

**Danica M. Jerotijević Tišma<sup>1</sup>**  
*University of Kragujevac*  
*Faculty of Philology and Arts*  
*English Department*

## **“I WENT TO VENICE AND WAS VERY WORRIED.” – DIFFERENTIAL SUBSTITUTION OF THE /V/-/W/ CONTRAST IN SERBIAN-ENGLISH INTERPHONOLOGY**

Previous findings of SLA research underlined the significance of differential substitution as one of the most common strategies for overcoming difficulties in pronouncing English consonants. The acquisition of /v/-/w/ contrast seems particularly demanding for Serbian EFL learners, due to diverse factors, the dominant one being mother tongue interference and the resulting perceptual assimilation. Having the previously stated in mind, as well as the scarcity of research in the Serbian scientific context, the present study aimed at investigating instances of differential substitution of the English labiodental fricative and the labial-velar approximant in relation to the phonetic context and the formality of task type. Both the perception and production of a total of 72 English-major students were tested using different instruments, including phonemic discrimination task, wordlist as well as paragraph reading, and the interview. The obtained data were quantitatively analyzed and supported by spectrogram illustrations of both native and non-native speakers of English. The results indicate a greater frequency of /v/-substitution ([w] being the main substitute) with phonetic context and task type playing a statistically significant role in differential substitution. Instances of substitution of the labial-velar approximant were not as frequent and are said to be characteristic of a specific group of learners (Serbian labiodental sonorant [v] being the main substitute). The results underscore the variability and complexity of the interlanguage system and the importance of valuing the process, and not merely the outcomes of L2 sound acquisition.

**Keywords:** /v/-/w/ contrast, substitution, interphonology, Serbian, English

### **1. INTRODUCTION**

The processes of speech perception defined as recognizing and interpreting speech sounds, i.e. analyzing their physical properties while simultaneously mapping them to the existing mental representations, and of speech production identified as the coordination of articulators to generate recognizable speech sounds, seem complex enough to undergo in a single language.

<sup>1</sup> danica.tisma@filum.kg.ac.rs

The situation becomes even more demanding when one or more foreign languages are learnt or acquired, since various factors contribute to the overall success or failure of the processes in question. The inevitable difference in the way native and non-native speakers perceive vowels or consonants was confirmed in numerous studies, especially on examples of the perception of /l/-/r/, but also /b/-/v/ and /s/-/θ/ contrasts (Guion et al. 2000: 2712; Brown 2000: 55-59). Failing to observe perceptual nuances, Dutch speakers have difficulty with final /v/, because they pronounce it as /f/, due to final obstruent devoicing in their L1 (Cutler et al. 2004: 3672). The phonological inventory of the native or foreign language alone will sometimes not be sufficient to explain the perception of sounds, as it will diverge depending on the analyzed contrast. The perception of new phonemic and phonetic contrasts will predominantly be shaped by the native phonological system, so the main issue is whether speakers, especially adults, are able to integrate multiple acoustic parameters into the perception of a non-native contrast.

The /v/-/w/ contrast is notorious for being problematic for EFL learners, regardless of the linguistic background. For instance, Piyamat and Deekawong (2021) report on the difficulties Thai EFL learners face with these sounds, while Ankerstein and Morschett (2013) previously detected the problems with German children's acquisition of the same English contrast. Mother tongue interference seems to intervene with the pronunciation of the /v/-/w/ in L2 English for Turkish learners, as well (Ercan 2018).

To the best of our knowledge, few studies have so far tackled any aspects of /v/-/w/ contrast acquisition by Serbian EFL learners, hence the primary aim of the present paper is to contribute to expanding the research done with other consonant classes, e.g. plosives (Jakovljević 2011; Čubrović 2011, 2013) and interdental fricatives (Sredović Trpeski 2013; Jerotijević 2014). Moreover, the present research seeks to offer explanations on instances of differential substitution (Brannen 2022) depending on the phonetic context and the formality of task type.

## 2. THEORETICAL BACKGROUND AND RATIONALE

Although it may seem rather straightforward, L2 perception is not a simple process of deciding whether two sounds from the native and foreign languages are similar or different, yet it requires a two-way distinction. The *Perceptual Assimilation Model* (Best 1994) is an offshoot of the direct-realist philosophical orientation advocating that humans directly perceive the external world using their senses. The model predicts that the difficulty in perceiving the sounds of a foreign language stems from the phonetic and articulatory similarities and differences between the native and a target sound. The newly acquired phoneme is assimilated into the already existing system of the native language, and the ability to distinguish and pronounce the target sounds is dependent on this particular assimilation process (Best et al. 2001: 777). If the phonetic contrast from the foreign language is assimilated into separate categories of the native language, it will result in an accurate

distinction. If the opposite happens, i.e. if the contrast is assimilated into one category of the native language, the accuracy of distinguishing the contrast will depend on how well the category has been assimilated (Best 1994: 190). The /v/-/w/ contrast in English and Serbian, investigated in the present research study, belongs to the *Category-Goodness Difference Type* (one of the four possible acquisition scenarios explained by the model (Best 1995: 124-125)), with the assumed outcome of being assimilated into the same L1 phone and the success of acquisition ranging from moderate to good.

According to the *Speech Learning Model* (Flege 1995), perception is the main cause of foreign accentedness, resulting from the so-called *equivalence classification*. Equivalence classification is a cognitive mechanism that allows people to perceive constant categories despite the inexhaustible variability of the speech output (Flege 1987: 49). As a result of the comparison and analysis of the L1 phonological system and the sounds of the target language, the sounds are categorized either as new, similar or identical. Challenges of clearly defining the previous concepts arise from the lack of objective instruments for measuring the degree of perceived phonetic distance (Flege 1995: 264). Identical sounds should not present a problem in acquisition. They share the same symbol of the International Phonetic Alphabet and do not differ acoustically or perceptually to a significant extent. They are subject to equivalence classification without negatively affecting the production. New sounds do not have a counterpart in the phonological system of the native language and therefore differ acoustically. Whereas new sounds are represented by a completely different IPA symbol, similar sounds are marked with the same symbol, but are acoustically diverse. It is precisely this subtle phonetic distinction that makes it especially problematic for a learner to perceive it and/or eventually produce it in a native-like manner. The currently explored /v/-/w/ contrast is particularly interesting because the two sounds can be regarded as similar even though it may not seem so, since /w/ is not integral to the Serbian phonological system. However, some authors have recognized labial-velar pronunciations in specific regional varieties (Petrović, Gudurić 2010). One more issue seems worth reconsidering here. Namely, the labiodental /v/ can be classified as similar, rather than identical, due to the reported differences in articulations in English and Serbian to be discussed in the next section of the paper.

Both of the models have been revised fairly recently, resulting in modified versions PAM-L2 (Best, Tyler 2007) and SLM-r (Flege, Bohn 2021) which basically expound on similar ideas and findings of the original versions. The revised version of the *Speech Learning Model* underscores the co-evolving of perception and production. The process of L2 category formation should be valued as exactly what the term suggests – the process and not the final result. Learners need to carefully consider L1 and L2 phonetic features and learner diversity is particularly emphasized. PAM-L2 and SLM-r dominate the research in SLA of phonology which is evident by a substantial number of studies (Flege et al. 2003; Aoyama et al. 2004; Bundgaard-Nielsen et al. 2011; Tyler

et al. 2014), many of which are very recent, as well (Laméris et al. 2023; Kim 2023; Zhou, Rato 2023; Aoyama et al. 2023).

### 3. *THE /v/-/w/ CONTRAST IN THE PHONOLOGY OF SERBIAN AND ENGLISH – SOURCES OF DIFFERENTIAL SUBSTITUTION*

Considering the phonological system of English and Serbian, /w/ is a central approximant, characterized either as bilabial (Roach 1990: 61) or labial-velar (Ladefoged 2003). Consonants that precede it will have rounded lips due to the anticipation of /w/. The phoneme /w/ does not have a special phonemic status in Serbian, but some authors list it as an allophone of the labiodental /v/ (Belić 1972: 54-55; Petrović, Gudurić 2010: 235). It should be noted that the formants of the approximants are not seen as a straight line, yet rather a curve on the spectrogram, and the previously mentioned characteristics of the formants are present in the parts of the spectrum where the features of the approximants are most typical (Lisker 1995: 130-132).

In Serbian, the upper incisors touch the moist part of the lower lip for /v/ (Miletić 1933: 41), while in English, the place of contact will vary depending on the following sound (Gimson 1978: 181). Belić (1972: 53) classifies /v/ as a fricative identifying its multiple articulations from bilabial to labiodental. The ambivalent nature of /v/ and the question of sonority is not unique to the Serbian language, and studies confirm it in other Slavic languages (Rubach 1993), as well as e.g. in Hungarian (Zsigri 1998), Swedish (Lombardi 1995). Both labiodental [f] and [v] in English have the main energy concentration at about 6000–8000 Hz, i.e. below 10000 Hz, but are often very faint due to low intensity, and the spectrum is flat with no peaks (Baart 2010: 71). The interesting features of the articulation of the /v/ in Serbian are certainly reflected on the spectrogram, so the sonorant nature of the sound will be marked by the presence of a voice bar and a slightly higher value of the first formant. The lowering of the first formant will indicate its semi-vowel-like nature and the articulation close to /w/, while the fricative realization will cause the appearance of scattered energy in the spectrum (Petrović, Gudurić 2010: 235). Its similarity to the vowel [u] is also mentioned in the literature (Simić, Ostojić 1996: 181). The labial-velar approximant [w] is likewise acoustically similar to the cardinal vowel [u], and its double articulation on the lips and soft palate will cause acoustic specificities (Ladefoged 2003: 140). Hardly anything is visible above the frequency of 1000 Hz, because the amplitude of the approximant decreases due to the narrowing of the vocal tract. Like the vowel [u], the approximant [w] will have a low first formant, but the second formant will also be significantly lower due to the movement of the tongue towards the back of the oral cavity and the rounding of the lips (Kent, Read 2002: 177).

Differential substitution is a process that occurs when an L2 learner replaces a target sound with a sound from the native language found to be

perceptually adequate enough (Brannen 2002). The resulting difference between the target L2 pronunciation and the learner's actual pronunciation can occasionally affect the intelligibility and impact the overall proficiency in the L2. Substitution is somewhat reminiscent of similar processes during the acquisition of phonology of a native language in children, which can be seen in the strategies, such as replacing fricatives and affricates with stop consonants, voicing obstruents before vowels and devoicing final obstruents, pronouncing sounds in the front part of the oral cavity, replacing liquid consonants with approximants, as well as the well-known substitution of plosives /p/ and /t/ (Ingram 1991: 57-59). For example, Chinese speakers choose /s/ and /z/ as substitutes for the English interdental fricatives /θ/ and /ð/ (Rau et al. 2009: 599), while Serbian speakers opt for the dental plosives /t/ and /d/ (Lee 2006: 132). However, it is traditionally assumed that speakers use only one fixed substitute, nevertheless, studies have shown that substitution is further characterized by variability (Rau et al. 2009: 582). Differential substitutes and the aforementioned fricatives are not the same sounds, but the confusion occurs due to the similarities in acoustic features (Ladefoged 2006).

Having the previously presented specificities of the /v/-/w/ contrast in Serbian and English in mind, the assumption is that some learners may find it difficult to establish two separate categories for the two sounds throughout the process of L2 sound acquisition and will potentially exhibit instances of differential substitution interchangeably.

## 4. METHODOLOGY

### 4.1. Aims

The overall aim of the present study is to explain instances of differential substitution of the /v/-/w/ contrast in Serbian-English interphonology from acoustic and articulatory perspectives. Furthermore, the goal is to investigate the accuracy of Serbian EFL learners' phonemic discrimination of the target contrast seeking to connect it to the actual production. Two factors are predominantly considered in production – phonetic context and formality of task type.

### 4.2. Research Questions

The current study is based on the following research questions:

- How accurate is Serbian EFL learners' phonemic discrimination of the /v/-/w/ contrast?
- How frequent is the differential substitution of the /v/-/w/ contrast depending on the phonetic context?
- How frequent is the differential substitution of the /v/-/w/ contrast depending on the formality of task type?
- How does the production of the /v/-/w/ contrast by non-native speakers compare to the production of native speakers based on the analysis of the spectrogram?

### 4.3. Participants

Bearing in mind that the current research was a part of a larger-scale study, a total of 72 first-year students at the English Department of the Faculty of Philology and Arts, University of Kragujevac (academic year 2013/2014), participated in the investigations of perception and production. All the relevant factors established in the literature were controlled in the sampling process, esp. gender, age, L1, speech or hearing disorders etc. The average age of our respondents was 18.97, the majority was female (65.3%), mostly coming from Kragujevac (27.8%), having previously attended grammar school (54.2%). A more detailed description of the sample can be found in Table 1.

**Table 1. Information on Participants**

Age	Gender	Place of Birth	Secondary Education
mean 18.97 min. 18 max. 22 st.dev.=0.43	Male 25 (34.7%) Female 47 (65.3%)  st.dev.=0.48	Kragujevac 20 (27.8%) Kraljevo 11 (15.3%) Jagodina 4 (5.6%) Raška 3 (4.2%) Kruševac 3 (4.2%) Despotovac 3 (4.2%) Other 28 (38.39%)  st.dev.=7.32	Grammar School 39 (54.2%) Economy 13 (18.1%) Technical Sciences 11 (15.3%) Medicine 4 (5.6%) Other 5 (6.9%)  st.dev.=1.25
N=72 (100%)	N=72 (100%)	N=72 (100%)	N=72 (100%)

In order to compare the production of the target consonants, two native speakers of the American English variety, male (Andover, Massachusetts) and female (Cleveland, Ohio), average age = 23.5 were also included. These speakers are not laymen in the linguistic sense, since both of them worked as instructors at the Department of English at the faculties of philology of the University of Novi Pazar and Kragujevac, respectively.

### 4.4. Instruments

To test the perception of the target contrast, a phonemic discrimination test was used (Silbert et al. 2015: 101), considering relevant findings from previous studies of consonant perception in the context of SLA (cf. Iverson et al. 2008; De Jong et al. 2009).

In the process of the vocabulary selection, we made sure that all the words were familiar in order to avoid focusing on the meaning instead of the perception of sounds. We used the guidelines for the B1 CEFR level<sup>2</sup>. The wordlist and minimal pairs were combined from several sources (Gimson 1978: 149-219; Jones 1979: 138-216; Jones 2012). The task in the phonemic discrimination test was to listen to a recording of a native speaker of American English

<sup>2</sup> Available on <http://www.englishprofile.org/wordlists>

reading words containing target sounds. The participants were provided with a list of minimal pairs and the task was to determine which of the two words was pronounced by the native speaker. Phonemic contrasts were given in the initial position due to the phonotactic restrictions of English, i.e. wherever it was possible to form a minimal pair.

The instruments used in production testing were threefold: a wordlist with target sounds as the most formal task type aimed to obtain careful pronunciation from the respondents (Čubrović 2013), two paragraphs to be read aloud (Diaz, Campos 2004; Paunović 2011) as a somewhat more informal task type, and the interview, which was aimed at eliciting spontaneous speech (Labov 1984). When choosing the material for recording, the choice of vocabulary was carefully considered. The task type variation was deemed necessary in order to avoid the predictability of the pronunciation of the recorded participants thereby losing potential phonetic specificities (Harrington 2010: 7). The wordlist comprised tokens from similar sources as for the perception tests (Gimson 1978: 149-219; Jones 1979: 138-216; Rau et al. 2009: 614-615; Jones 2012). Target phonemes were given in word-initial, medial, and final positions. Regarding the instrument for paragraph reading, two paragraphs were adapted from *The Story of Arthur the Rat* (Markham, Hazan 2002: 16) and *The Three Little Pigs* (Rau et al. 2009: 610-611). The questions in the interview resembled a semi-structured interview (Saunders et al. 2003: 245-246) in the sense that for most questions there was a sub-question in case the participant faltered or provided a minimal answer. The interview contained 37 questions, independently selected and compiled from several sources (Labov 1984; Rau et al. 2009: 615) in accordance with the recommendations of the relevant literature (Rapley 2004; Holstein, Gubrium 2004).

#### 4.5. Procedure

The perception was tested during the *English Phonetics* course in November 2013. Respondents listened to the recording of a native speaker and filled in lists with answers. Each word was read twice, or rarely three times, if there was an unpredictable disturbance. The testing of the production and the recording of the participants was performed during the winter semester of 2013/2014 academic year again within the *English Phonetics* course. A native speaker (Newark, New Jersey) pronouncing the target sounds in the words for the purposes of perception testing was recorded during the summer of 2013 in Jagodina, Serbia. The other two speakers whose production was later analyzed for comparison (Andover, Massachusetts and Cleveland, Ohio) were recorded in November 2015 in Novi Pazar and Belgrade. The recording was done using the Audacity 2.0.5 computer software<sup>3</sup>, Vivanco dm40 microphone, with 44.1 kHz sampling and 16-bit conversion, saved in WAV format for word lists and paragraph reading (Bettagere, Fucci 1999), and computer microphones for interviews (Ternström 2008: 10). The native speakers were recorded using the

3 Available on <http://audacity.sourceforge.net/>

Olympus Digital Voice Recorder VN-8600PC with a built-in microphone of the same sampling rate and conversion. In order to ensure a uniform quality of the recordings, all the previously mentioned considerations were listed beforehand in a set of recording instructions for interviews. The recordings were doublechecked and sent back for correction in case of error detection or malfunction. The recorded material was transcribed, annotated and segmented in the *Praat* software, version 6.02.3 (Boersma, Weenink 2016) by the author of the present paper, relying on oscillograms and spectrograms combined with the auditory method. The segmentation criteria followed the suggestions and instructions of previous studies (Peterson, Lehiste 1960; Machač, Skarnitzl 2009: 79, 92). Statistical processing of the perception and production testing results was done in the statistical program SPSS, version 20.0 (Field 2009).

## 5. RESULTS AND DISCUSSION

### 5.1. Perception Testing Results

The results of perception testing are presented in Table 2.

**Table 2. Phonemic Discrimination Results**

Target contrast	Phonemic Discrimination Results		
	Accuracy	N	St. dev.
[w] - [v]			
#_e	81.9%	72	0.387
#_aɪ	87.5%	72	0.333
#_ɜ (ɜ:)	90.3%	72	0.362
#_eɪ	91.7%	72	0.278
#_ɪə (ɪə)	90.3%	72	0.362
#_i:	88.9%	72	0.316
<b>Total</b>	88.43%	432	0.341

The average accuracy is around 88% which points to a high level of perceptual abilities of the chosen sample of participants, at least when it comes to the particular sound contrast. The percentage is the lowest in initial position preceding a short front-mid monophthong /e/, and the highest, interestingly enough, before the closing diphthong /eɪ/, beginning with the same vowel. Such a situation may have resulted from the specific nature of interlanguage itself, being variable and unpredictable. It may also indicate that the process of category formation is ongoing. ANOVA results demonstrate that there is no statistically significant difference in perception accuracy pertaining to different contexts ( $F=1.433$   $p=0.284$ ). Tukey's post hoc test of multiple comparisons singles out the initial position for the contrast [v]-[w]/#\_e as statistically significant compared to other positions ( $F=4.893$   $p=0.021$   $df=14$ ). The absence of absolutely correct perception of the given contrast is expected considering

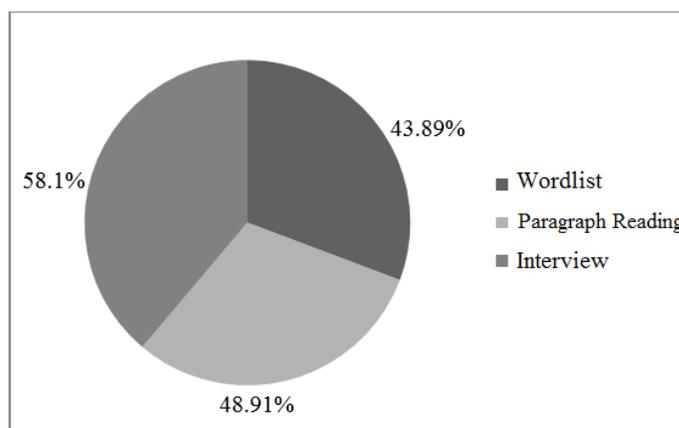
the fact that, following the selected theoretical framework, learners are expected to assimilate the given contrast to only one existing sound in the native phonological system.

## 5.2. Production Testing Results

### 5.2.1. Differential Substitution of the Labiodental Fricative /v/

The overall percentage of differential substitution in relation to the total number of realizations of the voiced labiodental fricative in the wordlist, paragraph reading and the interview, is presented in Graph 1.

**Graph 1. Differential Substitution of the Voiced Labiodental Fricative for Different Tasks**



Due to the similarity of the sounds and the resulting uncertainty, learners exhibit unsystematic pronunciation and hypercorrection. Graph 1 shows that almost one half of the total number of realized /v/ instances are substituted regardless of the task, with the interview showing the highest percentage of substitution (58.1%). It seems important to note that the number of participants who consistently pronounce labiodental fricatives is 23 or 31.94%. Furthermore, it was noted that one and the same speaker may produce different realizations in all the tasks. It is interesting that Serbian EFL learners resort to hypercorrection pronouncing the labial-velar approximant instead of the English labiodental fricative even in final positions, although the phonotactics of the English language does not allow this. It seems worth mentioning that the number of substitutions in the final position decreases in the paragraph reading task, then increases again in the interview. In the remaining contexts, we observe an upward trend in relatively regular intervals of about 3-5%. In this sense, it can be said that the decrease in formality had influenced the production of substitutes for the labiodental fricative.

To observe the aforementioned influence more closely, the results of the statistical testing are presented in Table 3 including the percentage counts of differential substitutes for /v/ regarding both phonetic context and task types.

**Table 3. Distribution of Differential Substitution of the Voiced Labiodental Fricative and the Results of Statistical Testing**

Phonetic Context	Task Types			Statistical Analyses
	Wordlist	Paragraph Reading	Interview	
#_V,C	[w] 16.13%	[w] 20.13%	[w] 25.69%	<b>Correlation</b> Context: R=-0.612 p=0.040 Tasks: R=0.587 p=0.048 <b>ANOVA</b> F=7.678 p=0.022 <b>Regression</b> Adjusted R <sup>2</sup> =0.625 Coefficients (non- standardized) Context: B=4.207 t=2.828 p=0.030 Tasks: B=4.035 t=2.713 p=0.035
V,C_V,C	[w] 15.59%	[w] 17.75%	[w] 18.17%	
V,C_#	[w] 12.17%	[w] 10.03%	[w] 14.24%	
Total	[w] 43.89%	[w] 48.91%	[w] 58.1%	
N	2046	2592	864	

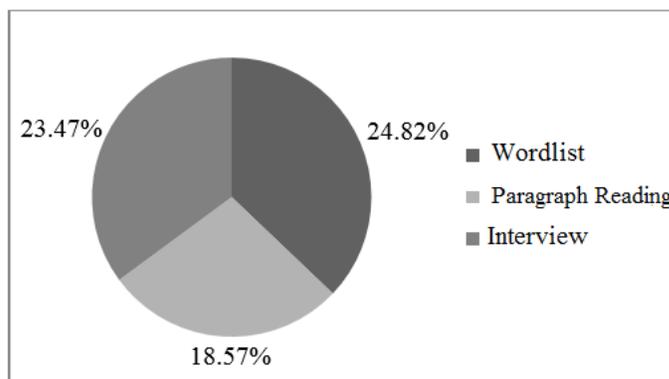
Judging by our corpus, the main substitute of the labiodental fricative is the hypercorrected articulation of the labial-velar approximant. The substitution occurs in initial positions prevaillingly, and less often in medial positions. As the formality of the task type decreases, the frequency of substitution increases. Hypercorrection of the labiodental fricative occurs in careful speech such as wordlist reading, being even more frequent in connected speech. The findings indicate that the correlation is statistically significant for both the context and the task type, and that it is positive in both cases, meaning that as informality increases, so does the number of substitutes. Moreover, there is a particular increase in the number of substitutes in the final compared to the initial positions. The ANOVA results show statistically significant results of multiple regression, so we can say that even 62.5% of the variability of substitutes

can be explained by independent variables, context and formality of the task. Both predictors are statistically significant, which can be seen from the data on unstandardized coefficients. It is assumed that for each subsequent context or task, the variability increases by about 4%. Based on examples from the corpus, it may be concluded that the formality of task type and phonetic context play a significant role in predicting the number of substitutes, which is consistent with earlier research on the influence of phonetic context (Hansen 2001) and the formality of task types (Diaz-Campos 2004; Rau et al. 2009).

### 5.2.2. Differential Substitution of the Labial-velar Approximant /w/

Compared to the labiodental fricative, the labial-velar approximant has significantly fewer examples of substitution (Graph 2).

**Graph 2. Differential Substitution of the Labial-velar Approximant for Different Tasks**



A more frequent occurrence of substitutes was found with seven participants only, i.e. with 9.72%. In the other examples and among the other respondents, the substitutes appear sporadically. The percentage of substitution is the lowest in paragraph reading task, and in the task of wordlist reading and interview the percentage was similar, pointing to no difference in two of the most extreme tasks regarding formality. Contrary to the assumptions, and unlike the situation with the labiodental fricative, the substitution of the English labial-velar approximant is somewhat constant, perhaps representing an individual variation characteristic of a certain group of speakers who repeat a similar pronunciation across task types. Such a state of affairs may be the result of inaccurate perception of the target sound, or the impossibility of making subtle modifications of the lips and tongue in order to generate the English approximant in question.

The percentage counts and the results of statistical analyses for the labial-velar approximant are presented in Table 4, showing variation both for the phonetic context and task types.

**Table 4. Distribution of Differential Substitution of the Labial-velar Approximant and the Results of Statistical Testing**

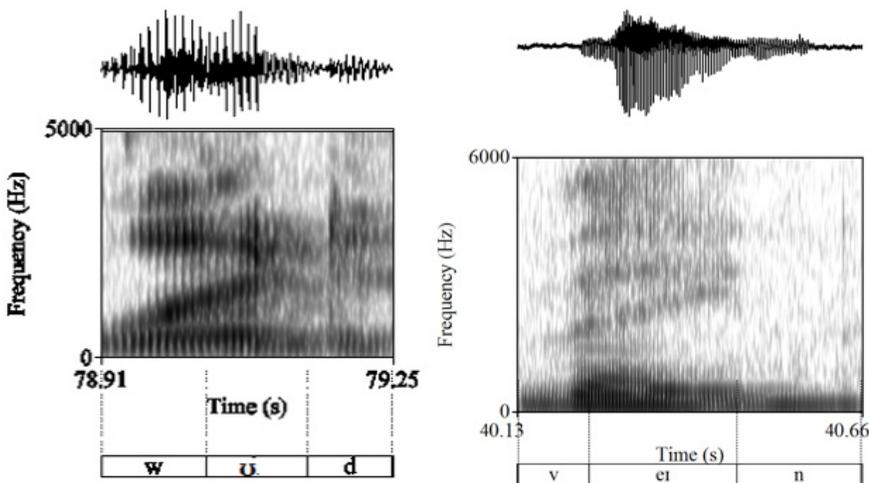
Phonetic Context	Task Type			Statistical Analyses
	Wordlist	Paragraph Reading	Interview	
#_V,C	[v] 10.11%	[v] 15.18%	[v] 14.17%	<b>Correlation</b> Context: R=-0.485 p=0.165 Tasks: R=-0.066 p=0.450 <b>ANOVA</b> F=0.474 p=0.663 <b>Regression</b> Adjusted R <sup>2</sup> =0.267 Coefficients (non-standardized) Context: B=-4.017 t=-0.964 p=0.406 Tasks: B=-0.335 t=-0.131 p=0.904
V,C_V,C	[v] 14.71%	[v] 3.39%	[v] 9.31%	
Total	[v] 24.82%	[v] 18.57%	[v] 23.48%	
N	1102	1381	720	

When we take a closer look at the substitution of the labial-velar approximant, we notice an interesting inconsistency in terms of phonetic context and tasks (Table 4), even though the results of statistical analyses do not point to the existence of the statistically significant difference. Namely, in the most formal task, the percentage of substitutions in the medial position is higher, while in the paragraph reading task, the given difference increases drastically for the initial position. In the interview, the number of substitutes increases in the medial position, and decreases in the initial position. Such oscillations in the number of substitutions depending on the context and the formality of the task type underscore the non-systematic nature of differential substitution in the interlanguage system, particularly when it comes to the labial-velar

approximant as a new sound non-existent in the phonological system of Serbian. In general, substitution is not present to a large extent, which means that the sound does not cause major problems pertaining to new category recognition, but it is necessary to examine other factors that could influence the choice of the substitute. More precisely, individual psychological factors should be included, as well as the age of onset and the level of L2 proficiency. The multiple regression test shows no statistical significance of the results, meaning that the variation related to the substitution of the labial-velar approximant is an outcome of random variability rather than the influence of the context and formality of the speaking style.

The previously presented quantitative data are best supported by spectrogram analysis. Hence, to begin with, Figure 1 shows the spectrogram of the labial-velar approximant and the labiodental fricative pronounced by a native American English speaker in initial positions to serve as an illustration for further comparison with the non-native realizations.

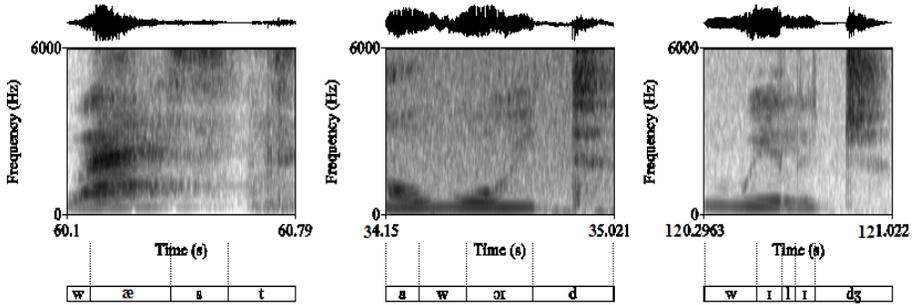
**Figure 1. Labial-velar Approximant /w/ and Labiodental Fricative /v/ Articulated by the Native Speaker**



The spectrogram on the left shows consistent, yet faintly visible formant structure with the typical lowering of the second formant due to tongue retraction. The frication is barely evident in the spectrogram on the right, characteristic of nonsibilant fricatives. Furthermore, both spectrograms display /w/ and /v/ with fully preserved voicing and a clearly visible voice bar.

Figure 2 displays the spectrograms of the initial and medial [v] produced by a non-native speaker across the three tasks.

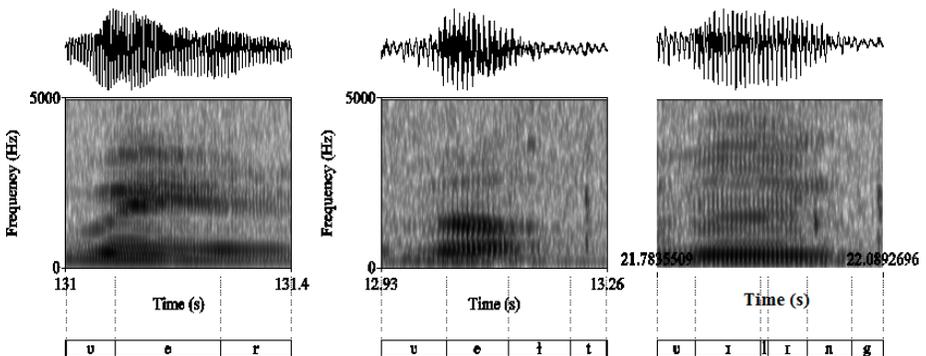
**Figure 2. Non-native Speakers' Substitution of [v] in the Wordlist, Paragraph Reading and the Interview (words from left to right: vast [væst], avoid [ə'vɔɪd], village ['vɪlɪdʒ])**



The velarized pronunciation is evidenced by the lowering of the second formant. A long period of occlusion is visible before the plosive and the plosive segment of the affricate, as well as the devoicing of final obstruents in all three examples.

Due to mother tongue interference, Serbian speakers occasionally realize the labial-velar approximant as Serbian labiodental [v]. In such cases the substitute for [w] is not the same as the English labiodental fricative, but its sonorant nature emerges as more dominant. There is no strict regularity of the labiodental sonorant substitute for the labial-velar approximant, and the variability is great. However, it can be found in all three tasks, most frequently in the initial prevocalic position (Figure 3).

**Figure 3. Non-native Speakers' Substitution of [w] in the Wordlist, Paragraph Reading and the Interview (words from left to right: where [weə-], wealth [weθ], willing ['wɪlɪŋ])**

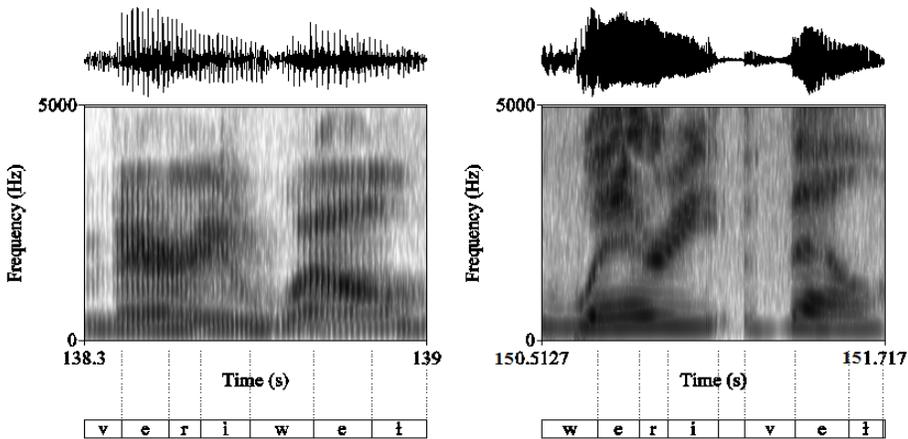


The first and third spectrograms show the initial approximant preserving voicing, while in the second picture the voicing begins, then seems to disappear, and returns again. The situation further confirms the unusual nature of the sound. In the first spectrogram there are elements of velarization and lip

rounding, however, in the recording it could clearly be heard that the speaker was pronouncing a substitute for the labial-velar approximant. In the second spectrogram, we notice that the speaker replaces the voiceless dental fricative with an alveolar plosive. The last spectrogram is interesting because at the end we see the typical combination alveolar nasal + plosive instead of a velar nasal.

It seems useful to compare the pronunciation of a native speaker and the pronunciation of a non-native speaker in the same example (combination: very well) (Figure 4).

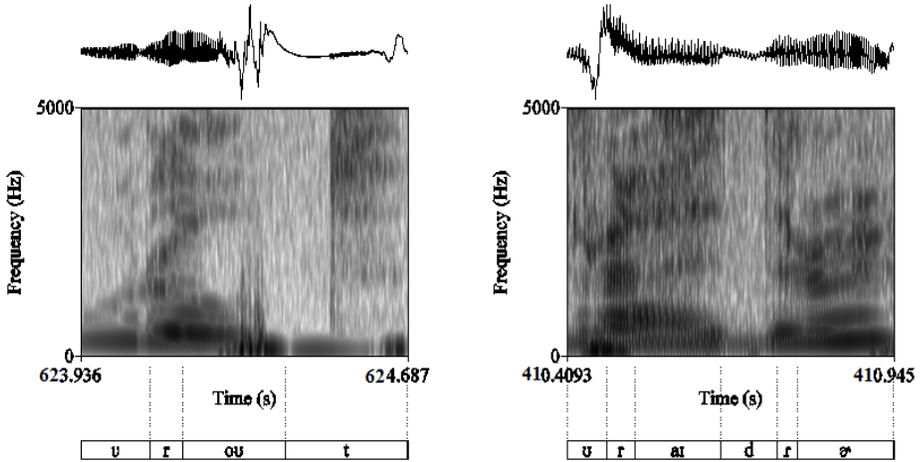
**Figure 4. The /v/-/w/ Contrast Produced by a Native (left) and a Non-Native Speaker (right) in the Same Example of Words**



The phrase *very well* [ˈveri weɪ] found in the corpus seems quite useful for comparison. Furthermore, it demonstrates the differences in the pronunciation of the two sounds. As such, it can be rather confusing for non-native speakers, resulting in inter-substitution. Velarization of the initial labiodental fricative is evident with the non-native speaker on the right, as well as the formant structure without the lowering of the third formant at the beginning of the second word in the sequence. The situation is exactly the opposite with the native speaker. Such examples are frequent in the corpus, especially if both sounds appear in a sequence or in neighbouring words.

Finally, Figure 5 shows a completely unacceptable, yet frequent, example of the realization [w] + [r] combination resulting from the specific orthographic conventions in English.

**Figure 5. Mispronunciation of the wr- Letter Combination as [w] + [r] (words from left to right: writer ['raɪtə] и wrote [rɔʊt])**



Although this is not directly related to the production of the bilabial approximant, it seems worth mentioning because it was encountered in the wordlist reading task and the interview. Although the [w] is not supposed to be pronounced, the speaker articulates both sounds, leading to a noticeable foreign accent. The given combination yields two realizations, either with the Serbian labiodental sonorant, or with the intense rounding of the lips in front of [r] resembling the vowel [ʊ]. The spectrogram on the right is particularly interesting because it exhibits an interesting realization of the word-medial plosive, articulated as a plosive first, and then as a tapped allophone. The initial sound preserves voicing which is evident by the presence of the voice bar.

## 6. CONCLUSION

The aim of the present study was to investigate instances of differential substitution of the /v/-/w contrast in the interphonology of Serbian and English. Both perception and production aspects were considered. In general, investigated sample of participants demonstrated high level of accuracy in the phonemic discrimination task. The substitutional variants were regarded in relation to the phonetic environment in which they occur and the formality of speech style, represented by different task types. Although it is well known that Serbian EFL learners mostly substitute interdental fricatives and a velar nasal, with alveolar plosives and the the combination of alveolar nasal + velar plosive respectively, [w] and [v] show interchangeable substitution, with the addition of the Serbian variant of the labiodental sonorant [ʋ] for [w]. This leads us to the conclusion that both new and similar sounds are subject to substitution or hypercorrection. Labiodental fricative substitution proved to be more frequent and dependent on the phonetic context and task formality. As far as the labial-velar approximant is concerned, we did not encounter a

high level of substitution, nor did we observe an increase in the number with a decrease in the for particular participants who had difficulties in the perception and adaptation of the articulators in the production.

Statistical significant difference was established only for the substitution of the labiodental fricative both for the phonetic context and the task type. However, the results underlined the variation of substitution in relation to the formality of the task type for /w/ also, though not statistically significant. The typical scenario for the increase in formality of the task was the resulting increase in the number of substitutes.

The variability and inconsistency of differential substitution as a strategy in overcoming difficulties in the pronunciation of both new and similar sounds may stem from the phonetic specificities of the target sounds and the relationship with the learners' mother tongue. The possible limitations of the study involve the design of the phonemic discrimination task, which might have pointed to different results had it been performed using an appropriate computer software. Future studies may concentrate on ways of helping learners cope with the perception and production difficulties by introducing various audio-visual methods and available online tools for pronunciation practice. Furthermore, the study has important pedagogical implications by potentially raising awareness of the teachers regarding the significance of the very process of acquisition and not the final outcome, with all the possible variations and complexities that may accompany it.

## REFERENCES

- Ankerstein, Morschett 2013: Ankerstein, C. A., Morschett, R. Do you hear what I hear? A comparison of phoneme perception in native and Saarlandian German non-native speakers of English, *Saarland Working Papers in Linguistics*, 4, 1-8.
- Aoyama et al. 2004: Aoyama K., Flege J. E., Guion S. G., Akahane-Yamada R., Yamada T. Perceived phonetic dissimilarity and L2 speech learning: the case of Japanese /r/ and English /l/ and /r/, *Journal of Phonetics*, 32, 233-250. [https://doi.org/10.1016/S0095-4470\(03\)00036-6](https://doi.org/10.1016/S0095-4470(03)00036-6)
- Aoyama et al. 2023: Aoyama, K., Hong, L., Flege, J.E., Akahane-Yamada, R., Yamada, T. Relationships Between Acoustic Characteristics and Intelligibility Scores: A Reanalysis of Japanese Speakers' Productions of American English Liquids, *Language and Speech*, 66(1), (pre-print online first). <https://doi.org/10.1177/00238309221140910>
- Baart 2010: Baart, J.L.G. *A Field Manual of Acoustic Phonetics*, US: SIL International.
- Belić 1972: Belić, A. *Osnovi istorije srpskohrvatskog jezika I – Fonetika*, Beograd: Naučna knjiga.
- Best 1994: Best, C. The emergence of native-language phonological influences in infants: A perceptual assimilation model, in: J. Goodman and H. Nusbaum (Eds.), *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words* (pp. 167-224), Cambridge, MA: The MIT Press.

- Best 1995: Best, C.T. A direct realist view of cross-language speech perception, in: Strange, W. (ed.), *Speech Perception and Linguistic Experience. Theoretical and Methodological Issues* (pp. 171-203), Baltimore: York Press.
- Best et al. 2001: Best, C. T., McRoberts, G. W., Goodell, E. Discrimination of Non-Native Consonant Contrasts Varying in Perceptual Assimilation to the Listener's Native Phonological System. *Journal of the Acoustical Society of America*, 109 (2), 775-794.
- Best, Tyler 2007: Best, C., Tyler, M. Non-native and second-language speech perception: Commonalities and complementarities, in: M.J. Munro and O.-S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production* (pp. 13-34), Amsterdam; Philadelphia: John Benjamins.
- Bettagere, Fucci 1999: Bettagere, R., Fucci, D. Magnitude-estimation scaling of computerized (digitized) speech under different listening conditions., *Perceptual and Motor Skills*, 88(3/2), 1363-1378.
- Boersma, Weenink 2016: Boersma, B., Weenink, D. Praat: Doing phonetics by computer (Version 6.02.3) [Computer program]. <http://www.praat.org/>. > 21. 6. 2016.
- Brannen 2002: Brannen, K. The role of perception in differential substitution. *Canadian Journal of Linguistics*, 47(1), 1-46.
- Brown 2000: Brown, C. The interrelation between speech perception and phonological acquisition from infant to adult, in: J. Archibald, editor, *Second Language Acquisition and Linguistic Theory* (pp. 4-63), Malden, MA: Blackwell.
- Bundgaard-Nielsen et al. 2011: Bundgaard-Nielsen R. L., Best C. T., Tyler M. D. Vocabulary size matters: the assimilation of second-language Australian English vowels to first-language Japanese vowel categories, *Applied Psycholinguistics*, 32, 51-67. <https://doi.org/10.1017/S0142716410000287>
- Čubrović 2011: Čubrović, B. Voice Onset Time in Serbian and Serbian English, *English Language Overseas Perspectives and Enquiries*, VIII, Spring, 9-18.
- Čubrović 2013: Čubrović, B. Initial and medial voiceless stops in Serbian: an acoustic analysis, *Nasleđe*, 26, 1-19.
- Cutler et al. 2004: Cutler, A., A. Weber, R. Smits, N. Cooper. Patterns of English phoneme confusions by native and non-native listeners, *Journal of the Acoustical Society of America*, 116, 3668-3678.
- De Jong et al. 2009: de Jong, K., Hao, Y.C., Park, H. Evidence for featural units in the acquisition of speech production skills: Linguistic structure in foreign accent, *Journal of Phonetics*, 37, 357-373.
- Díaz-Campos 2004: Díaz-Campos, M. Context of learning in the acquisition of Spanish second language phonology, *Studies in Second Language Acquisition*, 26, 249-274.
- Ercan 2018: Ercan, H. Pronunciation problems of Turkish EFL learners in Northern Cyprus. *International Online Journal of Education and Teaching (IOJET)*, 5(4), 877-893.
- Field 2009: Field, A. *Discovering statistics using SPSS*, UK: SAGE Publications Ltd.
- Flege 1987: Flege, J. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification, *Journal of Phonetics*, 15, 47-65.
- Flege 1995: Flege, J. E. Second language speech learning theory, findings and problems, in: W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233-277), Baltimore, MD: York Press.

- Flege et al. 2003: Flege J., Schirru C., MacKay I. Interaction between the native and second language phonetic subsystems, *Speech Communication*, 40, 467-491. [https://doi.org/10.1016/S0167-6393\(02\)00128-0](https://doi.org/10.1016/S0167-6393(02)00128-0)
- Flege, Bohn 2021: Flege, J. E., Bohn, O. S. "The revised speech learning model (SLM-r)," in R. Wayland (Ed.) *Second Language Speech Learning: Theoretical and Empirical Progress*, (pp. 3–83). Cambridge: Cambridge University Press.
- Gimson 1978: Gimson, A.C. *An Introduction to the Pronunciation of English*: ELBS.
- Guion et al. 2000: Guion, S., Flege, J., Akahane-Yamada, R., Pruitt, J.C. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants, *Journal of the Acoustical Society of America*, 107, 2711-2725.
- Hansen 2001: Hansen, J. Linguistic constraints on the acquisition of English syllable codas by native speakers of Mandarin Chinese, *Applied Linguistics*, 22, 338-365.
- Harrington 2010: Harrington, J. *Phonetic analysis of speech corpora*, Oxford: Wiley-Blackwell.
- Holstein, Gubrium 2004: Holstein, J.A., Gubrium, J.F. Active interviewing, in: Silverman D. (ed) *Qualitative Research: Theory, Method and Practice* (pp. 140-161), London: Sage.
- Ingram 1991: Ingram, D. Toward a theory of phonological acquisition, in: Miller J, editor. *Research on child language disorders: A decade of progress* (pp. 55-72), Austin, Tx: Pro-Ed.
- Iverson et al. 2008: Iverson, P. Ekanayake, D., Hamann, S., Sennema, A., Evans, B. G. Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers, *Journal of Experimental Psychology: Human Perception & Performance*, 34, 1305-1316.
- Jakovljević 2011: Jakovljević, B. Auditory Perception of Serbian and English Voiceless Stops by Serbian Speakers and Interference, *Zbornik za jezike i književnosti Filozofskog fakulteta u Novom Sadu I*, 47-55.
- Jerotijević 2014: Jerotijević, D. Uticaj fonetskog okruženja na usvajanje interdentalnih frikativa u englesko-srpskoj međujezičkoj fonologiji, *Zbornik sa V skupa mladih filologa Srbije, održanog 30. marta 2013., Savremena proučavanja jezika i književnosti, knjiga 1*, Filološko-umetnički fakultet, Univerzitet u Kragujevcu, 395-406.
- Jones 1979: Jones, D. *An outline of English phonetics*, Cambridge: Cambridge University Press.
- Jones 2012: Jones, D. *Cambridge English Pronouncing Dictionary*, 18 Ed. (PB +CD-ROM), Cambridge University Press.
- Kent, Read 2002: Kent, R. D., Read, C. *The acoustic analysis of speech*, London - San Diego: Whurr Publishers - Singular Publishing Group.
- Kim 2023: Kim, J.Y. Spanish-English Crosslinguistic Influence on Heritage Bilinguals' Production of Uptalk, *Languages*, 8(1):22. <https://doi.org/10.3390/languages8010022>
- Labov 1984: Labov, W. Field methods of the Project on Linguistic Change and Variation: in: J. Baugh & J. Sherzer (eds.), *Language in Use*. Englewood Cliffs: Prentice Hall, (pp. 43-70), Intensity. *GURT*, 84.
- Ladefoged 2003: Ladefoged, P. *Phonetic Data Analysis: An Introduction to Phonetic Fieldwork and Instrumental Techniques*, Maldan, MA: Blackwell.

- Ladefoged 2006: Ladefoged, P. *A course in phonetics* (5th ed.), Belmont, CA: Thompson Wadsworth.
- Laméris et al. 2023: Laméris, T. J., Li, K.K, Post, B. Phonetic and Phono-Lexical Accuracy of Non-Native Tone Production by English-L1 and Mandarin-L1 Speakers, *Language and Speech*, 66(1), (pre-print online first). <https://doi.org/10.1177/00238309221143719>
- Lee 2006: Lee, J. Typology of interdental fricatives with reference to loanword adaptation, *Studies in Phonetics, Phonology and Morphology*, 12(1), 127-148.
- Lisker 1995: Lisker, L. English /w, j/: Frictionless approximants or vowels out of place?, in: F. Bell Berti & L. Raphael (eds.), *Producing speech: contemporary issues. For Katherine Safford Harris* (pp. 129-142), Woodbury, NY. NY: AIP Press.
- Lombardi 1995: Lombardi, L. *Why Place and Voice are Different: Constraint-specific Alternations in Optimality Theory*, Univeristy of Maryland, College Park.
- Machač, Skarnitzl 2009: Machač, P., Skarnitzl, R. *Principles of Phonetic Segmentation*, Praha: Epoque Publishing House.
- Markham, Hazan 2002: Markham, D., Hazan, V. UCL Speaker database, *Speech, Hearing and Language: Work in progress*, 14. Department of Phonetics and Linguistics, 1-17.
- Miletić 1933: Miletić, B. Izgovor srpskohrvatskih glasova, *Srpski dijalektološki zbornik*, knjiga V, Beograd: Zadužbina Milana Kujundžića, 1-160.
- Paunović 2011: Paunović, T. Sounds Serbian? Acoustic properties of Serbian EFL students' speech, in: Kitis, E., N. Lavidas, N. Topintzi & T. Tsangalidis (Eds.) *Selected Papers from the 19<sup>th</sup> International Symposium on Theoretical and Applied Linguistics*, April 3-5 2009, Thessaloniki: Aristotle University, School of English, Department of Theoretical & Applied Linguistics, 357-369.
- Peterson, Lehiste 1960: Peterson, G. E., Lehiste, I. Duration of Syllable Nuclei in English, *Journal of the Acoustical Society of America*, 32, 693-703.
- Petrović, Gudurić 2010: Petrović, D., Gudurić, S. *Fonologija srpskoga jezika*, Beograd: Institut za srpski jezik SANU.
- Piyamat, Deekawong2021: Piyamat, B., Deekawong, K. Phonological Variations and Problems in English Pronunciation among Thai EFL Learners: A Case Study of Undergraduate Students at Huachiew Chalermprakiet University, *Liberal Arts Review*, 16(1), 70-84.
- Rapley 2004: Rapley, T. Interviews, in: C. Seale, G. Gobo, J.F. Gubrium and D. Silverman (eds) *Qualitative Research Practice*, Sage Publications: London, Thousand Oaks and New Delhi.
- Rau et al. 2009: Rau, V., Chang H., Tarone E. Think or Sink: Chinese Learners' Acquisition of the English Voiceless Interdental Fricative, *Language Learning*, 59(3), 581-621.
- Roach 1990: Roach, P. *English phonetics and phonology - A practical course*, Cambridge: Cambridge University Press.
- Rubach 1993: Rubach, J. *The Lexical Phonology of Slovak*, Oxford, UK: Oxford University Press.
- Saunders et al. 2003: Saunders, M., Lewis, P., Thornhill, A. *Research Methods for Business Students*, Harlow: Pearson Education.
- Silbert et al. 2015: Silbert, N. H., Smith, B. K., Jackson, S. R., Campbell, S. G., Hughes, M. M., Tare, M. Non-native phonemic discrimination, phonological short term memory, and word learning, *Journal of Phonetics*, 50, 99-119.

- Sredović-Trpeski 2013: Sredović-Trpeski, G. S. Usvajanje engleskih dentalnih frikativa /θ/ i /ð/ kod mladih učenika u učenju engleskog jezika kao stranog, *Nasleđe*, 10(26), 99-114.
- Ternström 2008: Ternström, S. Hi-Fi voice: observations on the distribution of energy in the singing voice spectrum above 5kHz, *Journal of the Acoustical Society of America*, 123, 3379.
- Tyler et al. 2014: Tyler M., Best C., Faber A., Levitt A. Perceptual assimilation and discrimination of non-native vowel contrasts, *Phonetica*, 71, 4-21. <https://doi.org/10.1159/000356237>
- Zhou, Rato 2023: Zhou, C., Rato, A. Syllable position effects in the perception of L2 Portuguese /l/ and /t/ by L1 Mandarin learners, *Second Language Research*, 39(2) (pre-print online first). <https://doi.org/10.1177/02676583221137713>
- Zsigri 1998: Zsigri, G. *Parametrical domains and clusters: An OT approach to voice agreement and final devoicing*, Ms., József Attila University, Szeged, Hungary.

Даница М. Јеротијевић Тишма

## “I WENT TO VENICE AND WAS VERY WORRIED.” – ДИФЕРЕНЦИЈАЛНА СУПСТИТУЦИЈА /v/-/w/ КОНТРАСТА У ЕНГЛЕСКО-СРПСКОЈ МЕЂУЈЕЗИЧКОЈ ФОНОЛОГИЈИ

Резиме

Претходни налази до којих су дошла истраживања у области усвајања страног језика потцртали су значај диференцијалне супституције као једне од најчешћих стратегија за превазилажење потешкоћа у изговору енглеских сугласника. Усвајање контраста /v/-/w/ чини се посебно захтевним за српске ученике енглеског као страног језика услед различитих фактора од којих је доминантан језички трансфер и њиме изазвана перцептивна асимилација. Имајући у виду претходно наведено, као и оскудност истраживања у српском научном контексту, наш рад има за циљ да истражи случајеве диференцијалне супституције енглеског лабиоденталног фрикатива и лабиовеларног апроксиманта у односу на фонетски контекст и формалност задатка при испитивању. И перцепција и продукција 72 студента енглеског језика тестирани су уз помоћ различитих инструмената, укључујући тест фонемске дискриминације, читање листе речи и пасуса, као и интервју. Добијени подаци квантитативно су анализирани и поткрепљени спектрограмским илустрацијама изворних и неизворних говорника енглеског језика. Резултати указују на већу учесталост супституције лабиоденталног фрикатива ([w] је доминантни супституент) при чему фонетски контекст и тип задатка играју статистички значајну улогу у диференцијалној супституцији. Случајеви замене лабиовеларног апроксиманта нису били толико чести и карактеристични су за одређену групу ученика (српски лабиодентални сонорант [v] јесте доминантни супституент). Резултати наглашавају варијабилност и сложеност међујезичког система и важност вредновања процеса, а не само коначног резултата усвајања гласова страног језика.

**Кључне речи:** контраст /v/-/w/, диференцијална супституција, међујезичка фонологија, српски, енглески

Примљен: 3. фебруар 2023. године  
Прихваћен: 13. јул 2023. године