Spectral Properties of American English Monophthongs

Biljana Čubrović

English Department, Faculty of Philology, University of Belgrade, Serbia

Professor of English Linguistics in the Department of English, Faculty of Philology, University of Belgrade in Serbia, Biljana teaches English Phonetics, Acoustic Phonetics, and History of English Language, at the undergraduate and postgraduate levels. In 2006, she was awarded the Certificate of Proficiency in the Phonetics of English by the International Phonetic Association at University College London. She was a Visiting Scholar at the postdoctoral level in the Department of Linguistics at Cornell University, Ithaca, NY in 2008-2009 and 2013-2015, where she carried out experimental phonetic research. She is the author of several books about the English language and English phonetics. A monograph entitled *Acoustic Investigations of Serbian and American English Vowel Inventories* was published in 2016. She was Chief Editor of *The Linguistics Journal* from 2013-2020 and Editor-in-Chief of the *Philologia Journal* and *Belgrade English Language and Literature Studies*. She has given invited talks in Japan, UK, Slovenia, Bosnia-Herzegovina and Serbia. She is one of the founders of the academic circle of international scholars *English Scholars beyond Borders*. Her research interests include: articulatory and acoustic phonetics, phonetics and phonology of English, French, and Slavic Languages, as well as second language acquisition and bilingualism. She founded *Laboratory for Phonetic Research* at the Faculty of Philology in Belgrade in 2017. She is currenty President of the Association of Foreign Languages and Literatures of Serbia.

Acknowledgements: This paper is a printed version of the plenary talk given at the Fourth International Conference *English Scholars Beyond Borders* at Leeds Beckett University, UK (29 June–1 July 2017). I would like to extend my gratitude to Dr. Naeema Hann and Prof. Roger Nunn for this invitation.

Abstract

This vowel study examines the spectral features of American English monophthongs in two groups of participants, English L1 speakers and Serbian L1 speakers, who reside in the USA and speak Mainstream American English (MAE). The two groups consist of nine participants each, who are all male and fully fluent speakers of English. Some differences in the realization of MAE monophthongs have been found in the two groups, even though both the length of residence and language exposure to English are significant in the Serbian L1 group. English tense/lax vowel pairs are not fully acquired in the group of Serbian L1 speakers, which reflects the vowel configuration present in Serbian as L1. Furthermore, the study shows that Serbian L1 speakers of English resist more modern pronunciation characteristics, like the low back merger in American English and keep the vowel contrast.

Keywords: vowels, American English, Serbian, acoustic analysis.

Introductıon

Vowel Inventories

Languages of the world vary in the size of their vowel inventories, as well as in the qualitative and quantitative features of the vowel segments that have been observed in them. Standard Serbian and American English are the two languages whose vowels are compared in

this vowel study in order to demonstrate how Serbian as L1 may influence the acquisition of English as L2.

Standard Serbian is commonly regarded as a language with a five-vowel inventory, which is seen as the preferred vowel configuration in the languages of the world (Maddieson, 1984). According to the UCLA Phonological Segment Inventory Database (UPSID; Maddieson, Precoda & Reetz, 2014), 20.8% of the surveyed languages (or 94 languages) make use of five vowels (Čubrović, 2016, p. 17). Segment-wise the Serbian vowel system is, therefore, not an unusual one. However, the use of pitch accents in Serbian makes the vowel system of Serbian rather complex. Furthemore, judging by the way pitch accents are traditionally represented in the literature, the Serbian language is rather unique. It is a commonplace to say that pitch accents affect the vowel duration of the five Serbian vowels without affecting the vowel quality, which remains constant and is taken for granted. Some recent vowel studies follow this approach (Krebs-Lazendic & Best, 2013). However, some notable empirical vowel studies of Serbian show that vowel quality is dependent upon vowel quantity, especially in vowels /e o a/ where these differences are more readily observed (Ivić & Lehiste, 1967, p. 58–59). Lehiste and Ivić's empirical study shows that spectral properties change when quantity changes (Lehiste and Ivić, 1963). For instance, the Serbian short vowel /e/ is different with regard to vowel quality from its long counterpart /e:/ in the way that the long vowel is higher and fronter. Graph 1 displays the acoustic data based on their main informant, where it is observed that all Serbian short vowels are different with regards to both qualitative and quantitative features. Serbian short vowels are represented with full triangles, whereas empty triangles are used to mark long vowels. Figure 1 demonstrates that short vowels are moved slightly toward the centre of the diagram. The vowel contrast between /e/ and /e:/ appears as the most salient on both axes, for instance.

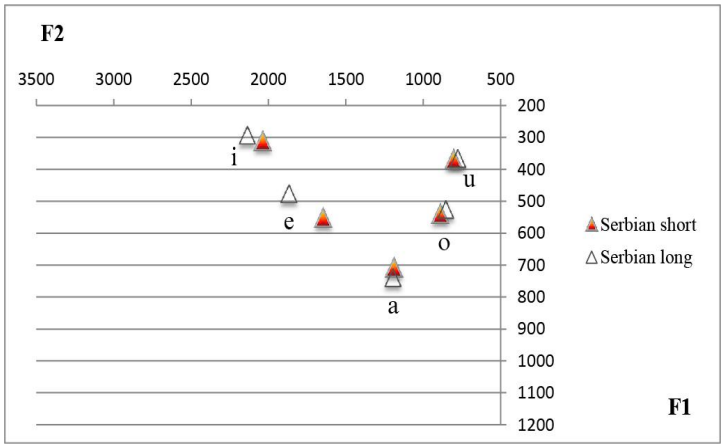


Figure 1: Serbian short and long vowel data based on the data from Lehiste and Ivić (1963, p. 82)

More comprehensive spectral changes dependent on the vowel duration are observed in Serbian mid vowels /e/ and /o/, as well as in the low /a/-vowel, as shown in the vowel diagram in Figure 1. The qualitative characteristics of Serbian high vowels /i/ and /u/ are less dependent on duration. Therefore, the quality of these two vowels is similar both in the case of /i/ and its long counterpart /i:/.

The American English (AE) vowel inventory is more diversified compared to the Serbian vowel inventory. A full set of AE monophthongs contains eleven different segments, /i ɪ e ɛ æ ʌ u ʊ o ɔ ɑ/ (Yavaş, 2011, p. 77–78), as exemplified in the following words *beat, bit, bait, bet, bat, but, boot, put, boat, bought*, and *pot* respectively. Even though these vowels are usually treated as monophthongs, some may be diphthongized, /e/ and /o/ in particular.

Furthermore, the following three are considered to be the main diphthongs of AE – /aɪ/, /aʊ/ and /ɔɪ/, as in *bite*, *bout* and *void* (Yavaş, 2011, p. 78). In comparison to standard Serbian, AE relies more on the differences in spectral properties of vowels with its five front vowels (*beat*, *bit*, *bait*, *bet*, *bat*), the central vowel (*but)*, and five back vowels (*boot*, *put*, *boat*, *bought*, *pot*).

Additionally, a much debated binary phonological grouping that includes the tense-lax distinction is often used when differentiating the vowels of American English (Lehiste & Peterson, 1961; Ladefoged & Maddieson, 1996). The vowels of *boot* and *put* are both described as high, back and rounded, which means that another feature is needed to distinguish between them. The introduction of tense-lax binary opposition solves this problem. Lax vowels are usually shorter than tense vowels, they are also lowered and more central in the vowel space compared to tense vowels and they require less muscular energy during their articulation. Yavaş (2011, p. 79) resorts to the tense-lax distinction when describing American English vowels, but he also points out that tense/lax division plays an important role in the accentuation of English words.

Experimental design

Two experiments were carried out in order to examine any differences between English L1 speakers and Serbian L1 speakers in the vowels of English. Nine monopthongs of American English were acoustically analysed in the two groups of experimental subjects. Lexical items examined were all monolyllabic words of English, earlier used by Bradlow (1993) in her comparison of English and Spanish vowels. The methods and procedures implemented in the two experiments will be discussed next.

Experiment 1: American English L1 group

Nine male speakers (ESs)[1] of American English took part in Experiment 1. At the beginning of the recording session, each participant was asked to fill in a questionnaire. The short survey included basic personal information like age, place of birth, other places of residence within the US as well as abroad (if they spent an extensive period of time living outside of the United States), mother's and father's native languages, languages spoken at home, and other languages studied. Completion of the questionnaire was crucial in the speaker selection process because we wanted to ensure that our participant cohort was as compact as possible. Because I was interested in monolingualism and English as the language of home and work, I was seeking experimental subjects who had those characteristics. I also wanted to ensure that the speaker cohort belonged to a broad dialect of American English also referred to as Mainstream American English that brings with it "the notion of a widespread, normative variety, or STANDARD DIALECT" (Wolfram and Schilling, 2015, p. 9).

The surveys show that the participants of this vowel study were from the American Northeast, except for ES2 (who was born in Minnesota, and also lived in Hawaii), but spent 10 years of his adult life in Ithaca, NY. All nine participants were students at Cornell University, Ithaca, NY, at the time of the recordings. Five were undergraduate students who elected to take a linguistics class, and four were graduate students of Linguistics at Cornell University. Their age ranged from 19 to 36 (average 23.4, median 21). With regard to the parents' language backgrounds, only two speakers had a parent whose native language is not English (ES1: Bosnian father, ES5: Dutch/Frisian father), but they have also lived in the US for an extensive period of time. All participants had exposure to other foreign languages, mostly taught in language classes. Table 1 summarizes this information:

[1] The speakers are marked as ES1-ES9.

Subject	Sex	Age	Birthplace	Language(s) spoken at home
ES1	M	19	New York City, NY	English
ES2	M	36	Mankato, MN	English, some French
ES3	M	19	Pittsburgh, PA	English
ES4	M	20	Cortland, NY	English, some Dutch and Frisian
ES5	M	20	Haverhill, MA	English
ES6	M	21	Columbia, MD	English
ES7	M	21	Manhasset, NY	English
ES8	M	28	Washington, DC	English, some Hebrew
ES9	M	26	Mt. Laurel, NJ	English

Table 1: Background information on American English L1 speakers

Eleven vowels of AE were recorded in the following monosyllabic words: *beat*, *bit*, *bait*, *bet*, *bat*, *but*, *boot*, *put*, *boat*, *bought* and *pot*. The vowels of *bait* and *boat* only served as experiment fillers and were eliminated from further analysis due to their diphthongal nature. The selected English forms have a CVC phonological structure, with an initial labial consonant, either /b/ or, in two cases, /p/. The final consonant is in all cases the coronal /t/ to eliminate possible effects of manner and place of articulation. This word list is also used in a comparative vowel study of English and Spanish (Bradlow, 1993, p. 34).

The set of forms containing the target vowels were embedded in the frame sentence "Say ___ again". The recordings were made in a sound-attenuated booth in the Phonetics Laboratory at Cornell University using Praat, Version 5.3.51 (Boersma & Weenink, 2013). The utterances with target forms embedded in the frame sentence were recorded three times in random order, giving a total of 243 tokens (9 speakers x 3 repetitions x 9 words, one for each vowel).

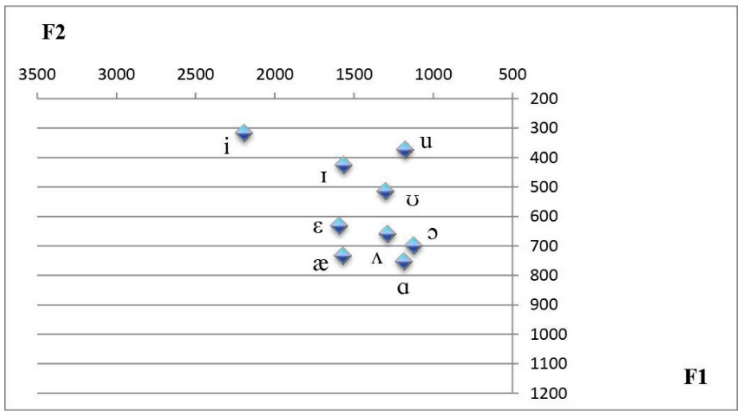
Participants were presented with the utterances on the computer screen (using Power Point), and with only one utterance on a slide at a time. Before the start of the recording, participants were given the opportunity to practice saying the experimental material. After they were acquainted with the materials, the participants were instructed to read the sentences "as naturally as possible". The experimenter monitored the recording level throughout the session so as to avoid weak or overloaded acoustic signals.

The nine vowels of English are identified by their mean location in the F1-F2 vowel space. Table 2 shows the numerical values of these two dimensions:

	/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/u/	/ʊ/	/ɔ/	/ɑ/
F1	317 (22)	426 (50)	632 (39)	733 (52)	660 (52)	374 (27)	515 (43)	699 (53)	753 (43)
F2	2192 (74)	1564 (66)	1593 (54)	1568 (49)	1288 (52)	1177 (123)	1300 (71)	1122 (89)	1183 (75)

Table 2. Mean values of F1 and F2 (with standard deviations) of English vowels produced by ESs

Graph 2 displays the acoustic measurements of English vowels, as produced by the participants in Experiment 1:



Graph 2: Vowel data for American English L1 group

Experiment 2: Serbian L1 group

Experiment 2 investigated the acoustic realizations of American English vowels in nine Serbian L1 speakers who live in the United States. Even though the period of time that they spent in the USA is significant, I hypothesize that their vowels of English will deviate from the vowel qualities of the speakers in Experiment 1. One of the significant factors influencing such deviations is L1 influence, which in this case is Serbian, one of the Slavic languages. We chose to mark these speakers SS1-SS9.

Similar to Experiment 1, each participant in Experiment 2 was asked to fill in a questionnaire before the recording session started. The Serbian L1 participants were asked to provide basic biographical data as well as the language(s) they speak at home and in their workplace. They were additionally asked to rate their own English fluency on a scale (1-5, 5 being the highest) at the time of relocation from Serbia and at the time of the recordings. Self-reported scores of 8 participants show that they consider themselves highly proficient in English (one participant did not provide data on his English language fluency). They were all born in Belgrade, Serbia (except for one participant, SS5, who was born in the south of Serbia, but lived in Belgrade for 27 years prior to moving to the US). All nine participants lived in Belgrade until they moved to the US. They lived in Atlanta, GA, at the time of the recordings. It may be argued that these speakers belong to a different dialect of American English, Southern American English, which features a number of regional dialectal characteristics. However, I will classify them as speakers of Mainstream American English, similar to the first group because their English did not show any marked regional characteristics. Their age ranges from 35 to 45 (mean age 39.7, median age 40). All experimental subjects had lived in the US for more than 12 years at the time of the recordings. Most speakers' place of residence was Atlanta only. However, SS2 and SS7 also lived outside of Atlanta. Speaker SS2 lived shortly in Augusta, GA, and Macon, GA, and SS7 lived in Greenville, SC, and Seattle, WA. The survey also shows that 6 out of 9 speakers speak English at home alongside with Serbian, and all but one speak exclusively English at work (SS5). Table 3 summarizes this information:

Table 3: Background information on Serbian L1 speakers

Subject	Sex	Age	L2 fluency then/now	Ls of home	Ls of work
SS1	M	40	4/5	Serbian	English
SS2	M	41	3/5	Serbian/English	English
SS3	M	40	2/5	Serbian/English	English
SS4	M	40	¼	English/Serbian	English

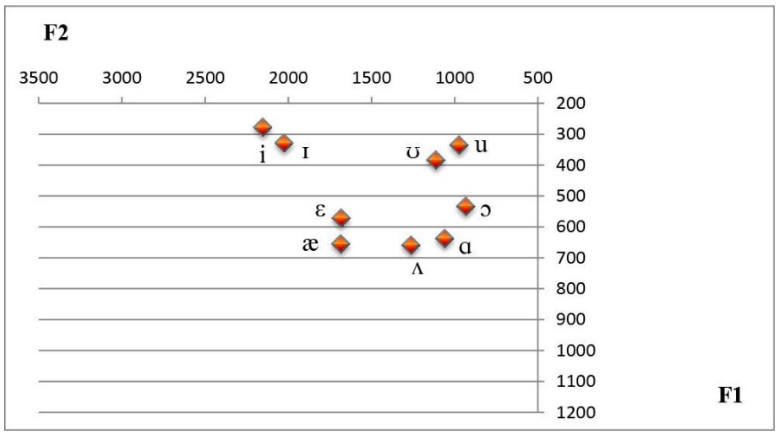
SS5	M	44	2/4	Serbian/English	English/Serbian
SS6	M	37	2/4	Serbian	English
SS7	M	45	¾	Serbian/English	English
SS8	M	36	2/4	Serbian	English
SS9	M	35	N/A	English/Spanish	English

Experiment 2 uses the same set of words, and the same methodology, as Experiment 1. Experiment 2 targets nine nuclei of AE monosyllabic words in a closed syllable: *beat*, *bit*, *bet*, *bat*, *but*, *boot*, *put*, *bought* and *pot*. The words were all recorded in the frame sentence "Say ___ again", repeated three times in random order, giving a total of 243 (9 speakers x 3 repetitions x 9 vowels) tokens for SSs. All recordings were made in a quiet room in Atlanta, GA, using Praat, Version 5.3.51 (Boersma & Weenink, 2013), with noise-cancelling headphones and a laptop computer. Participants read a set of sentences from Power Point slides, where only one sentence was presented at a time. They were also given the opportunity to familiarize themselves with the sentences before the recording started. Some participants needed more time than others, particularly with the word *bat*, that had to be explained by the experimenter so as to bring about the participants' association with Batman, for easier identification. After they had got acquainted with the materials, the participants were instructed to read the sentences "as naturally as possible". Table 4 displays the mean values of the first two vowel formants of Serbian L1 participants:

Table 4. Mean values of F1 and F2 (with standard deviations) of English vowels produced by SSs

	/i/	/ɪ/	/ɛ/	/æ/	/ʌ/	/u/	/ʊ/	/ɔ/	/ɑ/
F1	278 (26)	329 (62)	572 (52)	655 (89)	660 (52)	336 (33)	384 (46)	534 (61)	638 (64)
F2	2152 (116)	2025 (172)	1681 (78)	1684 (73)	1262 (70)	973 (100)	1112 (131)	933 (43)	1060 (124)

Graph 3 presents a two-dimensional diagram based on the acoustic measurements of vowels as realized in Experiment 2:



Graph 3: Vowel data for Serbian L1 group

Dıscussıon

Acoustic Overlaps and Deviations

The acoustic measurements obtained in this vowel study seem to differ in the two groups of participants, as hypothesized. However, some vowels in ESs and SSs resulted in being very similar acoustically. One of such examples is the high vowel /i/, which is alongside with /a/ and /u/ regarded as one of the point vowels in Stevens' Quantal theory of speech (1972). This theory assumes that there are some preferred regions for vowel production. Stevens also claims that the vowels articulated in these regions are best understood, and they are found almost universally in the vowel inventories of the languages of the world (Stevens, 1972, p. 56). This may be regarded as a viable explanation for the similarity between ES /i/ and SS /i/ vowels. These vowels are rendered as peripheral as possible in both participant groups for the English high vowel /i/. However, a slight decline in both formants of /i/ is observed in the Serbian L1 participant group.

The lax /ɪ/ seems more difficult to acquire by Serbian L1 speakers of English. Firstly, the vowel quality of the Serbian short /i/ is different from the English vowel quality. My analysis shows that the F2 of the Serbian L1 participants' /ɪ/ is higher by 461 Hz compared to the F2 mean value found for the ES productions in this study. This implies that the SS /ɪ/ is fronter compared to the ES vowel production, and thus simply closer to the tense /i/ vowel. This may be accounted for by the phonetic interference from Serbian as L1 resulting in more similar articulations of English /i/-/ɪ/ vowel opposition.

Moving on to the English back vowel pair /u/ and /ʊ/, spectral differences in both vowels are found in Serbian as L1 group. Assuming that /u/ is another point vowel, we could expect it to be acoustically similar in both experimental groups due to its universal presence in the languages of the world. However, the results of the present vowel study refute this claim. The F2 of Serbian L1 group /u/ is lower by 204 Hz compared to the other group, which makes it a more peripheral vowel. The ES articulations of /u/ are therefore more centralized. The average F1 value for the SS /u/ is slightly lower, yielding yet again a more peripheral vowel. Even though the two participants groups vary in both F1 and F2 acoustic measurements, their English vowel /u/ may be regarded as acoustically similar.

The lax /ʊ/ presents a difficult case for Serbian L1 group, even though they are long-term residents in the United States. The mean values of both F1 and F2 are lower in SS compared to the ES vowel articulation by 121 Hz and 188 Hz respectively. This means that the SS vowel /ʊ/ is generally higher and backer, and thus closer in quality to the English tense /u/. The vowel opposition /u/-/ʊ/ is evidently not acquired in SS group, similar to the /i/-/ɪ/ vowel pair.

The English /æ/ is considered a new phone to speakers of Serbian and on average it seems well acquired in the SS group. The F1 of SS /æ/ productions is lower than in the ES sample, but their F2 is higher compared to the ES vowel productions. In conclusion, the articulation of SS /æ/ is more fronted and raised in the SS group.

The vowel of *bat* is acoustically similar to the vowel of *bet*, the main underlying difference being in F1 values (101 Hz), whereas the difference in the F2 acoustic measurement seems insignificant (25 Hz) in the ES vowel production. The ES /æ/ is lower compared to /ɛ/. The productions of these two vowels are clearly separated in the vowel space in both experimental groups, but generally the spectral vowel measurements indicate that the SS vowels are slightly higher and fronter.

The acoustic analysis of English /ʌ/ indicates that this vowel is fully acquired (See Tables 2 and 4). Stevens' Quantal theory (1972) may account for the successful acquisition of this vowel as the vowel area where it is produced is considered to be a preferred vowel region.

And lastly, the articulation of the vowels /ɔ/ and /ɑ/ is discussed. In Mainstream American English, the two vowels are undergoing a vowel merger which is now complete in some regions of the United States. This vowel merger, known as a "a low back merger"[2], presupposes the suspension of the vowel contrast between /ɔ/ and /ɑ/. The abolition of the vowel contrast may be explained away as the backing and raising of /ɑ/. The acoustic measurements that were obtained for these two vowels may only be regarded as relative because the two vowels were not pre-classified as belonging to the /ɔ/ or /ɑ/-like vowels. The acoustic analysis of /ɔ/ and /ɑ/ will provide a full overview of all American English monophthongs. The relevant tokens in this vowel study still show an active vowel contrast between /ɔ/ and /ɑ/ in both participant groups. However, the contrast is more clearly preserved in the Serbian as L1 group, which implies a stronger resistance to the vowel merger.

Conclusıons

The acoustic analysis shows that several American English monophthongs are acoustically similar in the two participant groups investigated in this vowel study. The highest acoustic approximation is observed in English monophthongs /ʌ/ and /i/. Due to the fact that these vowels are generally regarded as point vowels in Stevens' Quantal theory of speech (1972) this finding is not surprising. On the basis of this result, it is only natural to presume that a similar finding is seen in the case of /u/, as the third point vowel. However, this is not the case.

The most significant deviations in the English L1 and Serbian L1 groups of speakers are found in the English tense/lax vowel pairs, where the participants of the latter group almost systematically decrease the difference between the two vowels in question. Serbian does not categorically use the quality difference between its long and short vowels, even though there are some implications that such vowels also exhibit spectral differences (Lehiste & Ivić, 1986; Čubrović, 2016).

And last but not least, this acoustic analysis shows that Serbian L1 speakers of English seem to resist more modern features of spoken American English realized in the form of low back merger. More recent developmental changes that are in progress do not seem to affect vowel productions in Serbian L1 speakers of English.

References

Boersma, P. and D. Weenink. (2013). *Praat: Doing Phonetics by Computer*, Version 5.3.51. [Internet]. Available at: http://www.fon.hum.uva.nl/praat/[12.09.2013].

[2] See Čubrović (2018) for a more detailed account of the low back merger.

Bradlow, A. R. (1993). *Language-specific and Universal Aspects of Vowel Production and Perception: A Cross-linguistic Study of Vowel Inventories*. Unpublished doctoral dissertation, Cornell University.

Čubrović, B. (2016). *Acoustic Investigations of Serbian and American English Vowel Inventories*. Belgrade: Faculty of Philology.

Čubrović, B. (2018). The Acquisition of Low Back Merger in Nonnative Speakers of American English". *Nasleđe,* 41, 17–26.

Ivić, P. and I. Lehiste. (1967). Prilozi ispitivanju fonetske i fonološke prirode akcenata u savremenom srpskohrvatskom jeziku III. *Zbornik Matice srpske za filologiju i lingvistiku,* 10, 55–93.

Krebs-Lazendic, L. and C. T. Best. (2013). First language suprasegmentally-conditioned syllable length distinctions influence perception and production of second language vowel contrasts. *Laboratory Phonology,* 4 (2), 435–474.

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

Lehiste & Ivić. (1963). Accent in Serbocroatian: An Experimental Study. *Michigan Slavic Materials*, No. 4. Ann Arbor: Department of Slavic Languages and Literatures.

Lehiste, I. and P. Ivić. (1986). *Word and sentence prosody in Serbo-Croatian*. Cambridge, MA: MIT Press.

Lehiste, I. and G. E. Peterson. (1961). Transitions, Glides and Diphthongs. *The Journal of the Acoustical Society of America,* 33 (1), 268–277.

Maddieson, I. (1984). *Patterns of Sounds, with a chapter contributed by Sandra Ferrari Disner*. Cambridge & New York: Cambridge University Press.

Maddieson, I., K. Precoda and H. Reetz. (2014). UCLA Phonological Segment Inventory Database (UPSID). [Internet]. Available at: http://web.phonetik.uni-frankfurt.de/upsid.html [26.11.2016].

Stevens, K. N. (1972). The quantal nature of speech: evidence from articulatory-acoustic data. In *Human communication: a unified view*, edited by Edward E. David and Peter B. Denes, 51–66. New York: McGraw-Hill.

Wolfram, W. And N. Schilling. (2016). *American English: Dialects and Variation*. Third Edition. Chichester, UK: Wiley Blackwell.

Yavaş, M. (2011). *Applied English Phonology*, 2nd edition. Chichester, UK: Wiley-Blackwell.