Journal of Phonetics 77 (2019) 100920

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/Phonetics

Research Article

Third-language learning affects bilinguals' production in both their native languages: A longitudinal study of dynamic changes in L1, L2 and L3 vowel production

Natalia Kartushina [a,*], Clara D. Martin [b,c]

[a] Department of Psychology, University of Oslo, Forskningsveien 3A, 0373 Oslo, Norway
[b] Basque Center on Cognition, Brain and Language, Paseo Mikeletegi 69, 20009 Donostia San Sebastián Spain
[c] Ikerbasque, Basque Foundation for Science, María Díaz de Haro 3, 48013 Bilbao, Spain

A R T I C L E I N F O

Article history:
Received 21 February 2019
Received in revised form 19 July 2019
Accepted 9 September 2019
Available online 26 September 2019

Keywords:
L2 production
Bilingual production
L2-immersion
Phonetic drift
Language switch
L3 learning

A B S T R A C T

This study examined the impact of a study abroad (SA) English program on English and native vowel production. Basque-Spanish bilingual adolescents were assessed on their vowel production in English, Basque and Spanish before the SA program, the day after the program was completed, and four months later. The results revealed that after the SA program, participants' English vowels were acoustically closer to English norms, revealing the effectiveness of SA programs in improving English vowel pronunciation. Yet, four months later, these benefits had faded, showing that regular input and active language use are required to maintain accurate pronunciation. SA also had effects on native production: bilingual participants showed assimilatory acoustic drift in both their languages towards the English vowel system; the extent of this drift was negatively correlated with improvements in English pronunciation. However, four months later, participants showed a 'return' drift towards their native norms. The results also revealed that usage frequency and switching habits played a 'protective' role: Frequent switching in bilinguals made the dominant native language less vulnerable to foreign-language influence. Our results suggest that factors related to the frequency and circumstances of native language use are key to authenticity in native language production.

1. Introduction

Foreign language learning is highly promoted in modern society. Students are encouraged to study abroad (SA) to achieve the most effective learning outcomes. It is known that second language (L2) learning can have side-effects on native language (L1) production, as L1 sounds may show drift from L1 norms towards (or away from) the sounds of the L2. Previous research has shown that factors such as L2 proficiency, L2 frequency of use and the frequency of L1-L2 switches (among others) modulate the amount of L1 drift (for a review see Kartushina, Frauenfelder, & Golestani, 2016). However, to the best of our knowledge, no study has examined changes in native production as a function of foreign language learning in individuals who have two native languages (simultaneous or very early bilinguals). Our study attempts to shed light on this by addressing four questions. First, does foreign language learning (here, a third language, L3) affect bilinguals' production? Second, if so, what factors determine which of the bilinguals' languages is more susceptible to change? Here, we examine, for the first time, the effects of SA English learning on the production of vowels in the two native languages of simultaneous Spanish-Basque bilingual adolescents and how these effects are modulated by the frequency of language use and switching. Our third question is whether a SA English program is effective in improving the pronunciation of L3-English vowels in the short and long term. This issue has not yet been examined, despite the popularity of SA programs. Finally, the fourth question is whether there is a relationship between the degree of change in native production and improvements in L3 pronunciation.

* Corresponding author.
E-mail address: Natalia.Kartushina@psykologi.uio.no (N. Kartushina).

https://doi.org/10.1016/j.wocn.2019.100920
0095-4470/

1.1. Effectiveness of foreign language immersion programs

Second language learning through immersion is considered to be the most efficient way to acquire a foreign language (see Lafford & Collentine, 2006 for a review). SA improves oral (Freed, So, & Lazar, 2003; Segalowitz & Freed, 2004) and verbal (Linck, Kroll, & Sunderman, 2009) fluency as well as grammatical (Howard, 2004) and semantic knowledge and awareness (Collentine, 2004) to a greater extent than study at home. These results are due, in particular, to an increase in the quantity and quality of language input, since SA students receive all instructions in the target language (Genesee, 1985) and, importantly, produced by native speakers in a variety of linguistic materials (Lafford, 2006).

However, the advantage of SA is less evident for L2 pronunciation. While some studies report slight gains in the production of a group of sounds (e.g., fricatives [ß, ð, γ] in Lord, 2010; or voiceless plosives in Mora, 2008) or a specific sound (e.g., word-final [l] in Díaz-Campos, 2004; interdental and uvular fricatives in George, 2014), others show no gain (e.g., vowel lengthening in Simões, 1996) or even impaired production (e.g., overall non-word reproduction in Lord, 2006). These null or negative results are at odds with other studies showing that the amount of native-input is a strong predictor of (the authenticity of) L2 production in long-term immersed speakers (Flege, 2009), suggesting that input from native speakers supports accurate L2 pronunciation. Note that the overwhelming majority of studies on the effectiveness of SA learning have examined the acquisition of Spanish consonants by L1-American English learners. To the best of our knowledge, no study has examined improvements in the quality of vowel pronunciation for other target languages.

Given that (1) English is a *lingua franca* in modern society, (2) SA programs for English are very popular in Europe (and around the world) and (3) English vowels represent a challenge for L2 learners with various L1s (e.g., Spanish, Italian, Russian, French, Mandarin, Japanese, Korean, e.g., see Bohn & Flege, 1992; Casillas & Simonet, 2016; Flege, MacKay, & Meador, 1999; Ingram & Park, 1997; Oh et al., 2011), it is of urgent importance to examine the effects of SA on English vowel pronunciation. Note that one of the most popular SA programs for learning English in Europe, the Erasmus program, frequently takes place in non-English-speaking countries (such as the Netherlands, Germany, and Poland). Therefore, it is particularly important to examine the effectiveness of SA in a non-target-language country. Previous research has shown that SA in a non-target-language country significantly improves general English proficiency (accurate selection of grammatical and lexical items) and capacity for lexical complexity (Llanes, Arnó, & Mancho-Barés, 2016). No study, however, has examined the effectiveness of SA in a non-target-language country on L2 pronunciation.

Our study fills a gap in the SA language-learning literature and assesses the effectiveness of a SA English program in the Netherlands for the production of English vowels by experienced Basque-Spanish bilingual learners of English. Note that in contrast to previous studies, the SA program examined in the current study did not include language instruction, but rather extensive English exposure and, in particular, intensive English use in a contextualized and communicative interaction setting (discussions about the social, economic and political situation in Europe and future challenges). Research on foreign-language pronunciation teaching in an 'at home' institution has shown that meaning-based activities focused on providing learners with opportunities for practice lead to larger pronunciation improvements in spontaneous speech than does decontextualized, controlled practice with no elaboration (Derwing, Munro, & Wiebe, 1998; Gatbonton & Segalowitz, 2005; Mora, 2008; Saito & Lyster, 2012; Saito, 2012). Accordingly, we expected that a SA Erasmus program focused on communication practice, as examined in this study, would also have beneficial outcomes for participants' pronunciation. Unlike previous SA studies, our study examines the production of all monophthong vowels in the target language and assesses the maintenance of production gains (if any) four months after the immersion period. Previous research on L2 speakers and learners immersed in an L2-speaking country has shown that the quality of L2 pronunciation degrades when L2 speakers return to their L1-speaking environment (e.g., they lose the ability to accurately produce English stop consonants twelve months after the end of the SA program, see Mora, 2008). This suggests that regular input from native speakers is required to maintain authentic L2 production (Sancier & Fowler, 1997; Tobin, Nam, & Fowler, 2017). Native Spanish speakers have difficulties producing some English vowel contrasts; in particular, the vowels /i/-/ɪ/ (e.g., Flege, Bohn, & Jang, 1997; Morrison, 2008), /æ/–/ɑ/ (Casillas & Simonet, 2016; e.g., Flege et al., 1997), and /ʌ/-/ɑ/ and /u/-/ʊ/ (Flege & Wayland, 2019). Given the results of previous studies on consonant production (Díaz-Campos, 2004; George, 2014; Lord, 2010), we expected that a 2-week SA English program would help Spanish-Basque bilingual students improve their pronunciation of English vowels, yet these pronunciation gains might show deterioration four months later, when they returned home and no longer enjoyed regular English input and extensive English use (Hypothesis 1).

1.2. Effects of foreign language learning on native production

The second main goal of our study was to establish whether English learning affects bilinguals' production in their native languages and, if so, what factors might modulate the strength of this impact on each language. Since, to date, no study has examined changes in simultaneous and early bilinguals' phonetic production as a function of L3 learning, we will summarize the literature on the effects of foreign language learning and language immersion on L1 production in L2 learners.

It is largely accepted that in bilinguals and L2 learners the L1 and L2 sounds coexist in a common phonological space and constantly interact with/co-influence each other (Best, 1995; Escudero, 2005; Flege, 1995). The best known phenomenon associated with L1 influence on L2 production is having a 'foreign' accent, defined as "phonological cues, either segmental or suprasegmental, which identify the speaker as a non-native user of the language" (Scovel, 1969, p. 38) Generally, accents in segmental production happen when an L2 sound *assimilates* to a phonetically similar L1 category, such that the latter is used to produce both the L1 and the similar L2 sound (e.g., the Japanese /r/ is used to produce the English

/r/ and /l/ sounds, Aoyama, Flege, Guion, Akahane-Yamada, & Yamada, 2004). However, an L2 sound can also *dissimilate* from the L1 category (to eventually establish a new category) if L2 speakers are able to discern phonetic differences between the similar sounds in the two languages. For example, highly-experienced native Italian speakers of English tend to exaggerate tongue movements when producing the English /eᶦ/ vowel in order to dissimilate/differentiate it from the similar sounding Italian vowel /e/[1] (Flege, Schirru, & MacKay, 2003). Assimilatory and dissimilatory drifts in the production of L2 sounds are dynamic and are already observed after a short period of language immersion (Levy & Law, 2010; Sancier & Fowler, 1997; Tobin et al., 2017).

L1 phonetic categories can also show assimilatory or dissimilatory drifts. The nature and the extent of L1 drift depend on a number of factors, including those related to L2 experience and L2 use (for a review see Kartushina, Frauenfelder, et al., 2016). In particular, a recent analysis of the literature suggests that greater experience with the L2 (related to higher L2 proficiency, but also to more frequent L2 use) is associated with more highly L2-accented L1 productions (i.e., assimilatory drift towards the L2), whereas little L2 experience and poor proficiency help keep L1 productions close to monolingual norms (Kartushina, Frauenfelder, et al., 2016). These observations support the Speech Learning Model (SLM), which states that as experience with an L2 increases (due to more frequent L2 use and input), so does its influence on the L1 (Flege, 1995).

Yet, recent studies suggest that the L2 starts affecting L1 production from the beginning of L2 learning and that this effect is further reinforced in L2-immersion contexts. For instance, Chang (2012) reported that novice SA language learners (immersed in an L2 environment) with no prior experience of a target language showed assimilatory drift in their L1 productions towards the phonetic properties of the L2 system. Note that this drift was not limited to specific L1 sounds, but operated at the systemic level: the vocalic space in native American English speakers showed a drift in the first formant (F1) towards the higher F1 values of the L2-Korean vowels (Chang, 2012; see, however, Lang & Davidson, 2017 who found no drift in the L1 production of American English learners of French). Similarly, even without immersion in an L2-speaking country, short extensive training with an L2 leads to assimilatory drift of L1 categories towards similar L2 sounds (Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2016; Schuhmann & Huffman, 2015). L1 assimilatory drift has been attributed to L2 learners making a strong perceptual association between similar L1 and L2 sounds; thus a change in the production of one target (L2) leads to a change in the production of the related sound in the other (native) language (Chang, 2012; Huffman & Schuhmann, 2016). Longer L2-immersion, which is often associated with more extensive L2 use in diverse contexts, leads to L2-accented L1 speech (for example, Lang & Davidson, 2017; Lev-Ari & Peperkamp, 2013; Chang, 2019) and, eventually, to L1 attrition, i.e., a decline in proficiency in the native language (De Leeuw, Schmid, & Mennen, 2007, 2010; Flege, 1987; Major, 1992; Mayr, Price, & Mennen, 2012).

[1] Yet, in Flege and colleagues' experiment, Italians' /eᶦ/ productions were not perceived as native-like by native English listeners.

Interestingly, even in experienced L2 speakers, short L2-immersion might result in an assimilatory L1 drift towards the phonetic properties of the L2 (Chang, 2013; Sancier & Fowler, 1997), although to a lesser extent than in novice L2 speakers (Chang, 2012). This assimilatory drift, however, is unstable, as L1 productions have been shown to *shift back to L1* monolingual norms when the immersion ends (Sancier & Fowler, 1997). However, in the latter case study, the end of the immersion period coincided with an absence of L2 use; so, it is not clear whether the 'return' to L1 norms was due to the lack of L2 use per se, or to extensive L1 exposure and use. A recent study that simulated the phonetic drifts reported in Sancier and Fowler's article using a dynamical systems approach suggests that phonetic drifts in speech production reflect changes in the amount of language use and exposure (Tobin et al., 2017). Thus, little or no L1 use during L2 immersion results in a remarkable drift of native sounds towards similar L2 sounds. Intensive L1 use and L1 exposure, on the other hand, protect the L1 from L2 influence, because L1 categories are 'updated' throughout and remain attuned to the phonetic features of the ambient L1 language. Therefore, L1 phonetic drift, resulting from L2 immersion, shifts back to monolingual norms with extensive L1 use. In sum, the above-presented studies of novice L2-learners and L2-speakers in immersion suggest that L2 learning prompts L1 assimilatory drift, which is strengthened in prolonged L2-immersion contexts. However, more research is needed to understand the effects of the L2 on the L1 when L2-immersion ceases, and, in particular, when L2 learning continues [in non-immersive forms], as is the case for numerous SA students.

Other studies have shown that L2 learning can lead to L1 dissimilatory drift. This dissimilatory drift has been observed, in particular, in early L2 learners (Flege & Eefting, 1987b; Harada, 2003), who are claimed to be more likely to perceive the L2 and L1 sounds as sufficiently dissimilar to lead to new category formation (SLM, Flege, 1995), and has been attributed to an enhancement in the phonetic differences between similar L1 and L2 sounds. For instance, early Spanish speakers of English (age of acquisition between 5 and 6 years) produced the Spanish /p/, /t/, and /k/ consonants with shorter VOTs than monolingual Spanish speakers, presumably in order to increase contrast with similar consonants (with long VOTs) in English (Flege & Eefting, 1987b). Analogous L1 drift has been observed in late L2 learners in the process of establishing new categories for L2 sounds (Flege & Eefting, 1987a; Huffman & Schuhmann, 2016; Huffman, Schuhmann, Keller, & Chen, 2017). In particular, a recent longitudinal study by Huffman and colleagues (2017) revealed that after a one-semester Japanese language course, native English learners showed a dissimilatory increase in their VOTs for voiceless native stops. The authors proposed that this dissimilatory drift served two purposes: first, to increase the distance from similar (short-lag) Japanese stops and, second to maintain the native voice-voiceless contrast, which would have been threatened had native voiceless stops drifted towards the Japanese ones.

In sum, a handful of studies investigating the effects of L2 learning on L1 phonetic production suggest that L2 learning leads to assimilatory drift in the production of native sounds, unless L2 learners have established new categories for L2 sounds and, hence, dissimilate L1 categories from them (i.e.,

Table 1
Summary of the questions addressed in this study with corresponding hypotheses for the short-term (the day after participants' return from the study abroad [SA] program, Time 2 [T2]) and long-term effects (four months after the SA, at Time 3 [T3]) on speech production.

Questions	Hypotheses for short-term effects	Hypotheses for long-term effects
Q1. Does SA improve the pronunciation of L3 vowels and, if so, do pronunciation gains persist four months later?	H1: Intensive exposure to native[1] L3 speakers and L3 use improve L3 pronunciation accuracy	H1: Slight deterioration in L3 pronunciation due to extensive use/exposure to the native languages (L1 and L2) and decreased exposure to L3 speech
Q2. Does intensive English use during the SA program lead to changes in bilinguals' native production (in Spanish and Basque) and, if so, are these changes maintained four months after the end of the SA period?	H2: L3 learning results in assimilatory drift of native production towards L3 norms	H2: Native sounds show drift (back) towards native norms
Q3. Are changes in native production modulated by the frequency of native language use and switching habits?	H3: The native language that is used less frequently is affected more. However, bilinguals who practice regular switching between their native languages show L3 influence on both their languages	H3: In L1 environments, i.e., after bilinguals return home, the language which is used more frequently shows more pronounced drift (back) to native norms. Bilinguals who practice regular switching between their native languages show similar change back to norms in both languages
Q4. Is there a relationship between the amount of drift in native production and the extent of improvement in L3 pronunciation?	H4: Bilinguals demonstrating greater improvement in L3 production at T2, i.e., who have started to or have already established new categories for L3 sounds (Flege & Eefting, 1987a), dissimilate native categories from similar L3 sounds, whereas bilinguals demonstrating little or no improvement show assimilatory drift	H4: Bilinguals who maintain SA-related improvement in L3 pronunciation, as indicated by no changes in L3 pronunciation accuracy between T2 and T3, show less pronounced change back to native norms

[1] Although, the SA program was in the Netherlands, the Spanish-Basque bilinguals had regular exposure to native speakers of English (e.g., teachers, program coordinators and students); in addition, they interacted with Dutch students, who were highly proficient in English.

increase the distance between the L1 and L2 categories). Surprisingly, to the best of our knowledge, only four studies have simultaneously examined changes in *L2 and L1 sound (consonant) production* as a result of L2-learning (Chang, 2013, 2019; Sancier & Fowler, 1997; Schuhmann & Huffman, 2015) and only one of these (Schuhmann & Huffman, 2015) investigated whether changes in L2 production were related to the extent of L1 drift. However, the results of this latter study were inconclusive due to the small number of participants ($n = 5$) and considerable individual differences. Our study fills this gap in the literature on foreign-native language contact and examines, longitudinally, the relationship between improvements in foreign-language production and the amount of change in native phonetic production in experienced (advanced) Basque-Spanish learners of English who completed a two-week intensive SA English learning program. Consistent with the presented literature, we predicted that, at the end of the SA program, Spanish-Basque speakers would exhibit assimilatory drift towards the English system in their native production; however, four months after the SA program, their native production was expected to revert to native norms (Hypothesis 2). In addition, we expected that the direction and the amount of drift in bilinguals' native vowel production would be related to improvements in their production of English vowels. Specifically, we anticipated that greater improvement would be associated with (greater) dissimilatory drift and little or no improvement would relate to (greater) assimilatory drift (Hypothesis 4 in Table 1).

Our study was designed to shed new light on how English language learning affects native production. Importantly, SA program participants tested in the current study were early bilinguals, giving us the opportunity to explore the effect of English learning on two (and not only one) native languages. Thus, the third aim of the study was to specifically explore, for the first time, any differences in the impact of L3 training on the two native languages. For this purpose, we examined two important (relevant to the current study) factors that have been shown to trigger changes in the production of both languages in bilingual speakers.

1.3. Factors affecting bilinguals' production

Simultaneous and early (before the age of three) bilinguals produce both L1 and L2 speech sounds in a manner that does not differ phonetically from the monolingual norms of their two languages (Barlow, 2014; Barlow, Branson, & Nip, 2013; Guion, 2003; MacLeod, Stoel-Gammon, & Wassink, 2009; Sundara, Polka, & Baum, 2006). A recent study has shown that even trilingual speakers approach the productions of monolingual speakers in each of their three languages (Schoormann, Heeringa, & Peters, 2017). These studies suggest that very early exposure to two or more languages enables listeners to partition their phonetic space in order to accommodate the speech sounds of their native languages in a monolingual-like way.

However, recent studies suggest that bilinguals' speech production can drift (from monolingual norms) in either of their languages if, on a daily basis, they use this language less frequently than the other. For instance, early Catalan-Spanish bilinguals who used Catalan (their L1) less frequently than Spanish, showed drift in their production of Catalan vowels towards similar Spanish categories (Mora, Keidel, & Flege, 2015; for effects on perception, see also Mora & Nadeu, 2012). Similarly, studies of highly proficient late L2 learners have shown that the extent of L1 use is related to the degree of foreign-accentedness in L1 speech, with less L1 use being associated with more accented L1 speech (De Leeuw, Schmid, & Mennen, 2007). Analogous results were obtained in a recent SA study where the absence of changes in students' L1 production was attributed to extensive use of the native language during the immersion period (Tobin et al., 2017). The authors hypothesized that the extensive use of a native language outside L2 classes limited and eventually prevented any L2 influence on the L1 production; they confirmed this hypothesis using a simulation model. A growing body of research on language contact in bilinguals suggests that not only the frequency, but also the circumstances of language use influence the authenticity of speech production. In particular, language-mixing (or code switching) has consistently been reported to affect bilinguals' production in both languages. For instance, among native Dutch highly proficient English speakers, those who used Dutch in language-mixing contexts were perceived to have a stronger foreign accent in Dutch, than those who tended to use Dutch in contexts with no language mixing (De Leeuw et al., 2007, 2010). Other research has associated switching costs with low inhibitory control (Linck, Schwieter, & Sunderman, 2012), that leads to greater activation of the language not in use (here, native) and, hence, greater L2->L1 influence (Lev-Ari & Peperkamp, 2013). In the same vein, using a switching paradigm, Antoniou and colleagues showed that, in language-switching trials, L2-dominant early bilinguals demonstrated assimilatory drift in L2 production towards similar L1 categories (Antoniou, Best, Tyler, & Kroos, 2011; for similar results in late L2 learners see Goldrick, Runnqvist, & Costa, 2014).

To sum up, research on bilinguals' speech production suggests that bilinguals' phonetic systems interact with each other and are susceptible to changes as a function of the frequency and circumstances of language use: the less a language is used, and the more switching to the other language that occurs, the less native-like production becomes. Consistent with these studies, we expected that the frequency and circumstances of language use (the self-reported frequency of switches from one language to another ranging from 'never' to 'all the time') would modulate the extent of L3 influence on bilinguals' native production. In particular, we hypothesized that Spanish-Basque bilinguals learning English in immersive contexts would exhibit more English influence when the native language was used less frequently and less influence when the native language was used more frequently (Hypothesis 3). In addition, we hypothesized that those bilinguals who practiced regular language switching (so the phonologies of their two languages interacted regularly) would exhibit L3-English influence in both their native languages, whereas non-switching peers would exhibit L3 influence in only one of their native languages (due to the lower likelihood of regular interaction between the two phonologies, see Hypothesis 3). All of the study hypotheses are summarized in Table 1 together with the four related questions.

2. Methods

2.1. Participants

Ten Spanish-Basque early bilinguals took part in the study (age range 16–17, four males). All bilingual participants were highly proficient in both their languages: they had been exposed to two languages either from birth ($n = 4$) or since before the age of 3 ($n = 5$, with the exception of one participant, who started learning Basque at the age of 6) and used both languages in daily life. Participants were exposed to Basque during child-care, in kindergarten and at primary school, while both Spanish and Basque were languages of learning at school.

An analysis of a language background questionnaire revealed that the majority of bilinguals (7 out of 10) used Spanish more frequently than Basque (see Table 2). To obtain this measure, we asked participants to estimate, on average, how much time they used/were exposed to each of their native languages ("Indica el porcentaje medio del tiempo que estás expuesto a/usando cada una de las lenguas"). Imbalances in the frequency of language use ranged from 20% to 75% for Spanish minus Basque and from 10% to 15% for Basque minus Spanish. One participant reported using both languages equally frequently. Male participants had the highest imbalances in frequencies (50%, 55% and 75%), indicating that the majority of male participants (3 out of 4) were highly Spanish dominant, as compared to the more balanced female participants. We computed a *language use* measure as the difference in frequencies of language use between Spanish and Basque. Zero indicated that both languages were used equally frequently; negative scores meant that Basque was used more frequently than Spanish, whereas positive scores meant that Spanish was used more frequently than Basque. The language use values ranged from −15 to 75 (Table 2).

An analysis of the switching habits questionnaire – adapted from Rodriguez-Fornells and colleagues' study (Rodriguez-Fornells, Kramer, Lorenzo-Seva, Festman, & Münte, 2012) – revealed that 8 out of 10 bilinguals regularly switched from one language to another while speaking either Spanish or Basque. Note that each participant reported different switching habits for each of his languages: for example, participant S3 reported switching frequently to Spanish when speaking Basque, but only very rarely to Basque when speaking Spanish. Two participants reported that they never switched from Spanish to Basque and two reported the opposite. All Spanish dominant bilinguals reported switching to Spanish when speaking Basque (either occasionally or frequently). There was no consistent switching pattern among the balanced group of bilinguals, however. Given that each participant reported different switching habits for each of his languages, we examined the switching habits separately for each language, and not conjointly as stated in the hypotheses (H3).

For each participant, we obtained two *switching scores* (one for each language); these scores ranged from 1-never to 4-frequently (as none of the participants reported '5-always', see Procedure for detail). Higher scores for a target language indicated more switches to the other language when speaking the target language.[2]

All participants had been exposed to some English at preschool (e.g., singing songs, color naming, number counting, etc.), however, their formal English learning started at school at the age of seven. Prior to the experiment, on a separate day, participants filled in the language background questionnaire and were interviewed, individually, by a native speaker of English to assess their English proficiency. The interviewer asked questions on different topics and rated participants' production on a scale from 1-very low to 5-close to native (the final scores were converted/doubled for ease of comparison with native language proficiency scores, which ranged from 1 to 10). The score reflected overall fluency, comprehensibility,

[2] For the statistical analyses the scores were recoded so that 1 corresponded to 0 and 4 corresponded to 3.

Table 2
Participants' linguistic profile.

ID	Sex	Age	Age of Acquisition (years)			Self-rating proficiency (max 10)			ENG proficiency (max 10)	Language use SP-BSQ	Relative language dominance
			SP	BSQ	ENG	SP	BSQ	ENG			
S1	F	16	2	0	4	10	10	8	6	−15	Balanced
S2	F	16	0	0	5	8	7	7	6	20	Balanced
S3	M	17	0	0	6	9	8	5	6	75	SpDomin
S4	M	17	0	1	3	8	7	5	6	50	SpDomin
S5	F	16	0	3	4	9.5	8	7	6	30	SpDomin
S6	F	17	0	6	7	8	8	5	4	0	Balanced
S7	F	16	0	3	3	10	9	8	6	50	SpDomin
S8	M	17	0	3	6	9	9	7	6	55	SpDomin
S9	M	16	0	0	3	10	9	7	8	−10	Balanced
S10	F	16	0	0	5	10	8	9	4	25	Balanced
Average		16.4	0.2	1.6	4.6	9.2	8.3	6.8	5.8	28	

Note: SP – Spanish, BSQ – Basque, ENG – English, SpDomin – Spanish Dominant, Language Use is the difference between the frequencies of Spanish and Basque use (negative values indicate more Basque use). English oral proficiency was assessed by a native English speaker on a scale from 1-very low, to 5-close to native, and reflected overall fluency, comprehensibility, intelligibility, accurate pronunciation and grammar use. The final English proficiency scores were converted/doubled for ease of comparison with native language proficiency scores, which ranged from 1 to 10.

intelligibility, accurate pronunciation and grammar use. As can be seen in Table 2, prior to the SA, participants' overall proficiency in English was intermediate.

At the time of testing, all participants attended the Anunciata College (San Sebastian, the Basque Country, Spain) and were enrolled in a SA Erasmus program. The SA program was an intensive two-week course on the future of Europe which was held in the Netherlands, where pupils interacted with professors, program leaders and peers from the host institution. For six to eight hours each day, pupils took part in debates, discussions and presentations about the future of social and political life in Europe. All communication and interactions were in English. There was no English language instruction teaching. Apart from the course, pupils used English to communicate with their host families, who were proficient in English. Note that when pupils met each other during the SA period, the majority of interactions were conducted in Spanish, as reported by their L3-English teacher, who accompanied them for the duration of SA course. Thus, during the SA period, Spanish was used more than Basque.

2.2. Stimuli

2.2.1. Words

In order to compare the acoustic properties of vowels across English, Spanish and Basque, we selected words with similar phonetic environments for the target vowel, i.e., the preceding consonant was either /b/ or /p/, and the following consonant was always /t/. For English and Spanish, we adapted[3] words from Bradlow (1995) study, where the author conducted an acoustic comparison of vowels between the two languages. The English words were all monosyllabic (beat, bit, bet, bat, pot, bought, put, boot, but), exemplifying the 9 monophthongal (here British) English /i, ɪ, e, æ, ɒ, ɔ, ʊ, u, ʌ/ vowels (Deterding, 1997). The Spanish words were dissyllabic (bita, beta, bata, bota, puta), exemplifying the five monophthongal Spanish /i, e, a, o, u/ vowels. In accordance with Basque phonotactics and syllabification, the following five Basque words were selected: pita, bete, batu, botu, putak. They exemplified the five monophthongal Basque /i, e, a, o, u/ vowels. It is important to note that, although all the words used in the Spanish condition also exist in Basque, we selected other non-cognate Basque words, so that the task would approximate a monolingual-Basque mode. All Spanish and Basque stimuli had penultimate stress. Each word was presented five times; therefore, there were five exemplars of each vowel for each language. In total, at each testing session, each participant read 95 words.

To refine our hypotheses on phonetic drift in native production, we examined vowel spaces in Spanish and English as produced by respective native speakers (similar to Chang, 2012). The Spanish vowel space was used as the reference, given that Spanish and Basque share the same vowel inventory, i.e., both have /i e a o u/, and that vowels are realized phonetically similarly in the two languages (Egurtzegi, 2013). Note, however, that, in order to increase the phonetic differences between the two languages, simultaneous and very early bilinguals might pronounce even very similar cross-language sounds differently (e.g., Quichua /i/ and Spanish /i/ in Guion, 2003). This hypothesis will also be tested in our analyses. As can be seen in Fig. 1, which depicts overall English (Deterding, 1997) and Iberian Spanish (Chládková, Escudero, & Boersma, 2011) vowel spaces, the English vowel space is considerably more extended for F1 and somewhat more extended for F2, compared to the Spanish vowel space. These differences in F1 are particularly noticeable in female speakers, with the average F1 being larger in English (612 Hz) than in Spanish (546 Hz) and the acoustic differences between similar cross-language vowels reaching 200 Hz in some cases, for example, /e/-/e/ 'bet', /a/-/ʌ/ 'but'. Cross-language differences in F1 between male speakers are less noticeable, nevertheless, the pattern is the same as that for female participants: the average F1 in English (478 Hz) is larger than in Spanish (460 Hz) and similar cross-language vowels are acoustically different, with slightly higher (e.g., /e/-/e/ 'bet', /a/-/ʌ/ 'but') or lower (e.g., /i/-/i/ 'beat') F1s. Therefore, if assimilatory drift operates at the systemic level (as in Chang, 2012, 2013), we would expect to find larger F1 values for native vowels produced after the SA period, with female participants having more pronounced drift than male participants. However, if the drift operates at a vowel-specific level, then

[3] The items 'bait' and 'boat' containing diphthong vowels were not used in the study.

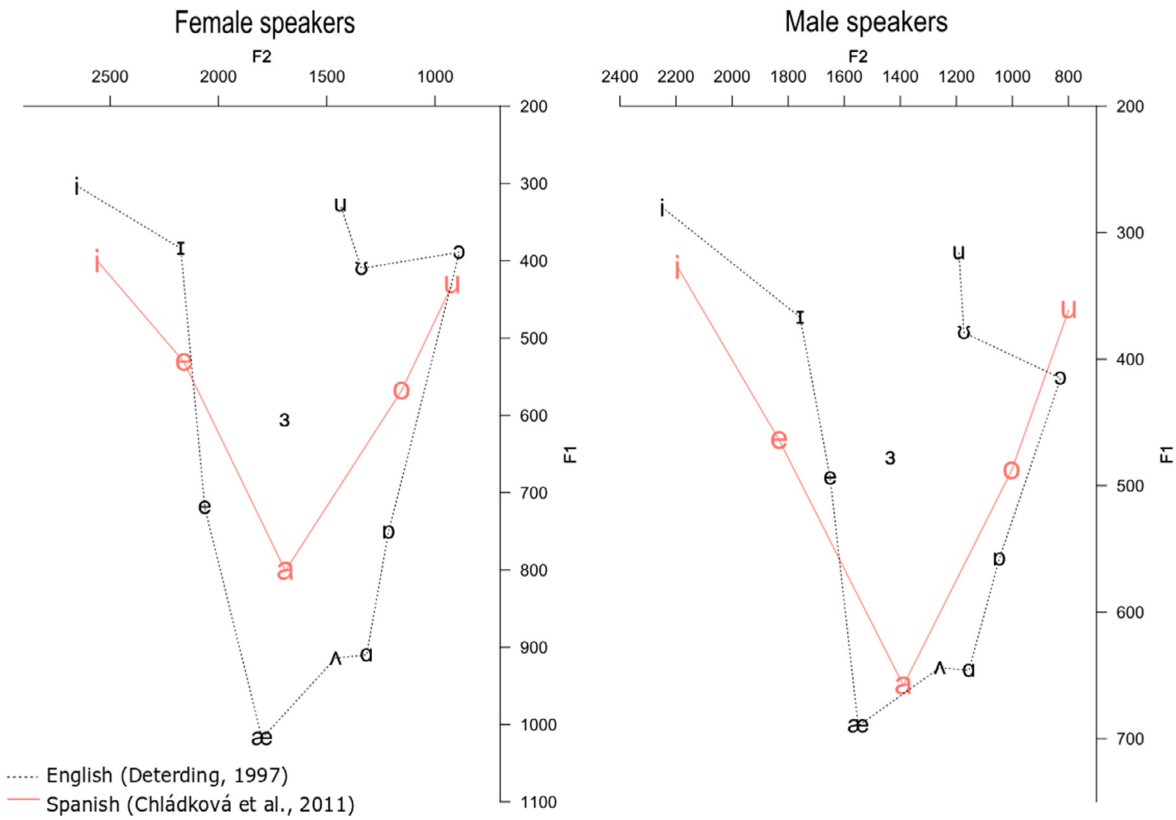


Fig. 1. F1/F2 acoustic spaces representing English vowels produced by 5 female and 5 male native British-English speakers (Deterding, 1997) and Spanish vowels produced by 10 female and 10 male native Spanish speakers from Madrid (Chládková et al., 2011).

we would expect higher F1 values for Spanish /e/, /a/, and /o/ vowels and lower F1 values for Spanish /i/ and /u/.

2.2.2. Audio recordings

Three male speakers were recruited to record the instructions for the participants: a native speaker of British English (36 years old), a native Spanish speaker (41 years old) from the Castile and León region (the standard Castilian language), and a native Basque speaker (29 years old) from Tolosa, where the Guipuzcoa dialect is spoken (the dialect of the participants in the experiment). Although the Basque speaker was a Basque-Spanish bilingual, he had first acquired Basque and considered himself to be Basque dominant. Each of the speakers had to read the instructions in his native language. For English, the instructions were as follows "Hello. Thank you for coming to our lab. In this task, you will have to read a few English words. On each trial, one word will be displayed on the screen. You will have to read this word out loud when a microphone appears on the screen. Please, try to read it as naturally as possible and at a moderate tempo. You will have to read the following English words: beat, but, bet, pot, put, bit, bat, bought, boot. Each word will appear five times. Please, push the SPACE bar to start". These instructions were translated and recorded in Spanish and Basque by native Spanish and Basque speakers, respectively; the word 'English' was changed to 'Spanish'/'Basque', and the English stimuli were replaced by Spanish/Basque stimuli. Recordings were carried out in a quiet room, using a Marantz PMD670 recorder and a Shure Beta 58A microphone, sampled at 22.05 kHz directly to 16-bit stereo .wav files.

2.3. Procedure

Each participant took part in one evaluation session and three testing sessions, i.e., T1, T2 and T3. At the evaluation session, held one week before the first testing session (T1), participants were interviewed by a native speaker of English to assess their English proficiency (see Participants section). The testing sessions took place at three time points: T1 – two months before the SA, T2 – the day after participants' return to Spain and T3 – four months after their return to Spain. In each testing session, participants performed, via the DMDX software (Forster & Forster, 2003), three reading tasks, one for each language, i.e., English, Basque and Spanish. The order of languages was counterbalanced. For each language, the same reading task was administered at T1, T2 and T3.

In each language, the reading task consisted of two phases: the instructions and the test. In order to trigger task performance in a language-specific mode (Grosjean, 2001), the instructions were presented, in the task-matched language, orally through the headphones fitted with a microphone (Sennheiser PC-350) and visually (written) on the screen. Recall that the instructions included the target words (see Stimuli section); therefore, prior to the test, all participants heard a native speaker producing the target words, and, thus, were familiarized with the stimuli (particularly important for English). Each test trial started with a 500-ms fixation cross that appeared in the middle of the screen; this was followed by a written target word with a picture of a microphone. The timeout was 2500 ms, that is, participants had 2500 ms, i.e., the length of the recording, to read the word. No audio information was available during the test phase. After the timeout, the next trial was initiated automatically. Within each language, each word was repeated five times; the stimuli were presented in a random order. Recordings were sampled at 22.05 kHz directly to 16-bit stereo .wav files; they lasted 12–15 min.

At the end of the third testing session, participants filled in a general language background questionnaire (to measure their daily use of Spanish and Basque) and a language-use questionnaire, that was designed to examine their language switching habits (if any). The language-use questionnaire – adapted from Rodriguez-Fornells and colleagues' study (Rodriguez-Fornells et al., 2012) – assessed, on a five-point scale from '1-never' to '5-always', the switching habits of Spanish-Basque bilinguals when speaking Spanish and Basque (see Appendix A).

2.4. Acoustic analyses of vowels produced in native (Spanish and Basque) and English (L3) languages

We performed acoustic analyses of the Spanish, Basque and English audio recordings (i.e., words) collected at T1, T2 and T3. In total, for each participant[4] we obtained 150 native (5 vowels × 5 repetitions × 2 languages × 3 sessions) and 135 English (9 vowels × 5 repetitions × 3 sessions) audio recordings. The recordings were verified for auditory quality, intensity and absence of noise (e.g., coughs, sneezes, sighs, etc.). Silent or unclear recordings were discarded from the analyses. Erroneous productions in English (e.g., 'bite' instead of 'beat') were also removed from the analyses. For each recording, the vowel stable portion was marked manually using Praat software (Boersma & Weenink, 2010). The onset was marked at the end of the formant transition following the initial consonant, where clear vowel tracks could be identified; the offset was marked at the beginning of the final formant transition to the consonant. F1 and F2 were computed at the midpoint of the vowel stable portion using an automated procedure in Praat. The following parameters were used: time step 0.01, maximum number of formants 5, maximum formant 5500 Hz for female and 5000 Hz for male participants, window length 0.025 s, preemphasis from 50 Hz. Extreme F1 and F2 values were detected with the *qqplot* function in R and were removed from the analyses. In total, 1430/1450 native (Basque and Spanish) and 1287/1305 English vowel tokens were retained for further analyses (0.014% of total data were discarded). The F1 and F2 of native tokens were separately submitted to statistical analyses.

In order to assess improvements in L3-English vowel production, we computed, for each testing session and for each vowel token, the *acoustic* Euclidian *distance* in F1/F2 space between the English vowels produced by our bilingual participants and the corresponding gender-matched English vowels produced by native speakers of British English, i.e., the norms (Deterding, 1997). Small distance values indicated an acoustic position closer to the target vowel. Extreme distance scores were detected using the *qqnorm* function; as a result, distance scores superior to 850 Hz ($n = 9$) were removed from the analyses. The remaining 1278 distance scores were used in the statistical analyses.

[4] Subject 3 was sick at T3 recording and could not attend; therefore, his data only contains recordings from T1 and T2 ($n = 190$ in total).

2.5. Statistical analyses

To answer the four research questions presented in Table 1, we performed two mixed effects regression model analyses in R (Core, 2012), using the *lmer* function from the lme4 package (Bates, Mächler, Bolker, & Walker, 2015), and two correlation analyses (for the last question). Each model included a number of fixed factors that differed across questions, and a common random factor, i.e., by-speaker intercept. To assess the significance of the main effects and their interactions, we used the *anova* function from the lmerTest package (Kuznetsova, Brockhoff, & Christensen, 2017) for all analyses. When multiple paired analyses were necessary, we used the *lsmeans* function implemented in the lsmeans package (Lenth, 2016), which applies the *Tukey* method to adjust the p-value for multiple comparisons.

2.5.1. Changes in English pronunciation accuracy across three testing sessions

The first model addressed the first question: **Q1**. Was the SA English program conducted in the Netherlands effective in improving the pronunciation of the English vowels and, if so, did these pronunciation gains persist, four months later? The *distance values* were fitted into a linear-mixed effects regression analysis; the fixed factors were Time (T1, T2 and T3), Gender (Male vs Female), Vowel (/i, ɪ, e, æ, ɑ, ɔ, ʊ, u, ʌ/), the three-way Time × Vowel × Gender interaction and its derivative two-way interactions. The random structure included by-Speaker random intercepts.

2.5.2. Acoustic changes in native (Spanish and Basque) vowel production across three testing sessions and the effect of native language use and switching habits on the extent of change

The second model addressed the second and the third questions: **Q2**. Did intensive English use during the SA program lead to changes in bilinguals' native production (in Spanish and Basque) and, if so, were these changes maintained four months after the SA? and **Q3**. Were changes in native production modulated by the frequency of native language use and switching habits? To answer these questions, F1s were fitted into the model, with the effects of Time (T1, T2 and T3), Gender (Male vs Female), Language[5] (Basque vs Spanish) and Vowel (/i/, /e/, /a/, /o/ and /u/) being included as fixed factors. The following interactions, stemming from our hypotheses, were also included as fixed factors. First, given that Spanish was used overall more frequently than Basque, we included a Time × Language interaction, expecting to find larger effects in Basque than Spanish. Second, to test whether the effects of Time differed across vowels and genders (as suggested by the acoustic comparison of the English and Spanish vowels in female and male speakers, see Fig. 1), we included a Time × Vowel and a Time × Gender interaction. Third, following our hypotheses on the effects of individual language use and switching habits (see H2 in Table 1), we included a Time × Use (language use measure) interaction, to examine the overall effect of Time as a function of the frequency of language use, and a Time × Lan-

[5] Although this paper does not examine differences in the production of speech sounds in bilinguals' native languages, i.e., between Spanish and Basque, we included Language as a fixed factor in our analyses since results from several studies have shown that simultaneous and very early bilinguals distinguish between similar L1 and L2 vowels in production (in order to enhance the phonetic differences between two vowel systems).

guage × Use interaction, to examine whether this effect differed between languages. Also, to test whether switching habits modulated the impact of Time on each language, we included a Time × Language × Switch (switching frequency scores) and a Time × Language interaction crossed with the factors Use and Switch (a four-way interaction) to test whether switching habits interacted with the frequency of language use. A similar mixed effects regression analysis was performed on F2s.

2.5.3. Relationship between improvements in English pronunciation and the amount of drift in native production

The next two correlation analyses addressed the fourth question: **Q4**: Were changes in bilinguals' native vowel production the day after the SA and four months later related to levels of improvement in and maintenance of English pronunciation immediately after the SA and four months later? To answer this question, we performed two correlation analyses: one for *short-term* changes (changes in F1 at T1 vs T2, T2–T1) and one for *long-term* changes (changes in F1 at T2 vs T3, T2–T3).[6] First, for each participant, we computed two measures of drift in native vowel production, short-term drift and long-term change. Higher values in short-term drift indicated larger assimilatory drift; higher values in long-term change indicated larger change back to native norms. Second, for each speaker we computed two measures of English pronunciation, a gain and a maintenance score, i.e., the difference in distance scores between T2 and T1 and between T3 and T2, respectively. A positive gain score indicated improvements in English pronunciation accuracy, whereas a negative gain score indicated deterioration in pronunciation. A positive maintenance score indicated improvements in English pronunciation four months after the SA, whereas a negative maintenance score indicated deterioration in English pronunciation four months after the SA. The first correlation analysis assessed the relationship between short-term changes in native production and gains in English, whereas the second correlation analysis assessed the relationship between long-term changes in native production and maintenance scores in English. Given our straightforward hypotheses, both correlation analyses were one-tailed.

3. Results

3.1. Changes in English pronunciation accuracy across the three testing sessions

The results of the distance scores analyses revealed a significant effect of Time, $F(2,1221) = 3.63$, $p = 0.027$, Vowel, $F(8,1216) = 51.80$, $p < 0.0001$,[7] Gender, $F(1,8) = 11.03$, $p = 0.01$, and significant Time × Vowel, $F(16, 1215) = 1.78$, $p = 0.028$, Vowel × Gender, $F(8,1216) = 18.66$, $p < 0.0001$ and Time × Vowel × Gender interactions, $F(16, 1216) = 1.91$, $p = 0.016$ (see Appendices B and C). As can be seen in Fig. 2, overall, the production of English vowels was acoustically closer to the target vowels after the SA period (+23 Hz, $\beta = 22.83$, se = 8.99, $t = 2.54$, $p = 0.030$) and, although numerically they appear to go back to their original position four months after the SA period (−20 Hz), this difference did not reach statistical significance ($\beta = -19.01$, se = 9.58, $t = -1.99$, $p = 0.12$), the effect size was $d = 0.38$.[8] Importantly, the differences between T1 and T3 were not significant ($p = 0.92$). The effect of Gender indicates that, overall, female participants had larger acoustic distances from target vowels than male participants.

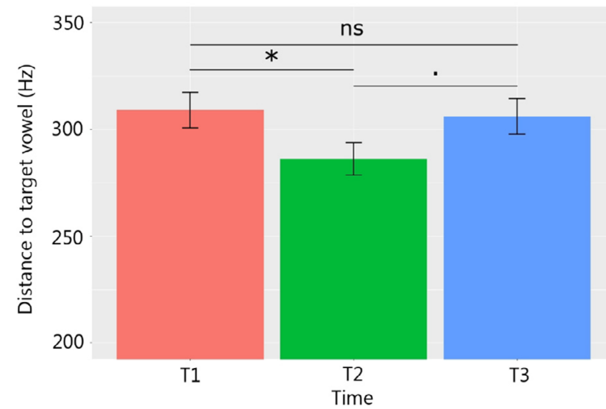


Fig. 2. Acoustic distance from L3-English vowels produced by Spanish-Basque bilingual participants to the target vowels produced by native British speakers at T1 – two months before the SA, T2 – the day after participants' return to Spain and T3 – four months after their return to Spain. Error bars indicate SEM. Asterisks indicate the significance level for a two-tailed test: * for $p < 0.05$, for $p = 0.12$; ns indicates a non-significant result.

The Time × Gender interaction was not significant ($p = 0.9$), suggesting that the benefits of the SA program and later deterioration in L3-English pronunciation were similar for male and female participants. However, as can be seen in Fig. 3 from the triple Time × Vowel × Gender interaction, improvements and deterioration in vowel production differed between male and female participants across vowels. This result is likely due to initial differences in pronunciation accuracy across English vowels between female and male participants.[9] The acoustic position of the English vowels in the F1/F2 space produced at T1 and T2 compared to the English norms can be seen in Appendices A and B.

3.2. Acoustic changes in native (Spanish and Basque) vowel production across three testing sessions

The results of the F1 analyses are summarized in Table 3. There was a significant effect of Time, with higher values of F1 at T2 than T1 (+13.5 Hz, $\beta = 15.55$, se = 3.87, $t = 4.02$, $p = 0.0002$), and lower values at T3 than T2 (−12.5 Hz, $\beta = -23.33$, se = 4.02, $t = -5.80$, $p < 0.0001$); the 1 Hz difference between T3 and T1 was not significant ($p = 0.13$), see Fig. 4. These results indicate that, after intensive L3-use, bilinguals' native vowel categories showed assimilatory drift towards the English vowel system, which, overall, has higher F1 than the native system; however, this drift was temporary: native vowel

[6] Correlation analyses were performed on F1 only, because there were no changes in F2 in native vowel production.

[7] Vowel specific effects across time points are beyond the scope of this paper. The interested reader can see Appendices A and B for the by-vowel figures.

[8] To estimate the effect size of the difference between the accuracies at T2 and T3, we used Lenhard and Lenhard (2016) calculator for computing the effect sizes from the test statistics. The following parameters were used: dependent mode of testing, $t = 1.99$ (the t-value of the lmeans test on the model, as reported in the manuscript), $n = 9$ (one participant did not take part in T3, see methods) and $r = 0.84$ (correlation between distance scores at T2 and T3).

[9] Although the authors found these vowel-specific differences between male and female participants across the three testing sessions very intriguing, further analyses and discussion of this result are beyond the scope of this paper.

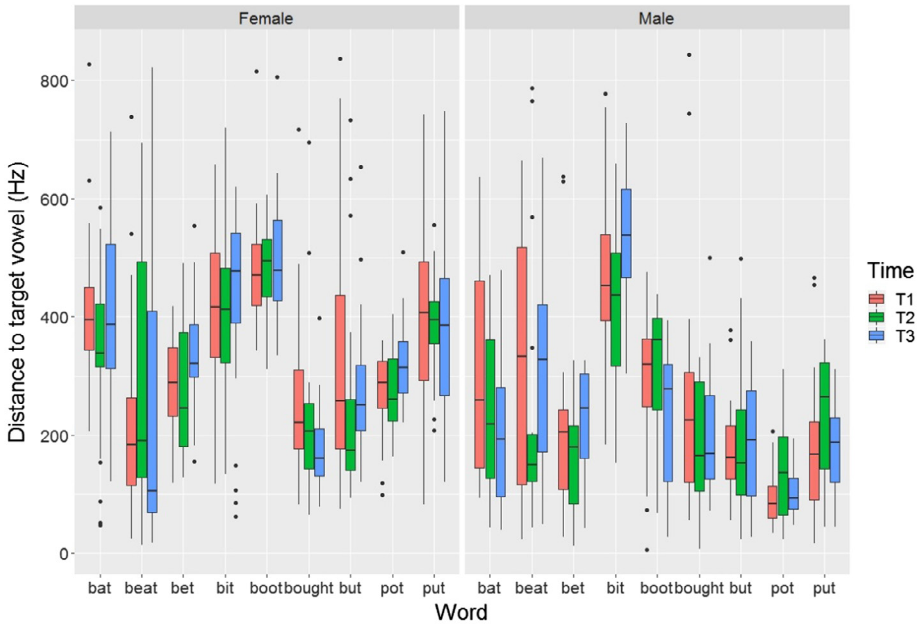


Fig. 3. Pronunciation of English vowels across tested words in female and male bilingual participants at T1 – two months before the SA, T2 – the day after participants' return to Spain and T3 – four months after their return to Spain.

categories returned 'back to norms' after a four-month period of regular language use in participants' native environment.

There was a significant effect of Language, with lower F1 in Basque (494.8 Hz) than in Spanish (497.3 Hz), and a significant effect of Gender, with female participants having higher values (+42 Hz) than male participants. There was also a significant effect of Vowel, indicating that the F1 differed across vowels. The absence of significant Time × Language and Time × Gender interactions indicated no difference in the effect of time between languages and genders, confirming that the assimilatory drift towards the larger F1 values of the English vowels was global. The marginal Time × Vowel interaction suggests that the effect of Time was relatively similar across vowels. As can be seen in Fig. 5, at T2, both male and female participants showed an increase in F1 across all native vowels, i.e., systemic drift towards the higher F1 values of English, and a relatively systemic decrease in F1 at T3, i.e., a return to native norms.

The results of a similar analysis on F2 revealed no effect of Time ($p = 0.9$), and no Time × Language ($p = 0.35$) interaction, indicating that intensive L3 use did not lead to changes in F2 in native vowel production. Participants' average F2s at T1 and T2 were 1665 Hz and 1666 Hz for female speakers and 1438 Hz and 1443 Hz for male speakers. There was an expected effect of Vowel, $F(4, 1384) = 9965.30$, $p < 0.0001$, and Gender, $F(1, 8) = 28.6$, $p = 0.0008$, with the vowels produced by female participants having, overall, higher F2 values (+213 Hz) than the vowels produced by male participants. Given that no effect of Time was found for F2, further analyses were performed on F1 only.

3.3. Role of native language use and switching habits on the extent of change in native production

3.3.1. Language use

The results shown in Table 3 did not reveal an interaction between Time and Use, suggesting that, overall, the effects of Time did not differ for participants who predominantly used either Spanish or Basque. However, a significant Time × Language × Use interaction suggests that these effects differed for the two native languages. In order to better understand how the frequency of language use modulated the effects of time on each language separately, we recoded the continuous variable Use into a categorical variable Group and assigned each participant to either the balanced or Spanish dominant group. Participants with low values of Use, <25 (those who used Spanish and Basque in a relatively balanced way, the maximum imbalance in frequency of use was 25%), were assigned to the balanced group, whereas participants with high values of Use, >30 (who used Spanish considerably more frequently than Basque, minimally 30% more) were assigned to the Spanish dominant group; there were five participants in each group (see Table 2 for details and Fig. 6).

As can be seen in Fig. 6, balanced bilinguals seemed to show similar effects of Time on both Basque and Spanish, whereas Spanish dominant bilinguals showed effects only on their Basque production. Multiple-comparison analyses confirmed this observation: at T2, balanced bilinguals showed an assimilatory drift for both Basque (+16 Hz, $\beta = 14.98$, se = 5.68, $t = 2.64$, $p = 0.023$) and Spanish (+18 Hz, $\beta = 18.27$, se = 5.65, $t = 3.23$, $p = 0.0036$), whereas Spanish dominant bilinguals showed an assimilatory drift in Basque only (+16 Hz, $\beta = 15.90$, se = 5.36, $t = 2.97$, $p = 0.0086$), while the numeric 3 Hz drift in Spanish did not reach significance ($p = 0.9$).[10] Four months after the SA period, at T3, balanced bilinguals showed a change back to native norms in both Basque (−23 Hz, $\beta = -21.92$, se = 5.71, $t = -3.84$, $p = 0.0004$) and Spanish (−26 Hz, $\beta = -28.63$, se = 5.70, $t = -5.01$, $p < 0.0001$), showing no differences between T3 and T1 for

[10] To perform multiple comparisons for the effects of Time as a function of Group and Language, *lsmeans* was run on a simplified version of the original model: the Language x Switch interaction was removed, and continuous factor Use was replaced by our new factor Group (Balanced vs Spanish dominant).

Table 3
Output of the anova function in R showing the statistical significance of the main effects and interactions tested in the model. Asterisks indicate the significance level of the factor: *** for $p < 0.001$, ** for $p < 0.01$, * for $p < 0.05$, for $p < 0.1$.

Factor	Sum Sq	Mean Sq	NumDF	DenDF	*F*	*P* (>*F*)
Time	40169	20084	2	1384.37	11.6895	9.241e−06***
Language	36506	36506	1	1390.28	21.2472	4.408e−06***
Gender	9878	9878	1	7.27	5.7491	0.046352*
Vowel	25635895	6408974	4	1384.04	3730.1505	<2.2e−16***
Time:Language	2178	1089	2	1384.16	0.6340	0.530640
Time:Gender	619	310	2	1385.36	0.1803	0.835070
Time:Vowel	24542	3068	8	1384.05	1.7855	0.075686
Time:Use	4325	1442	3	24.74	0.8391	0.485407
Time:Language:Use	14914	4971	3	1374.73	2.8934	0.034231*
Time:Language:Switch_center	62315	10386	6	1381.76	6.0447	2.945e−06***
Time:Language:Use:Switch_center	33403	5567	6	1341.48	3.2402	0.003641**

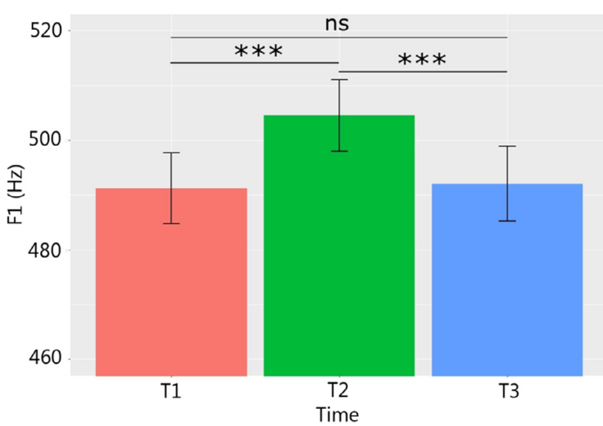


Fig. 4. Mean F1 of the native vowels produced by bilingual participants at different testing sessions: at T1 – two months before the SA, T2 – the day after participants' return to Spain and T3 – four months after their return to Spain. Asterisks indicate the significance level $p < 0.001$.

either language (7 Hz, $p = 0.4$ in Basque and 8 Hz, $p = 0.2$ in Spanish). Spanish dominant bilinguals, on the other hand, showed no change back to native norms at T3 in Basque (+1 Hz, $p = 0.6$). Likely because of the large variability of F1 at T3 in Basque, the numeric difference of 17 Hz between T3 and T1 did not reach significance ($p = 0.14$). In Spanish, the difference between F1 values at T2 and T3 was significant (−5 Hz, $\beta = -13.96$, se = 5.78, $t = -2.42$, $p = 0.042$), whereas the difference between F1 values at T3 and T1 in Spanish was not significant (2 Hz, $p = 0.10$).

3.3.2. Switch habits

Significant Time × Switch × Language and Time × Switch × Language × Use interactions (see Table 3) indicate that the effect of Time was modulated by the frequency of switches from one language to another and, importantly, that this effect differed across languages and depended on the frequency of language use. In order to facilitate the visualization and understanding of these interactions, we computed, for each participant, vowel and language, two *change* measures by subtracting the average F1 value at T1 from the average F1 at T2 and the average F1 at T2 from the average F1 at T3, to compute the respective short- and long-term changes in native vowel production. Fig. 7 visualizes the four term interaction that gives the most complete picture of this complex interaction between the amount and nature of changes in native vowel production and the frequency of language use and switching habits in Basque and Spanish. In balanced bilinguals, the frequency of switches had similar effects for both native languages and at both periods, i.e., for both short- and long-term changes (the day after participants' return from the SA program, at T2, and four months after the SA, at T3). In particular, in balanced bilinguals, switching played a 'protective' role during the period of English immersion: more switches were associated with less pronounced drift in both native languages, whereas four months after the end of the SA program, more switches led to a less noticeable 'recovery' of native norms. In Spanish dominant bilinguals, the effect of switching frequency was similar in both languages for short-term changes but diverged for Basque and Spanish for long-term changes. As can be seen in Fig. 7, in Spanish dominant bilinguals, the frequency of switches played a protective role for both languages during the immersion period, similar to the results seen for balanced bilinguals. However, four months after the SA, more frequent switches to Basque when speaking Spanish were associated with a larger change back to native Spanish norms. Recall, however, that there was no significant drift in Spanish after the SA program to start with, so this tendency should be interpreted with caution. In Basque, on the other hand, Spanish dominant bilinguals who more frequently switched to Spanish while speaking Basque showed less pronounced or no change back to Basque norms. Thus, these results suggest that frequent switches to Spanish while speaking Basque prevented Spanish Dominant bilinguals from recovering their Basque production even four months after the SA period.

3.4. Relationship between improvements in English pronunciation and amount of drift in native production

The results of the first correlation analysis between gains in English pronunciation and drifts in native production the day after participants' return from the SA program revealed a significant negative relationship ($r = -0.61$, $p = 0.031$), meaning that those participants who exhibited greater improvements in English pronunciation showed less assimilatory drift (towards English) in their native production (Fig. 8). The results of the second correlation analysis between maintenance of gains in English pronunciation and change 'back' to native norms four months after the SA program also revealed a significant positive relationship ($r = 0.72$, $p = 0.0094$). However, this correlation was likely driven by one of the participants who showed

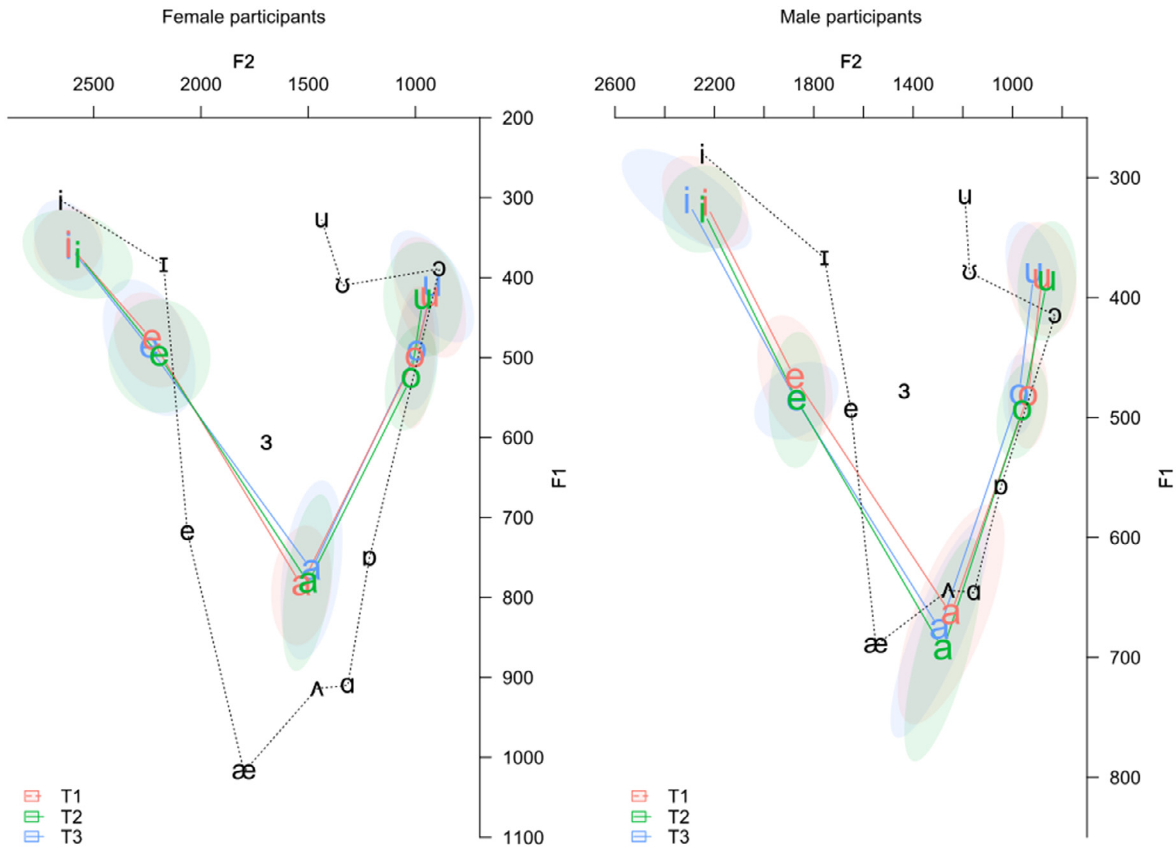


Fig. 5. The average position of the native (Spanish and Basque) vowels in the acoustic F1/F2 space produced by male and female participants at the three testing times along with the position of the English vowels (dotted line) produced by native British speakers of English (Deterding, 1997). The ellipses show the 68% confidence level.

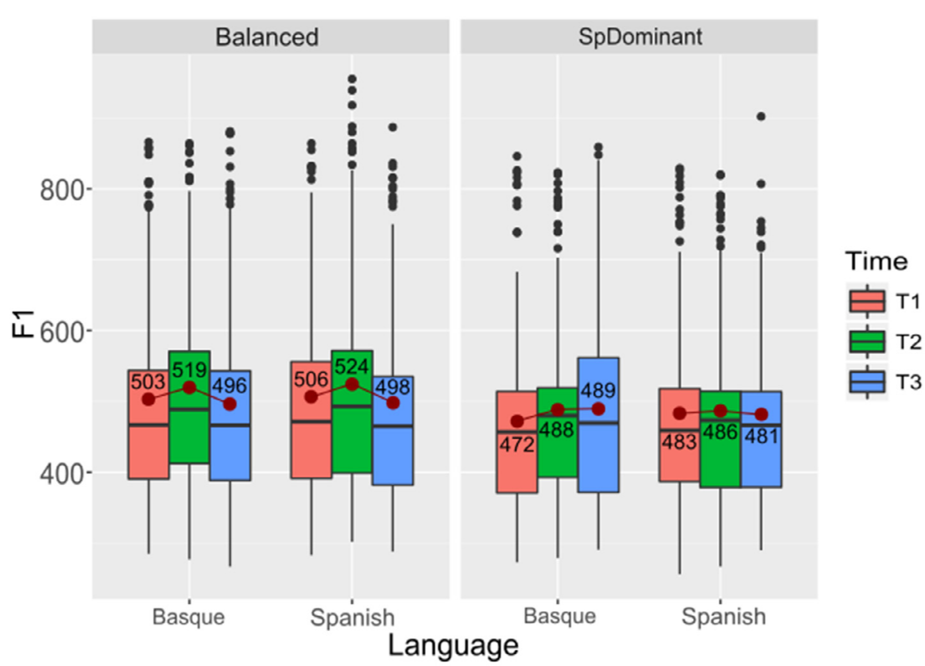


Fig. 6. Mean F1 of native vowels produced at different testing sessions by balanced and Spanish dominant bilingual participants in Basque and Spanish; T1 – two months before the SA, T2 – the day after participants' return to Spain and T3 – four months after their return to Spain. Dominance refers to the amount of language use.

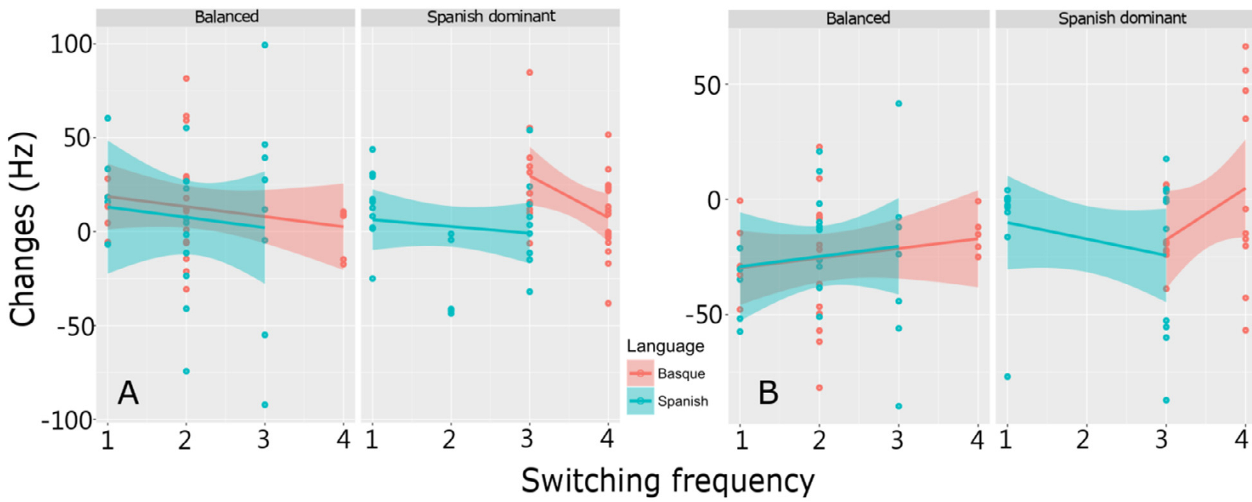


Fig. 7. Short-term (A) and long-term (B) changes in F1 in Basque and Spanish as a function of the frequency of switches (from 1-never to 4-frequently) from one language to another and the frequency of language use (Balanced vs Spanish dominant group). The absence of data points at a given switch frequency for a given language means that none of the participants reported switching at this frequency while speaking this language. Larger positive values for short-term changes indicate larger phonetic drift towards the English vowel system; larger negative values for long-term changes indicate larger phonetic drift towards the native vowel system.

extremely strong changes in native and English vowel production. When this participant was removed from the analyses, the correlation dropped to $r = 0.13$, suggesting that our data do not provide evidence for a relationship between the extent of deterioration in English pronunciation and change 'back' to native production norms. Additional robust correlation analyses (that allow us to take the behavior of outliers into account, without excluding them from the analyses) yielded similar results.[11] Additional exploratory correlation analysis (with the p-value set to 0.025) revealed a marginal negative correlation between gains in English pronunciation and the amount of change 'back to norms' in native production ($r = -0.59$, $p = 0.036$, see Fig. 8), indicating that those participants who showed more improvement in their English pronunciation tended to have less change back towards native norms four months after the SA period, likely because they had less pronounced drift in native production during the immersion period (see first correlation).

4. Discussion

4.1. Changes in English pronunciation accuracy one day after the SA program and four months later

The acoustic analysis of English vowels produced by Spanish-Basque bilinguals revealed that a two-week SA English program conducted in the Netherlands was effective in improving pronunciation of English vowels. On the day after their return to Spain, overall, participants' English vowel production was acoustically closer to the target English vowel norms. There were no differences in the amount of improvement between female and male participants, suggesting that all participants benefited from the SA program. Note that participants were not tested on their production of specific words chosen because they were frequently used words within the SA program, but rather on less frequently encountered words; this indicates that improvements reported in the current study reflect more accurate phonemic targets in overall English vowel pronunciation after the program, as compared to before. Our results for vowel production are in line with previous language-learning studies showing SA-related improvements in consonant production (Díaz-Campos, 2004; George, 2014; Lord, 2010) as well as in general English proficiency and the ability to handle lexical complexity in written English (SA in a non-anglophone country, Llanes et al., 2016). These results suggest that intensive foreign-language exposure and active foreign-language use help learners adjust their pronunciation to reach target norms, be they vowels or consonants. Note that our results are particularly encouraging for several reasons. First, the language program was relatively short (two weeks) as compared to traditional language SA programs (lasting from one to twelve months). Second, this program took place in the Netherlands, a non-anglophone country. Finally, and importantly, this SA program did not include language instruction (as provided to participants in previous SA research), suggesting that extensive English exposure and, in particular, intensive English use in a contextualized (discussions about social, economic and political situation of Europe and future challenges) communicative, goal-oriented classroom setting within the program framework were effective in improving participants' English pronunciation. These results are in line with a body of research on classroom-based second-language pronunciation teaching, showing that contextualized, meaning-based activities and communicative form-focused instructions with ample opportunities for practice lead to larger improvements in spontaneous speech than decontextualized, controlled practice with no elaboration (Darcy, 2019; Derwing et al., 1998; Gatbonton & Segalowitz, 2005; Mora, 2008; Saito & Lyster, 2012; Saito, 2012). Thus, a SA program in a non-anglophone country that does not include language instruction

[11] Given the presence of outliers in our data and the small sample size, we performed additional robust correlation analyses that allow us to take the behavior of outliers into account, without excluding them from the analyses. The results of the percentage bend correlation – *pball* function in MRS2 package (Mair & Wilcox, 2017) revealed a robust correlation coefficient $cor = -0.47$ (one-tailed $p = 0.08$) for the relationship between gains in English pronunciation and drift in native production at T2 (similar to previous analyses), and a non-significant robust correlation $cor = 0.37$ (one-tailed $p > 1$) for the relationship between maintenance of gains in English pronunciation and change 'back' to native norms, suggesting again, that our data do not provide evidence for a relationship between the extent of deterioration in English pronunciation and change 'back' to norms in native production.

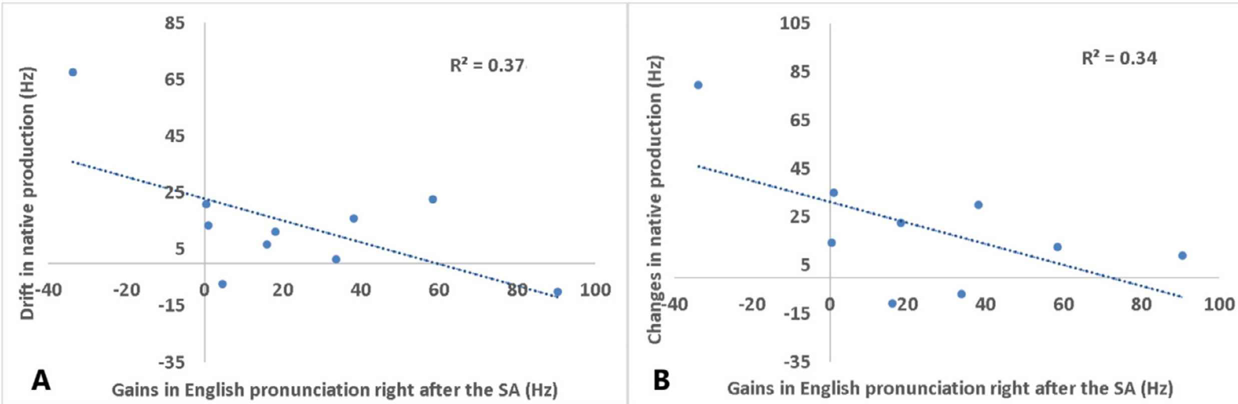


Fig. 8. Relationship between improvements in English pronunciation at T2 – the day after participants' return to Spain, and the amount of drift and change in native production (return drift) right after the study abroad period (A) and four months after their return to Spain (B), respectively. For (A), larger positive values on the y-axis indicate larger assimilatory drift towards English. Negative values on the y-axis indicate dissimilatory drift. For (B), higher values on the y-axis indicate larger changes back to native norms.

but is focused on communication and intensive interaction in a foreign language appears to be an effective means for improving segmental oral skills, i.e., vowel pronunciation as examined in this study.

However, improvements in English vowel pronunciation were not homogeneous across all tested vowels. As can be seen in Fig. 3 and as revealed by a significant vowel by time interaction, some vowels benefited from SA more than others. For example, while both female and male participants seemed to have improved their production of the /æ/ and /ɔ/ vowels (as in 'bat' and 'bought'), they appear to have deteriorated in their production of the /u/ vowel (as in 'boot'). These results are not explained by the 'more room for improvement' hypothesis, since all three vowels were approximately at the same distance from their targets before the SA program. The most likely explanation would be the 'general expansion' hypothesis; as can be seen in Appendices A and B, after the SA period, participants tended to approach English vowel targets by expanding their English vowel space. Thus, open front and central English vowels /æ/ and /ʌ/ ('bat' and 'but') became more open and front, whereas close back vowels /ɔ/ and /u/ ('bought' and 'boot') became 'more closed' and back. This general expansion strategy has been previously reported in novice L2 learners and experienced bilinguals. For instance, Chang (2012) revealed that, although American-English learners of Korean reached the Korean (long VOT aspirated stops) targets by the second week of a language-training course, they continued to increase VOT at the expense of native-like outcomes and also over-aspirated Korean stops by the end of the course. In the Guion (2003) study, Quichua-Spanish bilingual participants produced their Quichua vowels systematically higher than both Spanish vowels and Quichua monolingual norms, showing an upward extension of the vowel system. The author proposed that this shift was motivated by the need to create sufficient space for the Spanish vowels, even at the expense of less-native like outcomes for Quichua vowels. This kind of reorganisation of vowel space might be necessary for optimal accommodation of target vowels (Flege, 1995; Guion, 2003).

Four months after the SA program, bilingual participants showed a deterioration in their production of English vowels: acoustic distance to the target English vowels increased. Although, because of high between-speaker variability in the acoustic position of English vowels, this deterioration did not reach significance, the absence of statistically significant differences between the acoustic position of L3-English vowels before the SA and four months later suggests that the benefits of the SA program did not remain four months later; that is, participants returned to pre-SA levels. Note, however, that this 'return' to baseline did not generalize to all tested vowels. Pronunciation of some of the English vowels, which can be considered particularly difficult for Spanish learners of English, because they do not exist in the Spanish repertoire, and thus then to be assimilated to similar Spanish sounds, such as /ɪ/ ('bit') and /ɔ/ ('bought') in female participants and /æ/ ('bat') and /ʊ/ ('put') in male participants, showed persistent improvements and continued to improve even after the SA. The loss of pronunciation benefits four months after the SA program was likely due to a significant decrease in the amount of exposure to and active use of English. These results are consistent with the results of previous research on foreign-language learners immersed in an L2-speaking country (Sancier & Fowler, 1997; Tobin et al., 2017) and suggest that regular input from native speakers and active language use are required to maintain accurate sound pronunciation in a foreign language, at least for learners with intermediate levels of language proficiency, similar to those tested in this study.

4.2. Acoustic changes in native (Spanish and Basque) vowel production after the SA program and four months later

After the SA English program, bilinguals' native vowel categories showed an assimilatory drift towards the English vowel system, which has overall higher F1. Drift in native vowel production was observed only in F1 (larger vowel openness), with no changes in F2. Our results for Spanish-Basque bilingual learners of English are similar to those reported for American-English learners of Korean (Chang, 2012, 2013) and suggest that intensive foreign language use and immersion learning lead to systemic phonetic drift in native vowel production. The amount of drift was not dependent on the initial acoustic position of the native vowels with regard to English vowel space: both female and male participants showed similar amounts of drift, even though female participants' native vowels were acoustically further from the target English space.

As can be seen in Fig. 5, this drift did not operate at a vowel-specific, but rather at a systemic level, given that, in both male and female speakers, all native vowels showed larger F1 values after the SA period. The extent of this drift was similar to that reported for novice learners of Korean in Chang's studies, although, in the current study, participants spent less time in the foreign-speaking country than the English speakers in Chang's study.

Note that although participants were actively practicing English during their SA program (around 6–8 h per day), they continued using both their native languages (Spanish and, to a lesser extent, Basque) in their free time. Nevertheless, they showed drift in their native production, suggesting that native language use did not prevent this drift from happening. These results are at odds with previous studies showing that native language use prevents the influence of the ambient foreign language (Tobin et al., 2017). The likely explanation for this apparent discrepancy in the results between our study and those reported by Tobin and colleagues is the large difference between the amounts of foreign language use in the two studies (7 times more in our study). It is possible that when foreign language use reaches a certain level (a threshold), the native system cannot 'resist' the foreign-language influence and, thus, concedes. A similar dynamic has been observed in fluent bilinguals, whose phonetic production in one language is affected by the extent to which the other language is used (Mora & Nadeu, 2012; Mora et al., 2015).

Four months after the end of the SA program, participants showed a change back in native vowel production, that is, a return drift towards the acoustic positions they had exhibited before the immersion program. This drift was systemic (over the whole vowel space) and consistent: it generalized to both languages in both female and male participants. The influence of intensive English use and exposure, evident after the SA program, was no longer observable in bilingual students, likely because they considerably reduced their English use and regained more frequent use of their native languages. These results agree with those obtained in a case study on consonant production (Sancier & Fowler, 1997) and extend them to production of the whole vowel system, showing that authenticity in native language production is regained in a native language environment (see Linck et al., 2009 for a similar rebound in L1 verbal fluency six months after the immersion period). Taken together with the results from a recent study by Tobin and colleagues, who did not find drift in native production, our results suggest that phonetic accommodation (return drift) in native production happens only when significant changes (e.g., assimilation or dissimilation) are induced by foreign language use/immersion in a foreign language environment. In other words, accommodation to the phonetic norms of the native language (in a native language environment) occurs only when significant deviation from native norms has been induced by foreign language use. These results indicate, first, the robustness of the native production system, which is not easily influenced merely by exposure to a foreign language (see Fowler, Sramko, Ostry, Rowland, & Hallé, 2008, who found no effects on native production when a foreign language was overheard), but, also, its plasticity, because native production regains its initial state after being bathed in a native-language environment.

4.3. Role of native language use and switching habits on the amount of change in native production

Our results revealed that the size of the phonetic drift in native language production varied between languages as a function of the frequency of language use. After the SA program, balanced bilinguals, who used Basque and Spanish at relatively similar frequencies, showed an assimilatory drift towards the phonetic norms of English in both Basque and Spanish. By contrast, Spanish dominant bilinguals, who used Spanish more frequently than Basque (30% more), showed a similar-size drift in their Basque production only; their vowel production in Spanish remained almost intact (a non-significant 3 Hz assimilatory shift). These results are in line with the self-organizing dynamical system model, proposed by Tobin and colleagues (Tobin et al., 2017), and suggest that 'active' language use limited and eventually prevented a third language from influencing bilinguals' production in the more intensively used native language. Thus the effect of native language use does not generalize over both native languages, but is language specific. It is particularly interesting that only the vowel space of the less frequently used language showed an assimilatory drift towards the English vowel space, given that Basque and Spanish share the same vowel inventory (five vowels /i, e, a, o, u/ in each). Other research has shown similar dissociations between two languages in early (under 3 years of age) and simultaneous bilinguals. By producing similar cross-language sounds differently (Barlow et al., 2013; Fowler et al., 2008; Guion, 2003; Sundara et al., 2006), bilinguals effectively increase the phonetic distance between the two language systems. In sum, our results show that the effects of foreign (third) language use can dissociate between a bilinguals' two languages, but only if there is an imbalance in the frequency of use of these languages.

Analogously, the size of the return drift, i.e., the phonetic change back towards 'native norms' four months after the end of the SA program, was modulated by the amount of native language use, but differently for Spanish and Basque. Balanced bilinguals adapted their production in both their languages: after a four-month stay in a native language environment and usual language use, their vowel production in Spanish and Basque was similar to their production before study abroad in the Netherlands. Spanish dominant bilinguals, on the other hand, showed no adaptation to native Basque norms. Note that Spanish dominant bilinguals' production in Spanish was not affected by their intensive SA English language use and remained stable across the three testing sessions. However, there was a shift towards smaller F1 values at T3, suggesting that although the 3 Hz-drift towards higher F1 values at T2 was not significant, bilinguals significantly decreased vowel openess in native Spanish production when they were immersed again in their native environment. In Basque, apart from numerically higher F1 values, vowel production at T3, four-months after the SA program, showed increased variability relative to the production at T1, before the SA program. These results indicate that while some Spanish dominant participants 'recovered' from their intensive period of English use, other participants continued the assimilation process towards the higher F1 values of the English system. This suggests considerable variability in the process of speech 'recovery' for bilinguals' less used language.

Finally, the frequency of switches from one language to another modulated the size of the phonetic drift differently in balanced and Spanish dominant bilinguals. While in balanced bilinguals, the frequency of switches had similar short-term and long-term effects for both Spanish and Basque, it had diverging effects in Spanish dominant bilinguals. In particular, in both balanced and Spanish dominant bilinguals, switches played a 'protective' role during the period of English immersion: more switches were associated with less pronounced assimilatory drift in both native languages.[12] These results are in line with previous results for late bilinguals, showing that bilinguals' second language has stronger effects on the production of native sounds in speakers who frequently switch from L1 to L2 (De Leeuw, Schmid, & Mennen, 2010). Similarly, experimental (naming) studies have shown that bilinguals' production in both L1 and L2 is less accurate on switching trials, when bilinguals have to switch from one language to another (Antoniou et al., 2011; Goldrick et al., 2014), suggesting that co-activated languages pervasively influence each other's phonetic production. Together with the results of previous research, our findings suggest that switches allow the phonetic systems of the contacting languages to mutually update their acoustic targets and to adjust sound production if any changes have been induced by intensive foreign language use. This interpretation is supported, in particular, by our results in Spanish dominant bilinguals: those who practiced frequent switches from Basque to Spanish showed no or very small drift in Basque production after the SA program.

As far as long-term changes are concerned, our results revealed dissociations in the effects of switching frequency for balanced and Spanish dominant bilinguals. In balanced bilinguals, who showed L3-related changes in both native languages, the return change 'back towards native norms' was affected by the frequency of switches, with more switches leading to less noticeable change back to native norms in both languages. These results suggest that frequent contact between the two phonetic systems that were affected by intensive L3 use still prevented the 'recovery' of both four months after the SA program, as if they were mutually reinforcing each other's 'deviated' targets. In Spanish dominant bilinguals, similar to balanced bilinguals, more switches (from Basque to Spanish) resulted in less pronounced (or absent) change back to native norms in Basque, suggesting that frequent switches prevented Basque from 'recovering'. Yet, notice that Spanish dominant bilinguals who frequently switched to Spanish showed small assimilatory drift to start with, which might have contributed to less pronounced return drift at T3. We hypothesize that the absence of an overall return drift in Spanish dominant bilinguals should not be attributed to the same factor as the absence of drift in balanced bilinguals: in Spanish dominant bilinguals, the Basque phonetic targets did not have the 'opportunity' to recover (from the L3-English influence), given that Basque was used at least 30% less than Spanish and, even when speaking Basque, Spanish dominant bilinguals switched very frequently to Spanish, thereby preventing any changes in the production of the Basque vowels.

In sum, the results of our study suggest that frequency of language use and switching habits can independently influence phonetic development in multilinguals. In balanced bilinguals (who use both languages at similar frequencies,), switching protected native phonetics from the influence of the ambient language, whether it was foreign (English immersion during the SA program) or native (in Spain). It appeared that native phonetic targets were constantly 'updated' by bilingual contact/switching events. Spanish dominant bilinguals (who used Basque less frequently than Spanish) showed L3-induced phonetic drift in Basque vowel production only, with frequent switching to Spanish associated with less pronounced assimilatory and return drifts in Basque. This may suggest that Spanish dominant participants, who reported frequently switching back to Spanish while speaking (already non-dominant) Basque, simply had less 'opportunities' to recover Basque targets, due to overall low frequency Basque use; yet, large inter-speaker variability in return drift suggests that there may be other individually specific factors that modulate the speed of Basque recovery.

4.4. Relationship between improvements in English pronunciation and the amount of drift in native production

The results revealed a relationship between improvements in English pronunciation and the size of the phonetic drift towards the English vowel system. In particular, those participants who showed more improvements in English pronunciation showed less assimilatory drift (towards English) in their native production. These results are in line with the SLM (Flege, 1995) hypothesis and previous research on L2 learners, suggesting that L2 speakers dissimilate native sound categories from similar non-native sounds once they have discerned the phonetic differences between them and establish (or are in the process of establishing) new categories for L2 sounds (Flege & Eefting, 1987a, 1987b; Huffman & Schuhmann, 2016; Huffman et al., 2017). Our study suggests that bilinguals with greater improvements in L3 production at T2 might have started to establish new categories for L3 sounds, which launched a dissimilatory process of distinguishing native categories from similar L3 sounds. The amount of 'deterioration' in English pronunciation four months after the SA program was not related to the degree of drift in native production towards native norms. These results suggest that, in a native language environment, factors related to the frequency and circumstances of native language use, not foreign language use, modulate the authenticity of bilingual native speech production.

5. Conclusions

This study examined the effectiveness of the SA English program in the Netherlands for English vowel pronunciation in experienced Basque-Spanish bilingual learners of English and assessed whether intensive English use during SA had effects on native vowel production. Our results have shown that a SA program with no language instruction in a non-anglophone country appears to be an effective means for improving English vowel pronunciation, suggesting that classes featuring contextualized, communicative activities embedded in meaningful

[12] Recall that, due to the fact that each participant reported different switching habits for each of his languages, we examined switching habits separately for each language, and not conjointly as stated in the original third hypothesis. Thus, we were not able to answer the question as to whether bilinguals who regularly switch between their native languages show L3 influences on both of these languages. We did, however, examine the effects of switching on each language separately, and considered whether results were modulated by the frequency of that language use.

contexts that offer ample opportunities for practice enable learners to improve their foreign language production (similar to studies on foreign-language pronunciation instruction, see, e.g., Darcy, 2019; Derwing et al., 1998; Gatbonton & Segalowitz, 2005; Mora, 2008; Saito, 2012; Saito & Lyster, 2012); yet, the benefits of the SA program did not persist four months after the program, when participants returned to their home country, suggesting that regular foreign language use and input from native speakers might be required to maintain production gains in non-native pronunciation. More research is needed to understand what specific factors (e.g., classroom activities, foreign-language speech exposure or foreign-language practice in communicative settings) contribute to better preservation of the pronunciation improvements gained during the SA program. Similar to previous research (Chang, 2012, 2013; Mora, 2008; Sancier & Fowler, 1997), our results have shown that intensive foreign language use affects native vowel production: the day after the end of the SA program, bilinguals showed an assimilatory drift toward the English vowel system in both native languages. However, four months later, bilinguals showed a 'recovery' drift toward native norms, suggesting, in line with the dynamical systems hypothesis (Tobin et al., 2017), that intensive (native) language use and exposure induced adaptation to the phonetic norms of ambient speech.

Analyses of bilinguals' linguistic profiles revealed that the size of the phonetic drift in native language production varied between languages as a function of the frequency of language use and switching habits. In balanced bilinguals, who used Spanish and Basque at similar frequencies, the short- and long-term effects of the SA program were similar for both languages, whereas in Spanish dominant bilinguals, the effects of intensive English use were observed only for the Basque language, suggesting that cursory (irregular) native language use made this language more vulnerable to foreign language influence. The frequency of switches between the two native languages, on the other hand, played a 'protective' role during the immersion period: more switches were associated with less pronounced drift in both native languages, suggesting that switches allow bilinguals to maintain the phonetic targets of all native targets through constant 'updating' and mutual reinforcement across languages. Finally, our results revealed a negative relationship between the amount of improvement in English pronunciation and the size of the phonetic drift towards the English vowel system in native production, in line with the SLM (Flege, 1995) hypothesis. That is, greater improvement in English correlated with less assimilatory drift in native languages. This suggests that bilinguals who exhibit greater improvement in L3 production have started to establish new (at least phonetic) categories for English sounds and launched the process of dissimilating native categories from similar English sounds. Importantly, our results indicate that improvements in L3 production are not necessarily associated with any deterioration (foreign accentedness or attrition over the long-term) in native language pronunciation. Instead, they suggest that other factors related to the frequency and circumstances of native language use are key to the authenticity of native language production.

The results of this unique and pioneering study convincingly point to a complex interconnected relationship between the phonetic systems of three language systems in bilingual learners of English and demonstrate the sensitivity of phonetic systems to external factors of language use, such as, usage frequency and switching habits. More studies with larger sample sizes and other language combinations need to be conducted in order to further examine short- and long-term interactions between phonetic systems in multilinguals.

Acknowledgments

CM was supported by the Basque Government (BERC 2018-2021 program; PIBA18-29) and by the Spanish State Research Agency (AEI) through BCBL Severo Ochoa excellence accreditation (SEV-2015-0490), the Spanish Ministry of Economy and Competitiveness (MINECO) and the European Fund for Regional Development (FEDER) (PSI2017-82941-P; Europa-Excelencia ERC2018-092833). The authors thank Magda Altman for proofreading the article.

Appendix A

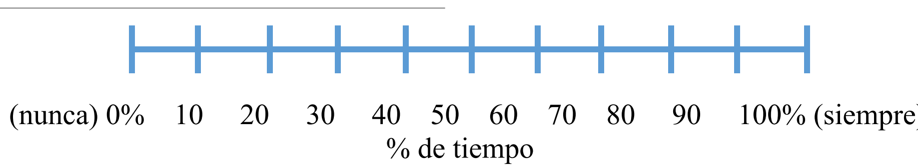
The language-use questionnaire – adapted from Rodriguez-Fornells and colleagues' study (Rodriguez-Fornells et al., 2012). Two dependent measures, used in the analyses were the answers to questions 3 and 6.

Ocho preguntas sobre el uso de Vasco y Español

Trate de contestar en qué medida las siguientes preguntas representan o se ajustan a su forma de hablar y expresarte en los idiomas que conoce (p. ej., Vasco–Español), en términos generales. Muchas de estas preguntas hacen referencia a si usted cambia o mezcla frecuentemente el vasco y el castellano en sus conversaciones. Cambiar o mezclar lenguajes es una característica muy particular de algunos entornos bilingües, como es el caso en el País Vasco. El siguiente cuestionario pretende investigar sobre dichos hábitos de cambio y mezcla de lenguas. Si tiene dudas sobre algunas respuestas, intente comparar su forma de hablar y expresarte con el de la mayoría, o de las personas que conoce bien.

GENERAL

1. Tiendo a mezclar idiomas durante una conversación (por ejemplo, cambio de español a vasco o lo contrario)



Si ha seleccionado 10% o más, responda por favor a todas las siguientes preguntas:

2. Cuando mezclo un idioma lo hago conscientemente

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Durante una conversación en VASCO

3. De manera general, durante una conversación en vasco, cambio al español

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

4. Tiendo a cambiar al español durante la conversación **de una frase a otra**.

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente.
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

5. Tiendo a cambiar al español durante la conversación **de una palabra a otra**.

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente.
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Durante una conversación en ESPAÑOL

6. De manera general, durante una conversación en español, cambio al vasco

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente.
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

7. Tiendo a cambiar al vasco durante la conversación **de una frase a otra**.

☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente.
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

8. Tiendo a cambiar al vasco durante la conversación **de una palabra a otra**.

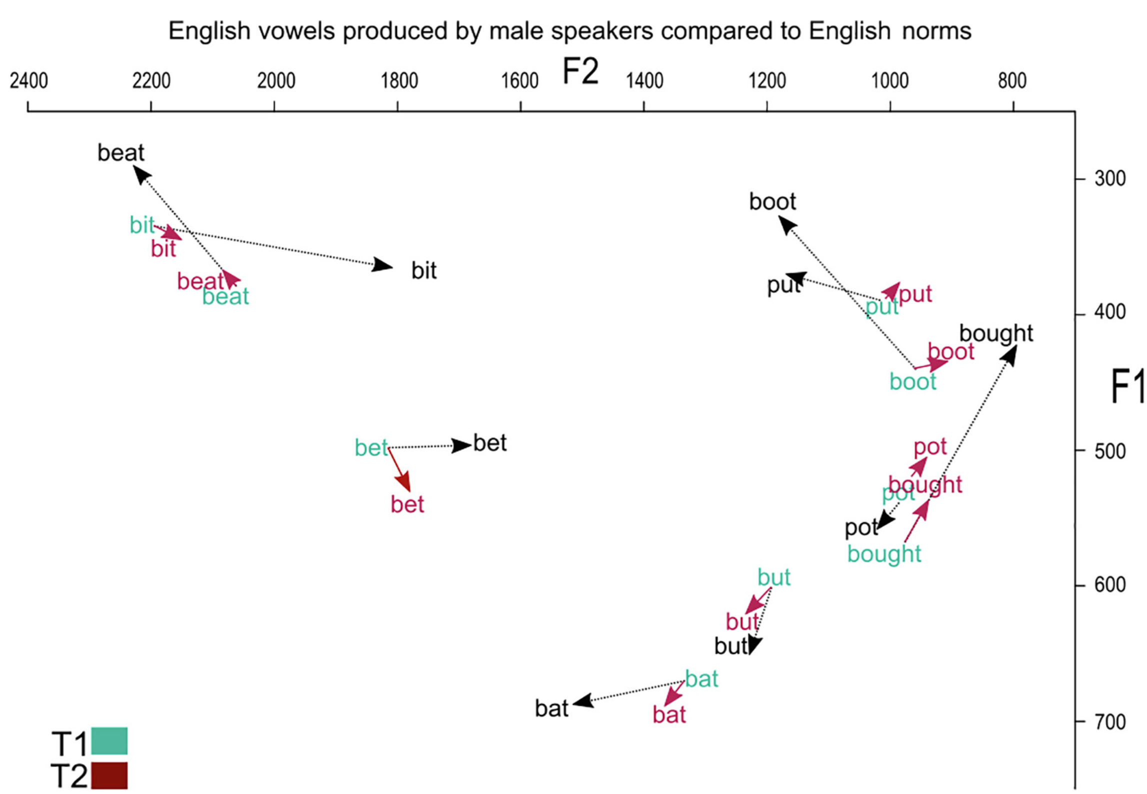
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

Cuando lo hago, lo hago conscientemente.
☐ nunca ☐ muy raramente ☐ ocasionalmente ☐ frecuentemente ☐ siempre

POR FAVOR, COMPRUEBE SI HA RESPONDIDO A TODAS LAS PREGUNTAS

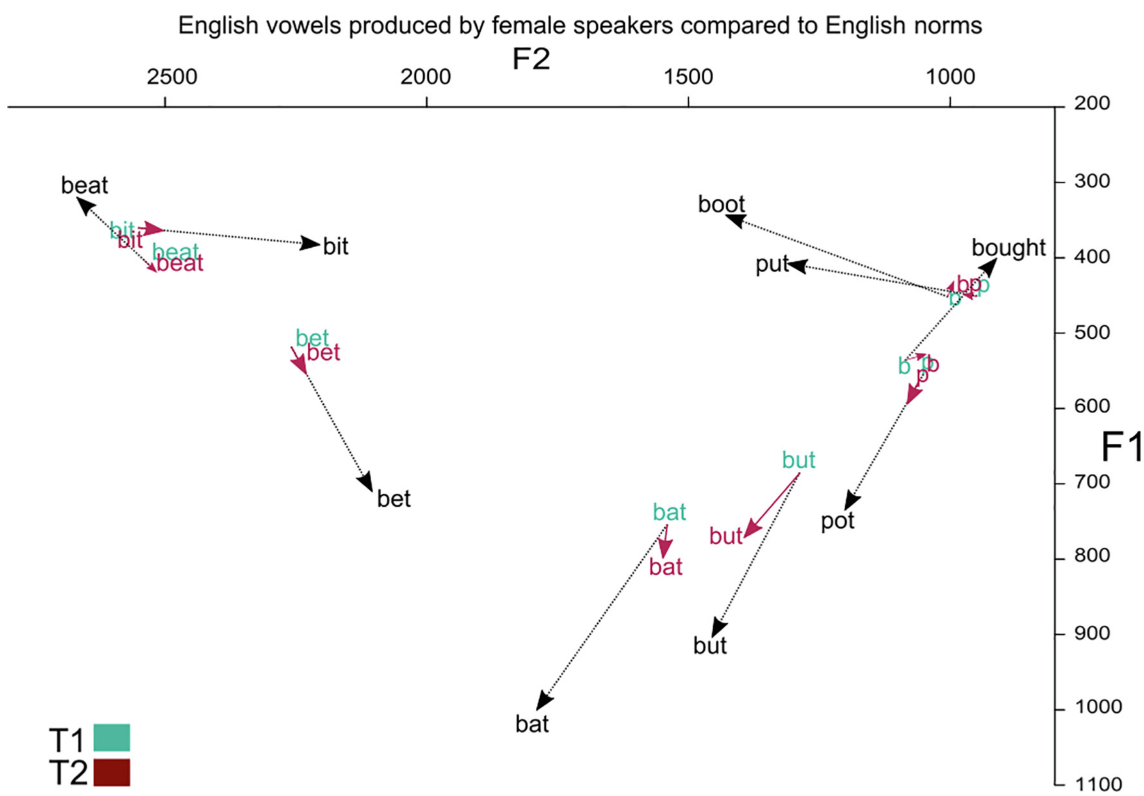
Appendix B

Acoustic position of English vowels produced by male bilingual Spanish-Basque participants at T1 – before the study abroad program, at T2 – the day after the SA program compared to the English norms. Black arrows indicate the direction of improvement needed for participants' production at T1 to reach the target vowel. Red arrows indicate participants' real direction of change from T1 to T2.



Appendix C

Acoustic position of English vowels produced by female bilingual Spanish-Basque participants at T1 – before the study abroad program, and at T2 – the day after the SA program compared to the English norms. Black arrows indicate the direction of improvement needed for participants' production at T1 to reach the target vowel. Red arrows indicate participants' real direction of change from T1 to T2.



References

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics, 39*(4), 558–570. https://doi.org/10.1016/j.wocn.2011.03.001.

Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., & Yamada, T. (2004). Perceived phonetic dissimilarity and L2 speech learning: The case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics, 32*(2), 233–250. https://doi.org/10.1016/S0095-4470(03)00036-6.

Barlow, J. A. (2014). Age of acquisition and allophony in Spanish-English bilinguals. *Frontiers in Psychology, 5*. https://doi.org/10.3389/fpsyg.2014.00288.

Barlow, J. A., Branson, P. E., & Nip, I. S. B. (2013). Phonetic equivalence in the acquisition of /l/ by Spanish-English bilingual children. *Bilingualism: Language and Cognition, 16*(01), 68–85. https://doi.org/10.1017/S1366728912000235.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: Theoretical and methodological issues* (pp. 171–204). Baltimore: York Press.

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.2. Retrieved from http://www.praat.org.

Bohn, O.-S., & Flege, J. E. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition, 14*(02), 131–158.

Bradlow, A. R. (1995). A comparative acoustic study of English and Spanish vowels. *Journal of Acoustic Society of America, 97*(3), 1916–1924.

Casillas, J. V., & Simonet, M. (2016). Production and perception of the English /æ/–/ɑ/ contrast in switched-dominance speakers. *Second Language Research, 32*(2), 171–195. https://doi.org/10.1177/0267658315608912.

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics, 40*(2), 249–268. https://doi.org/10.1016/j.wocn.2011.10.007.

Chang, C. B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics, 41*(6), 520–533. https://doi.org/10.1016/j.wocn.2013.09.006.

Chang, C. B. (2019). Language change and linguistic inquiry in a world of multicompetence: Sustained phonetic drift and its implications for behavioral linguistic research. *Journal of Phonetics, 74*, 96–113. https://doi.org/10.1016/j.wocn.2019.03.001.

Chládková, K., Escudero, P., & Boersma, P. (2011). Context-specific acoustic differences between Peruvian and Iberian Spanish vowels. *The Journal of the Acoustical Society of America, 130*(1), 416–428. https://doi.org/10.1121/1.3592242.

Collentine, J. (2004). The effects of learning contexts on morphosyntactic and lexical development. *Studies in Second Language Acquisition, 26*(2), 227–248. https://doi.org/10.1017/S0272263104262040.

Darcy, I. (2019). Improving pronunciation in spontaneous speech? A comparison of two instructional methods. Presented at the L2 Pronunciation Research Workshop: Bridging the Gap between Research and Practice, Barcelona.

De Leeuw, E., Schmid, M., & Mennen, I. (2007). Global foreign accent in native German speech. In *16th international congress of phonetic sciences* (pp. 1605–1608). Retrieved from http://eresearch.qmu.ac.uk/50/.

De Leeuw, E., Schmid, M. S., & Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition, 13*(01), 33. https://doi.org/10.1017/S1366728909990289.

Derwing, T. M., Munro, M. J., & Wiebe, G. (1998). Evidence in favor of a broad framework for pronunciation instruction. *Language Learning, 48*(3), 393–410. https://doi.org/10.1111/0023-8333.00047.

Deterding, D. (1997). The formants of monophthong vowels in standard Southern British English pronunciation. *Journal of the International Phonetic Association, 27*(1–2), 47–55. https://doi.org/10.1017/S0025100300005417.

Díaz-Campos, M. (2004). Context of learning in the acquisition of Spanish second language phonology. *Studies in Second Language Acquisition, 26*(02). https://doi.org/10.1017/S0272263104262052.

Egurtzegi, A. (2013). Phonetics and phonology. Basque and Proto-Basque. Language-Internal and Typological Approaches to Linguistic Reconstruction. Mikroglottika, 5.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*. Utrecht: LOT.

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

Flege, J. E. (2009). Give input a chance! In T. Piske & M. Young-Scholten (Eds.), *Input matters in SLA* (pp. 175–190). Retrieved from http://jimflege.com/files/Flege_inputmatters_2009.pdf.

Flege, J. E., Bohn, O.-S., & Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics, 25*(4), 437–470. https://doi.org/10.1006/jpho.1997.0052.

Flege, J. E., & Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*, 185–202.

Flege, J. E., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics, 15*, 67–83.

Flege, J. E., MacKay, I. R., & Meador, D. (1999). Native Italian speakers' perception and production of English vowels. *The Journal of the Acoustical Society of America, 106* (5), 2973–2987. https://doi.org/10.1121/1.428116.

Flege, J. E., Schirru, C., & MacKay, I. R. A. (2003). Interaction between the native and second language phonetic subsystems. *Speech Commun., 40*(4), 467–491. https://doi.org/10.1016/S0167-6393(02)00128-0.

Flege, J. E., & Wayland, R. P. (2019). The role of input in native Spanish Late learners' production and perception of English phonetic segments. *Journal of Second Language Studies, 2*(1). Retrieved from: http://jimflege.com/files/final_compact2.pdf.

Forster, K., & Forster, J. (2003). DMDX: a windows display program with millisecond accuracy. *Behavior Research Methods Instruments and Computers, 35*(1), 116–124. https://doi.org/10.3758/BF03195503.

Fowler, C. A., Sramko, V., Ostry, D. J., Rowland, S. A., & Hallé, P. (2008). Cross language phonetic influences on the speech of French-English bilinguals. *Journal of Phonetics, 36*(4), 649–663. https://doi.org/10.1016/j.wocn.2008.04.001.

Freed, B., So, S., & Lazar, N. A. (2003). Language learning abroad: How do gains in written fluency compare with gains in oral fluency in French as a second language? *ADFL Bulletin, 34*(3), 34–40.

Gatbonton, E., & Segalowitz, N. (2005). Rethinking communicative language teaching: a focus on access to fluency. *The Canadian Modern Language Review / La Revue Canadienne Des Langues Vivantes, 61*(3), 325–353. https://doi.org/10.1353/cml.2005.0016.

Genesee, F. (1985). Second language learning through immersion: a review of US programs. *Review of Educational Research, 55*(4), 541–561.

George, A. (2014). Study abroad in Central Spain: The development of regional phonological features. *Foreign Language Annals, 47*(1), 97–114. https://doi.org/10.1111/flan.12065.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less Nativelike. *Psychological Science, 25*(4), 1031–1036. https://doi.org/10.1177/0956797613520014.

Grosjean, F. (2001). The Bilingual's Language Modes. In Explaining linguistics. One mind, two languages: Bilingual language processing (pp. 1–22). Malden: Blackwell Publishing.

Guion, S. G. (2003). The vowel systems of Quichua-Spanish bilinguals. Age of acquisition effects on the mutual influence of the first and second languages. *Phonetica, 60*(2), 98–128.

Harada, T. (2003). L2 Influence on L1 Speech in the Production of VOT. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2003/papers/p15_1085.pdf.

Howard, M. (2004). On the role of context in the development of learner language. *ITL – International Journal of Applied Linguistics, 147*(1), 1–20. https://doi.org/10.2143/ITL.148.0.2002062.

Huffman, M. K., & Schuhmann, K. (2016). Effect of early L2 learning on L1 stop voicing. *Proceedings of Meetings on Acoustics Acoustical Society of America, 23*, 1–10. https://doi.org/10.1121/2.0000168.

Huffman, M. K., Schuhmann, K., Keller, K., & Chen, C. (2017). Interaction of drift and distinctiveness in L1 English-L2 Japanese learners 3517–3517. *The Journal of the Acoustical Society of America, 141*(5). https://doi.org/10.1121/1.4987389.

Ingram, J. C., & Park, S.-G. (1997). Cross-language vowel perception and production by Japanese and Korean learners of English. *Journal of Phonetics, 25*(3), 343–370. https://doi.org/10.1006/jpho.1997.0048.

Kartushina, N., Frauenfelder, U. H., & Golestani, N. (2016). How and when does the second language influence the production of native speech sounds: A literature review. *Language Learning, 66*(S2), 155–186. https://doi.org/10.1111/lang.12187.

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2016). Mutual influences between native and non-native vowels in production: Evidence from short-term visual articulatory feedback training. *Journal of Phonetics, 57*, 21–39. https://doi.org/10.1016/j.wocn.2016.05.001.

Kuznetsova, A., Brockhoff, P. B., & Christensen, R. H. B. (2017). lmerTest package: Tests in linear mixed effects models. *Journal of Statistical Software, 82*(13), 1–26. https://doi.org/10.18637/jss.v082.i13.

Lafford, B. A., & Collentine, J. (2006). *The effects of study abroad and classroom contexts on the acquisition of Spanish as a second language* (pp. 103–126). The Art of Teaching Spanish: Second Language Acquisition from Research to Praxis.

Lafford, B. A. (2006). The effects of study abroad vs. classroom contexts on Spanish SLA: Old assumptions, new insights and future research directions. Selected proceedings of the 7th conference on the acquisition of Spanish and Portuguese as first and second languages (pp. 1–25). Cascadilla Proceedings Project Somerville, MA.

Lang, B., & Davidson, L. (2017). Effects of exposure and vowel space distribution on phonetic drift: Evidence from American English Learners of French. *Language and Speech*. 0023830917737111.

Lenhard, W., & Lenhard, A. (2016). Computation of effect sizes. https://doi.org/10.13140/RG.2.1.3478.4245.

Lenth, R. V. (2016). Least-squares means: The R package lsmeans. *Journal of Statistical Software, 69*(1), 1–33. https://doi.org/10.18637/jss.v069.i01.

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics, 41*(5), 320–331. https://doi.org/10.1016/j.wocn.2013.06.002.

Levy, E. S., & Law, F. F. (2010). Production of French vowels by American-English learners of French: Language experience, consonantal context, and the perception-production relationship. *The Journal of the Acoustical Society of America, 128*(3), 1290. https://doi.org/10.1121/1.3466879.

Linck, J. A., Kroll, J. F., & Sunderman, G. (2009). Losing access to the native language while immersed in a second language evidence for the role of inhibition in second-language learning. *Psychological Science, 20*(12), 1507–1515. https://doi.org/10.1111/j.1467-9280.2009.02480.x.

Linck, J. A., Schwieter, J. W., & Sunderman, G. (2012). Inhibitory control predicts language switching performance in trilingual speech production. *Bilingualism: Language and Cognition, 15*(3), 651–662. https://doi.org/10.1017/S136672891100054X.

Llanes, À., Arnó, E., & Mancho-Barés, G. (2016). Erasmus students using English as a lingua franca: Does study abroad in a non-English-speaking country improve L2 English? *The Language Learning Journal, 44*(3), 292–303. https://doi.org/10.1080/09571736.2016.1198099.

Lord, G. (2006). Defining the Indefinable: Study abroad and phonological memory abilities. In C. Klee & T. L. Face (Eds.), *Selected proceedings of the 7th conference on the acquisition of Spanish and portuguese as first and second languages* (pp. 40–46). Somerville, Mass: Cascadilla Proceedings Project.

Lord, G. (2010). The combined effects of immersion and instruction on second language pronunciation. *Foreign Language Annals, 43*(3), 488–503. https://doi.org/10.1111/j.1944-9720.2010.01094.x.

MacLeod, A. A. N., Stoel-Gammon, C., & Wassink, A. B. (2009). Production of high vowels in Canadian English and Canadian French: A comparison of early bilingual and monolingual speakers. *Journal of Phonetics, 37*(4), 374–387. https://doi.org/10.1016/j.wocn.2009.07.001.

Mair, P., & Wilcox, R. R. (2017). Robust Statistical Methods in R Using the WRS 2 Package.

Major, R. C. (1992). Losing English as a first language. *The Modern Language Journal, 76*(2), 190. https://doi.org/10.2307/329772.

Mayr, R., Price, S., & Mennen, I. (2012). First language attrition in the speech of Dutch-English bilinguals: The case of monozygotic twin sisters. *Bilingualism: Language and Cognition, 15*(04), 687–700. https://doi.org/10.1017/S136672891100071X.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. *A Portrait of the Young in the New Multilingual Spain*, 241–263.

Mora, J. C., Keidel, J. L., & Flege, J. E. (2015). Effects of Spanish use on the production of Catalan vowels by early Spanish- Catalan bilinguals. In J. Romero & M. Riera (Eds.), *The Phonetics-phonology interface: Representations and methodologies* (pp. 33–54). Retrieved from DOI: 10.1075/cilt.335.02mor.

Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism, 16*(4), 484–500. https://doi.org/10.1177/1367006911429518.

Morrison, G. S. (2008). L1-Spanish speakers' acquisition of the English /i /–/I/ contrast: duration-based perception is not the initial developmental stage. *Language and Speech, 51*(4), 285–315. https://doi.org/10.1177/0023830908099067.

Oh, G. E., Guion-Anderson, S., Aoyama, K., Flege, J. E., Akahane-Yamada, R., & Yamada, T. (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), 156–157. https://doi.org/10.1016/j.wocn.2011.01.002.

R Core Team. (2012). R: A language and environment for statistical computing. Retrieved from http://www.R-project.org.html.

Rodriguez-Fornells, A., Kramer, U., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*. https://doi.org/10.3389/fpsyg.2011.00388.

Saito, K. (2012). Effects of instruction on L2 pronunciation development: A synthesis of 15 quasi-experimental intervention studies. *Tesol Quarterly, 46*(4), 842–854.

Saito, K., & Lyster, R. (2012). Effects of Form-Focused instruction and corrective feedback on L2 pronunciation development of/ɹ/by Japanese learners of English. *Language Learning, 62*(2), 595–633.

Sancier, M. L., & Fowler, C. A. (1997). Gestural drift in a bilingual speaker of Brazilian Portuguese and English. *Journal of Phonetics, 25*(4), 421–436.

Schoormann, H., Heeringa, W., & Peters, J. (2017). Standard German vowel productions by monolingual and trilingual speakers. *International Journal of Bilingualism*. 1367006917711593.

Schuhmann, K., & Huffman, M. K. (2015). L1 drift and L2 category formation in second language learning. ICPhS, Glasgow, Scotland, UK. Presented at the 18th International congress of phonetic sciences. Retrieved from https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS2015/Papers/ICPHS0850.pdf.

Scovel, T. (1969). Foreign accents, language acquisition, and cerebral dominance. *Language Learning, 19*(3–4), 245–253.

Segalowitz, N., & Freed, B. F. (2004). Context, contact, and cognition in oral fluency acquisition: Learning Spanish in at home and study abroad contexts. *Studies in Second Language Acquisition, 26*(2), 173–199.

Simões, A. R. M. (1996). Phonetics in second language acquisition: An acoustic study of fluency in adult learners of Spanish. *Hispania, 79*(1), 87–95. https://doi.org/10.2307/345617.

Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism, 9*(01), 97–114. https://doi.org/10.1017/S1366728905002403.

Tobin, S. J., Nam, H., & Fowler, C. A. (2017). Phonetic drift in Spanish-English bilinguals: Experiment and a self-organizing model. *Journal of Phonetics, 65*, 45–59. https://doi.org/10.1016/j.wocn.2017.05.006.