PRODUCTION ACCURACY OF L2 VOWELS: PHONOLOGICAL PARSIMONY AND PHONETIC FLEXIBILITY

ŠÁRKA ŠIMÁČKOVÁ
Palacký University Olomouc
sarka.simackova@upol.cz

VÁCLAV JONÁŠ PODLIPSKÝ
Palacký University Olomouc
sarka.simackova@upol.cz

Abstract
Ultimate attainment in foreign-language sound learning is addressed via vowel production accuracy in English spoken by advanced Czech EFL learners. English FLEECE–KIT, DRESS–TRAP, and GOOSE–FOOT contrasts are examined in terms of length, height, and backness. Our data show that, while being constrained by phonemic category assimilation (new vowel height distinctions are not created), the learners' interlanguage combines phonological parsimony (reusing L1 length feature to contrast L2 vowels) with phonetic flexibility (within-category shifts reflecting L1–L2 phonetic dissimilarity). Although achieving nativelike phonological competence may not be possible learners who acquire L2 in the prevailingly L1 environment, the Czech learners' implementations of English vowels revealed their ability to adjust for phonetic detail of L2 sounds.

Keywords: English as a foreign language, L2 phonology, ultimate attainment, vowels

1. Introduction

This study examines L2 pronunciation accuracy of advanced learners of English as a foreign language (EFL). While research on ultimate attainment in the domains of L2 phonetics and phonology is often carried out in immersion settings (e.g. the numerous AOL studies by Flege[1]), our goal is to estimate the limits on the final outcome of speech learning when L2 is learned without prolonged interactional input from a native speaker community. Speech sound learning under such input constraints suffers from serious disadvantages because frequent exposure to a variety of speakers from the same dialect would seem necessary for developing sensitivity to acoustic properties of L2 sounds, noticing L2–L1 differences, and

---

[1]   AOL stands for the Age of Learning, i.e. age at the onset of acquisition. The most frequently cited AOL studies include e.g. in Flege, Munro and MacKay (1995), Flege, Yeni-Komshian and Liu (1999), Mackay, I. R., Flege, J. E., & Imai, S. (2006).

refining L2 sound categories. Learning L2 English in one's home country, however, means that EFL learners only very sporadically experience exposure to authentic English used communicatively in a native English-speaking community. The native pronunciation models that such learners encounter interactively (e.g. native English teachers, personal acquaintances) and in media (e.g. via textbook recordings, YouTube, computer games, or TV shows) may be too diverse to allow them to develop stable sound representations close to one specific native English variety. Instead, their representations of L2 sounds are mainly shaped by frequent exposure to foreign-accented exemplars produced by their non-native teachers and peers. In addition, since EFL learning occurs in a pervasive L1 environment, the daily use of the L1 relative to L2 is typically very high. The interconnectedness of the learner's L1 and L2 sound patterns (Flege, Frieda & Nozawa 1997; Guion, Flege & Loftin 2000) results in L2 sound representations being constantly shaped by L1 interference due to the overwhelming prevalence of daily L1 use.

For the purposes of the current study participants were sought from a specific population of particularly successful EFL learners. They are best described as advanced L2 learners or, alternatively, as highly proficient late L1-dominant bilinguals, depending on whether we consider the fact that they actively continue to work on improving their L2 proficiency or the fact that they are linguistically fully competent in two languages (though predominantly using L1 in daily life). At the time of data collection the participants were students of English at an institution of higher education. They were learning English in an academic setting, being trained to become English language professionals, namely translators-interpreters, in a highly selective university programme. Such learners can be expected to have high levels of aptitude as well as to have received a great amount of formal exposure to the L2. At the same time, their language learning experience had been non-immersional, their L2 learning taking place in the L1 environment of their home country. If they experienced any more or less sustained or regular contact with a native speaking community, it was only in postadolescence; hence, they are best described as late bilinguals. By sampling L2 learners from such a population we hope to establish what the upper limits on foreign language sound learning might be, although, admittedly, the very idea of ultimate attainment, in the sense of a final fixed end-state, is an over-simplification and L2 (and possibly also L1) of these learners/bilinguals is likely to continue to change.

1.1. Measuring attainment – the monolingual reference point

A recent paper by Stoehr et al. (2017) highlights the role of language environment in which bilinguals function on a daily basis in shaping their ultimate linguistic competence. Arguing that the task of research into L2 phonology and phonetics is to establish how well "bilinguals have been able to adapt to the phonetic environment in which they actually acquire the L2" (p. 485), Stoehr et al. criticize

using the monolingual native speaker's performance as the reference point against which L2 speech acquisition is measured. In this view, while a comparison against monolingual native speakers may be suitable for learners immersed in L2 environment, it is less appropriate when one is examining speech of foreign language learners who have acquired the L2 in the L1 environment.

Another argument against monolingual linguistic competence as the reference point in evaluating outcomes of L2 learning is based on the findings of psycholinguistic research (Hopp & Schmid 2013, Stoehr et al. 2017). Researchers in bilingualism widely cite Grosjean's paper (1989) articulating the view that the linguistic competence of a bilingual does not comprise two separate language systems. Successful L2 learning eventually produces an individual competent in multiple languages (Cook 1992). In other words, the desired outcome of L2 learning is L2 mastery as well as retention of the previously learned L1 and other languages. However, adding a language to one's linguistic repertoire has consequences for the learner's overall linguistic competence, which manifests itself as inter-lingual interactions on all levels of language representation (e.g. Dussias & Sagarra 2007 for syntax, Meir, Walters & Armon-Lotem 2016 for morphosyntax), including phonological and phonetic representations. Research shows consistent differences between bilinguals and monolinguals both in how they represent speech sounds in their long-term memory (e.g. Barlow 2014, Fabiano-Smith & Barlow 2010 for children) and in how they process language during online production and perception (e.g. see articles in Deuchar 2016). Consequently, Hopp & Schmid (2013) or Stoehr et al. (2017) propose to rely on a bilingual reference point, arguing that since L2 learners' ultimate attainment can never mean full nativelikeness in the sense of achieving monolingual competence, bilinguals' linguistic performance should be compared to that of other bilinguals, e.g. late L2 learners to L1 attriters. Nonetheless, both studies (Hopp & Schmid 2013, Stoehr et al. 2017) include also (near) native monolingual control speakers.

While accepting the reservations about monolingual reference norms, we believe that, depending on the research goals, there may be good reasons to measure late L2 learners' speech against a native speakers' benchmark. Systematic comparisons of bilinguals' L1 and L2 performance in speech perception and production tasks to a monolingual baseline has led to the present-day understanding of bilingual speech sound representations. Thus, it is now well-established that, depending on factors such as age at the onset of acquisition, language proficiency, and language dominance, the L2 sound system shows a stronger or weaker influence of the L1 (among many others Caramazza et al. 1973, Flege 1991), that cross-language influences may be bi-directional with the L2 affecting the L1 (Chang 2012, Flege 1987), that bilinguals can form language-specific sound categories in the L2 which are (acoustically) distinct from corresponding L1 categories (Flege & Eefting 1987, Flege 1991), and that corresponding phonetic categories in a bilingual's L1 and L2 often show convergence but they may also diverge from each other, typically during early L2 learning that started before L1 categories are developed (Flege 2010).

As stated above, the current study aims to explore the limits on L2 speech learning by highly motivated learners in a foreign-language (FL) learning context typified by diminished interactional input. It is not concerned with how closely EFL learners approximate the L2 pronunciation they encounter in their immediate linguistic environment. Instead, it asks to what extent such learners can overcome the limitations set by that environment. This is done by describing the (near)-ultimate outcome of such learning in terms of differences from nativelikeness, i.e. from the native speakers' reference performance.

There are several reasons to expect at least some of our EFL learners to approximate, even if not completely match, nativelike speech. First, reduced interactional input available to these learners is to some extent compensated for by their both intensive (several hours a day at the time of data collection) and extensive (at least 10 years) instruction in English. As a result, the learners have achieved high proficiency in their L2 English, including oral proficiency as tested by CAE/CPE (Verhelst et al. 2009). Overall L2 proficiency has been shown to predict how nativelike L2 speech production is. More proficient L2 learners speak faster and make fewer errors (Kormos & Dénes 2004); they have better speech motor control (Nip & Blumenfeld 2015); they show greater segmental accuracy (Chakraborty, Domsch & Gonzales 2011) and adopt L2 articulatory patterns (e.g., Flege, Schirru & MacKay, 2003). In addition, based on our long-term experience with the target learner population, we believe motivation to be an essential factor in compensating for the input handicap. Empirical support for the impact of motivation on pronunciation learning in late L2 learners is found e.g. by Moyer (2004 2014). Typical learners in our population have no problem constructing an image of their ideal L2 self (Dörnyei 2009) as a fluent speaker with a specific nativelike English accent. Many of them are inclined to consider nativelike pronunciation to be important for their future profession of an English language translator-interpreter and desire to modify their pronunciation towards the chosen model (see Appendix for the learners' responses to three questions on attitudes to pronunciation included in a language experience questionnaire). Such learners are motivated to employ strategies for maximizing their exposure to L2, seeking out native English speakers outside school and exposing themselves to a variety of English-language media.

1.2. Cross-language interference

From the outset, our question about ultimate attainment was not whether there are L2 learner – native speaker differences in phonological competence but what is their nature. Nativelike acquisition of L2 phonology is relatively rare for late L2 learners due to firmly entrenched L1 sound patterns, even when L2s are learned in immersion contexts. Adding the unfavourable input conditions and low L2-to-L1-use ratios of non-immersion foreign language learning should make achieving authentic pronunciation impossible. While accent-rating studies do occasionally

report individual FL learners who sound undistinguishable from monolingual native speakers to native listeners' ears (Birdsong 2007, Bongaerts 1999), most successful FL learners' speech indeed reveals degrees of non-nativeness.

What eventually differentiates among bilinguals with respect to the authenticity of their L2 pronunciation is how developed their L1 phonological representations are at the start of learning L2 (Best & Tyler 2007). A completely developed L1 phonology of a late sequential bilingual has a firmer influence on L2 sound learning than a developing L1 phonology of an early simultaneous bilingual (McCarthy, Evans & Mahon 2013, Flege, Schirru & MacKay 2003). At the same time, early bilinguals' speech production shows bidirectional L1-L2 cross-language effects. For late bilinguals, on the other hand, a unidirectional influence of the L1 on the L2 is found (Baker & Trofimovich 2005), their L1 is more resistant to interference from L2.

The most influential models of L2 sound learning, the Speech Learning Model (e.g. Flege 1995) and the Perceptual Assimilation Model-L2 (Best & Tyler 2007), agree that L1 interference in L2 pronunciation, i.e. learners' inaccurate production of L2 segments, occurs because of inaccurate perception, where learners rely too much on L1 sound categories to process the incoming signal. Insufficient exposure to auditory input in the FL learning context complicates overcoming such L1 entrenchment. Further, central to SLM and consistent with PAM-L2 is the idea that phonetic categories[2] of both L1 and L2 are at the higher level of representation accommodated within a single phonological system. Consequently, a bilingual's/learner's phonology is comprised of both L2 and L1 phonetic categories and thus likely to differ from phonology of a monolingual speaker of either L2 or L1. Flege (1995) proposed that interlingual L1-L2 interactions take place at the subphonemic level of the phonetic categories. He further argued, with corroboration from empirical research, that one's phonetic categories remain flexible past puberty and the ability to form new phonetic categories remains available to adult learners. In Flege's model, the relative success in L2 phonetic learning depends on acoustic closeness of sounds in the L2 input to the existing L1 sounds: new sounds that are phonetically different from their closest L1 equivalents are more likely to be accurately acquired than phonetically similar sounds.

However, a shared phonological space also means that a bilingual's L2 phonology parsimoniously reuses the existing L1 phonemic elements (features, phonemes), i.e. there is also phonological cross-language influence. For example, L1-Czech voicing contrast or place features of obstruents can be transferred into L2-English although their phonetic implementation, e.g. English-specific VOT settings or the apico-alveolar rather than lamino-dental realization of /t, d/, must be newly learned. The early learned phonological structure seems to be quite resistant to change due to additional language learning. Attrition studies typically report phonetic adjustments rather than losses of phonologically relevant L1

---

2   PAM-L2 does not use acoustic categories but distal gestures.

distinctions (Mayr, Price & Mennen 2012; Schmid 2011; but cf. de Leeuw, Tusha & Schmid 2017). Studies of L2 speech learning show that learning a new L2 contrast which is not part of L1 phonology may prove impossible. Specific predictions about degrees of difficulty in acquiring contrasting L2 phonemes are made within the PAM-L2 framework.

This model allows for L1-L2 relationships both at the level of gradient phonetic knowledge and at the higher level of phonological representations. According to Best and Tyler (2007), "contrasts at the functional linguistic level of the L1 phonology and their relationship to phonological contrasts of the L2 are as important to perceptual learning as phonetic categories in the two languages" (p. 26). Perceptual assimilation (identification of a perceived L2 sound as an instance of an L1 category) that, at least initially, constrains L2-sound learning, results from the interplay of phonological and phonetic L1-L2 mappings. Different assimilation patterns are possible, of which the most relevant for the current study are (1) Two-Category assimilation (two members of an L2 contrast are assimilated to two different L1 phonemes), (2) One-Category assimilation with a Category-Goodness Difference (two members of an L2 contrast are assimilated to the same L1 phoneme but one is perceived as a less prototypical instance of the L1 phoneme than the other), and (3) One-Category assimilation (two members of an L2 contrast are assimilated to the same L1 phoneme and they are both equally (dis)similar to the L1 phoneme). The difficulty of acquiring the L2 sound contrast is expected to increase from the first to the last pattern. However, two-category assimilation is no guarantee that the phones instantiating the contrasting L2 phonemes will be perceived and implemented phonetically in a nativelike way. Learners' ability to treat them as distinct may rely on their associating the contrast with a non-target-like phonetic cue. For example, Spanish learners of L2 English have been shown to rely on duration in discriminating English /i/-/ɪ/, which native listeners primarily distinguish on the basis of spectral quality (Escudero & Boersma 2004). In the present study, we expect the Czech EFL learners (Czech being a vowel-quantity language) to produce a consistent durational difference but not necessarily a spectral difference between pairs of English vowels. As will be apparent from the contrastive comparison of English and Czech in the next section, this would be the case not because of the universal availability of the durational cue (e.g. Bohn 1995, Cebrian 2006, McAllister, Flege & Piske 2002) but because of the cross-language influence of the phonologically encoded vowel length in the learners' L1 (Ylinen et al. 2010).

When making predictions in the following section, we assume that memory representations of L2 sounds are based on the perception of L2 speech and its vis-a-vis interactions with existing L1 sound representations. We further assume that the same representations underlie L2 sound perception and production. Consequently, production data from an L2 learner are revealing about the learner's phonology and reflect the perceptual abilities that constrain it. The patterns of L1-L2 phonetic and phonological mappings that initially constrain perception of

L2 sounds are likely to change with exposure to L2 input. Analysing speech of advanced FL learners is informative about any enduring effects of the L1 sound system on L2.

1.3. English and Czech vowels: a cross-language comparison

The vowel inventories of the two languages in question differ in the number of phonemic categories and in the phonological features along which vowel phonemes are contrasted. In addition, even equivalent L1~L2 phonemes differ to a greater or lesser extent in how they are phonetically implemented. The following cross-linguistic comparison focuses on monophthongs.

   In the participants' L1, Czech, five contrasting vowel qualities specified in terms of height and backness combine with two degrees of length (Skarnitzl, Šturm & Volín 2016). This yields an inventory of ten vowel phonemes /iː, ɪ, ɛː, ɛ, aː, a, oː, o, uː, u/. Notice that, unlike the other long – short pairs, the members of the /iː/-/ɪ/ pair are represented by different IPA symbols. The qualitative differentiation of the long and short high front vowels was first documented in Podlipský, Skarnitzl & Volín (2009), who also noted a decreased /iː/-to-/ɪ/ duration ratio. Interestingly, the back pair of high vowels /uː/-/u/ exhibits a tendency towards a symmetrical change, although the qualitative differentiation of the two vowels is smaller compared to the front vowel pair (Skarnitzl & Volín 2012) and their short-to-long ration is greater, i.e. intermediate between that of /iː/-/ɪ/ and of the other three vowel pairs (Podlipský et al. 2009). In addition, the perception experiment in Podlipský et al. (2009) revealed a regional difference in attending to spectral vs. durational information in categorization of /iː/-/ɪ/. Listeners from Bohemia relied more on spectrum, while speakers from Moravia, the region of origin of our participants, relied more on the durational cue. The Czech EFL learners in this study can thus be expected to transfer the length feature into L2 English and differentiate English vowels as long vs. short.

   In the participants' L2, English, pairs of vowels also differ in duration. However, the temporal difference is phonetic rather than phonemic, accompanying a spectral differentiation of the vowels. Consequently, the distinction between vowels such as /iː/ and /ɪ/ or /uː/ and /ʊ/ has been treated as a lax – tense contrast (Giegrich 1992, Ladefoged & Johnson 2014). Perception experiments confirm that native English listeners attend primarily to vowel quality (Hillenbrand, Clark & Houde 2000, Morrison 2008, Kondaurova & Francis 2008) and only rely on duration as a secondary cue. Standard Southern British English (SSBE), the native reference accent used here, has a phonemic inventory of 12 monophthongs (Cruttenden 2014[3]). Six of these vowels are investigated here, namely the two lax – tense high vowel pairs /iː/-/ɪ/ and /uː/-/ʊ/, and the front non-high vowels /ɛ/-/æ/ contrasting in height.

---

[3]   The list of contrasting RP monophthongs in Cruttenden (2014) includes the SQUARE set, which is represented as containing a long monophthong.

The low front /æ/, represented as TRAP in Wells's lexical sets (1982), is exceptional in being phonetically longer than the other English lax vowels, especially before a voiced obstruent (Cruttenden 2014). This may have consequences for Czech learners of English who have been previously shown to perceive and produce the TRAP–DRESS contrast as a long–short distinction (Šimáčková 2003). In terms of quality, the vowel also presents a challenge to a Czech learner. There appear to be two options in how the quality of /æ/ can be treated vis-à-vis the learners' L1. First, since the non-high front region of the Czech vowel space is occupied by a pair of phonemes /ɛ/ and /ɛː/ undifferentiated in quality, a single-category assimilation of L2 /æ/ and /ɛ/ to the spectrally overlapping Czech /ɛ-ɛː/ can be predicted. The spectral overlap was observed in Czech learners' productions of English TRAP and DRESS in Šimáčková (2003). Alternatively, the Czech EFL bilinguals in the present study, who were tested 14 years later, may have picked up on the shift in the SSBE pronunciation of the TRAP vowel towards a lower and more retracted quality (Hawkins & Midgley 2005, Cruttenden 2014, Bjelaković 2017). In such a case, they might either assimilate percepts of TRAP to the Czech /ɛ/ or /ɛː/ but as poor exemplars of those categories, or they may assimilate them to the Czech low, non-back /a/ (or /aː/). The published reference values of F1 and F2 in Table 1 suggest that in terms of height (i.e. F1), the English /æ/ is closer to the Czech /a-aː/, whereas in terms of backness (F2) it is between Czech /ɛ/ and /a/.

Predicting the spectral mappings between the remaining five L2 target vowels and their L1 equivalents is more straightforward. The other member of the /ɛ/-/æ/ contrast, the English DRESS vowel corresponds to the Czech short /ɛ/. Its phonetic implementation is close to the Czech counterpart in height, though it is somewhat more front than the Czech vowel (see Table 1). The English lax–tense contrast between the high front FLEECE and KIT vowels maps on the contrasting long–short /iː/-/ɪ/ in Czech. The acoustic values in Table 1 confirm the phonetic closeness of the corresponding high front vowels, especially for English /i/ and Czech /iː/. English lax /ɪ/ appears somewhat more front compared to the Czech short /ɪ/. English and Czech are less close in the high back vowel pair. First, English /ʊ/ is spectrally more differentiated from its tense counterpart /u/ than the Czech short /u/ is from the long /uː/. Second, although both GOOSE and FOOT match Czech /u/ and /uː/ in vowel height, they are clearly different in terms of backness. It is well documented that the English tense /u/ and to some degree also the lax /ʊ/ have undergone fronting (Hawkins & Midgley 2005, Cruttenden 2014). Consequently, the vowels have become phonetically less similar to the Czech /uː/ and /u/, which are realized as fully back (Skarnitzl and Volín 2012). The cross-linguistic difference in the degree of retraction (represented by F2) is clearly evident in Table 1.

Table 1. Formant reference values for the target English vowels (Bjelaković 2017) and their closest Czech equivalents (Skarnitzl & Volín 2012)

Target English vowel	L1 Czech equivalent	F1 (Hz)		F2 (Hz)	
		RP	CZ	RP	CZ
FLEECE	iː	350	329	2623	2603
KIT	ɪ	457	492	2071	2251
DRESS	ɛ	636	678	1918	1793
TRAP	ɛ, ɛː a, aː	845	678, 710 773, 801	1663	1793, 1904 1503, 1418
GOOSE	uː	347	344	1852	757
FOOT	ʊ	444	415	1491	1004

To sum up, the contrasting English vowels and the inter-lingual assimilation patterns examined in the current study include:

1) FLEECE, KIT: Two-category assimilation (to L1 Czech /iː/, /ɪ/) based on vowel length and quality combines with a phonetic overlap of the corresponding L2-L1 categories.
2) GOOSE, FOOT: Two-category assimilation (to L1 Czech /uː/, /u/) based on length combines with phonetic dissimilarity of the corresponding L2-L1 categories. The phonetic realizations of the corresponding L2-L1 vowels differ spectrally.
3) DRESS, TRAP: (a) Both vowels assimilate to a single L1 phoneme (/ɛ/), although only one (DRESS) is phonetically similar to the Czech category, the TRAP vowel differing in spectrum as well as in duration, or (b) two-category assimilation (to L1 Czech /ɛ/, /ɛː/) based on length combines with phonetic dissimilarity of TRAP from Czech /ɛː/, or (c) two-category assimilation is possible based on quality and length (to /ɛ/, /aː/) or quality only (or to /ɛ/, /a/).

The relative accuracy of the three target contrasts in speech of beginning Czech EFL learners should then pattern in the following way: the FLEECE-KIT pair should be most clearly separated, showing both temporal and spectral differentiation, with FLEECE being realized closer to the L2 target. The GOOSE-FOOT vowels should show differentiation in the temporal dimension. Both vowels are likely to differ from the native English fronted targets. For the TRAP-DRESS contrast, alternative outcomes are possible: The vowels may not be differentiated at all, or they may differ only in duration. Alternatively, they may be clearly differentiated both temporally and spectrally, if TRAP assimilates to the Czech /aː/, or less clearly, if both assimilate to the Czech /ɛ/ and /ɛː/ and TRAP is recognized as a poor phonetic implementation of /ɛː/.

The question we are asking about advanced L2 learners are as follows: Can the phonological and phonetic constrains, defined for the initial state of beginner FL learning, be overcome as a result of prolonged and intensive FL learning experience leading to high L2 proficiency? Do highly proficient EFL learners stop

relying on L1 phonemic categories (such as contrastive length); do they expand the phonemic system (e.g. adding a height distinction) and the repertoire of phonetic categories (e.g. forming fronted [u] and [ʊ])?

2. Methodology

2.1. Reference data

Formant measurements of the Czech EFL bilinguals' target vowels were compared to the data from 7 female speakers of Received Pronunciation published in Bjelaković (2017). Detailed biographical information about the speakers in this corpus of read BBC news is provided in the paper. The speakers are older than the bilinguals in the current study, born between 1955 and 1968 (i.e. 49–62 years old at the time of the data collection.) The paper gives each speaker's mean F1 and F2 values in Hz. Standard British English pronunciation, RP, rather than American pronunciation was chosen as the reference point because of its wide representation in language teaching materials used in the Czech Republic.

The bilinguals' English vowels were further compared to the published formant values for equivalent Czech vowels (Skarnitzl & Volín, 2012). These reference data, based on recordings of read texts by 48 female speakers aged 20–30 years, comprise group means of F1 and F2[4]. The duration reference data were group means for six speakers (3 female), from a corpus of Czech Radio news recordings Skarnizl (2012).

2.2. Participants

The 20 Czech-English bilinguals were EFL learners, all young females between 19 and 27 years of age (M = 22). At the time of data collection they were enrolled in the bachelor programme 'English for Interpreters and Translators' at the Palacký University Olomouc. All were L1-dominant though highly proficient in their L2 (C1 or C2 in CEFR).

The stimuli were recorded by 5 native speakers of English: 2 speakers of SSBE (male, 40 and 52 yrs.), 3 speakers of North American English (1 male, 41 yrs; 2 female, 24 and 54 yrs.

2.3. Stimuli and procedure

The targets for analysis were 6 vowels in monosyllabic CVC words controlled for voicing of the post-vocalic consonant. Each high vowel occurred in six words, non-high vowels occurred in four words. The complete set included the KIT vowel

---

[4]   In the paper, the female values are represented only in a figure; the exact numerical values were kindly provided by the authors.

in *big, give, dig, fish, thick, sit*; FLEECE in *cheese, lead, leave, heat, cheat, niece*; GOOSE in *choose, lose, move, goose, shoot, soup*; FOOT in *good, hood, should, book, bush, look*; DRESS in *beg, bed, bet, neck*; TRAP in *bag, badge, back, match*.

Each target was placed in 2 sentences, once occurring sentence-initially (e.g. *Give them the money.*), once sentence-finally (e.g. *What did you give?*). In total there were 64 stimulus sentences and 24 fillers. Participants produced the sentences in a delayed repetition task during which they heard a stimulus sentence followed by a prompt *What should you say?* said by a different person. The participant responded using the quote frame *I should say, _* and repeating the stimulus, e.g. "I should say, Give them the money." The native speakers who provided the baseline read the stimulus sentences off a computer screen and produced them in the frame *I should say, _*. A subset of each native speaker's sentences was included in the elicitation instrument used with the learners.

2.4. Measurement and analysis

In each elicited vowel token, duration and frequencies of F1 and F2 were measured. The onset and offset of the vowels was determined manually from the waveform with a reference to the vocalic formant structure in the spectrogram. Formants were tracked using the Burg method in Praat (Boersma and Weenink, 2017), with the maximum formant value set to 3500 Hz for GOOSE and FOOT, to 3800 Hz for the other vowels and the number of formants set to 3. Subsequently, the mean F1 and F2 in the medial 50% of each vowel were computed in hertz. These acoustic measurements were then used to calculate the mean duration and the mean formant frequencies for each speaker's sentence-initial and sentence-final target vowel.

3. Results

3.1. Temporal differentiation

Figure 1 shows the learners' group mean duration of each vowel. Repeated Measures (RM) ANOVA with duration as the dependent variable and Vowel (FLEECE, KIT, GOOSE, FOOT, DRESS, TRAP) as the within-subject independent variable revealed its significant main effect [$F(5, 95) = 80. 21$, $p < .0001$]. According to a post-hoc Tukey test, vowels in each lax-tense pair significantly differed from each other ($p < .001$). Another RM ANOVA performed on long-to-short durational ratios found a significant effect of Vowel pair [$F(2, 38) = 44.33$, $p < .0001$], a post-hoc Tukey test confirming that the i/ɪ ratio to be significantly higher compared to the u/ʊ ratio, which was in turn higher than the æ/ɛ ratio ($p < .01$).

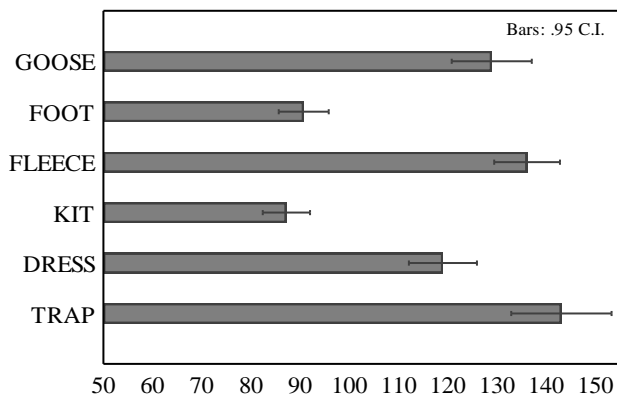**Figure 1.** Group mean durations of the Czech EFL bilinguals' target vowels in ms

3.2. Spectral differentiation

In Figures 2 and 3, learners' target vowels are represented in an F1-by-F2 vowel space. For this display, the measurements in Hz were converted into Bark and normalized so that height is represented as the distance of the F1 from F3 and retraction as the distance of F2 from F3. Individual learners' means for the six vowels are displayed in Figure 2, while Figure 3 shows each vowel token measured. For the sake of displaying the full front-back scale, the figures include measurements of an additional, back vowel (/ɔ/ in the word *thought*). A single token for each participant was measured. In Table 2, mean F1 and F2 values in Hz with standard deviations of the bilingual EFL group are juxtaposed to the native RP speaker data from Bjelaković (2017).

Table 2. Group mean F1, F2 values in Hz and standard deviations for 20 Czech EFL bilinguals in the current study and for the 7 RP speakers in Bjelaković, 2017

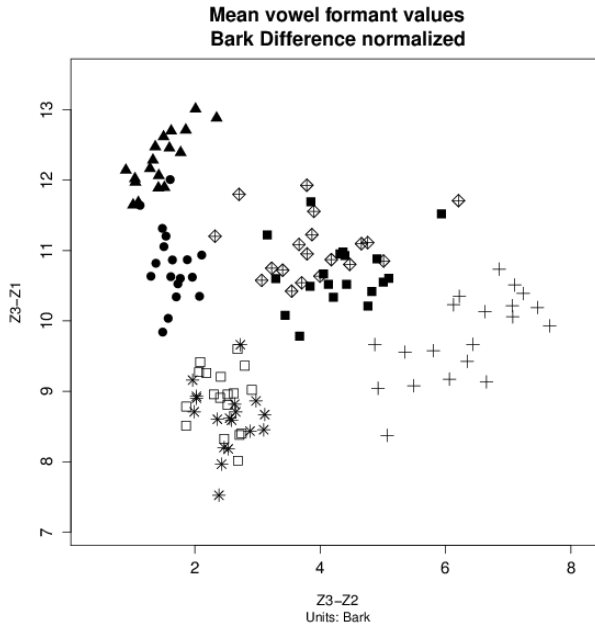|	Czech EFL bilinguals		RP speakers		Czech EFL bilinguals		RP speakers	
	F1		F1		F2		F2	
	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*	Mean	*(SD)*
FLEECE	383.7	*(25.8)*	349.7	*65.7*	2604.8	*123.1*	2623.0	*59.8*
KIT	467.5	*33.8*	457.4	*59.9*	2239.1	*154.4*	2071.1	*100.4*
DRESS	710.1	*75.7*	636.3	*94.2*	2013.0	*110.7*	1918.6	*61.1*
TRAP	743.3	*79.8*	844.9	*111.5*	1972.5	*162.9*	1663.4	*44.1*
GOOSE	405.6	*24.0*	347.0	*51.8*	1561.5	*208.3*	1852.7	*71.2*
FOOT	443.6	*36.5*	444.1	*73.4*	1458.8	*162.2*	1491.7	*97.4*

Figure 2. Czech EFL bilinguals' mean vowel formant values (Bark difference normalized). Legend: triangle FLEECE, dot KIT, empty square DRESS, asterisk TRAP, diamond GOOSE, filled square FOOT, plus THOUGHT
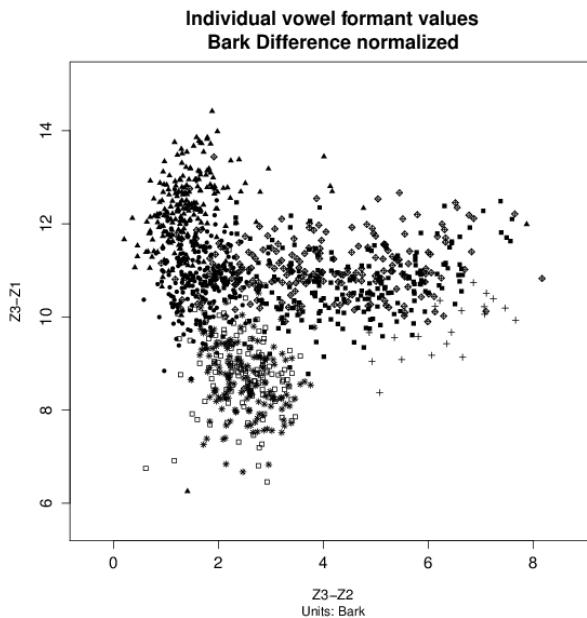

Figure 3. Czech EFL bilinguals' individual formant values (Bark difference normalized). Legend: triangle FLEECE, dot KIT, empty square DRESS, asterisk TRAP, diamond GOOSE, filled square FOOT, plus THOUGHT

In order to evaluate statistically the spectral differentiation of the learners' vowels as well as to compare them to native speakers' vowels, two separate RM ANOVAs were performed on mean F1 and F2 measurements. The within-subject variable Vowel (FLEECE, KIT, GOOSE, FOOT, DRESS, TRAP) and the between-subject variable Speaker Group (Bilinguals, RP Speakers) were included. The results for F1 showed a significant main effect for Vowel [$F(5, 125) = 340.00$, $p < .0001$] but not for Speaker Group. The interaction between the Speaker Group and Vowel as significant [$F(5, 125) = 10.856$, $p < .0001$]. A parallel main effect of Vowel [$F(5, 125) = 293.10$, $p < .0001$] but not Speaker Group, and a significant Vowel – Speaker Group interaction [$F(5, 125) = 19.624$, $p < .0001$) was found for F2. Post-hoc Tukey tests for both F1 and F2 confirmed that native speakers produced significant differences between the vowels in each pair ($p < .01$). For the learners, the post-hoc test results are summarized in Table 3. Their L2 /i/ was clearly constrasted with /ɪ/, having a significantly lower mean F1 and higher mean F2. The other two vowel pairs, /ɛ/-/æ/ and /u/-/ʊ/, did not differ in either dimension within the learners. The bilinguals differed significantly from the RP speakers with respect to two English vowels. They produced TRAP vowel with lower F1 and higher F2 values, i.e. the vowel was higher and more front compared to that of the RP reference data. Second, their L2 /u/ had a significantly lower F2, not showing the same degree of fronting as the native speakers' GOOSE vowel.

Table 3. Pairwise comparisons of mean F1 and F2. Columns 2 & 4 are within-subject comparisons within each vowel pair, Columns 3 & 5 comparisons of Czech EFL bilinguals and RP speakers from Bjelaković, 2017

Target vowels	F1 diff. from each other	F1 diff. from RP vowel	F2 diff. from each other	F2 diff. from RP vowel
FLEECE	p < .001	n.s.	p < .001	n.s.
KIT		n.s.		n.s.
DRESS	n.s.	n.s.	n.s.	n.s.
TRAP		p = .01		p < .001
GOOSE	n.s.	n.s.	n.s.	p < .001
FOOT		n.s.		n.s.

3.3. Assessing the degree of nativelikeness of the L2 data

Finally, using the nativelikeness criterion of bilinguals' values falling within 1 standard deviation of the native speakers's mean (e.g. Birdsong 2007), we compared individual mean F1s and F2s to the RP reference data. The number of bilinguals who satisfied this criterion for each vowel is given in Table 4, showing that more bilinguals were accurate in vowel height (F1) than in the front-back dimension (F2). The number of bilinguals who satisfied the criterion for both F1 and F2 was highest for FLEECE and FOOT. For F1, we further counted how many bilinguals satisfied the criterion for both contrasting vowels of each target pair.

Sixteen speakers approximated NS's values for both the FLEECE and KIT vowels, ten for FOOT and GOOSE, and only five for DRESS and TRAP. Two bilinguals produced RP-like F1 values for all 6 vowels.

Table 4. Number of bilinguals whose formant
values fall within 1 SD of the native speakers' group mean

	F1	F2	F1&F2
FLEECE	17	11	10
KIT	18	9	7
DRESS	13	7	5
TRAP	12	1	0
GOOSE	10	2	1
FOOT	18	13	11

3.4. The variability of L2 TRAP and DRESS

In Table 2, the bilinguals' F1 column showed higher standard deviations from the group mean for DRESS and TRAP relative to the other vowels, revealing increased between-speaker variability. Within-speaker variability in F1 across the vowels was assessed by a one-way ANOVA on individuals' SDs for each vowel, which found a significant main effect of Vowel [$F(5, 114) = 10.705$, $p < .001$]. A post-hoc Tukey test confirmed significantly higher SDs for DRESS compared to the remaining four vowels and significantly higher SDs for TRAP compared to three out of the four remaining vowels ($p < .05$), FLEECE being the exception. SDs for DRESS and TRAP did not differ. The variability within and across individual bilinguals in F1 of DRESS and TRAP can be observed in Figure 4, showing each speaker's mean, 1 SD from the mean (box), and 2 SDs from the mean (whiskers). A further perusal of individuals' raw formant (and durational) data reveals that lexical misrepresentation of TRAP and DRESS words is not uncommon for the bilinguals who pronounced some words belonging to the DRESS category with the TRAP vowel and vice versa. Figure 5 displays two scatter plots of F1 values against vowel duration of DRESS and TRAP produced by the speakers KM and KK, whose DRESS had one of the highest and the lowest SDs from the mean F1, respectively. Both figures show clusters of shorter vowels with lower F1 and longer vowels with higher F1, each cluster comprising both DRESS and TRAP words.
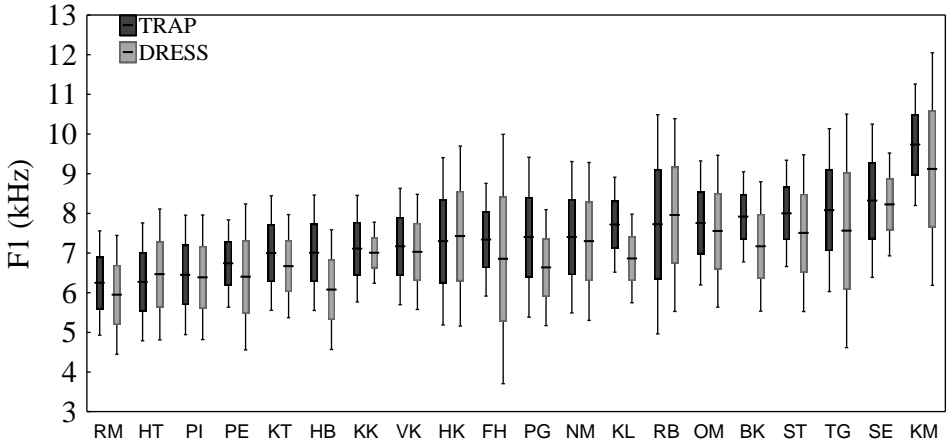
Figure 4. Individual mean F1 values (horizontal line), 1 standard deviation (box) and 2 standard deviations (bar) from the mean. Bilinguals are arranged according to F1 of TRAP from the lowest to the highest value
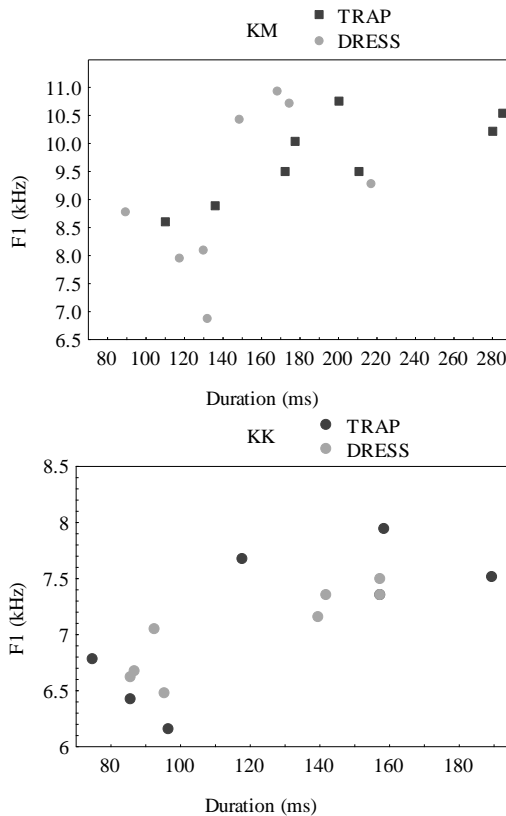

Figure 5. Scatterplots of F1 values against vowel duration for speakers KM and KK

3.4. Fronting of FOOT and GOOSE

The bilingual group's mean F2 of the GOOSE vowel is significantly lower relative to the RP speakers' value (Tables 1, 2) although one speaker's mean F2 of GOOSE (and FOOT) was in fact above 1 SD of the native speakers' value. Matching GOOSE and FOOT to their closest Czech equivalents (data from Skarnitzl & Volín 2012) reveals a difference between L1 and L2 high back vowels along the front-back dimension. A t-test for single means, testing the mean F2 values of the bilinguals' GOOSE vowel against the reference value of 757 Hz (long Czech /uː/) found the bilingual values to be significantly higher (t[19] = 17.27, p < .001). Likewise, a t-test for single means testing F2 values of bilinguals' FOOT vowel against the reference value of 1135 Hz (short Czech /u/) showed that they were significantly higher (t[19] = 8.93, p < .001). In sum, the Czech EFL bilinguals' English /u/ and /ʊ/ were more front (had higher F2) than the Czech native speakers' mean long /uː/ and short /u/ reported in Skarnitzl and Volín (2012).

4. Discussion

This is a study of cross-language L1 interference manifested in L2-vowel production of highly proficient EFL bilinguals. By comparing the acoustic measurements of bilingual vowels to native speaker reference data we address the question of limits on ultimate attainment in L2 sound learning under the input restrictions of the foreign-language settings. Assuming that cross-language interference involves L1-L2 interactions both at the higher-level phonological and lower-level phonetic representations, we selected three English vowel pairs for which we described, based on PAM-L2, differential degrees of L1 interference at the initial stages of L2 learning. We further suggested that phonological and phonetic learning means moving away from these L1 constraints and that speech of advanced bilinguals should evidence such changes.

Three pairs of contrasting English vowels were tested, namely /iː, ɪ/, /uː, ʊ/ and /æ, ɛ/. For each vowel pair we asked if and how clearly the bilinguals distinguished between the contrasting vowels and how closely each of the six bilinguals' vowels approximated production of reference native speakers of RP.

Our measurements showed that in all three pairs, the Czech EFL bilinguals relied on duration to differentiate between vowels. They used length to contrast FLEECE-KIT as much as, or even more than, they used it to contrast GOOSE-FOOT. They did so despite the reduction of the durational /iː/-/ɪ/ differentiation in their L1 related to increased spectral differentiation (Podlipský et al. 2009). The bilinguals also used duration to differentiate the L2 TRAP from DRESS. The smaller long-to-short duration ratio reported in group results cannot be interpreted as a weaker reliance on duration. Instead, individual bilinguals' data suggest that it is due to lexical misrepresentations. As illustrated by Figure 5, some DRESS

words were realized with extended vowel duration (and a relatively high F1) indicating that /æ/ was the target while some TRAP words had a relatively shorter vowel (and low F1) indicative of the /ɛ/ target.

With respect to vowel quality, the bilinguals' performance on FLEECE and KIT can be regarded as a baseline for evaluating the production accuracy of the other two pairs. In both English and Czech the high front vowels are differentiated by spectrum as well as duration. Spectrally, FLEECE overlaps with Czech /iː/ and KIT is very similar to Czech /ɪ/. Unsurprisingly, the bilinguals contrasted the English vowels both in length and quality. Further, for the group, the implementation of either vowel does not differ from the native speakers' implementation either in height or backness. Nonetheless, the bilinguals' production of KIT, which matches closely the Czech reference values of short /ɪ/ and appears somewhat more (though not significantly) front compared to the RP speakers' KIT, evokes Flege's (1995) reasoning about the relative difficulty of acquiring similar sounds completely. Out of the twenty bilinguals, ten produced FLEECE in a nativelike way, seven of which produced also nativelike KIT.

Compared to the FLEECE-KIT baseline, the bilinguals' production accuracy of GOOSE and FOOT is somewhat diminished. The vowels are contrasted by length but not by quality. The tendency towards some spectral differentiation between long and short /uː/-/u/ reported for the bilinguals' L1 Czech, is not observable in their L2 at all. In terms of nativelikeness, the lax /ʊ/ is better implemented than the tense /u/. In fact, with eleven bilinguals having nativelike pronunciation, /ʊ/ is the most accurately produced vowel of the whole set. Both vowels approximate better the native reference values when it comes to height, which is warranted by similarity between English and Czech. With respect to backness, bilinguals' FOOT is close to the RP realization while their GOOSE is different. And yet, a comparison with Czech /uː/ and /u/ shows for both vowels a substantial degree of fronting (in one case even an overshoot of fronting). This is an indication that the bilinguals succeeded in creating new spectral categories for their English high back vowels, distinct from the existing L1 phones. The bilinguals' L2 /u/ and /ʊ/, undifferentiated in height, are fronted together, which is sufficient for the accuracy of FOOT but not of GOOSE, with much more front realizations in modern RP.

Clearly, the bilinguals were least successful in production of DRESS and TRAP. The two vowels were separated in duration, TRAP being the "long" vowel (Šimáčková 2003), but not in vowel height or backness. Unlike in the case of GOOSE and FOOT, this may be a result of lexical misrepresentation rather than of phonological non-differentiation. In terms of vowel height (and duration), the bilinguals do seem to separate two phonetic targets but sometimes use [ɛ] for TRAP words and [æ] for DRESS words. The phonetic differentiation is then lost in individual means calculated across all TRAP and all DRESS tokens. Regarding the accuracy of each vowel, DRESS was closer to the RP reference values along both dimensions, reflecting the phonetic similarity of the English lax /ɛ/ the Czech

short /ɛ/. The bilinguals realized TRAP as a front vowel showing little, or rather, no retraction. Obviously, they did not assimilate the English vowel to the Czech /a-aː/. The results for vowel height discussed above suggest the bilinguals' perception of [æ]-phones as poor exemplars of /ɛ/.

To summarize the discussion, for the EFL bilinguals in this study, cross-language phonological interference deriving from early established phonemic categories continued to influence their vowel production even at the highest levels of proficiency. First, the phonemic vowel length discretely separated the vowels into long and short. Second, the FOOT-GOOSE and DRESS-TRAP contrasts showed that the bilinguals did not succeed in adding new vowel-height distinctions. The height differentiation of FLEECE and KIT can be interpreted as a transfer of L1 phonetic categories. In contrast, the phonetic implementation of the existing categories can shift towards more L2-like targets as evidenced by fronting of the GOOSE and FOOT vowels. The production results for TRAP signal the bilingual's ability to notice phonetic dissimilarity, which in this case led to destabilizing the /ɛ/-category and for individual items to less L2-like production (e.g. when neck is pronounced as [næk]).

5. Conclusion

It is not the case that there is one best outcome of L2 speech learning, nativelike L2 phonology. Depending on the circumstances of L2 learning, there are a number of ultimate attainment outcomes. This study shows that foreign language sound learning is constrained by phonological interference. For the Czech EFL bilinguals, the phonemic structure underlying L1 vocalic inventory defines the boundaries of their L2 system, in the sense that old phonemic categories are reused and new ones are difficult to set up. Such cross-language interference can be understood as parsimony: a bilingual late L2 speaker, who is fluent in both L1 and L2, has arrived at an optimal bilingual sound system in which phonological units such as features are not necessarily multiplied. Those units that have been already established are maximally exploited. However, as shown by our results for the fronting of the L2 high back vowels, even in the circumstances of limited access to interactional native input, these bilinguals show evidence of flexible phonetic learning, achieving an approximation to nativelike values for these vowels.

References

Baker, Wendy and Pavel Trofimovich. 2005. Interaction of native-and second-language vowel system (s) in early and late bilinguals. *Language and speech* 48(1). 1-27.
Barlow, Jessica A. 2014. Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology* 5. 288.

Best, Catherine T. and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language learning: In honor of James Emil Flege*, 13-34. Amsterdam: John Benjamins.

Birdsong, David. 2007. Nativelike pronunciation among late learners of French as a second language. In Ocke-Schwen Bohn and Murray J. Munro (eds.), *Language experience in second language learning: In honor of James Emil Flege*, 99-116.

Bjelaković, Andrej. 2017. The vowels of contemporary RP: vowel formant measurements for BBC newsreaders 1. *English Language & Linguistics* 21(3). 501-532.

Boersma, Paul and David Weenink. Praat: doing phonetics by computer [Computer program]. Version 5.4., 2014, retrieved 4 October 2014, http://www.praat.org/.

Bohn, Ocke-Schwen. 1995. Cross-language speech perception in adults: First language transfer doesn't tell it all. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 279-304. Timonium, MD: York Press.

Bongaerts, Theo. 1999. Ultimate attainment in L2 pronunciation: The case of very advanced late L2 learners. In David Birdsong (ed.), *Second language acquisition and the critical period hypothesis*, 133-159. Mahwah, NJ: Erlbaum.

Caramazza, Alfonso, et al. 1973. The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54(2). 421-428.

Cebrian, Juli. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34(3). 372-387.

Chakraborty, Rahul, Celeste Domsch and Maria D. Gonzales. 2011. Articulatory behaviors of nonnative speakers: Role of L2 proficiency and accent modification. *Perceptual and Motor Skills* 113(1). 311-330.

Chang, Charles B. 2012. Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics* 40(2). 249-268.

Cook, Vivian J. 1992. Evidence for multicompetence. *Language Learning* 42(4). 557-591.

Cruttenden, Alan. 2014. *Gimson's Pronunciation of English, 8th edition*. London and New York: Routledge.

Deuchar, Margaret. 2016. Cross-language effects in bilingual production and comprehension: some novel findings. *Bilingualism: Language and Cognition* 19(4). 706-709.

De Leeuw, Esther, Aurela Tusha and Monika S. Schmid. 2018. Individual phonological attrition in Albanian–English late bilinguals. *Bilingualism: Language and Cognition* 21(2), 278-295.

Dörnyei, Zoltán. 2009. The L2 motivational self system. In Zoltán Dörnyei and Ema Ushioda (eds.), *Motivation, language identity and the L2 self,* 9-11. Clevedon, UK: Multilingual Matters.

Dussias, Paola E. and Nuria Sagarra. 2007. The effect of exposure on syntactic parsing in Spanish–English bilinguals. *Bilingualism: Language and Cognition* 10(1). 101-116.

Escudero, Paola and Paul Boersma. 2004. Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition* 26(4). 551-585.

Fabiano-Smith, Leah and Jessica A. Barlow. 2010. Interaction in bilingual phonological acquisition: Evidence from phonetic inventories. *International journal of bilingual education and bilingualism* 13(1). 81-97.

Flege, James E. 2010. "Age" effects on second language acquisition. In Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel and Małgorzata Kul (eds.), *New Sounds 2010: Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech*, 113-118. Poznan, Poland.

Flege, James E. 1995. Second language speech learning: Theory, findings, and problems. In Winifred Strange (ed.), *Speech perception and linguistic experience: Issues in cross-language research*, 233-277. Timonium, MD: York Press.

Flege, James E. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America* 89(1). 395-411.

Flege, James E. 1987. The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics* 15(1). 47-65.

Flege, James E. and Wieke Eefting. 1987. Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication* 6(3). 185-202.

Flege, James E., Frieda, Elaina M. and Takeshi Nozawa. 1997. Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics* 25(2). 169-186.

Flege, James E., Munro, Murray J. and Ian R. MacKay. 1995. Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America* 97(5). 3125-3134.

Flege, James E., Schirru Carlo and Ian R.A. MacKay. 2003. Interaction between the native and second language phonetic subsystems. *Speech Communication* 40(4). 467-491.

Flege, James E., Yeni-Komshian, Grace H. and Serena Liu. 1999. Age constraints on second-language acquisition. *Journal of Memory and Language* 41(1). 78-104.

Giegerich, Heinz J. 1992. *English phonology: An introduction*. Cambridge University Press.

Grosjean, François. 1989. Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language* 36(1). 3-15.

Guion, Susan G., Flege, James E. and Jonathan D. Loftin. 2000. The effect of L1 use on pronunciation in Quichua–Spanish bilinguals. *Journal of Phonetics* 28(1). 27-42.

Hawkins, Sarah, and Jonathan Midgley. 2005. Formant frequencies of RP monophthongs in four age groups of speakers. *Journal of the International Phonetic Association* 35(02). 183-199.

Hillenbrand, James M., Clark Michael J. and Robert A. Houde. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America* 108(6). 3013-3022.

Hopp, Holger and Monika S. Schmid. 2013. Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics* 34(2). 361-394.

Kondaurova, Maria V. and Alexander L. Francis. 2008. The relationship between native allophonic experience with vowel duration and perception of the English tense/lax vowel contrast by Spanish and Russian listeners. *The Journal of the Acoustical Society of America* 124(6). 3959-3971.

Kormos, Judit and Mariann Dénes. 2004. Exploring measures and perceptions of fluency in the speech of second language learners. *System* 32(2). 145-164.

Ladefoged, Peter, and Keith Johnson. 2014. *A course in phonetics*. Nelson Education.

Mayr, Robert, Price, Sacha and Ineke Mennen. 2012. First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition* 15(4). 687-700.

Mackay, Ian R.A., Flege, James E. and Satomi Imai. 2006. Evaluating the effects of chronological age and sentence duration on degree of perceived foreign accent. *Applied Psycholinguistics* 27(2). 157-183.

McAllister, Robert, Flege, James E. and Thorsten Piske. 2002. The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *Journal of Phonetics* 30(2). 229-258.

McCarthy, Kathleen M., Evans, Bronwen G. and Merle Mahon 2013. Acquiring a second language in an immigrant community: The production of Sylheti and English stops and vowels by London-Bengali speakers. *Journal of Phonetics* 41(5). 344-358.

Meir, Natalia, Walters, Joel and Sharon Armon-Lotem. 2016. Bi-directional cross-linguistic influence in bilingual Russian-Hebrew children. *Linguistic Approaches to Bilingualism* 7(5). 514-553.

Morrison, Geoffrey Stewart. 2008. L1-Spanish Speakers' Acquisition of the English /i/-/ɪ/ Contrast: Duration-based Perception is Not the Initial Developmental Stage. *Language and Speech* 51(4). 285-315.

Moyer, Alene. 2014. Exceptional outcomes in L2 phonology: The critical factors of learner engagement and self-regulation. *Applied Linguistics 35*(4). 418-440.

Moyer, Alene. 2004. Accounting for context and experience in German (L2) language acquisition: A critical review of the research. *Journal of Multilingual and Multicultural Development*, *25*(1). 41-61.

Nip, Ignatius SB and Henrike K. Blumenfeld. 2015. Proficiency and linguistic complexity influence speech motor control and performance in Spanish language learners. *Journal of Speech, Language, and Hearing Research* 58(3). 653-668.

Podlipský, Václav J., Skarnitzl, Radek and Jan Volín. 2009. High front vowels in Czech: A contrast in quantity or quality? *INTERSPEECH-2009.* [Online] Available from http://www.isca-speech.org/archive/interspeech_2009. [Accessed on 10February 2017].

Schmid, Monika S. 2011. *Language Attrition*. Cambridge University Press.

Skarnitzl, Radek. 2012. Dvojí i v české výslovnosti [Two kinds of i in the pronunciation of Czech]. *Naše řeč [Our Speech] 95*(3). 141-154.

Skarnitzl, Radek and Jan Volín. 2012. Referenční hodnoty vokalických formantů pro mladé dospělé mluvčí standardní češtiny [Reference values of vocalic formants for young adult speakers of standard Czech]. *Akustické listy [Acoustic Letters]*18. 7-11.

Skarnitzl, Radek, Šturm, Pavel and Jan Volín. 2016. *Zvuková báze řečové komunikace: Fonetický a fonologický popis řeči [Sound basis of speech communication: Phonetic and phonological description of speech]*. Prague: Karolinum.

Šimáčková, Šárka. 2003. Engela's Eshes": Cross-linguistic perception and production of English [æ] and [ɛ] by Czech EFL learners trained in phonetics. In Solé, Maria-Josep, Daniel Recasens and Joaquín Romero (eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (ICPhS 2003), Barcelona, 3-9 August, 2003. 2293-2296.

Stoehr, Antje, et al. 2017. Second language attainment and first language attrition: The case of VOT in immersed Dutch–German late bilinguals. *Second Language Research* 33(4). 483-518.

Verhelst, Norman, et al. 2009. *Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.

Wells, John C. 1982. *Accents of English,* vol. 1. Cambridge: Cambridge University Press.

Ylinen, Sari, et al. 2010. Training the brain to weight speech cues differently: A study of Finnish second-language users of English. *Journal of Cognitive Neuroscience* 22(6). 1319-1332.

Appendix

A language experience questionnaire included three accent-related questions: about bilinguals' self-perceived foreign accentedness (Q1: *In your perception, how much of a foreign accent do you have in English?*), about their desire to improve their accent in English (Q2: *How important it is for you to improve your pronunciation?*) and about the importance of accent-free pronunciation (Q3*: In your opinion, how important is it for an interpreter to speak without a foreign accent?*). The bilinguals responded on a nine-point Likert scale, ranging from one, the least degree, to nine, the highest degree.

Speaker	Q1	Q2	Q3
LK	2	9	7
KH	2	9	6
ES	4	9	6
IP	4	8	7
TK	4	3	7
BH	4	7	7
TS	4	9	8
KK	4	6	6
MR	5	7	5
GT	5	9	6
KB	5	6	7
EP	5	7	4
HF	5	9	9
MK	5	7	6
MO	5	7	6
GP	6	8	4
KV	6	5	5
MN	6	8	7
TH	7	9	3
BR	8	9	8