New Vowel Category Acquisition in L2 Speakers of English: The Case of High Front and High Back Vowels

Biljana Čubrović, University of Belgrade, Serbia

Abstract

This article aims to look into the strategies that L2 learners of English with a Serbian language background develop in the acquisition of those pairs of English vowels whose qualitative characteristics are markedly different in English, but essentially the same in Serbian. The production experiment focusses on two groups of English speakers, L1 and L2, with the aim of comparing the English high front FLEECE/KIT vowel pair and high back GOOSE/FOOT vowel pair. We analyse the vowel production of five L2 speakers of English whose L1 is Serbian, and five L1 speakers of Mainstream American English. The investigation is centred around the extent to which the F1–F2 difference typical of English vowels is acquired by proficient L2 speakers. The results of the acoustic analysis show that the qualitative difference is acquired in the L2 speaker group between FLEECE/KIT and GOOSE/FOOT vowels, but also that L2 speakers rely on different strategies in the formation of new vowel categories.

Keywords: L2 vowel acquisition, high front vowels, high back vowels, English-Serbian analysis

1 Introduction

The vowel inventory of English is famously complex in both Mainstream American English (MAE) and Southern British Standard (SBS) pronunciation models. The vowel inventory of MAE is known to be less numerous, but its enormous dialectal diversity complicates this matter somewhat. Due to the increasing exposure of Serbian L1 speakers to MAE rather than SBS in recent decades, this paper is based on the assumption that Serbian EFL learners' vowels of English are comparable to those of MAE.

Most phonetic research carried out in the context of Serbian speakers of English used SBS as a desirable target for their EFL learners (Paunović 2002; Marković 2007; Dančetović 2017; Bjelaković 2018). More recently Bjekić (2022) tackled the issue of L2 vowel quality acquisition of Serbian EFL learners using the MAE vowel inventory as a target, which is a novel approach.

The English vowels whose quality proves difficult to acquire from the standpoint of Serbian EFL speakers are the vowels of FLEECE and KIT, as well as those of GOOSE and FOOT, among others. This paper examines the spectral features (F1–F2) of the four English vowels in order to find out whether the new phonological categories have been formed in English as L2 in a group of proficient EFL speakers of English whose L1 is Serbian.

2 Theoretical background

The relationship between vowel quality and vowel quantity in the languages of the world is an intricate one, and L2 learners face obstacles throughout the vowel learning process. Even in the case of quantity languages like Serbian, phonetic matters are not simple. Serbian vowel pairs like /e, e:/ and /o, o:/ clearly manifest a heavy influence of quantity on vowel quality. Other vowel pairs do not. Another question that is raised is the matter of which acoustic cue is stronger, primary and more influential: quality or quantity, and whether this is in any way predictable in any given language.

Serbian is traditionally described as a language that has five vowel pairs distinguished by phonological length. This would essentially mean that the two high front vowels of *pîta* (Eng. *(he) asks*) and *pìta* (Eng. *a pie*) have virtually the same spectral features, where the former is long and the latter vowel is short. The same applies to the two high back vowels, e.g. *rûka* (Eng. *an arm*) and *rùta* (Eng. *route*), where the difference once again is explained as a quantitative one,

without differences in the vowel quality. Lehiste (1970, 31) and Lehiste and Ivić (1986) claim that /i/ and /u/ do not show a marked influence of quantity on vowel quality and that the short and long categories of these vowels are distinguished by duration alone. The Serbian short /i/ and short /u/ vowels are not centralized and lowered in the vowel space in relation to their long counterparts. On the other hand, the qualitative difference between the vowels of *beat* and *bit*, or *food* and *foot*, in English is significant, alongside the quantity distinction. EFL learners of various language backgrounds find the English vowel contrasts challenging to acquire because, unlike their L1, English combines spectral cues with duration to form a single vowel category. Different mechanisms and strategies may be used in the process of vowel category acquisition in English depending on several factors, such as the linguistic experience of speakers whose vowel properties are studied, as well as their L1 background.

Spanish learners of English are similarly presented with difficulty when acquiring the vowel contrasts not found in Spanish as their L1. Casillas (2015) studied the production of the FLEECE/KIT pair in early and late learner groups and found that the vowel contrast was fully acquired in the group of early English language learners. The finding for late learners suggests that the contrast was not produced categorically, and that duration is a more salient acoustic cue than the F1–F2 spectral properties. Escudero and Boersma (2004, 580) found that "beginners seem to have trouble with the length distinction, whereas more experienced learners have developed a lexical length contrast". This implies that duration may be regarded as the primary acoustic cue that L2 learners resort to initially, and that the spectral features are acquired at a later stage.

Brazilian Portuguese learners are reported to struggle with the acquisition of the vowel quality of English high front vowels, as evidenced by Roberto Gonçalves and Silveira (2014). These vowels remain a challenge even for more proficient EFL learners who mostly rely on the quantity difference which is used categorically in Brazilian Portuguese.

Japanese and English differ markedly in the use of quality and quantity in producing L1 vowel distinctions (Hirata and Tsukada 2004). Oh et al. (2011) confirm that Japanese learners of English predominantly struggle with the lax vowel group of American English.

3 Methodology

3.1 Participants

Ten speakers took part in the experimental vowel study. The recordings were made in a sound-attenuated booth at the Cornell University Phonetics Lab and at the Belgrade Phonetics Lab using Praat on a Sony VAIO laptop computer. The experimental procedure is replicated from Čubrović (2016), this time with the aim of investigating and comparing L1 and L2 high front and high back vowels.

Five male speakers of MAE were recorded in part 1 of the experiment. Before the recordings were made, participants were asked to fill in a short questionnaire, which included questions related to personal data (age, place of birth, current and previous places of residence, and languages spoken at home). These speakers are marked as E1–E5 and their important data is shown in the table below.

Table 1. Basic information on the L1 MAE speakers.

Speaker	Age	Place of birth	Language(s) spoken at home
E1	19	New York City, NY	English
E2	20	Cortland, NY	English (some Dutch and Frisian)
E3	20	Haverhill, MA	English
E4	21	Columbia, MD	English
E5	21	Manhasset, NY	English

As can be concluded from Table 1, all experimental subjects are predominantly monolingual speakers born and raised in the American Northeast, with the exception of E2, who has one parent who is also a speaker of Dutch/Frisian. All speakers mostly use English in their everyday communication. All participants were also learners of foreign languages, and had exposure to these in a formal, classroom context. At the time of the recording, all speakers lived in Ithaca, NY.

The second group of speakers, who are native speakers of Serbian and proficient speakers of English, took part in the same experiment. This sample was deemed a representative sample of L2 MAE speakers.

Table 2. Basic information on L2 MAE speakers.

Speaker	Age	Place of birth	Language(s) spoken at home
S1	22	Belgrade, Serbia	Serbian
S2	21	Belgrade, Serbia	Serbian
S3	22	Belgrade, Serbia	Serbian
S4	21	Belgrade, Serbia	Serbian
S5	21	Belgrade, Serbia	Serbian

The group of L2 MAE is monolithic in the sense that they all reported they spoke MAE, but had not lived or spent any time in the areas where MAE has an L1 status. All five participants were also English majors at a public university in central Serbia. Their self-reported level of English was C1 at the time of the recordings. The experimenter verified that the L2 group was leaning towards MAE.

3.2 Materials and recording procedures

The acoustic experiment investigated the spectral features (F1 and F2) of four monophthongs of MAE in the following monosyllables: *beat*, *bit*, *boot*, and *put*. In addition to the four words listed above, included in the recordings were also the following tokens: *bet*, *bat*, *but*, *bought*, *pot*, *boat* and *bait*. These played the role of distractors. All eleven monosyllables share a characteristic CVC structure, with an initial labial consonant (voiced or voiceless) and a final coronal consonant /t/ so as to eliminate any potential effects of different places or manners of articulation. Hillenbrand at al. (2001) studied the effects of consonantal environment in English and observed highly significant effects of the phonetic environment, which has been avoided in the current vowel study by maintaining the same place of articulation of the final consonant. The initial consonant is not expected to exert any influence on the vowel quality.

The selected word forms were imbedded in the carrier sentence "Say _____ again". The utterances were recorded three times, in random order. The total number of utterances amounts to 330 (10 speakers x 3 repetitions x 11 word forms), 165 for L1 MAE group and 165 for L2 MAE group. As this paper focusses on the vowels in *beat*, *bit*, *boot*, and *put* , the total number of tokens analysed for the purposes of further analysis was 30 per vowel.

Participants were presented with the utterances on the computer screen, one at a time, and the pace of recordings remained stable. Once one carrier sentence was pronounced, the experimenter would change the slide that displayed the next token. Before the recordings were made, the participants were given instructions about the experimental procedures and provided time to familiarize themselves with the recording materials. After the short preparation stage, participants were asked to read the sentences as naturally as possible. The experimenter's task was to follow the recording level throughout the recording session so as to avoid any undesirable weak or overloaded acoustic signals that would impede acoustic analysis.

The MAE vowel inventory consists of eleven different segments, /i ɪ e ɛ æ ʌ u ʊ o ɔ ɑ/ (Yavaş 2011, 77–78), as in the following words *beat*, *bit*, *bait*, *bet*, *bat*, *but*, *boot*, *put*, *boat*, *bought*, and *pot*, respectively. The vowels of *bait* and *boat* may be diphthongized, even though they essentially belong to the category of monophthongs. The vowel inventory of MAE typically contains three diphthongs, as in *bite*, *bout* and *void* (Yavaş 2011, 78). Table 3 lists all the tokens recorded, but the ones marked bold were subjected to further acoustic and statistical analysis.

The full list of the words recorded is given in Table 3.

Table 3. English word list.

Word form	**MAE target vowel**	**Consonantal context**
beat	/i/	Labial_Coronal
bit	/ɪ/	Labial_Coronal
bait	/e/	Labial_Coronal
bet	/ɛ/	Labial_Coronal
bat	/æ/	Labial_Coronal
but	/ʌ/	Labial_Coronal
boot	/u/	Labial_Coronal
put	/ʊ/	Labial_Coronal
boat	/o/	Labial_Coronal
bought	/ɔ/	Labial_Coronal
pot	/ɑ/	Labial_Coronal

4 Analysis and discussion

The recordings were digitized at 22,000 Hz and labelled manually in Praat (Boersma and Weenink 2013). The spectral properties of vowels were extracted with the help of a Praat script (DiCanio 2013). Those formant measurements that deviated from the expected values underwent manual checking, and were corrected where generated erroneously. The number of mistracked formants was negligible.

F1–F2 graphs were formed so as to examine the vowel space characteristic of L1 MAE vowels in relation to those of L2 MAE. The first formant (F1) is inversely related to the vowel height, whereas the second formant (F2) relates to the degree of backness, e.g. the fronter the vowel, the higher its F2. As part of the F1–F2 graphs that follow, F1 is plotted on the vertical axis and F2 on the horizontal one, so these resemble the vowel diagrams that are traditionally used in articulatory phonetics. Each point in the F1–F2 diagram represents one repetition of one word token. Formant values were not normalized due to the fact that all speakers are male.

We first plotted F1 and F2 measurements for the L1 group (with one standard deviation) to show how short and long vowels spread in the vowel space, and to examine the vowel area for each of the four vowels studied. The graphs were made using NORM (Thomas and Kendall 2007). The acoustic data in Figure 1 shows that L1 speakers employ a specific area in the vowel space for each of the four vowels, and that there are no overlaps between the comparable pairs, i.e. *beat* vs. *bit* or *boot* vs. *put*.

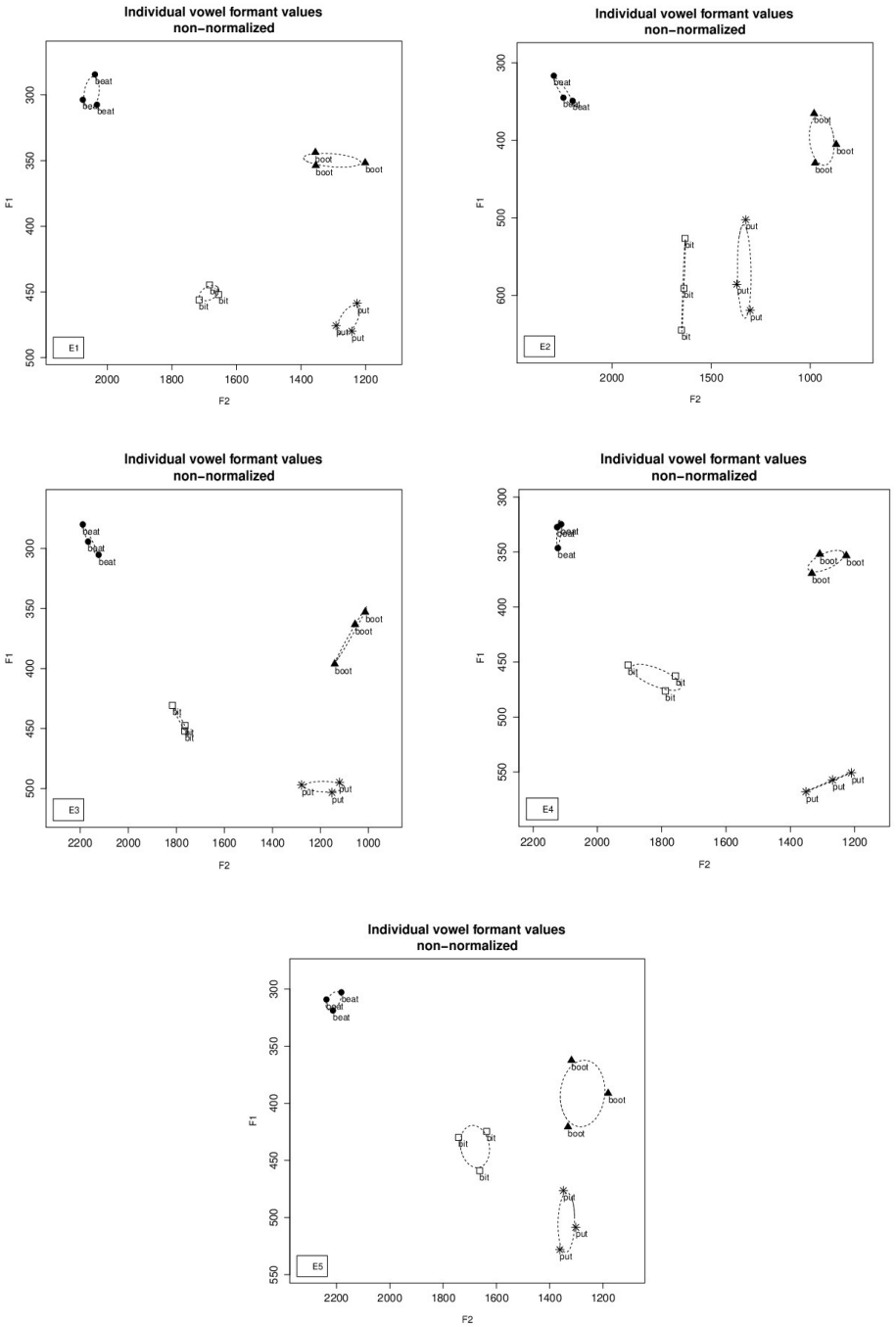

Figure 1. L1 high front and high back vowels for individual speakers (E1–E5).

Even though the four vowels are clearly separated in the vowel space for all five L1 speakers, some individual differences are observed. We will now look at the vowel space as used by all five individual speakers, E1–E5, and comment on any variations or similarities. The vowel of FLEECE is the most stable of all four and shows similarities with regard to frontness expressed by F2, which ranges from 2,112–2,294 Hz in speakers E1, E2, E3 and E4. The second formant of E1 speaker's FLEECE vowel is only slightly lower (F2 range is 2,038–2,075 Hz), which makes the vowel more peripheral and fronter. The F1 measurements for all five speakers are compact, ranging from 279–346 Hz.

The KIT vowel acoustic realizations seem similar in speakers E1, E3, E4 and E5 with regard to comparable F1 values (ranging from 424–476 Hz). Speaker E2 centralizes the KIT vowel more, which is marked by higher F1 values (526–644 Hz). The F2 range of KIT vowel is dispersed along the scale marking the degree of frontness (1,636–1,904 Hz). Speakers E3 and E4 have somewhat higher F2 values reducing the phonetic distance between FLEECE and KIT vowels.

The GOOSE vowel shows markedly more variation along the F2 scale, whereas measurement stability dominates in the F1 range (351–429 Hz). This implies that speakers' high back vowel varies on the degree of backness axis (F2 868–1,354 Hz). Speaker E2 produces the GOOSE vowel highest, followed by speaker E3, and at the other end of the backness scale speakers E4 and E5 shift their GOOSE vowel tokens to the central area of the vowel space.

The FOOT vowel is realized differently in the L1 speaker group. Speaker E2 pronounces it as a lower vowel and centralizes is more than the other speakers, thus bringing it closer to the KIT vowel in the front vowel area. The other four speakers have a tendency to use a similar range of F2 for KIT vowels, which to a certain extent overlaps with the F2 of GOOSE vowel. This finding results in the conclusion that for four out of five speakers in the L1 group the height of the tongue is a distinctive factor in the GOOSE/FOOT opposition.

Next we look at the four vowels as produced by five L2 speakers of English, marked S1–S5. Figure 2 shows that all five speakers have formed separate categories for the English vowels of FLEECE vs. KIT and GOOSE vs. FOOT, but also used different strategies in the new vowel category formation.

Speaker S1 has formed four different phonetic categories, but the distance in the vowel space between the long and short vowels is minimal. This speaker relies more heavily on vowel duration in distinguishing the English vowel pairs FLEECE/KIT and GOOSE/FOOT. This strategy may be attributed to the

transfer from Serbian as L1. The remaining four L2 speakers have formed separate categories for the four vowels in spite of the fact that Serbian, their mother tongue, does not recognize these.

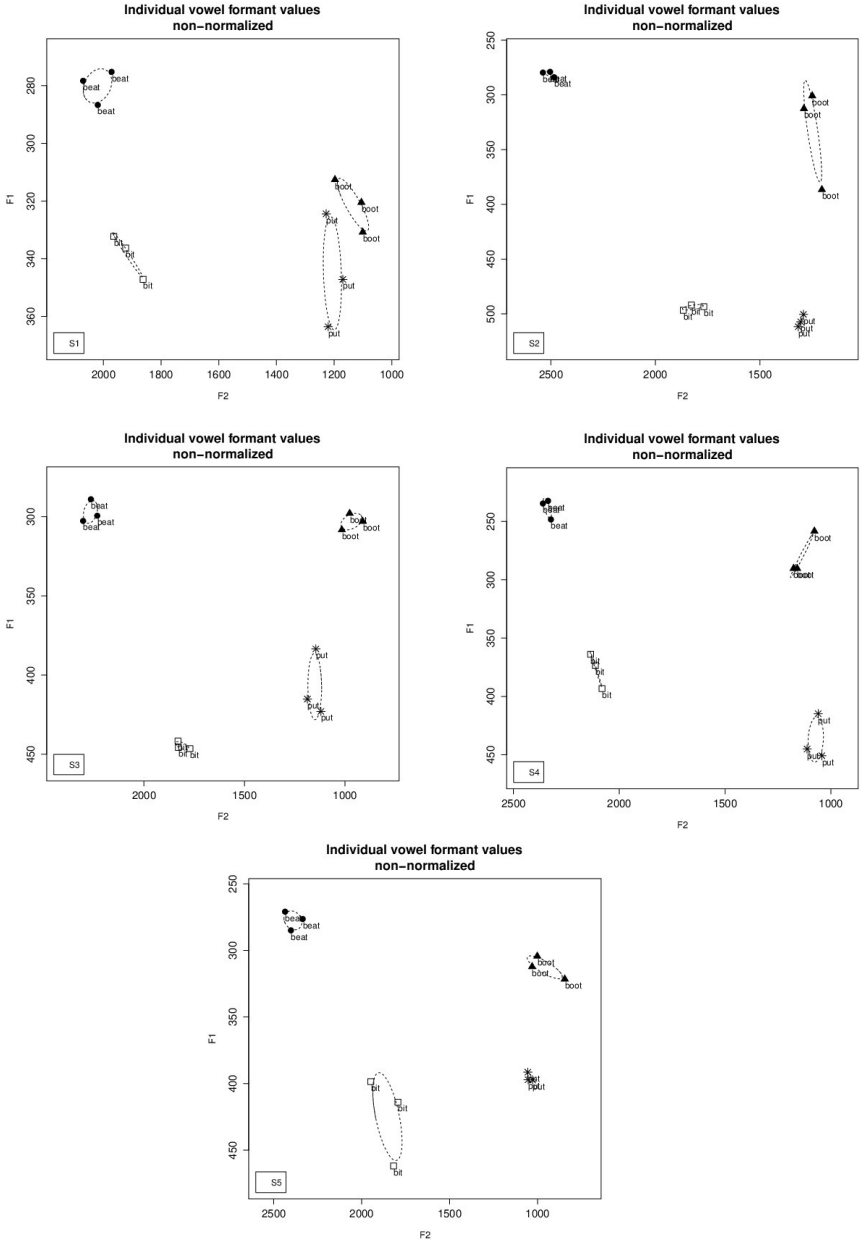

Figure 2. L2 MAE vowels.

It is commonplace to say that distinctions between short and long pairs of English high front and high back vowels are hard to acquire even for advanced EFL learners. These speakers typically find the quality of short vowels /ɪ ʊ/ especially burdensome. However, Figure 2 shows that the two high front vowels are well separated in the vowel space for the L2 speaker group.

For purposes of further discussion about any deviations from the L1 vowels, Figure 3 displays the L1 and L2 high front vowels. F1 values of the vowels of *beat* and *bit* for the L2 speaker group have somewhat higher values, which implies that the vowel itself is fronter and more peripheral in the vowel space in the speech of Serbian speakers of English. A high degree of dispersion on the F2 plane is observed in the L2 productions of the vowel of *bit* (from 1,767 Hz to 2,135 Hz). Speakers S2 and S3 have fully acquired the L1 vowel quality, which is clearly shown in Figure 2 above. All other L2 speakers produce a qualitative difference between the high front vowels, but their *bit* vowel is less centralized compared to the L1 vowel quality. Speaker E2 who belongs to the L1 group has distinctly higher values of F1 which may be accounted for by vowel lenition, as shown in Figure 1.

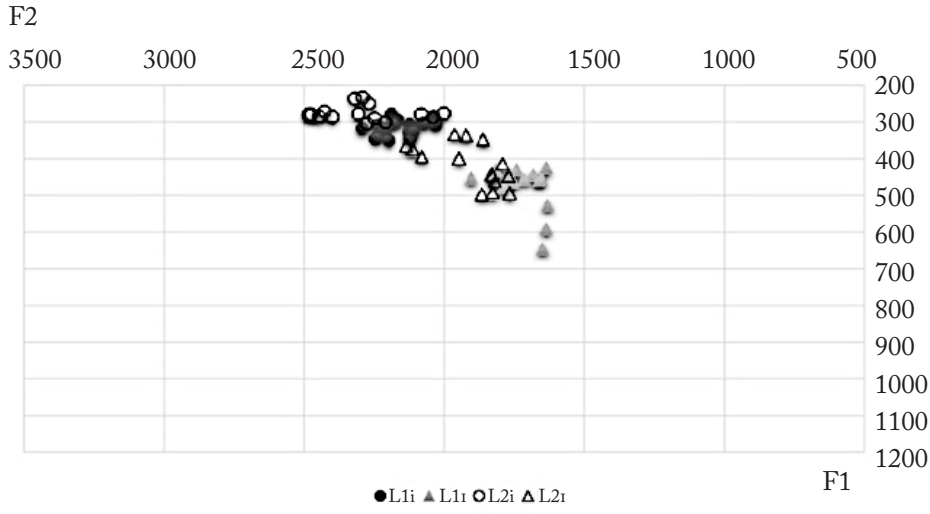

Figure 3. High front vowels in MAE L1 and L2.

The next pair of vowels to analyse are those of *boot* and *put*. Their spectral measurements are plotted in Figure 4. The L2 vowel of *boot* is characterized by somewhat lower F1 values, which points to an L2 vowel that is a higher vowel than its L1 counterpart. Some L2 tokens of the vowel in *put* evidently manifest overlapping with the L1 *boot* vowel, whereas only one L2 speaker acquired the vowel quality of the L1 *put* vowel (speaker S1). Figure 4 shows a certain degree of variability in both speaker groups. In conclusion, the L2 group has also formed a new vowel category for the FOOT vowel.

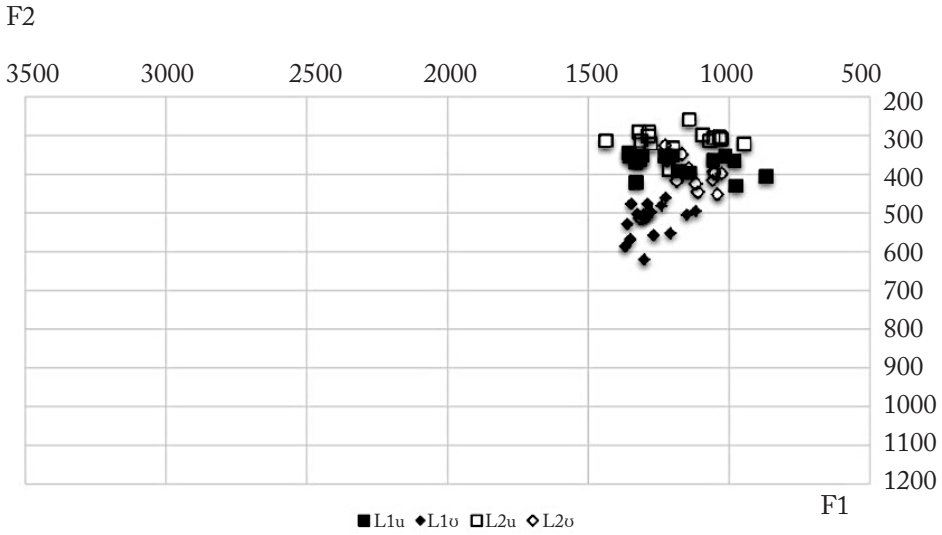

Figure 4. High back vowels in MAE L1 and L2.

4.1 Statistical analysis and discussion

In order to establish any differences in the vowel quality between the L1 and L2 speaker groups, a mixed-effects statistical model was run on the experimental data. The analysis was performed in R statistical software (2013) with the lme4 package (Bates et al. 2015). A separate model was run for each formant (F1 and F2) for each of the four monophthongs of English, with *Speaker Group* (L1 and L2) as a fixed effect and *Speaker* as a random effect.

The model returned statistically significant differences of speaker group on F1 for /i/ and on F2 for /i ɪ/. For the purposes of this paper, a p-value less than 0.05 (typically ≤ 0.05) in at least one of the formants (F1 or F2) was considered statistically significant. The summary of statistical findings is provided below in Table 4.

Table 4. Results of the mixed-effects model between L1 and L2 groups for F1 and F2.

Vowel	Pr(>|t|) F1	Pr(>|t|) F2
i	**0.00827 ****	**0.00707 ****
ɪ	0.0534	**0.0181 ***
u	0.126	0.104
ʊ	0.631	0.176

Significant codes: '**' 0.01 '*' 0.05

According to the results of the statistical analysis, the following vowels in the two groups of speakers do not show statistically significant differences: the vowels of *boot* and *put*. These vowel qualities may be rendered as fully acquired MAE vowels by the L2 group. The findings of this vowel study are strikingly different from a similar but more comprehensive study (Čubrović 2016) where the L2 speaker group consisting of long-term bilingual speakers of MAE who reside in the US only acquired the vowel of *but* of the nine monophthongs studied, but not one of the four vowels that are the focus of the present study. Bearing the two speaker groups in mind, the L2 group taken as a sample in the current paper is formed by undergraduate students majoring in English language, literature and culture. Bjekić and Čubrović (2021, 76) studied the MAE monopthtongs in a comparable experimental study with less advanced EFL speakers from another city in central Serbia, and found that "there is a significant difference in F1 and F2 between native and nonnative speakers for all English vowels except /i/". To sum up, trained language students performed better compared to the diaspora group or the less proficient EFL student group.

Figure 5 displays the F1 values with one standard deviation.[1] It can be seen from the standard deviation values in the graph that both L1 and L2 groups manifest a certain degree of variability.

1 The means of F1 and F2, and SD are provided in the Appendix.

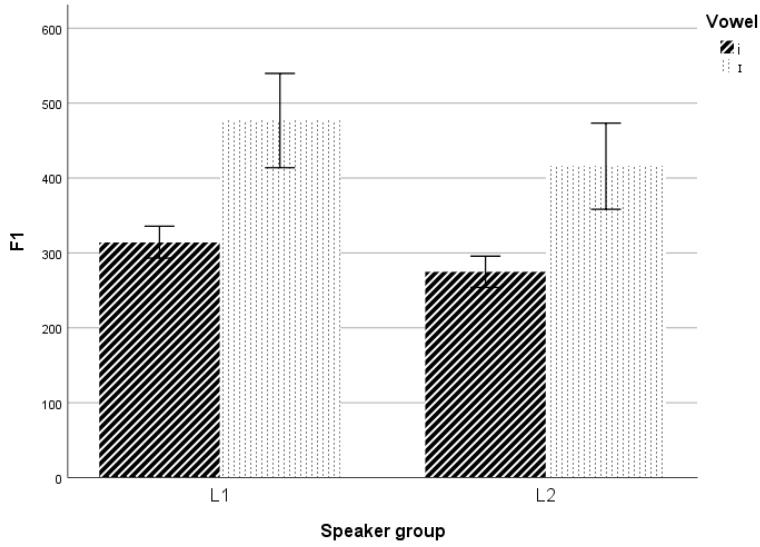

Figure 5. F1 in two speaker groups.

Figure 6 shows the values of F2 in the two groups of speakers, L1 and L2. More variability in F2 is evident in the L2 group as compared to the L1 group. This result may be attributed to high vowel variability and varying degrees of lenition in the GOOSE/FOOT vowels.

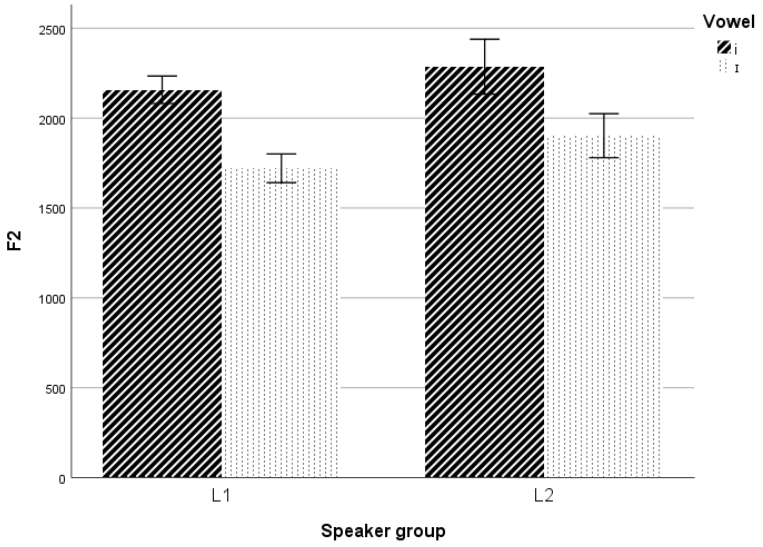

Figure 6. F2 in two speaker groups.

5 Conclusion

The two experiments in the present study add to the extensive body of acoustic research of L1 and L2 American English vowels, especially in the area of the production of high front and high back vowels. The results show a marked vowel variability in L1 American English vowels. The L2 speaker group, which included advanced speakers of English with Serbian as L1, has successfully formed new vowel categories for the KIT and FOOT vowels that do not overlap with the FLEECE and GOOSE vowels. However, the strategies used in the formation of new vowel categories vary in the L2 group, with at least one L2 speaker who seems to rely more heavily on vowel duration, a phonetic habit that has been transferred from Serbian as L1. The spectral analysis shows that the L2 group produced the GOOSE/FOOT contrast in a native-like fashion, i.e. that their productions did not differ from the L1 group with regard to F1 or F2 for each of the two vowels. The FLEECE and KIT vowels in the L1 and L2 groups still have some way to go before they are fully accommodated into the English vowel inventory.

References

Bates, Douglas, Martin Mächler, Ben Bolker, and Steve Walker. 2015. "Fitting linear mixed-effects models using lme4." *Journal of Statistical Software* 67 (1): 1–48. https://doi.org/10.18637/jss.v067.i01.

Bjekić, Ana. 2022. "Acoustic analysis of American English vowels in the production of native speakers of Serbian." [In Serbian]. PhD diss., University of Belgrade.

Bjekić, Ana, and Biljana Čubrović. 2021. "Acoustic characteristics of American English monophthongs in Serbian EFL speakers – A case study." *Philologia* 19 (1): 65–84. https://doi.org/10.18485/philologia.2021.19.19.6.

Bjelaković, A. 2018. "Formant measurement of Serbian speakers's English vowels." *Philologia* 16: 17–43. https://doi.org/10.18485/philologia.2018.16.16.2.

Boersma, Paul, and David Weenink. 2022. "Praat: Doing Phonetics by Computer, Version 6.2.09." Accessed September 12, 2022. [Internet]. http://www.fon.hum.uva.nl/praat/.

Casillas, Joseph. 2015. "Production and perception of the /i/-/ɪ/ vowel contrast: The case of L2-dominant early learners of English". *Phonetica* 72 (2–3): 182–205. https://doi.org/10.1159/000431101.

Čubrović, Biljana. 2016. *Acoustic Investigations of Serbian and American English Vowel Inventories*. Belgrade: University of Belgrade, Faculty of Philology.

Dančetović, Nikola. 2017. "Phonetic/Phonological perception and production of English monophthongs at tertiary level." [In Serbian]. PhD diss., University of Kragujevac. https://doi.org/10.13140/RG.2.2.20756.86408.

DiCanio, Christian. 2013. "Vowel acoustics script for corpus data analysis." http://www.acsu.buffalo.edu/~cdicanio/scripts/Vowel_Acoustics_for_corpus_data.praat.

Escudero, Paola, and Paul Boersma. 2004. "Bridging the gap between L2 speech perception research and phonological theory." *Studies in Second Language Acquisition* 26 (4): 551–85. https://doi.org/10.1017/S0272263104040021.

Hillenbrand, James M., Michael J. Clark, and Terrance M. Nearey. 2001. "Effects of consonant environment on vowel formant patterns." *Journal of the Acoustical Society of America* 109 (2): 748–63. https://doi.org/10.1121/1.1337959.

Hirata, Yukari, and Kimiko Tsukada. 2004. "The effects of speaking rate and vowel length on formant movements in Japanese." In *Proceedings of the 2003 Texas Linguistics Society Conference*, edited by Augustine Agwuele, Willis Warren, and Sang-Hoon Park, 73–85. Somerville, MA: Cascadilla Proceedings Project.

Lehiste, Ilse. 1970. *Suprasegmentals*. Cambridge, Mass.: The MIT Press.

Lehiste, Ilse, and Pavle Ivić. 1986. *Word and Sentence Prosody in Serbo-Croatian*. Cambridge, Mass.: The MIT Press.

Marković, Maja. 2007. "Contrastive analysis of acoustic and articulatory characteristics of vowel systems of English and Serbian." [In Serbian]. PhD diss., University of Novi Sad.

Oh, Grace E., Susan Guion-Anderson, Katsura Aoyama, James E. Flege, Reiko Akahane-Yamada, and Tsuneo Yamada. 2011. "A one-year longitudinal study of English and Japanese vowel production by Japanese adults and children in an English-speaking setting." *Journal of Phonetics* 39 (2): 156–67. https://doi.org/10.1016/j.wocn.2011.01.002.

Paunović, Tatjana. 2002. "Phonetic-phonological interference of Serbian in the perception and production of English vowels." [In Serbian]. PhD diss., University of Niš.

Roberto Gonçalves, Alison, and Rosane Silveira. 2014. "The production of English high-front vowels by Brazilian learners." *Fórum Linguístico* 11 (1): 9–22.

Thomas, Erik R., and Tyler Kendall. 2007. "NORM: The vowel normalization and plotting suite." Accessed October 23, 2022. http://ncslaap.lib.ncsu.edu/tools/norm/.

Yavaş, Mehmet. 2011. *Applied English Phonology*. 2nd ed. Chichester: Wiley-Blackwell.

Appendix

Descriptive statistics with number of tokens for each word form, mean values of F1 and F2, and standard deviations

F1

Descriptive Statistics				
Segment	**Speaker group**	**N**	**Mean**	**Std. Deviation**
i	L1 MAE	15	314.33	21.30
	L2 MAE	15	274.74	20.88
ɪ	L1 MAE	15	476.67	62.92
	L2 MAE	15	415.78	57.52
u	L1 MAE	15	374.02	27.41
	L2 MAE	15	309.98	27.17
ʊ	L1 MAE	15	520.38	46.31
	L2 MAE	15	418.21	56.66

F2

Descriptive Statistics				
Segment	**Speaker group**	**N**	**Mean**	**Std. Deviation**
i	L1 MAE	15	2,157.36	76.27
	L2 MAE	15	2,286.01	152.56
ɪ	L1 MAE	15	1,720.73	79.87
	L2 MAE	15	1,902.04	122.48
u	L1 MAE	15	1,176.28	162.25
	L2 MAE	15	1,089.28	126.45
ʊ	L1 MAE	15	1,276.95	75.03
	L2 MAE	15	1,155.51	99.77