The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish

Celia Gorba*, Juli Cebrian

*Corresponding author: celia.gorba@gmail.com

Universitat Autònoma de Barcelona, 08193 Bellaterra (Cerdanyola del Vallès) Barcelona, Spain

Abstract:

Some previous studies report that increased experience with a second language (L2) may result in a more target-like perception and production in the L2, as well as in a less native-like performance in the L1. The present paper aimed to 1) assess the role of L2 experience on L2 and L1 production of voiceless stops; 2) investigate the effect of L2 experience on L2 and L1 perception of voiceless stops; and 3) examine the relationship between perception and production. Three groups of English learners of Spanish differing in amount and type of L2 experience, as well as two groups of functional monolinguals, completed a production task and an identification task involving English and Spanish voiceless stops. The results revealed that the L2 speakers were more successful at producing than at perceiving Spanish stops accurately, with L2 experience having a positive effect on production. L2 experience was not found to affect performance in the L1, which could be related to an overall limited amount of L2 use even in an immersion setting. The results also showed a weak relationship between perception and production, which may partly be due to the different nature of perceptual and production measures.

Keywords

Crosslinguistic influence, speech perception, speech production, L2 speech, voice onset time, L2 experience

1. Introduction

The fact that a speaker's native or first language (L1) exerts an influence on a subsequently learned or second language (L2) is widely attested (e.g., Flege, 2007; Flege, MacKay, & Meador, 1999). The reverse type of influence, that is, the influence of the L2 on the L1, is also reported (e.g., Flege, 1987; Kartushina, Frauenfelder, & Golestani, 2016), although it has received less attention, particularly regarding the perception of L2 sounds. In addition, the degree of cross-linguistic influence may be modulated by factors such as the amount of L1 and L2 experience and use, and the degree of similarity between the L1 and L2 sounds (Best, & Tyler, 2007; Flege, & Bohn, 2021). This study aims to explore these issues further by contrasting how L1 English L2 Spanish speakers differing in the type and amount of L2 experience produce and also perceive voiceless stops both in the L1 and the L2, and compared to monolingual Spanish and monolingual English speakers. The following sections review the main issues and findings regarding the direction of crosslinguistic influence (CLI) and the effect of experience on CLI, the main claims of the relevant L2 speech models and the relationship between perception and production. Given that the focus of this paper is the perception and production of stops, the review of the literature will center mostly on these consonants. The characteristics of stops in Spanish and English are presented first.

1.1. Crosslinguistic differences in voice onset time

Voice onset time (henceforth VOT), i.e., the interval between the release of the stop and the onset of voicing, has been established as the main cue to the voiced-voiceless distinction in initial stops in a number of languages, including English and Spanish (Lisker, & Abramson, 1964). This study focuses on English and Spanish bilabial and velar voiceless stops, which share manner and place of articulation in both languages but differ in the use of VOT to contrast voicing (Lisker, & Abramson, 1964). Coronal stops are alveolar in English and dental in Spanish and will not be analyzed in this study. Table 1 summarizes the mean VOT values for English and Spanish bilabial and velar stops reported in the literature.

	/p/	**/b/**	**/k/**	**/g/**
American English (Lisker & Abramson, 1964)	58 (r: 20 - 120)	1/-101 (r: 0–5 / -130 – -20)	80 (r: 50 –135)	21/-88 (r: 0–35/ -150 – -60)
American English (Dmitrieva et al., 2015)	64.2 (SD: 18.2)	12.1/-107.3 (SD: 5/32)		
British English (Docherty, 1992)	42 (SD: 10.4)	15 (SD: 8.3)	62 (SD: 17.2)	27 (SD: 11.1)
Puerto Rican Spanish (Lisker & Abramson 1964)	4 (r: 0 - 15)	-138 (r: -235 – -60)	29 (r: 15– 55)	-108 (r: -165 – -45)
Castilian Spanish (Castañeda, 1986)	6.5 (SD: 6.4; r: 0 – 24)	-69.8 (SD: 24.7; r: -166 – - 24)	25.7 (SD: 10.5; r: 8.8 – 52)	-58 (SD: 26; r: -132 – -16)
Different dialects of Spanish (Dmitrieva et al., 2015)	14 (SD: 4.7)	-94.7 (SD: 31.5)		

Table 1. VOT means in ms for initial stops in American English, British English, Puerto Rican Spanish and Castilian Spanish reported in some previous studies. Where reported, SD = standard deviation, r = range.

English voiceless stops are produced with long-lag VOT. For instance, Lisker and Abramson (1964) reported that American English speakers produced /p/ and /k/ with a mean VOT duration of 58 and 80 ms, respectively. Similarly, the mean VOT values described for British

English speakers are 42 ms for /p/ and 62 ms for /k/ (Docherty, 1992). English word-initial voiced stops tend to be produced with short-lag VOT: Lisker and Abramson (1964) report mean VOT values of 1 ms for /b/ and 21 ms for /g/, while Docherty (1992) reports 15 ms and 27 ms, respectively. Still, voice-lead productions are also found for native English voiced stops (Dmitrieva, Llanos, Shultz, & Francis, 2015; Lisker, & Abramson, 1964). Regarding Spanish, voiceless stops are produced with short-lag VOT and voiced stops present voice-lead VOT. According to Lisker and Abramson (1964), Puerto Rican Spanish speakers produce /p/ with a mean VOT value of 4 ms, /b/ with -138 ms, /k/ with 29 ms and /g/ with -108 ms. Similarly, Castañeda (1986) reported mean VOT values of 6.5 ms, -69.8 ms, 25.7 ms and -58 ms for /p, b, k, g/, respectivley, in European Spanish. Importantly, short-lag VOT values signal a voiceless stop in Spanish and a voiced category in English.

1.2. Crosslinguistic influence and L2 experience

In the speech of bilingual and second language (L2) speakers, cross-linguistic influence, that is, the influence of one language on another, is a common phenomenon. Typically, the first language (L1) exerts an influence on the perception and production of the second language (L2), especially in the case of adult L2 learners (e.g., Antoniou, Best, Tyler, & Kross, 2010; Caramazza, Yeni-Komshian, Zurif, & Carbone 1973; Carlson, Goldrick, Blasingame, & Fink, 2016; Flege, 1991; Flege, Mackay, & Piske, 2002). For example, Flege (1991) investigated the production of English and Spanish /t/ by adult Spanish learners of English who had lived in the United States for 14 years. He found that the L2 speakers produced English /t/ with VOT values that were intermediate between the expected L1 and L2 values. Similarly, a group of Canadian French L1 English L2 bilinguals who had started learning English before the age of 7 produced English stops with significantly longer VOT than French stops, but significantly shorter VOT than English monolingual speakers (Caramazza et al., 1973).

Still, L2 speakers may achieve or approximate accuracy in their production and perception of the L2. Previous research has shown that increased experience, particularly a longer length of residence in an L2 setting, may result in a more target-like L2 (Bohn, & Flege, 1990; Flege, 1987; Flege, Bohn, & Jang, 1997; Jun, & Cowie, 1994; Levy, & Law, 2010; Stevens, 2001, among others). For instance, Stevens (2001) found that L1-English learners of Spanish in Spain improved their production of L2 voiceless stops (i.e., produced shorter VOT) to a greater extent than L2 Spanish learners in the US. Flege (1987) examined L1 and L2 production by different groups of L2 speakers: two groups of American English learners of French residing in the L1 setting but differing in years of learning and amount of time previously spent in an L2 setting, one group of American English speakers residing in France and one group of French speakers living in the US. Regarding the two L2 French speakers in an L1 setting, the less experienced group produced /t/ in English and French similarly, while the more experienced group distinguished English /t/ and French /t/, although the L2 /t/ was not target-like. By contrast, the L1-English L2-French speakers residing in France did not differ significantly from monolingual French speakers in their production of French /t/. Hence, Flege found evidence of a positive effect of L2 experience on L2 learning.

L2 experience, however, is a broad term that is used to refer to a variety of interrelated factors found to affect L2 (and L1) performance in addition to years of learning and learning setting. These include L2 proficiency, amount of L1 and L2 use, language immersion, instruction, motivation and aptitude, among others (Kartushina, Frauenfelder, et al., 2016; Piske, MacKay, & Flege, 2001; Purcell, & Suter, 1980). Piske et al. (2001) found that language use was the second most important factor determining L2 performance, after age at which L2 learning started. Amount of L1 and L2 use is in fact a factor that distinguishes the groups of L2 learners compared in Flege (1987). Recall that the L1-English L2-French speakers in Flege (1987) exhibited a positive effect of experience whereas the L1-Spanish L2-English speakers in Flege (1991) did not. This inconsistency may be explained by the fact that participants in Flege (1987) had native L2 speaking spouses and, consequently, made greater use of the L2, whereas half of the adult learners in Flege (1991) lived near the Mexican border, where there is a strong Spanish-English bilingual community. Similarly, Purcell and Suter (1980) found that a compound variable which combined years of residence in the L2 context and L2 use at home – measured by years living with native speakers of the target language – was one of the main predictors for degree of L2 accentedness. In fact, the revised version of the Speech Learning Model (SLM-r, see section 1.3 for the models' descriptions), proposes a variable that results from multiplying years of residence in the L2 setting by the proportion of L2 use (referred to as full time equivalent or FTE) as a more appropriate approach to quantifying L2 input. For instance, Aoyama and Flege (2011) investigated the acquisition of L2 liquids by Japanese learners of English and reported a correlation between years of FTE and their ability to discern between English /r/ and /l/. Finally, intensive L2 instruction may also enhance the effect of experience. For example, Casillas (2020) found that L1 English learners of Spanish substantially improved their production of L2 stops within a 7-week immersion program with intensive L2 instruction and L2 input (and suppressing L1 use) in an instructional setting. The present study investigates the perception and production of L1 and L2 stops by adult L2 learners differing in L2 experience, understood as the amount of L2 experience (years of L2 learning) and type of L2 experience (in the L1 or in the L2 setting).

Previous findings illustrate that CLI also affects L2 speakers' L1, showing that the L1 and the L2 are dynamic systems that interact in the course of the L2 learning process (Chang, 2012; Flege, 1995; see Kartushina, Frauenfelder et al., 2016, for a review). Kartushina, Frauenfelder et al. (2016) explain that absence of CLI may be found with simultaneous bilinguals and very early bilinguals (e.g., who started learning the L2 by the time they were two years old). These speakers may have distinct categories for each language and these categories may resemble those of native monolingual speakers of each language. In addition, later L2 learners with limited exposure to the target language may not be able to create separate L2 categories and use unmodified L1 categories for both L1 and similar L2 sounds (Flege, 1995). On the other hand, Kartushina, Frauenfelder et al. (2016) suggest that backward transfer (effect of L2 learning on L1 categories) occurs under two possible main circumstances. The first one is the case of L2 speakers at an advanced stage of L2 learning and who make predominant use of the L2 usually in an L2 immersion setting (e.g., Bergmann, Nota, Sprenger, & Schmid, 2017; Flege, 1987; Major, 1992; Harada, 2003). The second case involves learners at an early stage of L2 learning, characterized by an immersion setting, as discussed below (Chang, 2012; Kartushina, Hervais-Adelman, Frauenfelder, & Golestani, 2016; Sancier, & Fowler, 1997).

Regarding the former, Harada (2003) reported that heritage Japanese speakers in the United States were found to produce L1 voiceless stops with significantly longer VOT values than Japanese monolinguals, whereas their L2 production did not differ from that of English monolinguals. Major (1992) examined the production of /p, t, k/ by L1 English speakers who migrated to Brazil as adults and had spent between 12 and 35 years in the L2 country. Their productions of VOT in Portuguese, and in English, were found to deviate from those of monolingual speakers of these languages. Similarly, Flege (1987) found evidence that the two most experienced groups, that is, the Americans in France and the French speakers in the US, with around 12 years of immersion in the L2 setting, also differed from monolingual speakers in their L1 production. By contrast, Riney and Okamura (1999) found no effect of L2 experience on the production of L1 stops with highly advanced learners: L1-English speakers of Japanese in an immersion setting – as well as L1-Japanese L2-English speakers who had lived in an immersion setting but were back in Japan – produced L1 stops with native-like VOT values. It should be noted, however, that only five speakers were tested for each L1 and that all of them were linguists or language teachers. Thus, a reduced sample size and the participants' presumable metalinguistic knowledge may explain the absence of phonetic drift of the L1 towards the L2 in this case.

The second type of case of reported backward CLI involves relatively short periods of residence. For example, some studies investigating the effect of short (though repeated) stays in an L2 setting have found that the L1 VOT production may also drift towards more L2-like values after several months (4.5) in an L2 setting (Sancier, & Fowler, 1997). In addition, the learners' L2 production may become less target-like only a few weeks after returning from a short stay in the L2 setting (Sancier, & Fowler, 1997; Tobin, Nam, & Fowler 2017). L2 influence on the L1 has even been reported for novice L2 learners (Chang, 2012; Kartushina, Hervais-Adelman, et al., 2016). Chang (2012) investigated the production of L1 and L2 vowels and stops by American English speakers who were enrolled in a Korean language course in Korea. Phonetic drift of the L1 towards L2 values occurred as early as the second week of L2 classes, underscoring the dynamic nature of the L1 and the L2. Moreover, the influence of the L2 on the L1 was greater for beginner learners than for learners who already had some knowledge of Korean (Chang, 2013), which was attributed to a novelty effect, that is, a greater salience and heightened encoding of L2 stimuli in the context of a novel perceptual experience. Along similar lines, Kartushina, Hervais-Adelman, et al. (2016) report a drift in L1 vowel production toward L2 values only after one hour of vowel training involving articulatory feedback. It is possible that a combination of an immersion setting and L2 instruction or specialized training may have a greater impact on the possibility of modifying L1 and L2 patterns.

The studies reviewed in the previous paragraph show a type of CLI influence that involves phonetic drift from the L1 towards L2 values, also referred to as category assimilation (Flege, 2003, see section 1.3). CLI can also result in category dissimilation, that is, the L1 categories deflect away (i.e., become more dissimilar) from L2 categories in an attempt to maintain a distinction between the L1 and the L2. This also results in L2 speakers' L1 categories differing from those of L1 monolinguals. For instance, Flege and Eefting examined the production of voiceless stops in L1 Dutch and L2 English (1987a) and in L1 Spanish and L2 English (1987b). Unlike in English, voiceless stops in Spanish and in Dutch are unaspirated (i.e., produced with short-lag instead of long-lag VOT). Flege and Eefting (1987b) found that early Spanish learners of English made a distinction between L1 and L2 stops, but produced Spanish voiceless stops with shorter VOT than Spanish monolinguals did. Their English stops also had shorter VOT than English monolinguals. In addition, Flege and Eefting (1987a) reported that only those Dutch L2 English speakers who were judged to sound more nativelike in English showed a deflection in L1 Dutch stops. Thus, deflection of L1 categories may occur as a strategy to distinguish L1 and L2 sounds when L2 categories have been established but are still not target-like.

In brief, L2 speakers may approximate target-language VOT patterns more the greater the amount of experience with the L2, particularly in an L2 setting and with predominant L2 use (Flege, 1987; Flege, & Bohn, 2021). In addition, the L2 speakers’ L1 may also be affected as a consequence of L2 learning. This type of CLI may occur in immersion settings with highly proficient L2 speakers with a predominant L2 use and as the result of novelty effects. Kartushina, Frauenfelder et al. (2016) suggest that, in cases of L1 phonetic drift, early learners and bilinguals may have a tendency to deflect L1 and L2 categories away from each other as a consequence of L2 category creation, while later learners may experience category assimilation. Kartushina Frauenfelder et al. (2016) add that the type of impact may be influenced by the degree of similarity between L1 and L2 sounds, the type and amount of L2 experience and the level of proficiency.

Only a few studies have investigated the effect that L2 experience may have on L1 and L2 perception (Cabrelli et al., 2019; Cebrian, 2006; Dmitrieva, 2019; Major, 2010), particularly regarding stops (Flege and Eefting, 1987a, 1987b; Gorba, 2019), but some studies have compared L2 speakers and monolinguals (e.g., Caramazza et al., 1973; Hazan & Boulakia,). Hazan and Boulakia (1993) reported that French-dominant French-English bilinguals perceived the /p/-/b/ contrast with French-like values in both languages, thus showing an influence of the L1 on the L2. Other studies have reported that L2 speakers may use perceptual categories that are intermediate between native and target language values. For instance, Caramazza et al. (1973) evaluated the perception of voiced and voiceless stops by monolingual and bilingual speakers using an identification task involving synthetic stimuli ranging from -150 to 150 ms VOT. The participants were ten Canadian English monolinguals, ten Canadian French monolinguals and twenty L1-French L2-English bilinguals who had started learning the L2 by the age of 7. The results showed that the English monolinguals identified voiceless stops later in the continuum than French monolinguals, as expected. They also showed a sharper and more consistent identification slope, which indicated that English speakers were more sensitive to the VOT cue than French speakers. Regarding the bilingual speakers, the results showed that they performed similarly in both languages, with intermediate VOT values between those of monolingual speakers of English and French, and obtained less consistent slopes than the English monolinguals and sharper identification functions than the French monolinguals. Similarly, Flege and Eefting (1987b) reported that Spanish-English early bilinguals and late childhood bilinguals living in Puerto Rico had /t/-/d/ category boundaries that were shorter (i.e., more Spanish-like) than those of English monolinguals – whereas bilinguals since birth did not. Flege and Eefting (1987a) found that Dutch speakers of English of three different levels of proficiency had a significantly later category boundary in English than in Dutch, but the difference was small. These results indicate that L2 speakers and L1-exposed bilinguals tend to have single perceptual categories for their two languages, which was either the L1 category (Hazan, & Boulakia, 1993), or an intermediate cateogry (Caramazza et al., 1973; Flege & Eefting 1987a, 1987b). Hence, inability to develop separate categories may be explained by the fact that the L2 speakers were in their L1 country (Caramazza et al., 1973; Flege, & Eefting, 1987a) and exposed to L1-accented English (Flege, & Eefting, 1987b). Along similar lines, Ahn, Chang, DeKeyser and Lee-Ellis (2017) found that L1 Korean L2 English speakers started to perceive L1 contrasts less accurately the earlier their L1 input was reduced, underscoring the effect of relative exposure to the L1 and the L2.

Gorba (2019) contrasted the perception of L1 and L2 stops by three groups of Spanish learners of English differing in amount of L2 experience and learning setting, but with a similar amount of L2 instruction (about 13 years). The least L2 experienced group (INEXP) had never lived in an L2 setting, the group of experienced learners in Spain (EXP1) had spent an average of 9 months in an L2 setting and were back in their home country at the time of testing, and the most experienced group (EXP2) had been living in an L2 setting for a mean of 4 years and were still living in an immersion setting at the time of testing. It was found that all learner groups had numerically earlier – more Spanish-like – perceptual boundaries in English than English monolinguals, but this difference was only significant in the case of INEXP. When it comes to their L1, all groups presented longer VOT values than Spanish monolinguals, but only EXP2 had significantly later /p/-/b/ boundaries than Spanish monolinguals, indicating an influence of English on their L1. That is, instances of L1 influence on the L2 as well as L2 influence on the L1 were found to be related to amount of L2 exposure.

While results for production, despite some exceptions, tend to show an effect of experience on the L2, and also on the L1 for highly advanced learners in an immersion setting (in addition to cases of novelty effects, Chang, 2012), the results for perception are less consistent, probably due to an insufficient number of studies. One of the main contributions of the current paper lies in the fact that it investigates the effect of L2 experience not only on VOT production, like previous studies such as Flege (1987), but also on VOT perception, which has received scarce attention in the literature, particularly regarding L2 influence on the L1. In addition, L2 speakers' performance in their L2 is contrasted with their own performance in the L1, as well as with data from monolingual speakers of each language. The role of L2 experience is assessed by comparing groups of L2 learners that differ in amount and type of L2 experience, namely number of years learning the L2 and learning setting. Although factors such as age of onset of L2 learning, amount of L2 instruction, L2 use and language dominance are often interrelated with other forms of experience (Kartushina, Frauenfelder et al., 2016), the current paper will mainly focus on years of learning and learning setting. However, the possible role of additional factors will be addressed in the discussion. The first two sections of this paper have reviewed studies that presented different results for CLI and the possible factors that motivate them. The role of these factors is reflected in the major L2 speech models, which will be introduced in the next section, followed by a section reviewing the relationship between perception and production.

1.3. L2 speech models

Some current models of cross-linguistic perception and L2 speech, such as the Perceptual Assimilation Model (PAM, Best, 1995; PAM-L2, Best, & Tyler, 2007) and the Speech Learning Model (SLM; Flege, 1995, 2002, 2007; SLM-r: Flege, & Bohn, 2021), posit that CLI may be motivated by the coexistence of the L1 and the L2 in a common phonetic space. PAM claims that non-native phones are perceived or categorized in terms of native categories (i.e., assimilated to L1 phones), or may be uncategorized or even heard as non-speech sounds (Best, 1995). As a consequence, in the initial stages of L2 acquisition, the interaction between the two systems, that is, the categorization of non-native sounds in terms of L1 categories, may block the creation of new categories for L2 phones (Flege, 1995; Best, 1995). However, L1 categories established in childhood may also experience changes as a consequence of L2 acquisition (Flege, 1987). The SLM, as well as its recently revised version SLM-r, describes two ways in which, as a consequence of experience with the L2 (or FTE, in SLM-r terms), L1 and L2 categories may interact, with possible consequences for existing L1 categories. L1 and L2 categories may drift toward one another by a process of category assimilation, that is, the development of an L2 category in the vicinity of an L1 category may result in a merged category that presents characteristics that are intermediate between the two languages (Flege, 1987, 2002). According to the nature of the linguistic input received, the merged category may present more L1-like or L2-like characteristics. Alternatively, the acquisition of an L2 category may result in a process of category dissimilation, that is, the new L2 and existing L1 category may drift away from one another in order to maintain a difference between them, resulting in L1 – and possibly L2 – categories that differ from monolingual speakers' categories (Flege, 2002). Exposure to the L2 may thus help the creation of a new category for the L2 phone, but this new category may affect existing L1 categories. Support for Flege's hypotheses comes from studies that have found that L2 learners may perceive and produce L1 and L2 stops similarly with values that are intermediate between the two languages (e.g., Caramazza et al., 1973; Flege, 1987; Flege, & Eefting, 1987b; Williams, 1977). For instance, the late French-English bilinguals living in an L2 English setting in Flege (1987) had a single category for French /t/ and English /t/ (that is, their L1 and L2 productions did not differ significantly) and this intermediate merged category had VOT values that were significantly different from the mean VOT for monolingual French and monolingual English speakers. Regarding category dissimilation, as discussed above, Flege and Eefting (1986, 1987b) reported that early Spanish-English bilinguals produced Spanish /t/ with shorter VOT values than Spanish native speakers, presumably in order to make their L1 and L2 categories more phonetically distinct.

Conversely, the Second Language Linguistic Perception Model (L2LP, Escudero, 2005, 2009) rejects the idea of a common phonetic space and posits that bilingual speakers have separate L1 and L2 systems. Initially, the L2 system is an identical copy of the L1, which develops towards target-like values as phonetic information specific to the L2 becomes available to the learner. As a result, the L2 categories will gradually become distinct from those of the L1 and present more target-like values. Given that the two systems are separate, no direct interaction between the L1 and the L2 categories is expected, although intermediate perceptions – i.e., with values between the two languages – may take place. Escudero explains cases of intermediate perception in terms of Grosjean's (2001) language mode hypothesis, which claims that the state of language activation ranges on a continuum from a purely monolingual to a purely bilingual mode. Thus, according to the L2LP, CLI is motivated by the simultaneous activation of the two parallel systems. The results of some previous research showing that L2 learners perceive similar L1 and L2 phones differently can be interpreted in L2LP's terms as support for the existence of separate systems for the two languages (e.g., Casillas, & Simonet, 2018; Escudero, 2005; Gonzales, & Lotto, 2013). For example, Escudero (2005) tested her own hypothesis on the perception of /ɛ/-/æ/ – a contrast that the author reports exists both in Canadian English and Canadian French but which is implemented differently in terms of cue weighting – by Canadian L1-English learners of Canadian French in each language. It was found that the L2 learners perceived the contrast differently when it was presented in English as opposed to French, as listeners adjusted cue weighting – first formant (F1) and duration – to the language that was being tested. That is, the learners relied mostly on F1 in the French condition and on both F1 and duration in the English condition. This result was interpreted as evidence that L2 learners can establish new categories for the L2 that are separate from the L1 categories and that cue weighting can be adjusted according to language. Casillas and Simonet (2018) tested the perception of Spanish and English stops by simultaneous English-Spanish bilinguals in a study that carefully controlled language mode activation and found that bilinguals had separate L1 and L2 category boundaries for the /p/-/b/ contrast. Moreover, results also revealed that, after a short immersive experience – i.e., a seven-week intensive course –, beginning English learners of Spanish started to make a difference between L1 and L2 /p/-/b/ categorization. The data obtained in the present paper will be discussed in light of the SLM's and the L2LP's main premises.

1.4. The relationship between perception and production

Another theoretical issue that is addressed in the present study is the relationship between perception and production. Although it has been widely discussed in the literature, there is still no clear consensus regarding the nature of the relationship between the two modalities, that is, whether they develop in parallel or sequentially and, if so, in what order. In the case of L1 acquisition, links between perception and production characterize the process of L1 development (Kuhl et al., 2008). Some speech theories such as the Motor Theory (Liberman, & Mattingly, 1985) and the Direct-Realist approach (Fowler, 1986, 1990) posit that there is a link between the two dimensions, given that adults perceive speech and create phonetic categories based on articulatory gestures of their own speech. However, while many studies support a clear link between the two dimensions in the L1 (e.g., Brunner, Ghosh, Hoole, Matthies, Tiede, & Perkell, 2011; Fox, 1982; Newman, 2003), a lack of a connection between the two modalities has also been reported (Bailey, & Haggard, 1973; Shultz, Francis, & Llanos, 2012). For example, Newman (2003) reported that speakers that produced /p/ with the longest VOT values also showed a preference towards longer VOT values in a goodness rating task. Conversely, Schultz et al. (2012) found no link between the use of VOT and F0 in production and the weighting of these cues in an identification test. The difference between the outcomes of these two studies may be related to a difference in the perceptual measures used: a goodness measure of an already identified category vs. a measure of whether or not a given stimulus is identified as a given category. It is possible that a closer relationship between perception and production may be obtained with the former.

Regarding the L2, L2 speech theories such as the SLM (Flege, 1995) and the PAM-L2 (Best, & Tyler, 2007) assume that there is a link between the two dimensions and that accurate L2 perception tends to precede accurate L2 production. The revised version of the SLM (Flege, & Bohn, 2021), however, posits that perception and production co-evolve without precedence. It also claims that, even though there exists a strong bidirectional influence between the two modalities, their correspondence is never perfect, not even in the case of monolingual speakers. Empirical evidence, in fact, yields a variety of results. Some studies support the claim that perception precedes production (e.g., Bohn, & Flege, 1990; Flege, Mackay, & Meador, 1999, Nagle, 2018), while others suggest that L2 learners may produce L2 phones accurately even if they are not able to perceive them in a target-like manner (e.g., Caramazza et al., 1973; Flege, & Bohn, 1997; Flege, & Eefting, 1987a; Trofimovich, & John, 2011). For instance, Bohn and Flege (1990) found that L1-German L2-English speakers succeeded in distinguishing the /ɛ/-/æ/ contrast in an identification task, but they were not able to produce it accurately. Similarly, Nagle (2018) found that American English speakers learning Spanish in an instructional setting improved their identification of Spanish /p/ and /b/ to a greater extent than their production, which developed later in the course. Still, a greater amount of variability was observed in production than in perception, as some learners improved their production more rapidly than others. In other words, even though perception consistently preceded production, accuracy in perception and production did not develop at the same pace for all participants, resulting in different alignments between the two modalities for different learners throughout the process of L2 acquisition. Evidence of production development occurring prior to perception development has also been found (e.g., Caramazza et al., 1973; Flege, & Eefting, 1987a; Trofimovich, & John, 2011). Flege and Eefting (1987a) reported that Dutch learners of English produced the /t/-/d/ contrast similarly to English native speakers but were not able to discriminate it in a native-like manner. Similarly, Trofimovich and John (2011) reported that L1-Canadian French learners of English produced the /θ/-/t/ and /ð/-/d/ contrasts accurately, but showed difficulty in distinguishing them perceptually in an auditory priming task. The authors allude to the relative visual salience of the dentals and differences in word familiarity between the production stimuli and the perceptual stimuli among possible causes for the different results. They also report a notably greater variability in the L2 production of the English sounds than in native English production. In addition, it has been suggested that greater social or external pressure on production accuracy than on perceptual accuracy may explain a precedence of production over perception (Llisterri, 1995). For example, L2 speakers' difficulty perceiving the English /θ/-/t/ and /ð/-/d/ contrasts may have a smaller impact on native speakers' perception of L2 speech than an inaccurate production of the same contrasts.

Another finding reported in the L2 literature is the lack of a clear relationship between perception and production, especially regarding consonants (e.g., De Leeuw, Stockall, Lazaridou-Chatzigoga, & Gorba, 2019; Hattori, & Iverson, 2010; Peperkamp, & Bouchon, 2011; Sheldon, & Strange, 1982). For example, Sheldon and Strange (1982) investigated the relationship between the perception and production of the /r/-/l/ contrast by Japanese learners of English. They found that a subset of the participants were more successful at producing the contrast than at perceiving it, whereas the opposite was true for other participants. De Leeuw et al.'s (2019) study on phonotactic constraints also indicated that the two modalities may develop – at least to a certain extent – independently, as no relationship between accuracy of discrimination between word-initial /esp-/ and /sp-/ sequences and their production by Spanish learners of English was found. Moreover, an individual inspection of the data revealed that some participants performed more accurately in the perception task, whereas others were more accurate in production. Further, variables like L2 use and L2 proficiency were associated with a more target-like production, but these variables did not predict the discrimination accuracy.

The inconsitent results regarding the development of production and perception may be linked to cross-study differences in the tasks and measures used, the populations and target sounds tested, and the degree to which perception and production measures are comparable (Llisterri, 1995; Mack, 1989). Although cross-study differences make it difficult to pinpoint the effect of variables such as L2 experience, AOL and L2 proficiency, it has been found that experienced learners may present a greater alignment between the two dimensions than inexperienced learners (Bohn, & Flege, 1990) and that factors such as L2 use and L2 proficiency may have a clearer effect on production than on perception (De Leeuw et al., 2019). In addition, a stronger perception-production link has been reported for vowels (e.g., Bohn, & Flege, 1990) than for consonants, where often no clear link was established or, contrary to the SLM's predictions, production appeared to precede perception (e.g., Sheldon, & Strange, 1987; Williams, 1977). The degree of similarity between the L1 and the L2 phones may also influence the relationship between perception and production. For instance, Levy and Law (2010) found that /y/-/œ/ – two new phones – presented a clearer relationship between the two modalities than /u/-/y/ – a similar and a new phone, respectively. Another possible explanation for this apparent lack of consistency across studies may be related to the SLM-r's theoretical claim that the relationship between the two dimensions is never perfect, neither in bilingual nor in monolingual speakers. The present paper attempts to evaluate this relationship further by testing the perception and production of L1 and L2 stops by English learners of Spanish differing mainly in amount of L2 experience.

1.6. The present study

The present study examines the perception and production of L1 and L2 stops by English learners of Spanish differing in amount and type of L2 experience, that is, years of learning and experience living in the L2 setting. Thus, the goals of the present study are (1) to analyze the effect of amount and type of experience on L2 speakers' and production of L1 and L2 voiceless stops; (2) to investigate the influence of L2 experience on L1 and L2 perception of voiceless stops; (3) and to examine the relationship between perception and production in L1 and L2 stops also in light of differences in type and amount of L2 experience. The methodological contribution of this study, thus, stems from the fact that (1) both the perception and the production of stops by L2 learners are evaluated, (2) the learners' performance in both their L1 and their L2 is contrasted, and (3) the performance of the learners is compared to that of monolingual speakers of the L1 and of the L2. To our knowledge, only Gorba (2019) investigated the effect of experience in an L2 setting on the perception of both L2 and L1 stops.

Considering the findings of previous studies (e.g., Flege, 1987; Sancier, & Fowler, 1997), it is hypothesized that participants with more years of experience, particularly in an L2 setting, will present more target-like VOT values in the L2, and are more likely to show instances of L1 phonetic drift towards the L2. This assumption is in agreement with the SLM/SLM-r's claim that CLI is bidirectional. Moreover, if the SLM's claims are supported, we should expect that the learners with less experience in the L2 setting will present a single category for the two languages, given the similarity of the L1 and L2 phones, whereas the most experienced learners may have received enough exposure to the L2 so as to present separate L1 and L2 categories. By contrast, the L2LP would hypothesize that all learners – regardless of their experience in the L2 – would present separate L1 and L2 categories, although the L2 categories will resemble the L1's to a greater extent in the case of less experienced learners, as their L2 is still evolving towards L2-like values. Note, however, that the two models may be able to explain similar scenarios, as both of them can account for realizations that are intermediate between L1 and L2 values either in terms of the simultaneous activation of both the L1 and the L1 when language mode is not controlled (L2LP) or due to the process of category assimilation (i.e., merger hypothesis, SLM). Regarding the relationship between the two modalities in the L1, according to the Motor Theory (Liberman, & Mattingly, 1985) or the Direct-Realist approach (Fowler, 1986, 1990), a straightforward link should exist between the two dimensions, given that speakers base their perception on their own articulatory gestures. As for the L2, if production difficulty results from inaccurate perception, we may expect greater accuracy in production the more target-like the perception is (Flege, 1995, 2002), and a more target-like performance in perception than in production. Alternatively, the two modalities may show neither a precedence relationship nor a perfect alignment, as this may not be the case even regarding monolingual speakers (Flege, & Bohn, 2021).

The remaining of this paper presents the results of two experiments. Three groups of L1-English L2-Spanish speakers completed a production and a perception experiment under two conditions, namely in the L1 and in the L2. A group of functional monolinguals of each language also completed the experiment. Experiment 1 assessed the production of /p/ and /k/ in each language. Experiment 2 consisted of two forced-choice perception tests – one for the /p/-/b/ contrast and one for the /k/-/g/ contrast – which were presented in the two languages.

2. Experiment 1: Production of L1 and L2 voiceless stops

2.1. Methodology

2.1.1. Participants

A total of 51 participants completed the study, namely 41 young adult English learners of Spanish and 10 Spanish speakers (24 females (F) and 27 males (M), in their twenties and thirties). They included two groups of monolingual speakers (monolingual Spanish and monolingual English speakers) who acted as control groups, and three groups of English learners of Spanish. Two of these groups were living in their home country and differed in amount of L2 experience, namely the number of years learning English and previous experience in an L2 setting. In addition, a third group of English learners living in Spain and who had mostly learned the L2 in the L2 setting was tested. The characteristics of all five groups are described next and summarized in Table 2.

The English control group, henceforth ENCONT (*n*: 9; 4 F and 5 M), was made up of functional English monolinguals who were enrolled in undergraduate or graduate studies at Queen Mary university of London and were mostly from the London metropolitan area. The least experienced group of English learners of Spanish (INEXP; *n*: 11; 6 F and 5 M) had never lived in a Spanish-speaking country, although a few reported brief visits to Spain[1]. They were recruited from the highest levels of Spanish language courses at a Language Center at their institution or were in their first or second year of a BA in Hispanic Studies. They had started learning Spanish at a mean age of 13.4, for a mean of 3.8 years, and reported using Spanish an average of 12.3% of time during the week, as reported in a language background and use questionnaire that all participants completed (see section 2.1.2). A language dominance score was calculated using the Bilingual Language Profile (BLP, Birdsong, Gertken, & Amengual, 2012), which renders a score that ranges from -218 to 218 representing the two monolingual endpoints (negative for Spanish, positive for English), a value of zero indicating balanced bilingualism. INEXP obtained a dominance score of 108.7, that is, highly L1-dominant. The group living in London with some past experience in an L2 setting (EXPinUK; n = 10; 4 F and 6 M) were fourth-year students of Hispanic Studies who had spent at least one term (M: 9.4 months) in a Spanish-speaking country during the previous academic year. They started studying Spanish at a mean age of 15.5, had studied it for a mean of 7.5 years and their weekly percentage of L2 use was 15%. They obtained a dominance score of 106, that is, also clearly English dominant. There were no dialectal differences within the groups tested in the UK, as all participants were Londoners. The third group of L2 learners examined consisted of English language teachers living in Barcelona (EXPinSP; n = 10; 4 F and 6 M), who had spent a mean of 50.7 months, that is, 4.2 years, in the target-language setting and had a heterogeneous dialectal background[2]. Even though they had more experience living abroad, they started learning Spanish later, with a mean age of 18.9 and had received less formal instruction, namely a mean of 2.4 years in Spanish than the other two groups and used mainly English in their jobs. Moreover, their weekly use of Spanish was only somewhat higher than that of the English groups living in their home country (16.3%). As for language dominance, they were also highly English dominant, with a score of 116.5. Finally, the Spanish control group (SPCONT) completed the experiment in Spain. They were also speakers of Catalan – a language with a basically identical VOT system to Spanish (e.g., Julià i Muné, 1981) – and had studied basic English in school, although they rarely used it. Furthermore, they had never lived in an English-speaking country.

Group	*N*	Months in L2 setting	% L2 weekly use	AOL	Years of L2 instruction	Language dominance	Current Location
ENCONT	9	NA	NA	NA	NA	NA	UK
INEXP	11	none	12.3% (8; 3-27)	13.4 (3.7; 9-19)	3.8 (3.3; 1-7)	108.7 (27.3; 81-146)	UK
EXPinUK	10	9.4 (4.2; 4-12)	15% (6.5; 10-30)	15.5 (3.7; 11-20)	7.5 (5.3; 3-18)	106 (21.7; 82-130)	UK
EXPinSP	11	50.7 (27.3; 12-60)	16.3% (10.5; 7-37)	18.9 (2.1; 14-20)	2.4 (2.9; 0-7)	116.5 (26; 80-166)	Spain

[1] In order to use distinct labels for the three groups, for the sake of simplicity, this group will be referred to as “INEXP”, as they had no previous experience living in the L2 setting, even if they were not inexperienced in terms of L2 instruction.

[2] Four participants were from Southern England, one from Northern England, one from Scotland, two from Ireland, one from Canada, one from the East Coast of the United States and one from the South of the United States. Although, ideally, all participants should have the same dialectal background, participants from different English-speaking areas were included in the experiment due to the difficulty recruiting volunteers to participate in the study. It should be noted that dialectal differences were not expected to affect the L1 production of voiceless stops, and, in fact, no significant differences were observed between groups (see results section).

SPCONT	10	NA	NA	NA	NA	NA	Spain

Table 2. Group characteristics including number of participants, gender, L1, L2, mean number of months spent in L2 setting, mean % weekly L2 use, L2 age of learning (AOL), years of L2 instruction. Language dominance and location. Standard deviations and ranges are provided in parentheses.

2.1.2. Tasks and procedure

The production experiment consisted of two reading tasks – one in English and one in Spanish – in which target words were presented at the beginning of a carrier sentence. Each task had 32 sentences, including five beginning with /p/ and five with /k/, as well as fillers (see Appendix A for the complete elicitation list). Voiced stop production was also collected but will not be the object of study in this paper. Each sentence was read twice. The carrier sentences were “X is the next word” in the English task and the Spanish equivalent, i.e., “X es la siguiente palabra”, in the Spanish task. The testing groups completed both tasks in the same session whereas controls were only tested in their language. In order to control for language mode, instructions were provided in the testing language of each part of the experiment (the English and the Spanish part). Moreover, between the two tasks participants watched a short video (i.e., a short TedTalk) in the language of the second task in order to set participants to the right language mode (Grosjean, 2001). One of the authors of the current paper, an L1-Spanish highly-proficient near native speaker of English, was the experimenter in both language conditions.

The production of the Spanish learners living in the UK was recorded in a soundproof room at Queen Mary University of London. A Neumann TLM 103 microphone and a booth mixer Steinberg UR22 MKII were used. The participants in Spain were recorded at the speech lab at Universitat Autònoma de Barcelona, in an acoustically treated room using an Audio-Technica AT 2050 microphone and an Alesis Multimix 8 mixer. In both cases, recordings were made with the software Audacity and the files were saved in WAV format. A few recordings were made in a different institution also in Barcelona in a soundproof cabin using a Shure-SM58 microphone and a Marantz PMD-660 digital recorder. The sound files were also recorded in WAV format.

The perception experiment presented in this paper was also conducted in the same session, following the production task in each language (see section 3), that is, the production and perception tasks in a given language were completed consecutively. Order of completion of the tasks, that is, the language tested (L1, L2) and place of articulation (bilabials, velars), was counterbalanced across participants[3]. An online linguistic background questionnaire based on the BLP (BLP, Birdsong et al., 2012) was completed by participants prior to the testing session.

2.2. Results

VOT measurements were carried out manually using Praat (Boersma, & Weenink, 2016). Measurements were made from the onset of the burst to the beginning of voicing as indicated by the beginning of periodicity in the signal, considering both the spectrogram and the oscillogram.

[3] A two-way analysis of variance (IV: order, group; DV: mean VOT duration) was conducted in order to determine whether order of completion of the tasks had an effect on participants’ performance in production. No significant effect of order was found for English [/p/: $F(1) = 0.292; p > 0.05$; /k/: $F(1) = 2.813; p > 0.05$] nor for Spanish stop production [p/: $F(1) = 3.321$; $p > 0.05$; /k/: $F(1) = 1.744; p > 0.05$] and no interaction between order and group was found in any case ($p > 0.05$ in all cases). A similar analysis was conducted for perception (see section 3; IV: order, group; DV: category boundary) and the same result was replicated. No significant effect of order was revealed neither in the L1 [/p/-/b/: $F(1, 23) = 0.36; p > 0.05$.; /k/-/g/: $F(1, 23) = 2.032; p > 0.05$] nor in the L2 [/p/-/b/: $F(1) = 0.001$; $p = 0.99$; /k/-/g/: $F(1) = 0.095$; $p = 0.760$] and no significant interaction involving order was found ($p > 0.05$ in all cases).

In order to compare VOT production between groups, results were classified according to group, place of articulation and language. Figures 1 and 2 present the distribution of VOT productions obtained by each group in each language.

Considering first the production of stops in the L1, as can be observed from the boxplots (Figure 1), all groups presented a very similar distribution of VOT values in English. The mean productions for /p/ presented VOT values around 50 ms. ENCONT and INEXP had a mean of 52 ms (*SD* = 20), EXPinUK produced a mean VOT of 50 ms (*SD* = 18) and EXPinSP had the shortest mean with 46 ms (*SD* = 18). Just as in the case of /p/, all groups presented similar VOT means for /k/. ENCONT and EXPinSP had a mean VOT of 72 ms (*SD* = 21), whereas INEXP and EXPinSP obtained slightly longer VOT means, with 79 ms (*SD* = 18) and 76 ms (*SD* = 19), respectively.

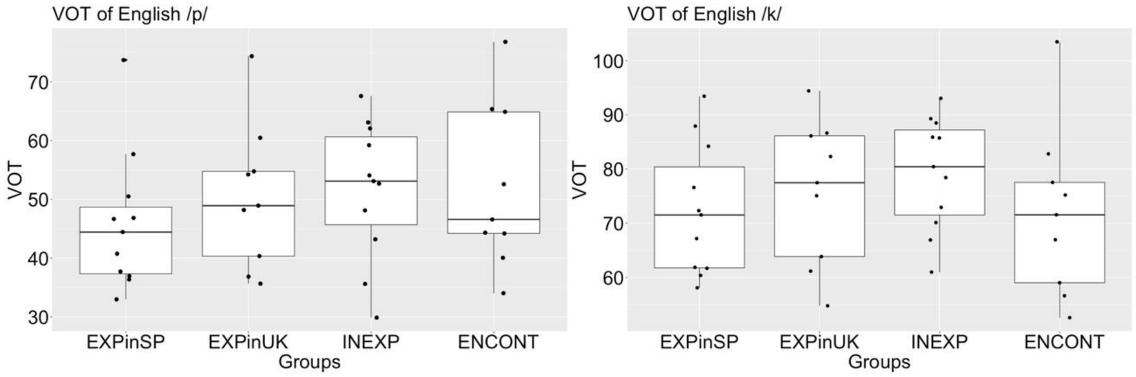

Figure 1. Boxplots of VOT production (in ms) for English /p/ and /k/ per group with jitter points illustrating individual data.

In order to assess if the L2 learner groups differed from each other and from the monolingual groups a series of linear mixed models were conducted with group (three learner groups and the corresponding monolingual group) as the independent variable and VOT as the dependent variable. Separate sets of analyses were conducted for each measure (VOT of English /p/, English /k/, Spanish /p/ and /Spanish /k/). In all cases the best fitting model, as determined by the lowest Akaike Information Criterion (AIC), was a linear mixed model with group as fixed factor, and using a random intercept for participant and for word. Random slopes were evaluated but were discarded because of a larger AIC. As expected from the similar results across groups, the linear mixed model revealed no significant effect of group for either English stop [/p/: $F(3, 36) = 0.562$; $p > 0.05$; /k/: $F(3, 36) = 0.78$; $p > 0.05$].[4] That is, all groups presented English-like VOT values, regardless of amount of L2 experience.

Regarding the production of Spanish stops, the L2 Spanish speakers tended to produce longer VOT than the native Spanish speakers (see Figure 2). EXPinSP produced the shortest VOT values for /p/ with 14 ms (*SD* = 11) – i.e., the closest to SPCONT, who presented 9 ms (range = 1.9 – 28.1) – followed by EXPinUK (*M* = 17, *SD* = 11), whereas INEXP presented the longest mean (28, *SD* = 23). As for /k/, EXPinSP and EXPinUK had the closest means to SPCONT (34 ms, range = 10 -70) with 46 ms (*SD* = 12) and 45 ms (*SD* = 14) of VOT, respectively, whereas INEXP presented 55 ms (*SD* = 21).

[4] See Appendix B for the coefficient estimates, standard errors and t-values for each model parameter regarding the analyses conducted on the production data.

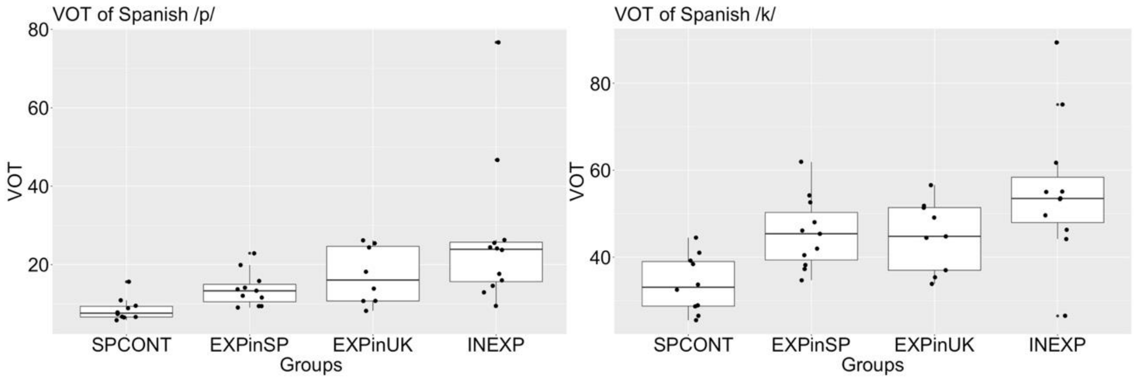

Figure 2. Boxplots of VOT production (in ms) for Spanish /p/ and /k/ per group with jitter points illustrating individual data.

The results of the linear mixed model analysis in this case revealed a significant effect of group on the production of both Spanish /p/ and /k/ [/p/: $F(3, 37.015) = 5.951$; $p < 0.01$; /k/: $F(3, 38.268) = 6.694$; $p < 0.01$]. In both cases, pair-wise comparisons with a Bonferroni correction showed that only INEXP differed significantly from SPCONT ($p < 0.01$ in the case of /p/, $p < 0.001$ regarding /k/). Moreover, regarding /p/, INEXP were also found to differ significantly from EXPinSP ($p < 0.05$). In short, results indicated a positive effect of years of learning and living in an L2 setting on Spanish stop production, given that only INEXP produced stops with significantly longer VOT values, and, thus, differed from the native Spanish speakers, and also from the English group living in Spain. Moreover, the individual inspection of the data revealed that a greater number of participants in the inexperienced group than in the two experienced groups differed from SPCONT.

Next, in order to determine whether English learners of Spanish produced L1 and L2 voiceless stops differently, and what the role of L2 experience may be in this respect, the English and Spanish data from each L2 learner group were contrasted (see Table 3 and Figure 3). In the case of /p/, all groups produced longer VOT values in English than in Spanish, INEXP being the group with the smallest difference between L1 and L2 stops (i.e., 24 ms, vs. 33 ms for EXPinUK and 32 ms for EXPinSP). The effects of group and language on the production of L1 and L2 /p/ and L1 and L2 /k/ were examined in a series of linear mixed models with VOT as the dependent variable, with group, language and their interaction as fixed factors, and a random intercept for participant and for word. Again, slopes were excluded because their inclusion in the model yielded larger AICs. For pairwise comparisons, Bonferroni adjustments were made. Results revealed a significant effect of language on /p/ VOT production [F (1, 8.051) = 193.644; $p < 0.001$]. The effect of group was only marginally significant [F (2, 28.011) = 3.02; $p = 0.065$], probably triggered by the marginally significant difference between EXPinSP and INEXP, as revealed by a pair-wise comparison ($p = 0.064$). All groups were found to produce significantly longer VOT values for /p/ in their L1 than in the L2 ($p < 0.01$ in all cases, as indicated by pairwise comparisons). The interaction between language and group reached significance [F (2, 577.037) = 6.012; $p < 0.01$], indicating that groups varied in the degree to which the two languages differed; specifically, the difference was smaller for INEXP than for the other two groups (i.e., 24 ms vs. 32-33 ms; see Table 3).

	/p/				**/k/**			
Group	**English *M***	**English range**	**Spanish *M***	**Spanish range**	**English *M***	**English range**	**Spanish *M***	**Spanish range**
EXPinSP	46 (18)	4.8 - 94	14 (11)	2.7 – 83.8	72 (21)	23.6 – 143.1	46 (12)	18.1 – 76.8

EXPinUK	50 (18)	15.8 – 118.1	17 (11)	1.8 - 50	76 (19)	39.6 – 124.1	45 (14)	12.9 – 109.4
INEXP	52 (20)	11.2-131.1	28 (23)	5.8 – 112.4	79 (18)	41.1 – 128.8	55 (21)	11 – 125.3

Table 3. Mean VOT productions and ranges (in ms) obtained by each learner group in English and in Spanish (Standard deviations are given between parentheses)

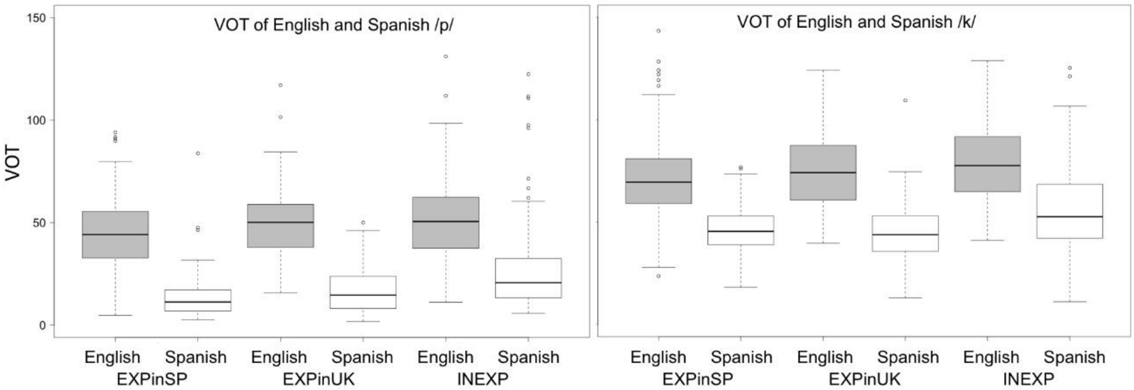

Figure 3. Production of /p/ and /k/ in English and in Spanish by each of the L2 groups.

Similarly, all learner groups produced longer VOTs for /k/ in English than in Spanish, especially the two groups with experience living in an immersion setting. The differences between L1 and L2 stops in this case were 24 ms, 31 ms and 26 ms for INEXP, EXPinUK and EXPinSP, respectively. In this case, the results of the linear mixed model revealed a significant effect of language on /k/ VOT production [F (1, 8.024) = 79.97; p < 0.001] but no significant effect of group [F (2, 28) = 2.479; p > 0.05], and no group × language interaction [F (2, 578) = 2.566; p > 0.05]. Bonferroni-adjusted pair-wise comparisons revealed a significant difference between the production of /k/ in Spanish and in English for all groups (p < 0.01). All in all, English learners of Spanish appeared to adjust the use of VOT production according to language, as they produced significantly longer VOT in English than in Spanish. This difference was found to be greater in L2 learners with experience in an immersion setting than in inexperienced learners, especially in the case of /p/.

2.3. Discussion

The results of the production experiment indicate a positive effect of L2 experience on the production of L2 Spanish voiceless stops, as INEXP, but not the two experienced groups, produced Spanish /p/ and /k/ with significantly longer VOT values than the Spanish controls. In the case of /p/, the inexperienced group also differed from the group in Spain. Furthermore, a greater number of participants in the inexperienced group than in the two groups with a greater amount of experience obtained a greater VOT mean than SPCONT's. This beneficial effect of residence in an L2 setting is in agreement with some previous studies (e.g., Bohn, & Flege 1990; Flege, 1987; Stevens, 2001), and can be linked to the greater exposure to the target language on the part of groups with greater L2 experience, and the concomitant greater amount of native input, in comparison to the group with no experience abroad. Recall that the latter had not spent time in a Spanish-speaking context whereas the experienced groups had spent an average of 9.4 months (EXPinUK) or 4.2 years (EXPinSP). Although the EXPinUK had spent a relatively short period of time and were tested when they were back in the UK, the input they received seems to have been enough to trigger the development of a more Spanish-like production. Additionally, or alternatively, the EXPinUK's formal learning of Spanish may have also contributed to this group's more target-like performance. This possibility was explored in a regression analysis conducted on individual factors, including number of months in an L2 setting, amount of L2 use, onset age of L2 learning, years of L2 instruction and dominance score[5]. However, the results did not reveal a significant effect of any of these factors on L2 performance. Thus the more target-like performance of the two more experienced groups cannot be linked specifically to the greater amount of instruction received by EXPinUK or to the role of the L2 setting in the case of EXPinSP, but possibly to a combination of the different sources of experience and the overall longer period of learning. In short, it appears that, as previous studies have shown (e.g., Casillas, 2020; Zampini, 1998), English learners are capable of producing Spanish voiceless stops with VOT values that resemble those of Spanish speakers. Still, learners with L2 experience still produced somewhat numerically longer VOT values than the Spanish controls, especially in the case of /k/. This numerical difference indicates that there is still some influence from the L1.

With respect to the production of L1 stops, no sign of phonetic drift towards the L2 was found: all learner groups presented native-like VOT values in English. Thus, this study did not replicate earlier findings such as those reported by Flege (1987), who found that a more target-like L2 VOT production often resulted in a less native-like L1 production. The lack of an effect of the L2 on the L1 may have to do with the fact that the highly experienced group in the present study had spent a considerably shorter period of time in the target language country than the highly experienced group in Flege's study (4.2 years and 12 years, respectively). Another important difference between the two studies has to do with amount of L1 and L2 use, found to be relevant factors in determining the amount of accentedness in L2 speech (Piske et al., 2001). The type of experience of the group living in the target language country in the present study may not reflect the expected experience in an immersion setting. Despite living in Spain, this group reported using mostly English at work and most of them had English-speaking partners, resulting in a relatively low L2 use (16.3 % weekly use, see Table 2). As a matter of fact, the regression analysis conducted on individual factors revealed that a smaller L2 use was related to longer VOT productions in the L1, thus showing a link between amount of L2 use and L1 performance. However, the three learner groups were similar in amount of L2 use and thus their performance in the L1 did not differ. The apparent lack of a strong influence of the L2 on the L1 can also be discussed in theoretical terms, as the fact that there was no change in the L1 while L2 production improved is consistent with the L2LP's premise that L2 phonological learning is independent of the L1 phonological system and provides no support for the SLM's claim that CLI is reciprocal.

The second issue addressed was whether learners are capable of maintaining separate categories for L1 and L2 stops or whether a single category was used for both languages. In this regard, a remarkable outcome is that all groups of learners appear to present separate L1 and L2 categories for both /p/ and /k/. Stops were consistently produced with significantly longer VOT in English than in Spanish. Interestingly, this was even found with the learners with no experience abroad, the only group that differed from the Spanish controls and for whom the difference between L1 and L2 stops was smallest. This outcome could be explained by the fact that language mode was accurately controlled in the production task, as participants had to read a series sentences in the language that was being tested and the elicitation for each of the languages was carefully separated. As a result, a more clearly monolingual mode might have been activated, thus preventing the use of intermediate categories (Escudero, 2005). It remains to be evaluated, however, how Spanish-like would L1-Spanish speakers judge the L2 learners' productions of Spanish /p/ and /k/ to be, particularly in the case of INEXP. Nevertheless, the fact that the

[5] The general effect of individual differences was tested by simultaneously testing all coefficients and the significance of each independent variable was assessed by conducting individual regression coefficients for each factor. With respect to L1 production, a significant effect of individual factors was revealed [F (5, 18) = 10.085; p = 0.000; R^2 = 0.737]. The individual coefficients indicated that a smaller use of the L2 [t (18) = -3.716; p = 0.002] and a shorter period of time spent in the L2 setting [t (18) = -2.609; p = 0.018] were related to using longer VOT values in the L1. No effect of the other variables was observed (p > 0.05). No significant effect of the individual factors was found for L2 production [F (5, 18) = 1.652; p = 0.197; R^2 = 0.315].

experienced groups' L1 and L2 productions did not differ from monolingual L1 speakers' production, and differed from one another, point to the development of separate categories for L1 and L2 sounds (Flege, 1987). The SLM (Flege, 1995) predicts that similar L2 phones – such as English and Spanish voiceless stops – are the most difficult type of phone to acquire, as they are initially assimilated to the L1 category, and category formation is, at least initially, blocked. With adequate exposure to the L2, L2 learners may be able to create a separate category that will evolve towards more target-like values. Results in this study support this prediction. In addition, the results of INEXP indicate that simply through L2 instruction the less experienced English learners of Spanish were able to make a difference between L1 and L2 stop production. It is possible that the use of a controlled task such as a sentence reading task may have favoured the accuracy of L2 performance. It would be, of course, interesting to explore if differences between L1 and L2 production were also observed in more spontaneous speech. Still, there is evidence that increased L2 experience improves L2 production. As mentioned above, the experienced groups presented VOT values that were comparable to those of Spanish speakers, whereas the least experienced learners produced Spanish stops with significantly longer VOT than Spanish controls. Results can also be discussed in terms of Escudero's (2005) L2LP, as the fact that all groups – regardless of L2 experience – presented significantly greater VOT values in English than in Spanish supports the existence of a separate L2 category that develops towards target-like values. Some L2 instruction was sufficient for L2 learners to distinguish L1 and L2 stops, as was revealed for the least experienced group. Moreover, L2 categories were more target-like in the case of the more experienced groups than in the case of the least experienced group, supporting Escudero's premise that the new L2 category will initially present L1-like values that will develop towards more target-like values. The following section will present the perception experiment and will, thus, provide a bigger picture of how English learners of Spanish acquire L2 phones.

3. Experiment 2: Perception of L1 and L2 voiceless stops

3.1. Methodology

3.1.1. Participants

Participants in this experiment were the same as the ones described in Experiment 1 (see Section 2.1.1).

3.1.2. Stimuli

A /p/-/b/ and a /k/-/g/ VOT continua with 17 stimuli each were created to test the identification of the two stop contrasts by English learners of Spanish. Vowel /i/ was selected to follow the stops because it has been found to be one of the most perceptually equivalent vowels in the two languages under study (Cebrian, 2019). Hence, first, a highly proficient L1-Spanish L2-English male speaker, who was a trained phonetician, was recorded producing a series of [bi], [pi], [p^hi] and [gi], [ki], [k^hi] syllables. The recording took place in an acoustically treated room at the Speech Laboratory at Universitat Autònoma de Barcelona, Spain, using an Audio-Technica AT 2050 microphone and an Alesis Multimix 8 mixer. The software used for the recording was Audacity.

A few steps were followed in order to identify the best token to be selected as the basis for each VOT continuum (/p-b/ and /k-g/) and to create the continua. First, secondary voicing cues, namely burst duration and intensity and F1 and F0 contour in the vowel, were neutralized by selecting a burst and a vowel that presented intermediate values between the speaker's voiced and voiceless stop productions. The intention was to obtain a single voicing-ambiguous burst and vowel to be used as the basis for all the stimuli in each continuum, one for each place of articulation (/b/-/p/, /g/-/k/). The VOT continua were created using Praat (Boersma, & Weenink, 2016). Stimuli with voice-lead VOT were created by the addition cycles of prevoicing of about 5 ms extracted from a selected prevoiced token and stimuli with positive VOT values were created by inserting portions of aspiration from a selected aspirated production. The objective was to create a continuum that went from a clearly voiced to a clearly voiceless stop, appropriate to test both English and Spanish perception, while limiting the number of steps to a practical number. The continua were piloted on six Spanish monolingual speakers and six English monolingual speakers in order to decide on an appropriate range to be used in both languages while maintaining the same number of steps. Following this pilot, the range of the /p/-/b/ contrast was set at -30 ms to 58 ms, and for /k/-/g/ at -25 ms to 72 ms. The resulting 17 steps in the continua were distributed at about 5 ms intervals, as illustrated in Table 4[6].

Stimulus	/p/-/b/ VOT steps	/k/-/g/ VOT steps
1	-30.4	-25
2	-25	-20.2
3	-20.9	-15.1
4	-14.8	-10.3
5	-9.9	-5.6
6	-5.1	16.4
7	8.6	21.9
8	13.8	26.9
9	18.7	31.7
10	23.7	36.6
11	28.8	41.7
12	33.6	46.9
13	38.5	50.2
14	43.3	56.4
15	48.7	61.7
16	53.4	66.3
17	57.9	72

Table 4. /p/-/b/ and /k/-/g/ continua: actual VOT values in ms.

3.1.3. Task and procedure

The bilabial and the velar stimuli were presented in separate two-alternative forced-choice identification tests – one for each place of articulation – and under two conditions, namely in English and in Spanish. The test used to examine bilabial stop perception was the same as the one completed by Spanish learners of English in Gorba (2019). The English and Spanish versions of the tests included the same stimuli but language-specific instructions and response alternatives were provided for each condition. The response options in all cases included keywords starting with the voiced or voiceless stop followed by the high front vowel /i/ and whose pronunciation was unequivocal from the spelling, to prevent orthographic ambiguities from influencing identification results (Hayes-Harb, Nicol, & Barker, 2010). Thus, the English /p/-/b/ test showed "b as in beetle" and "p as peeler" as response alternatives, whereas the options provided in Spanish were "b como en bicho" (*Eng*. "b as in bug") and "p como en pico" (*Eng*. "p as in beak"). As for the velar stop tests, the response alternatives were "g as in geeky" and "k' as in keeper" in English and "g como en guiso" (*Eng*. "g as in stew") and "k como en quito" (*Eng*. "k as in I remove") in

[6] Even though the two continua have the same number of steps, which vary in intervals of 5 ms, the difference between the two endpoints is greater regarding the /k/-/g/ continuum than in the case of the /p/-/b/ continuum. This difference stems from the fact that the burst used to create the velar continuum, which had a duration of 16.4 ms, is longer than the burst used in the bilabial continuum, with 8.6 ms, as a result of place of articulation. It should also be noted that there are no steps with 0 ms of VOT given that all stimuli include a burst, resulting in a VOT increment of 13.7 ms between the last prevoiced token and the first one with positive values in the case of the bilabial continuum and of 22 ms regarding the velar stop continuum.

Spanish.[7] Stimuli were presented in a random order and were repeated four times throughout the task, resulting in a total of 68 trials for each identification test (17 stimuli x 4 repetitions). Each stimulus was presented once on a given trial and could be replayed one more time if necessary.

3.2. Results

The number of times a stimulus in the continuum was identified as the voiceless category (/p/ or /k/), out of the total of four repetitions presented in the task, was calculated for every participant and every task. The identification functions obtained for each individual arranged by learner group and separated per language and place of articulation are presented in Appendix C. In order to calculate the category boundary for each participant, a logistic function for each participant's responses was calculated using the software SPSS Statistics (IBM Corp., 2015). The constant (b0) and slope (b1) of the logistic function were then used in the formula -LN(b0)/LN(b1) (Aliaga-García, & Mora, 2009; Gorba, 2019).

Figure 4 presents the distribution of the category boundaries obtained for each group for the English /p/-/b/ and /k/-/g/ contrasts. As can be observed, the performance of all four groups was very similar. All three L2 learner groups presented mean category boundaries for English /p/-/b/ that were similar to that of the native English speakers (ENCONT: *M*: 14 ms), even if the learners' boundaries were numerically greater, especially in the case of the experienced groups (EXPinSP: 16.1 ms; EXPinUK: 15.9 ms; INEXP: 14.5 ms). As for English /k/-/g/, ENCONT (*M*: 29.7 ms; range = 22.6 – 35) and EXPinSP (*M*: 29.4 ms) presented similar category boundaries, while EXPinUK's and INEXP's boundaries were located somewhat earlier in the continuum (27.3 ms and 27.2 ms, respectively). A one-way ANOVA exploring the effect of group, including the three groups of learners and the control group, was conducted for each place of articulation. The analysis did not find a significant effect for either the /p/-/b/ or the /k/-/g/ boundary [/p/-/b/: F (3, 37) = 1.162; p > 0.05; /k/-/g/: F (3, 37) = 595; p > 0.05]. The results of the statistical analyses, thus, suggest that English learners of Spanish had native-like L1 perceptual categories, unaffected by L2 experience.

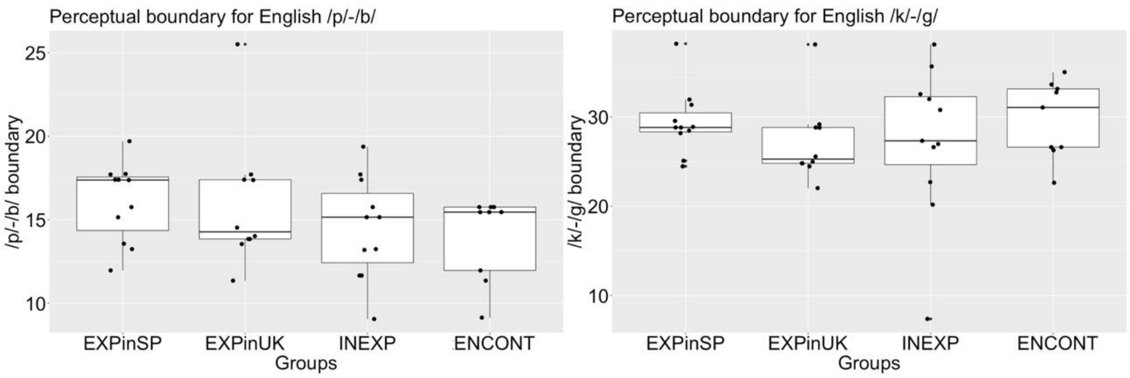

Figure 4. Perceptual category boundaries (in ms) for English /p/-/b/ and /k/-/g/ per group (jitter points illustrate individual data).

[7] An anonymous reviewer pointed out that the name of the letter <g> (/dʒi/ in English and /xe/ in Spanish) is different from the intended sound (/g/). Thus, despite the unambiguous keyword, the presence of the grapheme g in the response may have led to the activation of sounds other than /g/ and thus affected the results for /k/-/g/. While this is a limitation (which affects both the Spanish and the English tests), we expect the inclusion of a practice session before the experiment and the presence of unambiguous /g/ stimuli in the task may have made the intended sound clear. In fact, the results for /k/-/g/ are the expected when compared to the pattern observed for bilabials.

With respect to Spanish (L2) perception, the three groups of English learners of Spanish presented later category boundaries for /p/-/b/ than SPCONT (*M* = 3.4 ms; range = -7.9 – 9.4). EXPinSP presented the lowest category boundary (14.9 ms), followed by INEXP (15.7 ms) and EXPinUK (16.5 ms). In the case of /k/-/g/, English learners also needed greater VOT values to perceive /k/, as they had a later category boundary than SPCONT, which was located at 14.8 ms (range = 0 – 27.3). INEXP had the closest boundary to that of the native Spanish speakers (25.2 ms), followed closely by EXPinUK (25.6 ms). Contrary to expectations, EXPinSP yielded the latest category boundary (28 ms). In this case, the one-way ANOVA conducted on each pair of phones revealed a significant effect of group in both cases [/p/-/b/: F (3, 37) = 19.828; $p < 0.01$; /k/-/g/: F (3, 37) = 8.282; $p > 0.01$]. Bonferroni pair-wise comparisons showed that all groups differed significantly from SPCONT ($p < 0.01$ in all cases), whereas the learner groups did not differ from one another ($p > 0.05$ in all cases). That is, no effect of residence in an L2 setting was found on L2 stop perception, as all learner groups presented significantly later category boundaries regardless of their experience living in an immersion setting.

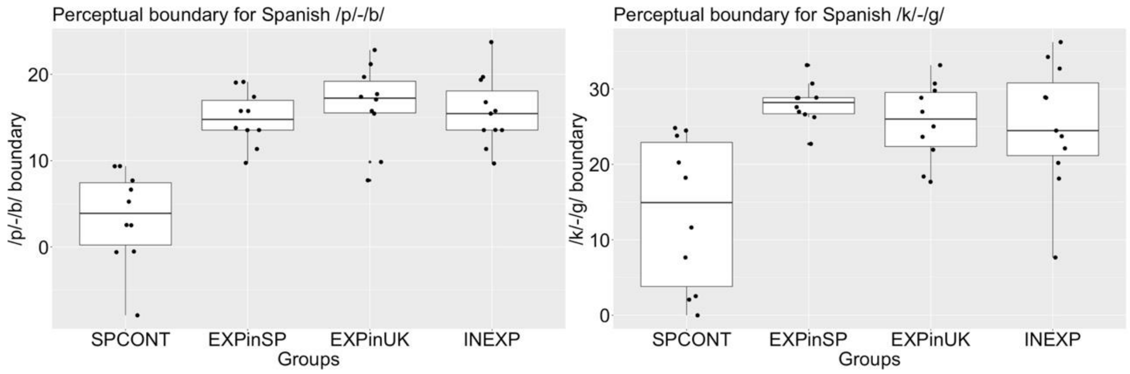

Figure 5. Perceptual category boundaries (in ms) for Spanish /p/-/b/ and /k/-/g/ per group (jitter points illustrate individual data).

Finally, the perception of L2 learners in the L1 and L2 was also compared (see Table 5). On the whole, the learners' category boundaries for both languages were very similar, albeit slightly shorter for Spanish /k/-/g/, and for Spanish /p/-/b/ in the case of EXPinSP. A two-way ANOVA – with category boundary as the dependent variable and language and group as the independent variables – was conducted for each place of articulation. The analysis confirmed the previous results, as no significant effect of group [/p/-/b/: F (2, 56) = 0.489; $p = 0.616$; /k/-/g/: F (2, 56) = 1.021; $p = 0.367$] emerged. In addition, no effect of language was found [/p/-/b/: F (1, 56) = 0.029; $p = 0.865$; /k/-/g/: F (1, 56) = 1.128; $p = 0.293$], as well as no significant interactions between group and language [p/-/b/: F (2, 56) = 0.645; $p = .528$; /k/-/g/: F (2, 56) = 0.031; $p = 0.970$]. These results show that English learners of Spanish perceived the stop contrast similarly in their L1 and L2. In fact, as none of the groups differed from ENCONT in English but all differed from SPCONT in Spanish, it appears that English learners of Spanish perceived L1 and L2 stops according to English values.

Group	**/p/-/b/ boundary**		**/k/-/g/ boundary**	
	English	**Spanish**	**English**	**Spanish**
SPCONT	-	3.4 (5.4)	-	14.8 (10.3)
EXPinSP	16.1 (2.4)	14.9 (3.1)	29.4 (3.7)	28 (2.8)
EXPinUK	15.9 (3.9)	16.5 (4.7)	27.2 (4.5)	25.6 (5.2)
INEXP	14.5 (3.1)	15.7 (4.1)	27.3 (8.5)	25.2 (8.3)

ENCONT	14 (2.5)	-	29.7 (4.3)	-

Table 5. Mean category boundaries (in ms) obtained by all groups in English and in Spanish. Standard deviations are provided in parentheses.

3.3. Discussion

The perception results showed that English learners of Spanish presented L1-like categories in both languages, indicating an influence of the L1 on the L2, whereas no evidence of L2 influence on the L1 was found. Since all groups presented L1-like L2 category boundaries – contrary to the results in the production experiment in this paper –, it appears that there was no effect of L2 experience on L2 perception. While this result runs counter to some previous findings (Bohn, & Flege, 1990; Cabrelli et al., 2019; Gorba, 2019; Major; 2010), a lack of an effect of residence in an L2 setting has also been reported in the literature (e.g., Cebrian, 2006; Flege, et al., 1994). For instance, Flege et al. (1994) evaluated the perception of dissimilarity between L1 and L2 vowels by Spanish learners of English and found that a group of learners who had spent about 7 years in the US did not differ significantly from another group with only 1.8 years of residence. Similarly, Cebrian (2006) observed no effect of prolonged residence in an L2 setting in a study in which Catalan speakers living in Toronto did not outperform Catalan learners of English living in Catalonia in L2 English vowel identification. All L2 learner groups in the current study perceived both L1 and L2 stops through a single L1-like category, showing no CLI effect on the L1. This result is not in agreement with some previous studies that reported an effect of living in an immersion setting on L1 perception (Cebrian, 2006; Dmitrieva, 2019; Gorba, 2019; Major, 2010). Differences between this and previous studies may be related to differences in length of residence and amount of L2 use. For instance, Cebrian (2006) found that Catalan speakers of L2 English living in Canada were not as accurate as Catalan learners of English in Catalonia in their identification of L1 front vowels. In that study, the L2 speakers living in target-language country had spent a considerably longer period of time in that context (a mean of 24 years) than the learners in Spain in the current study (4.2 years). It should be noted that a regression analysis conducted exploring the effect of several individual factors, including number of months in an L2 setting, did not yield any significant results neither in the L1 nor in the L2[8]. Thus, differences in length of residence in the L2 setting may not have been large enough to trigger changes in perception. Other factors may explain the little effect of experience on perception in the current study, as discussed below.

Additionally, and interestingly, the results obtained in the present study do not replicate the outcome of a similar previous perception study involving a mirror-image population, that is, Spanish leaners of English varying in L2 experience (Gorba, 2019). In that study, three groups of Spanish learners of English, with similar characteristics to the ones in the current study, completed the /p/-/b/ identification task described in the current paper. Two instances of an effect of experience were observed. First, the two groups of L1-Spanish L2-English learners with greater L2 experience presented perceptual category boundaries that did not differ from those of the monolingual English speakers. Only the inexperienced group presented significantly earlier – more Spanish-like – category boundaries than English controls in English. Secondly, the group of learners living in an L2 setting had significantly later – i.e., more English-like – category boundaries in their L1 than Spanish controls. That is, an effect of amount of experience, particularly in an L2 setting, was found on both L1 and L2 /p/-/b/ perception. This apparent contradiction with the present paper may be related to differences between the groups tested in Gorba (2019) and in the present study. While the L2 learners in that study and in the current study

[8] Just as in the case of production, a regression analysis (DV: category boundary; IV: months in an L2 setting, years of L2 instruction and dominance score) was conducted. Results did not show a sinificant result neither in the L1 [[F (5, 18) = 0.755; p = 0.549; R^2 = 0.173] nor in the L2 [F (5, 18) = 0.774; p = 0.581; R^2 = 0.177], nor for each individual factor (p > 0.05 in all cases).

had spent comparable amounts of time in their respective target-language countries, cross-study differences exist particularly in amount of L2 instruction and L2 use. A language background and use questionnaire showed that the Spanish learners of English in Gorba (2019) had studied English for a longer period of time and used the L2 to a greater extent than the English learners of Spanish in the current study, especially in the case of the learners in the L2 setting (16.3% weekly use in the current study and 54.6% in Gorba, 2019). In terms of L2 instruction, the group living in an English L2 setting in Gorba (2019) had learnt English for 13.9 years, whereas the group living in the Spanish L2 setting in the current study had only learnt Spanish for 2.4 years. These differences in amount of L2 instruction stem from the fact that English is a mandatory subject in the curriculum of Spanish primary and secondary schools, whereas that is not the case of Spanish in British schools. As for language use, English is more present in the media and in the entertainment industry than Spanish, resulting in a greater use of the L2 on the part of Spanish learners of English than for English learners of Spanish. Moreover, the group of English learners of Spanish living in Spain was made up mostly by English teachers who, therefore, used English almost exclusively at work. Hence, differences in general exposure to and knowledge of English vs. Spanish as a foreign language together with differences in amount of L2 use between the two populations may explain the diverging results of the two studies.

The discrepant results between the present study and Gorba (2019) may also be related to differences in the nature of the Spanish and English VOT contrasts. In English, VOT ranges from voice-lead VOT (phonologically voiced stops) to long-lag VOT (phonologically voiceless stops), and both voice-lead and short-lag VOT are possible in English voiced stop production, being short-lag VOT the most common (Lisker, & Abramson, 1964). By contrast, Spanish VOT ranges from voice-lead to short-lag VOT (Castañeda, 1986). Therefore, Spanish learners of English need to increase their VOT range to categorize L2 English phones, whereas English learners of Spanish already present an adequate VOT range to categorize L1 and L2 stops. As a result, the Spanish learners of English – such as the participants in Gorba (2019) – are exposed to greater VOT differences when acquiring English stops than English learners of Spanish – i.e., participants in the current study – are when learning Spanish. This may have helped the Spanish learners' shift their /p/-/b/ boundary forward as they gained L2 experience.

The lack of an apparent effect of L2 experience on L1 and L2 perception may also be related to an insufficient control of language mode in the experiment (Grosjean, 2001), as a result of testing the two languages with the same stimuli. Recall that the same stimuli were purposely created to test the two languages so as to be able to obtain comparable results, given that the inclusion of a different range for each language may have biased the number of voice-voiceless responses. In principle, language mode was controlled by using response alternatives that were specific to the language that was being tested (e.g., "b as in beetle" or "p as in peeler" as opposed to "b como en bicho" or "p como en pico", see section 3.1.1). In addition, participants completed a reading task and visualized a short video in the target language prior to the perception task. Still, the fact that the same continua were used in both the English and the Spanish identification tests implies that tokens with long-lag VOT – a feature of English but not of Spanish – were present in the Spanish task. This may have resulted in a bias towards a more English-like pattern of identification. We can also speculate that another potential bias towards English may come from a greater lexical interpretability of the sequence /Ci/ that constituted the stimuli of the identification task in English than in Spanish. For instance, the sequence /pi/ corresponds to a more common lexical item in English – "pea"/"pee" – than in Spanish – the mathematical number π, "pi". As for /bi/, both languages have a lexical representation – "bee" in English and "vi", "I saw" in Spanish. Nevertheless, the Spanish word is spelled with <v>, which may have reduced the activation of this lexical item, as the response category was spelled with <b> – i.e., "b as in bicho". As for /ki/ and /gi/, Spanish has no lexical representation for these sequences, whereas English does in the case of /ki/ – i.e., "key". It is unclear, however, how much this factor may influence the listeners as the /Ci/ sequences are also the first syllable in the response keywords.

Regarding the comparison between L1 and L2 perception, the results of the current experiment revealed no significant differences between the L1 and the L2 /p/-/b/ and /k/-/g/

perceptual category boundaries for any learner group. The fact that there were no differences in the perception of the two languages, along with the fact that the English groups differed from SPCONT in their L2 (Spanish) but not from ENCONT in their L1 (English), indicates that the L2 categories were assimilated to the L1, and probably resulted in a single shared perceptual category (Flege, 1995, 2002, 2007). However, this assumption conflicts with the findings of the production experiment, which indicated that English learners of Spanish produced L1 and L2 stops differently – i.e., with greater VOT values in English than in Spanish. This difference may be explained in terms of the possible problems with language mode in the Spanish task alluded to above. The activation of the English phonological system under the two task conditions may have resulted in similar identification responses across tasks. Alternatively, it is possible that the two dimensions require different control mechanisms and that are acquired – at least to a certain extent – independently (De Leeuw et al., 2019). The different results obtained in the two experiments will be discussed in the following section in terms of the relationship between the two modalities.

4. Relationship between perception and production

The results of Experiment 1 and Experiment 2 were compared in order to evaluate the relationship between production and perception of stops in the L1 and in the L2. First, an inspection of each individual's performance in the L1 perception and production tasks was carried out. In other words, an individual's mean VOT values for /p/ and for /k/ were compared to their perceptual category boundary for /p/-/b/ and /k/-/g/, respectively, both in the L1 and in the L2 (see the scatter plots presented in Figures 5 and 6, for the English and Spanish stops, respectively). In the case of the L1, no straightforward relationships were observed as participants that obtained the earliest category boundaries were not the same that presented the lowest /p/ and /k/ VOT means in production. To confirm this observation, correlation analyses between the perception and the production results were conducted for each point of articulation with the data for all three learner groups – a Pearson's correlation test if the data was normally distributed and a Spearman's correlation test if not. None of the tests revealed a significant correlation (/p/ and /p/-/b/ boundary: $r = -0.121$; $N = 31$; $p = 0.518$; /k/ and /k/-/g/ boundary: $\rho = 0.049$, $N = 31$, $p = 0.793$). Just as in the case of the L1, the inspection of each individual's production and perception measures revealed no clear relationships between L2 perception and production for neither place of articulation. Accordingly, the correlation analyses yielded no significant result (/p/ and /p/-/b/ boundary: $\rho = -0.085$, $N = 31$, $p > 0.05$; /k/ and /k/-/g/ boundary: $\rho = -0.039$, $N = 31$, $p > 0.05$)[9].

[9] Voiced stop production data was also collected, although it is not the focus of the current study. An anonymous reviewer suggested that the relationship between category boundary and voiced stop production also be investigated and reported. The results generally replicated the current finding. No significant results were found for English (/p/-/b/ and /b/: $r = -0.300$, $p = .101$; /k/-/g/ and /g/: $r = .161$, $p = 0.388$), nor for Spanish bilabial stops ($r = -0.240$, $p = .193$) and only a moderate correlation was revealed for velar stops ($r = 0.362$, $p = 0.049$).

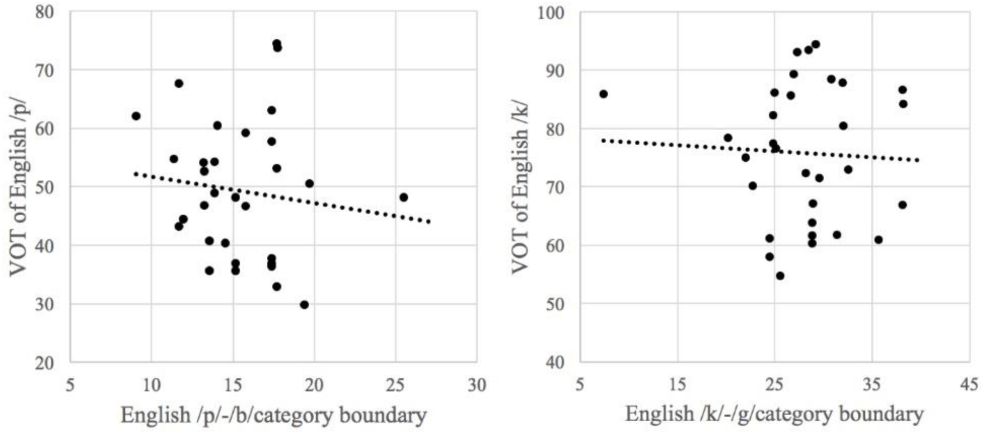

Figure 6. Scatter plots of the production and perception data for the English bilabial and velar stops.

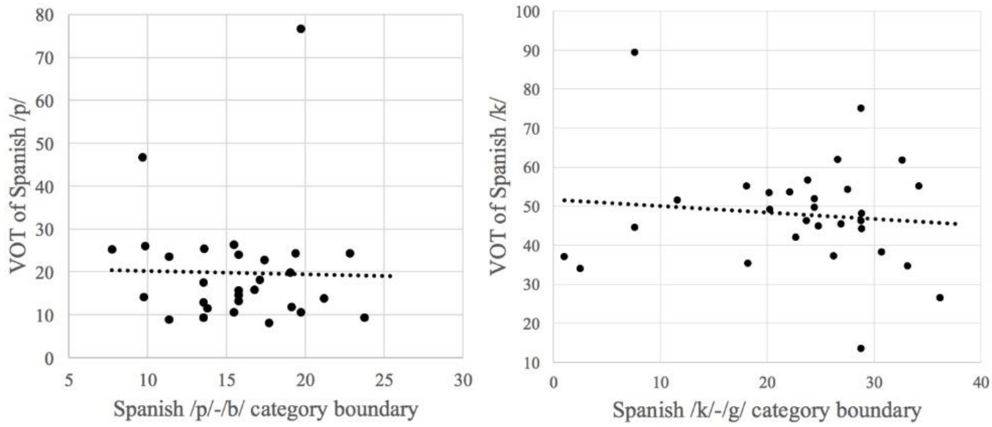

Figure 7. Scatter plots of the production and perception data for the Spanish bilabial and velar stops.

4.1. Discussion

The mean VOT productions of English learners of Spanish was compared to the category boundaries obtained for the bilabial and the velar contrasts, both in the L1 and in the L2. In line with some previous studies, no significant correlation was found between VOT perception and production, neither in the L1 (e.g., Bailey, & Haggard, 1973, Schultz et al., 2012) nor in the L2 (e.g., Bailey, & Haggard 1973; Gorba, 2016; Sheldon, & Strange, 1982; Williams, 1977). This finding is against the assumptions of some theoretical models of both L1 – such as the Motor Theory (Liberman, & Mattingly, 1985) and the Direct-Realist approach (Fowler, 1986, 1990) –, and L2 speech – like the SLM (Flege, 1995, 2002; but cf. Flege and Bohn's (2021) SLM-r) or the PAM (Best, 1995; Best, &, Tyler, 2007) – which assume that there is a link between the two dimensions. More specifically, the SLM and the PAM favour the view that perception generally leads production in L2 speech learning and some previous studies support this view (e.g., Bohn, & Flege, 1990; Nagle, 2019). The results of the current experiments actually suggest the opposite, since English learners of Spanish performed in a more target-like manner in production than in perception. Therefore, the lack of a straightforward link between the two dimensions could be explained by the fact that participants tended to perceive both L1 and L2 stops according to English values, whereas they made a difference between the two languages in production – i.e., they presented significantly longer VOT values in English production than in Spanish. It is thus possible that L2 speakers may succeed in learning to produce VOT with target-language values even if they continue to perceive the contrast in terms of the L1 distinction. This disassociation between perception and production has been found in some previous studies. Flege and Eefting

(1987a) reported that Dutch learners of English could produce the /t/-/d/ contrast similarly to English native speakers but failed at discriminating it. Similarly, Trofimovich and John (2011) found that Canadian French learners of English were capable of producing the English /t/-/θ/ and /d/-/ð/ contrasts even if they seemed unable to distinguish the dental and the alveolar obstruents perceptually. Llisterri (1995) suggests that it may not always be possible to predict production accuracy from perceptual abilities and vice-versa due to a variety of factors including the acoustic and perceptual characteristics of the sounds under study, the learners' characteristics, or the methodologies used.

In fact, the lack of an alignment between perception and production in the current study may be related to methodological differences concerning the measures used to assess each dimension and the extent to which they are comparable (Mack, 1989). For example, comparisons between measures may be complicated by differences in variability. Specifically, the measure used for production (mean VOT in ms) represents a relatively variable measure as speakers may use different amounts of VOT in different productions even of the same word. By contrast, the perception measure (category boundary), calculated based on the participants' reactions to a VOT continuum, is less likely to vary if listeners were to perform the identification task repeatedly. Yet, the mean of a more variable production measure was contrasted with the mean of a less variable perception measure. It should also be noted that the task used to assess perception was an identification test. Previous research suggests that a clearer relationship between the two dimensions may be found using a goodness rating task than an identification task, possibly because the former provides more specific information of what constitutes a representative token of a given category than the latter (e.g., Newman, 2003 vs. Shultz et al., 2012). In addition, as discussed above, the different results for perception and production may stem from an insufficient control of language mode activation in the perception task. The inclusion of English-sounding stimuli – i.e., from the long-lag VOT end of the continuum – in the Spanish test may have resulted in the interference of the L1 during the L2 tasks. In other words, following the language mode hypothesis (Grosjean, 2001), English could have been activated during the completion of the Spanish task, resulting in an L1-like perception of Spanish stops. Conversely, the production task may have been more successful at triggering a monolingual language mode in the target language, as participants were only exposed to instructions and sentences in the target language.

Finally, regarding the potential effect of L2 experience on the relationship between perception and production, no clear effect can be established. Given that all groups, regardless of L2 experience, lacked a clear link between the two modalities, it cannot be claimed that L2 experience modulated this relationship. Again, differences in the nature of the perceptual and production measures, as well as differences between the production and perception tasks in the efficacy of establishing stable language modes may obscure the potential effect of experience on the relationship between perception and production.

5. General discussion

The goal of this paper was to investigate the effect of L2 experience on the production and perception of L1 and L2 bilabial and velar voiceless stops. Thirty-two English learners of Spanish differing in amount and type of L2 experience and two groups of monolingual speakers were evaluated. The L2 Spanish learners included two groups of learners in an instructional setting in the UK (a group of advanced fourth-year undergraduate students of Spanish with some previous experience in an immersion setting (EXPinUK) and a group of first and second-year undergraduates with no previous experience in the L2 setting (INEXP)), as well as a group of English residents in Spain (EXPinSP). The participants produced word-initial voiceless stops in their L1 and in their L2 and completed a series of identification tasks involving stimuli from a VOT continuum, also in each language. Further, the relationship between the two dimensions – perception and production – was explored, both in the L1 and in the L2, a relationship that has yielded conflicting results in the literature. One of the main contributions of this paper lies in the fact that it investigates the effect of amount and type of L2 experience not only on VOT production, like previous studies such as Flege (1987), but also on VOT perception, which has received scarce attention in the literature, particularly regarding L2 influence on the L1. In fact, only a few studies have assessed the effect of L2 experience on L1 perception, and, to our knowledge, only Gorba (2019) investigated stops. In addition, the methodological approach of this paper also entails contrasting learners' performance in their L1 and their L2, and comparing it to that of monoligual speakers of each language.

The results of the first experiment showed a positive effect of L2 experience on L2 production, in agreement with previous studies (e.g., Flege, 1987; Stevens, 2001). The two groups with greater L2 experience, EXPinUK and EXPinSP, outperformed the more inexperienced learners in the production of L2 stops – i.e., the former tended to produce L2 stops with target-like VOT values, whereas the latter produced stops with longer VOT than Spanish controls. Furthermore, a greater number of L2 speakers differed from Spanish monolinguals in the case of the group with no experience in an L2 setting than in the case of the experienced groups. Thus, a greater amount of experience in terms of years of learning and residence in the L2 setting resulted in a more target-like VOT production in the L2. These results are in agreement with many previous studies showing a direct link between L2 experience and L2 learning (e.g., Bohn, & Flege, 1990; Flege, 1987; Flege, Bohn, & Jang, 1997; Lev-Ari, & Peperkamp, 2014; Levy, & Law, 2010; Munro, 1993; Stevens, 2001) and underscore the expected positive effect of prolonged L2 exposure posited by L2 speech models such as the SLM (Flege, 1995, 2002, 2007). However, support for a positive effect of L2 experience is weakened by the results of the second experiment: all three learner groups presented English-like perceptual category boundaries for Spanish stops. Fewer previous studies have examined the effect of L2 experience on L2 perception, as opposed to L2 production, and among these there is evidence of both a positive effect (e.g., Bohn, & Flege, 1990; Gorba, 2019), and of a lack of effect (e.g., Cebrian, 2006; Flege et al., 1994). Gorba (2019), a parallel perception study involving Spanish learners of English, reported that L2 experience enhanced perception in a more target-like manner. Differences in amount of L2 use and L2 instruction between the current study and Gorba (2019) may point to possible reasons for the discrepant results, as discussed next. Moreover, cross-linguistic differences related to the range of VOT used in English and in Spanish may also account for cross-study divergences, as will be explained below.

On the other hand, L2 experience was not found to influence the L1, neither in the case of production nor in perception, as all participants perceived and produced L1 stops according to English VOT values and, thus, did not differ from English monolinguals. Although relatively few previous studies have examined the effect that learning an L2 has on the L1, some evidence of L1 drift towards the L2 in L1 production (Flege, 1987) and in L1 perception (Gorba, 2019) after a long period of residence in the L2 setting have been reported[10]. Again, it is possible that the absence of an effect in the current study may be related to differences between this and previous studies in the populations tested. The L2 learners in an immersion setting examined by Flege (1987) had spent a longer period of time in an L2 setting than the most experienced group in the current study (12 years and 4.2 years, respectively). The Spanish learners of English residing in an L2 setting in Gorba (2019), for whom an effect of L2 experience on the perception of both L1 and L2 stops was observed, were comparable to the English learners of Spanish in this study in terms of time spent in the L2 setting. However, the two populations differed in amount of instruction and of L2 use. The Spanish learners of English in Gorba (2019) had received more formal instruction in the L2 (13.9 years vs. 2.4 years in the current study) and used the L2 more often (reported weekly use of 54.6% vs. 16%). This is in line with Piske et al.'s (2001) finding that, after starting age of learning, language use was the second most important factor affecting L2 development, and with Flege and Bohn's (2021) claim that formation of target-like phonetic categories is directly related to the quantity and quality of L2 input. Therefore, in addition to, or together with, L2 experience, other factors such as amount of instruction and L2 use play a role in the development of target-like L2 contrasts and the consequent effect that learning an L2 may

[10] See section 1.2 in the introduction for cases of L1 phonetic drift towards L2 at very early stages of learning (Chang, 2012; Kartushina , Hervais-Adelman et al., 2016).

have on the existing L1. In other words, the comparatively low L2 use on the part of EXPinSP may explain the absence of greater CLI in perception.

In addition, the different outcome in the present study and in Gorba (2019) may also be related to differences between the learners' L1, more specifically, to the different range of VOT used in English and in Spanish. The range of English is greater than that of Spanish, as English VOT ranges from voice-lead VOT to long-lag VOT and, in the case of Spanish, from voice-lead to short-lag VOT. Thus, a Spanish speaker's acquisition of the English stop voicing contrast involves the extension – towards greater values – of the VOT range. This may result in the L2 learners' moving forward their category boundary for voiced-voiceless stops, as it was the case in Gorba (2019). By contrast, acquiring Spanish stops does not involve an extension of the VOT range for English speakers, whose L1's VOT already includes prevoicing and short-lag – like in Spanish – as well as aspiration – unlike Spanish. Therefore, a displacement of their category boundary may not be triggered by acquiring Spanish.

It should also be noted that no differences were observed either in L1/L2 perception or in L1/L2 production between EXPinUK and EXPinSP despite EXPinSP's residence in Spain at the time of the experiment and notably longer stay (50.7 months vs. 9.4 months for the EXPinUK). In addition, the experienced group in the UK had received a greater amount of L2 instruction (7.5 years vs. 2.4 years), and were studying Spanish at the time of testing, unlike the EXPinSP learners. Moreover, despite living in different settings, the groups did not differ greatly in how much they used the L2. In other words, living in the L2 country did not result in a much greater L2 use on the part of the EXPinSP than on the part of EXPinUK (with a percentage of weekly L2 use of 16.3% and 15%, respectively). This is probably related to the fact that the experienced group in the UK were students enrolled in Spanish courses while the experienced group in Spain were English language teachers who used mostly English at work and, in most cases, also at home. Recall that a regression analysis did not yield a significant effect of either L2 instruction or L2 use on L2 perception and production (see footnotes 5 and 8). The regression analysis did reveal an effect of amount of L2 use and L2 instruction on L1 production, that is, a lower L2 use and smaller amount of instruction was related to longer VOT productions in English. It is possible, thus, that the amount of years of exposure to Spanish in Spain for EXPinSP and the combination of years of learning in the UK and a previous 9-month stay in Spain for EXPinUK may have been enough to influence the learners' production of L2 stops but the predominant use of their L1 in both cases may explain the absence of CLI affecting the L1. This finding supports the SLM-r's idea that is not just experience but its combination with L2 use that has the greatest impact on the learner's L2 – and L1. It is perhaps due to the similar L2 use across groups that differences between them were relatively small, particularly regarding the L1.

This paper also examined whether learners were capable of categorizing L1 and L2 stops differently. It was found that all English learners of Spanish produced English voiceless stops with significantly longer VOT values in English than in Spanish – although this difference was greater in the case of the two most experienced groups. In terms of perception, however, L2 speakers perceived L2 and L1 stops similarly – i.e., with English VOT values – regardless of L2 experience. The fact that their L1 and L2 categories are produced differently (i.e., with language-specific values) whereas voiceless stops are perceived similarly in the two languages may simply indicate that there is not necessarily a straightforward relationship between the perception and production of a VOT contrast. Particularly, the fact that accurate perception is not a requisite for accurate production lends little support to the predictions of some L2 speech models (Flege, 1995; 2002, 2007; Best, 1995; Best, & Tyler 2007), and may indicate that L2 perception and production are ruled by different mechanisms (De Leeuw et al., 2019; Hattori, & Iverson, 2010). In fact, it is possible that the L2 learners have a single category for L1 and L2 stops but that a distinction in production is achieved by the same mechanisms by which monolingual speakers realize some L1 categories differently depending on the social context (Flege, 1987; Labov, 1981). On the other hand, the different results for perception and production could have a methodological basis, as discussed below. In addition, the perception of the L2 speakers' production by native Spanish speakers has not been assessed. It is possible that, despite the significant differences in production, the L1 and L2 stops may not be perceived as different by native speakers. Still, in most cases, the learners' L2 Spanish productions were much closer to the monolingual Spanish values than to the monolingual English values, particularly for EXPinUK and EXPinSP, while the learners' L1 English productions closely resembled English monolingual values, suggesting a relatively stable distinction. Further research involving a judgement or rating task performed by monolingual Spanish speakers could be carried out so as to confirm the strength of the acoustic differences.

To summarize the findings concerning the relationship between L2 experience and cross-linguistic influence, it appears that phonetic drift occurred in the direction of the L1. The L1 influenced the L2, especially in the case of the group with no experience in an L2 setting – who produced significantly longer VOT values than Spanish controls. Moreover, all groups presented influence of English in their L2 perception, as they categorized Spanish stops with English-like values. As for the L1, no instances of phonetic drift towards the L2 were found, since groups produced native-like VOT values in their L1 and categorized stops similarly to English monolinguals. Regarding category formation, English learners of Spanish produced /p/ and /k/ differently in their L1 and L2, whereas they used the same English-like categories for both languages in perception.

The results of the current study can be interpreted following both the SLM's and the L2LP's premises. If we consider the production results first, it may appear that English learners of Spanish presented separate categories coexisting in the same phonetic space, in SLM/SLM-r terms, for voiceless stops in each language. However, the results of the perception experiment revealed no difference between L1 and L2 perception, given the use of L1-like categories in both languages. Although the learners' perceptual assimilation of Spanish stops to native English categories was not examined, we may speculate from the perception results that the Spanish stops were mapped onto the L1 categories. According to the SLM, under these circumstances a shared merged category may be used in the two languages. The merged category may then show more L1-like or L2-like features according to the input received. In fact, this category was implemented differently in perception (no distinction) and in production (distinct L1 and L2 productions). As mentioned above, it is possible that the distinct L1 and L2 realizations result from different phonetic implementations of the same shared L1-L2 category, as is attested for social context-dependent realizations in monolingual speech (Flege, 1987; Labov, 1981). On the other hand, the difference between perception and production may be linked to methodological issues, such as the greater control of the language mode in the production task than in the perception task, as discussed below.

Alternatively, the same results can be interpreted in terms of the L2LP (Escudero, 2005, 2009). Even though the L2LP addresses perception, the fact that participants performed differently in the Spanish and the English production task can also support the model's claim that the L1 and the L2 are separate systems. Moreover, in agreement with L2LP's predictions, L2 learners' categories for Spanish voiceless stops seem to be developing towards more target-like VOT values, as the groups that had lived in an immersion setting produced L2 stops more accurately than the learners with no prolonged experience in an L2 setting. The fact that L2 categories evolve as a function of L2 experience but L1 categories remain unchanged may also lend further support to the L2LP's claim that the two systems are independent and that L2 learning occurs without influencing the L1. If language mode was not sufficiently controlled in the Spanish task – that is, the presence of aspirated stimuli in the two languages and the greater lexical load of the stimuli in English could have activated an English mode in the Spanish categorization test –, the L2LP would predict that the English phonological system was activated, resulting in English-like responses. Thus, it appears that both models can explain the outcome of this experiment.

In brief, in spite of the fact that the SLM claims that the two systems coexist in a common phonetic space, whereas the L2LP argues that the two systems are separate, both models posit that L2 categories can improve and evolve towards target-like values. Moreover, variability in the performance of L2 learners depending on the nature of the input received is expected in both models – by means of a merged category according to the SLM and due to a parallel activation of the two systems in L2LP's terms. Further research where the L1 and the L2 modes are controlled is needed to test the current L2 acquisition models. Inducing both a monolingual and bilingual mode is necessary in order to test whether CLI is bidirectional – as the SLM claims – or whether the two systems are separate – as the L2LP posits. Ideally, highly proficient learners – that is, speakers with native-like L1 and L2 systems – should also be tested, given that a learner's category may still be developing and can present L1 features, as it is believed to be initially a copy of the L1 (Escudero, 2005, 2009).

The relationship between perception and production was analyzed by comparing each participant's category boundary with their mean VOT productions for each stop. Correlation analyses involving the perceptual and production data did not yield significant results, suggesting the lack of a close relationship – or a complex relationship – between the two modalities, both in the L1 and in the L2. This finding challenges the Motor Theory (Liberman, & Mattingly, 1985) and the Direct-Realist approach's (Fowler, 1986, 1990) claim that there is a straightforward link between the two modalities in the L1, as well as the assumptions of Flege's (1995, 2002) SLM and Best's (1995, Best, & Tyler, 2007) PAM, that L2 perception precedes production. In fact, some previous studies have found that L2 learners can produce L2 phones more accurately than they can perceive them (Flege, & Eefting, 1987; Flege, & Bohn, 1997; Trofimovich, & John, 2011). Trofimovich and John (2011) found that French Canadian learners of English were able to produce /θ/ and /ð/ accurately, but priming effects found in a lexical decision task showed difficulty in perceiving the /θ/-/ð/ contrasts. Like in the current study, no relationship between production and perception accuracy could be established. It has been pointed out that the link between perception and production may be affected by a variety of issues such as L2 proficiency, type of target structure, L1-L2 similarity, language differences in cue weighting for similar contrasts, greater detectability of production accuracy, and comparability of perception and production tasks and measures (Llisterri, 1995; Mack, 1989). Furthermore, the apparent lack of a perception-production link may result from the amount of individual variabilty reported in several studies (i.e., some participants may be more successful in perceiving L2 stops accurately than in producing them, whereas the opposite may be true for others; e.g., De Leeuw et al., 2019; Sheldon, & Strange, 1982). It should also be noted that, in fact, the most recent theoretical claims assume that, even though the two modalities may coevolve, the alignment between the two dimensions may never be exact (SLM-r, Flege, & Bohn, 2021).

The lack of a clear link between the performance of the L2 learners in the two modalities may also be related to methodological factors. It is possible that the perception and the production measures used in this study were, in fact, not fully comparable, as the production measure – amount of VOT production – allowed for greater variability than the perception measure – identification along a VOT continuum. Furthermore, the production task included full sentences (target words in carrier phrases), whereas the stimuli used in perception were isolated syllables. Thus, the production task had a greater lexical load than the perception task, which may have facilitated the activation of the intended language mode and resulted in a more target-like performance in the L2. In addition, the same perceptual stimuli – some of which presented aspiration, a feature of English not present in Spanish – were used to test the two languages, which may have also contributed to the activation of English.

The methodological differences between the perception and production tasks, albeit partly inevitable in a laboratory study, constitute a limitation of the current paper. Another consequence of the laboratory nature of this study is that it used a controlled method of data collection. For instance, differences in L1 and L2 production were observed, but different results may have been found if instead of using a reading task a more naturalistic task was utilized (e.g., spontaneous speech, map tasks). It remains to be seen if the L2 learners would still make a L1-L2 distinction in more spontaneous speech. Similarly, the use of a goodness rating task in perception may have also resulted in a different outcome, as previous studies using this type of task have found a clearer relationship between the two dimensions (e.g., Newman, 2003). Moreover, despite the overall large amount of data analyzed (production and perception data from 51 participants in two languages and for two places of articulation), the number of individuals per group – between nine and eleven participants– may not have been enough to capture consistent patterns of L2 acquisition by English learners of Spanish. It is possible that stronger and more consistent tendencies may have been observed given a greater sample size. Difficulties in locating participants with similar characteristics, level of experience, and linguistic background for each experience group prevented the recruitment of a larger pool of participants. In fact, it was also difficult to obtain groups that were more different in terms of other variables like language use. As mentioned above, L1 English speakers may be able to function in a foreign setting, particularly in a large cosmopolitan city, with a predominant use of their L1. These limitations call for more controlled carefully designed studies as well as pointing to the need for more naturalistic data, and open avenues for further research.

6. Conclusions

The present study has investigated the L1 and L2 perception and production of voiceless stops by English learners of Spanish differing in amount and type of L2 experience. Participants completed a perception and a production experiment in each language. The production experiment involved a sentence reading task including the target phones in initial position. As for perception, participants had to complete two identification tasks – one for /p/-/b/ and one for /k/-/g/. The production experiment revealed that English learners of Spanish produced L1 and L2 voiceless stops differently. Further, L2 production became more target-like with increased length of residence in the target-language setting and greater number of years of learning. Conversely, no effect of L2 experience was found on L1 production, given that all groups produced English voiceless stops with native-like VOT values. As for perception, all groups of English learners of Spanish perceived L1 and L2 stops with English-like VOT values, regardless of L2 experience. That is, no effect of L2 experience was found either on L1 or L2 stop perception. These findings show that accuracy in L2 production is possible despite failure to perceive in an L2 fashion. In addition, no clear relationship was found between the perception and production of stops either in the L1 or in the L2: participants who produced longer VOT values did not necessarily need longer VOT to identify voiceless stops. These results may partly be related to the difficulty of inferring production abilities from perceptual performance, and the lack of comparability across measures used to assess the two modalities. Additionally, methodological limitations may also explain some of the results, such as a greater control of language mode in the production task than in the perception task. Still, the differences between the least experienced group and the two experienced groups underscore the impact of L2 experience, either in terms of number of years of learning or of residence in an L2 setting, on L2 speech learning. In addition, even a smaller amount of experience in the case of the least experienced groups was able to promote a distinction in L1 and L2 production, highlighting the role of experience. Further research involving larger samples, more naturalistic data and greater control for unilingual and bilingual mode activation may help us understand better the nature and directionality of phonetic drift and the relationship between perception and production.

Acknowledgements

This work was supported by research grants from the Spanish Ministry of Economy and Competitiveness [Grant No. FFI2017-88016-P] and from the Catalan Government [Grant No. 2017SGR34].

Appendix A. List of sentences used in the production tasks

A.1. English production task

Read each of the following sentences twice:

1. Kitten is the next word
2. Peeler is the next word
3. Beaches is the next word
4. Watches is the next word
5. Gearbox is the next word
6. Cellphone is the next word
7. Keychain is the next word
8. Peaceful is the next word
9. Tiger is the next word
10. Pieces is the next word
11. Houses is the next word
12. Beefcake is the next word
13. Keenly is the next word
14. Girlfriend is the next word
15. Keeper is the next word
16. Music is the next word
17. Geezer is the next word
18. Razor is the next word
19. Beetle is the next word
20. Flawless is the next word
21. Peanut is the next word
22. Geekfest is the next word
23. Flipflop is the next word
24. Beetroot is the next word
25. Beating is the next word
26. Headphones is the next word
27. Keener is the next word
28. Geeky is the next word
29. Ancient is the next word
30. Peacock is the next word
31. Gearshift is the next word
32. Keyhole is the next word

A. 2. Spanish production task

Lee cada una de las siguientes frases dos veces:

1. Monte es la siguiente palabra
2. Pila es la siguiente palabra
3. Hombre es la siguiente palabra
4. Quicio es la siguiente palabra
5. Bizco es la siguiente palabra
6. Guinda es la siguiente palabra
7. Piso es la siguiente palabra
8. Mono es la siguiente palabra
9. Bicho es la siguiente palabra
10. Guiso es la siguiente palabra
11. Rata es la siguiente palabra
12. Birla es la siguiente palabra
13. Ante es la siguiente palabra
14. Quinta es la siguiente palabra
15. Orden es la siguiente palabra
16. Guía es la siguiente palabra
17. Pista es la siguiente palabra
18. Perro es la siguiente palabra
19. Biblia es la siguiente palabra
20. Guiño es la siguiente palabra
21. Uña es la siguiente palabra
22. Quita es la siguiente palabra
23. Uso es la siguiente palabra
24. Bici es la siguiente palabra
25. Móvil es la siguiente palabra
26. Quince es la siguiente palabra
27. Pico es la siguiente palabra
28. Cama es la siguiente palabra
29. Quise es la siguiente palabra
30. Susto es la siguiente palabra
31. Guita es la siguiente palabra
32. Pino es la siguiente palabra

Appendix B - Coefficient estimates, standard errors and t-values for each model parameter conducted on the production results

B.1. First analysis, L1 production, effect of group (three learner groups and the English control group). Estimates of fixed effects.

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	51.709	4.104	33.96	12.598	< 0.001
[Group=ENCONT]	0.396	5.580	36	0.071	0.944
[Group=EXPinSP]	-5.825	5.294	36	-1.1	0.279
[Group=EXPinUK]	-1.272	5.580	36	-0.228	0.821

Table B1. Coefficient estimates, standard errors and t-values for L1 production and effect of group (three learner groups and the English control group) for the production of English /p/. Reference group=INEXP.

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	79.295	4.451	28.758	17.814	< 0.001
[Group=ENCONT]	-7.544	5.751	36	-1.312	0.198
[Group=EXPinSP]	-7.001	5.456	36	-1.283	0.208
[Group=EXPinUK]	-3.527	5.751	36	-0.613	0.544

Table B2. Coefficient estimates, standard errors and t-values for L1 production and effect of group (three learner groups and the English control group) for the production of English /k/. Reference group=INEXP.

B.2. Second analysis, L2 production, effect of group (three learner groups and the Spanish control group). Estimates of fixed effects.

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	27.563	3.397	39.311	8.113	< 0.001
[Group=EXPinSP]	-13.835	4.576	37.031	-3.023	0.005
[Group=SPCONT]	-19.000	4.689	37.029	-4.052	< 0.001
[Group=EXPinUK]	-10.652	4.824	37.028	-2.208	0.033

Table B3. Coefficient estimates, standard errors and t-values for L2 production and effect of group (three learner groups and the Spanish control group) for the production of Spanish /p/. Reference group=INEXP.

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	55.433	3.324	39.189	16.675	< 0.001
[Group=EXPinSP]	-9.903	4.576	37	-2.164	0.037
[Group=SPCONT]	-21.544	4.811	39.172	-4.478	< 0.001
[Group=EXPinUK]	-10.498	4.824	37	-2.176	0.036

Table B4. Coefficient estimates, standard errors and t-values for L2 production and effect of group (three learner groups and the Spanish control group) for the production of Spanish /k/. Reference group=INEXP.

B.3. Third analysis, effects of group (three learner groups) and language (L1, L2). Estimates of fixed effects

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	27.5778	3.276	38.201	8.417	< 0.001
[Group=EXPinSP]	-13.849	4.275	35.501	-3.239	0.003
[Group=EXPinUK]	-10.667	4.506	35.494	-2.367	0.024
[language=English]	24.131	2.693	19.693	8.959	< 0.001
[Group=EXPinSP] * [language=English]	8.025	2.854	577.047	2.811	0.005
[Group=EXPinUK] * [language=English]	9.395	3.008	577.044	3.123	0.002

Table B5. Coefficient estimates, standard errors and t-values for production and effect of group (three learner groups and the Spanish control group) and language (English and Spanish) for the production of /p/. Dependent Variable: VOT. Reference group=INEXP, language=Spanish.

Parameter	Estimate	Std. Error	df	t	Sig.
Intercept	55.432	3.609	33.941	15.358	< 0.001
[Group=EXPinSP]	-9.903	4.285	35.818	-2.311	0.027
[Group=EXPinUK]	-10.497	4.517	35.818	-2.324	0.026
[language=English]	23.861	3.458	13.485	6.9	< 0.001
[Group=EXPinSP] * [language=English]	2.902	2.921	578	0.993	0.321
[Group=EXPinUK] * [language=English]	6.971	3.079	578	2.264	0.024

Table B6. Coefficient estimates, standard errors and t-values for production and effect of group (three learner groups and the Spanish control group) and language (English and Spanish) for the production of /k/. Dependent Variable: VOT. Reference group=INEXP, language=Spanish.

Appendix C. Identification functions per group, language and place of articulation.

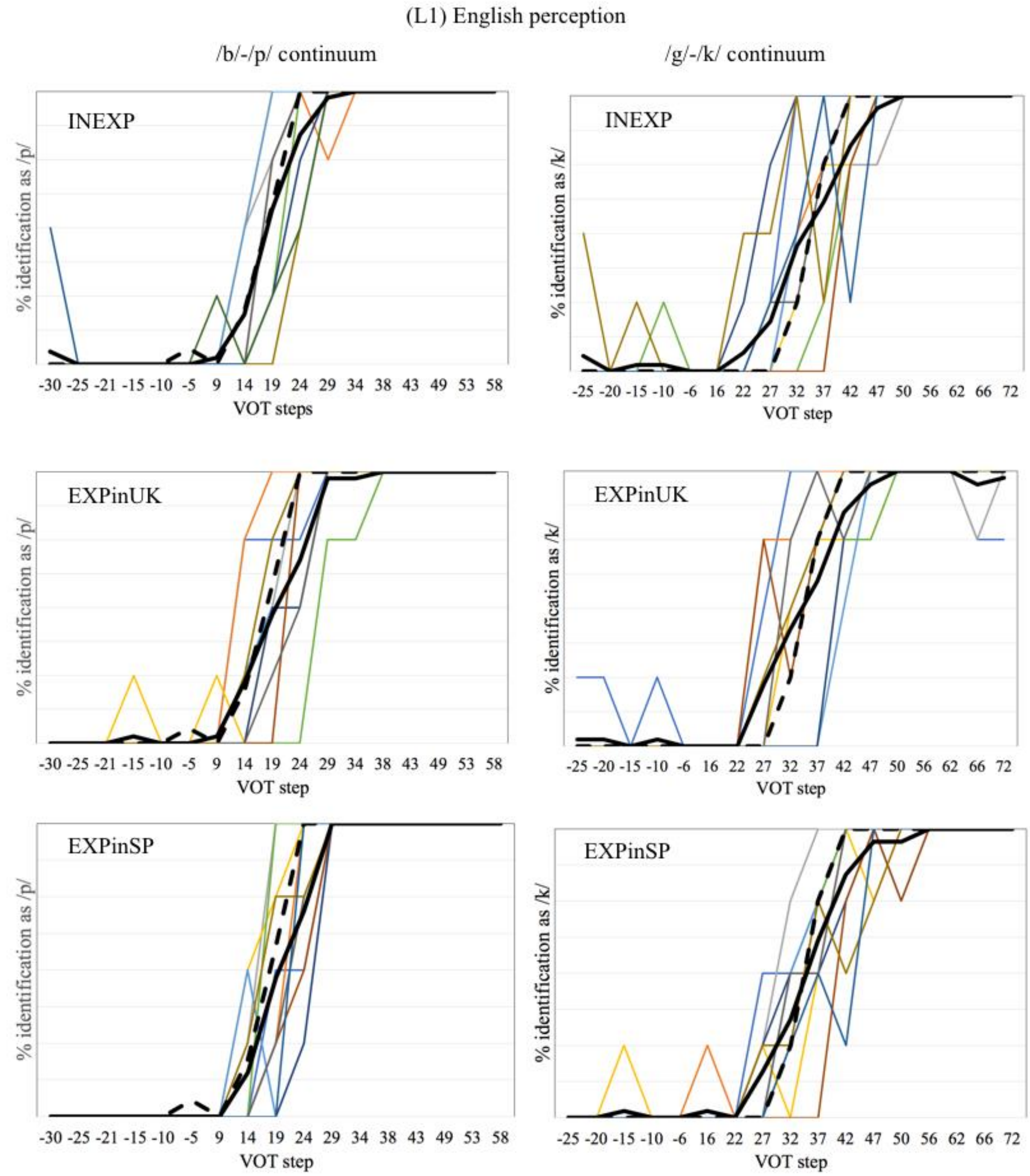

Figure C6. Identification functions obtained in the English identificatio tests by each individual in the three learner groups. Responses for bilabials are shown on the left, for velars on the right. The y axis shows each step in the VOT continuum (see section 3.1.2 for details) and the x axis indicates the % identification as voiceless stop. A dark thick line represents the mean identification function for each learner group, and a dashed line indicates the mean for the English monolinguals (ENCONT).

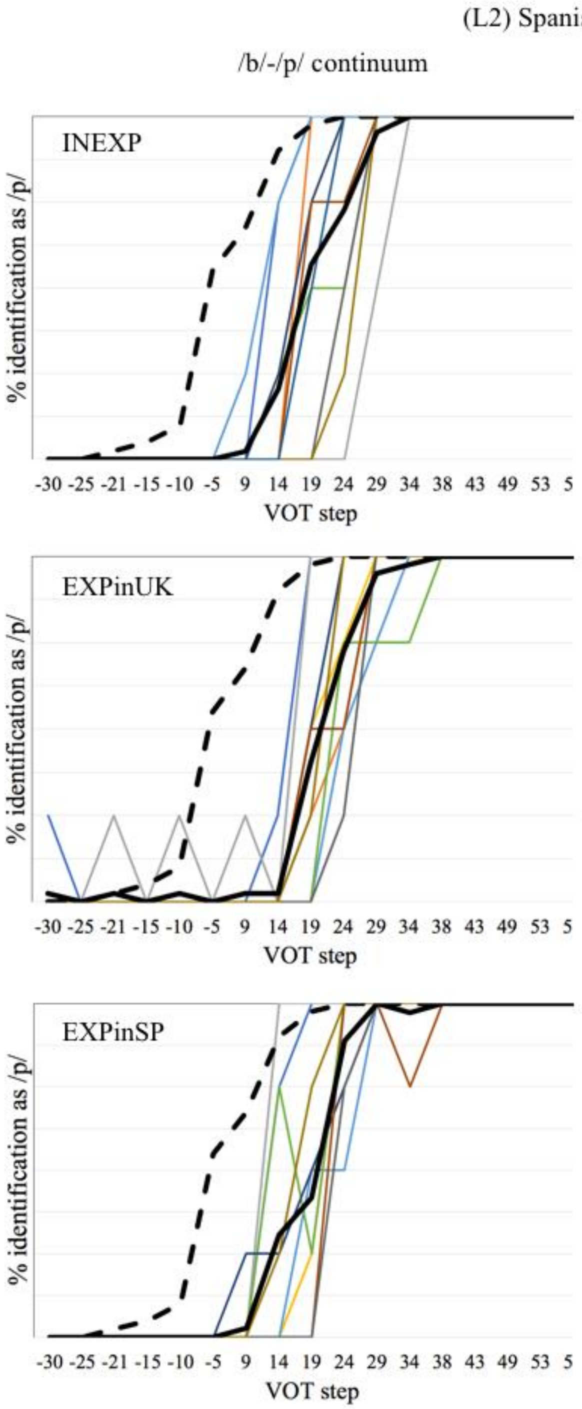

Figure C7. Identification functions obtained in the Spanish identification tests by each individual in the three learner groups. Responses for bilabials are shown on the left, for velars on the right. The y axis shows each step in the VOT continuum (see section 3.1.2 for details) and the x axis indicates the % identification as voiceless stop. A dark thick line represents the mean identification function for each learner group, and a dashed line indicates the mean for the Spanish monolinguals (SPCONT).

Appendix D. Supplementary data

Supplementary data to this article can be found online at

https://data.mendeley.com/datasets/fz4wsdtpvz/draft?a=1d95bf49-061e-43b9-97d3-88be9c3c4feb.

References

Abramson, A. S., & Lisker, L. (1973). Voice-Timing Perception in Spanish Word-Initial Stops. *Journal of Phonetics, 1*(1), 1–8. https://doi.org/https://doi.org/10.1016/S0095-4470(19)31372-5

Ahn, S., Chang, C. B., DeKeyser, R., & Lee‐Ellis, S. (2017). Age effects in first language attrition: Speech perception by Korean‐English bilinguals. *Language Learning*, *67*(3), 694-733. https://doi.org/10.1111/lang.12252

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, and B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production* (pp. 2–31). Newcastle upon Tyne, UK: Cambridge Scholars Publishing

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, *38*(4), 640–653. https://doi.org/10.1016/j.wocn.2010.09.005

Aoyama, K., & Flege, J. E. (2011). Effects of L2 experience on perception of English /r/ and /l/ by native Japanese speakers. Jo*urnal of the Phonetic Society of Japan*, *15*(3), 5-13.

Bailey, P. J., & Haggard, M. P. (1973). Perception and production: Some correlations on voicing of an initial stop. *Language and Speech,16*(3),189–195.

Bergmann, C., Nota, A., Sprenger, S. A., & Schmid, M. S. (2016). L2 immersion causes non-native-like L1 pronunciation in German attriters. *Journal of Phonetics*, *58*, 71-86.

Best, C. T. (1995). A Direct Realist View of Cross-Language speech perception. Speech Perception and Linguistic Experience. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 167–200). Timonium, MD: York Press.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Birdsong, D., Gertken, L. M., & Amengual, M. (2012). Bilingual language profile: An easy-to-use instrument to assess bilingualism. *COERLL, University of Texas at Austin*.

Boersma, P., & Weenink, D. (2016). *Praat: Doing Phonrtics by computer [Computer software]*. Version 6.0.19. Retrieved Februray 2016 from http://www.praat.org/.

Bohn, O. S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, *11*(3), 303–328. https://doi.org/10.1017/S0142716400008912

Brunner, J., Ghosh, S., Hoole, P., Matthies, M., Tiede, M. & Perkell, J. (2011). The Influence of Auditory Acuity on Acoustic Variability and the Use of Motor Equivalence During Adaptation to a Perturbation. *Journal of Speech, Language, and Hearing Research*, *3*, 727–739. https://doi.org/10.1044/1092-4388(2010/09-0256)

Cabrelli, J., Luque, A., & Finestrat-Martínez, I. (2019). Influence of L2 English phonotactics in L1 Brazilian Portuguese illusory vowel perception. *Journal of Phonetics*, *73*, 55–69. https://doi.org/10.1016/J.WOCN.2018.10.006

Caramazza, A., Yeni‐Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The Acquisition of a New Phonological Contrast: The Case of Stop Consonants in French-English Bilinguals. *Journal of the Acoustical Society of America*, *54*(2), 421–428.

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism*, *19*(5), 939–954. https://doi.org/10.1017/S1366728915000334

Casillas, J. (2020). Phonetic Category Formation is Perceptually Driven During the Early Stages of Adult L2 Development. *Language and Speech*, *63*(3). https://doi.org/10.1177/0023830919866225

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. https://doi.org/10.1016/j.wocn.2018.07.002

Castañeda, M. L. (1986). El VOT de las oclusivas sordas y sonoras españolas. *Estudios de Fonética Experimental*, *2*, 92–110.

Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387. https://doi.org/10.1016/j.wocn.2005.08.003

Cebrian, J. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, *145*, EL52-EL58. https://asa.scitation.org/doi/10.1121/1.5087645

Chang, C. B. (2012). Rapid and multifaceted effects of second-language learning on first anguage 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.

De Leeuw, E., Stockall, L., Lazaridou-Chatzigoga, D. & Gorba, C. (2019). Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. *Second Language Research*. https://doi.org/10.1177%2F0267658319886623

Dmitrieva, O. (2019). Transferring perceptual cue-weighting from second language into first language: Cues to voicing in Russian speakers of English. *Journal of Phonetics*, *73*, 128–143. https://doi.org/10.1016/J.WOCN.2018.12.008

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, *49*, 77-95. https://doi.org/10.1016/j.wocn.2014.12.005.

Docherty, G. J. (1992). *The timing of voicing in British English obstruents*. Berlin: Walter de Gruyter.

Escudero, P. (2005). *Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization*, Ph.D. thesis, LOT Dissertation Series, 113, Utrecht University

Escudero, P. (2009). Linguistic perception of “similar” L2 sounds. *Phonology in perception*, *15*, 152-190.

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. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, J. E. (1991). Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411. https://doi.org/10.1121/1.400473

Flege, J. E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. In

Winifred Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research* (pp. 233–277). Timonium, MD: York Press. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, J. E. (2002). Interactions between the Native and Second-Language Phonetic Systems. In P. Bunneister, T. Piske & A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode* (pp. 217–244). Trier: Wissenschaftlicher Verlag.

Flege, J. E. (2007). Language Contact in Bilingualism: Phonetic System Interactions. In J. Cole & J. I. Hualde (Eds.), *Laboratory Phonology* 9 (pp. 353–381) New York: De Gruyter Mouton.

Flege, J. E., & Bohn, O. S. (1997). Perception and production of a new vowel category by adult second language learners. *Second-Language Speech: Structure and Process (Studies on Language Acquisition)*, *13*, 53.

Flege, J. E., & Bohn, O. S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: theoretical and empirical progress* (pp. 3–83). New York, NY: Cambridge University Press.

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross‐language vowel perception. *The Journal of the Acoustical Society of America*, *95*(6), 3623–3641.

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. (1986). Linguistic and Developmental Effects on the Production and Perception of Stop Consonants. *Phonetica*, *43*: 155–171. https://doi.org/10.1159/000261768

Flege, J. E., & Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. S*peech Communication*, *6*(3), 185-202. https://doi.org/10.1016/0167-6393(87)90025-2522

Flege, J. E., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*(1), 67–83. https://doi.org/10.1016/S0095-4470(19)30538-8

Flege, J. E., MacKay, I. R. A., & 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., MacKay, I. R. A., & Piske, T. (2002). Assessing bilingual dominance. *Applied Psycholinguistics*, *23*(4), 567–598. https://doi.org/10.1017/S0142716402004046

Fowler, C. A. (1986). An event approach to the study of speech perception from a direct–realist perspective. *Journal of Phonetics*, *14*(1), 3–28.

Fowler, C. A. (1990). Calling a mirage a mirage: direct perception of speech produced without a tongue. *Journal of Phonetics*, *18*(4), 529–541.

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

Fox, R. A. (1982). Individual Variation in the Perception of Vowels: Implications for a Perception-Production Link. *Phonetica: International Journal of Speech Science*, *39*(1), 1. http://mendeley.csuc.cat/fitxers/21283f1243d1d945521e3843fb770fc5

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*, *24*(11), 2135–2142. https://doi.org/10.1177/0956797613486485

Gorba, C. (2019). Bidirectional influence on L1 Spanish and L2 English stop perception: The role of L2 experience. *The Journal of the Acoustical Society of America, 145*(6), EL587–EL592. https://doi.org/10.1121/1.5113808

Grosjean, F. (2001). The Bilingual's Language Modes. In J. Nicole (Ed.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 1–22). Oxford: Blackwell Publishing.

Harada, T. (2003). L2 Influence on L1 Speech in the Production of VOT. In M. J. Solé, D. Recasens & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp.1085–1088). Barcelona, Spain: Universitat Autònoma de Barcelona/Causal Productions.

Hattori, K., & Iverson, P. (2010). Examination of the relationship between L2 perception and production: an investigation of English /r/ - /l/ perception and production by adult Japanese speakers. In M. Nacano (Ed.), *Second Language Studies: Acquisition, Learning, Education and Technology*. Tokyo: Waseda University.

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French-English bilinguals. *Language and Speech, 36*(1), 17–38. https://doi.org/10.1177/002383099303600102

Hayes-Harb, R., Nicol, J., & Barker, J. (2010). Learning the phonological forms of new words: Effects of orthographic and auditory input. *Language and Speech, 53*(3), 367–381.

IBM Corp. (Released 2015). IBM SPSS Statistics for Windows, Version 23.0. Armonk, NY: IBM Corp.

Julià i Muné, J. (1981). Estudi contrastiu del oclusius de l'anglès i del català : un experiment acústic. *Estudi General: Revista de La Facultat de Lletres de La Universitat de Girona, 1*, 75.

Jun, S., & Cowie, I. (1994). Interference for "new" versus "similar" vowels in Korean speakers of English. *Ohio State University Working Papers, 43*, 117-130

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: L2 Influences on L1: A Literature Review. *Language Learning, 66*(S2):155-186. DOI: 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.

Kuhl, P., Rivera-Gaxiola, M. Conboy, B., Padden, D. & Nelson, T. (2008). Phonetic learning as a pathway to language: New data and native language magnet theory expanded. P*hilosophical Transactions of The Royal Society B Biological Sciences, 363*(1493), 979-1000.

Labov, W. (1981). Resolving the Neogrammarian Controversy. *Language, 57*(2), 267–308. doi:10.2307/413692

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

Lev-Ari, S., & Peperkamp, S. (2014). The influence of inhibitory skill on phonological representations in production and perception. *Journal of Phonetics, 47*, 36–46. https://doi.org/10.1016/j.wocn.2014.09.001

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

Liberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, *21*(1), 1–36.

Lisker, L., & Abramson, A. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*, *20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In K. Elenius & P. Branderud (Eds.). *Proceedings of the 13th International Congress of Phonetic Sciences, 4* (pp. 92-99). Stockholm: Published by the Congress organisers at KTH and Stockholm University

Mack, M. 1989. Consonant and vowel percpetion and production: Early English-French bilinguals and English monolinguals. *Perception and Psychophysics*, *46*(2) 187-200

Major, R. C. (1992). Losing English as a First Language. *Modern Language Journal*, *76*(2), 190.

Major, R. C. (2010). First language attrition in foreign accent perception. *International Journal of Bilingualism*, *14*(2), 163–183. https://doi.org/10.1177/1367006910363063

Nagle, C. L. (2018). Examining the temporal structure of the perception–production link in second language acquisition: A longitudinal study. *Language Learning*, *68*(1), 234-270.

Nagle, C. L. (2019). A longitudinal study of voice onset time development in L2 Spanish stops. *Applied Linguistics*, *40*(1), 86-107.

Nathan, G. S. (1987). On second-language acquisition of voiced stops. *Journal of Phonetics*, *15*(4), 313–322.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report. *Journal of the Acoustical Society of America*, *113*(5), 2850–2860. https://doi.org/10.1121/1.1567280

Piske, T., MacKay, I. R. A., & Flege, J. E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215. https://doi.org/10.1006/jpho.2001.0134

Purcell, E. T., & Suter, R. W. (1980). Predictors of pronunciation accuracy: A reexamination. *Language Learning*, *30*(2), 271–287.

Riney, T. J., & Okamura, K. (1999). Does bilingualism affect the first language?. *ICU Language Research Bulletin*, *14*, 101–113.

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. https://doi.org/10.1006/jpho.1997.0051

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261. https://doi.org/10.1017/S0142716400001417

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. https://doi.org/10.1006/jpho.1997.0051

Shultz, A. A., Francis, A. L., & Llanos, F. (2012). Differential cue weighting in perception and production of consonant voicing. *The Journal of the Acoustical Society of America*, *132*(2), EL95–EL101. https://doi.org/10.1121/1.4736711

Stevens, J. (2001). Study abroad learners' acquisition of the Spanish voiceless stops. *MIFLC Review*, *10*, 137-151.

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.

Trofimovich, P., & John, P. (2011). When three equals tree. In Trofimovich, P. & McDonough, K. *Applying Priming Methods to L2 Learning, Teaching and Research: Insights from Psycholinguistics* (pp.105–129). Amsterdam: John Benjamins.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, *21*(4), 289–297. https://doi.org/10.3758/BF03199477

Zampini, M. L. (1998). The Relationship between the Production and Perception of L2 Spanish Stops. *Texas Papers in Foreign Language Education*, *3*(3), 85-100.