DOI: https://doi.org/10.18485/philologia.2023.21.21.3
UDC: 811.111'243
811.111'342

GREAT ASPIRATIONS: EXAMINING VOT OF WORD-INITIAL VOICELESS STOPS IN ENGLISH AND SERBIAN IN SERBIAN EFL SPEAKERS

Nina Đukić[1]
University of Belgrade
Faculty of Philology
Belgrade, Serbia
https://orcid.org/0000-0001-7068-8072

Uloga aspiracije (pozitivno vreme nastupa zvučnosti) značajno se razlikuje u engleskom i srpskom jeziku. Dok pozitivno vreme nastupa zvučnosti ima distinktivnu funkciju u razlikovanju zvučnih i bezvučnih ploziva u engleskom, to nije slučaj u srpskom jeziku. Stoga bi jasan kontrast između dugih ploziva engleskog jezika i kratkih ploziva srpskog jezika mogao biti izazov za srpske govornike engleskog jezika. Ovaj rad ispituje stepen aspiracije kod bezvučnih ploziva u inicijalnoj poziciji u rečima na srpskom i engleskom jeziku kod dve ciljne grupe srpskih govornika engleskog: pet govornika srednjeg nivoa i pet naprednih govornika engleskog jezika. Osnovno pitanje istraživanja jeste u kojoj meri maternji jezik (L1) utiče na izgovor stranog jezika (L2) i na koji način se to ispoljava kod ove dve grupe ispitanika. Uzorci za analizu dobijeni su uz pomoć programa *Praat* (Boersma/Weenink 2022), snimanjem govora ispitanika koji čitaju rečenice sa ciljnim rečima u kojima se bezvučni plozivi nalaze u početnoj poziciji. Za svaki uzorak meri se vreme nastupa zvučnosti, odnosno interval između realizacije ploziva i početka fonacije sledećeg glasa, izraženo u milisekundama (ms). Rezultati pokazuju da se aspiracija naprednijih govornika engleskog podudara sa izvornim govornicima. U ovom ispitivanju zabeleženi su i neočekivani rezultati, konkretno L2 transfer, gde isti napredni govornici plozive na srpskom jeziku izgovaraju sa izraženom aspiracijom, što odstupa od uobičajenih fonetskih karakteristika izvornih govornika srpskog. Rezultati takođe pokazuju da, iako realizacija ploziva sa produženom artikulacijom kod govornika srednjeg nivoa engleskog nije podudarna izgovoru maternjih govornika, ova grupa ispitanika pravi jasniju razliku kada je u pitanju primena vremena nastupa zvučnosti u maternjem i stranom jeziku.

Ključne reči: aspiracija, vreme nastupa zvučnosti, L1 transfer, L2 transfer, srpski govornici engleskog.

1 Kontakt podaci (E-mail): nina.djukic@fil.bg.ac.rs

1. INTRODUCTION

Though learning a foreign language is primarily associated with acquiring the correct sentence structure or gradually expanding one's vocabulary, the process, in fact, begins with acquiring L2 phonetics and phonology. In other words, together with acquiring new lexis, learners simultaneously aim to produce and understand L2 sounds on a level that enables comprehensible communication. Specifically, phonetic features of a language include sounds, together with their contextual allophones or specific features of connected speech (Čubrović 2009).

However, the process of L2 sound acquisition entails the existence of an L1 set of phonemes that has been previously acquired through one's mother tongue. The relationship between these two sets has already been explored and linguistic evidence confirms the effect L1 has on L2 pronunciation and various phonetic features (Flege 1992; Flege *et al.* 1995). Such an effect is, in practice, reflected in foreign-accented speech. This phenomenon, thus, implies that the L2 speaker has yet to acquire and apply all phonetic features that are characteristic of that foreign language. The necessity of emulating a native-like accent, however, stands as a separate issue, since, as Zsiga explains, rules of pronunciation ought to be understood as "descriptive, rather than prescriptive" (2020: 48).

In terms of English as a foreign language, its pronunciation challenges vary depending on the learner's L1. That is to say, while some speakers struggle with a specific group of sounds e.g. /b/ and /v/ or dental fricatives /θ/ and /ð/ (Flege/Eefting 1987) others find the allophones of English or its intonation patterns rather challenging.

For Serbian speakers of English, one of the possible challenges could be one of the less evident features – aspiration. This feature is an integral part of speech for native English speakers, as it serves to distinguish between voiced and voiceless stops, which will be explained further in the following section. On the other hand, no records have been made regarding aspiration in the Serbian language, or at least not about its distinctive function as is the case in English. For this reason, the aim of this paper is to, firstly, assess to what degree the participants are able to use aspiration when speaking English and, secondly, examine the presence of aspiration in Serbian. Due to its alleged lack of presence in Serbian, it can be assumed that, as an example of L1 transfer, Serbian EFL speakers will have some issues acquiring and applying this feature fully.

2. THEORETICAL BACKGROUND

Aspiration is directly connected to the stop sounds of English and, specifically, to their manner of articulation. Namely, all English stops, /b d g p t k/, are produced as a combination of complex movements within the vocal tract, the key point being the highest degree of stricture – closure – where the articulators are in firm contact (Čubrović 2009: 35). This closure can form in several places within the tract, which is why the stops can be categorized according to their place of articulation. Regardless of their position, all stops go through the same three stages of articulation: approach, hold or compression and release stage (Čubrović 2009: 36). The final stage includes a sudden release of air with an explosive noise. Energy levels during the release, however, vary depending on their voicing.

Depending on the vibrations of the vocal folds, typically, these stops are grouped into voiced /b d g/ and voiceless /p t k/. Halle *et al.* introduce alternative terms for dividing stops by using the terms tense and lax stops (1957: 107), otherwise known as fortis and lenis stops (Čubrović 2009: 37). The reason why this kind of labelling is often deemed more precise is related to the exact manner of articulation of these two groups. With tense stops, /p t k/, there is significantly more pressure that builds up against the closure, resulting in a more powerful release (Halle *et al.* 1957: 107). On the other hand, lax stops, /b d g/, also include bursts, i.e., have their release stage. However, the key difference is the presence of aspiration, which is significantly more pronounced with tense and not lax stops. Though aspiration can be heard in speech, what remains is the question of recording and perceiving this feature on a waveform, i.e. on audio recordings. The aforementioned difference between aspirated and unaspirated consonants can, in fact, be found owing to VOT or voice onset time, which refers to the amount of time that elapses between the release of the consonant and the onset of voicing of the following sound (Zsiga 2020: 131). Clear segmentation might be a challenge, however, what remains as a clear rule is that we generally differentiate positive/long-lag, negative/short-lag and zero VOT where positive refers to a certain delay between the release of the plosive and the onset of vocal fold vibration. Negative VOT entails vocal fold vibration prior to the plosive release, while zero VOT means that the plosive release and vocal fold vibrations approximately coincide (Zsiga 2020: 189). In this way, by measuring VOT on a particular waveform, linguists have been able to explore the significance of aspiration and compare it across various languages and among different speakers (Halle *et al.* 1957; Cho/Ladefoged 1999; Kim 2011; Shimizu 2011).

Cho and Ladefoged examined 18 different endangered languages so as to discover VOT universals across different languages. These potential universals would effectively enable a somewhat accurate prediction of VOT values, depending on place of articulation of the consonant in question. In their pursuit, the authors discovered various patterns across examples from completely separate language families. One of the most significant universals for this particular study concerns VOT duration of stops, the authors stating that "velar stops always have a longer VOT (Cho/Ladefoged 1999: 218)". This regularity, however, applies only to languages that do not have uvular stops. Furthermore, in both aspirated and unaspirated stops, VOT is shortest before bilabial stops and intermediate before alveolar stops, with the exception of the unaspirated stops in Tamil and the aspirated stops in Cantonese and Eastern Armenian (Cho/Ladefoged 1999: 208)". Proof from 18 different languages led the authors to arbitrarily create four different classes of stops across languages, the main criterion being VOT duration with roughly set limits (in ms). The four categories are "unaspirated (30ms), slightly aspirated (50ms), aspirated (90ms) and highly aspirated stops (over 90ms)" (Cho/Ladefoged 1999: 223). The authors acknowledge that obvious exceptions exist, even in their own selected sample of languages. Nonetheless, such a generalization can help predict VOT values for different types of stops, speech cues necessary for their perception, as well as to predict the success rate of producing aspiration upon acquiring a foreign language.

The study conducted by Halle *et al.* presents on acoustic research of plosives and describes, firstly, the general phonetic features of English stops, illustrating their spectral features with sonograms. An equally important aspect of this research is a twofold

experiment, which includes both articulation and perception of plosives. Interestingly, the participants needed to identify English plosives in isolation and in syllable form, with alternating plosive positions (initial vs final), which proved feasible, but, nonetheless, a true challenge, according to the authors (1957: 108). Their findings imply that all correct answers were given owing to plosive transitions, i.e. in clearer contexts where they were followed or preceded by a certain vowel (Hale *et al.* 1957: 115). In other words, it was aspiration that proved to be one of the more significant factors of perception for native speakers.

For non-native speakers, however, the significance of aspiration varies depending on their mother tongue. As a long-established lingua franca, English has been examined as an L2 in context of various L1 languages. As Flege explained (1992), upon acquiring a foreign language, L1 and L2 contact is imminent, and, for this reason, transfer of various features arises as an interesting point of linguistic investigation.

A contrastive study conducted by Kim (2011) compares two languages with aspiration as their prominent feature: English and Korean. Namely, this paper examines the degree of aspiration used by Korean speakers of English with the aim of determining whether an accurate use of aspiration is directly related to language proficiency. The experiment is conducted with two native speakers (NS) of English and seven non-native (NNS), Korean speakers. Values collected from NS recordings are then used as a reference for those produced by NNS. Additionally, as our present study has no native speakers, the VOT values recorded by Kim will also be used as reference points here. Considering that VOT of Korean plosives are significantly longer than those of English (Kim 2011: 4), proficient speakers are expected to, consequently, shorten their VOT when speaking English. The results confirmed this initial hypothesis, together with the fact that less proficient speakers prolonged their VOT in English as well. However, an additional finding is that the initially mentioned contact between L1 and L2 (Flege 1992) is, in fact, mutual. The author has noted that, regardless of the NNS's English proficiency, Korean speakers shortened the VOT in their mother tongue which directly points at the cross language phonetic influence (Kim 2011: 8), i.e., that L1 is also susceptible to such effects.

Similarly, Shimizu conducts a more extensive analysis of Korean, Thai and Mandarin Chinese as opposed to English in terms of aspiration. The author identifies a possible challenge for NNS of English from this group of L1 languages since they have more complex types of aspiration: "Korean has a three-way contrast of tense, lax and aspirated stops. Thai also has a three-way contrast of voiced, voiceless unaspirated and aspirated stops, and Mandarin Chinese has a two-way contrast of voiceless unaspirated and voiceless aspirated stops in their stop series" (Shimizu 2011: 1818).

For this reason, this experiment included minimal pairs and triplets in respective L1 languages and English, all presented within carrier sentences. Acoustic analysis has shown that significant L1 transfer is present in the production of English plosives where NNS produce these sounds with a "considerable delay of voicing" (Shimizu 2011: 1818) and with strong aspiration. Another significant conclusion confirms that it is velar stops across all four languages are aspirated more prominently in comparison with bilabial and alveolar stops.

3. METHODOLOGY

A total of seven participants took part in this experiment. Three participants are in their first year of bachelor studies, studying for a degree that is not related to English (medicine and electrical engineering). Nonetheless, all three participants studied English during their primary and secondary education. The remaining four are all graduate students of the Faculty of Philology, University of Belgrade, particularly the English Department. All participants are female and their age ranges between 20 and 26. For the purpose of distinguishing between these two groups of participants, the three participants who do not have a degree in English will hereinafter be referred to as "advanced speakers" as opposed to "proficient speakers", who have obtained a degree, merely for the sake of labelling, without necessarily implying an accurate level of English.

The recording materials consist of target words with word-initial plosives in both English and Serbian, all of which were embedded into carrier sentences. For English, the words used were: *pat, cat, bat and tat*, while the Serbian counterparts included *Pata*, *tata*, *Kata* and *baba*. For the purpose of this research, only the tokens containing /p t k/ in Serbian and English were analyzed, while the two words with /b/ in word-initial position were used as distractors. All recordings were made directly in the latest version of the *Praat* software (Boersma/Weenink 2022), at the input frequency of 44100 Hz, while the participants read sentences shown on PowerPoint slides. Each participant read the sentences three times, which amounted to a total of 126 tokens to be analyzed for both languages, i.e. 63 in Serbian and 63 in English. After the recording was completed, each target word was cut from the original recording and analyzed separately as a token. All VOT values were then entered into a table in Excel, which was used to calculate mean values for each participant and draw conclusions.

4. RESULTS AND ANALYSIS

Prior to listing all values collected through the recordings of this research, it is necessary to present reference values for English stops produced by native speakers of English so as to compare and assess the results of this study. As it has been stated previously, the mean values stated here have been taken from Kim (2011) and will be used as reference points for drawing more reliable conclusions.

Table 1 confirms the universals regarding VOT duration, velar stops having the longest offset interval, followed by intermediate alveolar VOT and, finally, VOT of bilabial /p/ being the shortest (Cho/Ladefoged 1999: 208).

The measurement of VOT in our experiment was done for the interval between the release of the plosive and the onset of voicing of the following sound and expressed in milliseconds (ms). Since every target word had three repetitions per speaker, the values presented in the following table for each individual speaker is an average value for /p t k/. Firstly, all tokens were analyzed for advanced speakers (P1–P3 in Table 2, see Appendix A) and the average measurements for each participant were entered separately in Table 2 (see Appendix A).

Upon comparison with the reference values, it is evident that, though advanced speakers do produce a more prominent lag when it comes to word-initial stops, VOT is still not similar enough to the native speaker's level of production. Based on the

presented mean values, we can confirm that velar stops do have the most prominent VOT (Cho/Ladefoged 1999) in both respective languages. Mean VOT values for Serbian, too, remain in line with the initial hypothesis, as they are not prominent and the VOT duration is significantly shorter, unlike English.

The values for proficient speakers (P4–P7 in Table 3, see Appendix A) are firstly presented separately. What can immediately be noted is that the values are significantly closer if not fully in line with the native speakers' VOT values. However, there are a few surprising observations to be made: the average VOT value for the velar stop /k/ is the most prominent one, slightly longer than the native speakers themselves especially for P4 in both Serbian and English, presented in bold. On the other hand, the values for the Serbian /k/ are significantly higher than those of advanced speakers. A spectral representation of an unusually longer VOT in English and Serbian can be seen in Spectrogram 1 and Spectrogram 2, respectively (see Appendix B).

As participant P4 is a proficient EFL speaker and since the average VOT values for all four proficient speakers are significantly higher than those of P1–P3, it can be assumed that higher proficiency and increased exposure to a language with a long-lag aspiration feature can have an effect on the non-native speaker's VOT production in their own mother tongue. These results coincide with Kim's findings (2011) and in this way confirm the "bi-directional influence" (Grosjean 1989) of languages. In other words, in cases of significant exposure to a foreign language, that L2 is likely to affect a learner's L1 production as well.

To explore this matter further, what follows is a presentation of VOT values in Serbian produced by advanced as opposed to proficient speakers of English (Table 4, see Appendix A).

Based on the side-by-side comparison in Table 4, we can conclude that proficient speakers (P4–P7 in Table 4, see Appendix A) are more likely to assign longer VOT intervals even to Serbian plosives, mostly likely as a direct consequence of L2 transfer. However, Table 4 also confirms the existence of aspiration in the Serbian language for all 7 participants, though not as prominent and not as distinctive as in the English language.

The comparison of VOT in English between advanced and proficient speakers is presented separately in Table 5 (see Appendix A).

Regardless of NS reference values, it is immediately evident that proficient speakers have acquired the aspiration feature on a higher degree and have virtually no issues with applying it in their pronunciation. Conversely, though advanced speakers do have a stable command of English, in comparison with proficient speakers, they show a slightly less prominent use of aspiration in English.

5. FINAL REMARKS

The study presented in this paper examined VOT values of English and Serbian voiceless plosives produced by Serbian EFL speakers. Apart from direct comparison, one of the key aims of this research was to confirm whether a correlation exists between a speaker's language proficiency and the degree of positive VOT as expected in voiceless plosives. An additional aspect of this study was describing the status of aspiration in Serbian, as there are no clear records on this phonetic feature as it has no distinctive function in this language.

Though the results of this research answer the initial questions, they have also brought about the topic of L2 transfer. In the values presented earlier, it was to be expected that all non-native speakers who took part in the experiment would make a clear distinction between aspiration in English vs Serbian, as this feature is significantly less prominent in their mother tongue where it is not distinctive. Since the more proficient speakers have shown greater precision when producing aspiration, it can be maintained that a correlation between proficiency and accurate use of aspiration does exist.

Though the less proficient speakers have been somewhat less precise in terms of aspiration, they are more consistent when it comes to maintaining a clear distinction between aspiration in their L1 and L2. Conversely, the yielded values show that it is, in fact, the more proficient speakers that produce longer VOT in their mother tongue, which also points at the fact that, due to longer exposure and overall better command of English, L2 transfer has occurred over time, the interference causing cross phonetic influence between languages.

Since this study did not have its own pool of native speaker samples and had a limited number of participants, a larger sample of recordings needs to be made for both native and non-native speakers in order to draw more general conclusions.

REFERENCES

Boersma, P. and D. Weenink. 2022. *Praat: doing phonetics by computer* [Computer program]. Version 6.2.20. http://www.praat.org/

Cho, T. and P. Ladefoged. 1999. Variation and Universals in VOT: Evidence from 18 Languages. *Journal of Phonetics 27(2),* 207–229.

Čubrović, B. 2009. *Profiling English Phonetics*. Belgrade: Philologia.

Flege, J. and W. Eefting. 1987. Production and perception of English stops by native Spanish speakers. *Journal of Phonetics 15(1),* 67–83.

Flege, J. E. 1992. Speech Learning in a Second Language. In C. A. Ferguson *et al.* (eds.), *Phonological Development: Models, Research, Implications*. Timonium, MD: York Press, 565–604.

Flege, J. E. *et al.* 1995. Effects of Age of Second-Language Learning on the Production of English Consonants. *Speech Communication 16*, 1–26.

Grosjean, F. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language 36*, 3–15.

Halle, M. *et al.* 1957. Acoustic properties of stop consonants. *Journal of the Acoustical Society of America 29*, 107–116.

Kim, M-R. 2011. The Relationship between Cross-Language Phonetic Influences and L2 Proficiency in Terms of VOT. *Phonetics and Speech Sciences 3(3*), 3–10.

Shimizu, K. 2011. A Study on VOT of Initial Stops in English Produced by Korean, Thai and Chinese Speakers as L2 Learners. *International Congress on Phonetic Sciences XVII*, 1818–1821.

Zsiga, E. 2020. *The Sounds of Language*: *An Introduction to Phonetics and Phonology*. Malden, MA: Wiley-Blackwell.

APPENDIX A

The appendix contains five tables with average VOT values for different groups of participants in both English and Serbian, each specified below.

	p	t	k
Native Speakers	58	70	80

Table 1. Native speakers' mean VOT values for English (Kim 2011: 4)

	SRB			ENG		
participants	p	t	k	p	t	k
P1	13.08	15.95	39.33	57.42	49.61	64.07
P2	12.32	14.98	41.23	55.41	42.82	68.54
P3	12.97	15.65	43.03	56.09	45.03	60.05
AVG	12.79	15.53	41.2	56.31	45.82	64.22

Table 2. Advanced speakers' mean VOT values for Serbian & English

	SRB			ENG		
participants	p	t	k	p	t	k
P4	17.12	20.5	**70.32**	63.37	68.2	**84.17**
P5	16.32	19.02	59.67	61.54	69.13	83.87
P6	13.21	17.8	35.6	62.44	70.01	83.2
P7	16.87	20.02	60.89	62.89	66.78	84.03
AVG	15.88	19.335	56.5425	62.56	68.53	84.3525

Table 3. Proficient speakers' mean VOT values for Serbian & English

	SRB		
participants	p	t	k
P1	13.08	15.95	39.33
P2	12.32	14.98	41.23
P3	12.97	15.65	43.03
P4	17.12	20.5	70.32
P5	16.32	19.02	59.67
P6	13.21	17.8	35.6
P7	16.87	20.02	60.89

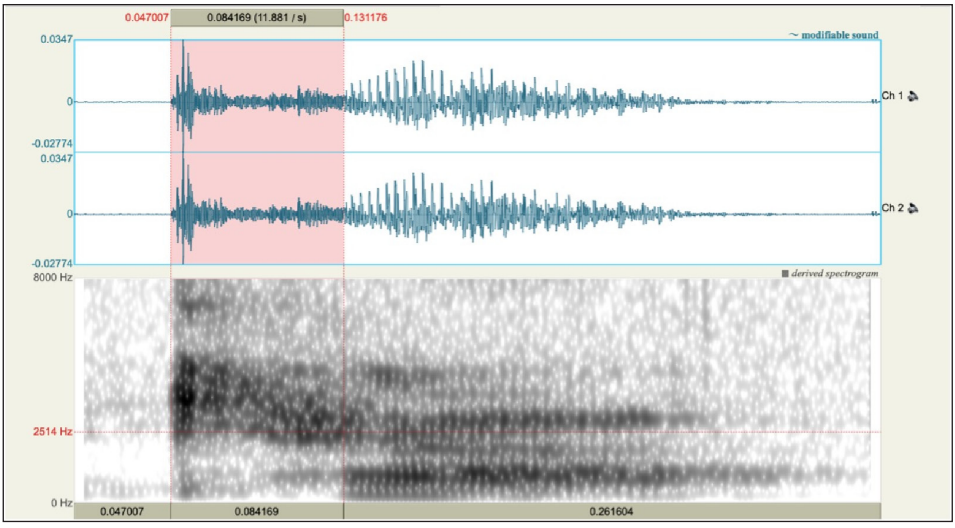
Table 4. All participants' VOT values for Serbian

	ENG		
participants	p	t	k
P1	57.42	49.61	64.07
P2	55.41	42.82	68.54
P3	56.09	45.03	60.05
P4	63.37	68.2	84.17
P5	61.54	69.13	83.87
P6	62.44	70.01	84.2
P7	62.89	66.78	84.03

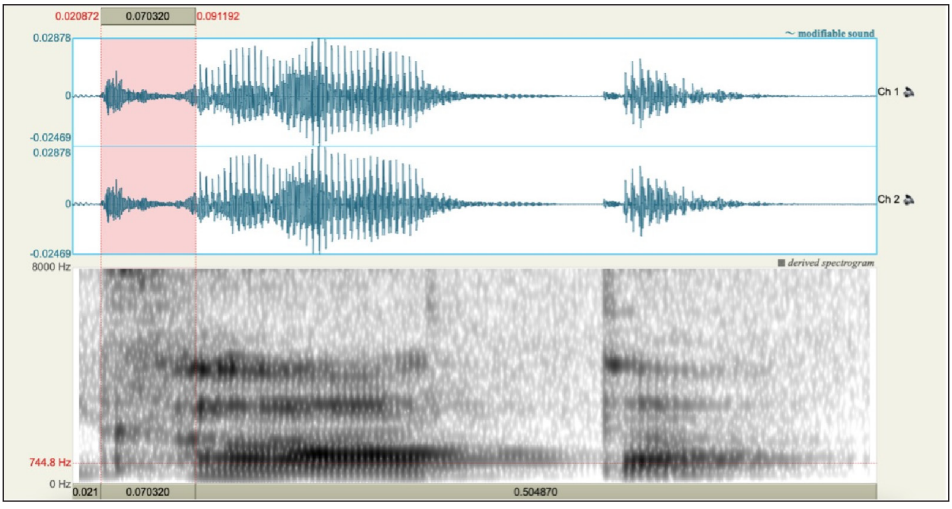
Table 5. All participants' VOT values for English

APPENDIX B

The appendix contains 2 spectral representations of recorded speech by speaker P4, who belongs to the proficient group of non-native speakers in this study.



Spectrogram 1. Word /kæt/ by participant P4



Spectrogram 2. Word /ká:ta/ by participant P4

SUMMARY

GREAT ASPIRATIONS: EXAMINING VOT OF WORD-INITIAL VOICELESS STOPS IN ENGLISH AND SERBIAN IN SERBIAN EFL SPEAKERS

The role of aspiration (positive VOT) in English and Serbian differs significantly. While positive VOT helps distinguish between voiced and voiceless plosives in English, it has no distinctive function in Serbian. For this reason, a clear contrast between long-lag stops of English and short-lag stops of Serbian might prove challenging for Serbian EFL students. For this reason, this paper examines the degree of aspiration in initial voiceless stops in both Serbian and English with two separate target groups: five intermediate and five highly proficient speakers of English. The aim of this study is to determine to what extent L1 affects L2 speech and how this is reflected in the two groups of speakers. All participants are asked to read sentences containing target words with voiceless stops in initial position while being recorded with the help of *Praat* (Boersma/Weenink, 2022). For every token, we measure VOT for the interval between the release of the plosive and the onset of voicing of the following sound, expressed in milliseconds (ms). The results show that more proficient EFL speakers reach native-like VOT values. This experiment, however, yielded some striking results, specifically the proof of significant L2 transfer, where the same proficient speakers pronounce their stops with a significant degree of aspiration in Serbian, which deviates from the phonetic habits of native speakers of Serbian. Results also show that, even though the use of long-lag stops in intermediate users does not resemble that of native speakers, they make a clearer distinction when it comes to applying VOT in their L1 and L2 respectively.

KEYWORDS: aspiration, VOT, L1 transfer, L2 transfer, Serbian EFL speakers.

ARTICLE INFO:
Original research article
Received: February 28, 2023
Revised: May 10, 2023
Accepted: November 1, 2023