of hits, misses, false alarms, and correct rejections per language group. Correct responses on identical pairs (e.g., version (1) and version (1)) were considered correct rejections, while hits were correct responses on different pairs (e.g., version (1) and version (2)). Although I paired up all the versions, I analyzed only those that were coupled with the original stimuli, that is, with the version (1), as these were of greatest interest to the study.

Listeners	Stimulus pair	Responses		Total
		Different (yes)	Same (no)	
English	Different	58 (Hit)	62 (Miss)	120
	Same	4 (False Alarm)	36 (Correct Rejection)	40
Mandarin	Different	21	27	48
	Same	1	15	16
Persian	Different	21	27	48
	Same	6	10	16
Serbian	Different	32	16	48
	Same	1	15	16

Table 5. The breakdown of the total number of responses received on hits, misses, false alarms, and correct rejections per each language group

D-prime values, which are indices of listeners' sensitivity to the signal, are given in Table 6. Table 6 also lists beta, A-prime, B"D, and c scores. The beta value indicates listeners' bias towards saying "yes" or "no". The A-prime value is a non-parametric indicator of discriminability. The value closer to 1 means better

discriminability. A B"D value denotes a non-parametric measure of bias, whereby the positive value means listeners' bias to respond "yes", while the negative value signals listeners' bias to say "no". Lastly, c value is yet another index of bias, which stands for the number of standard deviations from the midpoint between the two distributions.

Listeners	d-prime	beta	A-prime	B"D	c
English	1.19	2.12	0.80	0.47	0.63
Mandarin	1.19	2.46	0.8	0.61	0.75
Persian	0.14	1.03	0.56	0.02	0.23
Serbian	1.77	2.28	0.88	0.58	0.46

Table 6. d-prime, beta, A-prime, B"D, and c scores received from the SDT analysis

Serbian listeners showed the highest sensitivity for discriminating between the stimuli pairs. This was confirmed by both d- and A-prime score. Persian listeners showed the lowest sensitivity, whereas English and Mandarin speakers were identical in this respect. The exact binomial test showed that English listeners performed better than chance on the "same" pairs ($p < 0.01$), while they performed below the chance level on the "different" pairs ($p = 0.78$). The same results were observed for Mandarin listeners ("same" - $p < 0.01$, "different" - $p = 0.47$), while Persian listeners did not perform above the chance level on both "same" and "different" stimuli pairs ($p = 0.45$, $p = 0.47$, respectively). Serbian listeners performed better than chance on all the stimuli pairs ($p = 0.03$, $p < 0.01$).

6. Discussion

In the present study, I explored the perception of Serbian lexical pitch accents by English, Mandarin, and Persian speakers. I asked whether F0 alignment and F0 height were acoustic parameters robust enough for naïve listeners to discriminate between Serbian lexical pitch accents. The results reveal that English and Mandarin speakers can discriminate between Serbian lexical pitch accents reliably well, which can be attributed to the fact that they use these parameters to produce and perceive intonation units in their language. With respect to Mandarin, the main acoustic parameter of Mandarin tones is F0, or more specifically, F0 height and F0 contour (Jongman, et al., 2006). While Gandour's (1983) study attached greater importance to F0 contour than to F0 height, Massaro and colleagues (1985) found that both correlates are equally important. Mandarin speakers were therefore unsurprisingly capable of tuning into the F0 height and F0 alignment acoustic parameters to discriminate between lexical pitch accents. With regard to English, previous research has shown that pitch accents in English are often signaled by F0 alignment and F0 height (Grice, et al., 2017), which is why English listeners could retrieve the acoustic information from the stimuli and discriminate between Serbian lexical pitch accents, too. Due to the nature of the AX task used in the present study, this interpretation precludes the possibility of Mandarin and English

speakers relying on phonological representations of tones or pitch accents to contrast Serbian lexical pitch accents. Instead, the analysis supports the acoustic retrieval of the given parameters.

While Mandarin and English speakers did not have major difficulties on the task, Persian listeners performed below chance. This could be because Persian word-prosodic system is determined by stress rather than lexical pitch accents, and the main acoustic parameter of Persian stress is *duration* (Sadeghi, 2011). Therefore, Persian listeners were less sensitive to the stimuli than English and Mandarin listeners because they did not focus their attention to retrieving the given F0 acoustic parameters; hence, their discrimination level remined low. The finding sheds light on Persian word prosody, as it suggests that Persian belongs to stress-accented languages rather than pitch accent languages.

Lastly, F0 alignment and F0 height significantly influenced the perception of Serbian lexical pitch accents for Mandarin and English speakers, but not for Persian speakers. The fact that Persian speakers exhibited a type of "deafness" towards Serbian lexical pitch accents could suggest that F0 height and F0 alignment are not universally based acoustic properties of word-prosodic units, but that they are robust enough to indicate certain word-prosodic units, if the listeners rely on F0 alignment and F0 heigh in their L1 prosody. Therefore, results, confirm that listeners' L1 prosodic system governs the way in which non-native prosodic contrasts are perceived.

7. Conclusion

In the present study, I explored whether F0 acoustic parameters: F0 alignment and F0 height, were robust enough to influence the perception of lexical pitch accents. Serbian, Mandarin, English, and Persian speakers were required to carry out an AX discrimination on the stimuli recorded in Serbian. The stimuli consisted of short sentences whereby each sentence consisted of the target lexical item that possessed a lexical pitch accent. The target items were manipulated so that F0 alignment and F0 height were modified on each of the items. The modifications were then included in the AX discrimination tasks. Results revealed that F0 alignment and F0 height were reliable cues of lexical pitch accents for Serbian, Mandarin and English, but not for Persian listeners. This finding was attributed to the fact that listeners perceived Serbian lexical pitch accents through their L1 word-prosodic systems, relying on those acoustic parameters that were prominent in their language. Future studies are necessary to fully capture the perception of Serbian lexical pitch accents, as the present study had a low number of participants, and the main acoustic property investigated was F0, rather than F0, duration, the two main components of Serbian lexical pitch accents, as well as the vowel quality, which could be important for the perception of accents in Serbian.

Note: This paper was presented at the Fifth Belgrade International Meeting of English Phoneticians (BIMEP 2020), 20–21 March 2020, Faculty of Philology, University of Belgrade.

References

Abolhasanizadeh, V., Bijankhan, M., Gussenhoven, M. (2012). The Persian pitch accent and its retention after the focus. *Lingua*, *122*(13)*,* 1380–1394.

Beckman, M. E. (2012). *Stress and non-stress accent*. Netherlands Phonetic Archives. Berlin/Boston: Walter de Gruyter.

Beckman, M. E., Hirschberg, J., Shattuck-Hufnagel, S. (2005). In S-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 9–55). Oxford: Oxford University Press.

Dilley, L. C., Heffner, C. (2013). The role of F0 alignment in distinguishing intonation categories: evidence from American English. *Journal of Speech Sciences, 3*(1), 3–67.

Dupoux, E., Peperkamp, S., Sebastian-Galles, N. (2008). Persistent stress 'deafness': The case of French learners of Spanish. *Cognition*, *106*(2), 682–706.

Eslami, M., Bijankhan, M. (2002). Persian intonation system. *Iranian Journal of Linguistics*, *34*, 36–61.

Gandour, J. (1983). Tone perception in Far Eastern languages. *Journal of phonetics*, *11*(2), 49–175.

Godjevac, S. (2005). Transcribing Serbo-Croatian Intonation. In S-A. Jun (Ed.), *Prosodic typology: The phonology of intonation and phrasing* (pp. 146–171). Oxford: Oxford University Press.

Grice, M., Benzmüller, R. (1995). Transcription of German intonation using ToBI tones; the Saarbruecken system. *Phonus, 1,* 35–52.

Grice, M., Ritter, S., Niemann, H., Roettger, T. B. (2017). Integrating the discreteness and continuity of intonational categories. *Journal of Phonetics*, *64,* 90–107.

Hyman, M. L. (2006). Word-prosodic typology. *Phonology, 23*(2), 225–257.

Jongman, A., Yue, W., Moore, C. B., Sereno, J. A. (2006). Perception and production of Mandarin Chinese tones. In P. Li, Y. Shirai, R. Mazuka, M. Nakayama, L. H. Tan, E. Bates, O. J. L. Tzeng (Eds.), *The Handbook of East Asian Psycholinguistics* (pp. 209–217). Cambridge: Cambridge University Press.

Kohler, K. J. (2008). The perception of prominence patterns. *Phonetica*, *65*(4), 257–269.

Kügler, F., Gollrad, A. (2015). Production and perception of contrast: The case of the rise-fall contour in German. *Frontiers in psychology*, *6*, 1254.

Ladd, D. R. (1996). *Intonational Phonology*. Cambridge: Cambridge University Press.

Ladefoged, P., Johnson, K. (2014). *A course in phonetics*. Boston, MA: Cengage Learning.

Lehiste, I., Ivić, P. (1986). *Word and sentence prosody in Serbo-Croatian*. Cambridge, MA: MIT Press.

Levi, S. V. (2005). Acoustic correlates of lexical accent in Turkish. *Journal of the International Phonetic Association*, *35*(1), 73–97.

Liang, K., Zeger, S. L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika*, *73*(1), 13–22.

Macmillan, A. N., Creelman, D. (2004). *Detection theory: A user's guide*. Mahwah: Psychology press.

Massaro, D. W., Cohen, M. M., Tseng, C. (1985). The evaluation and integration of pitch height and pitch contour in lexical tone perception in Mandarin Chinese. *Journal of Chinese Linguistics*, *13*(2), 267–289.

Niebuhr, O., Winkler, J. (2017). The Relative Cueing Power of F0 and Duration in German Prominence Perception. In F. Lacerda, D. House, M. Heldner, J. Gustafson, S. Strömbergsson, M. Włodarczak (Eds.), *Interspeech 2017* (pp. 611–615). Rundle Mall, SA, Australia: Causal Productions.

Pierce, W. J. (2007). PsychoPy—psychophysics software in Python. *Journal of neuroscience methods, 162*(1-2), 8–13.

RStudio Team. (2019). *RStudio: Integrated Development for R.* RStudio, PBC, Boston, MA. http://www.rstudio.com/.

Sadat-Tehrani, N. (2009). The alignment of L+ H* pitch accents in Persian intonation. *Journal of the International Phonetic Association, 39*(2), 205–230.

Sadeghi, V. (2011). Acoustic Correlates of Lexical Stress in Persian. In L. W. Sum, E. Zee (Eds.), *Proceedings of the 17th International Congress of Phonetic Sciences (ICPhS XVII)* (pp. 1738–1741). Hong Kong: City University of Hong Kong.

Smiljanić, R., Hualde, J. I. (2000). Lexical and pragmatic functions of tonal alignment in two Serbo-Croatian dialects. *Chicago Linguistic Society, 36*(1), 469–482.

Stanojčić, Ž., Popović, LJ. (1992). *Gramatika srpskoga jezika* [*The grammar of the Serbian language*]. Belgrade: Zavod za udžbenike i nastavna sredstva.

Zec, D. (1999). Footed tones and tonal feet: rhythmic constituency in a pitch-accent language. *Phonology, 16*(2), 225–264.

Dušan Nikolić

Sažetak

EFEKTI AKUSTIČKIH PARAMETARA FREKVENCIJE OSNOVNOG TONA NA PERCEPCIJU SRPSKIH AKCENATA

Rad istražuje sposobnost percepcijskog razlikovanja srpskih akcenata od strane govornika engleskog, mandarinskog i persijskog jezika. Ove grupe govornika su izabrane zato što svaki od ovih jezika sadrži različit leksičko-prozodijski sistem. Na primer, srpski je jezik koji ima akcente, engleski jezik sadrži naglasak, mandarinski je tonski jezik, dok je persijski jezik sa naglaskom ili akcentima. Ispitanici nikada pre ovog eksperimenta nisu bili izloženi srpskom jeziku. Njihov zadatak je bio da urade test percepcijskog opažanja (tzv. AX zadatak) tako što su slušali kratke rečenice na srpskom u kojima 1) ciljna reč nije bila manipulisana ni na koji način, 2) vremensko ravnanje vrhunca frekvencije osnovnog tona (F0) sa ciljnom reči je izmenjeno, 3) vrhunac osnovnog tona ciljne reči je modifikovan, i 4) vremensko ravnanje i vrhunac osnovnog tona ciljne reči su zajedno modifikovani. Rezultati su pokazali da i vremensko ravnanje i vrhunac osnovne frekvencije utiču na percepciju srpskih akcenata kod govornika mandarinskog i engleskog, dok to nije slučaj sa govornicima persijskog jezika, što govori da se percepcija akcenata kod svih govornika studije odvija nedvosmisleno kroz prozodijski sistem njihovog maternjeg jezika.

Ključne reči:

akcenti, frekvencija osnovnog tona (F0), vremensko ravnanje, vrhunac, percepcija