Original scientific article
811.111'243'342.2
doi: 10.46793/NasKg2148.317JT

Danica M. Jerotijević Tišma[1]
University of Kragujevac
Faculty of Philology and Arts
English Department

CAN AUDIO-VISUAL TRAINING EQUALLY AFFECT PHONEMIC AND PHONETIC CONTRASTS? AN EXAMPLE OF L2 FRICATIVE PRODUCTION

The paper explores the effect of audio-visual perceptual training on Serbian EFL learners' production of novel phonemic and phonetic contrasts in L2, specifically focused on fricatives. Hence, the paper aims at discovering whether audio-visual training has equal effects at phonemic and phonetic levels, and also, whether the effect is the same at two different age/proficiency levels, 6th grade primary and 4th grade secondary school. In order to explore the phonemic level we concentrated on interdental fricatives, and for the phonetic level differences sibilant contrasts were included, following the predictions of the *Perceptual Assimilation Model* (Best 1994) and *Speech Learning Model* (Flege 1995). The testing for relevant acoustic information was performed prior to and immediately following the experimental period, when all the participants were recorded pronouncing a prepared sentence list containing target sounds. It consisted of measuring spectral moments, frication duration and comparison of spectrograms. The results of the audio-visual phonetic training proved especially beneficial for phonemic contrasts, i.e. interdental fricatives for both levels of age/proficiency, while sibilant contrasts showed insignificant progress. The age/proficiency level did not appear to be a significant predictor of the effect of audio-visual training. Along with the empirical results, the paper likewise presents pedagogical implications important for pronunciation teaching and highlights the significance of phonetic training in the Serbian EFL context in particular.

Keywords: perceptual training, fricatives, EFL, L2 sound production, interlanguage phonology

1. INTRODUCTION

The dynamicity of an interlanguage system is its greatest advantage for it can alter the way a learner processes L1 and L2 sounds and the relationship among them (Major 2001; Best, Tyler 2007). Some scholars have made a claim that the less experience one has with the native language, the less it is expected to interfere with target language learning (Best 1994), the latter being the view that favours younger learners' capacities over adults'. This of course is not a new idea in the history of SLA, especially if we take 'critical period' research into account (Lenneberg 1967). However, there are authors who believe adult

1 danica.tisma@filum.kg.ac.rs

learners have advantages of their own, such as maturation of cognitive ability that can benefit their performance in L2. Furthermore, empirical findings have proved that brain neuroplasticity can be preserved throughout the whole life span (Flege 1995).

After the communicative approaches emphasizing fluency over accuracy (Pennycook 1998), and the appearance of the idea of English as International Language (Jenkins 2000), speakers have seemingly become more tolerant to pronunciation errors, Moreover, the ultimate attainment of target-like pronunciation even started to be regarded as the misleading goal in learning (Derwing, Munro 2005). However, students on their part have continued to aspire to native-like production, still qualifying it as the ultimate outcome of learning a foreign language (Simon 2005).

Having the previously stated in mind, the present research[2] is inspired by the need to find ways to help EFL learners overcome the difficulties encountered in dealing with the phonological systems of L1 and L2 and eventually perhaps achieve the goal of native-like pronunciation. Thus, the current study aims to explore the possible effects of a specifically designed audio-visual phonetic training on Serbian EFL learners' production of similar and new contrasts in the interlanguage system at two age levels. In our case, the age variable can be regarded also as a proficiency level variable since the selected groups of participants differ not only by age but by the level of English, as well.

2. *THEORETICAL BACKGROUND: PAM (Best 1994) AND SLM (Flege 1995)*

Through an extensive set of hypotheses, the *Speech Learning Model (SLM)* (Flege 1995) proposed that an L2 sound that was similar to an L1 sound would be more difficult to acquire than an L2 sound that is relatively dissimilar to an L1 sound. The reason for this is fairly simple: encountering a similar sound, a learner would automatically assume that the sounds are identical, hence any potential learning would stop and no separate category would be formed. The author regards sounds that exist in both L1 and L2 phonological inventories yet do not share certain phonetic characteristics as *similar* sounds. *New* sounds, i.e. sounds non-existent in the mother tongue inventory, will be easier to acquire. Both phonological systems mutually affect each other, however, the positive side is that learners' perceptual ability remains adaptable and functional throughout the entire life span, which is one of the most groundbreaking claims of this model. Mother tongue and target language are intertwined via category assimilation or dissimilation (Flege 2003). Similarly to the formerly established phenomenon of equivalence classification, through category assimilation, a learner establishes a category that merges the phonetic properties of L1 and L2 sounds. Category dissimilation develops a new category that strays away from L1 and L2 sounds in order to maintain the difference in the phonological space.

2 The idea for the research was presented at the Belgrade Meeting of English Phoneticians (Bimep 2018), hosted by the Faculty of Philology, University of Belgrade.

The *Perceptual Assimilation Model (PAM)* (Best 1994) presents category assimilation patterns which help us determine learners' sound discrimination ability. Learners listening to L2 sounds perceive them either as speech or non-speech sounds. Speech sounds are further separated according to the possibility of assimilating to native phonetic categories. There are several types of assimilation: two-category assimilation predicting excellent discrimination, category goodness with very good to moderate perception, and single category assimilation anticipating poor discrimination of target sounds.

In recent years a considerable number of investigations has been performed in support of Flege's *Speech Learning Model* (Guion et al. 2000; Aoyama et al. 2004), as well as Best's *Perceptual Assimilation Model* (Harnsberger 2001).

3. *AUDIO-VISUAL PHONETIC TRAINING AND FRICATIVES IN ENGLISH AND SERBIAN*

During the articulation of fricatives, the air stream passes through a narrow channel formed by the articulators, creating thus an aperiodic turbulent noise or frication (Roach 1990: 48; Stevens 1998: 129; Ashby 2011: 58). English fricatives are classified as labiodental /v/, /f/, dental (or interdental) /ð/, /θ/, alveolar /z/, /s/, palato-alveolar (or post-alveolar) /ʒ/, /ʃ/ and glottal /h/, with /v ð z ʒ/ being voiced, while /f θ s ʃ h/ being voiceless. Some authors believe there is a difference between palato-alveolar and post-alveolar fricatives (Gimson 1978: 149), while others believe the terms can be used interchangeably (Ladefoged 2006: 12). Serbian classification is not that straightforward, however. Simić and Ostojić (1996: 190) classify /f/ as labial, /z/ and /s/ as dorso-dental /ʒ/, /ʃ/ as apico-palatal and /h/ as post-dorsal velar. Traditionally, /v/ is not regarded as a fricative, but as a sonorant. Miletić (1960: 29) provides a similar typology. Belić (1972: 53) classifies /s/ and /z/ as strident dental fricatives, /ʒ/, /ʃ/ as sibilant dentals. Sibilants /s/ and /z/ are thus primarily dental in Serbian (Simić, Ostojić 1996: 188; Petrović, Gudurić 2010: 237), sometimes recognized as dental-alveolar (Miletić 1960: 42). According to some other classifications, /ʒ/, /ʃ/ are classified as palatals (Stanojčić, Popović 1992; Petrović, Gudurić 2010: 241). There are no dental fricatives in Serbian. Hence, sibilant fricatives can be characterised as similar sounds in Serbian and English due to their diverse classifications, while interdentals can be described as new sounds, according to Flege's explanation (Flege 1995).

Properly implemented phonetic training not only helps learners concentrate on specific phonetic cues, but it also reinforces their memory by finding particular sound traces necessary for establishing more accurate categories for L2 sounds (Wang, Munro 2004). Typically, perception training aims at improving the most difficult contrasts, notoriously problematic for learners (Bradlow et al. 1997), while production training has the goal to improve target sounds production (Carey 2004).

In audio-visual phonetic training sessions learners are provided with visual materials comparing their production to those of native speakers, and corrective feedback, either direct or indirect, plays a crucial role (Wong 2013). Visual cues or articulatory gestures visible in the production of sounds help learners to perceive sounds (Hazan et al. 2005). The importance of visual cues is evident in the phenomenon of the McGurk effect which is based on acoustic information, as well as in lip reading practices (Nath, Beauchamp 2011).

Audio-visual training proved highly beneficial for the development of perception of children with specific language disorders and impairments (Meronen et al. 2013; Heikkilä et al. 2018), but also for individuals with hearing loss (Massaro, Light 2004; Lidestam et al. 2014; Moradi et al. 2017). When it comes to the Serbian scientific context, Marković (2014) investigated the role of visual input in the perception and production training of L2 English vowels by speakers divided into two groups, demonstrating the notable improvement of pronunciation in the group receiving visual along with auditory stimuli.

For the purposes of the present study, interdental and sibilant fricatives were chosen as target tokens due to the fact that interdental fricatives represent novel phonemic contrasts for Serbian EFL learners, while sibilant fricatives represent similar phonetic contrasts, since they do not share all the phonetic features with English equivalents, even though /s, z, ʒ, ʃ/ share the phonemic status in both languages.

4. METHODOLOGY

4.1. Research Questions

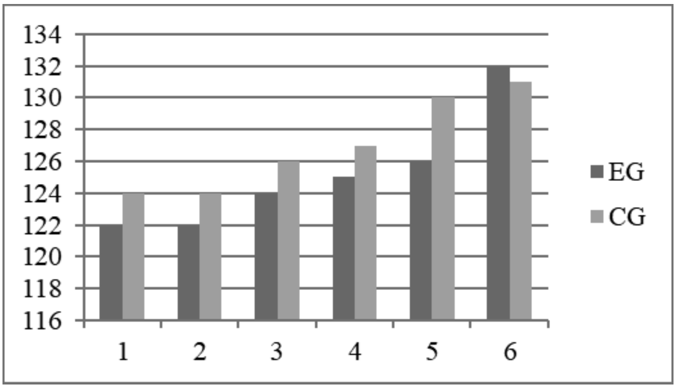
The present investigation is based on the following research questions:

- What is the effect of audio-visual phonetic training on the acquisition of L2 fricatives, i.e. do the fricatives pronounced by the experimental and control group show the same phonetic parameters in a post-test examination?
- Is the effect of the audio-visual phonetic training the same at two different age/proficiency levels?
- Is the effect of audio-visual phonetic training the same for phonemic and phonetic contrasts in L1 and L2?

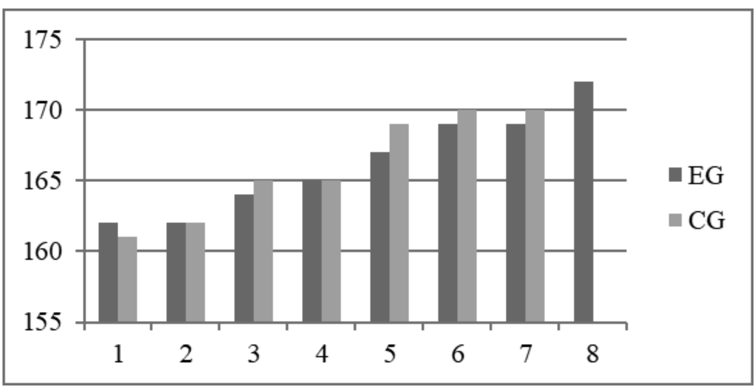
4.2. Participants

A total of 27 students participated in the study, 12 belonging to the first age group (6th graders Primary School ("Ljubiša Urošević" in Jagodina), 8 female, 4 male, mean age 11.65, divided into an experimental group (EG) of 6 participants: a control group (CG) of 6 participants) and 15 belonging to the second age group (4th graders Secondary School (Grammar School "Svetozar Marković" in Jagodina), 6 female, 9 male, mean age 17.95, divided into an experimental group of 8 participants, and a control group of 7 participants). The students in experimental and control groups were joined in such a way that there were no differences in L2 proficiency. To determine the current proficiency level of the participants and to enable equal division of the participants into experimental and control groups, the students underwent in-class paper-based testing[3] with their class instructor. The results of the testing were interpreted according to the available Cambridge Scale corresponding to the appropriate CEFR levels[4]. The students' points can be seen in Graph 1 and Graph 2 for the primary and secondary school participants, respectively.

Graph 1. Results of the Flyers Test for 6th Graders Primary School (EG and CG)



Graph 2. Results of the FCE Test for 4th Graders Secondary School (EG and CG)



The selected participants belonging to the group of 6th graders successfully passed the Flyers examination corresponding to the A2 level CEFR, whereas the selected participants in secondary school successfully passed FCE examination corresponding to B2 level CEFR. Since we had two age levels belonging to different levels of proficiency, we could equate the factors of age and proficiency only for the purpose of this research. By no means do we mean to say that the two factors are the same, yet that along with the different

3 Sample tests for Flyers can be found on http://www.cambridgeenglish.org/exams-andtests/flyers/preparation/, while First Certificate for schools sample test can be found on http://www.cambridgeenglish.org/exams-and-tests/first/preparation/.

4 Cambridge English Scale is available on http://www.cambridgeenglish.org/exams-and-tests/cambridge-english-scale/.

age, the proficiency levels of the participants were different as well. Thus we did not need to regard the two factors separately. There was quite a diverse set of participants in terms of their proficiency level, which was an interesting starting point for the investigation of their pronunciation mastery. The proficiency level of the participants was rather high for their age, yet it is understandable considering the fact that the majority of participants reported on having studied English in a private school additionally to the regular classes in their public school. For the sake of validity and reliability of group divisions, it seems important to note that no significant differences in proficiency were detected between the experimental and control groups before the beginning of the experimental period either with 6th graders ($t=-0.803$ $p=0.380$ $SD=1.60$) or 4th graders ($t=-0.552$ $p=0.598$ $SD=1.28$). All the participants came from the same hometown (Jagodina, Serbia), which reduced the chances for specific features resulting from regional variation.

4.3. Instruments

The chosen instrument for recording students' pronunciation of target sounds was a pre-planned sentence list containing words with target sounds in initial and final positions (two tokens for each position in CV and VC contexts). The only exceptions were a voiced post-alveolar fricative (no initial position is possible due to phonotactic restrictions), for which there were two tokens in final position, one in intervocalic and one in pre-nasal position; and a voiced interdental fricative, for which there was one token in final position and the other one in intervocalic position due to the lack of lexemes familiar to the 6th graders containing this particular consonant in final positions. A sentence instead of a word list was chosen in order to enable a more natural context for the pronunciation of target sounds. Hence, the instrument contained 24 sentences (8 for interdentals and 16 for sibilants) adjusted to the participants' proficiency levels. More precisely, the sentences were adapted to suit the needs of 6th graders (taking grammar and vocabulary into account)[5], while the same instrument was used for the older group, to ensure the validity of the testing, even though the sentences for the higher proficiency level could have been more complex. However, our focus was predominantly on target sounds. The choice of vocabulary and grammar in the sentences was only supposed to facilitate pronunciation. We avoided using a larger list of sentences so as to prevent fatigue and attention deficiency, especially with younger participants. An additional instrument, such as a word list, was likewise omitted because our aim was not to explore the effect of the selected tasks on participants' pronunciation. The same instrument was used both in pre-test and post-test examinations.

5 The expected list of vocabulary at level A2 can be found on http://www.cambridgeenglish.org/images/149681-yle-flyers-word-list.pdf.

4.4. Procedure

Audio-visual training. The audio-visual training for the experimental group included listening to native speakers and different recordings, watching videos of specific pronunciations, slow-motion movements of the articulators, etc. The combined sets of materials were mostly compiled from various websites[6] or excerpted from movies and TV shows. The training was structured in such a way that the instructor gradually introduced sounds and their features. After all the sounds were covered, the instructor introduced new materials with the occasional repetition of the already seen ones. Furthermore, the instructor provided necessary explanations, as well as corrective feedback. The control group merely received traditional audio training, including recordings of isolated words or simple texts taken from Roach (1990) and the accompanying recordings, also followed by the instructor's explanations and necessary feedback. The experimental period took place during the winter term of 2012/2013 school year, from October to December, for two hours a week (two months overall). The training sessions were designed as extracurricular activities for which the students received course credits. Both the experimental and control groups at both levels of interest had the same instructor. All the participants voluntarily agreed to take part in the examination.

It seems fitting to clarify the intention of the author in approaching the topic from the perspective of sound classification and results interpretation based on the phonological status of target sounds, yet with the awareness of the fact that the visual information may have been the most useful and accessible for the interdentals during the audio-visual training.

Data collection. The recording for obtaining information on students' pronunciation of target sounds was performed before and after the experimental period in the forms of pre- and post-tests for both experimental and control groups. The participants read the sentence list containing target sounds by repeating each sentence twice. All the recordings were made using the digital voice recorder Olympus VN-8600PC, sampling frequency 44.1 kHz, following the recommendations for recording procedure in the relevant literature (Harrington 2010: 13).

Acoustic Analysis. Acoustic analysis of the recorded data was performed in *Praat* (Boersma, Weenink 2017), version 6.0.36. For sibilants, it consisted of measuring spectral moments, intensity and frication (noise) duration. The spectral moments included centre of gravity, standard deviation, skewness and kurtosis, as applied in the previous research (Jongman et al. 2000). Centre of gravity, standard deviation, skewness and kurtosis are known in the literature as spectral moments. *Centre of gravity* depicts how high the frequencies in a spectrum are on average, while *standard deviation* shows how much the frequencies in the spectrum can deviate from the centre of gravity. *Skewness* demonstrates how much the shape of the spectrum below the centre of

6 An example of pronunciation demonstrations can be found on http://www.bbc.co.uk/learningenglish/english/features/pronunciation.

gravity is different from the shape above the mean frequency. Finally, *kurtosis* measures how much the shape of the spectrum around the centre of gravity is different from a Gaussian shape. Along with spectral moments, studies have shown the importance of relative amplitude and frication duration for the distinction of fricatives (Jones, McDougall 2009; Harrington 2010).

When it comes to interdental fricatives, their acoustic analysis included spectrogram interpretation and comparison, having been coded as either precise/accurate or imprecise/inaccurate pronunciation. More specifically, if the student pronounced an interdental fricative which could be seen as the existence of frication on a spectrogram instead of a differential substitute (alveolar plosive in this case, recognized by the silence, burst and characteristic frication noise following it), we regarded the pronunciation as accurate and counted it as such, regardless of whether it strayed away from native-like production. For the moment we were not interested in measuring spectral moments, since in most of the cases in the pre-test the pronunciation of interdentals was characterized by differential substitution, especially with the younger participants. Furthermore, interdentals are notorious for being very difficult to measure on the spectrogram so we estimated that this type of analysis would not be suitable for the present research. Hence we only measured spectral moments for those interdental fricative productions that resembled native-like pronunciation and provided results as evidence (however scarce) for the possibility of new category formation.

4.5. Statistical Data Analysis

Percentage counts, descriptive and t-test statistics were performed using IBM SPSS software, version 24.

5. RESULTS AND DISCUSSION

The mean results of the measurements of spectral moments and two other acoustic features (duration and intensity) relevant for a fricative description related to the first age/proficiency level are presented in Table 1. It seems useful to note that the measurements were obtained from the spectral slices in the medial portions of frication, where the noise was the most stable.

Table 1. Pre-Test and Post-Test Mean Results of Sibilants Production (Primary School)

Groups	Means	Centre of Gravity (Hz)	Standard Deviation (Hz)	Skewness	Kurtosis	Frication Duration (ms)	Intensity (dB)
Pre-test							
EG	/**s**/	7877	2398	-0.43	3.08	168	55.4
	/**z**/	8124	1985	1.29	2.58	154	49.2
	/ʃ/	3397	2430	-0.06	-1.26	195	59.2
	/ʒ/	3240	2235	-0.18	1.92	189	58.6

CG	**/s/**	8155	2012	0.14	-0.99	175	56.1
	/z/	8112	2457	2.92	1.95	172	51.2
	/ʃ/	3448	2245	-0.16	1.68	190	58.5
	/ʒ/	3285	2138	-0.09	-1.26	202	60.1
Post-Test							
EG	**/s/**	7245	2157	-0.29	2.78	182	54.8
	/z/	7520	2992	-0.18	2.42	176	49.8
	/ʃ/	3822	1788	1.24	1.08	188	61.5
	/ʒ/	3696	2012	-0.06	-0.27	196	57.2
CG	**/s/**	8217	1998	0.26	-1.17	185	57.4
	/z/	7996	2158	0.45	1.72	172	62.2
	/ʃ/	3242	2008	-1.32	-0.06	200	60.7
	/ʒ/	3118	1782	1.86	-0.09	212	59.6

If we compare the obtained data to the aforementioned results from the previous study by Jongman et al. (2000), we notice higher values for alveolar sibilants across experimental and control groups in both types of testing. This may point to a typically dental realizations of these sounds in Serbian, compared to alveolar localizations in English. Articulation located to the front of the mouth raises the value of the centre of gravity. Moreover, the higher frequency may be the result of a narrower passage of airstream through the front teeth. The lower frequency of post-alveolar sibilants characteristic of our Serbian sample might point to the articulation located further back than in the case of English, resulting from the inability of the students to perceive the subtle phonetic differences in the location of English and Serbian fricatives – post-alveolar as opposed to palatal. This is particularly interesting due to the fact that these localizations are very similar. Not rarely does it so happen that even the English sounds in question are classified as palatal. However, this close proximity causes difficulty in perception for the realisation of sounds in interlanguage phonology.

Comparing the frequencies of the investigated fricatives in post-test examination to the pre-test results with both groups of interest, we see only slight changes with the experimental group, especially regarding the lowering of frequency for alveolars and raising the frequency for post-alveolars. However, the changes are insufficient to approximate the values to the values of native speakers from the relevant study by Jongman et al. (2000). When it comes to the frication duration, a prolonged duration can be noticed in the interlanguage especially with /ʒ/, which may point to more attention paid to articulation due to the formality of the testing event. Overall, there are lower values of intensity within our sample, which may have been the result of participants' age and psychological factors, such as examination anxiety.

Table 2 presents the results of the same testing procedure for the higher proficiency/age level participants.

Table 2. Pre-Test and Post-Test Mean Results of Sibilants Production (Secondary School)

Groups	Means	Centre of Gravity (Hz)	Standard Deviation (Hz)	Skewness	Kurtosis	Frication Duration (ms)	Intensity (dB)
Pre-test							
EG	/s/	8578	2100	0.18	1.29	175	57.6
	/z/	8193	1890	0.95	1.38	182	58.9
	/ʃ/	3449	2455	1.29	1.17	240	60.7
	/ʒ/	3238	2317	-0.68	-0.84	229	62.5
CG	/s/	8256	2425	-0.98	-0.96	188	55.8
	/z/	7801	2196	-1.94	-1.83	185	58.4
	/ʃ/	3274	2243	-0.08	0.04	224	62.9
	/ʒ/	3382	2482	1.62	-1.31	232	58.4
Post-Test							
EG	/s/	7958	2007	0.25	-0.87	178	56.2
	/z/	7865	2297	-1.54	-0.96	184	52.9
	/ʃ/	3921	2221	-0.09	0.08	226	61.1
	/ʒ/	3816	2504	0.07	-1.12	218	58.9
CG	/s/	8145	2148	0.86	2.96	180	58.2
	/z/	8006	2367	0.93	1.93	189	56.4
	/ʃ/	3688	1965	-0.57	2.14	206	60.8
	/ʒ/	3574	2444	-1.74	-0.58	236	62.3

Generally, higher frequencies are noticeable for alveolars compared to the younger testing sample, however, the discrepancy from the native speakers remains considerable, which is something that the two age groups from our sample have in common. The frequency values for post-alveolars are likewise lower than the results by Jongman et al. (2000), for which the same explanation can be offered, i.e. this may be due to the tiny phonetic differences in the places of articulation between the sounds in English and Serbian. The differences in frequency may also have been the result of specific tongue configuration. To be more precise, the sampled participants perhaps used the front part of the tongue more actively to produce the alveolar sounds and the sides of the body of the tongue for post-alveolars in a slightly different way from the participants in the original English study. The measurements point to longer frication duration, which again may be interpreted as a sign of careful speech. Nevertheless, it should be noted that the participants read the same sentence twice in a row, yet it had no effect on frication duration. The overall intensity was slightly higher compared to the younger group of participants. Comparing experimental and control groups, there is actually no appreciable difference in post-examination, except in the case of a fairly negligible lowering of the alveolar centre of gravity of the experimental group in the post-test. The frequencies are likewise rather variable, which adds to the variable nature of interlanguage itself.

For the already explained reasons, we present the results of accuracy percentage for interdentals in Table 3. The remaining percentage not presented in

Наслеђе 48 • 2021 • 317–335

the table is reserved for differential substituents /d/, /t/, /ts/, /dz/, exceeding the scope of our present study.

Table 3. Pre-Test and Post-Test Results of Interdentals Production (Accuracy %)

Primary School (6th grade)				Secondary School (4th Grade)			
EG		CG		EG		CG	
/θ/	/ð/	/θ/	/ð/	/θ/	/ð/	/θ/	/ð/
Pre-Test							
20.83	37.5	25	33.33	34.38	37.5	40.63	39.29
Post-Test							
70.83	83.33	37.5	45.83	71.88	78.13	50	42.86
N=48		N=48		N=64		N=56	

The improvement of the experimental group when it comes to interdental fricative production is strikingly high, judging by the results presented in the table. The percentage of accurate production is slightly higher for the control group, too. However, the increase is more evident in the experimental groups for both levels of age/proficiency. This points to the positive effects of the implemented audio-visual training, especially regarding the voiced interdental counterpart. The slight enhancement of performance for the control group likewise underscores the significance of the application of any kind of explicit instruction, rather than no instruction at all.

The measurements of spectral characteristics of interdentals are available solely for the accurate realizations of interdental sounds (the ones approximating the native-like production), since the measurements for differential substituents would correspond to the acoustic features of plosives. The results of spectral features of interdental fricatives in the interlanguage are presented in Table 4.

Table 4. Mean Spectral Moments for Accurate Realizations of Interdentals

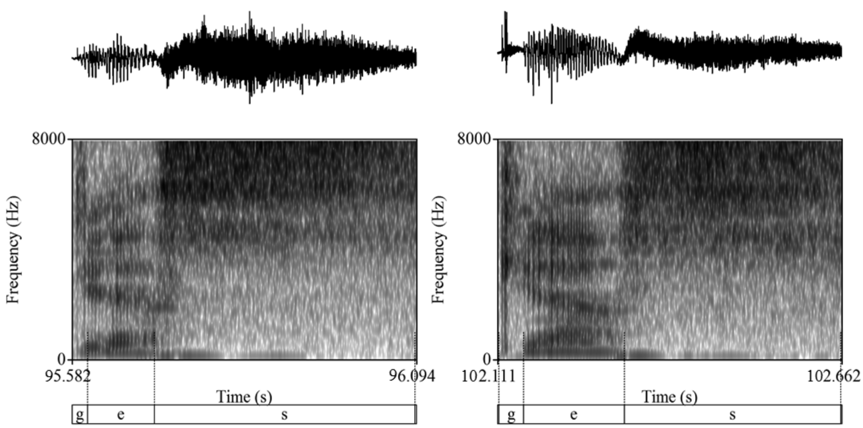
Groups (Means)			Centre of Gravity (Hz)	Standard Deviation (Hz)	Skewness	Kurtosis	Frication Duration (ms)	Intensity (dB)
Pre-test								
EG	6th	/θ/	6771	3047	0.64	-0.13	142	48.2
	4th		6133	3374	0.42	-0.85	194	52.7
	6th	/ð/	5874	4512	-0.09	-1.22	82	50.5
	4th		5745	3808	-1.24	-0.95	112	55.6
CG	6th	/θ/	6415	4397	0.58	0.08	155	49.7
	4th		6129	4088	0.74	-1.17	172	48.9
	6th	/ð/	5843	3236	-1.52	-0.58	92	44.9
	4th		5465	3153	-1.22	-1.14	115	54.2
Post-Test								
EG	6th	/θ/	5356	2826	0.75	1.58	138	52.1
	4th		5045	4176	-0.04	-0.92	118	55.6
	6th	/ð/	5118	3224	1.32	-1.08	89	55.8
	4th		5213	3768	1.07	0.84	99	52.7

CG	**6th**	/θ/	6775	3412	-0.98	0.18	149	50.9
	4th		5988	3533	-1.14	-0.06	182	55.8
	6th	/ð/	6112	4116	-0.08	1.35	85	49.6
	4th		5824	3396	1.29	1.15	104	54.7

Even though the spectrogram clearly shows frication, the results from the table demonstrate a significantly higher centre of gravity for both of the interdental fricatives across age levels, especially in the pre-test examination. Mind you, both the pre-test and post-test results from Table 4 represent accurate realizations only. The higher frequency of interdentals may have been the result of a greater pressure exerted upon upper incisors creating thus a narrower passage for the airstream, as well as of a different tongue configuration.

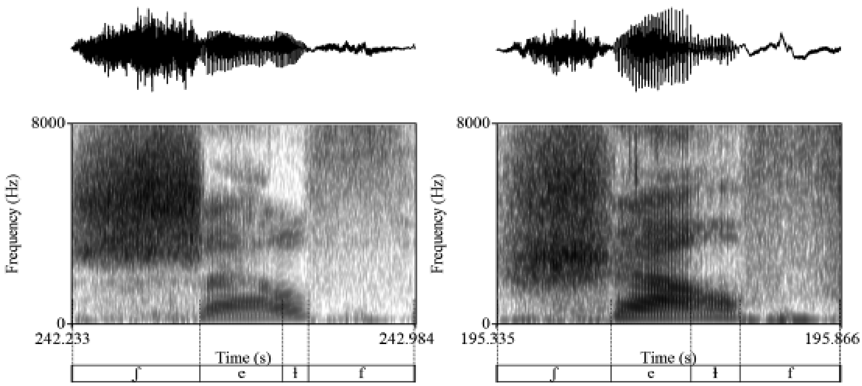
The visual representation of the scarce improvements noticed in the corpus of sibilant fricatives can be seen in Figure 1 and 2. They provide spectrogram illustrations of the comparison of pre-test and post-test productions of voiceless alveolar and post-alveolar sibilants produced by the 6th-grade Primary School participant (Figure 1) and the 4th-grade Secondary School participant (Figure 2).

Figure 1. Spectrogram Illustrations for Pre- (left) and Post-Tests (right) – Improved Alveolar Voiceless Sibilant



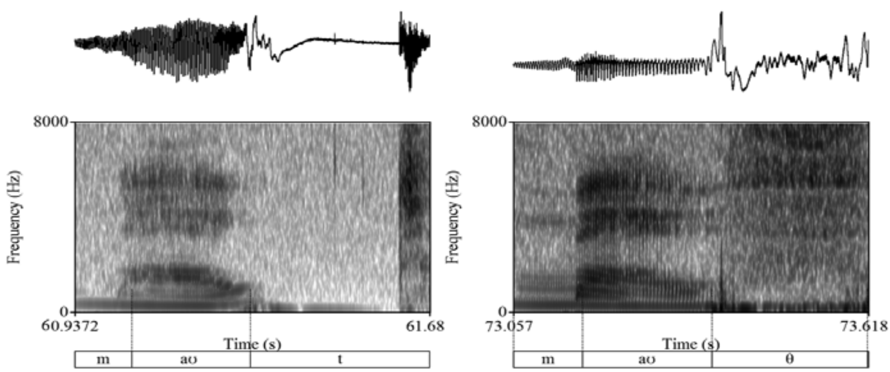
In Figure 1 we see the subtle lowering of the beginning of frication in the word *guess* which results in the lowering of the overall centre of gravity. Illustrations for post-alveolars in Figure 2 are given in reverse order, i.e. the post-test result is provided on the left. We notice a slightly higher frequency in the spectrogram here along with less intensity. However, examples like these are rarely found in our corpus.

Figure 2. Spectrogram Illustrations for Pre- (right) and Post-Tests (left) – Improved Post-Alveolar Voiceless Sibilants



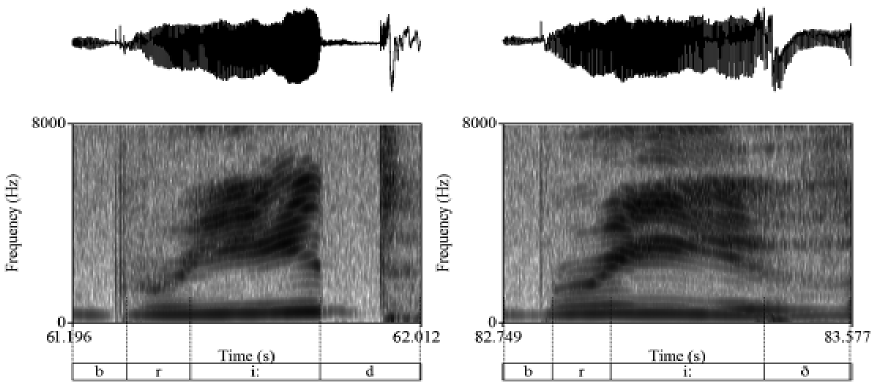
Figures 3 and 4 likewise present spectrogram illustrations of the younger and older group participants, respectively. The different spectral configuration is evident right away in the spectrogram on the right depicting the post-test production.

Figure 3. Spectrogram Illustrations for Pre- (left) and Post-Tests (right) – Improved Voiceless Interdental



In Figure 3 on the spectrogram to the left, we can see the silence, burst and frication portions characteristic of plosives. The visible formant transitions testify to the specific nature of alveolar plosives, i.e. the chosen differential substitute for the voiceless interdental fricative. The spectrogram on the right gives clear indications of the turbulent noise of the final sound of the word *mouth*.

Figure 4. Spectrogram Illustrations for Pre- (left) and Post-Tests (right) - Improved Voiced Interdental



The spectrogram on the left in Figure 4 likewise shows clear signs of an alveolar plosive which is characterized by the silent period and the burst followed by short frication. The presence of a voice bar tells us that we are dealing with the voiced sound. Similarly, to the previous Figure 3, in the spectrogram to the right we notice aperiodic noise, i.e. frication in the final section of the word *breathe*, which, along with the formant transitions and the presence of a voice bar, points to the existence of a voiced interdental fricative.

In the next section of the paper, the quantitative results of the statistical testing are presented, i.e. the comparison of the obtained results for all the fricatives tested. Table 5 presents the results of experimental and control groups comparison across age/proficiency levels, which was supposed to demonstrate whether there was a statistically significant difference in the results after the experimental period.

Table 5. Pre-test and Post-test Results Comparison at Two Age/Proficiency Levels

Sounds	Pre/Post Test T-test Results	
	6th Grade Primary School	4th Grade Secondary School
/s/	p=0.91 t=0.112 df[7]=47 SEoD[8]=10.396	p=0.59 t=0.544 df=59 SEoD=57.507
/z/	p=0.86 t=0.176 df=47 SEoD=10.416	p=0.22 t=1.289 df=59 SEoD=47.831
/ʃ/	p=0.94 t=0.075 df=47 SEoD=28.560	p=0.67 t=0.436 df=59 SEoD=61.028
/ʒ/	p=0.71 t=0.128 df=47 SEoD=24.758	p=0.49 t=0.715 df=59 SEoD=52.301
/θ/	**p=0.005** t=5.058 df=47 SEoD=480.554	**p=0.001** t=11.498 df=59 SEoD=188.083
/ð/	**p=0.003** t=3.839 df=47 SEoD=387.860	**p=0.001** t=27.339 df=59 SEoD=69.445

Judging by the results of t-test calculations, the statistically significant difference is detected with interdental fricatives only, for both levels of interest ($p<0.05$). This means that the audio-visual training implemented during the

7 Abbreviation for Degrees of Freedom
8 Abbreviation for Standard Error of Difference

experimental period had a positive impact on the improvement of realizations of interdental fricatives, while no significant effect was noticed with the other investigated fricatives. Such an outcome may underscore the fact that new sounds in the interlanguage system react better to intervention, i.e. are more likely to improve during a short-term audio-visual training. Perhaps a longer systematic application of audio-visual training could yield the same results with similar sounds, in our case alveolar and post-alveolar fricatives.

Finally, Table 6 presents results of the statistical testing of experimental group results in post-test examination comparing the two groups of participants.

Table 6. Post-Test Results Comparison Between Two Age/Proficiency Levels

Sounds	T-test Results Between Age Levels
/s/	p=0.18 t=1.460 df=55 SEoD=81.122
/z/	p=0.74 t=0.335 df=55 SEoD=95.011
/ʃ/	p=0.31 t=1.091 df=55 SEoD=87.786
/ʒ/	p=0.96 t=0.054 df=55 SEoD=151.763
/θ/	p=0.82 t=0.229 df=39 SEoD=129.761
/ð/	p=0.94 t=0.676 df=41 SEoD=170.384

No significant differences were noticed in any of the investigated consonants, which means that the audio-visual training had the same effect on the improvement of interdental fricatives, as well as no improvement in the production of alveolars and post-alveolars regardless of the current age/proficiency level of the participants.

6. *CONCLUSION*

To briefly summarize, the present paper aimed at investigating the effects of audio-visual training on the production of new and similar sounds in the interlanguage system, i.e. we attempted at examining whether the potential benefits of phonetic training are identical both at phonetic and phonemic levels. We likewise sought to find the answer to the research question regarding the influence of participants' age level (which, in our particular case, also meant different proficiency levels) on the effect of the implemented phonetic training.

The measurements obtained in the analysis show different phonetic features of similar sounds in the interlanguage compared to the results obtained from the native speakers (Jongman et al. 2000). This particularly relates to the generally higher centre of gravity for alveolars and lower centre of gravity for post-alveolars which was interpreted as the outcome of subtle differences in sound localisation in Serbian and English. Simultaneously, the audio-visual training did not significantly influence the improvement of realizations across age/proficiency levels which additionally confirms the assumed difficulty in the acquisition of similar sounds (Flege 1995). However, when it comes to interdental fricatives, defined as new sounds in the interlanguage, we noticed statistically significant improvement reflected in the decrease of the number of differential substituents in the post-test examination for both age/proficiency levels. This leads to the conclusion that the implemented audio-visual training had more beneficial effects on new sounds than on similar sounds in Serbian-English interlanguage phonology. Nevertheless, spectral measurements of the accurate realizations of interdental fricatives (approximating native-like production) showed the higher centre of gravity frequency values within our sample of participants, which we explained by different tongue configurations and exerted pressure on upper incisors. Yet it goes without saying that the effects of audio-visual training could have a lasting effect on L2 category formation in terms of the acquisition of auditory and acoustic features (Best 1994; Flege 1995).

The results of the paper related to the beneficial effect of audio-visual training confirm the findings from previous studies (e.g. Iverson, Evans 2007; Lidestam et al. 2014; Marković 2014).

The length of the training period represents one of the limitations of the present study. Namely, similar sounds perhaps require a longer training period than the new sounds, or the very audio-visual training implemented here was more suitable for the training of the pronunciation of interdentals. Furthermore, it seems that for the interdentals the visual information was considerably more meaningful for the participants than for the post-alveolars, for instance. The design of the instruments could likewise be altered for future research. Perhaps including a list of isolated words along with the contextualized examples represents a more favourable option. Moreover, the familiarity of the testing task might have caused a higher percentage of accuracy for interdentals, hence a different post-test word selection may be considered for future research.

Nevertheless, the results of the present research underscore important pedagogical implications related to the positive effects of phonetic training, even though not all the sounds showed equal improvement. After all, the non-existent sounds in the mother tongue phonological systems are more marked and thus more easily noticeable in interaction. Thus, teachers might, for starters, at least attempt at solving the problems with the new sounds by a more systematic application of the audio-visual training. The fact that both levels of age/proficiency benefited from the audio-visual training is also important for the enhancement of the teaching practice in Serbian EFL classrooms.

Future research may focus on a longer time-span of the training period, as well as on the long-term effects of audio-visual training.

REFERENCES

Aoyama et al. 2004: K. Aoyama, J. E. Flege, S. G. Guion, R. Akahane-Yamada, T. Yamada, , Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/, *Journal of Phonetics*, 32, 233–250.

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

Belić 1972: A. Belić, *Osnovi istorije srpskohrvatskog jezika I – Fonetika,* Beograd: Naučna knjiga.

Best 1994: C. Best, 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,* Cambridge, MA: The MIT Press, 167-224.

Best 1995: C.T. Best, A direct realist view of cross-language speech perception, *in*: Strange, W. (ed.), *Speech Perception and Linguistic Experience. Theoretical and Methodological Issues,* Baltimore: York Press, 171-203.

Best, Tyler 2007: C. Best, M. Tyler, 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*, Amsterdam; Philadelphia: John Benjamins, 13-34.

Boersma, Weenink 2017: B. Boersma, D., Weenink, Praat: Doing phonetics by computer (Version 6.03.6) [Computer program], http://www.praat.org/. > 25. 10. 2017.

Bradlow et al. 1997: R. R. Bradlow, D. B. Pisoni, R. Akahana-Yamada, Y. Tohkura, Training Japanese listeners to identify English /r/ and /l/: Some effects of perceptual learning on speech production, *Journal of the Acoustical Society of America*, 101, 2299-2310.

Carey 2004: M. Carey, CALL visual feedback for pronunciation of vowels: Kay Sona-Match, *CALICO J.* 21(3), 571–601.

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

Flege 1995: J.E. Flege, Second language speech learning theory, findings and problems, *in*: W. Strange (Ed.) *Speech Perception and Linguistic Experience: Issues in Cross Language Research*, Baltimore, MD: York Press, 233-277.

Flege 2003: J.E. Flege, Assessing Constraints on Second-Language Segmental Production and Perception, *in*: (eds.) A. Meyer, N. Schiller, *Phonetics and Phonology in Language Comprehension and Production, Differences and* Similarities, Berlin: Mouton de Gruyter, 319-355.

Gimson 1978: A.C. Gimson, *An Introduction to the Pronunciation of English*: ELBS.

Guion et al. 2000: S. Guion, J. E. Flege, R. Akahane-Yamada, J.C. Pruitt, 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.

Harnsberger 2001: D.J. Harnsberger, The perception of Malayalam nasal consonants by Marathi, Punjabi, Tamil, Oriya, Bengali, an American English listeners: A multidimensional scaling analysis, *Journal of Phonetics*, 29, 303-327.

Harrington 2010: J. Harrington, *Phonetic analysis of speech corpora*, Oxford: Wiley-Blackwell.

Hazan et al. 2005: V. Hazan, A. Sennema, M. Iba, A. Faulkner, Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English, *Speech Communication*, 47, 360–378.

Heikkilä et al. 2018: J. Heikkilä, E. Lonka, A. Meronen, S. Tuovinen, R. Eronen, P. Leppänen, U. Richardson, T. Ahonen, K. Tiippana, The effect of audiovisual speech training on the phonological skills of children with specific language impairment (SLI), *Child Language Teaching and Therapy*, *34*(3), 269–287. https://doi.org/10.1177/0265659018793697

Iverson, Evans 2007: P. Iverson, B.G. Evans, Auditory training of English vowels for first-language speakers of Spanish and German, *Proceedings of the 16th International Congress of Phonetic Sciences*, Saarbrücken, Germany, 1625-1628.

Jenkins 2000: J. Jenkins, *The Phonology of English as an International Language,* Oxford: Oxford University Press.

Jones, McDougall 2009: M. Jones, K. McDougall, The acoustic character of fricated /t/ in Australian English: a comparison with /s/ and /ʃ/, *Journal of the International Phonetic Association*, 39(3), 265-289.

Jongman et al. 2000: A. Jongman, R. Wayland, S. Wong, Acoustic characteristics of English fricatives, *Acoustical Society of America*, 108, 1252-1263.

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

Lenneberg 1967: E. Lenneberg, *Biological foundations of language*, New York: Wiley.

Lidestam et al. 2014: B. Lidestam, S. Moradi, R. Pettersson, T. Ricklefs, Audiovisual training is better than auditory-only training for auditory-only speech-in-noise identification, *Journal of the Acoustical Society of America*, 136, EL142–47.

Major 2001: R.C. Major, *Foreign accent: The ontogeny and phylogeny of second language phonology*, Mahwah, NJ: Laurence Erlbaum Associates.

Marković 2014: M. Marković, The effect of the visual component in the training of perception and production of vowels by Serbian learners of English, *u*: Gudurić, S., Stefanović, M. (ur.) *Jezici i kulture u vremenu i prostoru 3*, Novi Sad: Filozofski fakultet.

Massaro, Light 2004: D.W. Massaro, J. Light, Using visible speech to train perception and production of speech for individuals with hearing loss, *Journal of Speech, Language, and Hearing Research*, 47, 304–20.

Meronen et al. 2013: A. Meronen, K. Tiippana, J. Westerholm, T. Ahonen, Audiovisual speech perception in children with developmental language disorder in degraded listening conditions, *Journal of Speech, Language and Hearing Research*, 56, 211–21.

Miletić 1960: B. Miletić, *Osnovi fonetike srpskog jezika*. Beograd: Naučna knjiga.

Moradi et al. 2017: S. Moradi, A. Wahlin, M. Hällgren, J. Rönnberg, J.B. Lidestam, The efficacy of short-term gated audiovisual speech training for improving auditory sentence identification in noise in elderly hearing aid users, *Frontiers in Psychology*, 8: 368.

Nath, Beauchamp 2011: A. R. Nath, M. S. Beauchamp, Dynamic changes in superior temporal sulcus connectivity during perception of noisy audiovisual speech, *Journal of Neuroscience*, 31, 1704–1714.

Pennycook 1998: A. Pennycook, *English and the discourses of colonialism*, London, England: Routledge.

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

Roach 1990: P. Roach, *English phonetics and phonology - A practical course*, Cambridge: Cambridge University Press.

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

Simon 2005: E. Simon, How native-like do you want to sound? A study of the pronunciation target of advanced learners of English in Flanders, *Moderna Sprak*, 99, 12–21.

Stanojčić, Popović 1992: Ž. Stanojčić, Lj. Popović, *Gramatika srpskoga jezika. Udžbenik za I, II, II i IV razred srednje škole,* Novi Sad: Zavod za udžbenike.

Stevens 1998: K.N. Stevens, *Acoustic phonetics*, Cambridge, MA: MIT Press.

Wang, Munro 2004: X. Wang, M. Munro, Computer-based Training for Learning English Vowel Contrasts. *System*, 32, 539-552.

Wong 2013: W. S. Wong, *Training the Perception and Production of English Vowels /ɪ/-/iː/, /e/-/æ / and /ʊ/-/uː/ by Cantonese ESL Learners in Hong Kong*, PhD dissertation, Dept. Eng., The Chinese Univ. of Hong Kong, Hong Kong.

Даница М. Јеротијевић Тишма

МОЖЕ ЛИ АУДИО-ВИЗУЕЛНИ ТРЕНИНГ ПОДЈЕДНАКО УТИЦАТИ НА ФОНЕМСКЕ И ФОНЕТСКЕ КОНТРАСТЕ? ПРИМЕР: ПРОДУКЦИЈА ФРИКАТИВА СТРАНОГ ЈЕЗИКА

Резиме

Рад истражује ефекат аудио-визуелног перцептивног тренинга на изговор нових фонемских и фонетских контраста из Ј2 код српских ученика енглеског као страног језика, у конкретном случају фокусираног на фрикативе. У том смислу рад има за циљ да открије да ли аудио-визуелни тренинг има једнаке ефекте на фонемском и фонетском нивоу, као и да ли је ефекат исти када се у обзир узму два различита старосна узраста/нивоа постигнућа, шести разред основне школе и четврта година средње школе. Како бисмо подробније истражили фонемски ниво, концентрисали смо се на интерденталне фрикативе, а за разлике на фонетском плану укључили смо сибилантске контрасте, пратећи предвиђања *Модела перцептивне асимилације* (Best 1994) и *Модела учења говора* (Flege 1995). Испитивање релевантних акустичких информација извршено је пре и после експерименталног периода, када су сви учесници снимљени док изговарају унапред припремљену листу реченица која садржи циљне гласове, а састојало се од мерења спектралних момената, трајања фрикације и поређења спектрограма. Резултати аудио-визуелног фонетског тренинга показали су се посебно корисним за фонемске контрасте, тј. интерденталне фрикативе за обе групе испитаника, док су сибилантски контрасти показали безначајан напредак, што се донекле и дало очекивати. Чини се да узраст/ниво постигнућа није значајан предиктор ефекта аудио-визуелног тренинга. Уз емпиријске резултате, рад такође представља педагошке импликације важне за наставу изговора и значај фонетског тренинга нарочито у српском контексту учења енглеског као страног језика.

Кључне речи: перцептивни тренинг, фрикативи, енглески као страни језик, изговор гласова страног језика, међујезичка фонологија

Submitted: 16 June 2020
Accepted: 28 December 2020