



Biljana Čubrović

University of Belgrade

Faculty of Philology, English Department

Voice Onset Time in Serbian and Serbian English

Summary

In this paper, the acoustic facts of Voice Onset Time (VOT) are exemplified by looking at two virtually different languages in terms of recognizing VOT as a distinctive phonological parameter. Selected tokens of Serbian and Serbian English are recorded in carrier sentences and analyzed acoustically, as spoken by four proficient Serbian speakers of EFL. The results show that, although Serbian does not recognize VOT as a parameter creating phonological distinctions, advanced non-native speakers of English are capable of learning how to relate the oral and laryngeal gestures in order to produce more native-like pronunciations of English voiceless stops in the phonetic contexts where English /p t k/ are expected to have a long lag. Special attention is drawn to CV sequences whose VOT values deviate in the two languages, as well as to those where VOTs are similar, which can be used to raise the awareness of this phonetic phenomenon in a Serbian EFL learner.

Key words: VOT, voiceless stops, English, Serbian, Serbian English, pronunciation

Čas do začetka zvonečnosti v srbščini in srbski angleščini

Povzetek

Članek obravnava čas do začetka zvonečnosti pri dveh jezikih, kjer ta fonetični pojav nima enake fonološke razločevalne funkcije. Izbrane srbske in angleške besede so v stavkih prebrali štirje srbski govorniki angleščine kot tujega jezika. Posnetke smo akustično analizirali in rezultati so pokazali, da kljub temu da v srbščini čas do začetka zvonečnosti nima fonološke razločevalne vloge, se dobri nerojeni govorniki angleščine lahko naučijo, kako s pravilno artikulacijo čim bolj posnemati izgovorjavo rojenih govorcev angleščine pri nezvonečih zapornikih v okolju, kjer imajo angleški /p t k/ daljši čas nezvonečnosti. Posebna pozornost je namenjena soglasniško-samoglasniškim sklopom, pri katerih so časi do začetka zvonečnosti zelo različni v obravnavanih jezikih, kakor tudi tistim, kjer so si ti časi podobni. Na ta način lahko povečamo zavedanje tega fonetičnega pojava pri srbskih učencih angleščine kot tujega jezika.

Ključne besede: čas do začetka zvonečnosti, nezvoneči zaporniki, angleščina, srbščina, srbska angleščina, izgovorjava

Voice Onset Time in Serbian and Serbian English

1. Introduction

The parameter of Voice Onset Time (VOT), which is defined as the time interval between the stop release and the onset of vocal fold vibration for the following vowel (Lisker and Abramson 1964) has been a matter of debate in phonetic studies since it was first introduced in the 1950's in an attempt to deal with some heated issues in acoustically-based speech synthesis. Although the concept was originally designed for initial plosives, it was later implemented in other contexts, becoming the means of differentiating between voiced and voiceless stops in a large number of languages. A phonetic parameter like VOT was needed because current acoustic measurements at the time were insufficient to account for the absence of vocal fold vibration in typically voiced consonants.

All languages contain a category of stops in their phonemic inventories, which makes a stop a typical, optimal or ideal representative of the consonantal class. Various parameters are implemented when describing stops in the world's languages: phonation type, airstream mechanisms, relative timing of the onset of voicing and relative timing of velic closure. The relative timing of the onset of voicing is of interest in this article. Generally speaking, stops make use of at least three features in this domain: unaspirated, aspirated and pre-aspirated. The first two are significant for this article, as English and Serbian do not employ the class of pre-aspirated stops.

UCLA Phonological Segment Inventory Database (UPSID) presents results of a survey of 317 languages, claiming that the unaspirated voiceless category is found in 91.8% of languages. The unaspirated voiced stops are present in 66.9%, and the aspirated voiceless in 28.7% (Maddieson 1984, 27). The unaspirated voiceless category, as the most widespread one, seems to be most efficient from the aerodynamic and articulatory points of view, at least in word-initial positions. Due to their naturalness, Keating et al. (1983) claim that languages favour voiceless over voiced stops. Unaspirated categories are thus sometimes referred to as plain. Furthermore, statistics show that languages with two stop series are divided into two substantial categories: unaspirated voiceless/voiced contrast is evident in 117/162 languages (72.2%) and unaspirated voiceless/aspirated voiceless or unaspirated voiced/aspirated voiceless in 27 languages (Maddieson 1984). The issue of VOT continuum is therefore critical in a vast number of languages, but it is not the most widespread pattern. Serbian belongs to the former category, having a contrast between unaspirated voiced stops /b d g/, and unaspirated voiceless stops /p t k/. Furthermore, there is a difference in the place of articulation for /t/ in English and Serbian. Serbian /t/ has a dental articulation, whereas the English segment is produced on the upper alveolar ridge. Earlier research shows that there is variation in the effect alveolars have on VOT values, but velars repeatedly exhibit higher VOTs than labial stops. Many authors claim that the VOT descending scale ranges from velars to alveolars to labials in the speech of native English adults (Lisker and Abramson 1967; Klatt 1975; Zue 1976; Weismer 1979; Nearey and Rochet 1994).

The motivation to carry out the experiment with Serbian native speakers was sparked by a large number of papers studying VOT from different perspectives, acoustic, articulatory and perceptual, looking at both bilingual and multilingual language behaviour. Out of a solid number

of articles on the topic, I have chosen Lisker and Abramson's seminal article (1964), in which they examined 11 languages of the world, paying attention to their genetic and phonetic richness in order to create a representative language database. Word-initial prevocalic positions were studied both in isolated words and in connected speech. The results of Lisker and Abramson's study are as follows:

	Average	Range	No. of tokens
/p/	58	20:120	102
/t/	70	30:105	116
/k/	80	50:135	84

Table 1. VOT values for stops in isolated words.

	Average	Range	No. of tokens
/p/	28	10:45	24
/t/	39	15:70	26
/k/	43	30:85	25

Table 2. VOT values for stops in connected speech.

Several striking differences exposed in Tables 1 and 2 need to be commented upon. A significant difference between VOT values in isolated words and in connected speech should be attributed to the tempo of speech. It is a commonplace to say that more careful speech is relatively slow, and thus the temporal dimension is longer.

Lisker and Abramson (1964) launched the idea of differentiating voiced and voiceless stops by means of VOT in their attempt to discover the best measure by which it would be possible to separate the two phoneme categories. The reason for singling aspiration out is that it seems spectrographically unambiguous because it registers as noise. Moreover, it could ultimately be checked by speech synthesis experiments, popular at the time. The VOT continuum offers 3 categories pertaining to the stop voicing contrast: voicing lead (with negative VOT values), short-lag VOT (with zero or low positive VOT values), and long-lag VOT (with high positive VOT values), all measured in milliseconds.

2. Experiment Design

2.1 Method

A list of 27 English and Serbian words, monosyllables or disyllables, was recorded. Wherever possible, minimal or near minimal pairs, were used in order to neutralize the potential differences which could have been created by deviations in the phonetic environments in the English and

Serbian tokens. Nine vowels, both short and long, were analyzed in accented positions. They were invariably preceded by one of the voiceless plosives /p t k/. The selection of 27 phonetic contexts provides a common vocalic denominator typical of English and Serbian. The English vowel qualities under investigation are: /e i: ɪ u: ʊ ɔ: ɒ ʌ ɑ:/. Their Serbian approximations /i u o a/, affected by both long and short pitch accents, with the addition of the short Serbian /e/, are taken into account. The rationale behind the elimination of the long Serbian counterpart of /e/, as in the word *pêta* (Eng. *fifth*) from the recorded corpus is the lack of this vowel quality in English.

Each token was recorded three times in carrier sentences. All tokens were placed in accented positions and informants were instructed to stress them. The two female and two male Serbian speakers are all proficient speakers of English (English language and literature graduates). All four speakers have lived in Belgrade for more than fifteen years now. None of the speakers lived in an English speaking country for more than 8 months. Speakers' mean age was 30.7, ranging from 25-35 years of age.

Recordings were made in Praat, version 5.1.33, at a sampling rate of 22,050 Hz, using a Sennheiser Pc156 noise cancelling microphone. Recordings were analysed in the same software package, with the help of waveforms.

2.2 Results

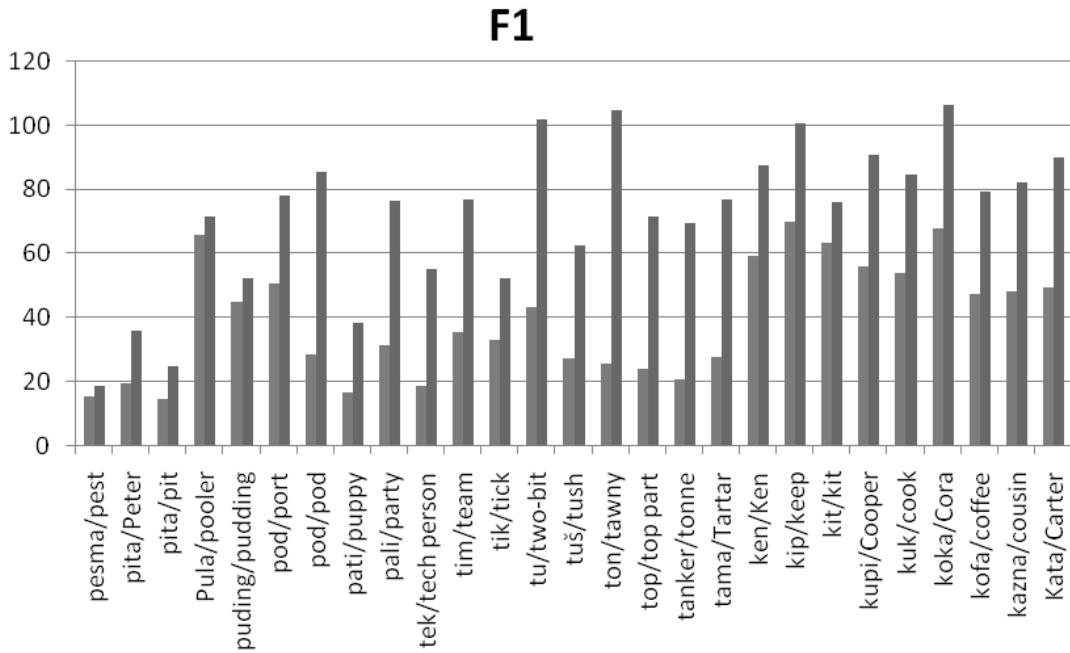
Each speaker's results were analysed separately for Serbian and Serbian English, bearing in mind common phonetic knowledge about how VOT functions in relation to other stop features (place of articulation, vowel type, etc.). For instance, the place of articulation seems to exert influence on VOT values. Velars, for instance, are significantly more aspirated than bilabials.

The following abbreviations are used for the four informants: F1 (female speaker no. 1), F2 (female speaker no. 2), M1 (male speaker no. 1), and M2 (male speaker no. 2). The main hypothesis postulated before the experiment is that VOT values are shorter for Serbian tokens than for Serbian English tokens, due to the fact that Serbian does not recognize aspiration as a distinctive feature of Serbian stops. Ranges of VOT values are given first for each individual speaker, followed by mean values for each CV sequence (presented in graphs underneath).

F1 VOT values range from 11-67 msec for the Serbian tokens containing /p/, 18-47 msec for the Serbian tokens having /t/, and 41-79 msec for /k/. The highest VOT mean value is found for the Serbian sequences /pu:-/, /tu:-/ and /ki:-/, and the lowest mean value is characteristic of /pi-/ , /te-/ and /ko-/. The VOT measurements are given in Graph 1 below for the first female speaker. VOT values for Serbian tokens are given in the first column (msec), and these are followed by the values for Serbian English tokens in column 2.

F1 VOT values range from 18-90 msec for the Serbian English tokens containing /p/, 42-111 msec for the Serbian English tokens having /t/, and 65-108 msec for /k/.

VOT values for Serbian English tokens are consistently higher in F1 speaker, which is clearly perceived in the graph.

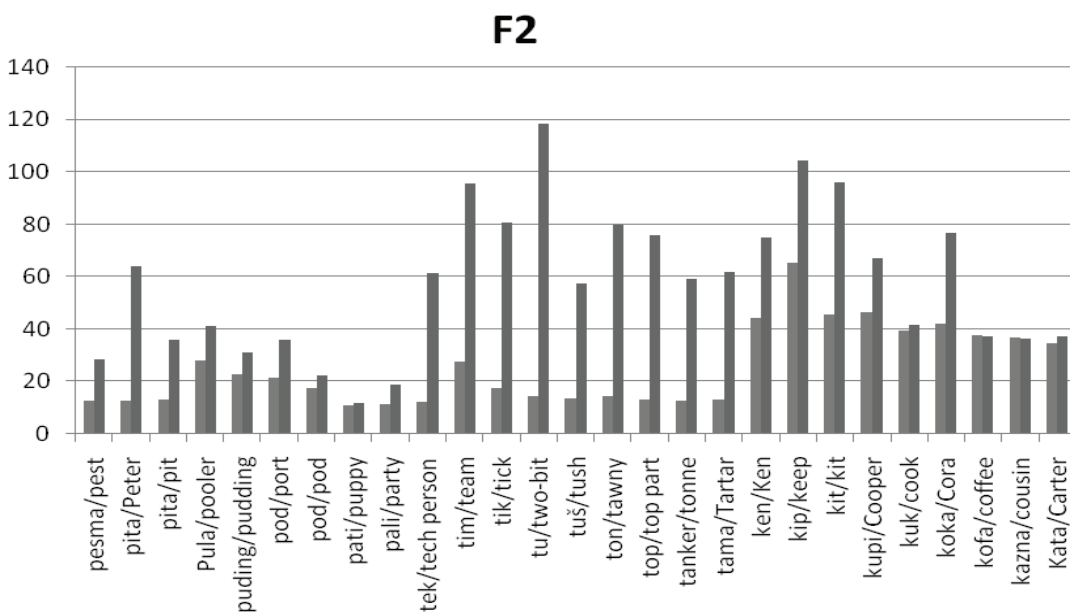


Graph 1. Mean VOT values for F1 speaker.

F2 VOT values range from 9-29 msec for the Serbian tokens containing /p/, 11-32 msec for the Serbian tokens having /t/, and 27-73 msec for /k/. The highest VOT mean value is found for the Serbian sequences /pu:-/, /ti:-/ and /ki:-/, and the lowest mean value is characteristic of /pa-/, /te-/, and /ka:-/.

F2 VOT values range from 10-95 msec for the Serbian English tokens containing /p/, 40-123 msec for the Serbian English tokens having /t/, and 27-153 msec for /k/.

VOT values for Serbian English tokens are higher in F2 speaker's production, which is clearly perceived in the graph.

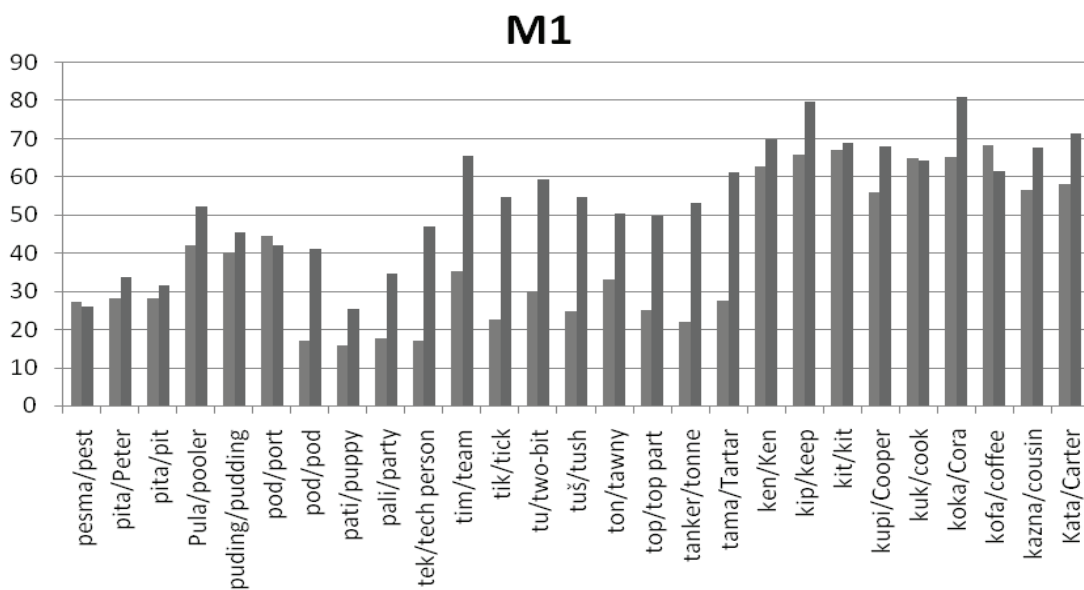


Graph 2. Mean VOT values for F2 speaker.

M1 VOT values range from 13-50 msec for the Serbian tokens containing /p/, 14-38 msec for the Serbian tokens having /t/, and 47-84 msec for /k/. The highest VOT mean value is found for the Serbian sequences /po:-/, /ti:-/, and /ko:-/, and the lowest mean value is characteristic for /pa-/, /te-/, and /ku:-/.

M1 VOT values range from 23-60 msec for the Serbian English tokens containing /p/, 44-67 msec for the Serbian English tokens having /t/, and 54-85 msec for /k/.

VOT values for Serbian English tokens are higher in M1 speaker's production, which is clearly perceived in the graph. The biggest differences are evident in the production of /t/ in Serbian English and Serbian.

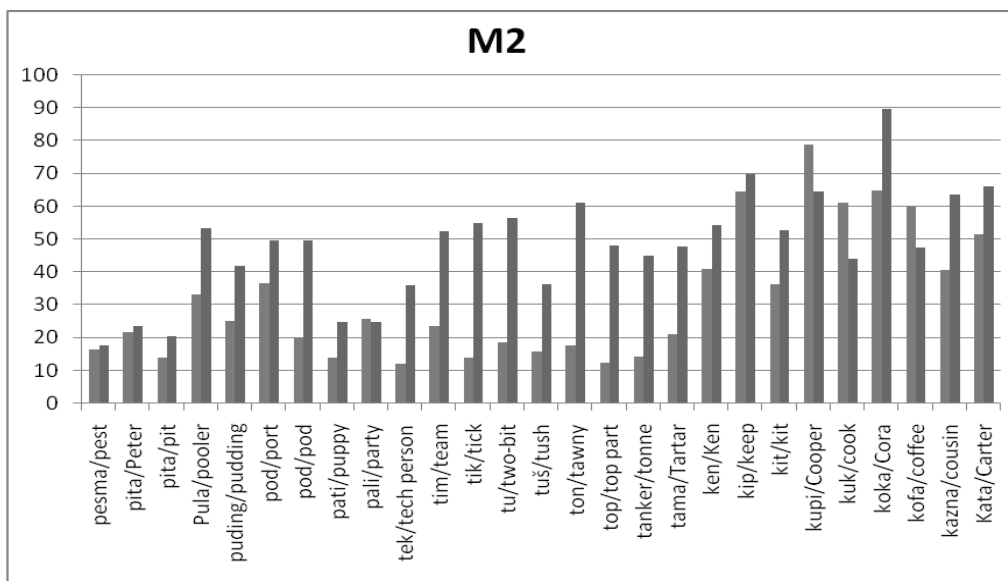


Graph 3. Mean VOT values for M1 speaker.

M2 VOT values range from 10-38 msec for the Serbian tokens containing /p/, 11-20 msec for the Serbian tokens having /t/, and 30-95 msec for /k/. The highest VOT mean value is found for the Serbian sequences /po:-/, /ti:/ and /ku:-/, and the lowest mean value is characteristic for /pi-/ and /pa-/, /te-/, and /ki-/.

M2 VOT values range from 13-56 msec for the Serbian English tokens containing /p/, 26-66 msec for the Serbian English tokens having /t/, and 40-93 msec for /k/.

VOT values for Serbian English tokens are higher in M2 speaker's production, which is clearly perceived in the graph. The biggest differences are evident in the production of /t/ in Serbian English and Serbian.



Graph 4. Mean VOT values for M2 speaker.

A great variation is noticed in the VOT values pertaining to Serbian and Serbian English. The most striking difference lies in the acoustic data for dental and alveolar /t/ in Serbian and Serbian English, respectively, which behave differently in the speakers' production. The average VOTs for both Serbian and Serbian English are given in Tables 3 and 4.

	Average	Range	No. of tokens
/p/	25	11:67	216
/t/	22	12:43	216
/k/	56	35:79	216

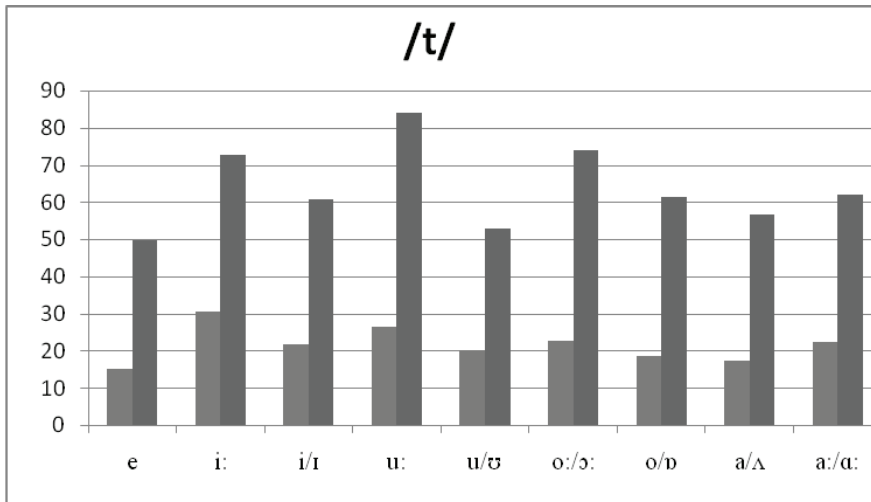
Table 3. VOT values for Serbian stops.

	Average	Range	No. of tokens
/p/	39	10:91	216
/t/	64	26:123	216
/k/	71	27:153	216

Table 4. VOT values for Serbian English stops.

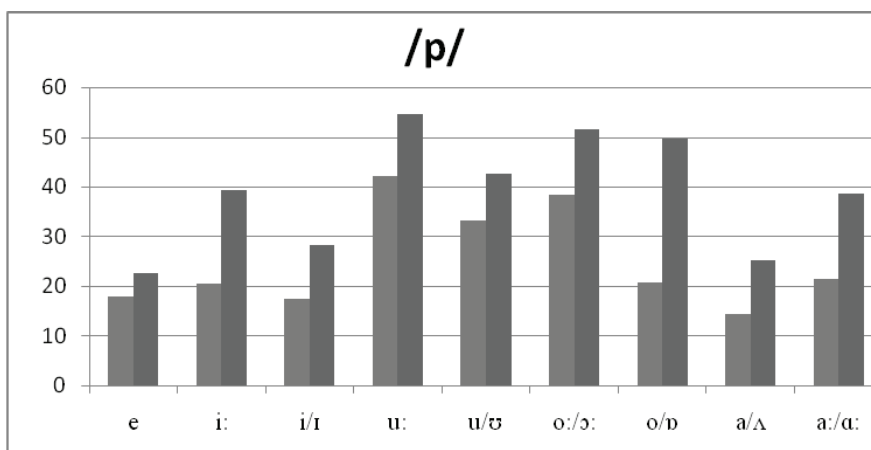
Serbian tokens with dental /t/ exhibit a much shorter VOT value compared to their Serbian English alveolar counterparts. As shown in the data, the dental articulations of /t/ in Serbian exert influence on the ranking of VOTs in an ascending order. Dentals have the lowest VOT values in Serbian, and they are very closely followed by labials. Velars expectedly have the longest voicing lag. All VOT values are positive in Serbian stops.

Graph 5 summarizes the differences in the production of Serbian and Serbian English /t/. VOT values are almost invariably significantly higher for Serbian English than for Serbian (the data in the first column refers to Serbian, whereas column 2 shows the values for Serbian English). The informants, being fluent speakers of English, have learnt how to acquire long-lag VOT values necessary for native-like English pronunciations. However, at lower levels, Serbian EFL learners need to be drilled into pronouncing alveolar /t/ articulations first.



Graph 5. Mean VOT values for Serbian and Serbian English /t/.

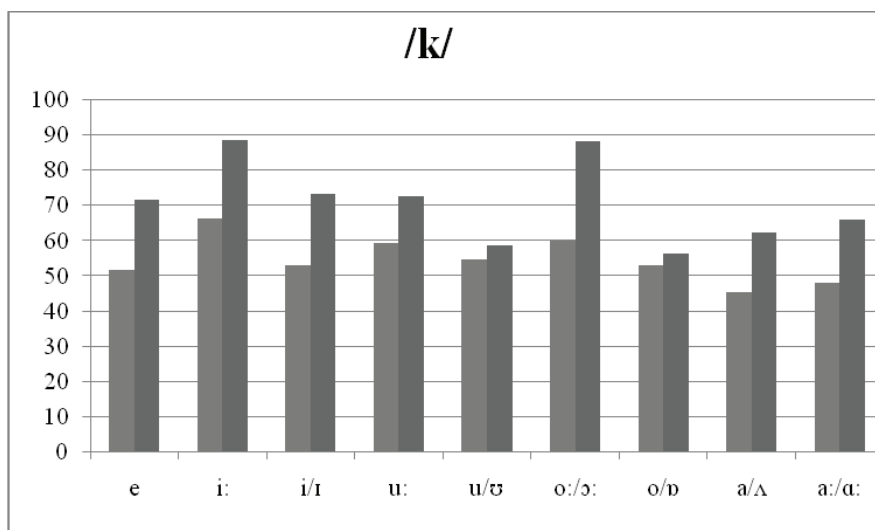
VOT measurements for bilabial stops in Serbian and Serbian English consistently deviate, with the exception of the sequence bilabial + /e/, where VOTs do not differ significantly. The informants have successfully acquired the long-lag VOT in their English pronunciation. According to Graph 6, CV sequences characterized by significant differences in VOT values are bilabial + /i: ɒ ʌ ɑ:/. These CV sequences should be treated separately in a Serbian EFL classroom by designing special pronunciation drills, dwelling on very simple vocabulary items, e.g. *peace, pot, pub, park*, etc.



Graph 6. Mean VOT values for Serbian and Serbian English /p/.

Labials and velars can be useful when learning how to relate laryngeal gestures in English as L2. By their nature, velars are characterized by high VOTs in many languages of the world. CV sequences of velar + short back vowel, according to the experimental data, have very similar VOT

values, and they can be used to raise awareness of the importance of aspiration in English (See Graph 7). A simple aspiration trick with a sheet of paper placed in front of the oral cavity whilst pronouncing a Serbian CV sequence should assist students in noticing how aspiration works even in their own mother tongue.



Graph 7. Mean VOT values for Serbian and Serbian English /k/.

3. Conclusion

Even though Serbian does not recognize aspiration as a distinguishing phonetic parameter, the experimental data shows that it is widely used in Serbian stop articulations. Serbian stop consonants with the longest lag are velars, as expected. However, dental and labial stops have quite similar VOT values, but the former stop category has a slightly longer lag. Such a phonetic state of affairs does not generate a foreign accent in Serbian speakers' English as such, due to the fact that Serbian EFL learners are required to learn how to pronounce English alveolar stops first.

Judging by the production of Serbian English stops as performed by the four participants in the present study, and considering the fact that the experimental conditions are artificial by default (influencing the VOTs to be longer than in connected speech), I claim that Serbian EFL learners can effectively acquire the VOTs necessary for native-like articulation of English stops. Velar stops are the best starting point as they are universally characterized by long-lag VOTs. Bilabials should be tackled in the second phase of the acquisition of English pronunciation. Due to the differences in the place of articulations, English alveolars should be handled last. This study shows that a number of stop+V sequences share similar VOT values in the two languages under investigation. Such sequences, especially if they are characterized by long-lag VOTs, can be productively utilized when teaching pronunciation to Serbian EFL learners. Although aspiration as such has not found its place in Serbian phonetic studies, this experiment shows that its presence in the phonological system of Serbian could undoubtedly be used to raise the phonological awareness of this phenomenon and trigger its usage in L2.

Bibliography

- Keating, P.A., W. Linker, and M. Huffman. 1983. Patterns in allophone distribution for voiced and voiceless stops. *Journal of Phonetics* 11: 277–90.
- Klatt, D.H. 1975. Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of Speech and Hearing Research* 18: 686–706.
- Lisker, L., and A.S. Abramson. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20: 348–422.
- . 1967. Some effects of context on voice onset time in English stops. *Language and Speech* 10: 1–28.
- Maddieson, I. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.
- Nearey, T.M., and B.L. Rochet. 1994. Effects of place of articulation and vowel context on VOT production and perception in French and English stops. *Journal of the International Phonetic Association* 24: 1–19.
- Weismer, G. 1979. Sensitivity of voice onset measures to certain segmental features in speech production. *Journal of Phonetics* 7: 194–204.
- Zue, V.W. 1976. *Acoustic characteristics of stop consonants: A controlled study* (Technical Report 523). Lexington, MA: Lincoln Laboratory, MIT.