The effect of foreign language experience on the categorization and production of native and non-native stops by Spanish learners of English

MA in Advanced English Studies: Multilingualism and Acquisition of English

Supervisor: Juli Cebrian Puyuelo

Celia Gorba Masip

July 2016

Universitat Autònoma de Barcelona

Acknowledgments

First of all, I would like to express my very great appreciation to my supervisor Dr. Juli Cerbrian for his guidance and support during the development of my Master's thesis. My special thanks are extended to Professor Maria Josep Solé, who has been of great help in the statistical design and analysis of my TFM. I would also like to show my gratitude to Dr. Núria Gavaldà, and Dr. Joan Carles Mora for their valuable advice, and Dr. María Machuca for her help in the recruitment of participants. Finally, I would like to thank my family and friends for their support and patience.

Table of contents

Index of Tables and Figures

Tables

Figures

Tables in Appendix C

Figures in Appendix C

Abstract: Previous research has assessed the effect of L1 experience on the categorization and production of L2 stops. However, the effect of L2 experience on the L1 has received much less attention. In addition, no previous studies have tested the same population on both perception and production and using modified natural stimuli. The present paper attempts to determine the effect of L2 experience on the perception and production of L1 Spanish and L2 English bilabial stops. A categorization task as well as a production task in each language was completed by experienced and inexperienced Spanish learners of English. Experienced learners were found to perceive and produce L2 bilabial stops more accurately than inexperienced learners, although such difference reached significance only in perception. As for L2 influence on the L1, experienced learners differed from inexperienced learners in the Spanish identification task. Nevertheless, no significant effect of experience was found in production. Moreover, experienced learners were found to categorize bilabials differently in each language, whereas inexperienced learners were not. Further, both groups seemed to produce L1 and L2 /p/ differently yet English and Spanish /b/ were produced with Spanish-like values. Finally, no relation between individual perception and production was observed.

1. Introduction

Voiced and voiceless stops in English differ from their Spanish counterparts in terms of VOT values. Spanish utterance-initial voiced stops are produced with voice-lead, whereas English utterance-initial voiced stops often present VOT values around zero However, instances of prevoicing in English have been found in previous research (Lisker and Abramson, 1964; Rosner, López-Bascuas, García-Albea and Fahey, 2000; Williams, 1977)[1]. In the case of utterance-initial voiceless stops in Spanish, voicing begins upon release or immediately thereafter. By contrast, English utterance-initial voiceless stops tend to be produced with voicing lag. Consequently, perceptual crossovers of utterance-initial stops in English present higher VOT values than in Spanish for all places of articulation (Abramson and Lisker 1973, Lisker and Abramson 1964).

A number of studies have tested crosslinguistic production and categorization of stops (Abramson and Lisker, 1973; Antoniou, Best, Tyler and Kross, 2010, 2011; Flege, 1987; Flege and Eefting, 1987; Lisker and Abramson, 1964; Lisker, Liberman, Erickson, Dechovitz and Mandler, 1977; Williams, 1977). For instance, Flege (1987) looked at the production of similar phones – including stops – in French and English, and assessed the effects of L1 on L2 and L2 on L1 as a result of L2 experience. Still, he based his research on production only. Williams (1977) considered both perception and production in Spanish and English, but he did not use natural speech to create the stimuli. To my knowledge, no previous work has looked at both production and perception of Spanish and English stops by the same population and using modified natural speech.

[1] See section 2.2. for a detailed description of Spanish and English VOT values.

In the next few pages, the present paper provides a review of the main literature that is relevant to the current study, including theoretical models concerning the acquisition of L2 phonology, VOT as the main cue for voicing regarding stops, other cues for voicing, the effect of language experience and the role of language mode. After that, the research questions addressed in this paper and their corresponding hypotheses are introduced. This is followed by a detailed methodology section that describes the design of the identification tests, as well as the production elicitation task. Then, the results of the perception experiments for each language and of the within group analysis are presented and discussed. The results of the production experiment and the corresponding discussion are presented next, and the relation between perception and production results is examined. After that, a general discussion and conclusions section relates the findings discussed in the previous chapters in light of the research questions of this study. Finally, the paper's last section acknowledges the limitations of the study, and points out possible issues for further research.

2. Theoretical framework

2.1. The acquisition of non-native phones: a theoretical approach

The present study is in line with previous theoretical approaches to L2 acquisition. A number of works have addressed L2 category formation, and the influence of the L1 on the acquisition of L2 phones (Flege, 1995; Flege, 2002; Best, 1995; Best and Tyler, 2007). Flege (1995) proposed the Speech Learning Model (SLM), which posits that L2 learners may fail to distinguish phonetic features of the L2 due to differences with their L1. According to SLM, the failure to authentically categorize distinct L2 phones may stem from two different reasons: (1) the assimilation of both sounds to a single L1 category, and (2) the inability to perceive L2 features that are not phonologically relevant in the L1. More recently, interactions between the L1 and the L2 have been explained on the basis of 'equivalence classification', which posits 'category assimilation' and 'category formation' (Flege, 2002). On the one hand, category assimilation is in line with SLM's view on the assimilation of an L2 to sound to an L1 existing category, that is the L1 and the L2 categories become more similar and new category formation is blocked. For instance, a Spanish-English bilingual may produce English voiceless stops with shorter VOT than monolinguals and Spanish voiceless stops with longer VOT than monolinguals. On the other hand, category dissimilation stands for the creation of an L2 category and the increase of the L1-L2 phonetic contrast – e.g. Spanish-English bilinguals may produce Spanish voiceless stops with shorter VOT than monolinguals and English voiceless stops with longer VOT than English monolinguals.

It appears, thus, that bilinguals may encounter difficulty in perceiving differences between the L1 and the L2 when it comes to phones that share the same phonological space (Flege, 1995). In this regard, we could expect Spanish-English bilinguals to categorize bilabial stops differently from English native speakers, as the /p/-/b/ contrast in both languages presents different features. Whereas Spanish /p/ is unaspirated and presents VOT values around 0 ms, English /p/ presents long-lag VOT values. Spanish /b/ presents voice-lead, whereas English /b/ may present voice-lead or short-lag VOT (Lisker and Abramson, 1964). Spanish and English stops are, in fact, similar phones – i. e. L2 sounds which share certain characteristics with L1 sounds, but are not identical (Flege, 1987). Previous studies (Flege, 1987; Flege and Eefting, 1987; Riney, and Okamura, 1999) found that similar phones are rarely categorized and produced authentically, as not even experienced L2 learners may be able to produce them according to monolinguals' values.

This miscategorization may be a result of equivalence classification, which is a useful tool for L1 category formation, as it provides the ability to filter out acoustic features that are not phonologically important (Flege, 1987). However, in the case of L2 phonology acquisition, equivalence classification can lead to the omission of L2 phonological features that are not present in the L1 due to L1 perceptual assimilation. For example, Flege (1987) found that, even though production of /t/ by highly experienced English learners of French did not differ significantly in terms of VOT from those produced by French monolinguals, they were not identical. Similarly, experienced French learners of English produced longer VOT in English than less experienced learners, but their productions of /t/ presented shorter VOT values than French monolinguals' /t/. By contrast, other studies found instances of bilingual speakers who produced initial stops authentically in both languages. Antoniou Antoniou, Best, Tyler and Kross (2010, 211) found that L2 dominant Greek-English bilinguals were able to produce initial stops in both languages with VOT values that resembled those of monolinguals.

So far, we have referred to L1 influence on the L2. In addition, Flege (1987) found evidence of both L1 influence on the L2, as well as L2 influence on the L1 regarding the production of similar sounds. Highly experienced French speakers of L2 English living in the US presented VOT values in their L1 that differed from monolinguals' productions and approximated the values of their L2. These findings gave rise to the 'merger hypothesis', which accounts for bilinguals' formation of a single phonological category of similar sounds for the L1 and the L2 with intermediate values to those of monolinguals of each language. In this sense, it may not be striking to find that Spanish-English bilinguals present a single category for /p/ and /b/.

It should be noted that SLM focuses mainly on the production of nonnative sounds, and assumes that perception leads production, that is, that an inaccurate perception of a nonnative sound will fail to guide the sensorimotor acquisition of a nonnative sound, and, thus, production will be inaccurate as well (Flege, 1995). However, whereas a number of previous studies have found a relation between perception and production (Newman, 2003; Perkell, Guenther, Lane, Matthies, Stockmann, Tiede, and Zandipour, 2004), others have failed to establish such relation (Williams, 1977). For instance, Newman (2003) found that those listeners who selected a token with longer VOT as a perceptual prototype of a stop consonant tended to show longer VOTs in their productions of stops. Nevertheless, Williams (1977) looked at both the production and perception of stops by Spanish-English bilinguals and obtained different evidence of categorization in each dimension; it was found that Spanish-English bilinguals perceived voiced and voiceless stops differently from monolinguals in both languages, but the values obtained in their production were close to those of monolinguals in each language.

Best (1995) postulated the Perceptual Assimilation Model (PAM), which focuses on the perceptual dimension. Contrary to SLM, PAM does not look at individual phonemes only, but it also explicitly considers discrimination of non-native contrasts (Best, McRoberts and Goodle, 2001). Best's model proposes three different ways in which a nonnative pair can be perceptually assimilated to the L1. In the first place, a nonnative pair may undergo Two Category assimilation, that is, it will be assimilated to a native pair. Secondly, the nonnative pair may undergo Single Category assimilation, i.e. it is assimilated to one single L1 phone – in such case one phone may fit better than the other, and there may be a Goodness Difference. Finally, assimilation may not take place at all. Thus, in PAM's terms, English /p/ and /b/ are likely to be assimilated as two separate categories which resemble their L1 – i.e. they may be assimilated to Spanish /p/ and /b/ – by Spanish learners, as the articulatory properties of bilabial stops in both languages are very similar. In this case, L2 learners may fail to acknowledge phonetic differences in pairs of the target language, because they filter out L2 acoustic properties that are not present in the L1 contrast. More recently, Best and Tyler (2007) postulated PAM-L2, which pointed out that both SLM and PAM referred to L2 and nonnative phones interchangeably. Such equivalence was assessed with an aim to "extend PAM's nonnative speech perception framework to L2 learners" (Best and Tyler, 2007: 15). It should be noted as well, that PAM considered only inexperienced listeners of an L2 (Best, 1995), whereas PAM-L2 also attempted to extend the previous model to more experienced L2 learners.

2. 2. VOT

Previous research established VOT as the main cue for the stop voicing distinction in a number of languages, including Spanish and English (Schertz, Cho, Lotto, Warner, 2015; Shultz, Francis, Llanos, 2012; Williams 1977). Abramson and Lisker (1964) reported an average value for initial /p/ in Spanish of 4ms and an average value for initial /b/ of -110 ms. By contrast, English /p/ was found to present an average VOT value of 58ms. In the case of English /b/, individual variation has been found. Abramson and Lisker (1964) found instances of speakers whose productions of /b/ presented prevoicing – an average of -101ms – and others whose /b/ presented VOT values around zero – in this case, an average of 1ms. It should be noted that Abramson's and Lisker's (1964) reported values for Spanish stops were based on the production of Puerto Rican Spanish. Rosner, López-Bascuas, García-Albea and Fahey (2000) tested the production of Castilian speakers. The reported mean VOT value for Spanish /b/ was -91.5ms, and /p/'s mean VOT value was 13.1ms. All in all, Spanish stops contrast voice lead and short lag VOT, whereas English stops tend to contrast short lag and long lag VOT.

The values presented above are true for production. However, perception tasks have not always been found to match performance in production. Williams (1977) tested the perception of bilabials as /p/ and /b/ by Spanish-English bilinguals and Spanish and English monolinguals – so as to obtain reference values. It was found that Spanish-English bilinguals produce /p/ and /b/ differently in each language, as they preserved the characteristics of Spanish and English voiced and voiceless bilabials, but they behaved similarly in both languages in the perception tasks. Spanish monolinguals' /b/-/p/ crossover point in a labeling task was located at -4ms, whereas English monolinguals' crossover point presented VOT values of +25ms. When the labeling task was administered to Spanish-English bilinguals, the average crossover location for Spanish /b/-/p/ was found at 12ms, whereas the crossover point for English was located at 10.5ms. Such a small difference between the values of the VOT boundary in both languages suggests that Spanish-English bilinguals performed according to Spanish categories of bilabial stops, since, as mentioned above, Spanish /p/ presents short-lag VOT values.

It should also be noted that there are a number of factors that influence VOT values. Place of articulation of voiceless stops has an impact on the duration of VOT in the following manner: voiceless stops present higher VOT as place of articulation moves back (Cho and Ladefoged, 1999; Lisker and Abramson, 1967; Thornburgh and Ryalls, 1998). Therefore, /k/ tends to present the longest VOT, while VOT values for /p/ are usually the shortest. Moreover, the height of the vowel that follows the stop influences its VOT values;

stops that precede high vowels usually present a longer VOT (Klatt, 1975). Other differences in VOT include the gender of the speaker (Berry and Moyle, 2011) and the position stops present in the speech, i.e. whether they appear in citation form – in such case values tend to be longer – or in running speech; and the position they occupy in the syllable, as well as stress (Lisker and Abramson, 1967).

Given that participants in this study are generally Spanish-Catalan bilinguals from a very early age – and usually since birth –, we should take into consideration the impact Catalan may have on their categorization of stops. Reported average values of Catalan VOT for /p/ are 3ms, whereas /b/ presents voicing lead (Julià, 1981 in Llisterri, 2016b). Needless to say, Catalan VOT values for bilabial stops are very close to those of Spanish reported by Abramson and Lisker (1964). Thus, their bilingual status should not interfere in their Spanish VOT values.

2. 3. Other cues for voicing

In spite of the fact that there is a general agreement upon the fact that VOT is the main cue for the stop voicing distinction in English and Spanish – that is, voicing during closure and voice-lag or aspiration – it does not unequivocally serve as the only acoustic basis for the categorization of two distinct phones (Lisker and Abramson, 1967). Other acoustic cues that influence the categorization of a stop as voiced or voiceless include burst spectrum, F0 and F1. Burst intensity has been found to be greater in voiceless stops than in voiced stops (Chordoff and Wilson, 2014). In fact, it has been suggested that burst spectrum has a great impact on goodness judgments of voiced and voiceless stops, especially when VOT values are ambiguous or non-prototypical (Chordoff and Wilson, 2014). Considering the importance of this cue, the present study has controlled for the intensity and duration of burst. In the methods section, a detailed explanation of the creation of the stimuli is provided. F0 onset and contour have also been reported to be a cue for voicing in English (Hazan and Boulakia, 1993; Whalen, Abramson, Lisker, and Mody, 1993); as well as F1 values (Hazan, and Boulakia, 1993).

2. 4. Language experience

Previous research has found L2 experience to have a considerable influence on L2 and L1 performance, including the production and perception of VOT of non-native phones (Flege, 1987; Miyawaki, Strange, Verbrugge, Liberman and Fujimora, 1975). Best and Tyler (2007) defined L2 experience as a minimum of 6 to 12 months of immersion in the L2. Yet in the present paper a period of a minimum of 3 months, accompanied by formal instruction in the L2, has been considered to be L2 experience.

Flege (1987) assessed the effect of experience on /t/ productions by English-French bilinguals. For the purpose of his research, a number of groups differing in English and French experience were tested. Groups included: English monolinguals, English native speakers with little experience in French, more experienced learners of French who had received formal education in their L2, the English native speakers living in France; French native speakers living in the US; and French monolinguals speakers living in France who had had little exposure to English. As mentioned above, findings suggest that experience has an impact on the production of nonnative VOT, as experienced L2 groups produced L2 /t/ more authentically than the less experienced groups. Nevertheless, English native speakers who had received formal education in English – but had less L2 experience – were found to produce French /t/ with VOT values that were slightly closer to monolinguals' values than the group living in France. All in all, it was found that the two groups of

English learners of French with a greater L2 experience had a separate category for English and French /t/; whereas those with less L2 experience tended to categorize stops in both languages according to their L1. Nevertheless, the group of L1 French speakers living in the US presented a merged category, as they produced stops in both languages with intermediate VOT values. Similarly, Miyawaki et al. (1975) determined that Japanese L1 speakers' effective discrimination between /l/ and /r/ requires exposure to the L2 phone at an early age, for only those experienced English learners who were exposed to /r/ before adolescence were able to discriminate it from /l/.

Flege (1987) also assessed the effect of experience on VOT productions of /t/ in the L1. It was found that those groups who had a greater experience in the L2 – both English and French native speakers – presented influence of their L2 on the L1. Thus, it appears that experience has a bidirectional effect on production of similar phones, that is, an effect of the L1 on the L2 and of the L2 in the L1 (Flege, 1987, 2002). Conversely, Riney and Okamura (1999) failed to find an effect of the L2 – i.e. English – on the L1 – Japanese – , as no evidence that native Japanese speakers changed their L1 VOT values to those of English was found.

2. 5. The role of language mode

The experiment carried out for the purpose of the present study involved two categorization tasks, which included the same stimuli, but were presented in different languages – i.e. Spanish and English –, as well as two production elicitation tasks in which participants were asked to read a number of sentences in each language. Given the nature of the tasks and the order in which they were presented (see Methods section), it was important to control for language mode, that is, a bilingual's – or multilingual's – state of activation of a language. Such state of activation varies across a continuum which ranges from a monolingual state to a bilingual state (Grosjean, 2001).

Antoniou, Best, Tyler and Kross (2011) tested the effect of language mode and code-switching on the production of stops by English-dominant Greek-English bilinguals whose L1 was Greek. It was found that language mode had an effect on their production of English stops, as Greek-English bilinguals produced stops in English with Greek-like VOT values when they were asked to code-switch.

However, control for language mode and task presentation in the target language has not always proven to be effective. Hazan and Boulakia (1993) tested the categorization of /p/ and /b/ of English-French in both languages using the same stimuli. Their results suggested that language dominance has a stronger effect on labeling behavior than language presentation.

Despite the controversial results, language mode has been controlled in the present research. During the administration of the experiment, it was intended to activate the corresponding monolingual mode in each task – i.e. the Spanish monolingual mode in the Spanish identification and production tasks, and the English monolingual mode in the English identification and production tasks – so that participants performed according to their /p/-/b/ categories in each language.

3. Goal and Research Questions

The goal of the present paper is to determine the effect of L2 experience on the categorization of utterance-initial bilabial stops by Spanish learners of English in both Spanish and English. Thus, this study aims to evaluate whether the L1 influences the perception and production of L2 stops, and also whether the L2 affects the categorization and production of L1 stops. The categorization of bilabial stops will be analyzed in terms of the location of the perceptual boundary between /p/ and /b/ in each language by means of a VOT continuum, and the mean productions per group will be compared. The main research questions addressed in the present study are the following:

1. Do Spanish experienced learners of English categorize /p/ and /b/ more authentically – i.e. along VOT values comparable to those found with English monolingual speakers – than less experienced EFL Spanish learners? In other words, is there a greater influence of the L1 on the categorization of Spanish /p/ and /b/ on the part of inexperienced learners of English?

2. Is there an influence of the L2 on the categorization of Spanish /p/ and /b/ on the part of Spanish learners of English – both experienced and inexperienced learners? If so, do experienced learners present a greater degree of L2 influence – i.e. does experience have an influence on the degree of L2 influence on the L1?

3. Do experienced learners of English produce L2 stops more authentically than inexperienced learners, that is, does degree of L2 experience have an effect on the authenticity of the production of L2 stops? If so, have they created a new category?

4. Do inexperienced learners of English produce L1 stops with VOT values closer to Spanish native speakers than experienced learners, that is, does degree of L2 experience have an effect on the authenticity of the production of L1 stops?

5. Do those participants who perceive stops more similarly to native speakers produce them more authentically? In other words, is there an individual relation between perception and production accuracy?

Experienced learners of English are expected to perceive English sounds more similarly to English native speakers than inexperienced learners of English. Experienced bilinguals are likely to have acquired new perceptual categories for English stops, which may present either VOT values close to those of monolingual speakers, or intermediate values to both languages; on the other hand, inexperienced learners are expected to perform according to their L1 categories. Thus, inexperienced learners of English will potentially place their category boundary between /p/ and /b/ at VOT values that approximate to those of Spanish monolinguals due to L1 influence.

When it comes to Spanish, experienced learners are more likely to present L2 influence on their L1 than inexperienced learners. By contrast, inexperienced learners' production of bilabials will most likely approximate those of Spanish native speakers. Given that experienced learners' categories may be intermediate to those of English and Spanish, one possibility is that they have a single category for both languages.

Just as in the case of perception, experienced learners of English are expected to produce L2 stops more authentically than inexperienced learners. Experienced learners' productions of English stops are predicted to be close to those of native speakers, and different from the ones they produce in Spanish. Thus, they are expected to have separate categories in each language. By The inexperienced group is predicted to produce English stops according to Spanish VOT values, that is, they are believed to present the Spanish production category in both languages. Therefore, inexperienced learners are not expected to present L2 influence on their L1. In spite of the fact that experienced learners are predicted to produce L2 stops similarly to native speakers in English, they are also expected to perform in their L1 like Spanish monolinguals.

A relation between accuracy in the production and perception of L1 and L2 stops is expected for both groups. That is, those participants who categorize /p/ more similarly to English native speakers are predicted to produce bilabial stops more authentically, and those subjects who perform like monolingual speakers in the Spanish identification task are expected to produce stops according to the Spanish VOT values.

4. Methodology

In order to test the categorization of voiced and voiceless stops in terms of VOT values, two identification tasks involving a VOT continuum were used for the purpose of the experiment. Stimuli were created by modifying natural speech. Identification of stops has previously been tested with synthetic speech, both in one language (Abramson and Lisker, 1973) and across two languages (Williams, 1977). To my knowledge, only Hazan and Boulakia (1993) tested crosslinguistic categorizarion of stops using modified speech in two languages, but their study involved English-French bilinguals.

Regarding stimuli creation for the present study, several approaches were taken until the desired natural sounding stimuli were obtained. Finally, a combination of a manual edition of the stimuli, as well as the use of a Praat script were used for the creation of the tokens, as explained below.

4.1. Stimuli

A phonetically trained male speaker of Spanish and English was recorded in an acoustically treated room at the Phonetics Laboratory at UAB using a high-quality Sony PCM-D50 recorder. The speaker produced a number of instances of aspirated, unaspirated and prevoiced bilabial stops followed by /i/, aiming at producing a series of tokens with perceptually equivalent vowels. The Spanish vowel /i/, which appears to be perceptually very close in both languages (Cebrian, 2015), was selected in order to control for a possible activation of an undesired language mode (Grosjean, 2001) due to the identification of the vowel as a category of a language other than the one being tested.

After the recording, all instances of /pi/ and /bi/ were extracted and analysed using Praat, version 5.3.56, (Boersma and Weenik, 2016). Amplitude and duration of bursts were measured. Measurements were made by looking at the release bar shown in the spectrogram, and considering zero-crossings to establish burst boundaries. Partially following Schuttenhelm's (2013) procedure, a burst that did not contain cues for voicing in terms of duration was selected and modified in order to create an ambiguous burst – i.e. one with intermediate values to /p/ and /b/ in terms of amplitude and duration. The ambiguous burst was intended not to contain perceptual cues for the identification of /p/ or /b/.

Accordingly, the burst that presented a duration closest to the average of 9 productions of stops containing the features concerned – i.e. 3 pre-voiced stops, 3 short-lag VOT stops, and 3 aspirated stops – was selected. The duration of the bursts ranged from 4.2ms to 12.7ms. The burst selected was extracted from a short-lag VOT stop and had a duration of 9.7ms. Even though it was slightly longer than the average duration of all tokens measured – which was about 8.4ms – this burst was the one that presented a length closest to the average. Once the burst had been extracted, its amplitude was adjusted using Praat so as to match the average value obtained from the nine productions, i.e. 63.5dB.

In order to create the continuum, two vowels were needed. Their intonation patterns, as well as their duration, were compared in order to find two perceptually equivalent instances of /i/ (see Figures 1 and 2). The selected vowels were extracted from two different phonetic contexts; one of the vowels was extracted from an aspirated production of /pi/, and the other was found in an instance of an unaspirated /pi/. The former was used for the creation of the stimuli that presented a positive VOT, whereas the latter was used for the creation of the tokens presenting prevoicing as well as a VOT value of 0.

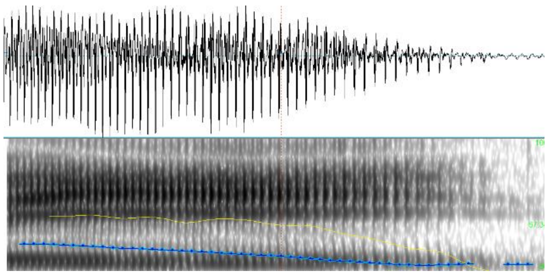
Figure 1. Vowel used to create unaspirated stimuli

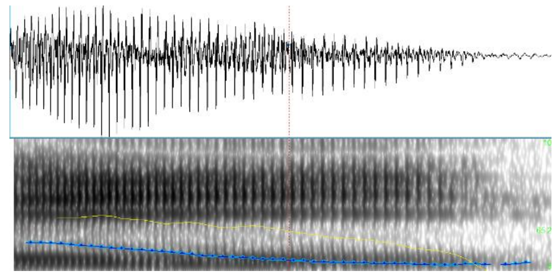
Figure 2. Vowel used to create aspirated stimuli

A total of 31 tokens which presented VOT values ranging from about -100ms to +100ms were created (see Table 1 for actual values). The stimuli varied in steps which ranged from about 5ms to 10ms. More specifically, steps that presented 100ms to 50ms of prevoicing as well as steps ranging from 50ms to 100 of aspiration varied in steps of about 10ms. Following the duration of steps in Williams' (1977) continuum, steps that were closer to a VOT of 0 ms – that is, those ranging from 45ms to 5ms of prevoicing and 5ms to 45ms of aspiration – varied in 5ms. Shorter steps were expected to help delimiting the category boundary more accurately. The 15 tokens with a negative VOT were created by adding cycles of prevoicing to the burst and the vowel extracted from an unaspirated context – which embodied the step with a VOT of 0ms. Cycles were carefully extracted at zero-crossings from the prevocing of a /bi/ that had been produced by the same speaker. On the other hand, the 15 aspirated stimuli were automatically created by means of a Praat script. The script was run twice in order to obtain stimuli of about both 10ms and 5ms (see Table 1). Once the 31 stimuli were built, intensity was normalized, and an identification task was created with Praat.

Step	VOT value	Step	VOT value
Step01	-98.6ms	Step17	14.1ms
Step02	-90.2ms	Step18	19.5ms
Step03	-80.4ms	Step19	24.8ms
Step04	-69.3ms	Step20	30.3ms
Step05	-60.6ms	Step21	35.5ms
Step06	-50.5ms	Step22	40.9ms
Step07	-45.2ms	Step23	45.9ms
Step08	-40.9ms	Step24	51.4ms
Step09	-35.2ms	Step25	56.3ms
Step10	-30.6ms	Step26	61.4ms
Step11	-25.9ms	Step27	72.2ms
Step12	-20.2ms	Step28	83ms
Step13	-15ms	Step29	93.5ms
Step14	-11.1ms	Step30	104ms
Step15	-5ms	Step31	109ms
Step16	9.7ms		

Table 1. Stimuli VOT duration in ms.

4. 2. Participants

A total of 22 Spanish learners of English completed the experiment. 11 of them – 10 females and 1 male – were first year students of English Studies. Their age ranged from 18 to 19, and none of them had lived or studied in an English speaking country previous to the experiment. They constitute the group of English learners with lower language experience and will henceforth be referred to as inexperienced learners. The remaining 11 Spanish learners of English – 5 males and 6 females – were 4th year students of Estudis Anglesos or graduate students who had lived and/or studied in an English-speaking country for at least three months. They embody the experienced English learners group, and their ages ranged from 22 to 41. The mean period of time spent in an English speaking country by the experienced group was 12 months. Moreover, 5 Spanish monolinguals – 3 males and 2 females – and 5 English monolinguals – 3 males and 2 females – completed the perception task in order to obtain a perceptual baseline for the categorization and production of bilabial stops in each language. Spanish monolinguals' ages ranged from 19 to 55, and English monolinguals' ages ranged from 28 to 38[2]. In the case of production, only 4 native speakers of each language completed the task due to organization and time issues. Participants ENS01 to ENS04 – 2 males and 2 females, whose ages ranged from 28 to 38 – completed the English production task. Participants SNS01 to SNS04 carried out the Spanish production task. Their ages ranged from 19 to 22.

4. 3. Task and procedure

After the stimuli were created, they were randomly sequenced and presented three times – a total of 93 tokens (31 stimuli x 3 repetitions) – in two identification tasks by means of Praat. Both tasks incorporated the same stimuli, but they were testing a different language, i.e. Spanish and English (see Figures 3 and 4 for an illustration of the tasks). Furthermore, the production of /p/ and /b/ in both languages was elicited. For the elicitation of the production data, participants read a list of sentences which included instances of initial /p/ and /b/ in different vowel contexts. More specifically, the production task aimed at eliciting 3 instances of words starting with /pi/ and /bi/, 2 instances of /pe/ and /be/ and 2 examples of /pa/ and /ba/ in each language, as well as 16 distractors. Only the /pi/-/bi/ productions were analyzed in the present research. The Spanish words were embedded in the carrier phrase 'X es la siguiente palabra' so that the stop appeared in utterance initial position, thus avoiding the possible assimilatory effects – e.g. spirantization of voiced stops – caused by previous sounds (Llisterri, 2016a). Similarly, the English words were included in the sentence 'X is the next word' (See Appendix E).

[2] See Appendix B for all participants' answers to the language background and use questionnaire.

It should be noted that the instructions for each task were provided in the corresponding language in order to trigger the activation of the desired language mode (Grosjean, 2001). Before the completion of the first task, participants were asked to fill in a linguistic background and language use questionnaire (see Appendix A).

In the first place, the testing groups – i.e. the experienced and the inexperienced participants – completed the Spanish task. The perception test consisted in the identification of each token as the first syllable sound in the Spanish word ‘pico’ or in ‘bicho’ Both the letter associated to /p/ and /b/, namely <p> and <b>, and the words mentioned above written in conventional Spanish orthography were provided as the two sole options for the identification of the given stimulus. Stimuli were presented once at every given trial, and participants had a chance to replay it one more time if necessary. As for the English task, it also presented the same stimuli three times in a randomized fashion. Participants had to label each token either as the first syllable in ‘Peter’ or in ‘beetle’. As was the case in the Spanish task, both letters <p> and <b> along with the words ‘Peter’ and ‘beetle’ respectively were provided in conventional English orthography. Additionally, six extra tokens were played before each task for practice, and, therefore, they were not analyzed.

Figure 3. Screenshot of the English categorization task in Praat.

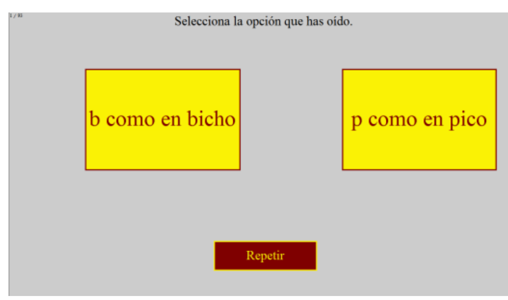

Figure 4. Screenshot of the Spanish categorization task in Praat.

Between the two identification tasks, the production of the subjects was recorded. First, participants were asked to read the list of Spanish sentences. Afterwards, they watched a short video in English so as to control for language mode, and read the English sentence list. Once the recordings were completed, the experienced and the inexperienced participants completed the identification test in English. The control groups – i.e. the English and the Spanish monolinguals – completed both the perception and the production tasks in their corresponding language only. The English native speakers, the experienced learners and the inexperienced learners completed the English version of the questionnaire, whereas the Spanish controls completed the Catalan version. Participants' answers to the questionnaires are reported in Appendix B. Once the experiment was completed, the results of the perception tasks were extracted for their analysis. Regarding the reading task, the VOT of all productions of /pi/ and /bi/ – 3 each for each task – was measured.

The experiment results are reported in the following sections. First, the results of the perception experiment will be presented, followed by the results of the production experiment.

5. Perception experiment

5. 1. Data analysis

For the purpose of analyzing the results obtained in the perception tasks, the mean values closer to 50% – that is to say, the ones that show more uncertainty – were considered in order to locate the category boundary. Moreover, the shapes of categorization curves were also compared between groups so as to assess certainty in the identification of bilabial stops. The steeper the identification function, the more certain participants are about their categorization; whereas the flatter it looks, the less certain they are. The percent identification of each of the stimuli in the continuum as /p/ by each group in English and in Spanish is presented in Figures 5 and 6 and Tables 2 and 4 below, respectively. Tables and figures include stimuli from -50.5ms to 61.4ms, as no variability was found at the remaining steps. /b/ responses have not been reported, as they are complementary to those of /p/.

A number of chi-square tests – i.e. one per comparison of each group to one another at the relevant steps for the categorization in each task – were carried out. The chi-squares were conducted on those steps at which at least one of the groups perceived /p/ between 40% and 60% of the times so as to reveal whether experience (independent variable) had an effect on the categorization of /p/ (dependent variable). Regarding the within-group between-language analyses, the results obtained by experienced and inexperienced learners in each language were analyzed by conducting a number of chi-square tests at steps 9.7ms, -5ms, -11.1ms and -15ms, given that these were the VOT steps where the category boundary is assumed to be located – i.e. they present identification values of /p/ closest to 50%. Language was the independent variable, whereas categorization of /p/ was the dependent variable. The results obtained in each identification task, followed by a within-groups comparison section will be presented below.

5.2. Identification of English stops

5.2.1. Results

Table 2 presents the results obtained by all groups in the English task – i.e. English native speakers (ENS), the experienced learners of English (EXP) and the inexperienced learners of English (INEXP) – in percentages. Figure 5 illustrates the categorization curves of /p/ of all groups in the English task. As both Table 2 and Figure 5 show, ENS heard /p/ 42% of the time when presented with stimulus with a VOT of 9.7ms. At this step, ENS presented more uncertainty in their categorization, as it was the closest value to 50%.

Except for step -11.1ms, ENS did not hear /p/ until step 9.7ms[3]. At step 17 (14.1ms), identifications as /p/ by ENS reached 92%. From step 19.5ms on, ENS identified /p/ unequivocally. The categorization curve obtained by ENS, except for a small peak at step 11.1ms, was very steep. In fact, it was the steepest line of all three groups.

Step value in ms	ENS	EXP	INEXP
-50.5ms	0%	0%	0%
-45.2ms	0%	3%	3%
-40.9ms	0%	0%	6%
-35.2ms	0%	3%	6%
-30.6ms	0%	3%	12%
-25.9ms	0%	12%	18%
-20.2ms	0%	9%	24%
-15ms	0%	12%	42%
-11.1ms	17%	15%	45%
-5ms	0%	18%	42%
9.7ms	42%	42%	58%
14.1ms	92%	82%	91%
19.5ms	100%	94%	100%
24.8ms	100%	100%	100%
30.3ms	100%	100%	100%
35.5ms	100%	100%	100%
40.9ms	100%	100%	100%
45.9ms	100%	97%	100%
51.4ms	100%	100%	100%
56.3ms	100%	100%	100%
61.4ms	100%	100%	100%

Table 2. Indentification of /p/ from step -50.5ms to step 61.4ms in the English task.

[3] Such small percentage in the case of step -11.1ms (17%), which in fact accounts for scarcely two cases , embodies the answer of one participant, ENS05.

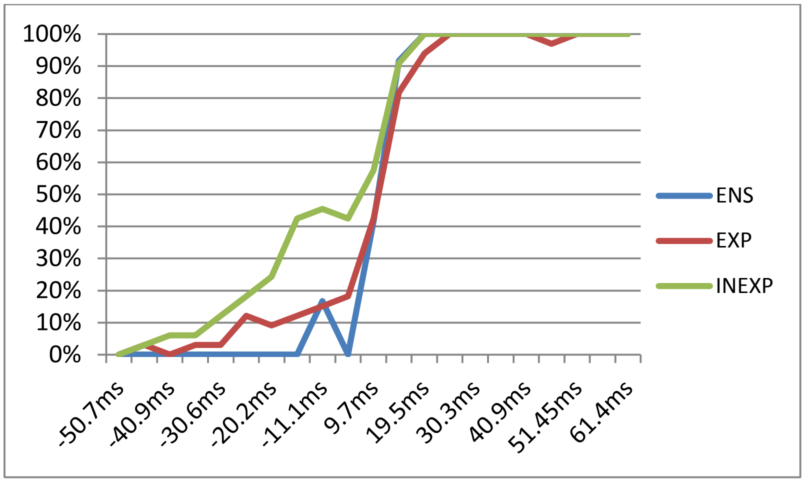

Figure 5. Categorization curve of /p/ from step -50.5ms to step 61.4ms in the English task.

Experienced learners' identification of /p/ also presented a mean value closest to 50% at step 9.7ms. They started hearing /p/ at step -45.2ms, although only 3% of the time. At step 25.9ms (15%) they started to hear /p/ increasingly. The experienced group started to hear /p/ 100% of the time from step 24.8ms on. As a result, their categorization curve for /p/ was not as steep as the one obtained by ENS.

When it comes to inexperienced learners of English, their closest mean values for the identification of /p/ was located between steps -15ms and 9.7ms. Just as the experienced group, inexperienced learners started hearing /p/ at step -45.2ms, but in their case the values obtained for the identification of /p/ increased more gradually along the continuum. Furthermore, they started hearing /p/ 100% of the times at step 19.5ms. Therefore, their categorization curve for /p/ in English shows as the most gradual of all three groups.

A total of twelve chi-square tests were conducted on steps -15ms, -11.1ms, -5ms and 9.7ms – at which at least one of the groups identified /p/ close to 50% of the times – in order to compare performance in the task between groups. Results are reported in Table 3. The chi-square test did not reveal any significant difference in the identification of /p/

among any of the three groups at step 9.7ms. At steps -5ms, -11.1ms and -15ms the chi-square did not show a significant difference between ENS and experienced learners. However, a significant difference was found the case of ENS and inexperienced learners and experienced and inexperienced learners (see Table 3 below).

English task				
Step	9.7 ms	-5 ms	-11.1 ms	-15 ms
Exp - Inexp	$\chi 2(1, N = 66) = 1.515, p = 0.218; p > 0.05$	$\chi 2(1, N = 66) = 4.591, p = 0.032; p < 0.05$ *	$\chi 2(1, N = 66) = 7.174, p = 0.007; p < 0.01$ **	$\chi 2(1, N = 66) = 7.639, p = 0.006; p < 0.01$ **
Exp - ENS	$\chi 2(1, N = 48) = 0.356, p = 0.551; p > 0.05$	$\chi 2(1, N = 48) = 3.117, p = 0.077; p > 0.05$.	$\chi 2(1, N = 48) = 0.027, p = 0.869; p > 0.05$	$\chi 2(1, N = 48) = 1.983, p = 0.159; p > 0.05$
Inexp - ENS	$\chi 2(1, N = 48) = 2.424, p = 0.119; p > 0.05$	$\chi 2(1, N = 48) = 8.984, p = 0.003; p < 0.01$ **	$\chi 2(1, N = 48) = 4.652, p = 0.031; p < 0.05$ *	$\chi 2(1, N = 48) = 8.984, p = 0.003; p < 0.01$ **

Table 3. Chi-square tests results for the English task.

5.2.2. Discussion

A significant difference between the inexperienced group and ENS, as well as experienced learners was found at steps -5ms, -11.1ms and -15ms. Nonetheless, difference in /p/ identification between experienced and inexperienced learners was not found to be significant at step 9.7ms. No significant difference was found between the experienced group and ENS. Results suggest that experienced learners of English identified initial /p/ in English similarly to ENS, but a difference in their identification function could still be observed. It should be noted that sample size with regard to the ENS group was very small, which may have resulted in the failure to find a significant difference between EXP and ENS.

When it comes to inexperienced learners, their identification of bilabials in English was found to differ significantly from that of ENS. In this case, it appears that inexperienced learners bilabial stops in English presented L1 influence. Such influence could be explained in SLM terms by the failure to adequately perceive features that are not phonologically relevant in the L1, in this case long-lag VOT (Flege, 1993). Therefore, in line with previous studies (Flege, 1987; Miyawaki et al, Antonious et al, 2010), experience appeared to have some influence on the formation of authentic categories of an L2 phoneme, given that EXP perceived L2 biliabial stops similarly to ENS, and that INEXP differ significantly.

Considering the fact that inexperienced learners' identification of initial /p/ in English differed significantly from that of ENS, we could claim that their miscategorization of English stops was a consequence of equivalence classification (Flege, 1987). In fact, they did not seem to have been able to create a category for /p/ in English yet. Both ENS' and experienced learners' /p/-/b/ category boundaries presented a VOT value of 9.7ms. Such a clear-cut categorization indicated certainty in their performance in the identification task. By contrast, the category crossover area of inexperienced learners of English was located between VOT values of -15ms and 9.7ms. A larger crossover area in their case, as opposed to the experienced group and ENS, pointed to uncertainty.

The location of the category boundaries differed from those found by Williams (1977), who reported that the English monolinguals' crossover point presented VOT values of 25ms, and the bilinguals' crossover point in English was located at 10.5ms. The fact that category boundaries were located earlier in the continuum in the present study may be due to the fact that the burst used to create the stimuli was extracted from a voiceless unaspirated production. Despite the efforts made to obtain a natural ambiguous burst that did not present any cues for voicing, the one used for the creation of the continuum was still slightly longer than the mean duration of all productions that were measured. It should be noted, however, that groups performed according to our expectations, but all of them consistently started to hear /p/ earlier in the continuum. All in all, as it was expected, experienced learners of English tended to perceive L2 bilabial stops more authentically than inexperienced learners of English

5.3. Identification of Spanish stops

5.3.1. Results

Table 4 presents the mean percentage of all /p/ responses by each group participating in the Spanish perception task, that is, Spanish native speakers (SNS), inexperienced learners of English (INEXP) and experienced learners of English (EXP), from step -50.5ms to step 61.4ms. Figure 6 illustrates the categorization curves of /p/ by the three groups also from step -50.5ms to step 61.4ms.

SNS started increasingly identifying /p/ at step 11 (-25.9ms) (13%). Steps that presented a value closest to 50% are -15ms (40%), -11.1ms (40%) and -5ms (33%). SNS started hearing /p/ 100% of the times at step 19.5ms, although at step 35.5ms they heard /p/ 93% of the times. The categorization curve obtained by SNS starts being very steep at step 9.7ms and, given that the percentage value of the previous steps increases gradually, it looks slightly curved until that point. It should also be noted that there is a flat line between steps 15ms 11.1ms.

Step value in ms	SNS	INEXP	EXP
-50.5ms	0%	0%	0%
-45.2ms	0%	0%	0%
-40.9ms	0%	3%	3%
-35.2ms	0%	12%	3%
-30.6ms	0%	9%	3%
-25.9ms	13%	6%	9%
-20.2ms	20%	27%	6%
-15ms	40%	45%	21%
-11.1ms	40%	45%	21%
-5ms	33%	52%	33%
9.7ms	87%	76%	76%
14.1ms	93%	97%	91%
19.5ms	100%	100%	100%
24.8ms	100%	100%	100%
30.3ms	100%	100%	100%
35.5ms	93%	100%	100%
40.9ms	100%	100%	100%
45.9ms	100%	100%	100%
51.4ms	100%	100%	100%
56.3ms	100%	100%	100%
61.4ms	100%	100%	100%

Table 4. Identification of /p/ from step -50.5ms to step 61.4ms in Spanish task.

Regarding experienced learners, they started hearing /p/ at step -40.9ms (3%). The percentage of the total identifications of /p/ increased gradually along the continuum and reached its value closest to 50%, in this case 33%, at step -5ms. At step 19.5ms, all participants started hearing /p/ 100% of the trials. They obtained a quite steep /p/ categorization curve, especially after step 11.1ms. Inexperienced learners also started hearing /p/ at step -40.9ms, but the number of /p/ responses increased earlier than in the case of experienced learners. The closest values to 50% regarding the identification of /p/ were located at steps -15ms (45%), -11.1ms (45%) and, especially, -5ms (52%). At step 18 (19.5ms), inexperienced leaners started unequivocally identifying /p/. Except for the bump at -35ms, INEXP patterned quite similarly to SNS. Moreover, at step 9.7ms all groups performed similarly, but between -25 and -11, the SNS and INEXP provided more identifications as /p/ than the EXP did.

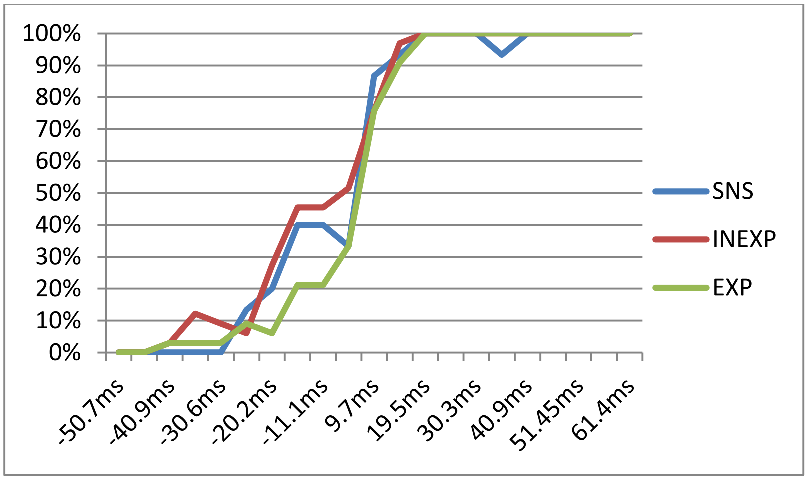

Figure 6. Categorization curve of /p/ from step -50ms to step 61.4ms in the Spanish task.

Nine chi-square tests were carried out in order to compare the differences in the identification of /p/ between groups. Results revealed that there was no significant effect of group on the categorization of /p/ at step -5ms ($p > 0.05$). However, at both at step -15ms ($\chi 2(1.\ N = 66) = 4.364.\ p = 0.037; p < 0.05$ *) and -11.1ms ($\chi 2(1.\ N = 66) = 4.364.\ p = 0.037; p < 0.05$ *) a significant difference was found between the identification of /p/ of experienced and inexperienced learners. All other comparisons yielded non-significant results (see Table A.1. in Appendix C for the results of all the chi-square tests performed).

5.3.2. Dicussion

Results show that experienced and inexperienced learners' identification of /p/ did not differ significantly from that of SNS. However, descriptively, percentages showed a difference of 19% at steps -11.1ms and -15ms between the experienced group and the SNS. A significant difference in this case may not have been found due to sample size limitations, since the SNS group was made up by only 5 subjects. Results suggest that experienced learners' categories of /p/ and /b/ in Spanish were close to those of SNS, although they were not identical. In any case, inexperienced learner's results were closer to SNS' than those of the experienced group are. Thus, there appears to be some influence of the L2 onto the L1 on the part of experienced learners. Still, unlike in the case of the production of French-English bilinguals studied by Flege (1987), it did not seem to be strong enough to result in a significant difference.

Contrary to what was expected, SNS performed more similarly to experienced learners than inexperienced learners' at step -5ms, although this difference between the testing groups was not found to be significant. These results may also have stemmed from sample size with regard to SNS. A significant difference was found between the performance of EXP and INEXP at steps -11.1ms and -15ms. Therefore, it appears that L2 experience had a certain negative effect on the perceptual categorization of L1 sounds, as lesser experience in the L2 seemed to relate to a greater similarity to L1 native speakers' identification of an L1 phonological contrast. A possible explanation for the difference in the identification of /p/ between experienced learners and SNS could have resulted from a partial activation of the English mode, given that aspirated tokens of /p/ were included, and the feature of aspiration is not present in Spanish stops, but in English. However, the inclusion of aspirated tokens was necessary in order to test both languages with the same continuum, as it would provide comparable results. In fact, previous studies have used continua which contained stimuli from voice lead to long lag VOT to test Spanish-English bilinguals (Abramson and Lisker, 1973; Williams, 1977).

The /p/-/b/ category boundaries of both SNS and inexperienced learners of English appeared to be located between -15ms and -5ms. The fact that their category boundary was not so clear-cut in their L1 could also be due to the fact that some of the stimuli presented contained aspiration. As it is true for the English categorization task, results in Spanish differed from those reported by Williams (1977) in a similar fashion. The category boundaries reported presented a VOT value of -4ms in the case of Spanish monolinguals, and 12ms in the case of Spanish-English bilinguals. Therefore, participants in the present study also consistently started to hear Spanish /p/ earlier in the continuum than it was expected, most probably due to the reasons stated above regarding the English task.

In short, as it was hypothesized, the /p/-/b/ categories of experienced learners of English appeared to be somewhat influenced by their English categories, as they perceived Spanish bilabial stops differently from Spanish monolinguals. Moreover, their category boundary was located at VOT values that are closer to the English monolinguals' values than the ones of inexperienced learners. As a matter of fact, inexperienced learners' categorization of bilabials was very close to that of Spanish native speakers.

5.4. Within-groups comparisons

5.4.1. Results

In order to compare the performance of each group in the two languages, the results obtained by each testing group in English and Spanish were analyzed by conducting a total of 8 chi-square tests on those steps that presented values close to 50% by at least one group in one language – namely, steps 9.7ms, -5ms, -11.1ms and -15ms. The chi-square tests did not reveal a significant difference in the categorization of /p/ across the two languages in the case of the inexperienced group at steps -15ms to 9.7ms ($p > 0.05$) (see Table A.2 for all results). Only one significant difference was found regarding experienced learners' performance between tasks; the categorization of /p/ in English and in Spanish at step 9.7ms was found to be significantly different ($\chi 2$ (1. $N = 66$) = 7.584. $p = 0.006$; $p < 0.01$ **).

5.4.2. Discussion

Experienced learners of English – that is to say, the group that was more likely to present L2 onto L1 influence – were found to pattern differently in each language. This difference between languages suggests that EXP may present different /p/-/b/ perceptual categories in each language. In line with Antoniou et al. (2011), who found that language mode had an impact on the production of L2 stops, language mode proved to have an effect on the crosslinguistic categorization of stops, although it may not have been completely controlled. As mentioned above, the fact that some of the stimuli included in the Spanish identification task presented voicing lag, may have had an influence on their performance. Conversely, inexperienced learners were found to perform similarly in both languages. This finding suggests that they may not have been able to create an L2 new category. Thus, as Best (1995) suggests, inexperienced learners of English seem to assimilate the /p/-/b/ contrast in English to the Spanish /p/-/b/ contrast, for they can be considered similar phones. In short, whereas EXP presented a tendency to perform differently in each language, INEXP appear to have used their L1 categories in both Spanish and English.

6. Production

6.1. Data Analysis

The VOT of all initial bilabial stops followed by /i/ produced by all participants in both languages – 3 words per language – was measured using Praat. Measurements were made from the onset of the burst to the beginning of voicing, and took into account zero crossings. Tables 5 and 7 show all mean productions of /p/ and /b/ per group and language. Tables A.3 and A.8 in the Appendix present the VOT means of each group per word in English and Spanish respectively.

VOT values of /p/ and /b/ obtained by all participants in each task underwent a normality test (Shapiro-Wilk Normality Test) and a homogeneity of variance test (Levene's Test for Homogeneity of Variance). Those comparisons that showed a normal and equal distribution underwent a One-way Analysis of Variance in order to assess the effect of group – ENS, experienced learners and inexperienced learners in the English task; and SNS, experienced learners and inexperienced learners in the English task – on VOT values of initial /p/ and /b/ –, which was measured in ms. If the Shapiro-Wilk normality test failed to reveal a normal distribution, a Kruskal Walis test was carried out.

As for the within-group analyses, the results obtained by each group in the two different languages were compared. A number of statistic tests were carried out in order to assess the effect of language (independent variable) on the production of /p/ (dependent variable) by each testing group; that is, whether experienced and inexperienced learners produced /p/ and /b/ differently in English and in Spanish. First, a Shapiro Normality Test was conducted for each comparison, as well as a Levene's Test for Homogeneity of Variance.

A description of the results obtained by participants in each language, followed by a discussion section, will be presented below. After that, the results of the within-group between-language analysis will be presented and discussed.

6.2. Production of English stops

6.2.1. Results

Mean values and the standard deviation of the production of /p/ and /b/ in the English task by group are presented in Table 5. Figures 7 and 8 illustrate the distribution of /p/ and /b/ productions by each group. ENS produced the longest average VOTs for /p/ (55.2ms), followed closely by the EXP (48.6ms) while INEXP produced somewhat shorter values (28.3ms). It seems that both L2 groups produced stops with greater aspiration than what is expected for their L1, and the experienced group approximated the ENS the most. EXP and INEXP presented a larger standard deviation (EXP: *SD* = 22.6; INXEP: *SD* = 25), than ENS (*SD* = 8.7). Moreover, ENS produced /b/ with an average of -18.1ms, whereas experienced and inexperienced learners produced a mean VOT that presented a long prevoicing (-69.6ms, in the case of experienced learners, and -70 ms in the case of inexperienced learners). All groups presented variability in their productions of /b/, including ENS, whose productions range from prevoiced to short-lag VOT. As mentioned above, a number of statistical tests were conducted in order to assess the effect of experience on VOT production.

	ENS		Experienced		Inexperienced	
	M	*SD*	*M*	*SD*	*M*	*SD*
/p/	55.2ms	8.8ms	48.6ms	22.6ms	28.3ms	25ms
/b/	-18.1ms	33.2ms	-69.6ms	46.5ms	-70ms	32.8ms

Table 5. Mean VOT and *SD* in ms per group in the English task.

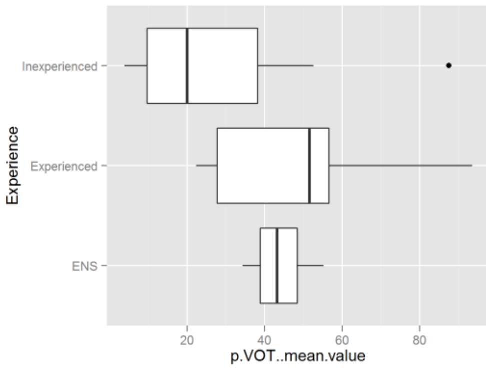

Figure 7. Boxplot of /p/ VOT mean value in ms per group in the Spanish task.

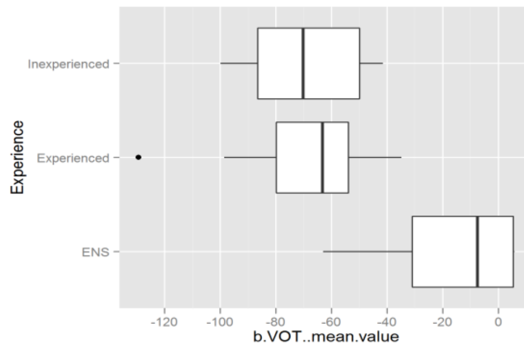

Figure 8. Boxplot of /b/ VOT mean value in ms per group in the English task.

No significant difference was found across groups in the English task in a one-way ANOVA testing the effect of group (ENS, EXP, INEXP) on the production of /p/ (VOT in ms (F (2,23) = 0.118, $p > 0.05$). In the case of /b/, Figure 8 shows that /b/ productions of experienced and inexperienced learners of English presented similar VOT mean values, whereas they differ to a great extent from those obtained by ENS. In fact, the one-way ANOVA revealed a significant difference between groups (F (2,23) = 0.004 **, $p < 0.01$). This significant difference can be accounted for by the contrast between ENS's average productions of /p/ (-18.1ms) and those of experienced (-69.6ms) and inexperienced learners (-70ms). It should also be considered that ENS showed a great variability in their productions of /b/, as some produced instances of /b/ prevoicing, whereas others did not (see individual results presented in Table A.5 in the Appendix).

Moreover, descriptively there was a difference across groups in the mean VOT values of disyllabic words. The sole disyllabic word in the English task showed a mean of 21.1ms in the case of the inexperienced group, and 35.7ms in the case of the experienced group, whereas ENS presented a mean VOT of 54.8ms (see Tables A.5 and A.6).

6.2.2. Discussion

The statistical tests did not show a significant difference in the production of /p/, between groups, whereas it did in the case of /b/, as both experienced and inexperienced learners differed from ENS. Results suggested that neither experienced nor inexperienced learners seem to have a /p/ category that is significantly different from that of ENS. Nonetheless, experienced learners' productions of initial /p/ presented a mean value that is almost identical to the one obtained by ENS – 48.6ms in the case of experienced learners and 55.2ms in the case of ENS –, whereas inexperienced learners' VOT mean value of /p/ was 28.3ms. Although this difference did not prove significant, the lower VOT values of the INEXP group's production may indicate that they were not as successful at producing native-like VOT values as the EXP group. It might be that the failure to find significant a difference was due to sample size, which was relatively small both in terms of number of participants and number of words. By contrast, experienced learners on the whole seemed to produce /p/ more authentically, especially regarding the similarity of the mean VOT values of /p/ that both present. Flege (1987) also found that more experienced learners produced L2 stops more authentically. In this case, it appears that experienced learners did not have L1 influence on their production of /p/.

In the case of English /b/, however, both EXP (69.6ms) and INEXP (-70ms) presented a mean VOT value of /b/ that was significantly longer than the mean obtained by ENS (-18.1ms). It may appear that experienced and inexperienced learners presented L1 influence on their production of initial /b/. However, an individual analysis revealed that some ENS – i.e. ENS02 and ENS04 – produced /b/ with a long prevoicing (see Table A.4 and A.5). Thus, prevoicing of initial /b/ could have stemmed from individual variation, as previous studies also found instances of prevoiced /b/ produced by ENS (Lisker & Abramson, 1964).

Regarding the mean production of the /p/ in Peter, descriptively, inexperienced learners produced a shorter voicing-lag. The fact that ‘Peter’ is a proper name that Spanish speakers usually hear in Spanish – e.g. in dubbed films or TV shows – may account for such difference.

In sum, both experienced and inexperienced learners produced /p/ with VOT values that did not differ significantly from those of ENS, even though INEXP’s average value was shorter. Results suggested that EXP produced /p/ authentically, and INEXP category of /p/ in production presented values close to those of ENS, although they were not identical. In the case of /b/, the mean VOT values obtained by both testing groups differed significantly from ENS. However, given that natives presented variability in their productions, it cannot be claimed that productions of /b/ by EXP and INEXP were not native-like.

6.3. Production of Spanish stops

6.3.1. Results

Table 6 shows the mean /p/ and /b/ VOT values and the standard deviation obtained by all groups in the Spanish task. All groups produced /p/ with a mean value of about 5ms, and /b/ with a long prevoicing: -96ms, in the case of SNS, -76.7ms in the case of EXP, and -73.2ms when it comes to INEXP. In the case of the production of /p/, little variation within groups was observed. In fact, all groups presented small *SD* values (see Table 6). As for /b/ experienced and inexperienced learners of English presented a greater variability than SNS (EXP: *SD* = 18.4; INEXP; *SD* = 21.9; SNS: *SD* = 9.9).

	SNS		Experienced		Inexperienced	
	M	*SD*	*M*	*SD*	*M*	*SD*
/p/	5.2ms	2.8ms	5ms	2.5ms	5ms	1.8ms
/b/	-96ms	9.9ms	-76.7ms	18.4ms	-73.2ms	21.9ms

Table 6. Mean VOT and *SD* in ms per group in the Spanish task.

As explained above, a one-way ANOVA and a Kruskal Walis one-way analysis of variance were conducted. In order to assess the effect of experience on the VOT productions of /p/, a Kruskal Walis test was carried out, as the Shapiro-Wilk normality test failed to establish a normal distribution ($W = 0.917$, $p = 0.039$; $p < 0.05$ *). The test did not show any significant difference across groups in the production of /p/ (H (2) = 0.008, $p = 0.995$; $p > 0.05$). In the case of /b/ productions in Spanish, the one-way ANOVA did not reveal any significant difference between groups either (F (2.12) = 0.142; $p > 0.05$) although, descriptively, SNS (-96ms) seemed to produce /b/ with a longer prevoicing than EXP (-76.7ms) and INEXP (-73.2ms).

6.3.2. Discussion

Both EXP and INEXP's mean productions of /p/ resembled the mean obtained by SNS. As a matter of fact, group did not show to have a significant effect on the production of Spanish /p/. Similarly, mean VOT values of /b/ of all groups in Spanish were very close, although SNS presented a slightly longer prevoicing than experienced and inexperienced learners. Despite this difference, the One-way Analysis of Variance revealed that experience did not have a significant effect on the production of /b/. Thus, it cannot be claimed that Spanish learners of English on the whole presented L2 influence in their production of /b/.

Nonetheless, a few instances of short-lag VOT productions of /b/ were found in the experienced group. Participant Experienced09 produced the Spanish word 'bicho', /bit͡ʃo/, (bug) with short-lag VOT (3.8ms), as well as 'villa', /biʎa/ or /biʝa/ (9.8ms). It should be noted that Experienced09 spent a total of 25 months in an English speaking country, which is the second longest period of time that any of the English learners participating in this study spent abroad. Only Experienced02 spent a longer period of time in an English speaking country, i. e. 30 months, (see Appendix B). It may be, then, that in the case of Experienced09, a longer immersion in English resulted in influence of the L2 on the L1 when it comes to the production of Spanish utterance initial /b/. However, more instances of productions of /b/ would be necessary in order to draw a reliable conclusion. In short, both groups produced L1 bilabial stops similarly to SNS. However, some exceptions were found, as there were few instances of short-lag productions of /b/ in the EXP group.

6. 4. Within-groups comparisons

6.4.1. Results

Tables 7 and 8 show the mean VOT values of /p/ and /b/ obtained by EXP and INEXP in each language respectively, as well as their corresponding standard deviation. The experienced group's mean productions of /p/ in English were longer than those in Spanish; they were short-lag in Spanish (5ms) and long-lag in English (48.6ms). The EXP group presented little variation in their productions of /p/ in Spanish (*SD* = 5), and were more inconsistent in English (*SD* = 22.6). Moreover, /b/ presents prevoicing in both languages (-76.7ms in Spanish and -69.6ms in English). Some variability was also observed in /b/ productions in both languages (see Table 7). The inexperienced group produced /p/ with a mean VOT value of 5ms in Spanish and 28.3ms in English. Productions of Spanish /p/ were very similar between participants (*SD* = 1.8), whereas English /p/ presented greater variability (*SD* = 25). INEXP's average VOT values of /b/ in each language also differed, although to a lesser extent; 5ms in the case of Spanish and 28.3ms in English. Inexperienced learners also produced /b/ with a long prevoicing, and presented a mean value of -73.2 in Spanish and -70ms in English. Very similar SD values were observed in productions of /b/ in both languages (see Table 8).

	/p/		/b/	
	M	*SD*	*M*	*SD*
Spanish	5ms	2.5ms	-76.7ms	18.4ms
English	48.6ms	22.6ms	-69.6ms	27.6ms

Table 7. Mean productions of /p/ and /b/ and *SD* by the experienced group.

	/p/		/b/	
	M	*SD*	*M*	*SD*
Spanish	5ms	1.8ms	-73.2ms	21.9ms
English	28.3ms	25ms	-70ms	21.1ms

Table 8. Mean productions of and /p/ and /b/ and *SD* by the inexperienced group.

In the case of /p/, a Shapiro normality test failed to reveal that the distribution was normal (EXP: $W = 0.837$, $p = 0.002$; $p < 0.01$ **; INEXP: $W = 0.684$, $p = 1.235e\text{-}05$; $p < 0.01$ **). Thus, a non-parametric Wilcoxon Sign-Rank Test per group was conducted. Given that in the case of the production of /b/ both groups presented a normal and homogeneous distribution (EXP: $W = 0.974$, $p = 0.802$; $p > 0.05$; INEXP: $W = 0.942$, $p = 0.227$; $p > 0.05$), a paired t-test per comparison was conducted.

The non-parametric tests showed that language had a significant effect on the production of /p/ for both groups (EXP: $W = 121$, $p = 2.835e\text{-}06$; $p < 0.01$ **; INEXP: $W = 111$, $p = 394e\text{-}03$; $p < 0.01$ **); in other words, both experienced and inexperienced learners produced /p/ differently in each language. Regarding /b/, the t-test results did not reveal any significant difference in the production of /b/ between languages for either group (EXP: t (20) = 0.354, p =0.726; $p > 0.05$; INEXP: t (29) = 0.707, p = 0.487; $p > 0.05$).

6.4.2. Discussion

As mentioned above, statistical tests showed that both experienced and inexperienced learners produced /p/ in Spanish and English with VOT values that differed significantly, whereas productions of /b/ did not prove to be significantly different across languages for either group. The significant effect of language on production of /p/ indicated that both experienced and inexperienced learners may present a separate category in each language. It should be noted that, descriptively, mean VOT values of /p/ of experienced learners (48.6ms) were closer to the mean obtained by ENS (55.2ms) than inexperienced learners' values were (28.3ms). Thus, in spite of the fact that both experienced and inexperienced learners of English produced /p/ differently in each language, EXP appeared to produce /p/ more authentically, that is, with VOT values closer to those of ENS, while INEXP seemed to have a /p/ category which presented an intermediate VOT value to that of ENS and SNS.

As for /b/, the Student t-tests did not show any significant effect of language on VOT production for either group. In fact, both experienced and inexperienced learners produced /b/ with long prevoicing. Thus, it appears that both groups shared the /b/ category in Spanish and English. In order to assess the impact of a shared category for /b/ in both languages, a couple of considerations need to be made. In the first place, the main contrast of bilabial stops in stressed initial position in English seemed to reside in the feature of aspiration; that is, in the presence of aspiration, /p/, which goes hand in hand with long-lag VOT; as opposed to the lack of it, /b/. Moreover, prevoicing of /b/ may also be present in native speakers' productions. Therefore, the fact that Spanish learners of English produced /b/ with a long prevoicing should not result in intelligibility problems, and, in fact, it may overlap with ENS' productions in some cases.

All in all, it appears that both EXP and INEXP produced English /p/ with longer VOT values than in Spanish, although EXP seemed to produce it more authentically. Nonetheless, both groups produced /b/ with similar VOT values – i.e. voice-lead VOT – in both languages.

7. Individual analysis. Relation between perception and production

7.1. Data Analysis

An individual descriptive analysis was conducted in order to assess the relation between production and perception. Given that the variables used for the analysis of the perception task were categorical and the variables of the production task were numerical, a statistical analysis was not attempted. Subjects were grouped together according to how they patterned in the perception and production tasks. Length of stay in an English speaking country of outliers were considered with a view to find a relation between degree of experience and inconsistencies with regard to the group means. In the case of perception, the category boundary of each individual was compared to the group's boundary. In order to assess the production dimension, mean VOT productions of all participants were compared to the mean of their group. Production and perception were compared so as to find individual parallel patterns in both dimensions; the categorical boundary of each subject was compared to their mean production of /p/ and /b/.

7.2. English tasks

Experienced learners of English appeared to be fairly consistent with their categorization of bilabials. Still, two outliers were found; Exp05 and Exp11 started categorizing /p/ earlier in the continuum (see Appendix D). Variation in this case did not appear to relate to length of stay in an English speaking country, as both outliers had spent a period of time close to the mean of the group (12 months).

With regard to production, a greater variability was found. A comparison between subgroups regarding productions of /p/ and /b/ showed that, in most cases, shorter VOT values of /p/ did not correspond to longer prevoicing of /b/ and vice versa, since participants that behaved similarly in the production of /p/ rarely patterned the same way in their production of /b/. Similarly, variation in length of stay did not seem to have an impact on production accuracy, given that participants within each subgroup did not share a similar amount of time spent in an English speaking country.

On the whole, a general relation between production and perception was not apparent in the experienced group, since, as mentioned above, participants were fairly consistent in perception, but not in production. Only one participant, Exp05, seemed to present a relation between perception and production, given that she produced /p/ with a shorter VOT than the group mean, and started to hear /p/ earlier in the continuum. By contrast, Exp11, who also started identifying /p/ earlier, produced /p/ with VOT values that matched the mean of the group, and /b/ presented a mean prevoicing shorter than the group mean. Therefore, no relation between production and perception could be established.

Contrary to experienced learners, the inexperienced group was not found to be consistent in either task (see Table 5 for mean VOT productions and *SD* per group and Appendix D for all responses in the identification task). However, identification group patterns did not match production, as participants within each subgroup produced /p/ with considerably different VOT values. Given that none of the inexperienced learners had lived or studied abroad, length of stay in an English speaking country could not be applied to account for variability.

7.3. Spanish tasks

In the Spanish perception task, some variability was found, especially regarding the inexperienced group. However, all groups showed to be very consistent in their productions of bilabial stops. Nevertheless, a relation between both dimensions was not observed.

The experienced group performed, on the whole, consistently in both tasks. However, a few participants, namely Exp05 and Exp11, started to hear /p/ earlier in the continuum. These exceptions did not seem to have any relation with the production task, since, as mentioned above, participants were very consistent in their production of /p/ and /b/ in terms of VOT values. Moreover, in spite of the fact that participants Exp05, Exp10 and Exp11 had spent a similar period of time in an English speaking country, their length of stay was not the shortest, but close to the mean. All participants' mean productions of /p/ presented a short-lag VOT, and all mean productions of /b/ had a long prevoicing. Only Exp09 was found to produce two instances of utterance initial /b/with a short-lag VOT, i.e. 'bicho' and 'villa'. In this case, Exp09 had spent a period of time substantially longer than the mean. However, his performance in the Spanish production task did not differ considerably from the overall's group performance.

The inexperienced group presented a greater variability in the perception task, but was fairly consistent in production. All participants' category boundary appeared to be located along the same VOT values. Some variability was found regarding the first step at which /p/ was heard, as well as the point at which /p/ was unequivocally categorized. Inconsistencies in perception, however, were not found to have a relation with production, given that all participants produced /p/ with a short-lag VOT and /b/ with a long prevoicing. In spite of the fact that individual mean VOT values of /b/ ranged from -111.4ms to -41.7ms, all instances were considered to be prevoiced bilabial stops.

In sum, none of the groups showed any clear relation between perception and production; participants' production and perception of initial bilabial stops patterned separately in both languages, as in very few occasions the same participants were found to perform similarly in both dimensions. Thus, it cannot be claimed that those participants whose perception of bilabials resembled that of ENS produced /p/ and /b/ more authentically.

These results are in line with Williams's (1977) findings, which also failed to establish a clear overall relation between perception and production. However, Newman (2003) found individual relations, as those subjects who performed more similarly to native speakers in the perception task tended to produce stops more authentically. The present paper also failed to establish such relation, although a bigger sample size, particularly in terms of the amount of production data analyzed, would have provided more reliable evidence.

8. General discussion and conclusions

The present paper has shed some light on the effect that L2 experience may have on the perception and production of initial bilabial stops of both the L1 and the L2. The analysis of production and perception tasks in both languages also accounted for category formation in the L2. However, a relationship between production and perception accuracy was not supported by the results of this study.

The first research question in this paper addressed the issue of whether experience had a positive effect on the perceptual categorization of L2 stops. Experienced learners were found to categorize /p/ more similarly to ENS than inexperienced learners, although not identically. Conversely, inexperienced learners performed according to Spanish VOT values in both identification tasks, and, thus, their L2 stops categories appeared to present L1 influence. These findings are in line with PAM's (Best, 1995) claim that inexperienced learners tend to perceptually assimilate an L2 contrast to an L1 contrast. Therefore, results suggested that experience may have a positive effect on the perceptual categorization of bilabial stops.

The second research question looked at the perception of L1 stops and category formation. Neither experienced learners nor inexperienced learners' identifications of /p/ were significantly different from those of SNS. However, descriptively, experienced learners seemed to differ slightly from SNS in the identification task. This difference may account for L2 influence, although it could have resulted from a partial activation of the English mode or task awareness. In any event, experience appeared to have less influence on the L2 than it does on the L1 in the perceptual dimension.

As for category formation, there was a tendency for experienced learners to perceive bilabial stops in a more native-like manner – they performed similarly to ENS and SNS in each task – which may indicate that L2 bilabial stops underwent category dissimilation (Flege, 2002). However, the L2 category would still present some L1 influence. As for inexperienced learners, they performed according to Spanish values in both tasks, which may account for a shared perceptual category in both languages.

The third RQ discussed in this paper considered the effect of L2 experience on the authenticity of L2 stops. Experienced learners produced /p/ authentically, but /b/ presented a VOT mean value very close to Spanish. The mean VOT values of /p/ production of INEXP were not as close to ENS' mean values as the ones obtained by EXP. In fact, they seemed to be intermediate to both languages. Therefore, it appears that experience tended to have a lesser effect in the accuracy of the production of L2 stops, given that the performance of both testing groups is more similar in this dimension. This finding is in line with Williams' (1977) results, which suggested that whereas experienced L2 speakers of English produced bilabial stops with different VOT values in each language, they performed more similarly in the perception task.

The fourth research question considered the effect of L2 experience on the production of L1 stops and category formation. Experience did not prove to have a significant effect on the production of L1 stops, given that /p/ and /b/ were produced similarly by all groups. Regarding category formation in production, experienced learners and, to a lesser extent, inexperienced learners of English produced /p/ differently in each language. This finding may indicate that they had created a category for /p/. When it comes to /b/, it appears that both groups shared a category in both languages, as all mean productions showed a long prevoicing. Therefore, English /b/ seemed to have been directly assimilated to Spanish /b/. However, it should be noted that some ENS produced instances of prevoiced /b/, and, thus, this feature should not be considered non-native.

The last research question considered a possible individual relation between production and perception of L1 and L2 stops. A descriptive analysis failed to establish such relation in both languages, given that, in most cases, those participants who perceived stops most similarly to monolinguals often did not produce /p/ and /b/ most authentically.

All in all, it appears that experience had a positive impact on both perception and production of L2 sounds. Overall, the experienced group outperformed inexperienced learners in both tasks. However, within-group variation in length of stay did not seem to have an effect on L1 and L2 perception of bilabial stops. Degree of experience in the L2 appeared to have a greater influence on perception, where the INEXP group differed from ENS to a greater extent than the EXP group did, than in production, where the two L2 groups patterned together, although the EXP group produced longer aspiration than the INEXP group. Moreover, experienced learners seemed to match ENS' production of /p/, but differ slightly in the identification task. As for the L1, a significant effect of experience was not found in either dimension. This finding is in line with Riney and Okamura's (1999) conclusions, as they also failed to find a substantial effect of L2 exposure on L1 production of stops. Nonetheless, descriptively, there appeared to be a certain degree of influence of English on the categorization of L1 bilabial stops. Furthermore, a relation between perception and production was not found, given that participants patterned separately in each dimension. It appears, thus, that a more accurate perception of stops did not involve a more authentic production.

9. Limitations and further research

In the first place, the present research used a relatively small sample size. A greater number of participants, and more instances of productions of stops would have provided more robust evidence. Despite the attempt to recruit homogeneous groups, group arrangement was based on length of stay abroad. Perhaps, potential confounding variables such as age of acquisition, motivation for learning and proficiency in a language other than Spanish, English and Catalan may have had an effect on participants' performance.

Moreover, a wider variety of acoustic contexts would have shed more light on the categorization of L2 stops. Thus, other points of articulation as well as different vowel contexts may be analyzed in further research. Furthermore, the VOT continuum used for the identification task could have presented a bias towards a voiceless identification. As discussed above, an early identification of /p/ may have stem from the fact that the burst used for the creation of the stimuli was longer than the average duration of all bilabial productions. A more rigorous control of the burst will be considered for the creation of future VOT continua. Similarly, other acoustic cues for voicing, such as F1 and F0 should possibly be considered.

As for task design and order in which tests were carried out, completion of the identification task previous to production may have raised awareness of the experiment's goal. Awareness may have had an influence on the production of stops, which may account for the fact that participants were overall successful producing /p/. A more spontaneous task – which would result in less controlled productions – will be considered for further study, in order to test whether learners would produce L2 stops as accurately. In addition, the inclusion of aspirated stimuli in the Spanish task may have partially activated the English language mode. All limitations concerning task design and creation of the continua, as well as the inclusion of other variables, such as point of articulation and different vowel environments, will open new lines for further research.

References

Abramson, A.S, and Lisker, L. (1973). Voice-timing perception in Spanish word-initial stops, *Journal of Phonetics*, 1: 1 – 8.

Antoniou, M., Best, C. T., Tyler, M. D. and Kross, C. (2010). Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2. *Journal of Phonetics*, 38: 640-653-

Antoniou, M., Best, C. T., Tyler, M. D. and Kross, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code switching. *Journal of Phonetics*, 39: 558 – 570.

Berry, J. and Moyle, M. (2011). Covariation among vowel height effects on acoustic measures. *J. Acoustical Society of America*, 130 (5).

Best, C. T. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech perception and linguistic experience: issues in cross-language research*. Timonium, MD: York Press.

Best, C. T., McRoberts, G. W. and Goodle, E. (2001). Discrimination of non-native consonant contrasts varying in perceptual assimilation to the listener's native phonological system, *J. Acoustical Society of America*, 109 (775).

Best, Catherine T. & Michael D. Tyler (2007). Chapter 2. Nonnative and second-language speech perception: Commonalities and complementarities. In M. J. Munro and O. S. Bohn (Eds.), *Second language speech learning: The role of language experience in speech perception and production*, Amsterdam: John Benjamins: 13 – 34.

Boersma, P. and Weenik, D. (2016). Praat: Doing phonetics by computer. Retrieved from www.praat.org

Cho, T. and Ladefoged, P. (1999). Variation and Universals in VOT: evidence from 18 languages. *Journal of Phonetics*, 27: 207 – 229.

Chordoff, E. and Wilson, C. (2014). Burst spectrum as a cue for the stop voicing contrast in American English, *J. Acoustical Society of America*, 136: 2762.

Cebrian, J. (2015). Reciprocal measures of perceived similarity. In The Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow, 104: 1 – 9

Flege, J. E. (1987). The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification. *Journal of Phonetics*, 15: 47 – 65.

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

Flege, J.E. (1995). Chapter 8: Second Language Speech Learning Theory, Findings, and Problems. In Winifred Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, Timonium, MD: York Press: 233 – 277.

Flege, J.E. (2002). Interactions between the native and second-language phonetic systems. In P. Burmeister and T. Piske and A. Rohde (Eds.), *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Trier: Wissenchaftlicher Verlag: 217-243.

Grosjean, F. (2001): The Bilingual's Language Modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing*, Oxford: Blackwell: 1–22.

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

Klatt, D. H. (1975). Some effects of context on voice onset time in English stops, *Language and Speech*, 10: 1 – 28.

Lisker, L. and Abramson, A. S. (1964). A Cross-Language Study of Voicing in Initial Stops: Acoustical Measurements. *Word*, 20 (3): 384 – 421

Lisker, L. and Abramson, A. S. (1967). Some effects of context on voice onset time in English stops, *Language and Speech*, 10: 1 – 28.

Lisker, L, Liberman, A. M., Erickson, D. M, Dechovitz, D., Mandler, R. (1977). On pushing the voice-onset-time (VOT) boundary about. *Language and Speech*, 20: 209 – 216.

Llisterri, J. (2016b). *Las características acústicas de los elementos segmentales*. Retrieved June 6, 2016, from Departament de Filologia Espanyola, Universitat Autònoma de Barcelona Web site: http://liceu.uab.es/~joaquim/phonetics/fon_anal_acus/caract_acust.html

Llisterri, J. (2016a). *Español peninsular - Consonantes. Sistema fonológico y principales alófonos*. Retrieved July 4, 2016, from Departament de Filologia Espanyola, Universitat Autònoma de Barcelona Web site: http://liceu.uab.es/~joaquim/phonetics/fon_esp/IPA_cons_sp.html

Miyawaki, K., Strange, W., Verbrugge, R., Liberman, A. M, Jenkins, J. J., and Fujimura, O. (1975). An effect of linguistic experience: The discrimination of [r] and [1] by native speakers of Japanese and English. *Perception & Psychophysics*, 18 (5): 331–340.

Newman, R. S. (2003). Using links between speech perception and speech production to evaluate different acoustic metrics: A preliminary report, *J. Acoustical Society of America*, 113: 2850 – 2860.

Perkell, J., Guenther, F. H., Lane, H., Matthies, M. L., Stockmann, E., Tiede, M., and Zandipour, M. (2004). The distinctness of speakers' productions of vowel contrasts is related to tehri discrimination of the contrasts, *J. Acoustical Society of America*, 116: 2338 – 2344.

Riney, T. J., and Okamura, K. (1999). Does Bilingualism affect the First Language?, *ICU Language Research Bulletin*, 14: 101 – 113.

Rosner, B. S., López-Bascuas, L. E., García-Albea, J. E., Fahey, R. P. (2000). Voice-onset times for Castilian Spanish initial stops, *Journal of Phonetics*, 28: 217 – 224.

Schertz, J., Cho, T., Lotto, A., Warner, N. (2015). Individual differences in phonetic cue use in production and perception of a non-native sound contrast, *Journal of Phonetics*, 52: 183 – 204.

Schuttenhelm, Lisan (2013). *Perception of the non-native phone [g] in Dutch: where lies the VOT boundary?* (unpublished MA Thesis, Universitet van Amsterdam).

Shultz, A. A., Francis, A. L., Llanos, F. (2012). Differential cue weighning in perception and production of consonant voicing, *Acoustical Society of America*, 132 (2): 95 – 101.

Thornburgh, D. F., Ryalls, J. H. (1998). Voice onset time in Spanish-English bilinguals: early versus late learners of English, *J. Commun. Disord*. 31: 215 – 229.

Whalen, D. H., Abramson, A. S., Lisker, L., and Mody, M. (1993). F0 gives voicing information even with unambiguous voice onset times, *Acosutical Society of America*, 47: 36 – 49.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psycopyhysics*, 21 (4): 289 – 297.

Appendix A. Questionnaires.

Questionnaire

Participant Code: ________________________________

Experiment date and time: ____

Personal Information

- Name: ____
- Age: ____
- Occupation: ____
- Place of birth: ____
- Parents' place of birth: ____
- Place of residence: ____
- Previous place(s) of residence (where you have lived for at least a few months; indicate when and for how long):

  ____

  ____

  ____

Language and language use information

- Native language: __________________________
- Parents' native language: ____

- Indicate how often you use SPANISH in your daily interactions with the following groups:
  - At home: never rarely sometimes often always
  - With friends: never rarely sometimes often always
  - At university: never rarely sometimes often always
  - At work: never rarely sometimes often always

- Indicate how often you use CATALAN in your daily interactions with the following groups:
    - At home: never rarely sometimes often always
    - With friends: never rarely sometimes often always
    - At university: never rarely sometimes often always
    - At work: never rarely sometimes often always

- Indicate how often you use ENGLISH in your daily interactions with the following groups:
    - At home: never rarely sometimes often always
    - With friends: never rarely sometimes often always
    - At university: never rarely sometimes often always
    - At work: never rarely sometimes often always

Qüestionari

Codi del participant: ___________________________________

Data i hora de l'experiment: ____

Dades personals:

- Nom: ____
- Edat: ______
- Ocupació: ____
- Lloc de naixement: ____
- Lloc de naixement dels pares: ____
- Lloc de residència: ____
- Altres llocs de residència (on hagis viscut almenys un parell de mesos. Indica quan i per quant de temps):

  ____

  ____

  ____

- Has cursat alguna assignatura de fonètica anglesa o de pronunciació de l'anglès (incloent Fonètica i Fonologia Anglesa I i II?_______________

Llengües i ús de la llengua

- Llengua materna: _________________________
- Llengua materna dels pares: ____

- Indica amb quina freqüència utilitzes l'ESPANYOL diàriament amb els grups següents:

- A casa:

  mai quasi mai de vegades amb freqüència sempre

- Amb bels amics:

  mai quasi mai de vegades amb freqüència sempre

- A la universitat:

mai quasi mai de vegades amb freqüència sempre

- A la feina:

mai quasi mai de vegades amb freqüència sempre

- Indica amb quina freqüència utilitzes el CATALÀ diàriament amb els grups següents:

- A casa:

mai quasi mai de vegades amb freqüència sempre

- Amb bels amics:

mai quasi mai de vegades amb freqüència sempre

- A la universitat:

mai quasi mai de vegades amb freqüència sempre

- A la feina:

mai quasi mai de vegades amb freqüència sempre

- Indica amb quina freqüència utilitzes l'ANGLÈS diàriament amb els grups següents:

- A casa:

mai quasi mai de vegades amb freqüència sempre

- Amb bels amics:

mai quasi mai de vegades amb freqüència sempre

- A la universitat:

mai quasi mai de vegades amb freqüència sempre

- A la feina:

mai quasi mai de vegades amb freqüència sempre

Appendix B. All participants questionnaire answers.

	Age	Place of birth	Other places of residence	Time spent in an English speaking country	Use Sp at home	Use of Sp with friends	Use of Sp at university	Use of Sp at work	Use Cat at home	Use of Cat with friends	Use of Cat at university	Use of Cat at work	Use En at home	Use of En with friends	Use of En at university	Use of En at work
ENS01	33	Adelaide (Aus)	London, Perth, Barcelona (1 year)	since birth	rarely	rarely	never	N/A	N/A	N/A	N/A	N/A	always	often	always	N/A
ENS02	38	Chesterfield (UK)	Hull (UK), Brighton, Manchester, Barcelona	since birth	sometimes	sometimes	sometimes	never	N/A	N/A	N/A	N/A	often/alwa	often/always	often/always	Always
ENS03	28	Bristol (UK)	Mollerussa (3 years), London (4 years), Sabadell (1 year)	since birth	N/A	N/A	N/A	N/A	often	sometimes	never	sometimes	often	often	always	always
ENS04	28	Tennessee (US)	South Carolina, North Carolina, Yuyao (1 year), Dominican Repulic (1 year) Barcelona (1 year)	since birth	never	never	never	rarely	N/A	N/A	N/A	N/A	always	always	always	always
ENS05	28	Rugby (UK)	Granada (1 year), Japan (1 year), Barcelona(1 year)	since birth	often	often	sometimes	sometimes	always	often	often	sometimes	often	often	often	always
Exp01	22	Sant Joan de Vilatorrada	Davis, California	9 months	always	always	always	rarely	sometimes	always	always	rarely	never	rarely	sometimes	often
Exp02	24	Barcelona	Berkeley,(CA) Toronto (Canada)	30 months	often	rarely	rarely	rarely	often	always	always	often	sometimes	sometimes	rarely	always
Exp03	23	Vinaròs	Cerdanyola del Vallès, London	9 months	never	sometimes	sometimes	rarely	always	often	often	never	never	rarely	often	always
Exp04	25	Valladolid	Parma (1 year), L'Aquila (1 year), Genova (1 year), Canberra (Aus, 2 months), Toronto (2 months)	4 months	often	often	sometimes	sometimes	never	never	never	never	sometimes	often	often	sometimes
Exp05	22	Reus	Edinburgh (9 months)	9 months	never	sometimes	rarely	sometimes	always	always	sometimes	sometimes	rarely	sometimes	always	always
Exp06	23	Sabadell	Dublin	9 months	sometimes	sometimes	often	often	often	sometimes	rarely	rarely	never	rarely	often	often
Exp07	41	Medellín (Colombia	Morgantown (1 year), Barcelona (1 year)	12 months	always	often	sometimes	sometimes	never	never	never	never	never	sometimes	often	often
Exp08	22	Vic	Edinburgh (4 months)	4 months	never	sometimes	often	N/A	always	often	often	N/A	never	rarely	often	N/A
Exp09	25	Vic	Manchester (11 months), York, UK (14 months)	25 months	never	never	rarely	rarely	always	sometimes	often	often	rarely	often	often	often
Exp10	38	Barcelona	Glasgow, Scotland (12 months)	12 months	always	sometimes	sometimes	never	never	sometimes	sometimes	sometimes	never	sometimes	often	often
Exp11	23	Barcelona	Castelló d'Empúries, Toronto (10 months)	10 months	often	often	often	sometimes	always	often	sometimes	often	rarely	rarely	sometimes	rarely
Inexp01	18	Guipúzcoa	Zarautz, Hendaya, Barcelona	N/A	always	often	often	always	never	never	rarely	never	sometimes	sometimes	sometimes	sometimes
Inexp02	19	Girona	Barcelona	N/A	rarely	never	sometimes	never	always	always	always	always	never	rarely	sometimes	never
Inexp03	19	Barcelona	Terrassa	N/A	always	often	often	sometimes	never	rarely	sometimes	never	never	sometimes	always	often
Inexp04	19	Barberà del Vallès	-	N/A	always	always	often	often	never	rarely	often	often	never	rarely	often	sometimes
Inexp05	18	Romania	Les Borges Blanques (Lleida 9 years)	N/A	rarely	often	always	always	never	sometimes	sometimes	rarely	never	rarely	often	rarely
Inexp06	18	Barcelona	L'Escala (Girona)	N/A	rarely	rarely	rarely	sometimes	always	always	often	often	never	rarely	often	sometimes
Inexp07	19	Lloret de Mar	Barcelona	N/A	rarely	sometimes	sometimes	N/A	often	rarely	sometimes	N/A	never	sometimes	often	N/A
Inexp08	18	Manresa	-	N/A	rarely	never	rarely	never	always	always	sometimes	often	never	never	sometimes	often
Inexp09	18	Terrassa	-	N/A	always	always	often	always	often	sometimes	sometimes	never	never	rarely	sometimes	often
Inexp10	18	Mallorca	Barcelona (9 months)	N/A	never	sometimes	sometimes	N/A	always	often	often	N/A	never	rarely	sometimes	N/A
Inexp11	19	Girona	Cerdanyola del Vallès (9 months)	N/A	rarely	often	often	sometimes	often	sometimes	sometimes	often	never	rarely	sometimes	sometimes
SNS01	21	Manresa	Manresa	N/A	always	sometimes	rarely	N/A	never	often	often	N/A	never	rarely	never	N/A
SNS02	22	Terrassa	Rubí	N/A	always	always	always	always	rarely	rarely	rarely	rarely	never	rarely	never	never
SNS03	20	Barcelona	Rubí	N/A	always	often	often	N/A	rarely	soemtimes	rarely	N/A	never	rarely	never	N/A
SNS04	19	Sallent	L'Escala (Girona)	N/A	never	rarely	never	sometimes	always	always	always	often	never	never	never	never
SNS05	55	Spain	Spain	N/A	always	often	often	often	never	sometimes	sometimes	sometimes	never	never	never	never

Appendix C. Tables and Figures.

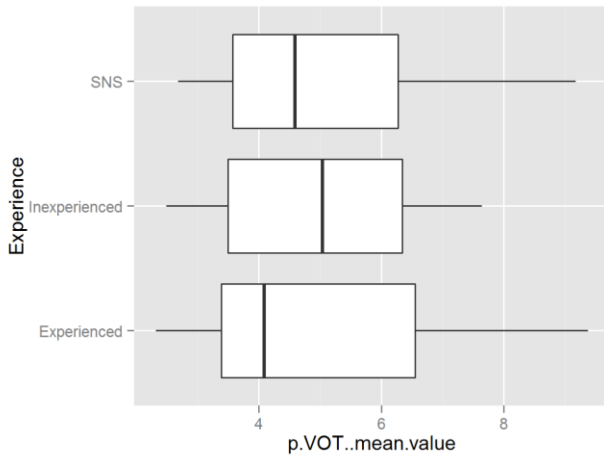

Figure A. 1. Boxplot of /p/ VOT mean in the Spanish task.

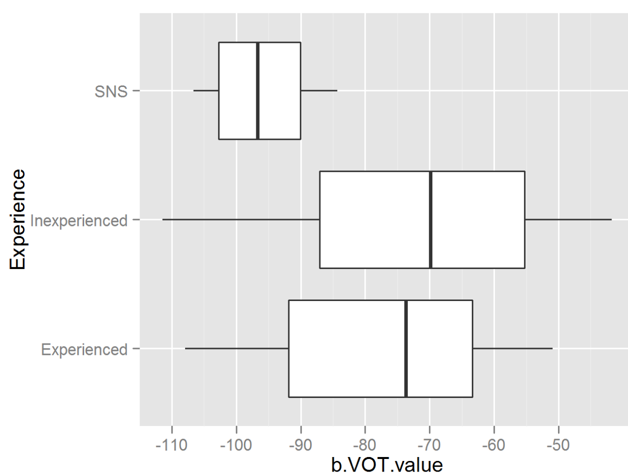

Figure A. 2. Boxplot of /b/ VOT mean value in the Spanish task.

Spanish task			
VOT	-5ms	-11.1ms	-15ms
Exp - Inexp	$\chi 2(1, N = 66) = 2.233$, $p = 0.135$; $p > 0.05$	$\chi 2(1, N = 66) = 4.364$, $p = 0.037$; $p < 0.05$*	$\chi 2(1, N = 66) = 1.2$, $p = 0.037$; $p < 0.05$*
Exp - SNS	$\chi 2(1, N = 48) = 0$, $p = 1$; $p > 0.05$	$\chi 2(1, N = 48) = 1.843$, $p = 0.175$; $p > 0.05$	$\chi 2(1, N = 48) = 1.2$, $p = 0.175$; $p > 0.05$
Inexp - SNS	$\chi 2(1, N = 48) = 1.373$, $p = 0.241$; $p > 0.05$	$\chi 2(1, N = 48) = 0.125$, $p = 0.724$; $p > 0.05$	$\chi 2(1, N = 48) = 1.2$, $p = 0.724$; $p > 0.05$

Table A. 1. Chi-square tests results for the Spanish task.

English - Spanish					
VOT	14.1 ms	9.7 ms	-5 ms	-11.1 ms	-15 ms
Experienced	$\chi 2(1, N = 66) = 1.158$, $p = 0.282$; $p > 0.05$	$\chi 2(1, N = 66) = 7.584$, $p = 0.006$; $p < 0.01$ **	$\chi 2(1, N = 66) = 1.981$, $p = 0.159$; $p > 0.05$	$\chi 2(1, N = 66) = 0.407$, $p = 0.523$; $p > 0.05$	$\chi 2(2, N = 66) = 0.982$, $p = 0.322$; $p > 0.05$
Inexperienced	$\chi 2(1, N = 66) = 1.065$, $p = 0.302$; $p > 0.05$	$\chi 2(1, N = 66) = 2.455$, $p = 0.117$; $p > 0.05$	$\chi 2(1, N = 66) = 0.547$, $p = 0.459$; $p > 0.05$	$\chi 2(1, \mathrm{N} = 66) = 0$, $p = 1$; $p > 0.05$	$\chi 2(1, N = 66) = 0.062$, $p = 0.804$; $p > 0.05$

Table A. 2. Chi-square tests results for the between-language comparisons.

	beat	peel	beetle	peace	Peter	beer
ENS	-21.7ms	55.1ms	3.2ms	55.7ms	54.8ms	-35.8ms
Experienced	-82.4ms	57.3ms	-73.9ms	51.7ms	35.7ms	-52.5ms
Inexperienced	-71.7ms	33ms	-73.4ms	30.7ms	21.1ms	-64.9ms

Table A. 3. Mean VOT in ms per group and per word in the English task.

	beat	peel	beetle	peace	Peter	beer	mean /p/	mean /b/
ENS01	9.4ms	51ms	2.8ms	44.1ms	43.3ms	3.1ms	46.1ms	5.1ms
ENS02	-103ms	38.2ms	2.2ms	33.1ms	31.9ms	-88.3ms	34.4ms	-63ms
ENS03	4.4ms	39.8ms	6.2ms	40.1ms	41.1ms	6.6ms	40.3ms	5.8ms
ENS04	1.9ms	91.3ms	1.8ms	105.5ms	103.1ms	-64.7ms	100ms	-20.3ms
mean	-21.7ms	55.1ms	3.2ms	55.7ms	54.8ms	-35.8ms	55.2ms	-18.1ms

Table A. 4. ENS production results per word.

	mean /p/	mean /b/
ENS01	46.1ms	5.1ms
ENS02	34.4ms	-63ms
ENS03	40.3ms	5.8ms
ENS04	100ms	-20ms
Exp 1	40.4ms	-85ms
Exp 2	26.5ms	-60.5ms
Exp 3	22.3ms	-61.5ms
Exp 4	26.9ms	-98.4ms
Exp 5	28.6ms	-63.2ms
Exp 6	56.3ms	-74.6ms
Exp 7	93.5ms	-37.4ms
Exp 8	56.8ms	-34.7ms
Exp 9	76.9ms	-129.3ms
Exp 10	51.5ms	-73.8ms
Exp 11	54.5ms	-47.1ms
Inexp1	9.7ms	-99.8ms
Inexp2	3.8ms	-41.4ms
Inexp3	19.9ms	-70.1ms
Inexp4	9.6ms	-83.1ms
Inexp5	17.1ms	-83ms
Inexp6	38.2ms	-61.5ms
Inexp7	38.3ms	-89.7ms
Inexp8	87.6ms	-48.6ms
Inexp9	6.5ms	-51.3ms
Inexp10	27.8ms	-46.7ms
Inexp11	52.6ms	-94.7ms

Table A. 5. All participant mean productions of /p/ and /b/ in the English task.

	monosyllabic	disyllabic
ENS	55.4ms	54.8ms
Experienced	54.5ms	35.7ms
Inexperienced	31.9ms	21.1ms

Table A. 6. Mean VOT in ms per group and number of syllables.

Participant	Experience	VOT
ENS01	ENS	43.3ms
ENS02	ENS	31.9ms
ENS03	ENS	41.1ms
ENS04	ENS	103.1ms
Exp01	Experienced	23.8ms
Exp02	Experienced	15ms
Exp03	Experienced	17.4ms
Exp04	Experienced	4.5ms
Exp05	Experienced	4.4ms
Exp06	Experienced	43.9ms
Exp07	Experienced	100.9ms
Exp08	Experienced	51.6ms
Exp09	Experienced	75.6ms
Exp10	Experienced	39.9ms
Exp11	Experienced	15.1ms
Inexp01	Inexperienced	17.8ms
Inexp02	Inexperienced	1.4ms
Inexp03	Inexperienced	9.4ms
Inexp04	Inexperienced	4.1ms
Inexp05	Inexperienced	3.1ms
Inexp06	Inexperienced	49.1ms
Inexp07	Inexperienced	34ms
Inexp08	Inexperienced	50.8ms
Inexp09	Inexperienced	8.3ms
Inexp10	Inexperienced	23.2ms
Inexp11	Inexperienced	31.2ms

Table A. 7. Mean VOT production of ‘Peter’ per participant in ms.

	pino	vino	bicho	pipa	pillo	villa
SNS	7.7ms	-89.3ms	-91.9ms	3.6ms	4.3ms	-98.ms
Experienced	5.3ms	-79.6ms	-81.8ms	4.5ms	5.4ms	-68.6ms
Inexperienced	5.2ms	-74.9ms	-76.2ms	4.5ms	5.2ms	-68.6ms

Table A. 8. Mean VOT in ms per group and per word in the Spanish task.

	mean /p/	mean /b/
SNS01	2.6ms	-84.3ms
SNS02	5.3ms	-101.4ms
SNS03	9.1ms	-106.6ms
SNS04	3.8ms	-91.9ms
Exp01	3.2ms	-107.9ms
Exp02	9.3ms	-50.9ms
Exp03	2.3ms	-91.6ms
Exp04	3.7ms	-93.3ms
Exp05	4ms	-73.6ms
Exp06	2.6ms	-63ms
Exp07	6ms	-85.2ms
Exp08	3.5ms	-68.3ms
Exp09	6.4ms	-53.9ms
Exp10	9.3ms	-92.1ms
Exp11	4.7ms	-63.6ms
Inexp1	2.5ms	-111.4ms
Inexp2	6.2ms	-51.4ms
Inexp3	6.2ms	-86.5ms
Inexp4	5ms	-82.5ms
Inexp5	3ms	-69.8ms
Inexp6	4.6ms	-87.5ms
Inexp7	6.4ms	-65.1ms
Inexp8	3.9ms	-55.6ms
Inexp9	7.6ms	-54.8ms
Inexp10	6.8ms	-41.7ms
Inexp11	2.4ms	-99.4ms

Table A. 9. All participant mean productions of /p/ and /b/ in the Spanish task.

Appendix D. All responses to the English identification task.

Participant	ENS01	ENS02	ENS03	ENS04	ENS05	Exp01	Exp02	Exp03	Exp04	Exp05	Exp06	Exp07	Exp08	Exp09	Exp10	Exp11	Inexp01	Inexp02	Inexp03	Inexp04	Inexp05	Inexp06	Inexp07	Inexp08	Inexp09	Inexp10	Inexp11
Language	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En	En
98.651 ms	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
90.252 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80.432 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69.314 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60.646 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
50.507 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45.227 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	1	0	0	0	0
40.957 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0
35.230 ms	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
30.622 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	2	0	0
25.979 ms	0	0	0	0	0	0	0	0	0	2	0	0	0	0	0	2	0	0	0	0	0	0	3	0	3	0	0
20.235 ms	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	1	0	0	0	1	0	0	3	1	3	0	0
15.033 ms	0	0	0	0	0	0	0	0	0	1	0	0	0	0	1	2	0	0	3	0	0	2	3	3	3	0	0
11.144 ms	0	0	0	0	2	0	0	0	0	1	0	0	0	1	0	3	0	1	3	0	0	3	3	2	3	0	0
5.092 ms	0	0	0	0	0	2	0	0	0	2	0	0	0	0	0	2	0	1	3	1	0	1	2	2	3	1	0
9.735 ms	0	1	1	1	2	1	1	1	0	3	1	1	1	1	1	3	0	3	3	1	0	3	3	2	3	1	0
14.142 ms	3	1	3	1	3	3	3	3	2	3	3	2	1	1	3	3	3	3	3	3	2	3	3	3	3	2	2
19.586 ms	3	3	3	3	3	3	3	3	2	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3
24.867 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
30.397 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
35.581 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
40.925 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
45.926 ms	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
51.454 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
56.300 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
61.457 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
72.235 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
83.025 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
93.550 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
104.027 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
109.028 ms	3	3	3	3	3	3	3	3	3	3	3	3	5	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Appendix E. All responses to the Spanish identification task

Participant	Exp01	Exp02	Exp03	Exp04	Exp05	Exp06	Exp7	Exp08	Exp09	Exp10	Exp11	Inexp01	Inexp02	Inexp03	Inexp04	Inexp05	Inexp06	Inexp07	Inexp08	Inexp9	Inexp10	Inexp11	SNS01	SNS02	SNS03	SNS04	SNS05
Language	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp	Sp
98.651 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
90.252 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
80.432 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
69.314 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
60.646 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50.507 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45.227 ms	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
40.957 ms	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
35.230 ms	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	2	0	2	0	0	0	0	0	0	0
30.622 ms	0	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0	0
25.979 ms	0	0	0	0	2	0	0	0	0	1	0	0	0	0	0	0	0	1	0	1	0	0	0	1	0	0	1
20.235 ms	0	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0	0	3	1	3	0	1	0	1	1	0	1
15.033 ms	0	0	0	1	1	0	0	0	0	2	3	0	0	3	1	0	0	3	2	3	1	2	2	1	1	0	2
11.144 ms	0	0	0	0	1	0	0	0	0	3	3	0	0	2	1	0	2	3	1	3	2	1	2	2	0	0	2
5.092 ms	2	0	1	2	2	0	0	0	0	1	3	1	3	3	0	0	0	3	3	2	1	1	0	3	1	0	1
9.735 ms	2	3	2	0	3	3	2	2	3	2	3	1	3	3	2	1	3	3	3	3	1	2	3	2	3	3	2
14.142 ms	3	3	3	2	3	3	2	2	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	2
19.586 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
24.867 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
30.397 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
35.581 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	2	3	3
40.925 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
45.926 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
51.454 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
56.300 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
61.457 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
72.235 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
83.025 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
93.550 ms	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
104.027 ms	3	3	3	3	3	3	3	3	2	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
109.028 ms	3	3	3	3	3	3	3	1	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3

Appendix F. Sentences used for the production task.

English

Read the following sentences:

1. Rose is the next word.
2. Beat is the next word.
3. Peel is the next word.
4. Dog is the next word.
5. Chair is the next word.
6. Shirt is the next word.
7. Read is the next word.
8. Sheep is the next word.
9. Pen is the next word.
10. Cat is the next word.
11. Queen is the next word.
12. Ban is the next word.
13. Pet is the next word.
14. Sock is the next word.
15. Beetle is the next word.
16. Mess is the next word.
17. Peace is the next word.
18. Peter is the next word.
19. Bat is the next word.
20. Leaf is the next word.
21. Beer is the next word.
22. Love is the next word.
23. Ben is the next word.

24. Pat is the next word.
25. Jeans is the next word.
26. Like is the next word.
27. Dress is the next word.
28. Better is the next word.
29. Leak is the next word.
30. Pan is the next word.

Spanish

Lee las siguientes frases:

1. Reloj es la siguiente palabra.
2. Pino es la siguiente palabra.
3. Silla es la siguiente palabra.
4. Vino es la siguiente palabra.
5. Cosa es la siguiente palabra.
6. Vela es la siguiente palabra.
7. Madre es la siguiente palabra.
8. Padre es la siguiente palabra.
9. Carta es la siguiente palabra.
10. Pena es la siguiente palabra.
11. Casa es la siguiente palabra.
12. Bicho es la siguiente palabra.
13. Taza es la siguiente palabra.
14. Bolso es la siguiente palabra.
15. Pipa es la siguiente palabra.
16. Mesa es la siguiente palabra.
17. Perro es la siguiente palabra.

18. Móvil es la siguiente palabra.

19. Iglesia es la siguiente palabra.

20. Pala es la siguiente palabra.

21. Café es la siguiente palabra.

22. Sala es la siguiente palabra.

23. Niño es la siguiente palabra.

24. Tele es la siguiente palabra.

25. Pillo es la siguiente palabra.

26. Villa es la siguiente palabra.

27. Ola es la siguiente palabra.

28. Marco es la siguiente palabra.

29. Beso es la siguiente palabra.

30. Gato es la siguiente palabra.