811.163.41'342.1:811.111'342.1
811.111'243
https://doi.org/10.18485/sj.2024.29.1.9

DANICA M. JEROTIJEVIĆ TIŠMA*
University of Kragujevac
Faculty of Philology and Arts

Оригинални научни рад
Примљен: 17. 10. 2023.
Прихваћен: 12. 12. 2023.

VARIABLE PRODUCTION OF THE RHOTIC APPROXIMANT IN THE INTERLANGUAGE OF SERBIAN AND ENGLISH

The rhotic approximant exhibits considerable degree of variability in the production of native speakers, let alone speakers of English as a foreign language. Hence, the present study aimed at investigating the variability in the production of the rhotic approximant in the Serbian-English interlanguage phonology. Furthermore, the goal was to investigate acoustic features relevant for the description of the aforementioned variability, including formant frequencies and duration. Independent factors such as phonetic context, gender and proficiency were likewise taken into consideration. To answer the proposed research questions, a total of 28 English-major students participated in the study, recording their careful pronunciation of words containing target tokens. The results indicated three dominant variants of the rhotic approximant: a trilled [r], retroflex [ɹ] and flapped [ɾ]. Gender proved to be a statistically significant predictor of formant frequencies and duration, with female speakers possessing lower formant frequencies and shorter durations. Proficiency was not a statistically significant factor of variation, except between B1 and C1 levels CEFR for F2, whereas the phonetic context determined the variability in F1 and F3. The results complement the ongoing theoretical research in the field of interlanguage phonology and point to important pedagogical implications underlined in the concluding segments of the paper.

Keywords: rhoticity, approximants, phonetics, EFL, interlanguage

* danicajerotijevic@gmail.com

1. BACKGROUND AND RATIONALE

Previous studies dealing both with acoustics and sociolinguistic variation have underscored high levels of variability in the production of the English rhotic approximant (Guenther et al. 1999; Hashi et al. 2003). The variability is evident with regards to different dialects, among speakers and within the same speaker even. Nevertheless, studies within the broader field of second language sound acquisition are conspicuously scarcer, especially in the Serbian EFL context. Namely, investigations done so far discuss the variability in terms of producing postvocalic /r/ or not, thus pointing to the preferred pronunciation model, i.e. Southern British Standard or General American (Grubor, Hinić 2011; Čubrović Bjelaković 2020; Janevska 2022). Thereby the present study aims to contribute to the existing studies by offering an acoustic account on the variability in the production of the rhotic approximant across phonotactically plausible contexts in the interlanguage system of L1 Serbian and L2 English.

Pronunciation has long been a five-minute end-of-class activity and the consequences of almost completely neglecting it in foreign language teaching have been confirmed by numerous studies (Derwing, Munro 2005; Hurtado, Estrada 2010). The lack of effective pronunciation teaching leads to a complete avoidance of pronunciation-related activities or the teaching is reduced only to explaining basic features at the segmental level. Errors in the production of English sounds among Brazilian students are attributed primarily to inadequate pronunciation teaching (De Godoy et al. 2006). The situation is similar with Turkish (Hişmanoğlu 2006), Chinese (Chujo 2012) or Finnish students (Lintunen 2004).

Interlanguage studies have paid special attention to the difficulties students face in acquiring the articulatory and acoustic properties of the sounds of a foreign language and have defined the most important factors that influence the accurate perception and production of the target sounds (Nakayama, Yamagushi 2003; Moyer 2004). External factors include various socio-cultural factors, such as age, speaking style, geographical region, ethnicity or social status of the speaker, and internal include factors such as transfer and factors of universal development, as well as the phonetic environment (Romaine 2003: 410). Negative interference or language transfer remains one of the dominant factors in explaining learner errors, even though the interpretation of its influence has altered throughout the years.

The theoretical models selected in the present paper share the same assumption that perception is guided by the native phonological system, i.e. that non-native speakers will assimilate target phonemes into the native, or more familiar ones. On the one hand, the *Speech Learning Model* (Flege 1995) recognizes acoustic specificities in the speech signal as indispensable units of perception, while on the other hand, the *Perceptual Assimilation Model* (Best 1994) focuses perception on the articulatory movements that produced the speech signal. One of the important

differences is the way the two models view the importance of the phonetic environment in which the target contrast is found within a word. According to the *Speech Learning Model*, the position in the word in which the target phoneme is found is of key importance for perception, because non-native speakers will identify allophones of a foreign language with allophones or sounds in the native language, and the accuracy of perception will directly depend on the acoustic similarity or difference between native and target phonemes in the same position (Flege 1995: 238–239). Perception is thus related to phonetic rather than abstract phonological differences in the sound inventories of the two languages. In predicting the assimilation of non-native sounds, Flege's model assumes equivalence at the level of phonetic categories, i.e. it is based on the comparison of allophonic variations in the same environment in two languages, while the *Perceptual Assimilation Model* includes similarity and difference at both phonetic and phonological levels, with the default consistency of phonemes in different positions (Best 1994). The two models have relatively recently received their modified and updated versions: PAM-L2 (Best, Tyler 2007) and SLM-r (Flege, Bohn 2021) grounded on the findings of the original versions. SLM-r underlines the importance of recognizing the co-evolving of both perception and production. The process of L2 category formation should be regarded as an ongoing process and the evaluations and measurements of the progress should be done taking this into consideration. There is no final product, the very process is what counts as relevant. Each learner is a specific individual with their own abilities, strategies and skills. The findings of PAM-L2 and SLM-r have been confirmed in numerous studies (e.g. Tyler et al. 2014; Laméris et al. 2023; Kim 2023; Zhou, Rato 2023; Aoyama et al. 2023).

2. PRODUCTION OF APPROXIMANTS IN SERBIAN AND ENGLISH

During the articulation of approximants, the articulators are close to each other, but the vocal tract is not narrowed as much as to produce turbulence in the airflow (Ladefoged 2006: 15; Ashby 2011: 62–63). The articulation depends a lot on the following vowels. There are four approximants (semivowels or glides) in English: palatal /j/, postalveolar retroflex /r/, labial-velar /w/ and lateral /l/. Some authors classify the phoneme /r/ as postalveolar (Gimson 1978: 205), and others as an alveolar approximant (Ladefoged 2006: 15). It should be mentioned that in English a given consonant is characterized by secondary articulation, i.e. both labialization and velarization (Odgen 2009: 91).

In many languages and dialects around the world, rhotic sounds are known for their extraordinary phonetic variability; some languages possess only one, while others have more rhotic consonants mostly differing in the manner rather than in

the place of articulation (Ladefoged, Maddieson 1996: 237). The phonological distinction of the duration of the rhotic consonant in languages will contrast the extremely short consonant realized by hitting the tip of the tongue against the alveoli with a multiple vibrant.

Rhotic sounds are usually characterized by a low third formant, which is a well-known feature of the American retroflex approximant [ɹ] (Lindau 1978). The lowered third formant at around 2000 Hz is also specific for Russian /r/ (Fant 1960: 70–75; Kavitskaya 1997: 751–754).

In English, there is one phoneme /r/, which usually has two different realizations, retroflex [ɹ], where the tip of the tongue rises up and bends towards the dorsal part of the oral cavity, and bunched [ɹ], with the tongue tightly contracting and bendsing towards the back of the oral cavity (Zhou et al. 2008: 4471). However, the perceptually relevant acoustic variations between the two pronunciations are subtle. For example, a slightly greater difference was observed between the fourth and fifth formants during retroflex articulation, which actually corresponds to the size of the rear part of the oral cavity, which is larger when pronouncing the retroflex approximant.

By raising the tip of the tongue and quickly hitting the alveolar ridge, the so-called tapped [ɾ] is realized, which appears as an allophonic variation of the alveolar plosives /t/ and /d/, most often, but not exclusively, in the American variant (Kahn 1980: 94). The medial alveolar nasal may also fairly frequently be replaced by a tapped [ɾ] (Ladefoged 2006: 168). There are several important acoustic indicators of the tapped pronunciation of alveolar sounds (Dungan et al. 2007: 3167). Namely, the occlusion occurs between two vowels and the intensity decreases in relation to the surrounding vowels. At the end of the tapped sound, there is usually a high-intensity burst release, and a drop in the value of the fourth formant. Another allophonic variation is flapped [ɽ], realized when the /r/ sound is between two vowels and is almost identical to the tapped variant.

In Serbian, /r/, /l/, /j/ are most often classified as alveolar sonorants, although there are examples of classifying /v/ and /j/ as semivowels, and /l/ and /ʎ/ as lateral sonants (Miletić 1960: 44). According to Simić and Ostojić (1996: 196) /r/ and /l/ are apico-postdental, and /j/ and /ʎ/ are dorso-palatal. Petrović and Gudurić (2010: 166) classify /r/, /l/, /ʎ/ as oral sonorants. Articulations of the phoneme /r/ are significantly different in English and Serbian, because in English the given sound is a retroflex postalveolar, while in Serbian it is an alveolar vibrant. Belić (1972: 54–55) finds that liquid /r/, although it is primarily alveolar, can be formed wherever dental consonants are formed and is accompanied by drumming, i.e. trembling.

Although they are typologically more marked than plosives, approximants are still contextually less marked due to the speaker's efforts to keep the duration of the obstruction short, while still avoiding the devoicing of the plosive (Ohala

1994: 4144–4147). Approximants are phonetically less marked intervocalically, but more marked in initial positions, because the sonority does not increase (Shea, Curtin 2011). Foreign language learners may have difficulty with approximants in initial position, more so than in intervocalic position.

It should be noted that the formants of approximants are not visible as a straight line yet as 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).

When it comes to the duration of the rhotic approximant [ɹ], the spectrograms show great variability given the variety of pronunciation of the mentioned sound and the relatively slow movement of the body of the tongue (Hayward 2000: 166–167). The approximant is characterized by a low value of the third formant, which decreases even more when secondary articulations, such as velarization and labialization, are added. Due to labialization, the third formant is usually below 2000 Hz, and sometimes it drops to 1500 Hz (Ladefoged 2006: 54). Several studies dealt with the acoustic characteristics of approximants in English, usually examining their sonorant and consonant nature measured by appropriate acoustic parameters (Lisker 1957; Lehiste 1962; Espy-Wilson 1992; Chase 2017). The values of the first three formanst for /r/ in English are 300–350 Hz (F1), 1000–1200 Hz (F2) and 1600-1750 Hz (F3).

The /r/ in Serbian is likewise characterized by substantial variability, and the same speaker can pronounce the sound differently depending on an occasion. Its realization is also conditioned by regional stratification (Simić, Ostojić 1996: 182). In the initial position or in the dominantly consonantal environment an epenthetic vowel /ə/ may sometimes be found during the articulation of /r/ (Petrović, Gudurić 2010: 198).

3. METHODOLOGY

3.1. The Aims of the Study and Research Questions

The present paper aims at investigating the acoustic variability of the English rhotic approximant produced by Serbian EFL students, i.e. the variability occurring in the interlanguage phonology of Serbian and English. The formulated aims of the study stem from the observed complexity of the interlanguage production with variations in the pronunciation of /r/ being one sole segment in the overall range of variations of consonant production.

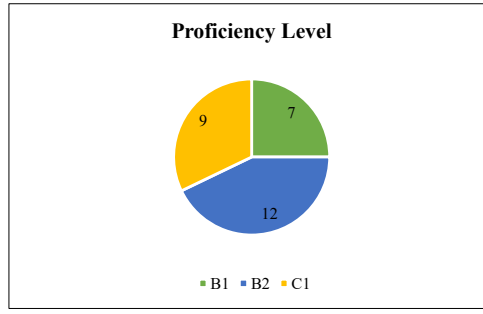
Having the proposed goals in mind, the empirical investigation was based upon the following research questions:

- What are the relevant acoustic parameters for describing the production of the English rhotic approximant in Serbian-English interphonology?
- What are the possible variants of the English rhotic approximant in the production by Serbian EFL learners?
- Are there any specific phonetic environments that trigger variability more than others?
- Does variability in any way depend on factors such as gender or target language proficiency?

3.2. Participants

A total of 28 second-year English-major students (14 male, 14 female, average age 20.71) at the Faculty of Philology and Arts, University of Kragujevac participated in the study. All the participants have signed the written informed consent prior to the recording procedure. Although initially there were more participants in the project, due to the insufficient number of representative recordings by male speakers, the final sample was rounded to 28, to ensure the equal distribution of both genders. All the participants opt for General American variety and the results of diagnostic testing[1] of their proficiency in English can be seen in Graph 1.

Graph 1. *Participants' Overall Proficiency in English*



The participants were at the very beginning of their second year of English studies at the tertiary level and have successfully passed the *English Phonetics* exam.

[1] The test was combined by the author of the paper from the reliable sources available at https://www.learnenglishteam.com/free-english-level-tests-downloadable-pdf/.

3.3. Instruments and Procedure

The primary instrument for eliciting the production of the rhotic approximant from the chosen sample was a formal task of a wordlist with 40 words in total (Appendix 1) containing the target sound in various positions (Yavaş 2002; Liu 2011). Each word was pronounced three times in a row and the participants were advised to pronounce the words as naturally as possible. The second or the third attempt were opted for in the analysis in the majority of cases, since it was in these examples that the speakers sounded more relaxed and relatively spontaneous. Immediately before the actual reading began, the participants were asked a few general questions (e.g. How are you feeling today? What is your favorite hobby? Do you have a pet? etc.) to relax themselves as much as possible. Before finalizing the wordlist for this particular recording project, the participants underwent another recording procedure using the interview technique to elicit spontaneous speech several months earlier. The interview served as a specific type of diagnostics for the relevance and justification of the present research, since it was precisely then that a considerable degree of variability in the production of the English rhotic approximant was noticed. Understandably, we made sure the words in the list were familiar to students.

The focus of the present paper was not on postvocalic /r/ and r-colored vowel as much as it was on the actual approximant realizations in initial and medial positions. However, ten specific contexts were included in the wordlist as well with the [ɚ] immediately following different consonant sounds. The reason behind this decision lies in the fact that appreciable variation was detected in these positions, as well. These examples were observed as particular triggers for pronunciation variations during the interview. Namely, the r-colored vowel is pronounced with similar variants to the approximants in initial and medial positions which is why the measures for the final position were included in the analysis.

The quantitative description of the number of analyzed examples per specific phonetic context can be seen in Table 1.

Table 1. *Target Token Distribution in the Wordlist*

Phonetic Context	Number of Examples in the Wordlist	Total Number of Recorded Items	Number of Analyzed Tokens
[r]/#_V	6	504	168
[r]/V_V	7	588	196
[r]/C_V	17	1428	476
[ɚ]/C_#	10	840	280

The recording was performed using the Olympus VN-8600PC recording device with an inbuilt microphone, with 44.1 kHz sampling height and 16-bit conversion saved in WAV format (Bettagere, Fucci 1999). The average duration of a single recording was 198s. The procedure took place at the beginning of the winter semester 2021/2022. The recorded material was transcribed, annotated and segmented in *Praat*, version 6.3.10 (Boersma, Weenink 2023) by the author of the paper, relying on waveforms and spectrograms in combination with the auditory method (Golafshani 2003). The criteria for segmentation followed the suggestions and instructions of previous studies for approximants (Machač, Skarnitzl 2009: 79, 92). Based on the relevant literature, we selected the description of the acoustic parameters of consonants for analysis together with examples of studies in which they were previously investigated, i.e. the distribution and frequency of formants, as well as duration.

Statistical processing of the production results including independent samples t-test and ANOVA was done in *SPSS*, version 20.0 (Field 2009).

4. RESULTS AND DISCUSSION

The results of the acoustic measurements are presented in Table 2, including formant frequencies and duration in different positions.

Formant frequencies are given in Hz of course and the duration is described in miliseconds. Formant frequencies were measured in the central most stable parts of the sound trying to avoid the coarticulatory effects of the ensuing vowel. However, due to the approximant nature of the analyzed sound, it must be noted that the coarticulatory effects were often difficult to avoid. The mean values and range were considered for initial (I), medial (M) and final (F) positions where applicable. The range was important for it directly points to a considerable degree of variability among individual speakers. The flapped variant has no value in initial and final positions since the variant was observed in intervocalic positions only.

Table 2. *Acoustic Characteristics of the Rhotic Variants*

Phonetic Features		Variants					
		trill [r]		retroflex [ɹ]		flapped [ɾ]	
		Male	Female	Male	Female	Male	Female
F1 (mean) F1 (range)	I	426 382–465	432 375–488	382 324–475	364 300–382	/	/
F1 (mean) F1 (range)	M	401 372–464	412 390–485	362 305–442	365 308–401	452 412–490	442 400–512
F1 (mean) F1 (range)	F	387 348–432	402 372–474	312 285–345	325 305–392	/	/
F2 (mean) F2 (range)	I	1424 1285–1522	1510 1220–1602	1442 1154–1612	1348 1128–1542	/	/
F2 (mean) F2 (range)	M	1445 1195–1946	1382 1055–1508	1452 1224–1618	1414 1258–1645	1520 1114–1685	1485 1225–1685

F2 (mean) **F2 (range)**	**F**	1385 1082–1786	1278 1105–1408	1185 1005–1528	1208 1042–1612	/	/
F3 (mean) **F3 (range)**	**I**	2612 2148–2941	2654 2215–2885	1922 1724–2254	1815 1614–2302	/	/
F3 (mean) **F3 (range)**	**M**	2478 1986–2648	2544 2002–2789	1844 1624–2268	1832 1687–2398	2540 2015–2756	2602 2407–2902
F3 (mean) **F3 (range)**	**F**	2008 1716–2375	1988 1785–2412	1825 1788–2365	1794 1655–2556	/	/
Duration (mean) **Duration (range)**	**I**	68 52–98	74 45–90	62 39–88	57 41–92	/	/
Duration (mean) **Duration (range)**	**M**	62 46–102	66 48–95	50 35–82	52 36–88	8 6–12	10 7–14
Duration (mean) **Duration (range)**	**F**	85 68–104	92 72–112	76 59–98	80 52–102	/	/

Generally, by comparing the results of the present study with the ones found in studies with native speakers of English (Espy-Wilson 1992; Chase 2017), higher values can be observed for all three formants. This may be explained by a particular phonetic environment on the one hand, but also by the specific nature of the interlanguage productions, on the other hand. F1 values exceed the native speakers' frequency by 100 Hz in certain cases, while the differences in F2 rise up to 400 Hz in some cases. The third formant (F3) is particularly interesting because, even though it should be lowered due to retroflex articulations, for some speakers goes above 2800 Hz. The participants are undergoing the process of learning and acquisition during which the phonetic specificities of the two languages interact and mix. Their system of sound perception must be incredibly adept to recognize the subtle phonetic nuances of the target sounds simultaneously being hindered by numerous external and internal factors. The range values likewise point to the individual differences in sound articulation.

When it come to the duration, we must, first and foremost, comment on one particular issue important for future studies. Namely, due to the heavy coarticulatory effects and sounds overlapping, it was extremely demanding to decide on the actual sound boundary necessary to determine the duration. However, the measures were taken carefully ensuring minimal and negligible sound overlaps. The rhotic variants are pronounced the longest in final positions, which was expected in a wordlist as a representative of formal and careful speech. Understandably enough, the sounds are the shortest in medial positions. The flaps are the shortest, again, as expected, since they assume one very brief tap of the tongue tip against the alveolar ridge. There is an obvious consistency in the duration differences depending on speakers' gender, with female speakers' mean values pointing to shorter articulations.

The three variants of the rhotic approximant produced by the chosen sample of participants are the trill and vibrant [r] resembling the mother tongue pronunciation, the native-like retroflex [] testifying to the formation of the new phonetic category in the interlanguage and an allophonic variant from L2, flapped intervocalic [ɾ], which was not as frequent in the corpus. However, the existence of the variant may indicate the perceptual confusion in learners and underscore that the process of acquisition is in progress.

The percentage counts for the appearance of different variants depending on the position in the word are presented in Table 3.

Table 3. *Distribution of Variants Based on the Phonetic Environment*

Phonetic Context	Variants		
	trill [r]	retroflex [ɹ]	flapped [ɾ]
[r]/#_V	36.9%	63.1%	/
[r]/V_V	9.69%	65.82%	24.49%
[r]/C_V	58.82%	41.18%	/
[ɚ]/C_#	32.86%	67.14%	/

The majority of the sample opts for the trill and retroflex variants, except in certain intervocalic positions. It seems interesting to note that the post-consonantal environment triggers the trill variant to a greater percentage than the retroflex articulation, even though the latter is dominant in initial, intervocalic and final positions. The consonants that trigger trill pronunciations, judging by the particular wordlist chosen in the present paper, are voiceless fricatives, mostly interdental and postalveolar, and bilabial plosives.

In order to investigate the mutual effects of articulations and certain independent factors proposed in the research questions, the data were analyzed using the corresponding statistical analyses the results of which can be found in Table 4.

Table 4. *Results of Statistical Analyses*

Factor	Dependent Variables				
	F1	F2	F3	Duration	Variants
Context	F=10.800 p=0.001	F=0.568 p=0.0574	F=10.849 p=0.001	F=1.542 p=0.227	F=2.852 p=0.051
Gender	t=23.509 p=0.001	t=25.965 p=0.001	t=23.869 p=0.005	t=4.680 p=0.001	t=1.075 p=0.292
Proficiency	F=0.577 p=0.569	F=5.100 p=0.014	F=1.465 p=0.250	F=2.058 p=0.149	F=1.274 p=0.297

Judging by the results of ANOVA, there is a statistically significant difference between the values of the first and third formants in the investigated phonetic contexts ($p<0.05$). Tukey's post-hoc multiple comparisons calculations demonstrated the statistically significant difference between the initial and medial (post-consonantal more precisely) positions for F1, and intervocalic and post-consonantal medial

positions for F3. Duration and the distribution of variants did not show statistically significant difference depending on the phonetic context. This points to the conclusion that the pronunciation of the rhotic approximant displays considerable inconsistency and is highly dependent on individual differences.

Gender proved to be a statistically significant factor in determining the formant frequency variations and duration. Namely, female speakers generally show lower values of formant frequencies and shorter durations, which is somewhat unexpected and may depend upon the very sample selected for this particular study. No statistically significant difference was observed for the distribution of variants depending on gender, which means that both male and female speakers in the sample use different variants without any appreciable consistency.

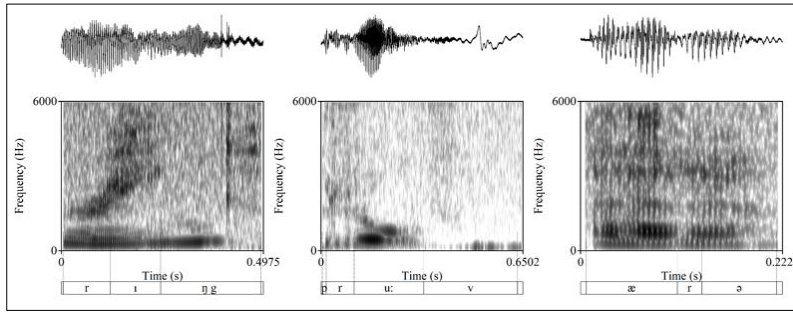
A single speaker may opt for different variants on various occasions. The triggers may not only stem from the phonetic context, but there may be other factors, as well.

Proficiency, as an independent variable, proved to be statistically significant for F2 values, and the post-hoc comparisons showed the significant difference between B1 and C1 CEFR, as examples of the lowest and highest analyzed levels. No further statistically significant difference was detected pertaining to different proficiency levels, which seems interesting because it indicates that pronunciation variations may not necessarily accompany the advancement in the overall L2 proficiency.

To illustrate the degree of variability in the production of the rhotic approximant, the following figures show waveforms and spectrograms of the articulations we found relevant enough for the support of quantitative measurements.

Figure 1 illustrates the realizations of the three possible variants found in the corpus (from left to right: trill, retroflex and flapped). *Praat* textgrids do not support all the phonetic symbols, thus the sounds are all marked as [r] in the spectrograms.

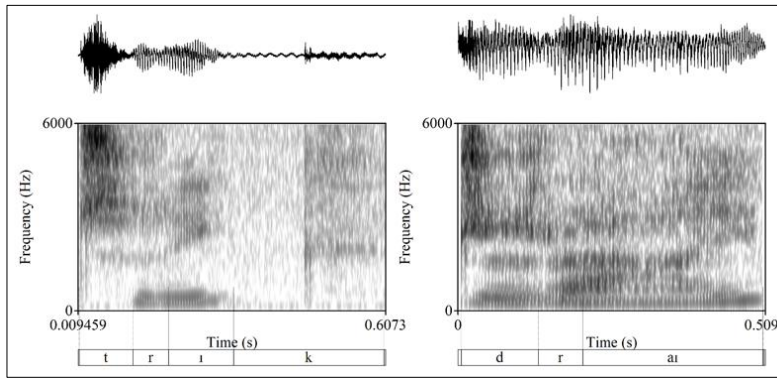
Figure 1. *Three Variants of the Rhotic Approximant in Initial and Medial Positions*



Going from left to right, in the first spectrogram we notice an upward curve in the third formant transitioning into the following high front monophthong, yet the third formant is not very close to the second one. A more pronounced lowering of the third formant is expected for the rhotic approximant here. The strikes of the vibrant are clearly visible when the image is zoomed in more closely. The first spectrogram likewise shows the pronunciation of the velar nasal as a combination of a nasal and a voiced velar plosive. The spectrogram in the middle displays a retroflex articulation and F2 and F3 are closer together. The particular speaker is at C1 CEFR and her pronunciation is fairly native-like. The third spectrogram displays the flapped variant and the visible interval is very short.

Postalveolar sequences represent a specific challenge for Serbian EFL learners, and the rhotic articulations following the retracted /t/ and /d/ articulations are shown in Figure 2.

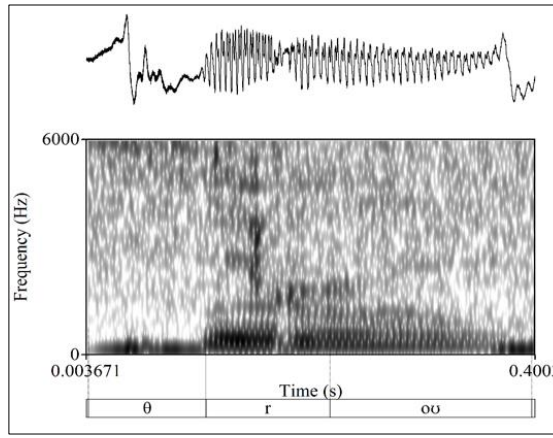
Figure 2. *Trill Variant in Postalveolar Sequences*



The spectrograms show intense frication in the first sound of both sequences, and the sound is followed by a trill pronunciation which is evident by F2 and F3 being quite apart. Periodic strikes of the vibrant are likewise evident in the illustrations. Although a combination of an affricate-like articulation in place of retracted alveolar plosives followed by a vibrant seems demanding, the speakers in the spectrogram opt for it probably due to the closeness of articulators.

In post-consonantal positions, particularly when the rhotic approximant follows a fricative, some speakers in our sample consistently choose the trill articulation. One such example is shown in Figure 3.

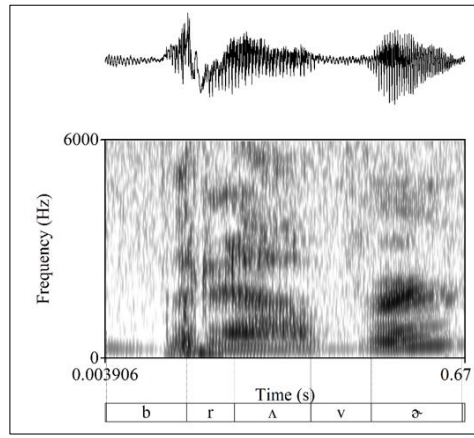
Figure 3. *A Trill Variant Following the Voiceless Interdental Fricative*



The spectrogram clearly displays the periodic strikes of the vibrant sounds which is so strong that it transfers the voicing on the preceding interdental fricative, which is evident by the voice bar. The interdental is likewise heavily fricativized, almost resembling a sibilant.

It was not rarely found in the corpus that the same speaker uses different pronunciation option within the same word. This again points to the complexity and dynamicity of the interlanguage system. The spectrogram showing a post-consonantal trill and a post-vocalic retroflex can be seen in Figure 4.

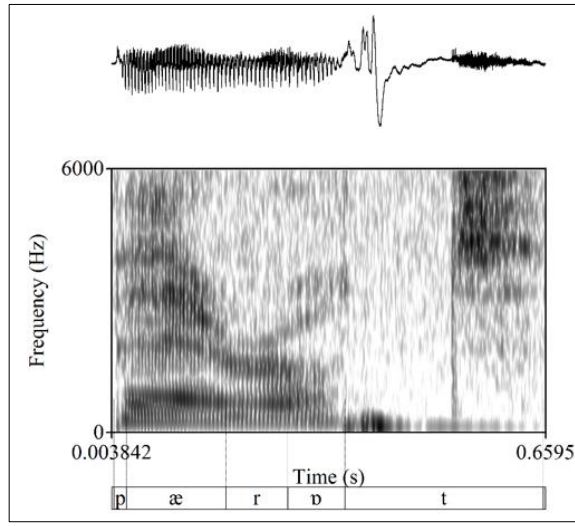
Figure 4. *Different Variants within the Same Word*



Immediately after a bilabial voiced plosive the speaker pronounces the vibrant [r] sound and the tongue strikes are obvious. The speaker likewise opts for [v] as a substitute for the voiced interdental fricative ending in a retroflex articulation of the r-colored vowel. The entire sequence preserves voicing throughout.

Finally, Figure 5 displays an interesting situation where the speaker pronounces the rhotic approximant as a retroflex, yet inserts a rounded vowel instead of the [ə].

Figure 5. *Vowel Change after the Retroflex Pronunciation*



The situation displayed in the previous image may either be explained by the speaker trying to make the pronunciation easier or by the specific orthographic conventions in English. Namely, the particular word is spelled with an ”o” grapheme, which is why, due to the sound-to-grapheme correspondence in Serbian, the L2 learner may have been following the spelling conventions of the native language as an alternative explanation to trying to make the retroflex pronunciation easier by inserting an epenthetic vowel.

CONCLUSION

The present research aimed at shedding light on the importance of investigating the variability of the rhotic approximant in the interlanguage system. Just as this particular sound displays variability in English, the situation is similar in the interlanguage, as well. The chosen sample of participants which included English-major students from B1 to C1 CEFR levels of proficiency showed an inconsistent and

variable production of the rhotic approximant in different phonetic environments. The three observed variants were a highly vibrant trill pronunciation [r] similar to the one in L1, the retroflex articulation [ɹ] similar to the one in L2 and the flapped [ɾ] variant present in L2 as an allophonic variation. The values of the obtained formant frequencies were notably higher than the ones suggested in the literature for native speakers (e.g. Chase 2017).

Gender proved to be a statistically significant factor in predicting the variation in formant frequencies and duration, with female speakers showing the tendency towards lower frequencies and shorter durations of the target sound. Phonetic context was not statistically significant for all formant frequencies yet merely for F1 and F3 in relation to the initial vs. post-consonantal position and intervocalic and post-consonantal position. The auditory analysis indicated that voiceless fricatives trigger trill pronunciation more often than other consonants. Proficiency did not count as a significant predictor of variation, except for F2 values between the lowest B1 and highest C1 investigated levels.

The results underscore the complexity of the interaction of L1 and L2 phonetic features in the interlanguage system. Learners are in the process of category formation and the process is characterized by a substantial degree of variability in the production (Flege, Bohn 2021). This variability should be regarded as favorable and acceptable since it indicates the learning in progress, and it is the very process that should be appreciated rather than the expected final product. Having the previously stated in mind, the results of the present study contribute to the explanations of interlanguage complexity and theories of acquisition in the context of Serbian-English interphonology. Furthermore, the results point to the necessity of reconsidering the existing teaching practice that focuses on individual sounds produced, more often than not, regarding them as correct or incorrect.

The possible limitation of the paper lies in the choice of the instrument, since the list of words should perhaps have been more extensive and should have included various phonetic contexts and more examples. Future research may consider the perception of the rhotic approximant and also explore other external factors shaping EFL learners' articulation.

REFERENCES

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

Ashby 2011: Ashby, P. *Understanding Phonetics (Understanding Language)*. New York: Routledge.

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, 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* Vol. 88(3, Pt 2). 1363–1378.

Boersma, Weenink 2023: Boersma, B., Weenink, D. Praat: Doing phonetics by computer (Version 6.3.10) [Computer program]. http://www.praat.org/. > 29. 5. 2023.

Chase 2017: Chase, B. P. *An Acoustical Analysis of the American English /l, r/ Contrast as Produced by Adult Japanese Learners of English Incorporating Word Position and Task Type*. Unpublished Master's Thesis. USA: Brigham Young University.

Chujo 2012: Chujo, J. The necessity of systematic English phonetic education at the tertiary level in Japanese education. In A. Stewart & N. Sonda (eds.), *JALT2011 Conference Proceedings*, Tokyo, JALT.

Čubrović, Bjelaković 2020: Čubrović, B., Bjelaković, A. Pronunciation Model Selection, or Do You Speak American?. *BELLS90 Proceedings*, 1, 139–151.

De Godoy et al. 2006: De Godoy, B., Gontow, S., Marcelino, C. *English Pronunciation for Brazilians: The sounds of American English*. São Paulo, SP: Disal Editora.

Derwing, Munro 2005: Derwing, T. M., Munro, M. J. Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly,* 39(3), 379–397.

Dungan et al. 2007: Dungan, M., Morian, K., Tucker, B. V., Warner, N. Fourth formant dip as a correlate of American English flaps. *Journal of the Acoustical Society of America*, 121, 3167.

Espy-Wilson 1992: Espy-Wilson, C. Acoustic measures for linguistic features distinguishing the semivowels /w j r l/ in American English. *Journal of the Acoustical Society of America*, 92 (2), 736–757.

Fant 1960: Fant, G. *Acoustic Theory of Speech Production,* Mouton: The Hague.

Field 2009: Field, A. *Discovering statistics using SPSS.* UK: SAGE Publications Ltd.

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, 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.

Golafshani 2003: Golafshani*,* N. Understanding Reliability and Validity in Qualitative Research. *The Qualitative Report*, 8(4), 597–606.

Grubor, Hinić 2011: Grubor, J., Hinić, D. Pronunciation Tendencies towards British and American English in EFL Teachers in Serbia: Are Teachers Immune to Global Influences?. *Nasleđe*, 17, 299–310.

Guenther et al 1999: Guenther, F. H., Espy-Wilson, C. Y., Boyce, S. E., Matthies, M. L., Zandipour, M. and Perkell, J. S. Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America*, 105(5), 2854–2865.

Hashi et al. 2003: Hashi, M., Honda, K., Westbury, J. Time-varying acoustic and articulatory characteristics of American English [r]: a cross-speaker study. *Journal of Phonetics*, 31, 3–22.

Hayward 2000: Hayward, K. *Experimental Phonetics.* New York: Longman.

Hurtado, Estrada 2010: Hurtado, L. M., Estrada, C. Factors influencing the second language acquisition of Spanish vibrants. *The Modern Language Journal*, 94, 74–86.

Janevska 2022: Janevska, M. Produkcija postvokalskog /r/ kod srpskih studenata anglistike. *Philologia Mediana*, 14(14), 429–448.

Kahn 1980: Kahn, D. *Syllable-based generalizations in English phonology*. New York: Garland.

Kavitskaya 1997: Kavitskaya, D. Aerodynamic constraints on the production of palatalized trills: The case of the Slavic trilled [r], in: *Proceedings from the 5th European Conference on Speech Communication and Technology*, 2, 751–754.

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

Ladefoged 2006: Ladefoged, P. *A course in phonetics* (5th ed.). Belmont, CA: Thompsom Wadsworth.

Ladefoged, Maddieson 1996: Ladefoged, P., Maddieson, I. *The Sounds of the World's Languages*. Cambridge, MA, and Oxford: Blackwell Publishing.

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

Lehiste 1962: Lehiste, I. Acoustic studies of boundary signals. *Proceedings of the Fourth International Congress of Phonetic Sciences*. Hague: Mouton & Co., 178–187.

Lindau 1978: Lindau, M. Vowel features. *Language*, 54, 541–563.

Lintunen 2004: Lintunen, P. *Pronunciation and Phonemic Transcription: A study of advanced Finnish learners of English.* Turku: University of Turku.

Lisker 1957: Lisker, L. Some cues to the voiced-voiceless distinction among the intervocalic stops in English. *Haskins Laboratories 24th Quarterly Progress report*. Appendix 5.

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.

Liu 2011: Liu, Q. Factors Influencing Pronunciation Accuracy: L1 Negative Transfer, Task Variables and Individual Aptitude. *English Language Teaching*, 4(4), 115–120.

Machač, Skarnitzl 2009: Machač, P., Skarnitzl, R. *Principles of Phonetic Segmentation*. Praha: Epocha Publishing House.

Moyer 2004: Moyer, A. *Age, accent and experience in second language acquisition: An integrated approach to critical period inquiry.* Clevedon: Multilingual Matters.

Nakayama, Yamagushi 2003: Nakayama, K., Yamagushi, K. The confusion of alveolar and palato-alveolar fricatives by Japanese learners of English. *Bulletin of the Yamagata University, Cultural Science*, 15(2), 207–220.

Odgen 2009: Ogden, R. *An introduction to English Phonetics*. (Edinburgh Textbooks on the English Language). Edinburgh University Press.

Ohala 1994: Ohala, J. J. Speech aerodynamics, in: R. E. Asher and J. M. Y. Simpson (eds), *The Encyclopedia of Language and Linguistics*. (pp. 4144–4148). Oxford: Pergamon.

Petrović, Gudurić 2010: Petrović, D., Gudurić, S. *Fonologija srpskoga jezika*. Beograd: Institut za srpski jezik SANU.

Romaine 2003: Romaine, S. Variation, in: C. Doughty and M. Long (eds), *The handbook of second language acquisition*. (pp. 409–435). Malden, MA: Blackwell.

Shea, Curtin 2011: Shea, C. E., Curtin, S. Experience, representations and the production of second language allophones. *Second Language Research*, 27(2), 229–250.

Simić, Ostojić 1996: Simić, R., Ostojić, B. *Osnovi fonologije srpskog književnog jezika*. Beograd: Univerzitet u Beogradu.

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

Yavas 2002: Yavas, M. VOT patterns in bilingual phonological development, in: Themes in Clinical Linguistics Windsor, F., Kelley, L & N. Hewlett (Eds.), Lawrence Erlbaum Assoc., Inc. (B)

Zhou et al. 2008: Zhou, X., Espy-Wilson, C. Y., Tiede, M., Boyce, S., Holland, C., Choe, A. An MRI-based articulatory and acoustic study of 'retroflex' and 'bunched' American English /r/ sounds, *Journal of the Acoustical Society of America*, 123, 4466–4481.

Zhou, Rato 2023: Zhou, C., Rato, A. Syllable position effects in the perception of L2 Portuguese /l/ and /ɾ/ by L1 Mandarin learners, *Second Language Research*, 39(2) (pre-print online first). https://doi.org/10.1177/02676583221137713

ВАРИЈАБИЛНОСТ У ПРОДУКЦИЈИ РОТИЧКОГ АПРОКСИМАНТА У СРПСКО-ЕНГЛЕСКОМ МЕЂУЈЕЗИКУ

Р е з и м е

Продукција ротичког апроксиманта и у реализацији код изворних говорника показује значајан степен варијабилности, тако да је више него очекивано да ће ситуација бити слична код говорника енглеског као страног језика. Стога је примарни циљ овог рада истражити варијабилност у изговору ротичког апроксиманта у српско-енглеској међујезичкој фонологији. Такође, циљ је био да се истраже акустичке карактеристике релевантне за опис поменуте варијабилности, укључујући фреквенције прва три форманта и трајање. Независни фактори као што су фонетски контекст, пол и ниво постигнућа такође су узети у обзир при анализи. Како бисмо одговорили на предложена истраживачка питања, укупно 28 студената енглеског језика учествовало је у истраживању, а њихов пажљив изговор речи које садрже циљне гласове снимљен је и анализиран у софтверу за акустичку анализу говора. Резултати указују на три доминантне варијанте ротичког апроксиманта у међујезику: вибрант, ретрофлексни изговор и брзи удар, тзв. пљескави изговор. Пол се испоставио као статистички значајан предиктор разлика у вредностима формантских фреквенција и трајања. Ниво постигнућа није статистички значајан фактор варијације, осим између B1 и C1 нивоа Заједничког европског оквира за живе језике за други формант, док се уз помоћ фонетског контекста могла предвидети варијабилност првог и трећег форманта. Резултати нашег рада употпуњују текућа теоријска истраживања у области међујезичке фонологије и указују на важне педагошке импликације потцртане у завршним сегментима рада.

Кључне речи: ротичност, апроксиманти, фонетика, енглески као страни језик, међујезик.

Даница М. Јеротијевић Тишма

Appendix 1: Wordlist

Rest	Mirror	Dry	Brother
Rope	Parrot	Drive	Stabber
Rust	Error	Trick	Eager
Ruby	Borrow	Train	Razor
Wrist	Crush	Bring	Sniffer
Wrong	Crack	Throw	Greater
Array	Agree	Throat	Pleasure
Sparrow	Prize	Frame	Paler
Berry	Strut	Shrink	Slipper
Prove	String	Froze	Flasher