Biljana Čubrović University of Belgrade, Serbia 2019, Vol. 16 (1), 15–28(164) revije.ff.uni-lj.si/elope doi: 10.4312/elope.16.1.15–28 UDC: 811.111(73)'243:81'342.41

Duration as a Phonetic Cue in Native and Non-Native American English

ABSTRACT

This vowel study looks at the intricate relationship between spectral characteristics and vowel duration in the context of American English vowels, both from a native speaker (NS) and non-native speaker (NNS) perspective. The non-native speaker cohort is homogeneous in the sense that all speakers have Serbian as their mother tongue, but have been long-time residents of the US. The phonetic context investigated in this study is /bVt/, where V is one of the American English monophthongs /i I u $\upsilon \varepsilon \approx \Lambda \circ \alpha$ /. The results of the acoustic analysis show that the NNS vowels are generally longer than the NS vowels. Furthermore, NNSs neutralise the vowel quality of two tense and lax pairs of vowels, /i I/ and /u υ /, and rely more heavily on the phonetic duration when prononuncing them.

Keywords: vowel; duration; vowel quality; native; non-native; American English

Trajanje kot razločevalna fonetična lastnost v ameriški angleščini rojenih in tujih govorcev

POVZETEK

Pričujoča študija samoglasnikov obravnava zapleteno razmerje med spektralnimi značilnostmi in trajanjem samoglasnikov v ameriški angleščini z vidika rojenih in tujih govorcev. Skupina obravnavanih tujih govorcev je homogena, saj so vsi rojeni govorci srbščine, ki že daljše obdobje živijo v ZDA. Študija proučuje fonetični kontekst /bVt/, kjer je V eden od enoglasnikov v ameriški angleščini /i I u $\upsilon \varepsilon \alpha \land \sigma \alpha$ /. Izsledki akustične analize pokažejo, da so samoglasniki tujih govorcev običajno daljši od samoglasnikov domačih govorcev. Sledi ugotovitev, da tuji govorci nevtralizirajo kakovostno razliko med pari samoglasnikov /i I/ in /u υ /, in da se pri izgovarjavi bolj zanašajo na fonetično trajanje.

Ključne besede: samoglasnik; trajanje; kakovost samoglasnikov; tuji govorci; rojeni govorci; ameriška angleščina



1 Introduction

The vowels of American English (AE) differ in their durations, and the different durations are often said to be phonetically realised as tense or lax. Even though vowel duration in American English has been thoroughly explored, it still offers a fine terrain for further research in several different niches. First, there is intrinsic duration that is studied by Black (1949) and Lehiste and Peterson (1961), who found that open vowels are longer than the vowels produced with a more close jaw opening. To illustrate this point, we can say that the American English vowel / α /, that is more open, is intrinsically longer than the close vowel /i/. Another important and well-investigated phonetic characteristic of English vowels pertains to the duration that is brought into connection with the voicing of the following consonant (House and Fairbanks 1953; House 1961; Chen 1970). The vowel of *bead* is realised as longer than the vowel duration, all else being equal.

Second, some extralinguistic factors may influence vowel duration: vowels produced by female speakers are generally longer than the vowels produced by male speakers (Hillenbrand et al. 1995; Holt, Jacewicz, and Fox 2015). Sociolinguistic influences may also have impact on vowel duration in English. In a recent study of Southern African-American English compared to White American English, Holt, Jacewicz and Fox (2015) claim that African-American speakers in the same geographical area as White Americans produce longer vowels.

2 Earlier Research on the Duration of American English Vowels

One of the early studies that aims at establishing the relationship between frequency, intensity and duration of vowels characteristic of the reading style in American English was carried out by Black (1949). We will here only point to the results that Black (1949) offered in relation to the phonetic duration. A sample of 16 male speakers was analysed in a voiceless phonetic environment, where 11 English vowels were recorded in isolation at a specific pace in the context where the vowel is preceded by /t/ and followed by a voiceless /p/, in short /tVp/. This article provides a good comparison ground against the corpus specifically designed for the purposes of the current study that utilises a similar phonetic context, namely /bVt/. Table 1 provides the averages of the phonetic duration for 9¹ vowels, which are read from a magnetic tape.

TABLE 1. Mean values of vowel duration produced by 16 male speakers (adapted and taken from Black (1949, 218)).

Vowel	i	I	3	æ	Λ	u	υ	э	a
Mean duration in ms	159	135	153	208	154	200	153	209	192

We only give data for the 9 AE vowels that are the subject of the current study.

Peterson and Lehiste (1960) studied the influence of the consonants that precede and follow an AE vowel (a monophthong or diphthong). They used two sets of data, and here we will present the measurements only of the larger set that involves 30 monosyllabic minimal pairs and 10 additional disyllabic minimal pairs (Peterson and Lehiste 1960, 693–94). Table 2 systematises data for the vowels that are the subject of the current study.

TABLE 2. Mean values of vowel duration produced by five speakers (adapted and taken from Peterson and Lehiste (1960, 702)).

Vowel	i	I	3	æ	Λ	u	σ	Э	a
Mean duration									
in ms	240	180	200	330	N/A	260	200	310	260

Peterson and Lehiste (1960) delved further into every individual consonantal context looking at the magnitude of influence of the preceding and following consonants. They arrived at the conclusion that the impact of the preceding consonant on the duration of the syllable nucleus is negligible, but the following consonant has a significant impact on the duration of the vowel. Table 3 provides the mean durations of short and long nuclei in different consonantal contexts. We have singled out only the stop consonants that close the syllable in question. It is evident that the ratio of the duration of vowels in front of voiceless and voiced consonants is approximately 2:3 for American English in favour of voiced segments. The importance of Peterson and Lehiste's research study lies in the fact that they systematically examined all phonetic environments and came to a robust conclusion that voicing is an important cue in distinguishing long and short vowels of AE.

TABLE 3. Mean values of short and long nuclei as a function of English stops (adapted and taken from Peterson and Lehiste (1960, 702)).

	/p/	/b/	/t/	/d/	/k/	/g/
Short	138	203	147	206	145	243
Long	188	307	210	318	200	314

House's study (1961) contributes to the overall phonetic research in the sense that he investigated vowel durations across phonetic contexts, similar to Peterson and Lehiste (1960). Each of the 12 AE vowels is followed either by a voiced or voiceless consonant (three stops, one affricate and three fricatives) in the speech of three male talkers. The vowel speech sound occurs in a stressed syllable of a disyllabic nonsense word. House (1961) provides duration ratios for different phonetic contexts (voiced vs. voiceless consonants that affect vowel durations), different characters of vowels (tense vs. lax), and compares cumulative durations for groups of vowels (close/mid/open tense or lax) in different consonantal contexts (stop, affricate, fricative). House (1961, 1176) finds

that there is a systematic progression in vowel durations depending on vowel features and phonetic environment. Table 4 gives an overview of values resulting from his study.

	Tense	Lax
Close	150	120
Mid	170	
Open	220	150

TABLE 4. Mean duration in ms (rounded to the nearest 10 ms) after voiceless consonants (taken and adapted from House (1961, 1176)).

The longest vowel duration is observed in open tense vowels followed by voiced fricatives -400 ms (stops shorten the vowel duration most, all else being equal, and affricates are positioned in between these two classes of consonants). However, close lax vowels followed by a voiceless stop are characterised by the shortest duration that is only one quarter of that mentioned above, at 100 ms.

TABLE 5. Mean duration in ms (rounded to the nearest 10 ms) after voiced consonants (taken and adapted from House (1961, 1176)).

	Tense	Lax
Close	320	220
Mid	350	
Open	360	250

Hillenbrand, Clark and Houde (2000) came to interesting findings in their study of the effects of duration on vowel recognition in American English. The pairs of vowels including /i–I/, /u– υ / and /I–e– ε / are minimally affected by duration because their spectral features are different enough to distinguish between them, unlike / α – σ – Λ / and /e– ω / that are significantly affected by variable duration.

This paper looks into the vowel duration as a cue that plays a role in distinguishing between American English vowels as produced by two groups of speakers, native and non-native. To the best of our knowledge, this is the first study of its kind that compares the American English spoken by Serbian expatriates with native speakers of this variety. The Serbian vowel inventory is traditionally described as one of the commonest vowel systems: a five-vowel system that comprises /i e a o u/. These five vowels are combined with four pitch accents (long and falling, long and rising, short and falling, and short and rising). Some more recent approaches, starting with Jakobson (1937 [1962]), propose a novel approach to the Serbian vocalic system, where quantity and pitch are factored out as two distinct dimensions. Such a view was adopted in several other studies on Serbian pitch accent (Browne and McCawley 1965, Inkelas and Zec 1988). We will adopt the latter approach in this vowel study and regard Serbian as a quantity

language following Lehiste (1970), who claims that short and long vowels may also differ in their spectral characteristics (see also Čubrović 2016, 26–29).

3 Experiment

3.1 Participants

Ten native speakers of Serbian who live in the United States and five native speakers of American English took part in the experiment. All ten participants are male.

At the beginning of the recording session, each participant was required to fill in a questionnaire. The Serbian participants were asked to report the length of residence (LOR) in the United States and language(s) spoken at home. The Serbian participants were also 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 recording in the States. All ten participants were born in Belgrade, Serbia (except for one participant who was born in the south of Serbia, but lived in Belgrade for 27 years prior to moving to the US), and continued to live in Belgrade until they moved to the States. They all live in Atlanta, GA, and their age ranges from 35–44. Nine of them had lived in Atlanta for more than 12 years at the time of the recording. Seven out of ten speakers mostly speak Serbian at home, and all participants use exclusively English at work.

Native speakers of English were asked to report on their place of residence and languages spoken. All five lived in the North-East of the United States at the time of the recording. Three were undergraduate students at Cornell University, Ithaca, NY, and two were employees (former and present) of the same university. Table 6 summarises this information.

Subject	Sex	Age	L2 fluency (self-evaluated) then/now	Place of residence	Length of residence	Language(s) spoken at home
NNS						
GV	М	40	4/5	Atlanta, GA	12	Mostly Serbian
SG	М	41	3/5	Atlanta, GA	23	Serbian/English
MR	М	40	2/5	Atlanta, GA	14	Mostly Serbian
MS	М	40	1/4	Atlanta, GA	15	Mostly English
IS	М	44	2/4	Atlanta, GA	15	Mostly Serbian
NC	М	37	2/4	Atlanta, GA	16	Mostly Serbian
VG	М	42	2/3	Atlanta, GA	8	Mostly Serbian
NN	М	36	2/4	Atlanta, GA	13	Mostly Serbian
MP	М	35	N/A	Atlanta, GA	14	English/Spanish
UZ	М	45	3/4	Atlanta, GA	16	Mostly Serbian

TABLE 6. Background information on participants.

NS	NS								
MB	М	19	/	New York City, NY	1	English			
ТС	М	70	/	Ithaca, NY	/	English, some French			
PI	М	19	/	Pittsburgh, PA	/	English			
MI	M	73	/	Ithaca, NY	/	English			
NI	М	20	/	Cortland, NY	/	English			

3.2 Materials and Recording Procedures

The acoustic experiment targets nine vowels of American English (AE) in the following monosyllabic words: *beat, bit, bet, bat, but, boot, put, bought* and *pot*. The words were all embedded in the frame sentence "Say _____ again", and repeated three times in a random order, giving a total of 270 (10 speakers x 3 repetitions x 9 vowels) tokens for Serbian NSs and 135 (5 speakers x 3 repetitions x 9 vowels) tokens for English NSs, totalling 405 repetitions.

All NNS recordings were made using Sennheiser noise-cancelling headphones and a Sony laptop computer running Praat, Version 5.3.51 (Boersma and Weenink 2013). The NSs of American English were recorded in a sound-attenuated booth in the Cornell University Phonetics Lab. Participants were given sets of sentences in a Power Point presentation, and only one sentence was presented on a slide at a time. They were also given the opportunity to familiarise themselves with the sentences before the recording started. After they had got acquainted with the materials, the participants were instructed to read the sentences "as naturally as possible".

3.3 Analysis and Discussion

The recordings were digitised at 22,000 Hz and analyzed using the Praat software for acoustic analysis of speech (Boersma and Weenink 2013). All elicited materials were first manually labelled and vowel segmental acoustic features measured with the use of a script written by DiCanio (2013). This script generated eight acoustic measures: vowel duration, F1, F2, F3, centre of gravity, standard deviation, skewness, and kurtosis. Even though duration is in the focus of the present study, F1 and F2 will also be shown in order for vowels to be fully analysed.

The vowels will be analysed in groups in the following sections – /i I/, /u v/, / $\epsilon \propto \Lambda$ /, and /2 a/.

3.3.1 High Vowels /i 1/

The first pair of vowels are those of *beat* and *bit*. In AE, they are most often described as tense and lax, respectively. NSs clearly differentiate them by vowel quality, which is shown in Figure 1.



Figure 1. NS /i i/.

NNSs have a tendency to merge /i/ and /I/. This phenomenon may be accounted for by the fact that Serbian language background speakers rely heavily on the phonetic duration when distinguishing between these two vowels. They transpose this phonetic property from L1 (Serbian) into L2. The merger is not observed in speaker SG, who has the longest LOR in the States (23 years). Similarly, the acoustic characteristics of /I/ of speaker MS, whose LOR is 15 years, approximate the NSs production of this vowel.

The duration measurements for the native and non-native participants are looked at next. The mean value for the tense vowel /i/ in the native speaker group is 112.5 ms (SD 15.1 ms). Its lax counterpart has a mean value of 88.6 ms (SD 14 ms) in the same group of participants.



Figure 2. NNS /i i/.

The non-native speakers' tense vowel is significantly longer at 132.51 ms and with a larger SD (30.05 ms). The non-native speakers' lax vowel duration is only marginally different from the native speakers' average duration measurement at 87.6 ms (SD 20.74 ms).

It seems worth noting the standard deviations that are significantly higher in the nonnative speaker group, which points to vowel duration instability in this participant group.

3.3.2 High Vowels /u v/

The next pair of vowels are the vowels of *boot* and *put*. In AE, /u/ is tense and / υ / is lax. NSs clearly separate the two vowels in the vowel area, which is shown in Figure 3. For native speakers of English there is no overlapping of /u/ and / υ / when vowel quality is concerned.



Figure 3. NS /u v/.

NNS vowels are realised differently and their tendency to merge the two vowels is reported in this vowel study. The nonnative speaker participants consistently produce F1 of $/\upsilon$ / with a lower frequency. This makes the lax $/\upsilon$ / a higher vowel and closer to /u/ in the NNS group. NNSs obviously disregard the quality difference between the two English vowels and rely more on the phonetic duration, similar to the /i/-/I/ pair, which is displayed in Figure 4 below.

As for the duration measurements, NS tense /u/ is on average 125 ms long (SD 15.8), while its lax counterpart is 73.8 ms long (SD 9.9 ms). NNSs tend to produce the tense vowel as even longer. The mean value of /u/ in the NNS group amounts to 154.5 ms (SD 35.9 ms). The lax / υ / is realised as significantly shorter, and its average value in the NNS group is 80.3 ms (SD 12). To conclude, NNSs seem to rely on the phonetic duration when producing the AE tense /u/. The SD for the



Figure 4. NNS /u v/.

NNS vowel is also larger when compared to the NS realisation of this vowel which points to the vowel duration instability and hesitation on the part of the NNSs.

3.3.3 The Vowels $/\epsilon \propto \Lambda/$

In the next group of vowels, we first investigate the phonetic characteristics of two vowels $|\varepsilon|$ and $|\omega|$. It has been noted that there is more variation in the acoustic vowel space for $|\varepsilon|$ and $|\omega|$ even in the group of NSs. There is a general tendency, though, for $|\varepsilon|$ to be produced with a lower F1, which makes it a higher vowel than $|\omega|$ in AE native speech. The three tokens of all three vowels that have consistently lower values of F1 are all produced by speaker MB. This may be due to his vocal tract length, which can be longer in tall people. This reduction in F1 values in one speaker may be seen as his idiosyncratic characteristic.

In order to provide a comprehensive overview of AE monophthongs in this paper, the vowel $/\Lambda/$ is also displayed on Figure 5, but it clearly does not overlap with $/\epsilon \alpha/$ in the native speaker group.

The NNS participants in this study tend to merge $/\epsilon/$ and $/\alpha/$, i.e. they do not clearly differentiate between the two. The vowel $/\alpha/$ is a new sound to Serbian language speakers. A similar finding can be observed in other Slavic languages. Slovenian learners of English, for example, find this vowel contrast in BE the most challenging (Stopar 2015, 89; Komar 2017, 163). However, there are nine tokens of the *bat* vowel in Figure 6 that have higher F1 values and they are repetitions of three speakers, SG, NN and UZ. Two of these participants have a relatively long LOR in the States, of 23 and 16 years. The speaker NN has lived in the States for 13 years, which is slightly under the mean value for LOR in the study (13.5 years). These nine tokens of *bat* approximate the NS spectral characteristics.







Figure 6. NNS / $\epsilon \approx \Lambda$ /

As for vowel duration, the three English vowels are given here in ascending order, from the shortest to the longest: $\epsilon \propto \Lambda$. The average values for the NS vowels are consistently shorter than for the NNSs' ones. As expected, the standard deviations are larger in the NNS group. All duration measurements for the three vowels are given in Table 7.

NS	3	æ	Λ
Mean	110.4	161.6	100.6
SD	10.7	13.6	11.7
NNS	3	æ	Λ
Mean	120.8	173.1	116.2
SD	16.3	31.1	21.8

TABLE 7. Duration measurements for $/\epsilon \propto \Lambda/$.

3.3.4 The Vowels /o a/

Both /3/ and $/\alpha/$ are described as back vowels in AE. Neutralisation of these two vowels is observed in many regions of the US and Canada, and is known as *cot-caught* merger. As a result of this phonological process, the two vowels become one (see Čubrović 2017; 2018).

The NS participants of this study mostly differentiate the vowels in question, i.e. the vowels in the words *bought* and *pot*. The NNSs utilise the same speech habit as NSs. However, the /ɔ/ vowel is articulated with a lower F2 by most subjects in the NNS cohort. This implies that the NNS /ɔ/ is a more back vowel than in native AE speech.

The vowel of *pot* shares the same phonetic characteristics as /a/ in the two groups of participants of this study. The values for both formants are lower in NNSs. This vowel is, therefore, produced as a higher vowel and with a greater degree of backness, see Figures 7 and 8.



Figure 7. NS /9 a/.



Figure 8. NNS /ɔ a/.

A brief look at the duration measurements provides an interesting finding: the vowel of *bought* is significantly longer in the NNS group at the mean value of 199.5 ms (SD 56.9 ms) compared to the NSs' mean value of 156 ms (SD 14 ms). This is consistent with the findings for the tense /i u/, where NNSs also rely heavily on vowel length.

The vowel of *pot* is shorter in the NNS group (117.3, SD 23) than in the NS group of participants (123.2 ms, SD 13.7), which runs counter to the general rule that NS vowels are consistently longer compared to NNS vowels.

4 General Discussion and Conclusions

The tables that follow summarise the mean values of the vowel duration for all nine vowels investigated in this research study in both NS and NNS groups. The non-native speaker group data is given first, followed by the average duration measurements for the native speaker group. The third row displays the duration ratio for the two groups of participants (NNS vowel duration/NS vowel duration).

	i	I	3	æ	Λ	u	σ	э	a
NNS	132.5	87.6	120.8	173.1	116.2	154.5	80.3	199.5	117.3
NS	112.5	88.6	110.4	161.6	100.6	125	73.8	156.2	123.2
	1.17	0.98	1.09	1.07	1.15	1.23	1.08	1.27	0.95

TABLE 8. Duration ratio.

The vowel duration ratio analysis shows that NNS vowels are consistently longer in duration than their NS counterparts, with the exception of /1/ and / α /. This is also shown in Figures 9 and 10.



FIGURE 9. NS vowel duration with SD.

One of the significant findings of this small-scale vowel study is the neutralisation of vowel-quality of several vowels in the NNS production, which leads to a heavy

reliance on phonetic duration as a single, most important phonetic cue. The vowels /i/ and /u/ are reported to undergo a sort of a spectral merger in the NNS production, therefore NNS have to rely more heavily on the phonetic duration. It is a matter of debate why NNS articulate /o/ as a very long vowel in AE. One of the reasons might be the influence of British English that these speakers were taught at school, where /o:/ and /v/ form a vowel pair, similar to /i: I/ and /u: v/.



FIGURE 10. NNS vowel duration with SD.

Last but not least, the open vowel $/\alpha$ / is intrinsically long so both native and nonnative speakers produce it this way. As a new sound to the NNS group, its vowel quality is somewhat more difficult to acquire, but its universal vowel duration is a good place to start.

References

- Black, John W. 1949. "Natural Frequency, Duration, and Intensity of Vowels in Reading." *Journal of Speech and Hearing Disorders* 14: 216–21. https://doi.org/10.1044/jshd.1403.216.
- Boersma, Paul, and David Weenink. 2013. *Praat: Doing Phonetics by Computer*. Version 5.3.51. Accessed June 20, 2013. http://www.praat.org/.
- Browne, E. Wayles, and James D. McCawley. 1965. "Srpskohrvatski akcenat." *Zbornik za filologiju i lingvistiku* 8: 147–51.
- Chen, Matthew. 1970. "Vowel Length Variation as a Function of the Voicing of the Consonant Environment. *Phonetica* 22: 129–59. https://doi.org/10.1159/000259312.
- Čubrović, Biljana. 2016. *Acoustic Investigations of Serbian and American English Vowel Inventories*. Belgrade: Faculty of Philology.
- —. 2017. "Low Back Merger in Native and Non-Native Speakers of American English." The Linguistics Journal 11 (1): 222–31.
- 2018. "The Acquisition of Low Back Merger in Nonnative Speakers of American English." Naslede 41: 17–26.
- Dicanio, Christian. 2013. "Script." Accessed February 16, 2019. http://www.acsu.buffalo.edu/~cdicanio/ scripts/Vowel_Acoustics_for_corpus_data.praat.
- Hillenbrand, James M., Laura A. Getty, Michael J. Clark, and Kimberly Wheeler. 1995. "Acoustic Characteristics of American English Vowels." *The Journal of the Acoustical Society of America* 97 (5): 3099–3111. https://doi.org/10.1121/1.409456.

- Hillenbrand, James M., Michael J. Clark, and Robert A. Houde. 2000. "Some Effects of Duration on Vowel Recognition." *The Journal of the Acoustical Society of America* 108: 3013–22. https://doi. org/10.1121/1.1323463.
- Holt, Yolanda Feimster, Ewa Jacewicz, and Robert Allen Fox. 2015. "Variation in Vowel Duration Among Southern African American English Speakers." *American Journal of Speech-Language Pathology* 24: 460–69. https://doi.org/10.1044/2015_AJSLP-14-0186.
- House, Arthur S. 1961. "On Vowel Duration in English." *The Journal of the Acoustical Society of America* 33 (9): 1174–78. https://doi.org/10.1121/1.1908941.
- House, Arthur S., and Grant Fairbanks. 1953. "The Influence of Consonant Environment upon the Secondary Acoustical Characteristics of Vowels." *The Journal of the Acoustical Society of America* 25 (1): 105–13. https://doi.org/10.1121/1.1906982.
- Inkelas, Sharon, and Draga Zec. 1988. "Serbo-Croatian Pitch Accent: The Interaction of Tone, Stress, and Intonation." *Language* 64 (2): 227–48.
- Jakobson, Roman. (1937) 1962. "On the Identification of Phonemic Entities." In *Selected Writings I*, 418–25. The Hague: Mouton.
- Komar, Smiljana. 2017. "The Relationship between the Perception and Production of Four General British Vowels by Slovene University Students of English." *Linguistica* 57 (1): 161–70. https://doi. org/10.4312/linguistica.57.1.161-170.
- Lehiste, Ilse, and Gordon E. Peterson. 1961. "Transitions, Glides and Diphthongs." *The Journal of the Acoustical Society of America* 33 (3): 268–77. https://doi.org/10.1121/1.1908638.
- Lehiste, Ilse. 1970. Suprasegmentals. Cambridge, MA: MIT Press.
- Stopar, Andrej. 2015. "Perception of Four General British Vowels by Slovenian University Students of English as a Foreign Language." In *PTLC2015: Proceedings of the Phonetics Teaching and Learning Conference, London*, 5–7 August 2015, 87–90. London: Phonetics Teaching and Learning Conference.