Article

Foreign-Language Phonetic Development Leads to First-Language Phonetic Drift: Plosive Consonants in Native Portuguese Speakers Learning English as a Foreign Language in Brazil

Denise M. Osborne 1 and Miquel Simonet 2,*

1 Department of Languages, Literatures, & Cultures, State University of New York at Albany, Albany, NY 12222, USA; dmosborne@albany.edu
2 Program in Second Language Acquisition and Teaching, University of Arizona, Tucson, AZ 85721, USA
* Correspondence: simonet@arizona.edu

Abstract: Fifty-six Portuguese speakers born and raised in Brazil produced Portuguese words beginning in one of four plosives, /p b k g/. Twenty-eight of them were monolinguals (controls), and the rest were learners of English as a foreign language (EFL). The learners were also asked to produce English words beginning with one of four plosives, /p b k g/. We measured the plosives’ voice onset times (VOT) to address the following research questions: Do foreign-language learners, whose exposure to native English oral input is necessarily limited, form new sound categories specific to their additional language? Does engaging in the learning of a foreign language affect the phonetics of one’s native language? The EFL learners were found to differ from the controls in their production of Portuguese voiced (but not voiceless) plosives—prevoicing was longer in learner speech. The learners displayed different VOT targets for voiced (but not voiceless) consonants as a function of the language they were speaking—prevoicing was longer in Portuguese. In EFL learners’ productions, English sounds appear to be fundamentally modeled on phonologically similar native sounds, but some phonetic development (or reorganization) is found. Phonetic development induced by foreign-language learning may lead to a minor reconfiguration of the phonetics of native language sounds. EFL learners may find it challenging to learn the pronunciation patterns of English, likely due to the reduced access to native oral input.

Keywords: second language acquisition; phonetics; VOT; Portuguese; English

1. Introduction

When acquiring a second language (L2), most of us find it difficult to learn its pronunciation. Most people who have learned a L2 have “an accent” in their L2—the pronunciation patterns of L2 learners typically differ from those of monolingual speakers of the L2 (Colantoni et al. 2015; Piske et al. 2001; Simonet 2016; Chang 2019; Wayland 2021). This seems to be true also of many bilinguals who learned their L2 as children (early sequential bilinguals), people who live in bilingual societies (language-contact communities), and people who migrated to a foreign land where the L2 is spoken. Whereas, learning a L2 early and using it often diminishes the saliency of one’s accent, it does not always eliminate it (Piske et al. 2001). Perhaps surprisingly, some bilinguals develop “an accent” in their first or native language (L1)—they have pronunciation patterns in their L1 that differ from those of monolingual speakers of the L1 (Kartushina et al. 2016b). In extreme cases, the effects of the L2 on the L1 lead to a phenomenon known as L1 attrition, which we, following others, define as the reduction or decrease of one’s fluency and proficiency in a language, the loss of skill (Schmid 2011, pp. 11–17; Köpke and Schmid 2004, p. 5). Some scholars have concluded that the two languages of bilinguals coexist in a common representational...
network, thus influencing each other (Grosjean 1989). This has been hypothesized for phonological and phonetic knowledge, as well (Flege and Bohn 2021; Flege 1995).

Much research on L2 pronunciation development (and its limits) focuses on bilinguals immersed in their L2 (Flege 2007, 2018; Flege and Bohn 2021), and research on the potential effects of L2 learning in L1 pronunciation (or L1 phonetic drift) is mostly concerned with bilinguals who seem to be dominant in their L2 and who, in some cases, experience a reduced L1 use (Kartushina et al. 2016b; Hopp and Schmid 2013; de Leeuw et al. 2010; Major 1992; de Leeuw et al. 2018). However, many people who study a L2 do so in communities where the L2 is not commonly spoken, perhaps in their own home country—the L2 is thus a foreign language for them. In such cases, learners continue to be immersed in their L1 and use it daily, and they seldom use and practice their L2. Often, learners’ experience with their L2 is limited to the classroom setting. The population the present study investigates consists of native speakers of Portuguese—born, raised, and residing in Brazil—learning English as a foreign language (EFL). These learners rarely interact with native English speakers, and they began learning English as adults.

Our study addresses the following research questions: Do foreign-language learners develop pronunciation patterns that resemble those of native speakers of the foreign language? Or, more narrowly (and technically), do foreign-language learners, whose exposure to native oral input in their L2 is very limited, form new sound categories specific to the L2 that approximate target sounds as produced by native speakers of the L2? Moreover, does engaging in the learning of a foreign language affect the phonetics of one’s native language? Our study is concerned with foreign-language phonetic learning and native language phonetic drift. We further ask whether these two phenomena are connected.

1.1. Review of the Literature
1.1.1. L2 Phonetic Development

L2 phonetic development may be investigated in a variety of ways. One of them consists of assessing the strength of “nonnative accent” in L2 learners, as judged by a panel of native-speaking listeners of the learners’ target language. Research on the phenomenon of “nonnative accent” has identified several factors that appear to modulate the degree to which L2 learners’ pronunciation approximates that of native speakers of the L2, and these include L2 age of acquisition, L1/L2 use, length of L2 experience, motivation, formal instruction, and individual language-learning aptitude (Piske et al. 2001). This research suggests that people who began learning their L2 as adults, particularly if they continue to use their L1 often, are unlikely to acquire the pronunciation patterns of the L2 in a way that closely approximates native speaker norms (e.g., Flege et al. 1997). A second way in which L2 phonetic development has been investigated is by comparing speech samples produced by L2 learners and monolingual controls by means of acoustic analysis, and this research also suggests that late L2 learners are rather unlikely to produce speech samples that do not systematically differ from those of monolingual controls (e.g., Flege 1991). Native-likeness may be unlikely for adult L2 learners, but this does not mean that L2 phonetic development is impossible (e.g., Flege et al. 1995). Rather than asking the extent to which L2-learner speech samples differ from those of monolinguals of the L2, one could ask to what extent learners’ L2 samples differ from those they produce in their L1. From this perspective, L2 learning has to do with their having formed sound categories specific to the sounds of their L2—i.e., separate from their L1 categories—even when such new categories may not be identical to those of native speakers of the L2 (Casillas and Simonet 2018, p. 63). This is the perspective we take in the present study. Our study asks whether late L2 learners who seldom use their L2 develop phonetic categories specific to their L2, and we address this question by comparing speech samples across L2 learners’ two languages.

Much research on L2 speech development is concerned with populations fully immersed in the L2, such as migrants (Tsukada et al. 2004, 2005; Flege et al. 2003, 2006; MacKay et al. 2001). Many adult learners, however, study their L2 in their home country, typically in classroom settings. Such learners tend to remain dominant in their L1, use their
L1 much more often than they use their L2, and their experience with their L2 is limited to the classroom. Relatively few studies have examined L2 phonetic development in such populations, comparing L1 and L2 productions in a within-speaker design (Dmitrieva et al. 2020; Solon 2016, among others). For instance, Solon (2016) analyzed the production of Spanish /l/ by native English speakers studying Spanish as a L2 in the United States (US). English /l/ is darker (i.e., more pharyngealized) than Spanish /l/, particularly in syllable-coda position. In fact, English has two allophones for /l/ in complementary distribution, one used in syllable onset position and one in syllable-coda position. In a cross-sectional study, Solon (2016) found that learners, as they became more proficient in Spanish, tended to approximate native-Spanish phonetic norms and to reduce the acoustic difference between syllable-onset and coda allophones, which they seemed to transfer in the earlier stages of learning. Interestingly, even the most novice learners in Solon’s sample seemed to differentiate between their L2 and the L1 /l/, thus displaying evidence of L2 phonetic development. Dmitrieva et al. (2020) investigated the production of obstruents by native English speakers learning Russian as a L2 in the US. The learners in the Dmitrieva et al. sample also produced slightly different phonetic categories for the obstruents in their L1 and L2.

These findings suggest that L2 phonetic development is possible in cases in which one would think it unlikely. L2 learners may form new (i.e., L2-specific) phonetic categories for their L2 sounds. This does not mean that the new categories closely resemble those of native speakers of the target language. How exceptional are the findings discussed above? In the present study, we replicate and extend these findings with a different, but comparable, population of foreign-language learners.

1.1.2. L1 Phonetic Drift

A growing body of research has documented the existence of differences between monolingual and bilingual speech attributed to the influence of the L2 phonetic system on the L1 (Stoehr et al. 2017; Chang 2012; Mayr et al. 2020; Kartushina et al. 2016a; Flege and Eefting 1987; Flege 1987; Guion 2003; Major 1992; Sancier and Fowler 1997; Mora and Nadeu 2012; de Leeuw et al. 2010; Bergmann et al. 2016; Ulbrich and Ordin 2014; Fowler et al. 2008). The phonetic influence of the L2 on the L1 has been termed L1 drift (Dmitrieva et al. 2020; Sancier and Fowler 1997; Mayr et al. 2020; Chang 2013), and this is the term we use here.

Some bilinguals become dominant in their L2 and seldom use their L1. For instance, migrants may fully immerse themselves in the culture of their L2 after leaving their home country and moving to a L2-speaking country. For some migrants, this may lead to a phenomenon known as L1 attrition, the reduction or decrease of one’s fluency and proficiency in their L1. The phonetic consequences of L1 attrition have been documented in a number of studies (Flege 1987; Major 1992; de Leeuw et al. 2010, 2018; Hopp and Schmid 2013), but L1 attrition is not the focus of our investigation. It has been shown that L1 phonetic drift may be observed even in the absence of L1 attrition—that is, in bilinguals who continue to use their L1 often and remain very fluent in it (Dmitrieva et al. 2020; Sancier and Fowler 1997; Chang 2012; Mayr et al. 2020). This type of drift is what we are interested in: An effect of the phonetic patterns of L2 on those of the L1 that does not come about as a result of a reduction or decrease of a learner’s fluency and proficiency in their L1. The literature, however, has not consistently distinguished between L1 attrition and L1 drift (in the absence of attrition) (Schmid 2011, pp. 11–17).

In a review of the literature, Kartushina et al. (2016b) identified several factors that appear to modulate the nature and size of L1 drift. The chief among them is the age of L2 acquisition, and a close second is the L2/L1 use. According to this review, late L2 learners are less likely than early learners to experience L1 drift. The later in life a L2 is learned, the less likely it is to influence the L1. In late learners, L1 drift, when found, tends to be the result of full interlingual equivalence classification and assimilatory in nature, to use the terminology of the Speech Learning Model (SLM) (Flege 1995; Flege and Bohn 2021). For
instance, late learners may produce a single phonetic category for L1 and L2 sounds that, in the speech of monolingual speakers of each language, are similar but not identical. This merged category, therefore, differs both from the one used by monolingual speakers of the L2 and the one used by monolingual speakers of the L1. An intergroup difference in this direction—one in which bilinguals produce the sounds of two languages as more similar to each other (or even fully merged, identical) than those produced by monolingual speakers of those languages—is evidence of assimilatory L1 phonetic drift (Fowler et al. 2008; Flege 1987). In early learners, on the other hand, L1 phonetic drift is more likely to be the result of new-category formation and dissimilatory in nature. For instance, early learners may produce two quite different phonetic categories for L1 and L2 sounds that, in monolingual speech, are only slightly different, thus magnifying such interlingual phonetic difference in their pronunciation. This sort of intercategory deflection tends to affect the L2 sound more than the corresponding L1 sound (Flege et al. 2003), but may affect both (Mora and Nadeu 2012). These may be general tendencies, but there is no principled reason to expect that assimilatory drift is specific to late learners and dissimilatory drift to early learners. The general observation, at any rate, is that the age of L2 acquisition is associated with the L1 phonetic drift. The second factor that modulates interlingual phonetic interactions is the L1/L2 use. People who use their L2 much more often than their L1 may, in time, become dominant in their L2, and this may lead to their developing “an accent” in their L1, modifying their L1 pronunciation patterns (Mayr et al. 2020; Hopp and Schmid 2013; Major 1992; de Leeuw et al. 2010). Other factors may include speech register, L2 proficiency and experience, cognate status (Amengual 2012), and the most recent linguistic environment in which bilinguals have been immersed (Sancier and Fowler 1997; Simonet 2014; Simonet and Amengual 2019).

A small body of research suggests that the L1 drift may be found even in late L2 learners who continue to use their L1 often and, in some cases, seldom use their L2 (Sancier and Fowler 1997; Dmitrieva et al. 2020; Kartushina et al. 2016b; Chang 2012). Two studies are particularly relevant here since they focus on populations like ours, classroom learners in a foreign-language setting. Chang (2012) found evidence of L1 drift in L1 English learners of Korean who were taking a 6-week Korean language course in Korea, and Dmitrieva et al. (2020) found it in L1 English learners of Russian who were studying their L2 in North America. Dmitrieva et al. (2020) examined a constellation of acoustic correlates of plosive voicing, both of plosives in word-initial and in word-final position. For word-initial plosives, VOT was analyzed, and, for word-final plosives, the authors measured preceding vowel duration, stop closures, frication, and the duration of the voicing period during closure. Chang (2012) examined the acoustics of both plosives and vowels. Regarding the plosives, both VOT and f0 at onset were measured; for the study of vowel timbre, both F1 and F2 were analyzed. Whereas all the participants in Chang’s study were novice L2 learners, the Dmitrieva et al. speakers varied in proficiency between relatively novice to intermediate. In addition, whereas Chang’s participants were learning their L2 in the country where the L2 is spoken (and thus were likely to be exposed to their L2 outside the classroom), the Dmitrieva et al. speakers were learning their L2 in their home country (and were rarely exposed to their L2 outside the classroom). Given the factors that seem to modulate L1 drift (Kartushina et al. 2016b), including age of acquisition and use, one would not have readily anticipated that the populations investigated in Chang (2012) and Dmitrieva et al. (2020) would show evidence of L1 drift, but they did. How exceptional are these findings? Can we replicate them by investigating comparable populations?

1.1.3. The Plosives of Portuguese and English

Regarding plosive consonants, the phonemic inventories of Portuguese and English are identical. First, both languages contrast a set of phonologically voiced stops with one of phonologically voiceless stops. Second, both languages contrast three sets of plosives varying in place of articulation: Velars, coronals, and bilabials. In sum, both Portuguese and English have /p t k b d g/. However, the phonetic substance of these phonemes
differs between the two languages. To explain some of these differences, we focus on a single phonetic feature (or acoustic metric), **voice onset time** (VOT), and on how this feature is manifested in utterance-initial position. VOT is an acoustic feature that measures the asynchrony between two acoustic landmarks that correspond to two articulatory events involved in the production of plosives: Articulatory release and the onset of vocal fold vibration (Lisker and Abramson 1964, 1967; Abramson and Whalen 2017). VOT varies as a function of phonological voicing, such that voiced and voiceless consonants differ in terms of VOT patterns (among other features), but it is also affected by place of articulation (Cho and Ladefoged 1999).

In utterance-initial position, the phonologically voiced plosives of Portuguese present a period of prevoicing (i.e., the onset of voicing precedes articulatory release), whereas phonologically voiceless consonants present a brief voiceless period following articulatory release and no voicing during closure (i.e., the onset of voicing follows articulatory release, but such voicing lag is brief) (Lousada et al. 2010; Major 1987; Sancier and Fowler 1997). That said, at least one study reports to have found aspirated voiceless plosives in Brazilian Portuguese—voiceless plosives whose voicing lag is unusually (variably) long (Alves et al. 2008). Portuguese, therefore, is a “true voicing” language (Kirby and Ladd 2016; Beckman et al. 2013), that is, a language that contrasts plosives that present voicing during articulatory closure with plosives that do not (Lousada et al. 2010; Sancier and Fowler 1997; Major 1987, 1992). English, on the other hand, is an “aspirating” language (Beckman et al. 2013). In utterance-initial position, phonologically voiced English plosives present a brief period of devoicing and lack voicing during closure (i.e., the onset of modal voicing follows articulatory release by a few milliseconds), and phonologically voiceless plosives have a relatively long period of aspiration, also lacking voicing during closure (i.e., the voicing lag period is very long since the devoiced burst is followed by a period of voiceless aspiration) (Lisker and Abramson 1967). In English, the phonological contrast between voiced and voiceless plosives is phonetically implemented with the presence (vs. absence) of aspiration rather than voicing. In order to make the association between the phonetics and phonology more transparent, some scholars have proposed that the “voicing” contrast in English does not actually involve phonological voicing (i.e., a [voice] distinctive feature) but aspiration (i.e., a [spread glottis] distinctive feature) (Beckman et al. 2011, 2013). We would thus say that Portuguese is a [voice] language and English is a [spread glottis] language.

A small body of literature has explored VOT in the speech productions of Portuguese–English bilinguals (Major 1987, 1992; Sancier and Fowler 1997). Major (1987) found that native Portuguese speakers learning English as a foreign language in Brazil produced English /p t k/ with a much shorter voicing-lag period than native English-speaking controls. Major (1992), on the other hand, analyzed the Portuguese and English productions of a group of native English speakers fully immersed in Brazilian culture, having resided in Brazil between one and three decades. Major’s findings showed that some of the bilinguals produced Portuguese /p t k/ with shorter voicing-lag periods than native Portuguese controls and English /p t k/ with shorter voicing-lag periods than monolingual English controls. This is an example of assimilatory L1/L2 influence leading to L1 drift. Finally, Sancier and Fowler (1997) investigated both the English and Portuguese productions of a single L1 Portuguese speaker who learned English as a L2, a very proficient L2 learner. The speaker was recorded in both languages in three different occasions: After 4 months in the US, immediately after a 2.5-month stay in Brazil, and once again after 4 months in the US. The speaker produced English /p k/ with a much longer voicing-lag period than Portuguese /p k/, which showed that the speaker had formed L2-specific VOT categories. Moreover, there was some systematic variation between the recording sessions, such that voicing-lag periods were longer (in both languages) after 4 months in the US than after 2.5 months in Brazil. The latter suggests that recent phonetic exposure may serve to recalibrate VOT targets. The present study is concerned with the Portuguese and English productions of a group of L1 Portuguese learners of English as a foreign language who remain immersed in their L1 and have never travelled to an English-speaking country. Our
population is, therefore, the same population investigated in Major (1987), but different from the populations investigated in the other two studies (Sancier and Fowler 1997; Major 1992).

1.1.4. The Learning of English in Brazil

In public and private schools in Brazil, English has been a compulsory subject from the 6th grade to the final year of high school since January of 2020. These changes were implemented according to proposals made by the new Base Nacional Comum Curricular (BNCC, basenacionalcomum.mec.gov.br, accessed on 22 June 2021), a normative document from the Education Ministry of Brazil that defines the essential disciplines taught in K-12. The BNCC had previously established that schools should include at least one compulsory foreign language in their curriculum (Law n. 9.394 of December 1996), but the guidelines did not establish which foreign language should be taught. In 2005, a change in the BNCC (Law n. 11.161) turned Spanish into the compulsory foreign language to be taught in public schools.

Even though English is currently the most frequently taught foreign language in public and private Brazilian schools, its instruction and learning have faced several challenges. Some of the challenges identified in the literature concern the large classes, a lack of resources, teachers with insufficient training and proficiency in English, and the fact that instruction is primarily restricted to grammar (Santos 2011). The first national-level research on the teaching and learning of English was conducted only recently (British Council 2019). This study was intended to provide baselines for compulsory English instruction in Brazil. The findings of the study suggested that the challenges reported in previous studies (e.g., Santos 2011) were shared across the country. The study highlighted two main obstacles for the teaching of English in public schools: The lack of teachers with specialized training, and the lack of a curriculum that focused on the social use of the language. It revealed that about half of the English teachers who taught in public schools did not hold a degree in English language or the teaching of English; 81% of the English teachers in the study complained about the lack or the unsuitability of textbooks and course materials; and less than one fourth of the classrooms had access to the internet.

Along with the challenges for effective English teaching in public schools, there is a general belief that the learning of English in public schools is ineffective and that, if someone wants to learn English, they have to study in a language school (Silva 2004). Language schools are private schools, they are believed to be better equipped, to have fewer students in the classroom, better trained instructors, and more resources (Polidário 2014). This belief is shared among students and teachers.

The data of the present study were collected at one of the branches of Cultura Inglesa, (www.culturalinglesa.com.br accessed on 22 June 2021) a language school franchise that works in partnership with the British Council. Cultura Inglesa opened its first school in 1934 in Rio de Janeiro, the capital of Brazil at the time (Tavares 2018). It is currently one of the most popular language schools in Brazil, with more than 70 branches across the country. The school offers its students exchange programs and a variety of English proficiency exams published by Cambridge English Qualifications (e.g., FCE-First Certificate in English). Their English teachers are consistently engaged in training, courses, and seminars—both in Brazil and abroad. Students attend two 80-min classes per week, each class has a relatively small number of students, and classrooms are equipped with computers and projectors. Cultura Inglesa uses the communicative approach with no translation to Portuguese and uses books from international publishers.

1.2. The Current Study

We analyze the production of voiced and voiceless plosives in two languages, Portuguese and English. Native Portuguese speakers were recruited for our production study. Some were monolingual Portuguese speakers, and others were learning English as a foreign language (EFL) in a private language school in Brazil. The monolinguals were recorded
only in their native language and served as controls; the EFL learners were recorded in both Portuguese (L1) and English (L2).

Firstly, we ask whether EFL learners differ from monolingual Portuguese speakers in their production of Portuguese plosives, with a focus on VOT. If the experience of actively learning a L2 leads to modifications in L1 pronunciation patterns (L1 phonetic drift), we would find that learners and monolinguals differ in the way they pronounce Portuguese sounds. If, on the other hand, L2 learning (in this population of Brazilian EFL learners) does not lead to L1 phonetic drift, we would not find any significant differences between the two groups of native Portuguese speakers. Our hypothesis, given the findings in the contextual literature, was that the Brazilian EFL learners in our study would be unlikely to display any effects of L1 phonetic drift, since they seldom use their L2, continue to be immersed in their L1, and rarely interact with native English speakers.

Secondly, we ask whether EFL learners develop new phonetic categories specific to their L2, English. To address this question, we compare the VOTs of Portuguese and English plosives produced by the EFL learners. Do these learners use the same VOT categories for their two languages or different ones? If the learners transferred the phonetic categories of the L1 into their L2 and had failed to develop new VOT targets for their L2, we would find that they produced a single prevoiced VOT category for voiced plosives and a single short-lag VOT category for voiceless plosives in both languages. Note that we do not ask whether the EFL learners pronounce English sounds in the same way native English speakers do (Major 1987), but, rather, whether their English VOT categories differ from their own Portuguese VOT categories. The focus, therefore, is on new-category formation (Flege 1995; Flege and Bohn 2021), not nativelikeness. Our hypothesis, given the findings in the literature, was that the Brazilian EFL learners would fundamentally transfer their native categories to their L2 and would be unlikely to have developed phonetic categories specific to their L2. In sum, our working hypotheses were the null hypotheses.

2. Method

2.1. Sample

A sample of 56 adults participated in an elicited production study. All of the participants were native speakers of Portuguese born and raised in Brazil. The participants were divided into two groups according to whether they were learning English as a foreign language (EFL) or not. Twenty-eight participants were EFL learners and, at the time of the study, were enrolled in English classes in a private language school. This was our experimental group. The remaining 28 participants did not consider themselves learners of English, at least not at (or before) the time of the study. This was our control group. The difference between the two groups is that between a group of emerging bilinguals (i.e., learners actively engaged in the task of studying a foreign language in school) and one of functional monolinguals.

The control group consisted of adults raised as monolingual speakers of Portuguese (N = 28). All of the participants in this group were born and raised in the state of Minas Gerais, most of them in Araxá. Other cities of origin included Bambuí, Campos Altos, Frutal, Ibiá, Perdizes, São Gotardo, Três Marias, Uberaba, and Uberlândia. They were recruited in a variety of locations around the city of Araxá, mostly in a university setting. The median age of the control group was 29 years old, with 18 and 55 being the minimum and maximum ages, respectively. Twenty-one of the members of this group were women, and seven were men. Nine of the participants had obtained a postgraduate degree, 12 had graduated from college, and seven had a high school diploma. None of the members of this group reported having had any significant exposure to English—they had never studied English or traveled to an English-speaking country. Some had studied a Romance language—Spanish, mostly—for a few months or up to a year. Immediately after their participation in the production study, they were asked if they were able to produce a full sentence in English (any sentence). None were able to do so.
The target, experimental group consisted of adults raised as monolingual Portuguese speakers (N = 28). All of the members of this group were born and raised in the state of Minas Gerais, most of them in Araxá. Other cities of origin included Belo Horizonte, Ibiá, Ponto Nova, São Gotardo, São João del Rei, and Uberaba. The median age was 26, the minimum age was 18, and the maximum was 50. Eighteen of them were women and 10 were men. Two participants in this group had obtained a postgraduate degree, 19 had graduated from college, and seven had a high school diploma. The participants in this group were enrolled as EFL students in the Araxá branch of Cultura Inglesa (culturainglesaaraxa.com.br accessed on 22 June 2021). The EFL learners in our sample were recruited and tested in the language school.

In addition to English, some of the EFL learners reported having studied some Spanish or Italian, but none had been studying any language (besides English) in the months preceding the study. We collected additional relevant data from the participants in the EFL group, most of the data pertained to their experience as EFL learners and their English proficiency. We asked them whether they had ever had a native English speaker as a teacher—fifteen of them (54%) had—and how long had they been taking English classes at Cultura Inglesa—this ranged from one to 21 years, with 4 years being the median age (the 25th percentile was two and the 75th percentile was seven). None of the participants had ever visited any English-speaking country.

A survey was administered to all EFL learners. The survey asked participants to rate their English proficiency on a 1 (poor) to 7 (excellent) scale in each of the following skills: Grammar, listening, pronunciation, reading, speaking, and vocabulary. The mean score for all these skills was 4.2 (SD = 1.2). The seven skills were highly correlated with each other (r = 0.5–0.8), and a reliability analysis yielded a high score: Cronbach’s α = 0.93. The survey also asked their estimated percentage of English use in the following environments: With friends, at home, on the internet, in the media (including music and television), in online classes, at school, and at work. The overall percentage of use of English, averaged over all settings and learners, was 38% (SD = 13.5%). Percentage scores were generally not correlated with each other, and they varied from a high percentage of use of English at the language school (82%) to a low use with friends (16%). At work, the mean percentage was 41.2%, and this was the setting that induced the largest variance in the sample (SD = 38%). The survey also asked participants to rate their motivation to learn English on a 1 (disagree) to 7 (agree) scale in response to the following prompts: I am learning English because it will help me find a better job (M = 6.1), I am learning English because I love English or American culture (M = 5.6); Learning English makes me feel important (M = 5.3); I do not ever want to stop learning English (M = 6.4); I like learning English (M = 6.5); When I speak English, I do my best to avoid using Portuguese (M = 5.6); When I speak English I try to imitate the English or American accent (M = 5.6). Unsurprisingly, the motivation questions did not reliably measure the same construct, Cronbach’s α = 0.78. Overall, the survey suggests that the participants in our sample were highly motivated to learn English and that their use of English was mostly restricted to the language school. In terms of their self-assessed proficiency, the participants rated themselves, on average, as intermediate learners, but a variety of proficiency levels is represented in the sample.

Finally, general English proficiency was assessed by means of a brief cloze test focusing on grammar and vocabulary (Brown 2002; Tremblay 2011), the St. George’s International English Placement Test (stgeorges.co.uk/online-english/online-english-test accessed on 22 June 2021). This test is comprised of 40 individual sentences, all of them very brief, out of which one word has been substituted by a blank. Four options are given to test takers to fill in each blank—this is a multiple-choice test. Out of 40 points, our participant sample obtained a median score of 26.5 (SD = 7.9). The minimum score was 10 and the maximum was 38; the 25th percentile was 19.75 and the 75th percentile was 31.25. The self-assessed proficiency scores were positively correlated with the results of the cloze test, r = 0.5, 95% CI [0.15, 0.73], p = 0.007. This further suggests that we were able to recruit learners from a range of English proficiency levels, from relative novice to advanced learners.
2.2. Instrument

The main experiment was an elicited production task with both auditory and visual prompts presented simultaneously. For each target word—that is, in each experimental trial—participants heard an acoustic stimulus (or auditory prompt) and they saw on a computer screen both a written rendering of the target word (or orthographic prompt) and a drawing that represented the word meaning (or figure prompt). In other words, the participants simultaneously received three types of prompts to elicit their production of each target word. For instance, for the Portuguese word *pato* “duck,” the participants heard a recording by a native speaker of the utterance *pato* é *a palavra* “duck is the word,” played over headphones. Simultaneously, the computer screen showed an orthographic rendering of the target word, <pato>, and a line drawing representing the bird. Line drawings were creative commons figures or are in the public domain, they were all outline drawings in black and white. The simultaneous presentation of prompts in three modes was done to ensure that all EFL learners, including the beginners, had the best chance to recognize the English word they were being asked to produce (this method was probably redundant in the Portuguese production task).

The control group of Portuguese monolinguals were asked to produce only the Portuguese words, and the EFL learners were asked to produce both the Portuguese (L1) and English (L2) words in two separate sessions. The present study focuses on the production of bilabial and velar plosives. Dental (or alveolar) plosives were not included since, in Brazilian Portuguese, they are known to be pronounced as postalveolar affricates when followed by high front vowels (Barbosa and Albano 2004; Albano 2001, pp. 68–86). This study is concerned with the voicing contrast as manifested in VOT. In sum, we examine both voiced and voiceless plosives in two places of articulation, bilabial and velar.

In our materials, the target plosives appeared always in word- and utterance-initial position to be able to reliably measure prevoicing. For each of the four plosives (and each of the two languages), we chose 20 words that began with that plosive. For half of those 20 words, the target plosive was followed by a low vowel; the other half was followed by a high vowel, either front or back. We were not interested in assessing the potential role of contextual vowels (Lousada et al. 2010; Nearey and Rochet 1994; Yavaş and Wildermuth 2006), but we included such variation for the sake of generalizability to all vowel contexts. We obtained a balanced number of observations of the four target consonants, /p b k g/.

All in all, we manipulated phoneme, vowel context, and language, and we controlled for utterance and word position (initial). Target words were placed in a constant carrier phrase: __ é *a palavra* (Portuguese), __ *is the word* (English). When possible, we used minimal pairs contrasting in voicing, such as *pond-bond* (English) and *panda-banda* (Portuguese), in both languages. Most English words were monosyllabic and most Portuguese words were disyllabic. In the Portuguese disyllabic words, lexical stress occurred in word-initial position. These design principles resulted in a list of materials comprising 80 words per language: 20 (lexical items) × 4 (phonemes). The list of target words is found in Table 1.

The auditory stimuli were recorded from one male talker of each language. The talkers were asked to read out loud a list of utterances. The utterances were comprised of the target word in a constant carrier phrase: __ é *a palavra* (Portuguese), __ *is the word* (English). The talkers were also asked to record the question Qual é *a palavra*? (Portuguese) or *What is the word?* (English) at the end of the recording session. To record the auditory stimuli, the talkers sat inside a sound-attenuated booth on the campus of the University of [Removed for Review]. The stimuli were recorded with a Fostex DC-R302 digital recorder and a Shure SM10A head-worn dynamic microphone. The signal was digitized at 44.1 kHz and 16-bit quantization. The talkers read the entire list of materials in their native language three times. One rendering of each target item was selected to be used as auditory stimuli. The sound files were normalized for peak intensity at 75 dB. The talker who produced the English materials was a native speaker of English born and raised in [Removed for Review]. When he was recorded, he was 22 years old and did not speak any language other than English. The talker who produced the Portuguese materials was a native speaker of...
Portuguese born and raised in the city of São Paulo, Brazil. At the time of the study, he was living in [Removed for Review]. An exchange student at the University of [Removed for Review], the Portuguese talker had been in the US for 8 months when he was recorded. He assessed himself as being an intermediate English learner.

Table 1. Word lists produced by participants.

| Phoneme | Portuguese | English |
|---------|------------|---------|
| /p/     | panda, pata, pato, pano, puca, pala, palha, pasta,       | pond, pack, par, part, pan, pat, park, |
|         | papa, passo, pia, pica, picho, pilha, pinga, pico,      | pox, pot, path, pea, pin, peep, peach, |
|         | pingo, pulo, puca, puxo.                                 | pig, pit, pull, Pete, put, push.      |
| /b/     | basta, baba, baço, Bia, bica, bicho, bilha, binga,      | bond, back, bar, Barb, ban, bat, bark, |
|         | bico, bingo, burro, bucha, bucho.                         | box, bat, bath, bee, bin, beep, beach, |
| /k/     | case, cabo, quintu, guilha, quina, quica,               | kit, kiss, cool, keys, coo.           |
|         | quincha, quita, Quito, cuspa, cuolo, cura.               |                                      |
| /g/     | gale, gato, gala, gama, gata, gana, ganuro, gasto,     | goth, garb, got, god, guard, gap, gas, |
|         | gaze, garbo, guinho, guilha, guina, guiga,              | Gary, gash, gab, good, gill, guilt, gig, |
|         | guinchu, guitar, Guido, gume, Guto, gula.               | girl, git, gift, goose, geese, goo.   |

2.3. Procedure

Speech productions were elicited, as explained above, by three types of simultaneous prompts: An auditory prompt (a recording by a native speaker of the language), a written prompt (an orthographic rendering of the word), and a figure prompt (a conceptual representation of the word in the form of a line drawing). Each trial consisted of the simultaneous presentation of the three prompts followed by a recording of the question Qual é a palavra? (Portuguese) or What is the word (English). The question was to be followed by the participant’s production of the target utterance. A new trial began every 6 (Portuguese) or 7 s (English). The timing of the trials was determined arbitrarily. There were 80 trials per language, presented in random order within each session.

The Portuguese controls provided speech samples only in their native language and thus participated in a single experimental session. The participants in this group were recruited by a native speaker of English born and raised in the state of New York, a research confederate, who visited the language school and invited students to participate. All conversations between the confederate and the participants took place in English. This presumably encouraged the participants to situate themselves in English mode for the English session. The confederate asked the participants a list of background questions—including the proficiency, use, and motivation questions reported above—and then asked them to take the English cloze task. Finally, the confederate administered the elicited production task.

When the participants were done with the English portion of the study, they were approached by the first author, a native Portuguese speaker. She invited them to participate in an “additional study on Portuguese,” the second experimental session. They were invited to stay in the room or to return after a brief break.

All conversations between the first author of the study and the participants took place in Portuguese to encourage them to switch to their native language mode. In both sessions, the randomized presentation of the prompts was managed by a stimulus presentation software, PsychoPy2 (Peirce 2007; Peirce et al. 2019). The survey and cloze test data were collected in paper format.
2.4. Data and Analyses

The acoustic metric of choice in this study is VOT. Since we obtained samples of both English and Portuguese voiced and voiceless plosives, one would expect to find the full range of values, from prevoicing (negative VOT) to aspiration (long lag VOT). The segmentation of the acoustic material was done by hand by the first author, who utilized both waveform and spectrographic displays to locate the acoustic landmarks of interest. Segmentation was done in Praat (Boersma 2001). For each target consonant, the first author placed a mark at the onset of the burst that corresponded to the release of articulatory closure and another one at the onset of modal voicing. Both acoustic landmarks were adjusted so that they occurred at upwards zero-crossings in the waveform. If the onset of modal voicing precedes the burst, VOT is a negative value, and this indicates the presence of prevoicing. If, on the other hand, the onset of modal voicing follows the burst, VOT is a positive value, and this indicates voicing lag. A very long lag is suggestive of the presence of aspiration.

The data set had a theoretical ceiling of 6720 observations. Each participant in the Portuguese control group produced 80 tokens, all of them in Portuguese: 80 (words) \times 28 (speakers) = 2240 tokens. Each participant in the EFL learners group produced 80 Portuguese tokens and 80 English tokens: 80 (words) \times 2 (languages) \times 28 (speakers) = 4480 tokens. The actual data set comprised 6714 observations, as six tokens were either discarded due to the presence of noise in the recording or simply not recorded due to experimental error, such as a trial for which the participant did not produce a response.

Inferential statistics were conducted on by-speaker averages. We calculated the average VOT per speaker per condition. We call this metric \( mVOT \) for mean VOT. This resulted in three data sets, which were then combined into larger data sets to conduct a variety of statistical comparisons across speaker groups and conditions. The first data set comprised the Portuguese control data: It included four average values per participant, one per plosive /p b k g/: 28 (speakers) \times 4 (phonemes) = 112 observations. The second data set comprised the Portuguese (L1) productions of the EFL learners: It included four average values per participant, 28 (speakers) \times 4 (phonemes) = 112 observations. The third data set comprised the English (L2) productions of the EFL learners, 28 (speakers) \times 4 (phonemes) = 112 observations. Each of these values is an average over 20 observations. In sum, the data set comprising 6714 raw VOT measurements was reduced, by means of by-speaker and by-condition averaging, to 336 observations. Data reduction and wrangling were done with an \( R \) script (R Core Team 2018), with the functions provided by the package \textit{tidyverse} (Wickham et al. 2019). Data analyses were conducted in \textit{Jamovi} (The Jamovi Project 2020), a free open-source GUI for \( R \). The \( R \) packages used in \textit{Jamovi} were \textit{afex} (Singman et al. 2020), \textit{emmeans} (Lenth 2018), and \textit{esci} (version 0.9.1 for Jamovi, written by Robert J. Calin-Jageman). See \textit{jamovi.org/library} (accessed on 22 June 2021) for a list of available modules. Synthetic data and code, including the \textit{jamovi} files, may be made available to readers interested in reproducing our analyses. Readers may contact the corresponding author.

3. Results

This section first reports on the results concerning the VOT values of Portuguese plosives as produced by our two groups of Portuguese native speakers, the EFL learners and the controls. This is a between-group comparison that keeps the language constant, Portuguese. Secondly, we report on a statistical comparison of the VOT values of both the Portuguese (L1) and English (L2) plosives produced by the EFL learners. This is a within-group comparison of plosives in two languages.

3.1. Portuguese Productions: Between-Subjects Comparison

The Portuguese productions of both groups of native speakers, the controls and the EFL learners, were compared against each other. The dependent variable was \( mVOT \) (ms), and the factors were \textit{place} (bilabial, velar), \textit{voicing} (voiced, voiceless), and \textit{group} (controls,
EFL learners). The descriptive statistics (mean and standard deviation, M (SD)) were as follows. For /b/, mVOT values for the controls were −112 (22) ms, and learners’ values were −117 (25) ms. For /g/, the mean for the controls was −83 (22), and the learners’ mean was −100 (18). For /p/, the controls’ mean was 11 (6) and learners’ mean was 12 (6). For /k/, the mean for the controls was 55 (11), and learners’ mean was 61 (10).

The data were submitted to a mixed-design ANOVA with place and voicing as within-subject factors and group as a between-subjects factor. This is a (2) × (2) × 2 design. The α criterion was set at 0.05. The ANOVA yielded main effects of both voicing, F(1,54) = 2373, p < 0.0001, η²G = 0.94, and place, F(1,54) = 484, p < 0.0001, η²G = 0.48. There was a marginally significant effect of group, F(1,54) = 4.4, p < 0.05 [0.0413], η²G = 0.03. There were two significant two-way interactions: A voicing and place interaction, F(1,54) = 120, p < 0.0001, η²G = 0.151, and an interaction between voicing and group, F(1,54) = 12.5, p < 0.001 [0.0008], η²G = 0.08. In general terms, voiced plosives had negative VOT values and voiceless ones had positive values, as one would expect (Lousada et al. 2010). Velar plosive means were “displaced to the right” relative to bilabial means; that is, /k/ (M = 60, 95% CI [53, 62]) had a longer voicing lag than /p/ (M = 12 [7, 16]) and /g/ had a shorter prevoicing period (M = −92 [−96, −86]) than /b/ (M = −109 [−114, −105]).

Figure 1. Mean (and 95% CI) of Portuguese mVOT values plotted as a function of place (bilabial, velar), voicing (voiced, voiceless), and speaker group (EFL learners, controls). Data come from 56 native speakers of Portuguese, 28 of whom are foreign language learners of English.
3.2. Learner Productions: Within-Subject Comparison

This section reports on the results of a comparison between the Portuguese (L1) and English (L2) productions of the 28 EFL learners in our sample. The Portuguese monolingual controls are excluded from this analysis. The dependent variable was \( mVOT \) (ms), and the predictors were place (bilabial, velar), voicing (voiced, voiceless), and language (Portuguese, English). It is important to remember that all factors were within-subject factors—the factor language compared the English (L2) and Portuguese (L1) productions of a single group of speakers. The descriptive statistics, mean (and standard deviation), were as follows: For /b/, the mean VOT for L1 productions was −117 (25), and the L2 mean was −93 (36). For /g/, the L1 VOT mean was −100 (18), and the L2 mean was −62 (39). For /p/, the L1 VOT mean was 12 (6), and the L2 mean was 19 (14). For /k/, the mean VOT for L1 tokens was 61 (10), and the L2 mean was 61 (17).

The \( mVOT \) data were submitted to a repeated measures ANOVA with place, voicing, and language as within-subject predictors. This is a \( (2) \times (2) \times (2) \) design, and the \( \alpha \) criterion was set at 0.05. The ANOVA revealed significant main effects of voicing, \( F(1,27) = 674, p < 0.0001, \eta^2_G = 0.89 \), place, \( F(1,27) = 346, p < 0.0001, \eta^2_G = 0.36 \), and language, \( F(1,27) = 25, p < 0.0001, \eta^2_G = 0.12 \). Voiced plosives had, on average, negative VOT values, and voiceless plosives had, on average, positive ones, a mean difference of approximately 131 ms (SE = 5). Relative to bilabials, velars were generally “displaced to the right,” a mean difference of about 35 ms (SE = 2); that is, /k/ had a longer voicing lag period than /p/, and /b/ had a longer prevocing period than /g/. Finally, there was a difference between English and Portuguese values such that, in general, English values were “displaced to the right” relative to Portuguese ones, a mean difference of about 17 ms (SE = 3). Since there were several significant interactions between the factors, the main effects may not be interpreted on their own. There were two two-way interactions: voicing by place, \( F(1,27) = 42, p < 0.0001, \eta^2_G = 0.05 \), and voicing by language, \( F(1,27) = 27, p < 0.0001, \eta^2_G = 0.08 \). The voicing by place interaction seemed to be due to the fact that the difference between velar and bilabial plosives was larger in the voiceless set, \( M_{\text{diff}} = 45, t(53) = 18, p^{\text{tuckey}} < 0.0001 \), than in the voiced set, \( M_{\text{diff}} = 24, t(53) = 10, p^{\text{tuckey}} < 0.0001 \). The voicing by group interaction was due to the fact that the prevocing period was significantly longer in the Portuguese voiced plosives than in the English ones, \( M_{\text{diff}} = 31, t(50) = 7.2, p^{\text{tuckey}} < 0.0001 \), whereas voicing lag was similar for the two languages in the voiceless set, \( M_{\text{diff}} = 3.5, t(50) = 0.8, p^{\text{tuckey}} > 0.05 \) [0.85]. It seems that, in this data set, voiced consonants differ as a function of the language spoken by the learners, whereas voiceless consonants do not.

The omnibus ANOVA also yielded a three-way interaction. There were no significant effects of language neither for /p/, \( M_{\text{diff}} = 7, t(65) = 1.5, p^{\text{tuckey}} > 0.05 \) [0.82], not for /k/, \( M_{\text{diff}} = 0.2, t(65) = 0.05, p^{\text{tuckey}} > 0.05 \) [1]. In addition, whereas there were effects of language for both /b/ and /g/, language effects were larger for the velars, \( M_{\text{diff}} = 38, t(65) = 8.2, p^{\text{tuckey}} < 0.0001 \), than for the bilabials, \( M_{\text{diff}} = 24, t(65) = 5.2, p^{\text{tuckey}} < 0.0001 \). To summarize, voiceless plosives had, on average, a positive VOT in both L1 and L2 productions, and voicing lag was particularly long for /k/. VOT values in L1 and L2 voiceless plosives did not differ from each other. Voiced plosives had, on average, a negative VOT in both L1 and L2 productions, and the prevocing period was longer for /b/ than for /g/. Most importantly, the prevocing period was longer in the Portuguese productions than in the English productions. There were, therefore, effects of language in the voiced set. It seems that the EFL learners utilized the same VOT targets for the English and Portuguese voiceless plosives. However, they seem to have separate VOT targets for the English and Portuguese voiced plosives. Figure 2 plots average VOT values as a function language, place, and voicing.
The EFL data presented here suggest that learners develop a “compromise” VOT category for English voiced plosives (Casillas 2021; Flege 1991, p. 395). This compromise category seems to be based on the native Portuguese category, and thus presents significant prevoicing, but it also seems to approximate, to some extent, the native English category, hence the shorter negative VOT. There is, however, an alternative explanation. Casillas (2021) suggests that the “compromise” VOT values that are sometimes found in the literature on bilingual speech production may result from averaging extreme values and not from the actual development of intermediate or compromise categories in bilingual speech. According to Casillas, bilinguals may be producing “performance mismatches”—when speaking in their less dominant language, bilinguals may fluctuate between L1-like and L2-like tokens. The presence of L1-like tokens in the pool of L2 productions could thus alter the resulting average, displacing it in the direction of L1 categories. Translated to our findings, Casillas’ interpretation would be that our EFL learners present, on average, shorter negative VOT values for their L2 plosives than for their L1 plosives since we have averaged both English productions with Portuguese-like VOT values (i.e., with prevoicing just as long as their native Portuguese plosives) and English productions with English-like VOT values (i.e., with short-lag VOT). Therefore, the averages reported above would not represent any actual production target of the learners but would be the result of averaging extreme values. By hypothesis, our EFL learners could have been producing both prevoiced and short-lag VOT values in their English plosives, and thus the intermediate category we reported as the average would simply indicate the possible presence of performance mismatches. The long prevoicing in the Portuguese (L1) plosives, on the other hand, would simply come from the lack of any short-lag VOT tokens in this pool—there would not be any performance mismatches in the native language.

To address this issue, we focused on the English (L2) productions of the EFL learners only and, particularly, on the voiced plosives, /b g/. There were 1117 voiced plosives in the English data set. Of these, 901 were produced with prevoicing and 216 were produced with short-lag VOT. There were 1120 voiced plosives in the Portuguese data set and, of these, only 19 were produced with short-lag VOT. Performance mismatches appear to occur in both the L1 and L2 but, as Casillas hypothesized, performance mismatches seem to be more common in the L2. This would indeed displace the average VOT of the English plosives in our data set closer to zero, towards the prototypical English category. The difference between the two languages was found to be significant: 98.3% (95% CI [97.4, 98.9]) of the

![Figure 2. Mean (and 95% CI) mVOT values plotted as a function of place (bilabial, velar), voicing (voiced, voiceless), and language (L2 English, L1 Portuguese). Data come from 28 native speakers of Portuguese learning English as a foreign language in Brazil.](image-url)
Portuguese voiced productions were prevoiced, whereas only 80% ([78.2, 82.9]) of the English voiced productions were prevoiced—which is a difference of 17.6% ([15.2, 20.1]) (Cumming 2013, pp. 399–401). A mixed-effects logistic regression model confirmed that the proportion of short-lag VOT tokens was larger in English than in Portuguese: $\hat{\beta} = 2.62$, $z = 5.04$, $p < 0.001$. The English voiced plosives produced by the EFL learners may not, after all, have a shorter prevoicing period than their Portuguese voiced plosives. It could be that about 20% of their English productions were produced with short-lag VOT.

How long is the prevoicing period of the English voiced plosives that are indeed prevoiced in the speech of the EFL learners? Is it just as long as that of their own Portuguese productions? To answer these questions, we selected all of the English and Portuguese voiced plosives, /b g/, that had been produced with negative VOT, prevoicing. Tokens produced with short-lag VOT were excluded from this analysis. In this sample, there were 1101 Portuguese tokens and 901 English tokens. Then, we calculated by-speaker and by-condition averages. The descriptive statistics for this subset were as follows: The mean VOT for English /b/ was $-106$ (SD = 31) and that for Portuguese /b/ was $-118$ (25); the mean VOT for English /g/ was $-94$ (24) and that for Portuguese /g/ was $-103$ (18). The dependent variable, $mVOT$ (ms), was submitted to a repeated measures ANOVA with language (English, Portuguese) and place (bilabial, velar). Recall that only voiced plosives with prevoicing were included in this analysis. The ANOVA yielded significant main effects of place, $F(1,27) = 25$, $p < 0.0001$, $\eta^2_G = 0.07$, and of language, $F(1,27) = 8.1$, $p < 0.001$ [0.0083], $\eta^2_G = 0.04$, but there was no significant interaction between the two factors, $F(1,27) = 1.2$, $p > 0.05$ [0.29], $\eta^2_G = 0.001$. Bilabials had, on average, a longer prevoicing period than velars, $M_{diff} = 14$, $t(27) = 5$, $p_{tuckey} < 0.0001$. Most importantly, on average, the prevoicing period of the English plosives was shorter than that of the Portuguese plosives, $M_{diff} = 11$, $t(27) = 3$, $p_{tuckey} < 0.001$ [0.008]. Figure 3 plots average VOT values as a function of language and place.

The difference between English and Portuguese /b/ in this data subset was, on average, 12 ms, 95% CI [3.2, 20.9]. The standardized mean difference, corrected for bias, was $d_{avg} = 0.42$, 95% CI [0.13, 0.74]. The correlation between the paired measures was $r = 0.69$. As for /g/, the difference between the English and Portuguese VOT values in this subset was, on average, 9 ms, [1.6, 16.3]. Corrected for bias, the standardized mean difference was $d_{avg} = 0.42$, 95% CI [0.10, 0.78], and the correlation between the paired measures was $r = 0.62$. 

![Figure 3. Mean (and 95% CI) mVOT values plotted as a function of place (bilabial, velar) and language (L2 English, L1 Portuguese). Data come from 28 native speakers of Portuguese learning English as a foreign language in Brazil. The sample includes only voiced plosives with prevoicing.](image-url)
To summarize, even when excluding the tokens that have been produced with short-lag VOT, the average prevoicing length was found to be larger in Portuguese than in English in the speech produced by EFL learners. It seems that, in addition to performance mismatches, EFL learners produced intermediate or compromise VOT categories.

3.2.2. Effects of Proficiency or Use?

The present subsection explores the possible role of L2 proficiency and amount of usage on the production of English plosives by EFL learners. The analyses we have reported in preceding subsections suggest that only an analysis of voiced plosives is likely to reveal any effects of proficiency or use. Firstly, EFL learners seem to have longer prevoicing production targets for voiced plosives in their native Portuguese than monolinguals speakers of Portuguese do. There is no difference, however, between the two groups with regards to voiceless plosives. Secondly, EFL learners seem to have longer prevoicing targets for voiced plosives in their native Portuguese than they do for voiced plosives in English, their L2. There is no difference between L1 and L2 productions for this group with respect to the voiceless plosives. All this suggests that, if we are to find any effects of English proficiency or use on speech production in this sample, we are likely to find them only in the voiced plosives. Therefore, for the analyses reported in this subsection, we focused exclusively on the voiced plosives produced by the EFL learners.

We conducted two sets of analyses. On the one hand, we investigated the potential effects of proficiency and use on the phonetic characteristics of English voiced plosives. For these analyses, we concerned ourselves only with the English voiced plosives produced by the 28 EFL learners in our sample, their L2 plosives. We asked whether English proficiency or use (or both) led to differences in the VOT of the English plosives produced by EFL learners. We hypothesized that increases in English proficiency and use are associated with a shorter length of prevoicing in the English voiced plosives produced by the learners. On the other hand, we analyzed the potential effects of proficiency and use on the size of language mode effects in the voiced plosives produced by the learners. To obtain our dependent variable, we subtracted the mean VOT of a given learner’s Portuguese voiced plosive from that of their own corresponding English plosive. We asked whether English proficiency or use (or both) led to differences in the size of the difference between mean L1 and L2 VOT values (for voiced plosives only). We hypothesized that increases in English proficiency and use are associated with larger differences (i.e., larger effects of language mode) between L1 and L2 voiced plosives.

The first set of analyses focused on mean VOT values in the production of English /b/ and /g/. The first analysis was concerned with English /b/. We obtained the mean VOT for each speaker and regressed it against a set of predictors. The chosen predictors were as follows: Measured English proficiency (i.e., the score resulting from the cloze test; range = 0–40), self-assessed English proficiency (i.e., the average of a given learner’s various self-assessed proficiency scores; range = 1–7), and self-assessed amount of English use (i.e., the average of a given learner’s various estimated usage scores: range = 0–100). These values, four per participant (one metric and three predictors), were submitted to a linear regression model. The overall fit of the model was poor, $R^2 = 0.16$, and the results yielded a series of null findings: Measured proficiency, $\beta = -0.21, t = -0.21, p > 0.05$ [0.84], self-assessed proficiency, $\beta = -1.29, t = -0.17, p > 0.05$ [0.87], and use, $\beta = -0.95, t = -1.49, p > 0.05$ [0.15]. The second analysis focused on English /g/. Mean VOT of /g/ was regressed against the proficiency and use predictors: Measured English proficiency, self-assessed English proficiency, and self-assessed amount of English use. Once again, the overall fit of the linear regression model was poor, $R^2 = 0.19$, and none of the predictors yielded a significant result: Measured proficiency, $\beta = -0.75, t = -0.71, p > 0.05$ [0.48], self-assessed proficiency, $\beta = -0.53, t = -0.07, p > 0.05$ [0.95], and use, $\beta = -0.97, t = -1.47, p > 0.05$ [0.15]. In sum, there is no evidence that English proficiency (neither measured nor self-assessed) or use affect VOT production in English words in EFL learners whose native language is Portuguese, at least not for /b/ or /g/.
The second set of analyses focused on the size of language mode effects, that is, on the size of the difference in mean VOT between English and Portuguese voiced plosives. For the first analysis in this group, we obtained, for each of the 28 EFL learners, the mean VOT difference between English (L2) and Portuguese (L1) /b/ productions. The mean difference values were regressed against three predictors: Measured proficiency, self-assessed proficiency, and self-assessed amount of English use. The overall fit of the linear regression model was extremely poor, \( R^2 = 0.09 \), and none of the factors were found to account for any significant amount of variance: Measured proficiency, \( \beta = -0.59, t = -0.67, p > 0.05 \) \[0.49\], self-assessed proficiency, \( \beta = 0.66, t = 0.1, p > 0.05 \) \[0.92\], and use, \( \beta = -0.51, t = -0.09, p > 0.05 \) \[0.37\]. The second analysis was concerned with the mean VOT difference between English (L2) and Portuguese (L1) /g/ productions. The same three predictors were used in a linear regression model. The model had a relatively poor fit, \( R^2 = 0.25 \), and none of the predictors were found to be significant: Measured proficiency, \( \beta = -1.22, t = -1.37, p > 0.05 \) \[0.18\], self-assessed proficiency, \( \beta = 5.16, t = 0.75, p > 0.05 \) \[0.45\], and use, \( \beta = -1.10, t = -1.95, p > 0.05 \) \[0.06\].

We found no evidence that either English proficiency or the amount of English use affected speech production in L1 Portuguese EFL learners. There was no evidence that the length of the prevoicing period of English voiced plosives is modulated by any of the experience indicators we used. The size of the difference between the length of the prevoicing period of Portuguese (L1) and English (L2) voiced plosives did not appear to change as a function of proficiency nor amount of L2 use.

4. Discussion

4.1. Summary of Findings

The present study focused on two main data comparisons and a secondary analysis. In all comparisons, the dependent variable was VOT (ms), and the target sound categories were, also in all cases, velar and bilabial voiced and voiceless stop consonants. The two main data comparisons were as follows. First, two groups of native Portuguese speakers—born, raised, and living in Brazil—produced the target sounds in their native language. One of the two groups comprised active EFL learners and the other, functional monolinguals, with no prior (substantial) exposure to English. The first comparison was concerned with contrasting the productions of the two speaker groups in their native language—a between-speakers comparison. Second, we obtained comparable L2 (English) production data from the EFL learners. Thus, the second main comparison was concerned with contrasting the L1 and L2 speech productions of this particular group—a within-speaker comparison. The secondary analysis, made possible by having recruited EFL learners of various English proficiency levels, explored the potential effects of EFL experience and proficiency on the speech productions of the EFL learners.

Were there phonetic differences between the Portuguese stops produced by L1 Portuguese EFL learners and those of Portuguese monolinguals? In our data, there were significant, albeit small, effects of speaker group. The voiceless plosives did not differ as a function of group, but the voiced ones did. In particular, the EFL learners were found to produce Portuguese /b/ and /g/ with longer prevoicing (negative VOT) than the monolinguals. While varying in VOT length, it is important to keep in mind that the average (and mode) Portuguese voiced plosive in both groups was prevoiced. Secondarily, VOT was found to be modulated by place of articulation and voicing, as one would expect (Lousada et al. 2010; Cho and Ladefoged 1999).

Was there evidence of the EFL learners having developed phonetic categories specific to English, that is, separate from those of their Portuguese? In our data, we found significant, albeit small, effects of language. The voiceless consonants were not modulated by language, but the voiced ones were. The English (L2) voiced plosives were found to present significantly shorter negative VOT (prevoicing) than the Portuguese (L1) consonants. Once again, we point out that, while varying in VOT length, the mode (and average) voiced plosive of a Brazilian EFL learner, in both Portuguese and English, is prevoiced. We
conducted a series of statistical analyses as a follow-up of this finding. We found that EFL learners produced a higher proportion of tokens with short-lag VOT in their L2 than in their L1. Prevoiced tokens were still in the majority. The difference in prevoicing length between L1 and L2 tokens remained even after discarding the short-lag VOT tokens—it remained in a data subset that included only truly prevoiced tokens. In other words, a difference in the proportion of tokens with or without prevoicing did not fully explain the language effects found initially with respect to prevoicing length.

The third analysis, a secondary one in the context of our study, was concerned with the possible effects of English proficiency (or experience) on pronunciation. Two analyses were conducted. First, English proficiency was not found to be associated with the variation in VOT measurements in the English data. Second, English proficiency was not found to be associated with the size of the language effect, that is, the size of the difference between English and Portuguese negative VOTs for the voiced plosives was not correlated with English proficiency. In sum, experienced EFL learners did not seem to differ systematically from inexperienced learners in terms of their VOTs. To be clear, EFL learners did differ from each other in their pronunciation patterns—they were not a homogeneous group. However, interlinear variation could not be explained with the meta information we gathered from our EFL learners.

4.2. Interpretation and Implications

4.2.1. L2 Phonetic Development

Do L1 Portuguese EFL learners develop phonetic categories specific to English sounds, that is, separate from those of their Portuguese sounds? Or, more broadly, do foreign-language learners, whose exposure to native L2 oral input is necessarily limited, form new sound categories specific to their L2? The answer is a nuanced “yes”. We found that, for the most part, L1 Portuguese EFL learners used similar, if not identical, phonetic categories for both of their languages. Firstly, there was no evidence of the formation of an aspirated category to be used in English voiceless plosives. EFL learners used ostensibly the same phonetic category, a short- to mid-lag VOT, for all voiceless plosives. Secondly, EFL learners also failed to develop a short-lag phonetic category to be used in English voiced plosives. The learners mostly produced prevoiced tokens both in their L1 and in their L2—that was their mode production pattern. The learners’ productions do not resemble those of native English speakers. The pronunciation of plosives was clearly modeled on the phonetics of their native language. However, and this is important, there were significant, albeit small, differences in the length (and proportion) of prevoicing as a function of language. This suggests that the EFL learners were employing two prevoiced phonetic targets, one for Portuguese voiced plosives and one for English voiced plosives.

Our results are in line with comparable research with foreign-language learners in classroom settings (Dmitrieva et al. 2020). Dmitrieva et al. (2020) found that native English speakers learning Russian in the US produced Russian voiceless plosives mostly with aspirated VOT and Russian voiced plosives mostly with short-lag VOT (Russian resembles Portuguese, and not English, in its use of VOT categories in voicing contrasts: Russian’s voiced plosives are prevoiced and voiceless plosives have short-lag VOT). The learners in the Dmitrieva et al. study modeled the phonetics of their L2 plosives on those of their L1. However, Dmitrieva et al. (2020) also found evidence of L2-specific pronunciation. For instance, one third of Russian voiced plosive tokens were prevoiced and the period of aspiration in the voiceless plosives was shorter in Russian than in English productions. Evidence of phonetic development in foreign-language speech consisted of small, but significant, subcategorical modifications of L1 sounds, like it did in our study.

To make sense of these findings, we make use of some of the basic principles of the SLM (Flege 1995; Flege and Bohn 2021; Flege et al. 2021). Other theoretical models, such as the L2LP (van Leussen and Escudero 2015; Escudero 2005), could be used as well—albeit with some modifications to our explanation. We rely on a single model for explanatory simplicity, as it is not our goal here to compare L2 speech models. Most
L2 speech researchers would agree that (emergent) bilinguals develop “equivalences” or representational connections between the sounds of their L1 and those in the L2 input they are exposed to, which they must mentally store (Simonet 2016). According to the SLM, speakers possess a single representational system for sounds—a common storage space for L1 and L2 phonetic categories (Flege 1995; Flege and Bohn 2021). Emergent bilinguals tend to categorize L2 sounds in the input they receive as a function of the sounds already in their system, thus utilizing mechanisms optimized for the processing of their L1 sounds. This process is typically referred to as “equivalence classification” (Flege 1987), and it arguably results in a warped representational space which, though containing both L1 and L2 sounds, is typically modeled after L1 sounds. The effects of equivalence classification are evident in the voiceless set in our study. We failed to find any difference in VOT between the L1 and L2 voiceless plosives of our EFL learners. It would seem that, in this learner population, English voiceless plosives are classified as instances of Portuguese voiceless plosives, blocking the potential development of an aspirated category specific to English. This is an instance of full equivalence classification, a process that results in a single L1/L2 phonetic category for voiceless plosives. What about voiced plosives? Since Portuguese voiced plosives are prototypically prevoiced, it is not surprising that the L1 Portuguese EFL learners in our study also prevoiced the English voiced plosives they produced. This is another instance of equivalence classification. However, in this case, learners did manage to develop two subcategories or two types of prevoiced plosives: One has very long prevoicing and is specific to Portuguese, and the other has shorter prevoicing and is used when speaking English. This seems to be an instance of partial equivalence classification, resulting in two subcategories. While EFL learners might use two different phonetic targets for their L1 and L2 voiced plosives, such phonetic targets are likely still associated in their mental representation—they are “equivalent” at some level.

Why is there full equivalence classification in the voiceless set but partial equivalence classification in the voiced set? Or, in other words, why is phonetic development restricted to the voiced set? At this juncture, we can only speculate, since we had not predicted this particular finding. To begin to address these questions, we would like to direct the reader’s attention to a body of findings on the malleability of VOT categories. At least three studies have examined the role of speech rate on VOT in a variety of languages (Kessinger and Blumstein 1997; Magloire and Green 1999; Beckman et al. 2011), and these studies elucidate some facts about VOT categories. It has been found, on the one hand, that aspirated and prevoiced categories are amenable to the effects of speech rate. Short-lag VOT categories, on the other hand, are not. Kessinger and Blumstein (1997) analyzed VOT data from voiced and voiceless plosives in three languages: French, English, and Thai. This research study found that the VOT of French voiced (but not voiceless) plosives is affected by speech rate—in their data, prevoicing was lengthened in slower speech. It was also found that the speech rate altered the VOT of English voiceless (but not voiced) plosives—the aspiration period was longer in slower speech. Finally, in Thai, which has a three-way contrast, speech rate was found to affect both aspirated and prevoiced plosives—slower speech lengthened both the prevoicing and aspiration periods. In all three languages, the short-lag VOT categories were unaffected by manipulations in speech rate. On a different, but related note, Tobin et al. (2017) found, in the speech of proficient Spanish–English bilinguals, an asymmetry between English (aspirated) and Spanish (short-lag VOT) voiceless plosives. In the Tobin et al. study, bilinguals’ productions in both of their languages were recorded in two separate sessions, one in an English-speaking country and one in a Spanish-speaking country. English /p t k/ differed between the two sessions—the aspiration period was shorter when Spanish was the ambient language—while Spanish /p t k/ did not. Interestingly, the authors noted that “the absence of an effect of ambient language on Spanish VOTs in this investigation suggests that the shorter VOTs may somehow be more stable and resistant to accommodation than longer VOTs” (Tobin et al. 2017, p. 52) They also note that this pattern has been documented in various other studies, even if it may not have been explicitly commented on (Antoniou et al. 2011; Chang 2012).
We postulate, based on these observations, that short-lag VOT categories are less malleable (or more resistant to modification) than both prevoiced and aspirated categories, and that this may account for the asymmetry in our findings. Perhaps short-lag VOT categories result from the in-phase coordination of two articulatory events, timed to occur simultaneously, while both prevoiced and aspirated categories result from anti-phase coordination, events timed to occur in a sequence (Browman and Goldstein 2000). In addition, perhaps sequential articulatory coordination is more malleable—more amenable to the development of subcategories in L2 speech—than simultaneous coordination. An alternative explanation makes use of distinctive features in phonology, privative features in particular (Beckman et al. 2011, 2013). Under this account, aspirated categories possess a [spread glottis] featural specification, and prevoiced categories have a [voice] featural specification. Short-lag VOT categories, on the other hand, lack any laryngeal featural specifications. This account has been used to explain the facts regarding speech rate discussed above (Beckman et al. 2011): Categories with a featural specification (but not those lacking one) are affected by speech rate. A possible extension of this account to our findings is this: L2 learners may be more likely to develop phonetic subcategories specific to their L2 for sounds that have an explicit featural specification (aspirated and prevoiced categories, but not short-lag VOT categories), whereas unspecified sounds could be more resistant to the formation of subcategories.

An anonymous reviewer notes that asymmetrical behavior between voiced and voiceless plosives had already been observed in the phonological literature on L2 acquisition and L1 drift situations (Schwartz 2020). Several studies have found cross-linguistic phonetic influence in the voiced (but not the voiceless) set in bilingual speakers whose two languages differ in terms of how they implement the laryngeal contrast in the plosives—that is, people who speak both an “aspirating” language, such as English, and “true voicing” language, such as Portuguese (e.g., Gabriel et al. 2018; Kang et al. 2016; Podlipský et al. 2020; see Schwartz 2020 for a review). It seems that such bilinguals are more likely to assimilate (or cross-linguistically equate) prevoiced and underlyingly voiced short-lag VOT plosives than they are to assimilate underlyingly voiceless short-lag VOT and aspirated plosives. In other words, the plosives in the voiced set seem to be cross-linguistically more similar (in the bilinguals’ behavior) to each other than those in the voiceless set. Schwartz (2020) speculated that this asymmetry revealed a phonological similarity between prevoiced and underlyingly voiced short-lag VOT plosives that does not exist between aspirated and underlyingly voiceless short-lag VOT plosives. According to Schwartz, plosives have three featural levels or nodes of representation: Closure, voice, and vocalic onset. Aspirated plosives have the featural specification [fortis] at all three levels; underlyingly voiceless plain (i.e., short-lag VOT) plosives, on the other hand, receive the featural specification [fortis] only at the voice onset level; and both prevoiced and underlyingly voiced plain plosives lack any featural specification. This phonological account captures the idea that underlyingly voiced and prevoiced plosives are phonologically identical, whereas the two plosives in the voiceless set are phonologically different. In Schwartz’s (2020) account, differences in phonological representation explain the asymmetrical patterns of cross-linguistic influence in the plosives produced by bilinguals. Note, however, that we have found that the Brazilian learners of English in our study produced a single VOT category for both their Portuguese (L1) and their English (L2) voiceless plosives, whereas the voice consonants were cross-linguistically different (very similar to each other, but still different). Our data suggest that cross-linguistic assimilation or equivalence classification could actually be stronger in the voiceless set than in the voiced set. It is thus not clear whether our results are in line with Schwartz’s (2020) observation or not. This issue requires further research. What seems to be important is that asymmetrical cross-linguistic behavior between phonologically voiced and voiceless plosives seems to be found recurrently in studies of bilingual phonetic behavior (Schwartz 2020), and our data come to corroborate this observation.
The published research study most similar to ours is that of Dmitrieva et al. (2020). Their results were comparable to ours, but there is one aspect of Dmitrieva et al. learners’ experience with their L2 that differs from our learners’ experience with English. All of the Russian learners in the Dmitrieva et al. sample had exclusively attended language classes taught by native Russian speakers. On the other hand, only about half the EFL learners in our sample had ever had a native English-speaking teacher—never consistently. The role of input in L2 speech has been considered important for some time (Flege 2008), but input continues to be notoriously difficult to assess, which might have encouraged researchers to focus on other things, such as age of acquisition (Flege 2018). At this juncture, we must at least mention the possibility that the EFL learners in our sample may have never received enough “normative” English input to be able to develop native-like, L2-specific phonetic categories. The English input some of our learners received was undoubtedly sufficient to develop new grammatical and lexical norms—in fact, some of the learners in our sample were relatively proficient in English. However, if such input was “accented” (in that it was produced by other native Portuguese speakers, EFL learners themselves) it may have never comprised a sizable amount of aspirated voiceless plosives or short-lag voiced plosives. In the absence of large amounts of such input, how could we expect that foreign-language learners would ever develop native-like phonetic targets in their L2? What seems to be worth mentioning, though, is that the EFL learners in our sample did develop L2-specific phonetic targets for one of the consonant sets investigated, even if such targets were modeled on L1 sounds and very different from target L2 norms. It is not surprising that the EFL learners did not “sound like” native English speakers, but it is interesting that some phonetic development took place, even if only on the margins. L2 immersion may be needed for nativelikeness, but our data suggest that immersion is not crucial at the initial stages of phonetic-category formation (Dmitrieva et al. 2020).

4.2.2. L1 Phonetic Drift

Are there phonetic differences between the Portuguese plosives produced by L1 Portuguese EFL learners and that of Portuguese monolinguals? Or, less narrowly, does engaging in the learning of a foreign language affect the phonetics of one’s native language? The answer, once again, is a nuanced “yes”. We found a phonetic difference between monolingual speakers of Portuguese and EFL learners. The difference concerned the length of the prevoicing period in the voiced plosives. The EFL learners had a longer prevoicing period than the monolinguals. This seems to be evidence of L1 phonetic drift, a modification to one’s native pronunciation resulting from foreign-language learning. Evidence of this sort was also found in a comparable study (Dmitrieva et al. 2020). In our study, the voiceless consonants were not at all affected, and even the voiced consonants were affected only marginally—prevoicing was the norm for both groups of Portuguese speakers. Still, a significant difference between EFL learners and monolinguals was found. Furthermore, there is an apparent connection between our findings concerning L2 development and our findings concerning L1 drift. Since it is the same set of plosives, that is, voiced plosives, that showed evidence of L2 phonetic development (i.e., the formation of a new subcategory) and L1 phonetic drift, it is reasonable to postulate that the two findings are connected.

We speculate that, to “make room” for a new prevoiced plosive specific to English, L1 Portuguese EFL learners shortened the prevoicing period of the English voiced plosive and lengthened the prevoicing period of their native voiced plosive. This seems to be an instance of dissimilation, one of the possible epiphenomena, according to the SLM, of new-category formation (Flege et al. 2003; MacKay et al. 2001; Simonet 2011; Flege and Eefting 1987). Category dissimilation or deflection has been found in other studies, but only in early, proficient bilinguals. For instance, Flege and Eefting (1987) found that Spanish/English bilinguals’ Spanish voiceless stops had a shorter VOT period than those of Spanish monolinguals, likely since bilinguals also possessed a long-lag VOT category for their English voiceless stops. Since, in our data, the formation of a new subcategory specific to the L2 results in (or co-occurs with) a modification of the corresponding L1 category, we
conclude that a reorganization, in the form of internal deflection, of the phonetic system of the EFL learners resulted in an instance of L1 phonetic drift. Therefore, it seems that dissimilation is also possible in late L2 learners.

Dmitrieva et al. (2020) found evidence of L1 phonetic drift associated with L2 phonetic development, as we did, but in their study L1 phonetic drift was assimilatory, not dissimilatory—learners modified their L1 phonetic categories in the direction of assimilation to the corresponding L2 category rather than away from it. The Dmitrieva et al. findings are relatively unique in that evidence for both L1 drift and L2 phonetic learning were found for classroom learners in a foreign-language setting (see Chang 2012; Kartushina et al. 2016a). Findings comparable to these had been reported for advanced learners in immersion (or immigrant) settings, as reviewed in the Introduction (de Leeuw et al. 2010, 2018; Baker and Trofimovich 2005; Tobin et al. 2017). Our findings are in line with the Dmitrieva et al. study and add support to their observation that foreign-language learners, who have only very limited exposure to native English oral input, may also undergo L1 phonetic drift. L1 drift may (or may not) be the results of a ‘novelty effect’ (Chang 2012). In our data, L1 drift did not seem to be associated with English experience—it was, therefore, not exclusive to the production of novice learners. As mentioned, what seems to be new about our findings—different from Dmitrieva et al., for instance—is that we documented the existence of dissimilatory (rather than assimilatory) L1 phonetic drift in late L2 learners who seldom use their L2. Kartushina et al. (2016b), in their recent review of the literature, conclude that, in late L2 learners, “the production of L1 categories seems to be unaffected, that is, the L1 categories remain unchanged” (p. 168). They also attribute dissimilation exclusively to early learners. Our findings suggest that the Kartushina et al. statements are a simplification of the facts. L1 categories may be affected by L2 speech learning even in late learners who continue to be dominant in their L1, and dissimilation is also possible in this population. Perhaps a more accurate observation is that L1 drift of a dissimilatory nature depends on the formation of L2-specific sound categories. In other words, dissimilation is a possible result of L2-specific category formation. If a learner, including a late learner, forms a new category specific to their L2, dissimilation is a possible aftereffect.

An anonymous reviewer points out that our EFL learners provided the Portuguese speech data immediately after they had provided the English data. One could postulate that what triggered the lengthening of prevoicing in the Portuguese voiced plosives relative to the English prevoiced tokens were carry-over effects from the speakers’ having participated in the English experimental block immediately before providing the Portuguese data, rather than bona fide cross-linguistic influence from the L2 on the L1. Grosjean (2011) distinguishes between two types of cross-linguistic influence, transfer and interference. Transfer is the permanent, static influence of one language’s features on the other—an influence at the level of linguistic competence or long-term representation. Interference, on the other hand, is the ephemeral influence or temporary intrusion of a feature of one language on the other, an influence exclusively at the level of performance or processing. Research has shown that the effects of (dynamic) interference may go beyond those of transfer and are not necessarily affected by bilingual language dominance (Simonet 2014; Simonet and Amengual 2019). Are the L1 drift effects we captured in our current study the result of transfer or of interference? We are afraid we cannot answer this question here, and we must acknowledge that our experimental design results in the presence of a confound. Only future research may resolve this issue. The effects of L2 phonetic development on L1 phonetic drift we have captured in our study could be the result of transfer, interference or both (Grosjean 2011; Simonet 2014). At any rate, we note that both transfer and interference are forms of cross-linguistic influence.

We conclude that, specifically through partial equivalence classification (i.e., the formation of new subcategories specific to the L2), L2 phonetic development leads to L1 phonetic drift, and it may do so even for foreign-language learners in classroom settings, who have only minimal chances to receive native input in their target language. Our findings are in line with Dmitrieva et al. (2020), among others (e.g., Schwartz 2020). Both
sets of findings, Dmitrieva et al. and ours, suggest the existence of a connection between partial equivalence classification and L1 drift. Our findings are novel in that they show that L1 drift may be dissipulatory in nature even in L2 learners who continue to use their L1 daily. L2-specific new-category formation and concomitant dissipatory L1 phonetic drift are not an exclusive property of life-long bilinguals living in an L2 immersion setting.

5. Conclusions

Fifty-six native speakers of Portuguese produced Portuguese words beginning in one of four plosives, /p b k g/. There were two main groups of Portuguese speakers in the sample. One group comprised monolingual speakers and the other, learners of English as a foreign language. The learners (but not the monolinguals) were also asked to produce English words beginning in one of four plosives, /p b k g/. We measured the VOT of all target word-initial stops. We found that the learners produced a single VOT category for voiceless plosives in both Portuguese and English but two subcategories of voiced plosives, one in Portuguese and one in English. While voiceless stops were prevoiced in both languages, the English stops had, on average, a shorter period of prevoicing than the Portuguese stops. This was interpreted as evidence of the learners having formed a phonetic subcategory specific to the L2. We also found that the two groups of Portuguese speakers differed from each other precisely in the duration of the prevoicing period in the voiced plosives. The English learners produced voiced Portuguese plosives with a longer period of prevoicing than the monolinguals. Their voiceless plosives did not differ. This was interpreted as evidence of dissipatory L1 phonetic drift. We postulated that the English learners had been able to develop a L2-specific subcategory for their English voiced consonants by (also) dissipating their Portuguese voiced consonant from such new subcategory. Such interlingual dissimilation resulted in L1 drift. L2 phonetic development (new-category formation) and L1 drift are possible in foreign language learning, even when exposure to target-language native phonetic norms is severely restricted.

Author Contributions: Conceptualization, D.M.O. and M.S.; Data curation, D.M.O.; Formal analysis, M.S.; Funding acquisition, D.M.O.; Investigation, D.M.O. and M.S.; Methodology, D.M.O. and M.S.; Writing—original draft, M.S. All authors have read and agree to the published version of the manuscript.

Funding: The authors acknowledge the receipt of funds to purchase equipment (digital recorder and microphone, $839.98) from the Graduate and Professional Student Council, Research and Project (REaP) Grant, of the University of Arizona. The grant was awarded to D.M.O. in 2018.

Institutional Review Board Statement: The study was conducted according to the guidelines of the Declaration of Helsinki and approved by the Institutional Review Board (Human Subjects Protection Program) of the University of Arizona (Protocol UAR 1404294632, approved on 18 April 2014).

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: Synthetic data are available from the corresponding author on request. The data are not publicly available because the authors failed to ask participant permission to publicly share their data. The code to reproduce all analyses is also available on request.

Conflicts of Interest: The authors declare no conflict of interest.

References

Abramson, Arthur, and Douglass Whalen. 2017. Voice Onset Time (VOT) at 50: Theoretical and Practical Issues in Measuring Voicing Distinctions. Journal of Phonetics 63: 75–86. [CrossRef]

Albano, Eleonora. 2001. O Gesto e Suas Bordas: Esboço de Fonologia Acústico-Articulatória Do Português Brasileiro. Campinas: Mercado de Letras.

Alves, Mariane, Izabel Seara, Fernando Pacheco, Simone Klein, and Rui Seara. 2008. On the Voiceless Aspirated Stops in Brazilian Portuguese. In Computational Processing of the Portuguese Language. Berlin: Springer, pp. 248–51.

Amengual, Mark. 2012. Interlingual Influence in Bilingual Speech: Cognate Status Effect in a Continuum of Bilingualism. Bilingualism: Language and Cognition 15: 517–30. [CrossRef]
Antoniou, Mark, Catherine Best, Michael Tyler, and Christian Kroos. 2011. Inter-Language Interference in VOT Production by L2-Dominant Bilinguals: Asymmetries in Phonetic Code-Switching. *Journal of Phonetics* 39: 558–70. [CrossRef] [PubMed]

Baker, Wendy, and Pavel Trofimovich. 2005. Interaction of Native- and Second-Language Vowel System(s) in Early and Late Bilinguals. *Language and Speech* 48: 1–27. [CrossRef] [PubMed]

Barboza, Plinio, and Elenora Albano. 2004. Brazilian Portuguese. *IIPA* 34: 227–32. [CrossRef]

Beckman, Jill, Péter Helgason, Bob McMurray, and Catherine Ringen. 2011. Rate Effects on Swedish VOT: Evidence for Phonological Overspecification. *Journal of Phonetics* 39: 39–49. [CrossRef]

Beckman, Jill, Michael Jessen, and Catherine Ringen. 2013. Empirical Evidence for Laryngeal Features: Aspirating vs. True Voice Languages. *Journal of Linguistics* 49: 259–84. [CrossRef]

Bergmann, Christopher, Amber Nota, Simone Sprenger, and Monika Schmid. 2016. L2 Immersion Causes Non-Native-like L1 Pronunciation in German Attriters. *Journal of Phonetics* 58: 71–86. [CrossRef]

Boersma, Paul. 2001. Praat, a System for Doing Phonetics by Computer. *Glot International* 5: 341–45.

British Council. 2019. Políticas Públicas Para o Ensino de Inglês: Um Panorama Das Experiências Na Rede Pública. Available online: https://www.britishcouncil.org.br/sites/default/files/bncc_portuguesbx.pdf (accessed on 22 June 2021).

Browman, Catherine, and Louis Goldstein. 2000. Competing Constraints on Intergestural Coordination and Self-Organization of Phonological Structures. *Bulletin de La Communication Parlée* 5: 25–34.

Brown, James. 2002. Do Cloze Tests Work? Or Is It Just an Illusion? *Second Language Studies* 21: 79–125.

Casillas, Joseph. 2021. Interlingual Interactions Elicit Performance Mismatches Not ‘Compromise’ Categories in Early Bilinguals: Evidence from Meta-Analysis and Stop. *Languages* 6: 9. [CrossRef]

Casillas, Joseph, and Miquel Simonet. 2018. Perceptual Categorization and Bilingual Language Modes: Assessing the Double Phonemic Boundary in Early and Late Bilinguals. *Journal of Phonetics* 71: 51–64. [CrossRef]

Chang, Charles. 2012. Rapid and Multifaceted Effects of Second-Language Learning on First-Language Speech Production. *Journal of Phonetics* 40: 249–68. [CrossRef]

Chang, Charles. 2013. A Novelty Effect in Phonetic Drift of the Native Language. *Journal of Phonetics* 41: 520–33. [CrossRef]

Chang, Charles. 2019. The Phonetics of Second Language Learning and Bilingualism. In *The Routledge Handbook of Phonetics*. Edited by William Katz and Peter Assmann. Abingdon: Routledge, pp. 427–47.

Cho, Taehong, and Peter Ladefoged. 1999. Variation and Universals in VOT. *Journal of Phonetics* 27: 207–29. [CrossRef]

Colantoni, Laura, Jeffrey Steele, and Paola Escudero. 2015. *Second Language Speech: Theory and Practice*. Cambridge: Cambridge University Press.

Cumming, Geoff. 2013. *Understanding the New Statistics: Effect Sizes, Confidence Intervals, and Meta-Analysis*. London: Routledge.

de Leeuw, Esther, Monika Schmid, and Ineke Mennen. 2010. The Effects of Contact on Native Language Pronunciation in an L2 Migrant Setting. *Bilingualism: Language and Cognition* 13: 33–40. [CrossRef]

de Leeuw, Esther, Aurela Tusha, and Monika Schmid. 2018. Individual Phonological Attrition in Albanian-English Late Bilinguals. *Bilingualism: Language and Cognition* 21: 278–95. [CrossRef]

Dmitrieva, Olga, Allard Jongman, and Joan Sereno. 2020. The Effect of Instructed Second Language Learning on the Acoustic Properties of First Language Speech. *Languages* 5: 44. [CrossRef]

Escudero, Paola. 2005. *Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization*. LOT Dissertation Series 113. Utrecht: Utrecht University.

Feige, James. 1987. The Production of ‘New’ and ‘Similar’ Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics* 15: 47–65.

Feige, James. 1991. Age of Learning Affects the Authenticity of Voice-Onset Time (VOT) in Stop Consonants Produced in a Second Language. *The Journal of the Acoustical Society of America* 89: 395–411. [CrossRef] [PubMed]

Feige, James. 1995. Second Language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winnifred Strange. Timonium: York Press, pp. 229–73.

Feige, James. 2007. Language Contact in Bilingualism: Phonetic System Interactions. In *Laboratory Phonology* 9. Edited by Jennifer Cole and José Hualde. Berlin: Mouton de Gruyter, pp. 353–80.

Feige, James. 2008. Give Input a Chance! In *Input Matters in SLA*. Edited by Thorsten Piske and Martha Young-Scholten. Bristol: Multilingual Matters, pp. 175–90.

Feige, James. 2018. A Non-Critical Period for Second-Language Learning. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne Nyvad, MMichaela Hejna, Anders Hojen, Anna Jespersen and Mette Sorensen. Aarhus: Aarhus University, pp. 501–41.

Feige, James, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, pp. 3–83.

Feige, James, and Wieke Eefting. 1987. Production and Perception of English Stops by Native Spanish Speakers. *Journal of Phonetics* 15: 67–83. [CrossRef]

Feige, James, Naoyuki Takagi, and Virginia Mann. 1995. Japanese Adults Can Learn to Produce English /sl/ and /l/ Accurately. *Language and Speech* 38: 25–55. [CrossRef]
Flege, James, Elaina Frieda, and Takeshi Nozawa. 1997. Amount of Native-Language (L1) Use Affects the Pronunciation of an L2. *Journal of Phonetics* 25: 169–86. [CrossRef]

Flege, James, Carlo Schirru, and Ian MacKay. 2003. Interaction between the Native and Second Language Phonetic Subsystems. *Speech Communication* 40: 467–91. [CrossRef]

Flege, James, David Birdsong, Ellen Bialystok, Molly Mack, Hyekyung Sung, and Kimiko Tsukada. 2006. Degree of Foreign Accent in English Sentences Produced by Korean Children and Adults. *Journal of Phonetics* 34: 153–75. [CrossRef]

Flege, James, Katsura Aoyama, and Ocke-Schwen Bohn. 2021. The Revised Speech Learning Model (SLM-r) Applied. In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by Ratree Wayland. Cambridge: Cambridge University Press, pp. 84–118.

Fowler, Carol, Valery Sramko, David Ostry, Sarah Rowland, and Pierre Hallé. 2008. Cross Language Phonetic Influences on the Speech of French-English Bilinguals. *Journal of Phonetics* 36: 649–63. [CrossRef]

Guion, Susan. 2003. The Vowel Systems of Quichua-Spanish Bilinguals. *Phonetica* 60: 98–128. [CrossRef]

Hopp, Holger, and Monika Schmid. 2013. Perceived Foreign Accent in First Language Attrition and Second Language Acquisition: The Impact of Age of Acquisition and Bilingualism. *Applied Psycholinguistics* 34: 361–94. [CrossRef]

Kang, Yoonjung, Sneha George, and Rachel Soo. 2016. Cross-Language Influence in the Stop Voicing Contrast in Heritage Tagalog. *Heritage Language Journal* 13: 184–218. [CrossRef]

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Frauenfelder, and Narli Golestani. 2016a. Mutual Influences between Native and Non-Native Vowels in Production: Evidence from Short-Term Visual Articulatory Feedback Training. *Journal of Phonetics* 57: 21–39. [CrossRef]

Kartushina, Natalia, Ulrich Frauenfelder, and Narli Golestani. 2016b. How and When Does the Second Language Influence the Production of Native Speech Sounds: A Literature Review. *Language Learning* 66: 155–86. [CrossRef]

Kessinger, Rachel, and Sheila Blumstein. 1997. Effects of Speaking Rate on Voice-Onset Time in Thai, French, and English. *Journal of Phonetics* 25: 143–68. [CrossRef]

Kirby, James, and Dwight Robert Ladd. 2016. Effects of Obstruent Voicing on Vowel F0: Evidence from ‘True Voicing’ Languages. *The Journal of the Acoustical Society of America* 140: 2400–2411. [CrossRef]

Köpke, Barbara, and Monika Schmid. 2004. Language Attrition: The next Phase. In *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*. Edited by Monika Schmid, Barbara Köpke, Merel Keijzer and Lina Weilemar. Studies in Bilingualism. Amsterdam: John Benjamins, pp. 1–43.

Lenth, Russell. 2018. Emmeans: Estimated Marginal Means, Aka Least-Squares Means (Version 1.4.7). Available online: https://cran.r-project.org/package=emmeans (accessed on 22 June 2021).

Lisker, Leigh, and Arthur Abramson. 1964. A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word* 20: 384–422. [CrossRef]

Lisker, Leigh, and Arthur Abramson. 1967. Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech* 10: 1–28. [CrossRef] [PubMed]

Lousada, Marisa, Luis Jesus, and Andrea Hall. 2010. Temporal Acoustic Correlates of the Voicing Contrast in European Portuguese Stops. *Journal of the International Phonetic Association* 40: 261–75. [CrossRef]

MacKay, Ian, James Flege, Thorsten Piske, and Carlo Schirru. 2001. Category Restructuring during Second-Language Speech Acquisition. *The Journal of the Acoustical Society of America* 110: 516–28. [CrossRef]

Magloire, Joel, and K. P. Green. 1999. A Cross-Language Comparison of Speaking Rate Effects on the Production of Voice Onset Time in English and Spanish. *Phonetica* 56: 158–85. [CrossRef]

Major, Roy. 1987. English Voiceless Stop Production by Speakers of Brazilian Portuguese. *Journal of Phonetics* 15: 197–202. [CrossRef]

Major, Roy. 1992. Losing English as a First Language. *The Modern Language Journal* 76: 190–208. [CrossRef]

Mayr, Robert, David Sánchez, and Ineke Mennen. 2020. Does Teaching Your Native Language Abroad Increase L1 Attrition of Speech? The Case of Spaniards in the United Kingdom. *Languages* 5: 41. [CrossRef]

Mora, Joan, and Marianna Nadeu. 2012. L2 Effects on the Perception and Production of a Native Vowel Contrast in Early Bilinguals. *International Journal of Bilingualism* 16: 484–500. [CrossRef]

Nearey, Terrance, and Bernard Rochet. 1994. Effects of Place of Articulation and Vowel Context on VOT Production and Perception for French and English Stops. *Journal of the International Phonetic Association* 24: 1–18. [CrossRef]

Peirce, Jonathan. 2007. PsychoPy—Psychophysics Software in Python. *Journal of Neuroscience Methods* 162: 8–13. [CrossRef]

Peirce, Jonathan, Jeremy Gray, Sol Simpson, Michael MacAskill, Richard Höchenberger, Hiroyuki Sogo, Erik Kastman, and Jonas Lindelöv. 2019. PsychoPy2: Experiments in Behavior Made Easy. *Behavior Research Methods* 51: 195–203. [CrossRef]

---

**Note:** The list includes citations for various research articles and books that discuss topics related to language acquisition, phonetics, and bilingualism. These references are typical of those found in academic journals and books in the fields of linguistics and applied linguistics.
Piske, Thorsten, Ian MacKay, and James Flege. 2001. Factors Affecting Degree of Foreign Accent in an L2: A Review. *Journal of Phonetics* 29: 191–215. [CrossRef]

Podlipský, Václav, Sára Simáčková, and Kateřina Chládková. 2020. Phonetic Drift Reveals Interconnected Phonological Representations in Simultaneous Bilinguals: A Case Study of English and Czech Stop Consonants. *International Journal of Bilingualism* 25: 789–99. [CrossRef]

Polidório, Valdomiro. 2014. O Ensino de Língua Inglesa No Brasil. The English Teaching in Brazil. *Revista Tracessias* 8: 340–46. 

R Core Team. 2018. *R: A Language and Environment for Statistical Computing (Version 4.0.1).* Vienna: R Core Team. Available online: https://cran.r-project.org (accessed on 22 June 2021).

Sancier, Michele, and Carol Fowler. 1997. Gestural Drift in a Bilingual Speaker of Brazilian Portuguese and English. *Journal of Phonetics* 25: 421–36. [CrossRef]

Santos, Eliana S. S. 2011. O Ensino Da Língua Inglesa No Brasil. BABEL: Revista Eletrônica de Línguas e Literaturas Estrangeiras 1: 1–7.

Schmidt, Monika. 2011. *Language Attirnation*. Cambridge: Cambridge University Press.

Schwartz, Geoffrey. 2020. Asymmetrical Cross-Language Phonetic Interaction: Phonological Implications. *Linguistic Approaches to Bilingualism*. [CrossRef]

Silva, Kleber. 2004. ‘Inglês Só Se Aprende Na Escola de Idiomas’: Importantes Reflexões No Que Concerne Ao Ensino/Aprendizagem de Língua Inglesa. Paper presented at VII Congresso Brasileiro de Linguística Aplicada, São Paulo, Brazil, October 10–14.

Simonet, Miquel. 2011. Production of a Catalan-Specific Vowel Contrast by Early Spanish-Catalan Bilinguals. *Phonetics* 68: 88–110. [CrossRef]

Simonet, Miquel. 2014. Phonetic Consequences of Dynamic Cross-Linguistic Influence in Proficient Bilinguals. *Journal of Phonetics* 43: 26–37. [CrossRef]

Simonet, Miquel. 2016. The Phonetics and Phonology of Bilingualism. In *Oxford Handbooks Online: Scholarly Research Reviews*. Oxford: Oxford University Press, pp. 1–25.

Simonet, Miquel, and Mark Amengual. 2019. Increased Language Co-Activation Leads to Enhanced Cross-Linguistic Phonetic Convergence. *International Journal of Bilingualism* 24: 208–21. [CrossRef]

Singman, Henrik, Ben Bolker, Jake Westfall, Frederik Aust, and Mattan Ben-Shachar. 2020. *Afex: Analysis of Factorial Experiments* (Version 0.27-2). Available online: https://cran.r-project.org/package=afex (accessed on 22 June 2021).

Solon, Megan. 2016. Do Learners Lighten Up? *Studies in Second Language Acquisition* 39: 801–32. [CrossRef]

Stoehr, Antje, Titia Benders, Janet van Hell, and Paula Fikkert. 2017. Second Language Attainment and First Language Attrition: The Case of VOT in Immersed Dutch–German Late Bilinguals. *Second Language Research* 33: 483–518. [CrossRef] [PubMed]

Tavares, Ana C. B. 2018. Como Ensinar Cultura: A Abordagem de Aspectos Culturais No Ensino de Língua Inglesa Sob a Ótica Do Professor Brasileiro. Unpublished Ph.D. dissertation, Universidade do Porto, Porto, Portugal.

The Jamovi Project. 2020. Jamovi (Version 1.2.22). Available online: https://www.jamovi.org (accessed on 22 June 2021).

Tobin, Stephen, Hosung Nam, and Carol Fowler. 2017. Phonetic Drift in Spanish-English Bilinguals: Experiment and a Self-Organizing Model. *Journal of Phonetics* 65: 45–59. [CrossRef]

Tremblay, Annie. 2011. Proficiency Assessment Standards in Second Language Acquisition Research: ‘Clozing’ the Gap. *Studies in Second Language Acquisition* 33: 339–72. [CrossRef]

Tsukada, Kimiko, David Birdsong, Molly Mack, Hyekyung Sung, Ellen Bialystok, and James Flege. 2004. Release Bursts in English Word-Final Voiceless Stops Produced by Native English and Korean Adults and Children. *Phonetics* 61: 67–83. [CrossRef] [PubMed]

Ulbrich, Christiane, and Mikhail Ordin. 2014. Can L2-English Influence L1-German? The Case of Post-Vocalic /r/. *Journal of Phonetics* 45: 26–42. [CrossRef]

van Leussen, Jan-Willem, and Paola Escudero. 2015. Learning to Perceive and Recognize a Second Language: The L2LP Model Revised. *Frontiers in Psychology* 6: 1–12. [CrossRef] [PubMed]

Wayland, Ratree, ed. 2021. *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge: Cambridge University Press.

Wickham, Hadley, Mara Averick, Jennifer Bryan, Winston Chang, Lucy D’Agostino McGowan, Romain François, Garrett Grolemund, Alex Hayes, Lionel Henry, Jim Hester, and et al. 2019. Welcome to the Tidyverse. *Journal of Open Source Software* 4: 1686. [CrossRef]

Yavaş, Mehmet, and Renée Wildermuth. 2006. The Effects of Place of Articulation and Vowel Height in the Acquisition of English Aspirated Stops by Spanish Speakers. *International Review of Applied Linguistics in Language Teaching* 44: 251–63. [CrossRef]