Universitat Autònoma de Barcelona

Crosslinguistic influence in L2 speakers: The effect of language experience on L1 and L2 VOT production and perception

Celia Gorba Masip

Crosslinguistic influence in L2 speakers: The effect of language experience on L1 and L2 VOT production and perception

Celia Gorba Masip

Ph.D. Dissertation

Supervised by Dr. Juli Cebrian

Programa de Doctorat en Filologia Anglesa

Departament de Filologia Anglesa i Germanística

Universitat Autònoma de Barcelona

May 2020

Universitat Autònoma de Barcelona

Acknowledgements

For the last three and a half years, I have been completing my Ph.D., which has resulted in the present manuscript. It has been an exciting learning journey that I could not have successfully completed without the excellent guidance and support of my supervisor, *Dr. Juli Cebrian*, who I have known and admired since I was an undergraduate student. Thank you very much for all your thorough feedback, suggestions, corrections and emotional support, Juli. I would also like to thank Juli Cebrian for giving me the chance to be part of his research group, which has enriched my knowledge and research abilities beyond the scope of my Ph.D.

I am also very thankful to the Spanish Ministry of Economy and Competitiveness [Grant No. FFI2017-88016-P] and to the Catalan Government [Grant No. 2017SGR34] for their financial support to Juli's projects, which has allowed the team to conduct our research and given us the opportunity to attend conferences all over the world. And, of course, I would also like to thank my colleagues from the research group, *Dr. Núria Gavaldà*, *Dr. Angelica Carlet* and *Dr. Susana Cortés* for sharing their ideas with the team and for listening to mine. I have learnt a lot working with you these past few years.

I would also like to thank *Dr. Joan Carles Mora* for his useful advice and feedback during these last few years and for inviting me to participate in the Barcelona L2 Speech Meetings, where I have met other researchers in the field from whom I have learnt a lot. I am also thankful to Joan Carles for helping me recruiting participants and for allowing me to use the facilities at the Universitat de Barcelona to test some of the participants.

I also value *Dr. Maria Machuca*'s advice and kindness throughout the years and for her help with the segmentation of stops. Thank you very much, Maria.

During my PhD years, I have been lucky to do two research stays abroad. I would like to thank *Dr. Esther de Leeuw* for accepting me at Queen Mary University of London and letting me use their lab, where I collected half of my data. I am also very thankful to *Dr. Mirjam Broersma* for welcoming me at Radboud University and inviting me to give a talk at the very final stages of my Ph.D. The feedback was very much appreciated.

A very special thanks to all the participants who volunteered to take part in the experiment even if they would not get an economic compensation. I would also like to thank *Carles Salse*, who assisted me in the recordings I made at the Servei del Tractament de la Parla i el So.

I am very thankful to the English Department at UAB for trusting me and giving me the invaluable chance to work with them by choosing me to hold a PIF position. This pre-doctoral position has allowed me to work full time on my Ph.D. and to gain some experience as a university teacher. I would not have been able to finish in less than four years otherwise. I would also like to thank all my colleagues in the English Department for their kindness and support.

I would also like to thank *Dr. Hortènsia Curell* for helping me find the last participants I needed to finish my data collection.

A very special thank you to *Andrea Huerta*, *Jessica McDaid*, *Núria Gavaldà* and *Àngel Barranqueras*, who I consider not just my colleagues but also my friends, for their emotional support through my Ph.D. years.

I cannot but thank my parents, *Lola Masip* and *Rafa Gorba*, and my brother, *Amado Gorba*, for their support. It was my mom, who, when I was just a small child somehow guessed my fate by calling me *doctora de las palabras* ('the doctor of the words'). It would have been scary if her guess had been more accurate and actually called me 'the doctor of the sounds'. In any case, my mom, as well as my father, has always encouraged me to study and do what I liked and has been incredibly supportive during my Ph.D. years (even if, as she says, and she may have a point, the academic world is very tough these days).

I would also like to thank my partner, *Aida Alsina*, the person who has put up with my occasional anxiety the most and has always been incredibly supportive, even during my research stays, when we were apart. She has also been my guinea pig on several occasions, as I have asked her to try my tests more than once.

Finally, I would like to thank all the cats in foster care that have kept my company during the tedious and tough writing months: *Kiara*, *Roberto*, *Rona*, *Carmen*, *Nissan*, *Iker*, *Matías* and *Sam*. Thank you all for providing me with fluffy writing breaks.

Sabadell, May 2020

Abstract

The acquisition of a second language (L2) may bring about crosslinguistic influence (CLI), which can be either progressive – previously acquired languages influence later acquired ones – or regressive – later acquired languages influence previously acquired ones. Previous studies have found instances of progressive (e.g., Antoniou, Best, Tyler & Kross, 2010; Zampini, 1998) and regressive CLI (e.g., Bergmann, Nota, Sprenger & Schmid 2016; Harada, 2003), as well as bidirectional CLI (e.g., Flege, 1987; Major, 1992). One of the main factors that may modulate CLI is L2 experience – understood as amount of time spent in an immersion setting. More specifically, L2 experience may result in a more target-like perception and production in the L2, as well as in a less native-like performance in the first language (L1) due to an influence of the L2 (e.g., Flege, 1987; Levy & Law, 2010). This study investigates the effect of L2 experience on both the perception and production of L1 and L2 initial stops. Moreover, since both modalities were examined, the relationship between perception and production was also investigated. Two mirror-image populations were tested, including Spanish learners of English and English learners of Spanish who differed in amount of L2 experience and learning setting. One group of inexperienced learners, one group of moderately experienced learners living in their home country, one highly experienced group in an immersion setting and one control group of functional monolinguals were tested for each language. Participants had to complete a perception experiment and a production experiment involving initial stops in each language. The perception tasks consisted of two forced-choice identification tests – one for bilabial stops and one for velar stops – which presented stimuli varying on a VOT continuum. The same stimuli were presented in the two conditions – i.e., in English and in Spanish. The production experiment consisted of a carrier sentence reading task presenting stops in absolute initial position in each language.

Regarding the perception experiment, different results were revealed for the Spanish learners of English and for the English learners of Spanish. A bidirectional effect of L2 experience was found on the Spanish groups, since L2 experience resulted in more target-like L2 category boundaries but also in less native-like L1 category boundaries. Spanish learners tended to categorize L1 and L2 stops similarly and only on one occasion – the moderately experienced group's /k/-/g/ categorization – were L1 and L2 category boundaries found to differ significantly. No effect of L2 experience was found on the English groups, who perceived both L1 and L2 stops with English-like VOT values. In the case of production, L2 experience had a positive effect on the L2 of both populations, as the experienced groups tended to produce more target-like values than the inexperienced groups. Still, L2 experience may have a less straightforward effect on voiced stops than on voiceless stops. None of the Spanish groups presented English-like values for voiced stops and the moderately experienced English participants outperformed the highly experienced participants – probably due to differences in amount of L2 instruction. No evidence of the effect of L2 experience on L1 production was found. Contrary to the absence of separate L1 and L2 categories observed for perception, L2 learners tended to produce Spanish and English stops differently – i.e., with greater VOT values in English than in Spanish –, although in the case of /b/, out of the Spanish groups, only the most experienced group made a significant difference. Finally, regarding the relationship between perception and production, no straightforward relationship was found at an individual level, since participants who presented later category boundaries did not always use greater VOT values in production. On the other hand, the overall results of the two experiments suggest that the English learners were more successful at producing L2 stops accurately than at perceiving them, opposing the assumption of the Speech Learning Model (Flege, 1995) or the Perceptual Assimilation Model (Best & Tyler, 2007) that L2 perception precedes production. An overall more target-like L2 performance in production than in perception may stem from methodological issues related to language mode control. The production task – a reading task – presented the target phones in a language specific context, whereas, in the production experiment, stimuli were presented in isolated syllables. Moreover, those isolated syllables – /pi/, /bi/, /ki/ and /gi/ – are more meaningful in English than in Spanish, which may have resulted in the – total or partial – activation of English in both the English and the Spanish condition. In sum, L2 experience was found to influence both the L2 and the L1, although a greater effect of the L1 on the L2 was revealed than the other way around. A greater amount of L2 experience might be needed to observe a clearer influence of the L2 on the L1, as the most experienced groups had lived in an immersion setting for a relatively short period of time – i.e., about 4 years. Differences between the two populations tested indicate that CLI may take place differently according to the order of acquisition of the languages involved and that factors other than L2 experience – such as L2 use or L2 instruction – may play a role.

Keywords: L2 perception and production, L2 experience, relationship between perception and production, crosslinguistic influence, Spanish stops, English stops.

Resumen

La adquisición de una segunda lengua (L2) puede resultar en influencia interlingüística, que puede ser tanto progresiva – de una lengua adquirida con anterioridad a otra adquirida posteriormente –, como regresiva – de una lengua adquirida posteriormente a otra adquirida anteriormente. Estudios previos han mostrado ejemplos de influencia lingüística progresiva (e.g., Antoniou, Best, Tyler & Kross, 2010; Zampini, 1998) y regresiva (e.g., Bergmann, Nota, Sprenger & Schmid 2016; Harada, 2003), así como bidireccional (e.g., Flege, 1987; Major, 1992). Uno de los factores que pueden modular la influencia interlingüística es la experiencia en L2 – entendida en este estudio como el tiempo vivido en un contexto de inmersión lingüística. Concretamente, la experiencia en L2 puede resultar en una percepción y producción más precisas en la L2, pero menos nativas en la primera lengua (L1) a causa de la influencia de la L2 (e.g., Flege, 1987; Levy & Law, 2010). Este estudio investiga el efecto de la experiencia en L2 tanto en la percepción como en la producción de las oclusivas en posición inicial. Asimismo, el hecho de que este estudio examine las dos modalidades permite que se investigue la relación entre percepción y producción. Seis grupos de hablantes de una segunda lengua que se diferencian en su experiencia en L2, así como en el orden de adquisición de las lenguas – hablantes de español como L1 e inglés como L2 y hablantes de inglés como L1 y español como L2 –, participaron en el estudio. Por cada L1, participaron en el estudio un grupo de aprendices de L2 sin experiencia, un grupo de aprendices con una experiencia moderada en L2 que residen en su país de origen, un grupo con una alta experiencia en L2 en un contexto de inmersión lingüística y un grupo de monolingües funcionales. Los participantes completaron un experimento de percepción y otro de producción de oclusivas en cada lengua. El experimento de percepción consistía en dos tests de identificación – uno para las oclusivas bilabiales y otro para las oclusivas velares – que presentaban estímulos de un continuo de VOT. Los mismos estímulos fueron empleados para examinar la percepción de ambas lenguas – i.e., inglés y español. El experimento de producción consistía en la lectura de frases que presentaban oclusivas en posición inicial absoluta.

Los hablantes de español como L1 y los hablantes de inglés como L1 obtuvieron diferentes resultados en el experimento de percepción. Por un lado, los hablantes de español como L1 e inglés como L2 presentaron indicios de interferencia lingüística

bidireccional, ya que la experiencia en L2 resultó en una percepción más precisa en la L2, pero menos nativa en la L1. Además, cabe destacar que los españoles aprendices de inglés categorizaron las oclusivas de la L1 y la L2 de manera similar, a excepción de la categorización de /k/-/g/ por parte del grupo con una experiencia moderada. Por otro lado, no se encontraron indicios de influencia de la experiencia en L2 en los hablantes de inglés como L1, puesto que estos categorizaron tanto las oclusivas del inglés como las del español según los valores de VOT del inglés. En el caso de producción, los resultados indican que la experiencia en L2 tuvo un efecto positivo tanto en los hablantes de español como L1 como en los hablantes de inglés, dado que los grupos con experiencia produjeron valores de VOT más precisos que los grupos sin experiencia. Sin embargo, el efecto de la experiencia en L2 en las oclusivas sonoras es menos clara que en las oclusivas sordas. Ninguno de los grupos de hablantes de español como L1 presentaron valores de VOT semejantes a los del grupo de monolingües de inglés y, en el caso de los hablantes de inglés como L1, el grupo con una experiencia moderada obtuvo valores de VOT más cercanos a los de los españoles monolingües que el grupo con más experiencia – posiblemente a raíz de diferencias relacionadas con el aprendizaje formal del español. No se encontraron evidencias claras del efecto de la experiencia en L2 en la producción de la L1. Al contrario que en percepción, los participantes produjeron las oclusivas del español y del inglés con valores de VOT diferentes, es decir, con valores mayores en inglés que en español. No obstante, en el caso de la producción de la /b/ por parte de los grupos españoles, esta diferencia solo llegó a ser significativa para los aprendices con más experiencia. Finalmente, en cuanto a la relación entre percepción y producción, no se pudo establecer una relación clara entre ambas modalidades a nivel individual, ya que los participantes que presentaban valores de VOT mayores en percepción no siempre los presentaban también en producción. Sin embargo, los resultados globales de ambos experimentos indican que los hablantes de inglés como L1 y español como L2 obtuvieron valores de VOT más semejantes a los de los hablantes nativos de la L2 en producción que en percepción, en contraposición a las premisas del *Speech Learning Model* (Flege, 1995) o del *Perceptual Assimilation Model* (Best & Tyler, 2007), que sostienen que la adquisición de la percepción en L2 precede a la adquisición de la producción. Una producción en L2 generalmente más precisa que la percepción en L2 puede deberse a problemas relacionados con el control la activación de la lengua que se está analizando. El experimento de producción – una tarea de lectura – presentaba los sonidos estudiados en un contexto monolingüe, mientras que, en el caso del experimento de percepción, los estímulos se presentaron en sílabas aisladas. Además, dichas sílabas – /pi/, /bi/, /ki/ and /gi/ – tienen más significado en inglés que en español, por lo que el inglés pudo haberse activado – total o parcialmente – tanto en el test en inglés como en el test en español. En resumen, los resultados de esta tesis indican que la experiencia en L2 puede influir la L2 y la L1, aunque se ha observado un efecto mayor de la L1 en la L2 que de la L2 en la L1. Es posible que se requiera una mayor experiencia en L2 para poder observar una influencia de la L2 en la L1 más clara, puesto que incluso los grupos más experimentados en este estudio tenían una experiencia en L2 relativamente corta – i.e., de unos 4 años. Las diferencias entre los hablantes de español como L1 y los hablantes de inglés como L1 sugieren que la influencia interlingüística puede tener lugar de diferente manera según el orden de adquisición de las lenguas y que otros factores, además de la experiencia en L2 – como el uso o aprendizaje formal de la L2 – pueden influir en ella.

Palabras clave: Percepción y producción de L2, experiencia en L2, la relación entre percepción y producción, influencia interlingüística, oclusivas del español, oclusivas del inglés.

Resum

L'adquisició d'una segona llengua (L2) pot resultar en influència interlingüística, que pot ser tant progressiva – d'una llengua adquirida amb anterioritat a una altra adquirida posteriorment –, com regressiva – d'una llengua adquirida posteriorment a una altra adquirida anteriorment. Estudis previs han mostrat exemples d'influència lingüística progressiva (e.g., Antoniou, Best, Tyler & Kross, 2010; Zampini, 1998) i regressiva (e.g., Bergmann, Nota, Sprenger & Schmid 2016; Harada, 2003), així com bidireccional (e.g., Flege, 1987; Major, 1992). Un dels factors que poden modular la influència interlingüística és l'experiència en L2 – entesa en aquest estudi com temps viscut en un context d'immersió lingüística. Concretament, l'experiència en L2 pot resultar en una percepció i producció més precises en la L2, però menys natives en la primera llengua (L1) a causa de la influència de la L2 (e.g., Flege, 1987; Levy & Law, 2010). Aquest estudi investiga l'efecte de l'experiència en L2 tant en la percepció com en la producció de les oclusives en posició inicial. Així mateix, el fet que aquest estudi examini ambdues modalitats permet que s'investigui la relació entre percepció i producció. Sis grups de parlants d'una segona llengua que es diferencien en la seva experiència en L2, així com en l'ordre d'adquisició de les llengües – parlants d'espanyol L1 i anglès L2 i parlants d'anglès L1 i espanyol L2 –, participaren en l'estudi. Per cada L1, hi participaren un grup d'aprenents de L2 sense experiència, un grup d'aprenents amb una experiència moderada en L2 que residien al seu país d'origen, un grup amb una alta experiència en L2 a un context d'immersió lingüística i un grup de monolingües funcionals. Els participants completaren un experiment de percepció i un de producció d'oclusives en cada llengua. L'experiment de percepció consistia en dos tests d'identificació – un per les oclusives bilabials i un altre per las oclusives velars – que presentaven estímuls d'un continu de VOT. Els mateixos estímuls van ser emprats per examinar la percepció d'ambdues llengües – i.e., anglès i espanyol. L'experiment de producció consistia en la lectura de frases que presentaven oclusives en posició inicial absoluta.

Els parlants d'espanyol L1 i els parlants d'anglès L1 van obtenir diferents resultats en l'experiment de percepció. D'una banda, els parlants d'espanyol L1 i anglès L2 presentaren indicis d'influència lingüística bidireccional, ja que l'experiència en L2 va resultar en una percepció més acurada en la L2, però menys nativa en la L1. A més a més, cal destacar que els espanyols aprenents d'anglès categoritzaren les oclusives de la L1 y

la L2 de manera similar, amb l'excepció de la categorització de /k/-/g/ per part del grup amb una experiència moderada. D'altra banda, no es trobaren indicis d'influència de l'experiència en L2 en els parlants d'anglès L1, donat que aquests categoritzaren tant les oclusives de l'anglès com les de l'espanyol segons els valors de VOT de l'anglès. En el cas de la producció, els resultats indiquen que l'experiència en L2 va tindre un efecte positiu tant en els parlants d'espanyol L1 com en els parlants d'anglès L1, ja que els grups amb experiència produïren valors de VOT més acurats que els grups sense experiència. No obstant, l'efecte de l'experiència en L2 en les oclusives sonores és menys clara que en el cas de les oclusives sordes. Cap dels grups de parlants d'espanyol L1 presentaren valors de VOT semblants als del grup de monolingües d'anglès i, en el cas dels parlants d'anglès L1, el grup amb una experiència moderada va obtenir valors de VOT més propers als dels espanyols monolingües que el grup amb més experiència – possiblement arran de diferències relacionades amb l'aprenentatge formal de l'espanyol. No es trobaren evidències clares de l'efecte de l'experiència en L2 en la producció de la L1. Al contrari que en percepció, els participants produïren les oclusives de l'espanyol i de l'anglès amb valors de VOT diferents, és a dir, amb valors més grans en anglès que en espanyol. No obstant, en el cas de la producció de la /b/ per part dels grups espanyols, aquesta diferència només va ser significativa pels aprenents amb més experiència. Finalment, pel que fa a la relació entre percepció y producció, no es va poder establir una relació clara entre ambdues modalitats a nivell individual, ja que els participants que presentaven valors de VOT majors en percepció no sempre els presentaven també en producció. Tanmateix, els resultats globals d'ambdós experiments indiquen que els parlants d'anglès L1 i espanyol L2 van obtenir valors de VOT més propers a aquells dels parlants natius en producció que en percepció, en contraposició a les premisses de l'*Speech Learning Model* (Flege, 1995) o del *Perceptual Assimilation Model* (Best & Tyler, 2007), que sostenen que l'adquisició de la percepció en L2 precedeix l'adquisició de la producció. Una producció en L2 generalment més precisa que la percepció en L2 pot ser deguda a problemes relacionats amb el control l'activació de la llengua que s'està analitzant. L'experiment de producció – una tasca de lectura – presentava els sons estudiats en un context monolingüe, mentre que, en el cas de l'experiment de percepció, els estímuls es presentaren en síl·labes aïllades. A més, aquestes síl·labes – /pi/, /bi/, /ki/ and /gi/ – tenen més significat en anglès que en espanyol, de manera que l'anglès va poder haver-se activat – totalment o parcial – tant en el test en anglès com en el test en espanyol. En resum, els resultats d'aquesta tesi indiquen que l'experiència en L2 pot influir tant la L2 com la L1, tot i que s'ha observat

un efecte major de la L1 en la L2 que de la L2 en la L1. És possible que es requereixi una major experiència en L2 per poder observar una influència de la L2 a la L1 més clara, ja que fins i tot els grups més experimentats en aquest estudi tenien una experiència en L2 relativament curta – i.e., d'uns 4 anys. Les diferències entre els parlants d'espanyol L1 i els parlants d'anglès L1 suggereixen que la transferència interlingüística pot donar-se de diferent manera segons l'ordre d'adquisició de les llengües i que altres factors, a part de l'experiència en L2 – com l'ús o l'aprenentatge formal de la L2 – hi poden influir.

Paraules clau: Percepció i producció de L2, experiència en L2, la relació entre percepció i producció, transferència interlingüística, les oclusives de l'espanyol, les oclusives de l'anglès.

Table of contents

List of Figures

List of Tables

1. Introduction

Research in L2 speech acquisition has shown that the L1 and the L2 interact, since, as some L2 speech models claim, both systems coexist in a common phonological space (Best & Tyler, 2007; Flege, 1995). As a result, the L1 may influence the L2 and the L2 may as well have an effect on the L1, as previous studies have revealed (e.g., Caramazza, Yeni-Komshian, Zurif & Carbone, 1973; Flege, 1987; Hazan & Boulakia, 1993; Sancier & Fowler, 1997; Williams, 1977). This interaction between the two systems hinders the acquisition of new separate categories for L2 sounds, which may initially be perceived as realizations of an L1 category (Best & Tyler, 2007; Flege, 1995, 2002, 2007).

Flege (1987) points out that not all L2 phones are acquired the same way and taxonomizes L2 phones from an L1 standpoint as 'new' – i.e., without a clear L2 counterpart –, 'similar' – i.e., with a clear L1 counterpart, sharing some acoustic properties, although differing in others – and 'identical' phones – shared between the two languages. According to the Speech Learning Model (SLM, Flege, 1995, 2002, 2007), similar phones are considered to be the most difficult to acquire by an L2 learner given that the similarity between the L1 and the L2 phone may prevent the creation of a new L2 category. As a matter of fact, similar phones are believed to be rarely produced with target-like values. By contrast, new phones, which may initially be produced less accurately than similar phones, are more likely to be successfully acquired given enough experience with and exposure to the target language, precisely because they are more distinguishable from the closest L1 phone(s) than similar phones are.

The sounds studied in the present thesis are English and Spanish bilabial and velar stops, which can be classified as similar phones. The main feature that distinguishes

English and Spanish stops is the use of VOT to contrast voicing. English voiceless stops present long-lag VOT, that is, there is a delay in the onset of voicing after the release of the burst, whereas Spanish stops present short-lag VOT – i.e., voicing starts after the release of the stop. As for voiced stops, they are produced with voice-lead VOT – i.e., voicing starts before the release of the stop – in Spanish, whereas in English they are often produced with short-lag VOT, although instances of voice-lead productions have been reported in the literature (e.g., Lisker & Abramson, 1964).

L2 experience – understood as the amount of time spent in an L2-speaking country in some previous studies (e.g., Cebrian, 2006; Flege, 1987; Flege, Bohn & Jang, 1997) and in the present thesis – has been found to be a factor that helps producing and perceiving L2 phones more accurately (e.g., Bohn & Flege, 1990; Flege et al., 1997; Lev-Ari & Peperkamp, 2013; Levy & Law, 2010). Moreover, L2 experience has also been reported to result in an influence of the L2 on the L1, both in perception and in production (e.g., Cebrian, 2006; Dmitrieva, 2019; Flege, 1987). The present thesis was partly inspired by Flege (1987), which tested the production of L1 and L2 /t/ – a similar phone – by L1-English L2-French and L1-French L2-English speakers. Flege's study also investigated the production of the French vowels /i/ and /y/ – a similar and a new phone, respectively. This thesis will focus on stops – both voiced and voiceless – of two different places of articulation – and will investigate both perception and production.

The main goal of this study is to examine the effect of L2 experience on L2 and L1 perception and production of English and Spanish bilabial and velar stops by L1-Spanish L2-English and L1-English L2-Spanish speakers. Moreover, the relationship between perception and production of stops – both in the L1 and in the L2 – will be investigated. It is hypothesized that L2 experience will result in a more target-like production and perception of L2 stops, but that it may also bring about less native-like perception and production of the L1 due to L2 influence.

In order to do so, a total of eight mirror-image groups – four L1-Spanish and four L1-English groups – have participated in the study. Participants included a control group of functional monolinguals, an inexperienced group of L2 learners, a group of moderately experienced L2 learners living in an L1 setting and a group of experienced L2 learners living in an immersion setting for each L1. In order to test perception, participants had to complete a labelling task for each pair of stops in both languages. As for production, participants had to complete a sentence reading task in English and in Spanish. The control groups only completed the experiment in their language.

This study will shed light on the effect of L2 experience on L1 perception, which has received little attention in previous research. The fact that both perception and production are examined will allow us to investigate the relationship between the two modalities in the L1 as well as in the L2, as conflicting results have been reported in the literature (e.g., Perkell, Guenther, Lane, Matthies, Stockmann, Tiede & Zandipour, 2004; Rallo Fabra & Romero, 2012; as opposed to Lev-Ari & Peperkamp, 2013; Sheldon & Strange, 1982; Shultz, Francis & Llanos, 2012). Moreover, this thesis will examine the influence of L2 experience on the relationship between perception and production in the L2 and in the L1, which is also an understudied topic in previous research. Furthermore, the inclusion of two mirror-image populations in terms of L1 and L2 allows the investigation of the direction of language learning, which is key to understanding the process of acquisition of L2 speech. Examining two populations will also enable us to make stronger claims and generalize the findings in the event that comparable results are obtained by the matching L1-Spanish and L1-English groups. If contradictory results are found for two language populations, we will be able to seek language specific trends in crosslinguistic influence and to investigate individual factors – such as amount of L2 instruction, L2 use or age of L2 acquisition – that may differ across groups.

2. Literature review

This section will review aspects that are relevant to the research topics discussed in this Ph.D. dissertation. First, the main L2 speech acquisition models will be summarized. Next, previous studies that have shown instances of crosslinguistic influence in L2 speakers – including both the effect of the L2 on the L1 and the effect of the L1 on the L2 – will be discussed. Moreover, the research on the main factors that contribute to crosslinguistic influence will be reviewed, paying special attention to L2 experience. This will be followed by a section on the relationship between perception and production, both in the L1 and in the L2. The chapter ends with a description of the main cues for voicing in stop sounds, particularly voice onset time (VOT), comparing its use in Spanish and English.

2.1. L2 Speech Acquisition Models

First language (L1) speech acquisition, as well as second language (L2) speech acquisition – in infancy and adulthood – has been widely addressed in the literature. Research on L1 acquisition has examined the biological grounds of language learning as well as the process of acquisition (e.g., Chomsky, 1957; Lenneberg, 1967). Regarding speech, research in the area of L1 acquisition attempted to establish which skills and which phonetic and phonological abilities are innate in children and how children learn to focus on language-specific features and gradually lose the ability to attend to features or cues that are not relevant for the target language (e.g., Best, 1994, 1995; Kuhl, 1993). The acquisition of second language speech is different from the acquisition of first language speech in two main ways: (1) There is a difference in cognitive maturation (e.g., Lenneberg, 1967) (2) and the learner already has an existing phonological system – i.e., the L1 – through which non-native phones may be categorized (e.g., Iverson & Kuhl, 1995; Wode, 1978). Regarding maturational constraints, early language acquisition theories, such as the so-called Critical Period Hypothesis, claimed that both first and second language speech cannot be mastered after adolescence due to biological and maturational constraints (e.g., Lenneberg, 1967; Long, 1990; Patkowski, 1990; Scovel, 1988). In a similar fashion, researchers like Oyama (1976, 1978) defend the existence of a Sensitive Period – i.e., a less strict variant of the Critical Period – for the acquisition of the phonological system of an L2. As for the effect of the L1 on L2 acquisition, the Perceptual Magnet Effect model claims that the L1 phonological prototypes have a 'magnet effect' on non-native phones, which are non-prototypical exemplars – i.e., non-native phones are categorized as prototypes of the most similar L1 category – (Kuhl, 1991). Similarly, some theories claim that the L2 is perceived through the L1's phonological 'grid', resulting in a deviant perception from that of native speakers (Trubetskoy, 1969; Wode, 1978). Current L2 speech theories attempt to disentangle the processes that intervene in L2 acquisition and explain why certain non-native features are rarely mastered by L2 speakers. This is done by addressing the initial stages of learning an additional language – paying special attention to the role of the L1 – and studying the different stages in the process of acquisition up to final attainment (e.g., Best & Tyler, 2007; Escudero, 2005, 2009; Flege, 1995, 2002, 2003, 2007). Moreover, other factors such as language experience or age of learning are taken into consideration in some of the most influential L2 models (e.g., Flege, 1995, 2002, 2003, 2007). Even though most L2 acquisition theories address the effect of the L1 on later learnt languages, recent studies have gained an interest in investigating the effect that acquiring an L2 may have on the existing L1 (e.g., Flege, 1987; Fowler, Sramko, Ostry, Rowland & Hallé, 2008; Major, 1992). Still, L1 speech attrition has received less attention than morphosyntactic attrition. This section will review the main L2 speech acquisition models, including Kuhl's Native Language Magnet Model (NLMM, (Kuhl, 1991, 2000), Flege's Speech Learning Model (SLM, Flege 1995, 2002, 2003, 2007), Best and Tyler' Perception Assimilation Model (PAM, Best, 1994, 1995; PAM-L2, Best and Tyler 2007), Grosjean's Language Mode Theory (1997, 1998, 2001, 2012) and Escudero's Second Language Linguistic Perception Model (L2LP, 2005, 2009).

2.1.1. Native Language Magnet Model

In order to explain the process of speech acquisition, Kuhl (1991, 2000) established a difference between categorical perception and language-specific development – what she called 'mapping'. Whereas categorical perception, which is universal, is also shown by animals and does not seem to apply only to speech, language mapping is exclusive to humans and is language specific. These assumptions challenged the belief that children's ability to parse speech phonetically relies on a genetic mechanism that evolved just for the sake of speech processing. Therefore, the innate ability to process speech is not based on Universal Grammar, but on the strategies used to recover and process speech in the ambience.

Moreover, Kuhl and her colleagues (Iverson & Kuhl, 1995; Kuhl, 1991, 2000; Kuhl, Conboy, Coffey-Corina, Padden, Rivera-Gaxiola, & Nelson, 2008) proposed the Native Language Magnet Model (NLM), which posits that experience with the ambient language produces a filter with language-specific acoustic dimensions through which speech is perceived. Language mapping results in the Perceptual Magnet Effect; when infants are exposed to speech sounds, prototypes – i.e., good instances of a phonetic category – function as a magnet for other stimuli within the same category. In other words, all stimuli considered as belonging to the same L1 category – both prototypical and non-prototypical – are perceived as instances of a given category. As a result of L1 mapping, learning a second language becomes more difficult, since L2 specific features of speech will be perceived through the L1's filter. Thus, L2 speech acquisition involves the development of two different mappings – i.e., an L1 and an L2 mapping.

2.1.2. Speech Learning Model

Flege challenged the critical period hypothesis and posited that 'the mechanisms and processes used in learning the L1 sound system […] remain intact over the life span and can be applied to L2 learning' (Flege, 1995: 239). In Flege's view, there are causes other than age that might prevent an L2 learner from speaking the target language in a native-like manner, such as age of learning (AOL), age of arrival (AOA) to a new country, amount and type of input received and amount of language use (e.g., Flege, 1981, 2007; Flege, Yeni-Komshian & Liu, 1999; Piske, Mackay & Flege, 2001). These factors, along with others such as L2 experience, may determine the relationship between the L1 and the L2 and how they interact (see section 2.3). It is this interaction that determines which sounds will most likely be pronounced successfully and which ones will be more difficult to master.

In order to understand how L1 sounds may affect the acquisition of L2 sounds, Flege described L2 phones from an L1 standpoint as 'new', 'similar' and 'identical' L2 phones (e.g., Bohn & Flege, 1992; Flege, 1987, 1988a, 1988b 1991c). New sounds are those which differ considerably from L1 phones in acoustic terms. At first, they may be perceived as the closest L1 phone, but eventually they are expected to be distinguished from the L1 phones and be acquired as a new category. For example, the English vowel /æ/ is a new phone for a German learner and the closest sound in their L1 is /ɛ/ (Bohn & Flege, 1992). Conversely, similar sounds are those which have a clear L1 counterpart, since they share most acoustic features. For instance, an English /p/ is a similar L2 phone for a Spanish speaker, given that place and manner of articulation are shared between the L1 and the L2, but there is a difference in terms of VOT duration. In spite of the fact that they resemble an L1 phone, according to Flege, similar phones are the most difficult type of sounds to acquire and are rarely produced 'authentically', that is, in a native-like manner. The equivalence classification hypothesis posits that similar phones will be perceived as the closest L1 phone and, therefore, a new L2 category will not be formed initially because these phones will be readily substituted by the L1 counterpart (Flege, 1987, 1988a, 1988b). Still, language experience may help to create a new category for the L2 sound. Identical phones are those which are shared with the L1 and, thus, do not need to be learnt. For example, English and Dutch /ɪ/ are acoustically the same (Flege, 1988b).

Flege's proposals gave shape to the Speech Learning Model (SLM, Flege, 1995, 2002, 2003, 2007[1]), which thoroughly addressed the issue of how the L1 and the L2 interact. According to the SLM, L1 and L2 phonetic categories are comprised in a common phonological space and interact in two ways, namely through 'category assimilation' and through 'category dissimilation'. Category assimilation will take place when category formation for an L2 phone is blocked, whereas category dissimilation occurs when a new L2 category is formed and it needs to be distinguished from the existing one. It should be noted that the SLM claims that perception leads production and that, therefore, the inability to correctly perceive an L2 sound will lead to its inaccurate production (see section 2.4 for further detail). As a matter of fact, according to Flege, category formation depends on the perceived phonetic (dis)similarity between the L2 and the L1 phones. Thus, the closest an L2 phone is to an L1 sound, the more likely it is for

[1] A revised version of the SLM is in progress and will most likely be published in the next few months. However, it has not been published in time to be discussed in the present thesis.

category formation to be blocked. A possible outcome will be that the L2 category will be ‘assimilated’ to an existing L1 phone. L1 and L2 categories may then merge. According to the nature of the input received, the merged category will show properties that resemble the L1 or the L2 to a greater or lesser extent. In other words, the same category will be realized using properties that approximate to the L1 to a greater extent when the input received is L1-like, whereas the merged category will resemble the L2 to a greater extent when the input is L2-like. This hypothesis was referred to as the Merger Hypothesis (MH) by Flege (1987). Flege found evidence supporting the MH. Native French speakers of English living in the US produced English /t/ with shorter VOT values – i.e. more French-like – than English monolinguals, but also produced longer VOT values in French – i.e., more English-like – than French monolinguals. In other words, their /t/ productions in the L1 and in the L2 did not differ significantly. Thus, the interference between L1 and L2 categories can be bidirectional and the L1 phonetic categories established in childhood may experience changes after the acquisition of an L2.

Conversely, when the L2 phone is accurately discriminated from its closest L1 phones, a new category will be formed. It is in this case when category dissimilation may take place by increasing the phonetic differences between the newly established L2 phone and the most similar L1 phone. As a result, L1 and L2 categories may not be identical to those of monolingual speakers. For instance, Flege and Eefting (1986, 1987) found that early Spanish-English bilinguals produced English stops with shorter VOT values than English speakers. Still, their productions of /t/ also presented shorter VOT values than those of Spanish monolinguals, presumably in order to make their L1 and L2 categories more phonetically distinct.

2.1.3. Perceptual Assimilation Model

Best proposed the Perceptual Assimilation Model (PAM), which postulated a direct realist view of non-native speech perception (Best, 1995). According to PAM (Best McRoberts & Sithole, 1988; Best, 1994, 1995), speech perception is based on articulatory gestures, while the SLM claims that it is based on acoustic-phonetic cues. PAM initially addressed the perception of non-native contrasts by naïve listeners, whereas PAM-L2 (Best & Tyler, 2007) makes a distinction between beginners – which perceive L2 phones as functional monolinguals – and experienced speakers. The L2 model aims to explain the perceptual learning of L2 learners by extending PAM's premises and replying to the SLM's postulates (Flege, 1995, 2002, 2007). Although Best and colleagues' proposal has some commonalities with the SLM – e.g., both models consider that the L1 and the L2 coexist in a common space – the two models differ in important ways. Whereas the SLM addresses individual phones, the PAM focuses on pairs of phonological contrasts. Moreover, the SLM pays attention to L2 production, whereas the PAM is directed to cross-language perception (Bohn & Best, 2010). Moreover, the PAM-L2 rejects the idea of the existence of mental representations of L2 categories, as they are believed to be unnecessary. Both L1 and L2 phones are directly perceived using 'articulatory gestures' and, therefore, mental representations of phonetic categories are not needed for L2 perceptual learning to take place (Best & Tyler, 2007).

As it has been mentioned above, the PAM-L2 focuses on pairs of L2 contrasts and how they are perceived by L2 learners. When an L1 and an L2 category are similar, assimilation of the L2 phone to the most similar L1 phone – in articulatory terms – is expected to occur (Best & Tyler, 2007). Assimilation of an L2 contrast to the L1 may take place in the following ways:

(1) Two L2 phones that constitute an L2 contrast are perceived as two equally acceptable L1 phones – two category assimilation. In this case excellent discrimination of the two phones is expected. The PAM-L2 does not address this particular kind of assimilation, presumably because discrimination between the two phones is considered to be sufficient already at the initial stages of L2 acquisition.

(2) Assimilation of an L2 pair to the L1 may also take place through single category assimilation, which involves the assimilation of both categories in an L2 contrast to a single L1 phone. As a result, a poor discrimination of the L2 pair is expected initially. An L2 learner will need to create a new category for at least one of the phones in order to be able to distinguish them. The creation of a new category will depend on the similarity of the L2 phones to the L1. If the new phone is already regarded as a good exemplar of the equivalent L1 phone, the creation of a new category is less likely to occur. Moreover, the need to learn the contrast for communicative purposes – i.e., if it is used contrastively in many words, or if the words where the phones appear are high-frequency – will also play a role in the creation of new categories.

(3) L2 assimilation to the L1 may also take place through category goodness. In this case, the L2 pair is assimilated to a single L1 category, but one of the L2 phones is perceived as a better exemplar of the L1 phone. In this case, moderate discrimination is expected. PAM-L2 added that a new phonetic and phonological L2 category is likely to be formed for the most deviant phone, whereas the better exemplar of the L1 phone will be perceived as equivalent to the existing L1 category, both phonetically and phonologically. Eventually, the learner may create a new category for the 'better-fitting' phone, although this will depend on the degree of similarity between the L2 phone and the existing L1 phone.

(4) Furthermore, only one phone may be perceived as a good exemplar of the L1, whereas the other one may be left uncategorized – uncategorized vs. categorized. In this case, no further perceptual learning is expected, as the two phones will be distinguished from all other L2 categories, either through two category assimilation or uncategorized-categorized. Another possibility would be that one L2 phone is perceived as deviant from the closest L1 category. In this case, the L1 and the L2 phones will be considered as two realizations of the same category, that is, as equivalent at a phonological level, but they will be perceived to be different at a phonetic level. An example would be the case of English and French /r/.

(5) Finally, assimilation of an L2 contrast to the L1 may not take place if the naïve listener does not perceive any of the phones as belonging to any L1 category. In the event that both uncategorized L2 categories are distant from one another – from the perspective of the L1 phonological space – the learner will be able to distinguish them and will likely create new categories for each of the phones. However, if the sounds that constitute the contrast are close to each other, the learner will have greater difficulty in perceiving the two phones as different and a single new category will be created.

2.1.4. Language Mode Theory

Grosjean's Language Mode Theory (Grosjean, 1997, 1998, 2001, 2012) was proposed as an attempt to provide an answer to why bilinguals speak differently depending on interlocutors and context. Language mode refers to 'the state of activation of the bilingual's languages and language processing mechanism, at a given point in time […]' (Grosjean, 2001: 2). There is a language mode continuum which ranges from purely monolingual – where one language is activated and the other one is deactivated, although never completely – to a completely bilingual mode, where both languages are operative.

In a bilingual mode, the speaker uses one base language but parallelly activates the other one – to a greater or lesser extent. The base language and the degree of activation of the parallel language operate separately. Therefore, there could be a change in the base language that is being used or a change in the degree of activation of the parallel language without one affecting the other. Furthermore, the selection of the base language and the degree of activation of the language that has not been selected depend on series of aspects, including the interlocutor – i.e., whether they are monolingual or bilingual speakers –, the situation – i.e., whether it requires one or both languages –, the speaker's language dominance, or the linguistic and sociolinguistic characteristics of community where they live (Grosjean, 2001). Thus, a monolingual mode will be activated if the interlocutor or the situation are monolingual. For instance, a bilingual child will activate a monolingual mode when speaking to a monolingual parent. Conversely, if the interlocutor is a bilingual speaker and feels comfortable mixing languages, a bilingual mode will be activated. This may happen, for example, when a bilingual child is speaking to a bilingual parent. If the topic of the conversation cannot be covered in the primary language, code-mixing is even more likely to be happen and, thus, a bilingual mode will be activated. An intermediate language mode will be activated as a result of a combination of some of the factors described above (e.g., if the interlocutor can speak both languages but is not fully proficient in one of them, if the interlocutor does not like mixing languages or if the situation is formal).

Moreover, the ability of the bilingual speaker to deactivate the language that is not being used also depends on a number of factors that may be either intrinsic to the speaker or related to context. It should also be noted that movement along the language continuum may take place at any given point according to the factors mentioned above. Individuals may differ in their tendency to travel along the VOT continuum, as some may rarely be in a bilingual mode (i.e., mix languages) whereas others may tend to activate both languages frequently. One individual factor that plays a key role on language mode activation and, thus, on code-mixing is language dominance. Bilinguals who are highly-dominant in one language – which are referred to by Grosjean as highly language dominant bilinguals – tend to activate their dominant language to a greater extent when the language used is their weaker language than the other way around. In fact, highly language dominant bilinguals will only be able to deactivate their dominant language in a very controlled monolingual context in their weaker language. It should also be noted that language mode may change from monolingual to bilingual if the bilingual speaker/listener is exposed – or believes that might be exposed at some point – to elements of the language that is not being used. This process can also apply to readers who are exposed to lexical elements from the other language.

2.1.5. Second Language Linguistic Perception Model

More recently, Escudero proposed the Second Language Linguistic Perception Model (L2LP, Escudero, 2000, 2005, 2009), which, just as the PAM-L2, addresses L2 perception. In line with other models, such as the SLM, the PAM-L2 or the NLMM, the L2LP claims that the L1 is the baseline for L2 acquisition. However, Escudero's model differs substantially from the SLM and the PAM-L2 in that it rejects the idea of a common phonological space and posits that bilingual speakers have separate L1 and L2 systems. As a matter of fact, the two systems are believed to be separate from the initial stage. This is explained through the 'Full Copying Hypothesis', which claims that, in order to create a separate system for the L2, the L1's perceptual mapping is copied. The new system will then develop through learning and the mapping will be adjusted in order to present more target-like phonetic properties. Just as in L1 acquisition, L2 learners will learn through ‘auditory-driven category formation and lexicon-driven category boundary shifts’ (Escudero, 2005: 161).

As a result of full copy and phonemic equation – i.e., the mapping of the L2 categories on the L1 –, the initial stage of L2 acquisition may involve three possible scenarios for L2 sound perception, namely a ‘similar scenario’ – i.e., there is a phonologically equivalent but phonetically different L1 category for the L2 phone –, a ‘new scenario’ – the L1 does not a have a phonologically equivalent category to the L2 phone – and the ‘subset scenario’ – one L2 category is equated to two L1 categories. Regarding the acquisition of similar phones, L2 learners will create a copy of the most similar L1 category, which will then develop to become more target-like. The similar phone scenario would involve categories such as the English and French /t/ and /d/. An L1-French learner of English will equate English /t/ to French /t/ and English /d/ to French /t/. Thus, L2 development will require a shift in the category boundary between the two categories in order to account for phonetic differences between the two languages. As for the new scenario, the L2 phone may be mapped to an existing L1 category, which may already have been equated to another L2 category. For example, both English vowels /i/ and /ɪ/ will be equated to the L1 Spanish vowel /i/, resulting in too few representations. Perceptual development would involve the creation of ‘new categories and perceptual mappings’ as well as the ‘integration of the new categorized dimensions with already categorized dimensions’ (Escudero, 2005: 125). The subset scenario involves the opposite process: one L2 category may have been mapped onto two L1 categories, resulting in too many representations. For instance, the L2 Spanish vowel /e/ may be equated to both /ɪ/ and /ɛ/ by an L1-Dutch speaker. In this case, L2 development will require a shift in category boundary and the reduction of the lexical and perceived categories. Given the complexity of the tasks required for perceptual learning to take place, the L2LP claims that the new scenario is the most challenging one for an L2 learner, followed by the subset scenario, whereas similar phones allow the reusing of existing L1 categories. This view opposes Flege's (e.g., 1987) claim that similar phones are more difficult to master than new phones. It also should be noted that the L2LP posits that L2 learners can successfully develop L2 categories in all three scenarios, which would result in a target-like phonological system and in an optimal distinction between the L1 and L2 systems. Even though the L1 and the L2 systems are separate, intermediate perceptions – i.e., with values between the two languages – may occur. Escudero's view on L1-L2 interference is in line with Grosjean's Language Mode theory (Grosjean, 1997, 1998, 2001) and claims that any intermediate perception is caused by the simultaneous activation of the two parallel systems, which is triggered by the nature of the input received.

In this section, frequently cited L2 speech acquisition models have been reviewed, including the NLM, the SLM, the PAM-L2, the Language Mode Theory and the L2LP. Flege's SLM and Escudero's L2LP are especially relevant for this research as the sounds that are investigated are similar phones – i.e., sounds that share most phonetic traits but differ in others. Moreover, the testing of both the L1 and the L2 may result in evidence in favour or against some of the main hypotheses proposed by the SLM – including the merger hypothesis – and may shed some light on how the two languages interact, a matter on which the SLM and the L2LP have different views. Finally, Grosjean's Language Mode Theory and the L2LP's understanding of how language activation is the main source of intermediate perceptions between the L1 and the L2 phonological systems are crucial in this study because language activation needs to be triggered according to the language that is being tested in order to obtain reliable results. After having presented the main theoretical aspects of L2 acquisition and how the L1 and the L2 may interact during the process of L2 speech learning, the following section will offer a review of previous studies that have found crosslinguistic influence in L1 and L2 speech, especially regarding stops.

2.2. Crosslinguistic influence in L2 speakers

As most L2 speech models predict, the acquisition of a second language may bring about crosslinguistic influence (CLI, e.g., Best & Tyler, 2007; Flege, 2002; Grosjean, 2012). Previous studies have reported several ways in which the L1 and the L2 may – or may not – interact, namely L1 influence on the L2, bidirectional L1-L2 interaction, unidirectional L2 influence on the L1 and no L1-L2 interaction. Given the large number of studies on CLI, this review is limited mostly to studies on VOT.

The most common type of crosslinguistic interaction in the literature – when it comes to both production and perception – is L1 influence on the L2 (e.g., Caramazza, et al. 1973; Carlson, Goldrick, Blasingame & Fink, 2016; Flege, 1990, 1991b; Flege, Mackay & Piske, 2002; Hazan & Boulakia, 1993; Zampini, 1998). For example, Caramazza et al. (1973) found that, although early Canadian French-English bilinguals could produce voiceless stops with aspiration, they failed to modify voiced stops, as they seemed to be using their French prevoiced categories in the L2. Similar results have been found for English learners of Spanish. Zampini (1998) investigated the production of /p/ and /b/ by English learners of Spanish and found that, at the end of an undergraduate phonetics course, they had improved their production of Spanish /p/, although it still had longer – more English-like – VOT values. Nevertheless, the learners were not able to produce Spanish /b/ consistently with prevoicing and, in fact, their Spanish /b/ did not differ from their English /b/ in terms of VOT. Flege (1990) found that Spanish adult learners of English living in the US produced intermediate VOT values for /t/ in English, that is, they used shorter, more Spanish-like VOT values. Similar findings have been reported regarding perception. Hazan and Boulakia (1993) found that French-dominant bilingual users of English perceived the /p/-/b/ contrast in English similarly to French monolinguals, that is using their L1 categories. Participants in this study included mostly adolescents in an international school in France, but also their teachers.

Some works have tested both the L1 and the L2 and have found a bidirectional effect, that is, an influence of the L1 on the L2 as well as an influence of the L2 on the L1 (e.g., Caramazza, et al. 1973; Flege, 1987; Fowler et al., 2008; Hazan & Boulakia, 1993; Major, 1992; Sancier & Fowler, 1997; Williams, 1977). For instance, Fowler et al. (2008) analyzed the production of /p, t, k/ by French-English simultaneous bilinguals and found that their VOT productions in French were longer than those of French monolinguals. Still, their VOT values in English were shorter than those reported for English monolingual speakers. In spite of the fact that the VOT values of their L1 and L2 were getting closer, bilingual speakers had separate categories for French and English. Similarly, Major (1992) examined the production of /p, t, k/ by L1 English speakers who immigrated to Brazil as adults. Their productions of VOT in both Portuguese and English were found to deviate from those of monolinguals of both languages. Furthermore, Major found a negative correlation between L1 and L2 production accuracy. In other words, a less native-like production in English was correlated with a more native-like production in Portuguese. However, English and Portuguese VOT values were not identical. Hazan and Boulakia (1993) also examined bilingual production of voiced stops. French-English bilinguals were reported to produce /p/ with significantly shorter VOT values than English monolinguals and /b/ with significantly longer prevoicing (more detailed results are provided in the section 2.3**)**. Flege (1987) tested late English-French bilinguals differing in terms of L2-experience. It was found that the most experienced groups presented a bidirectional influence in their production of /t/, as intermediate VOT values were used in both languages; that is, English stops were produced with shorter VOT – more French-like – values whereas French /t/ was produced with longer VOT than we would expect for a French speaker due to an influence from English.

Previous studies have also found bidirectional L1-L2 influence in perception. Williams (1977) analyzed Spanish-English early bilingual's /p/-/b/ perceptual categorization using both an identification and discrimination task that presented stimuli ranging on a VOT continuum. Bilinguals were found to vary from both English and Spanish monolinguals. In most cases, participants seemed to be using the same criteria for both languages. However, according to their results, participants could be split into two groups: those who seemed to be using intermediate values in both languages – that is, a boundary in Spanish with higher VOT values than Spanish monolinguals' and a boundary in English with lower VOT values than English monolinguals' – and those who appeared to use Spanish values in both cases. The latter, thus, seem to be an example of unidirectional L1 influence on the L2. Moreover, Caramazza et al. (1973) found that early Canadian French-English bilinguals perceived L1 and L2 stops similarly.

A few studies have also found a unidirectional L2 influence on the L1, especially regarding bilinguals living in a second language context (e.g., Bergmann, Nota, Sprenger & Schmid 2016; Harada, 2003). Harada (2003) assessed VOT production by Japanese families living in the US. Early bilinguals were reported to produce voiceless stops with significantly longer VOT values in Japanese, whereas their production of English stops did not differ from those of English monolinguals. Still, they had separate L1 and L2 categories. Bergmann et al. (2016) found that L1-German L2-English speakers in an immersion setting produced German /a:/ and /l/ with values that differed significantly from German monolinguals. However, their L2 was not tested, so it cannot be claimed that crosslinguistic influence was unidirectional.

Finally, some studies have failed to find any crosslinguistic interaction in bilingual speakers (e.g., Abramson & Lisker, 1973; Antoniou, Best, Tyler & Kross, 2010, 2010; Williams, 1977). For example, Antoniou et al. (2010) found that Greek-English heritage speakers living in Australia produced velar and coronal stops in initial position in a native-like manner both in English and in Greek. Similarly, Williams (1977) failed to find any significant differences in the production of English and Spanish /p/ between bilingual and monolingual speakers of each language. Still, some differences were observed regarding voiced stops, as bilinguals were found to prevoice /b/ more frequently than monolinguals. In spite of the fact that previous works have reported that English monolingual speakers may also produce /b/ with voice-lead (e.g., Lisker & Abramson, 1964), the use of negative VOT values in voiced stops seems to be more common among Spanish-English bilinguals, as they may be transferring the use of prevoicing from their L1.

In perception, Abramson and Lisker (1973) also failed to find differences in the voiced-voiceless identification of stops between Spanish-English bilinguals who had recently moved to the US and monolinguals of each language. As a matter of fact, bilinguals were reported to have different categories for English and Spanish stops. Even though the study did not test monolingual speakers, the authors claim that their perception of English and Spanish stops matches the monolingual values reported in the literature.

So far, the different types of crosslinguistic influence – and the lack of it – have been discussed. When there is no crosslinguistic interaction, we expect that L2 speakers present separate L1 and L2 categories with native/target-like values (e.g., Antoniou et al., 2010). When crosslinguistic influence takes place, bilinguals may present separate intermediate categories (e.g., Casillas & Simonet, 2018; Fowler et al., 2008; Gonzales &

Lotto, 2013; Harada, 2003; Williams, 1977). L2 speakers may also present a single shared category, which may be more L1-like (e.g., Caramazza et al., 1973; Hazan & Boulakia, 1993; Zampini, 1998) or present intermediate values between the two languages – i.e., a merged category, in SLM terms (e.g., Flege 1987; Williams 1977). To my knowledge, no previous research has found instances of an L2-like single category.

Some previous studies have also found evidence for separate L1 and L2 categories (Antoniou et al., 2010; Casillas & Simonet, 2018; Elman, Diehl & Buchwald, 1977; Flege, 1990; Harada, 2003; Zampini, 1998). For instance, Antoniou et al. (2010) found that Greek heritage speakers in Australia produced both L1-Greek and L2-English initial /p, t, b, d/ in a native-like manner, although that was not always the case in medial position. Previous research has also reported instances of separate categories that differed from those of monolinguals of one of the two languages. Flege (1990) found that late Spanish learners of English produced native-like L1 initial /t/ but presented intermediate values in the L2. The opposite finding was reported in Harada (2003), as Japanese-English early bilinguals produced English voiceless stops authentically but presented significantly longer VOT values in their L1 than Japanese monolinguals. Regarding perception, Casillas and Simonet (2018) found that simultaneous English-Spanish bilinguals presented separate L1 and L2 /p/-/b/ category boundaries. Moreover, the authors reported that, after a seven-week intensive course, beginning English learners of Spanish already started to make a difference between the L1 and L2 categorization of /p/-/b/. Zampini (1998) reported similar results for English learners of Spanish, who, after completing a course on Spanish phonetics, presented separate L1 and L2 /p/-/b/ boundaries, as an identification test using the non-words 'pada-bada' revealed. That is, the English learners presented significantly later category boundaries for English than for Spanish. However, it should be noted that the same task was completed three times during the course and, thus, results may be influenced by task-familiarity.

As for L2 speakers using a single category in both languages, Hazan and Boulakia (1993) reported that French dominant early French-English bilinguals produced both French and English /b/ with prevoicing. Flege (1987) found that late French-English bilinguals in an immersion setting presented a merged category, that is, produced /t/ with intermediate values in their L1 and L2 that did not differ significantly from one another. It should be noted that some studies have found differences between perception and production. For example, Williams (1977) reported that Spanish-English bilinguals perceived the English and the Spanish /p/-/b/ contrast similarly but made a difference in production. Moreover, Caramazza et al.'s (1973) early English-French bilinguals categorized L1 and L2 stops similarly but produced English /p/ with greater VOT values. Still, they produced /b/ with French-like values in both languages. The fact that opposing results have been found for perception and production in some previous studies will be further discussed in section 2.4.

It has become evident that previous research has reported different findings when it comes to the nature of crosslinguistic influence in L2 users and how they categorize L1 and L2 sounds – i.e., using separate or shared categories. However, it should be noted that the studies reviewed above differ in methodological terms as well as in the type of subjects that were tested. Regarding methodological differences, some studies have used synthesized stimuli in order to test perception (e.g., Caramazza et al., 1973; Williams, 1977), whereas others have implemented modified natural stimuli (e.g., Casillas & Simonet, 2018; Hazan & Boulakia, 1993) or natural stimuli (e.g., Bohn & Flege, 1993).

Moreover, language mode has not been regulated in the same manner across studies. Some studies have controlled it providing instructions in the target language (e.g., Caramazza et al., 1973; Williams, 1977). More recently, Gonzales and Lotto (2013) and Casillas and Simonet (2018) have gone a step forward when it comes to controlling language mode, as the testing tokens were embedded in a word that sounded in a more L1 or L2 manner according to the language that was being tested. More specifically, both languages were tested using the words 'bafri' and 'pafri' and included the same /pa/-/ba/ VOT continuum. The English tokens were appended the syllable [fɹi] – i.e., with an English pronunciation –, whereas the Spanish ones were appended [fɾi] – i.e., with a Spanish pronunciation. It should be noted, though, that Casillas and Simonet (2019) did not counterbalance the order of testing of the languages, as Spanish was presented first to all participants, which may have resulted in an 'acoustic recalibration' in the second test. Bohn and Flege (1993) also revealed the importance of 'language set' – in this case, meaning different VOT ranges – since they found that Spanish-English bilinguals, as well as Spanish and English monolinguals, presented the same language effect in their /t/-/d/ categorization. All groups labelled short-lag /t/ as /t/ more often when it was presented along with short-lag /d/ than when it was presented with long-lag /t/.

As for differences between the populations under study, some studies have investigated early bilinguals (e.g., Caramazza et al., 1973; Hazan & Boulakia, 1993), others have focused on adult L2 learners (e.g., Flege, 1987; Major, 1992) and some have compared both (e.g., Flege, 1990; Fowler et al., 2008). The factor assessed in some studies was L2 experience (e.g., Flege, 1987), whereas in others it was language dominance (e.g., Carlson et al., 2016; Hazan & Boulakia, 1993). Moreover, there is a difference in the language setting where participants were tested across the works reviewed. Due to the great variability in the methodology and the type of subjects that were investigated, it has been difficult to find clear trends in the results. Still, some generalizations can be attempted. It appears that only early bilinguals have been found to be able to separate L1 and L2 categories and produce both categories with native-like values (e.g., Antoniou et al., 2010; Flege, 1990). It also seems that L2 speakers are better able to separate L1 and L2 categories in production than they are in perception (e.g., Caramazza et al., 1973; Williams, 1977). Still, those studies that have used modified natural stimuli and have controlled the language set using distinctively target-like sounding stimuli have found small yet significant differences in the perception of the two languages tested (e.g., Casillas & Simonet, 2018; Gonzales & Lotto, 2013). The following section will discuss the main individual factors that have been reported to affect crosslinguistic influence in L2 speakers.

2.3. Individual factors influencing crosslinguistic interaction

This section will examine the most relevant individual factors reported in the literature, including age of learning of the target language (AOL), age of arrival to the L2 setting (AOA), language dominance, language use, amount of L2 instruction, gender, motivation and inhibitory skill, most of which are reviewed in Piske et al. (2001). A subsection is devoted to L2 experience – understood as amount of time spent in an L2 setting –, which is the main factor investigated in the present dissertation.

A vast number of works have studied the effect of AOL in L2 performance (e.g., Fathman, 1975; Flege, 1991a; Flege & Fletcher, 1992; Flege et al., 1999; Flege, Munro & MacKay, 1995; Mackay, Flege, Piske & Schirru, 2001; Miyawaki, Strange, Verbrugge, Liberman & Fujimora, 1975; Moyer, 1999; Oyama, 1976; Patkowski, 1990; Piper & Cansin, 1988; Seliger, 1975; Suter, 1976; Tahta, Wood & Loewenthal, 1981; Thompson, 1991) and, as a matter of fact, Piske et al. (2001) found it to be the most determinant predictor of L2 accentedness. Flege (1991a) found significant differences between

Spanish children and adult learners of English regarding /t/ production. Late learners used intermediate values in English, whereas early learners used native-like values in the production of /t/. Moreover, Miyawaki et al. (1975) determined that an effective discrimination between /l/ and /r/ by Japanese L1 speakers requires exposure to the L2 phone from an early age, given that only those experienced learners of English who were exposed to /r/ before puberty were able to effectively discriminate it from /l/.

In the same vein, AOA – which in some cases occurs at the same time as AOL – has been found to have an effect on L2 perception and production accuracy (e.g., Flege, Mackay & Meador, 1999; Flege et al., 1995). For example, Flege et al. (1995) studied the L2 production of late Italian learners of English differing in AOA to Canada. They found that, after a certain AOA, late learners were not able to produce word-final /d/ in a native-like manner. Similarly, Flege et al. (1999) found that the later Italians arrived in Canada, the less accurately they produced and perceived English vowels.

Moreover, language dominance has been reported to have an effect on accuracy of L2 production and perception (Elman et al., 1977; Flege et al., 2002; Hazan & Boulakia, 1993). Hazan and Boulakia (1993) studied stop perception and production of French-English speakers varying in language dominance. They found that French-dominant bilinguals used shorter VOT values for /p/ and longer prevoicing for /b/ in both languages than English-dominant speakers. As for perception, French-dominant bilinguals seemed to rely mostly on duration cues – i.e., VOT – whereas English-dominant bilinguals showed greater variability in cue-weighting, given that some of them relied on spectral cues. Similarly, Flege et al. (2002), found a correlation between language dominance and L2 accentedness ratings.

When it comes to language use, conflicting results have been reported in the literature. On the one hand, some studies have found that amount of L2 use can be a predictor of L2 degree of accentedness, although often not on its own. Early studies by Purcell and Suter (e.g., 1980) found that a compound variable which combined years of residence in the L2 context and years living with native speakers of English was tightly related to L2 accuracy in production. Moreover, Tahta et al. (1981) reported that language use at home accounted for 26% of variance in degree of L2 accentedness in early bilinguals, but only for 9% when considering late and early learners together. Furthermore, a recent study by Casillas (2019) showed that L1-English late learners of English improved their perception and production of Spanish stops after a 7-week immersion session in which L1 use was prohibited. However, other studies have failed to find a relationship between amount of L2 use and accuracy in L2 production (e.g., Elliott, 1995; Flege & Fletcher, 1992; Mackay et al., 2001). Regarding perception, Mora and Nadeu (2012) found a negative effect of L2 use on the L1, as it was found to be negatively correlated to accurate perception of the L1 /e/–/ɛ/ contrast in Catalan-Spanish bilinguals. As for the L1, some previous studies have found that a reduced use may result in a more target-like L2 pronunciation (e.g., Flege, Murray & Munro, 1995; Flege et al., 1999; Guion, Flege & Loftin, 2000; Piske et al., 2001). For instance, Flege et al. (1999) reported that Korean learners of English that did not use their L1 often were less L1-accented than those who did. Moreover, Piske et al. (2001) found that L1 use was, along with AOL, the only clear predictor of L2 accentedness.

Formal instruction has also been found to have an effect on L2 pronunciation. For example, Flege and Fletcher (1992) found that number of years of L2 instruction was a good predictor of L2 accentedness. Nevertheless, it was not established as the most relevant variable, since it only accounted for 5% of the variance in the accentedness ratings. In fact, some studies have failed to find an effect of L2 instruction on pronunciation accuracy (e.g. Elliott, 1995; Flege, et al., 1999; Thomson, 1991). Elliot

(1995) and Piske et al. (2001) attribute the little influence of L2 instruction on L2 pronunciation to the fact that it has traditionally received scant attention in the L2 classrooms.

The effect of gender on L2 speech acquisition has also been investigated in previous studies (e.g., Elliott, 1995; Flege & Fletcher, 1992; Flege, et al., 1995). It has generally been found that gender is not a significant factor in L2 speech acquisition by itself, as gender effects on degree of L2 accentedness differed according to AOL (e.g., Piske et al., 2001; Flege & Fletcher, 1992; Elliott, 1995).

Similarly, motivation has not been found to be a determining effect in degree of L2 accentedness, as not all previous studies have found that it has a significant effect. On the one hand, Moyer (1999) found a correlation between motivation and pronunciation ratings. Some other studies (e.g, Flege et al., 1995; Flege et al., 1999) found that motivation had an effect on L2 accentedness, but that it was small compared to other individual factors. On the other hand, Thompson (1991) failed to find an effect of motivation on L2 pronunciation accuracy.

Recently, the effect of inhibitory skill on L2 speech has been investigated (e.g., Darcy, Mora & Daidone, 2014; Lev-Ari & Peperkamp, 2013). Darcy et al. (2014) reported that inhibitory skill may be positively related to a more accurate L2 perception and production, as the processing of the most relevant phonetic information would be enhanced by the listener. As for the L1, Lev-Ari and Peperkamp (2013) found that bilinguals with lower inhibitory skill presented greater influence of the L2 on the L1 than those with a greater inhibitory skill. As it has already been mentioned, L2 experience is the main factor investigated in this dissertation. For this reason, the following subsection will review previous findings regarding the effect of L2 experience on both the L2 and the L1.

2.3.1. The effect of language experience and language setting on crosslinguistic interaction

Second language experience – understood as amount of time spent in an L2 setting and which may also be referred to as length of residence – has been widely found to be one of the main factors causing L1-L2 influence in bilingual speakers (e.g., Bohn & Flege, 1990; Flege, 1987; Levy & Law, 2010; Lev-Ari & Peperkamp, 2013; Miyawaki et al., 1975). As a matter of fact, PAM-L2 (Best & Tyler, 2007), which extended PAM's premises to L2 learners, stressed the effect that immersion contexts have on second language acquisition as opposed to a classroom setting. The main difference between the two settings lies upon the fact that formal instruction provides restricted input and 'little or unsystematic conversational experience with native speakers' (Best & Tyler, 2007: 11), whereas L2 immersion can provide extensive native input. Therefore, classroom based L2 learners should not be equated to bilinguals learning an L2 in an immersion context. Still, both types of L2 users differ from monolinguals in that both have some kind of exposure to the target language. Hence, the need to compare all three kinds of speakers – namely L2 experienced learners, L2 inexperienced leaners and monolinguals – so as to effectively assess the effect of second language experience on both the target language and the existent language.

In order to make such comparisons, we first need to determine what it means to be an experienced L2-learner. Previous studies have found that a short L2 experience of 6 months in an immersion setting may already result in an L2 perceptual benefit over inexperienced learners (e.g., Aoyama, Flege, Guion, Akahane-Yamada & Yamada, 2004; MacKain, Best & Strange, 1981). For this reason, the PAM-L2 framework suggests that the experienced-inexperienced learner cut-off should be placed at 6-12 months (Best & Tyler, 2007).

Another important factor that we need to consider is when the L2 immersion took place – i.e., whether it is present or past at the time of testing and the recency of the experience. In this regard, Sancier and Fowler (1997) found that recency of language experience influenced both L1 and L2 production accuracy. Their case study involved a Brazilian Portuguese speaker of English that showed a bidirectional influence on VOT production. The subject was found to produce stops more accurately as a function of recent language experience. After spending a few months in the US, the informant produced English stops in a more native-like manner, whereas Portuguese stops were produced with less native-like VOT values. The opposite was true after a stay in Brazil. The current thesis will compare L2 learners living in a target-language setting with L2 learners who are living in an L1 country after a stay in an L2 setting. Next, the results of previous research on the effect of L2 experience on the L2 and on the L1 will be discussed.

2.3.1.1. The effect of L2 experience on the L2

As previous research points out, L2 experience has an effect on the acquisition of L2 categories, both in perception and production (e.g., Bohn & Flege, 1990; Flege, 1984, 1987; Flege et al., 1997; Jun & Cowie, 1994; Lev-Ari & Peperkamp, 2013; Levy, 2009; Levy & Law, 2010; Munro, 1993). Bohn and Flege (1990) assessed the perception of /i/-/ɪ/ – two similar phones – and /ɛ/-/æ/ – a similar and a new phone – by German learners of English differing in L2 experience. Subjects were divided into two groups, namely a group with a short L2 experience – less than a year –, a more experienced group – with participants that had spent a mean of 7.5 years in an English-speaking country – and a monolingual control group of each language. The more experienced German learners of English, but not the less experienced group, were found to perceive /ɛ/-/æ/ in a target-like fashion. The experienced group and the English monolinguals categorized the contrast based on spectral cues, whereas the inexperienced group relied mostly on duration. However, no significant differences between the two German groups were found regarding the perception of /i/-/ɪ/. In this case, both German groups relied on duration, contrary to English speakers, who relied more on quality.

Similarly, Flege et al. (1997) found that German, Spanish, Mandarin and Korean experienced L2 speakers of English – who had been living in the United Stated between 5 and 9 years at the time of testing – perceived the /ɛ/-/æ/ and /i/-/ɪ/ contrasts in a more native-like manner – that is, relying on spectral cues rather than on duration – than inexperienced speakers – who had spent less than a year in the United States. A study previous to this Ph.D. dissertation (Gorba, 2016, 2018) found that Spanish learners of English with a moderate L2 experience – a mean of 12 months – outperformed Spanish learners of English with no L2 experience abroad in the perception of English /p/-/b/. In other words, the English /p/-/b/ category boundary obtained by the experienced learners was more similar to that of English native speakers than that of inexperienced learners. Nevertheless, Cebrian (2006) did not find a significant effect of L2 experience on the L2 perception of /i, ɪ, eɪ, ɛ/. L1-Catalan L2-English speakers who had been living in Canada for a mean of 24 years did not outperform Catalans in Spain in their ability to identify L2 vowels. Moreover, it was observed that both the experienced and the inexperienced Catalan speakers of English relied mostly on duration cues – rather than on spectral differences, as L1-English speakers did – in order to distinguish between tense and lax English vowels.

L2 experience has also been found to improve the production of L2 sounds. Previous works have reported that the longer bilinguals have lived in an L2 context, the more native-like their VOT productions are (e.g., Lev-Ari & Peperkamp, 2013; Levy & Law, 2010). Similarly, Flege et al. (1997) found that the experienced learners of English with diverse linguistic backgrounds described above produced vowel /ɪ/ more accurately than inexperienced learners with the same L1 background. However, no group differences were found in the production of vowel /i/, as /ɪ/ is a new phone in all the languages tested – except for German – and /i/ is a similar phone in all the L1s that were studied. This finding is in line with SLM's predictions, which posit that similar phones will be more difficult to acquire than new phones, as they will be assimilated to the closest L1 phone and will rarely be produced authentically – even with L2 experience.

Moreover, Flege (1987) assessed the effect of experience on VOT production of /t/, /y/ and /u/ by English-French bilinguals. For this purpose, English-French/French English speakers differing in L1 and in L2 experience were tested. Several groups were included: English monolinguals; English native speakers with little L2 experience in French – 9 months in an immersion setting – living in the US at the time of testing; moderately experienced learners of French – 1.3 years in an immersion setting – living in the US who had received formal education in their L2; English native speakers who had been living in France for 11.7 years with less formal instruction than the previous group described; French native speakers who had been living in the US for 12.2 years and French monolinguals speakers living in France who had had little exposure to English. Flege's findings indicate that experience has an impact on the production of nonnative similar phones, given that experienced L2 groups produced L2 /t/ and /u/ with values that approximated the values of French controls more than the less experienced group did. As a matter of fact, the two groups of French learners in the US produced significantly longer VOT values for /t/ than French monolinguals, whereas the Americans in France – i.e., the most experienced group – did not. It should be noted, though, that, numerically, all English groups presented greater VOT values than French controls. As for the L1-French L2-English group living in the US, they produced stops in both languages with intermediate VOT values and presented a shared merged category. Regarding French /y/ – classified as a new phone by Flege – only the least experienced English group differed significantly from French monolinguals, whereas the rest of the L2-experienced English groups performed in a target-like manner. This finding is again consistent with Flege's hypothesis that similar phones – like /t/ and /u/ – are more difficult to acquire than new phones – like /y/ –, as the former will probably be assimilated to the closest L1 phones and will rarely be produced authentically.

2.3.1.2. The effect of L2 experience on the L1

So far, the effect of L2 experience on L2 perception and production reported in previous research has been discussed. Although little research has been conducted on the effect of L2 experience on the L1, this section will review previous studies on the matter (Cabrelli, Luque & Finestrat-Martínez, 2019; Cebrian, 2006; Dmitrieva, 2019; Flege, 1987; Gorba, 2016; Major, 2010). Living in an immersion context may deprive the L2 user from sufficient L1 input and bring about L1 phonological attrition. This was the case in Major's (2010) study, which assessed the ability to detect foreign speech by Brazilian Portuguese speakers living in Brazil as opposed to Brazilian Portuguese speakers living in the United States. Although no significant differences were found regarding the identification of non-native speech – as opposed to native speech –, there was a difference in accentedness ratings. Brazilians in the US rated nonnative Portuguese speakers with significantly lower accentedness scores than those living in Brazil. That is, L2 immersion caused a certain degree of attrition in the ability to assess foreign accentedness, but not to the extent to be unable to discern between native and nonnative speech. Similarly, Cebrian (2006) found that L1-Catalans living in Catalonia outperformed L1-Catalans living in Toronto when it comes to the identification of the L1 vowels /i, e, ei, ɛ/. Moreover, the study prior this dissertation (Gorba, 2016, 2018) reported above, also found an effect of L2 experience on L1 /p/-/b/ perception, as experienced leaners appeared to be moving their Spanish category boundary towards English-like values.

As for production, Flege (1987) also investigated the effect of experience on the VOT production of /t/ in the L1. It was found that the groups that had the greatest L2 experience – both among the L1-English and L1-French speakers – presented L2 influence on the L1 – i.e., produced L1 stops differently from monolingual L1 speakers. By contrast, Gorba (2016) failed to find any significant effect of experience on L1 production. Both experienced and inexperienced groups produced Spanish stops like the control group. As for perception, although descriptively experienced learners had shifted their boundary towards a more English-like identification, their perception did not differ significantly from that of Spanish monolingual speakers. It should be noted, though, that Gorba (2016) tested moderately experienced speakers – who had spent only 12 months in an English-speaking country – in an L1 setting, that is, immersed in the L1.

All in all, it appears that most studies focusing on L2 experience have found that this factor has a significant effect on L2 perception (e.g., Bohn & Flege, 1990; Flege et al., 1997) and production (e.g., Flege, 1987; Flege et al., 1997), as L2-experienced learners tend to use more target-like categories than inexperienced learners. Cebrian (2006), however, failed to find a positive effect of experience on the identification of L2 English /i, ɪ eᶦ, ɛ/, since Catalan speakers in Canada did not outperform those living in Catalonia. Moreover, no differences between the two groups were observed in the perceived similarity of the L2 vowels between Catalan and English. Similarly, even though Flege and Bohn (1990) found a positive effect of L2 experience on the categorization of the /ɛ/-/æ/ contrast, they failed to do so in the case of /i/-/ɪ/. In a similar vein, Flege (1987) found that, in spite of the fact that the most experienced L1-English

L2-French speakers approximated native English speakers in their productions of the similar phones /t/ and /u/, they appeared to fail at producing them in a native-like manner. Conversely, /y/ – classified as a new phone – was produced with similar values to French monolinguals by all the groups of English speakers tested except the least experienced one. The fact that experienced learners outperformed inexperienced learners in some occasions but performed similarly in others may have to do with the type of phones investigated. As Flege (1987) points out, the effect of L2 experience may vary according to the type of phones – similar vs. new – under study, being the former more difficult to acquire than the latter.

When it comes to the effect of L2 experience on the L1, some research has found that experienced learners present L2 influence on L1 perception (e.g., Major, 2010; Cebrian, 2006) and production (e.g., Flege, 1987). Out of the studies reviewed above that investigate the effect of L2 experience on L1 production, only Gorba (2016) failed to find a significant difference in L1 production between experienced and inexperienced Spanish learners of English. However, the experienced group in this study had only lived in the target language setting for a mean of 12 months, as opposed to the other studies cited, which presented considerably longer lengths of residence in an immersion setting. Moreover, the experienced learners were tested in Spain, as their L2 immersion period, which took place the previous academic year, had finished several months before they were tested. Considering Sancier and Fowler's (1997) findings, which revealed an effect of recency of L2 experience on L1 production, the fact that they had been living in an L1 setting for several months before the testing time may have resulted in a re-attunement of L1 perception to native-like values.

This section has reviewed the most important individual factors influencing crosslinguistic interaction that have been reported in the literature. All except inhibitory skill have been considered directly or indirectly in the design and the analysis of the experiments presented in the present thesis, as it will be shown in Chapter 4 ('Methodology'). In the first place, gender was counterbalanced across all groups. As it has been mentioned before, L2 experience – i.e., amount of time spent in the L2 setting – is the main factor under analysis, as L1-Spanish L2-English and L1-English L2-Spanish varying in amount – i.e., length of residence – and type – i.e., past or present – L2 experience were tested on their L1 and L2 production of stops. As for AOA, all participants – who were young adults – had moved to the L2 country after adolescence. Regarding motivation, great differences between groups are not expected, given that all participants were either studying a degree in their L2 or were immersed in an L2 setting. Thus, all should be highly motivated. Participants had to complete a linguistic background questionnaire, as well as the Bilingual Language Profile (BLP, Birdsong, Gertken, & Amengual, 2012). Information about their AOL, amount of L2 use, amount of L2 instruction and language dominance was collected. It was later investigated whether these factors also had an effect on L1 and L2 stop perception and production (see Chapter 8 on 'The effect of individual factors on stop perception and production'). The following section will discuss the theories and findings related to the relationship between perception and production.

2.4. The relationship between perception and production

2.4.1. The relationship between perception and production in the L1

The relationship between perception and production has been widely discussed in the literature. Some speech theories assume that there is an alignment between the perception and the production of speech sounds. For example, the Motor Theory

(Liberman, Cooper & Harris, 1962; Liberman, Cooper, Shankweiler & Stuudert-Kennedy, 1967; Liberman & Mattingly, 1985) and the Direct-Realist approach (Fowler, 1986, 1990) posit that adults perceive speech and create phonetic categories based on articulatory gestures of their own speech. However, conflicting findings have been reported when it comes to the nature of this relationship. Whereas some studies have supported the existence of a link between perception and production in the L1 of adult speakers (e.g., Brunner, 2011; Cooper & Lauritsen, 1974; Fox, 1982; Newman, 2003; Perkell et al., 2004), others have failed to establish a clear relationship between the two dimensions (e.g., Ainsworth & Paliwal, 1984; Bailey & Haggard, 1973; Shultz et al., 2012).

Perkell et al. (2004) analyzed the contrast distance between /ɑ/ and /ʌ/ and between /u/ and /ʊ/ in the perception and production of young adult speakers of American English. They found that that the more accurately a speaker discriminated the contrasts, the more distinctly the speaker produced them. Similarly, Fox (1982) reported that individual variation in the perception of English vowels – analyzed using a scaling task in which participants rated the similarity/dissimilarity of two tokens – was related to each participant's acoustic vowel space – i.e., tongue height and advancement. Finally, in a study involving stops, Newman (2003) observed that speakers who produced the sequence /pa/ with the longest VOT values also showed a preference towards longer VOT values in a goodness rating task. Newman concluded that individual differences in perception were related to individual differences in production.

On the other hand, other studies report no relationship between the perception and production of VOT. Bailey and Haggard (1973) failed to find any significant correlations between the perception and the production of VOT and of f0 in the voiced-voiceless stop contrast. Perception was tested with a goodness rating task (appropriate to exaggerated) and the measures used in production were VOT means, the difference between voiced and voiceless production and f0 values. More recently, Shultz et al.'s (2012) study on VOT and f0 weighting in perception and use in production replicated the same results – i.e., no significant correlations were found between perception and production values. In short, the relationship between perception and production is unclear given the opposing results of previous studies. The following section will review the findings related to the relationship between the two dimensions in the L2.

2.4.2. The relationship between perception and production in the L2

Some L2 speech theories – like the SLM and the PAM-L2 – claim that there is a link between the two dimensions and that L2 perception precedes L2 production (Flege, 1995; Best & Tyler, 2007). Flege's (1995) SLM posits that difficulties in the production of L2 sounds stem from the influence of L1 categories on L2 perception. Thus, perception should precede production in L2 acquisition. Best and Tyler (2007) agree with Flege's hypothesis and assume that L2 production may continue to adjust to L2 target-like phonetic and phonological features over a longer period of time than perception (Best & Tyler, 2007). However, Flege (1999) also predicts that perception and production in the L2 may not be perfectly aligned, but clearer correlations may be found as participants gain experience in the L2. Similarly, Bohn and Flege (1997) claim that the two dimensions do not progress in parallel and that, as a matter of fact, production may as well lead perception in early stages of L2 speech acquisition.

Some previous studies have found a relationship between perception and production in the L2 (e.g., Baker & Trofimovich, 2006; Flege et al., 1997; Flege, et al., 1999; Jia, Strange, Wu, Collado & Guan, 2006; Levy & Law, 2010). For example, Levy and Law (2010) investigated the link between perception and production accuracy of the

French vowels /i, y, u, œ, o, a/ by American English learners of French. Perception was tested with an identification test and production using a repetition task. Production results were analyzed by means of both acoustic analyses and native speaker judgements. A modest correlation between the perception and production measures was found. The contrast /y/-/œ/ presented the most straightforward evidence for the relationship between the two modalities, as those participants who categorized them separately more often also obtained better intelligibility judgements by the L1 speakers. A somewhat weak relationship was reported by Rallo Fabra and Romero (2012), who studied the relationship between accurate discrimination and production of L2 English vowels by Catalan speakers differing in L2 proficiency. English vowels were paired as follows: /ɑ-ʌ/, /ɛ-æ/, /i-ɪ/ and /u-ʊ/. The measure used in perception was overall A' discrimination score for each English pair – which takes into account the proportion of hits and false alarms. Vowel production was evaluated by English native speakers and was quantified using the mean bidirectional error, which was calculated by adding up the confusion rates of the two vowels in each pair – i.e., the percentage of times that a vowel was perceived as its counterpart. A general analysis including all groups did not reveal a significant correlation between the two variables. Still, a within group analysis showed a moderate, yet significant, correlation in the case of the learners with intermediate proficiency, whereas that was not the case regarding low and high proficiency learners. This difference between groups and the fact that the only correlation found was weak indicates the complexity of the relationship between the two dimensions.

Some previous studies have obtained evidence that supports the SLM's assumption that perception leads production in L2 acquisition (Borden, Gerber & Milsark, 1983; Casillas, 2019; Flege et al., 1999; Levy & Law, 2010). For instance, Flege et al. (1999) studied the perception-production relationship in L2 vowel perception by

L1-Italian L2-English speakers. Perception was tested using a discrimination task and production with an intelligibility test presented to native speakers. Results revealed that accuracy of L2 production was determined by accurate L2 perception. Moreover, Levy and Law's (2010) study reported above also found evidence that perception may precede production, as inexperienced American learners of French were more accurate at perceiving the /y/-/u/ contrast than at producing it. Only a few studies involving consonants provide support for the assumption that L2 perception leads L2 production. For example, Borden et al.'s (1983) study on liquids found that adult Korean learners of English were more accurate at perceiving the English /r/-/l/ contrast in a phonemic identification task than at producing it. Regarding stops, Casillas (2019) found that late L1-English learners of Spanish participating in a seven-week immersion program were found to change their /p/-/b/ perceptual boundary towards Spanish-like values earlier than their production boundary, which was calculated by standardizing the VOT values of the two phones and averaging them together.

Conversely, some previous studies involving stops have found that production may precede perception (e.g., Bohn & Flege, 1990, 1997; Caramazza et al., 1973; Flege & Eefting, 1987; Trofimovich & John, 2011; Zampini, 1998). For example, Caramazza et al. (1973) reported that Canadian L1-French L2-English early bilinguals were able to produce L1 and L2 voiceless stops differently but identified the stop contrasts according to L1 values. Similar results were reported by Flege and Eefting (1987), who found that Dutch learners of English could produce the /t/-/d/ contrast similarly to English native speakers but were unable to discriminate it correctly. In the same vein, Trofimovich and John (2011) found that L1-Canadian French learners of English were able to produce the /θ/-/t/ and /ð/-/d/ contrasts accurately, but showed difficulty in distinguishing them in perception. Regarding vowels, Bohn and Flege (1990, 1997) found that experienced L1-

German learners of English produced the /ɛ-æ/ contrast with English-like values, but differed from English monolinguals in perception, as they did not rely on spectral cues to the same extent. Inexperienced learners were not able to perceive nor to produce the English contrast accurately. The fact that experienced learners outperformed inexperienced learners in production to a greater extent than in perception indicates that L2 production developed earlier than L2 perception.

Yet, a fair amount of research has actually failed to find a clear relationship between L2 perception and L2 production at an individual level, especially regarding consonants (e.g., Baese-Berk, 2019; Darcy & Krüger, 2012; Gorba, 2016; Lev-Ari & Peperkamp, 2013; Peperkamp & Bouchon, 2011; Sheldon & Strange, 1982; Williams, 1977; Zampini, 1998), although some studies on vowels also failed to find a link between the two dimensions (e.g., Darcy & Krüger, 2012; Peperkamp & Bouchon, 2011). For instance, Sheldon and Strange (1982) analyzed the identification and production of English /l/-/r/ by adult Japanese speakers living in the US. Results suggested that some learners were more successful at producing the contrast than at perceiving it, whereas the opposite was true for other learners. As for stops, Williams (1977) reported that Spanish-English bilinguals performed like native speakers in production – tested with both a word list and a sentence reading task – but perceived bilabial stops using a shared category in both languages, as the results of an identification and a discrimination test in each language revealed. However, perceptual categories were more Spanish-like for some speakers and more English-like for others. In the same vein, a study prior to this thesis (Gorba, 2016) did not find a clear relationship between the perception – tested with an identification test – and production – assessed with a carrier sentence reading task – of English /p/-/b/ by Spanish learners of English. A link between the two dimensions could not be established neither for inexperienced nor for experienced learners, as those participants who presented a perceptual category boundary with greater VOT values did not always use more aspiration in production.

All in all, it seems that the nature of the relationship between perception and production in the L1 and, especially, in the L2 is not fully understood, as contradictory results have been reported in the literature. Some studies have found a relationship between the two dimensions – i.e., accuracy in perception is correlated to accuracy in production (e.g., Flege et al., 1997; Flege et al. 1999; Levy & Law, 2010). Still, it should be noted that in many cases the studies that have found a relationship between the two dimensions did not find significant results in all the phones analyzed and, when they did, moderate correlations were often reported (e.g., Levy & Law, 2010; Rallo Fabra & Romero, 2012). Another matter that has been investigated in previous studies is whether perception precedes production or whether production precedes perception. Some findings suggest that accurate L2 perception is a requirement for accurate L2 production (e.g., Borden et al., 1983; Flege et al., 1999), whereas others indicate that L2 learners may be able to produce an L2 sound in a native-like manner but differ from native speakers of the L2 in perception (e.g., Bohn & Flege, 1997; Caramazza et al., 1973; Flege & Eefting, 1987). Nevertheless, some previous studies have not found a clear relationship between the two dimensions at an individual level, since accuracy in perception was not correlated to accuracy in production – i.e., participants who were target-like in production were not accurate in perception and the other way around (Gorba, 2016; Lev-Ari & Peperkamp, 2013; Sheldon & Strange, 1982; Williams, 1977).

This lack of agreement in the literature could stem from various reasons. As Bohn and Flege (1997) point out, different criteria have been used to evaluate participants' performance in perception and production. For example, some studies analyzed perception using an identification test (e.g., Caramazza et al., 1973; Levy & Law, 2010), whereas others implemented a discrimination test (e.g., Flege & Eefting, 1987; Flege et al., 1999; Rallo Fabra & Romero, 2012) or both (e.g., Williams, 1977). Regarding production, some have implemented a repetition task (e.g., Levy & Law 2010), whereas others have used reading tasks (e.g., Bohn & Flege, 1997; Williams, 1977). Moreover, some studies used acoustic measures to analyze production and relate it to the perception measures (Caramazza et al., 1973; Levy & Law, 2010; Williams, 1977), whereas others used goodness rating tasks (e.g., Flege et al, 1999). Llisterri (1995) suggested that the type of sound – i.e., consonants as opposed to vowels – may also have an effect on the relationship between the two dimensions. As a matter of fact, most of the studies reviewed above that found a relationship between perception and production and, especially, that perception leads production investigated vowels (e.g., Flege et al., 1999; Bohn & Flege, 1990; Borden et al., 1983). Conversely, most studies involving consonants either found that production may lead perception (e.g., Caramazza et al., 1973; Flege & Eefting, 1987) or failed to establish a clear relationship between both dimensions (e.g., Baese-Berk, 2019; Baker & Trofimovich, 2006; Gorba, 2016; Sheldon & Strange, 1987; Williams, 1977; Zampini, 1998). The nature of the sound in relation to the L1 – that is, whether it is a similar or a new phone – may also play a role (Bohn & Flege, 1997). For example, Levy and Law (2010) found that the discrimination of the new French sounds /y/ and /œ/ – two new phones – by English speakers was more accurate than that involving one or two similar sounds /y/ and /u/.

Factors related to the population tested may also have an impact on the relationship between perception and production. Previous studies have also found differences between experienced and inexperienced learners (e.g., Bohn & Flege, 1990, 1997). Bohn and Flege (1990, 1997) reported that inexperienced German learners of English were able to discriminate /ɛ/ and /æ/ but could not make a clear distinction in production, whereas experienced learners were more accurate than the inexperienced in both dimensions. Finally, some studies tested early bilinguals (e.g., Caramazza et al., 1973; Williams, 1977), whereas others tested adult L2 learners (e.g., Bohn & Flege, 1997; Levy & Law, 2010; Sheldon & Strange, 1987). Given the number of variables that vary across the studies described above, it is difficult to find a pattern that accounts for the differences in the reported results. The present thesis aims to study further the relationship between perception and production of bilabial and velar stops in adult L1-Spanish L2-English and L1-English L2-Spanish learners differing in amount of L2 experience and language setting. The following section will describe the main cues to the voicing contrast in stops in Spanish and in English.

2.5. Voicing cues in stops

2.5.1. VOT

The feature studied in the present thesis is voice onset time (VOT), that is, 'the interval between the release of the stop and the onset of phonation' (Abramson & Lisker, 1973). Previous works have established VOT as the main cue for stop voicing distinction in English and Spanish, as well as in other languages (Lisker & Abramson, 1964; Schertz, Cho, Lotto & Warner, 2015 et al., 2015; Shultz et al., 2012; Williams, 1977). Nevertheless, even though both Spanish and English have two categories for each point of articulation – i.e., a voiced and a voiceless category – the two languages make a different use of VOT: whereas voiceless English stops are aspirated – present long-lag VOT – and voiced stops are unaspirated – generally produced with short-lag VOT –, the Spanish contrast is based on the presence of prevoicing – voice-lead VOT – in voiced stops and the use of short-lag VOT in voiceless stops. Still, a small percentage of short-lag production for voiced stops may be found in Spanish, as reported by Flege and Eefting (1987) and for other languages with a similar VOT system to Spanish, like Polish (Keating, 1984). Previous studies have measured the VOT duration of stops in English and in Spanish (e.g., Castañeda, 1986; Docherty, 1992; Lisker & Abramson, 1964; Rosner, López-Bascuas, García-Albea & Fahey, 2000). Given that the present study investigates initial stops, the VOT measurements of initial voiced stops reported in previous research will be reviewed (see Table 2.1).

Regarding the bilabial contrast, Lisker and Abramson (1964) investigated the mean VOT production of stops in several languages using a word list and reported an average value for initial /p/ in Puerto Rican Spanish of 4 ms and a mean VOT for initial /b/ of -138 ms. Castañeda (1986) investigated initial stops in Castilian Spanish using a word list and found that /p/ was produced with a VOT value of 7 ms, whereas /b/ presented a mean of -70 ms. Similarly, Rosner and Fahey (2000) found that Castilian Spanish speakers produced initial /b/ in isolated words with a mean VOT value of -91.5 ms and initial /p/ with a mean of 13.1 ms. Conversely, /p/ in American English was reported to have a mean VOT of 58 ms (Lisker & Abramson, 1964). As for /b/ production, individual variation was found; whereas some speakers' productions of /b/ presented a VOT of 1 ms, others tended to prevoice /b/ and presented a mean VOT of -101 ms. Docherty (1992) investigated initial stops in British English using a reading task and reported a mean VOT of 42 ms for /p/ and 15 ms for /b/. Although instances of prevoiced tokens were found for all voiced stops, they were discarded.

Velar stop production follows a very similar pattern. Lisker and Abramson (1964) found that Spanish speakers produce /k/ with a mean VOT of 29 ms and /g/ with a mean of -108 ms. In the case of Spanish, Castañeda reported VOT values of 26 ms for /k/ and -58 ms in the case of /g/. Moreover, Rosner et al. (2000) found that /k/ was produced with a VOT mean of 29 ms and /g/ presented a mean value of -74 ms. Regarding American English, it was found that speakers produced /k/ with a mean VOT value of 80 ms. When it comes to the production of /g/, an average VOT of 21 ms was reported. However, as with /b/, a subset of speakers produced voice-lead VOT – with a mean of -88 ms (Lisker & Abramson, 1964).

Regarding coronal stops, Lisker and Abramson (1964) found that initial /t/ in Spanish presented a VOT value of 9 ms, whereas /d/ was produced with a VOT of -110 ms. Regarding Castilian Spanish, /t/ was produced with a mean VOT value of 10 ms, whereas /d/ presented -78 ms (Castañeda, 1986). Similarly, Rosner et al. (2000) found that Spanish speakers produced initial /t/ with a mean of 14 ms and /d/ with -92 ms. In the case of American English, /t/ presented long-lag values, with a mean of 70 ms in initial position, whereas, just as in the case of /b/ and /g/, /d/ could be produced either using short-lag VOT – i.e., with a mean of 5 ms – or voice-lead VOT – i.e., -102 ms (Lisker & Abramson, 1964). As for British English, /t/ was reported to be produced with a mean VOT of 64 ms and /d/ with 21 ms (Docherty, 1992). In short, Spanish contrasts short-lag VOT – voiceless stops – with voice-lead VOT – voiced-stops –, whereas English contrasts long-lag VOT – voiceless stops – with short-lag or voice-lead VOT – voiced stops. It should be noted that Spanish /t/ and /d/ are dental stops, whereas English /t/ and /d/ are alveolar stops (Casillas, Díaz & Simonet, 2015). Due to this difference in place of articulation in the two languages under study, this thesis will focus on /p/, /b/, /k/ and /g/.

Language	Study	/p/	/b/	/t/	/d/	/k/	/g/
English	American English (Lisker and Abramson, 1964)	58	1/-101	70	5/-102	80	21/-88
	British English (Docherty, 1992)	42	15	64	21	62	27
Spanish	Puerto Rican Spanish (Lisker, & Abramson, 1964)	4	-138	9	-110	29	-108
	Castilian Spanish (Castañeda, 1986)	7	-70	10	-78	26	-58
	Castilian Spanish (Rosner et al., 2000)	13	-92	14	-92	29	-74

Table 2.1. VOT means in ms for initial stops in English and Spanish reported in some previous studies.

Some previous studies have also tested the perceptual categories of stops in English and in Spanish. As has already been mentioned, Williams (1977) analyzed /p/-/b/ perception in monolingual and bilingual speakers of each language with a VOT continuum that was implemented using both an identification and a discrimination test. The category boundaries for the voicing contrast – i.e., the point at which listeners start identifying stimuli as belonging to another category – for Spanish and English monolinguals were calculated on the basis on their responses along the continuum Spanish monolinguals obtained a /p/-/b/ category boundary of -4 ms, whereas the English monolinguals' boundary presented VOT values of 25 ms. Similarly, Abramson and Lisker (1970), tested Spanish and English speakers using an identification test and reported a category boundary for the bilabial contrast with a VOT value of 14 ms for Spanish and of 25 ms for English. As for the velar contrast, Abramson and Lisker (1970) found that Spanish speakers had a category boundary with a VOT value of 25 ms, whereas the English category boundary was located at 42 ms.

As explained above, English voiced stops have been found to present short-lag values and prevoicing. Although short-lag VOT seems to be the general pattern, some English dialects tend to use voice-lead often. Speakers of Southern American English

(SAE) have been reported to use voice-lead rather than short-lag VOT. Hunnicutt and Morris (2016) found that 78% of the tokens analyzed in their study analyzing speakers of SAE were prevoiced. Herd, Torrence and Cariño's (2016) study suggested that the use of prevoicing may also depend on ethnicity. African Americans used voice-lead 90% of time, whereas only 35% of the voiced stops produced by European Americans presented voice-lead VOT. In a similar way, Scottish English speakers also alternate the use pevoicing with short-lag VOT in voiced stops. Moreover, Scottish English presents shorter VOT values for voiceless stops than Standard British English (Scobbie, 2006). It should be noted that speakers living in areas in England close to the Scottish border have also been reported to use similar VOT values to Scottish English speakers. Nevertheless, these differences with Southern British English seem to be decreasing among the younger generations (Docherty, Watt, Llamas, Hall & Nycz, 2011). Finally, Antoniou et al. (2010) reported that voiced stops in Australian English are often produced with short-lag VOT, but that they may as well be produced with voice-lead VOT by some speakers.

It should be noted that VOT values are also influenced by other factors, such as place of articulation, context, gender and speech rate. As discussed above, long-lag and short-lag bilabial stops present shorter VOT than coronal stops, whereas velar stops present the longest VOT values (Lisker & Abramson, 1964; Poch, 1985). Voiceless stops tend to present greater VOT values when followed by high vowels than when followed by low vowels (Klatt, 1975; Poch, 1984). Moreover, voiceless stops appearing in words in isolation present longer VOT values than those embedded in a sentence (Lisker & Abramson, 1967). Some studies have also reported an effect of gender in VOT production. In English, women have been found to produce longer VOT than males, especially in the case of voiceless stops (Koenig, 2000; Swartz, 1992; Whiteside & Irving, 1997). As for Spanish, Poch (1985) observed differences in the duration of /p/ and /t/ in stressed position between women and men, as the former produced longer VOT values than the latter when the stop was followed by the vowel /i/. Conversely, Rangel, Torres and Mojica (2016) did not find a significant effect of gender on VOT production in Spanish stops. Regarding speech rate, it has been found that VOT values increase as speech rate decreases in English long-lag VOT, whereas shorter prevoicings have been reported in Spanish voice-lead VOT with a faster speech rate. As for short-lag categories, no clear effects of speech rate have been found in either language (Kessinger & Blumstein, 1997; Magloire & Green, 1999; Miller, Green & Reeves, 1986; Theodore, Miller & DeSteno, 2009; Volaitis & Miller, 1992). It should also be noted that a greater variability in terms of VOT duration has been found for voiced stops than for voiceless stops, as the duration of prevoicing in voiced stops can vary from relatively very long to relatively short (Abramson & Lisker, 1964; Castañeda, 1986)

2.5.2. Other cues for voicing

Although VOT has been established as the main cue for voicing in initial stops in Spanish and in English, other acoustic properties have been found to play a role in the voiced-voiceless contrast. Secondary voicing cues for stops include the burst – more specifically burst intensity and duration –, F1 – i.e., onset frequency and transition duration – and f0 – i.e. frequency after the release.

Regarding the burst, voiceless stops have been found to present greater mean intensity as well as greater intensity at the beginning of the burst than voiced stops in English (Chodroff & Wilson, 2014; Parikh & Loizou, 2005). Regarding duration, previous research reported that the burst of Spanish voiceless stops tended to be longer than that of voiced stops (Martínez Celdrán, 1991; Villamizar, 2002). In fact, Spanish voiced stops, especially /b/, have sometimes been found to lack a clear burst (Villamizar, 2002). Machuca (1997) also found an absence of a burst in voiceless stops in Spanish in spontaneous speech.

Furthermore, F1 onset frequency and transition duration have been reported to have an effect on voicing perception (Benkí, 2005; Hazan & Boulakia, 1993; Liberman, Delattre & Cooper, 1958). Higher frequencies at F1 onset influence stop perception towards a voiceless identification both in English and in Spanish (Benkí, 2005). Similarly, shorter F1 transitions indicate voicelessness. Finally, f0 values after the release have been found to play a decisive role in the perception of voice and voiceless stops, especially in cases of ambiguous VOT, as higher f0 values have been associated with a bias towards voiceless stops (Idemaru & Holt, 2011; Whalen, Abramson, Lisker & Mody, 1993). However, it should be noted that f0 has been reported to be a relevant cue in short-lag and long-lag VOT productions, but not in the case of prevoiced stops, both in Spanish and in English (Llanos, Dmitrieva & Francis, 2013).

In short, even though VOT is the main cue for voicing in English and Spanish stops, other cues such as burst intensity and duration, f0 and F1 play a role in the voiced-voiceless distinction. The present study has attempted to neutralize these secondary cues in the perception of voicing in stops in order to create a VOT continuum to test L1/L2 English and L1/L2 Spanish perception (further detail on the creation of the continua is provided in section 4.2.1).

This section has offered a review of the literature on key aspects of L2 acquisition – including the most cited L2 speech models, the possible manners in which the L1 and the L2 may interact and the main factors that influence CLI –, the relationship between perception and production and voicing cues in stops, particularly regarding the two languages under study. The next chapter will state the goal and contribution of the current thesis and will present the research questions addressed, as well as its hypotheses.

3. Goal, research questions and hypotheses

3.1. Goal

The goal of the present thesis is to investigate the effect of L2 experience – a compound variable including both amount of L2 experience (i.e., length of residence in a target language setting), as well as learning setting (i.e., instructional or immersion setting) partially following Flege's (1987) design – on the perception and production of L2 and L1 stops. More specifically, the perception and production of bilabial and velar stops by L1-Spanish L2-English and L1-English L2-Spanish speakers varying in L2 experience will be analyzed (see section 4.1 for a description of participants). This study also aims at determining whether L2 learners use a single shared category or separate categories in the L1 and in the L2. Moreover, the relationship between perception and production of L1 and L2 stops will be explored. The sections below present the research questions and sub-questions, as well as the main hypotheses, of this thesis.

3.2. Research questions

- RQ.1. What is the effect of L2 experience on L2 stop perception and production?
  - SUB-RQ.1.1. Do differences in length of residence in the target language country and learning setting affect the perception of L2 stops?
  - SUB-RQ.1.2. Do differences in length of residence in the target language country and learning setting affect the production of L2 stops?
- RQ.2. What is the effect of L2 experience on L1 stop perception and production?

  - SUB-RQ.1.1. Do differences in learning setting and length of residence in the target language country affect the perception of L1 stops?
  - SUB-RQ.1.2. Do differences in learning setting and length of residence in the target language country affect the production of L1 stops?
- RQ.3. Do bilinguals have separate categories or a shared category in their L1 and L2? Do length of residence and learning setting play a role in category formation?
- RQ.4. What is the relationship between the perception and production of stops?
  - SUB-RQ.4.1. Is there a relationship between the perception and production of stops in the L1?
  - SUB-RQ.4.2. Is there a relationship between the perception and production of stops in the L2?
  - SUB-RQ.4.3. Do length of residence and learning setting influence the relationship between perception and production?

3.3. Hypotheses

RQ.1 addresses the effect of L2 experience on L2 stop perception and production. L2 experience has been found to result in a more accurate perception of L2 phones (e.g., Bohn & Flege, 1990; Flege et al., 1997) and, specifically, of L2 stops (e.g., Gorba, 2016, 2018). Thus, amount of L2 experience is hypothesized to result in a more native-like perception of the L2 (SUB-RQ.1.1). Differences may be found between experienced learners immersed in the target language country and experienced learners living in their home country. Being constantly exposed to the L2 may result in a more native-like perception and the target language is more likely to be activated – in Grosjean's (2001) terms – in an immersion setting than in an instructional setting. Conversely, the limited input received in an instructional setting may result in a greater L1 influence on L2 perception. Regarding production (SUB-RQ.1.2.), a positive effect of L2 experience is also expected, that is, L2 experienced learners are predicted to produce L2 stops more accurately than inexperienced learners (e.g., Flege, 1987; Lev-Ari & Peperkamp, 2013; Levy & Law, 2010). Just as in the case of perception, learners living in an immersion setting are predicted to produce stops in a more native-like manner than those living in the home country.

Regarding the L1, it can be hypothesized that the greater the amount of L2 experience the more likely it is that L1 stop production and perception is affected (RQ.2). Previous studies have found that highly experienced L2 learners may present L2 influence on their L1 production (RQ 2.2); the greater their amount of L2 experience, the less native-like their L1 production will be (e.g., Flege 1987). Setting is also expected to play an important role. Previous research found that L2 learners living in the target language country presented a greater effect of the L2 on L1 production (e.g., Flege, 1987). Moreover, Sancier and Fowler's (1997) study pointed out that the recency of the stay abroad had an impact on L1 production accuracy, given that the L1 presented a greater L2 influence right after a stay in the L2 setting than after having been immersed in the L1 setting for a few months. Therefore, it is hypothesized that experienced participants living in the target language country may show evidence of L1 attrition in production, whereas moderately experienced participants in their home country, as well as inexperienced learners, will present a smaller or no L2 influence on their L1 production.

Little attention has been paid to the effect of L2 experience on L1 perception, as, to my knowledge only a few studies have investigated it (SUB-RQ.2.1; e.g., Cebrian, 2006; Dmitrieva, 2019; Major, 2010). Cebrian found that inexperienced Catalan learners of English outperformed experienced Catalan learners of English in Canada in the identification of L1 front vowels. It should be noted that some studies have found that a short period of exposure to the L2 in an instruction setting may also bring about changes in the perception of L1 stops (Casillas & Simonet, 2018; Tice & Woodly, 2012). Therefore, it is possible that inexperienced learners may also present L2 influence in their L1 perception. Still, we expect L2 influence on the L1 to be smaller in an L1 setting than in an L2 setting (e.g., Cebrian, 2006)

Regarding the creation of a new L2 category (RQ.3), the SLM expects category formation to be blocked initially in the case of similar L1 and L2 phones – such as it is the case of /p/, /b/, /k/ and /g/ in Spanish and in English – because the learner is not able to discern cross-linguistic phonetic differences (Flege, 1992, 2003, 2007). As a result, the L2 phone might be assimilated to the closest L1 phone and the existing phonological category may merge with the L2 and present characteristics of both languages (Flege, 1987). This outcome would support the SLM's merger hypothesis. However, the SLM posits that phonetic learning may take place as L2 learners gain experience. Thus, learners with a greater amount of L2 influence are expected to have a shared category presenting values closer to the L2, whereas participants with less experience, especially those in the L1 setting, are expected to present a category with intermediate or L1-like values. The category may as well present somewhat different numerical values according to the nature of the input. That is, L2 learners may present more L1-like values when tested in the L1 and values closer to the L2 when tested in the L2. According to the SLM, another possible outcome would be that the L2 learner eventually created a separate L2 category and presented two separate categories with target-like values (Flege & Eefting, 1986, 1987). As the category dissimilation hypothesis claims, the L1 category may be modified and enhance its differences with the L2 in order to be effectively distinguished from the new L2 category – e.g., English learners of Spanish may present greater VOT values than

English monolinguals. As stated above, L2 experience is sometimes required for category formation, especially in the case of similar phones – like English and Spanish stops. Therefore, only L2 learners with L2 experience might be able to create a new L2 category. In short, category assimilation is expected to take place in L2 learners, resulting in merged categories. The merged categories will present more L2-like values as participants gain experience and may be implemented somehow differently in the L1 and in the L2. However, the most experienced learners may be able to create a new L2 category, which may also influence the L1 category through the process of dissimilation. Thus, an effect of L2 experience is expected on category formation, being the creation of a new category more likely in the case of experienced learners.

Differences between the L1 and the L2 of L2 learners can also be discussed in L2LP's terms (Escudero, 2000, 2003, 2005). According to the L2LP, the L1 and L2 systems are separate. However, initially, the L2 phonological system is a copy of the L1, resulting in identical L1 and L2 categories. As the L2 phonology of the learner develops, their L2 categories will become more target-like, eventually reaching values that are perfectly distinguishable from the L1's. Thus, according to Escudero's model, experienced learners will present target-like VOT values which will clearly differ from the L1. The L2 phonological system of inexperienced learners may still be developing, which may result in more similar L1 and L2 categories than those of experienced learners. Therefore, the degree of distinctiveness between the L1 and L2 categories is expected to be modulated by L2 experience, being the highly experienced learners the best able to make a clear distinction. It should also be noted that the model expects the L1 to remain native-like. An intermediate implementation of the L1 and L2 categories, that is, a perception – or production, if we extend the L2LP's premises to both modalities – with intermediate values between the two systems of the learner is expected if the two phonological systems are activated by the nature of the input, that is, if the language mode is not set to the desired monolingual mode at the time of testing.

RQ.4 addresses the relationship between the perception and production of L1 and L2 stops. According to the main speech acquisition models, we would expect that there is indeed a straightforward relationship between both dimensions. When it comes to the L1 (SUB-RQ.4.1), the Motor Theory (Liberman et al., 1962; Liberman, et al., 1967; Liberman & Mattingly 1985) and the Direct-Realist approach (Fowler, 1986, 1990) assume that speakers produce and perceive speech on the basis of articulatory gestures and that, thus, there is a very tight link between the two modalities. Therefore, individual differences between adult speakers in perception should be correlated to individual differences in production. However, conflicting results have been reported in the literature (e.g., Newman, 2003; as opposed to Shultz et al., 2012). If the Motor Theory and the Direct-Realist approach are right, we will find a relationship between the two modalities in the L1, at least in the case of monolingual speakers. However, results will replicate Shultz et al.'s (2012) findings if no link between the perception and production of L1 stops is found. Despite the discrepancies in previous studies, it is hypothesized that there will be a clear link between perception and production of VOT in the L1. That is, participants that present later category boundaries will also produce greater VOT values. L2 experience may also modulate this relationship in L2 speakers (SUB-RQ.4.3). If L2 experience has a greater influence on one of the two modalities, experienced learners may not present a clear relationship between L1 VOT values in perception and production. Conversely, inexperienced learners may present a clearer relationship between the two dimensions in the L1, as their L1 will have been affected by the L2 – and therefore moved in the direction of the target language – to a lesser extent than in the case of experienced learners.

As for the L2 (SUB-RQ.4.2), L2 speech acquisition models such as the SLM (Flege, 1995) and PAM-L2 (Best & Tyler, 2007) posit that there is a link between perception and production and assume that perception leads production in L2 acquisition. As a matter of fact, a number of previous studies support this prediction (e.g., Casillas, 2019; Levy & Law, 2010). Nevertheless, other studies have found that L2 speakers may be able to produce L2 stops accurately but perceive them using their L1 categories (e.g., Bohn & Flege, 1997; Flege & Eefting, 1987) or have failed to establish a clear relationship between the two dimensions (e.g., Gorba, 2016; Williams, 1977). If the theoretical assumptions by the SLM and the PAM-L2 are right, we will find that L2 learners are more target-like in perception than in production. However, L2 experience is expected to modulate this relationship (SUB-RQ.4.3). That is, we expect that perception and production in the L2 will align as participants gain experience, possibly reaching a native-like alignment between the two modalities in the case of the highly experienced groups. It should also be noted that some previous studies have in fact found that L2 production accuracy may precede L2 perception accuracy (e.g., Bohn & Flege, 1990; Caramazza, et al. 1973; Flege & Eefting, 1987; Zampini, 1998).

However, the results of this thesis will replicate those obtained by some other studies if they fail to establish a straightforward relationship between the two dimensions – i.e., that some participants present a more target-like perception and others are more accurate in production (e.g., Lev-Ari & Peperkamp, 2013; Sheldon & Strange, 1982; Williams, 1977). In spite of the fact that there is not an agreement in the literature, it is hypothesized that there is a link between the two dimensions and, following Flege (1995) and Best and Tyler's (2007) assumption, that accuracy in perception precedes accuracy in production. At an individual level, just as in the case of the L1, it is expected that a participant that presents a later category boundary in the L2 will also use greater VOT values in production.

So far, the main research questions of this thesis and its main hypotheses have been presented. It should also be noted that the nature of this study, which investigates L1-Spanish learners of English as well as L1-English learners of Spanish, also allows to make comparisons between the two mirror-image populations and to test the effect of directionality of language learning. Similarly, the acquisition of voiceless stops as opposed to voiced stops can also be compared, as both types of phones are tested. Even though the two populations under study are comparable in terms of second language experience, it is possible that Spanish learners of English and English learners of Spanish encounter different difficulties in the acquisition of L2 stops. The fact that English presents a greater VOT range – from voice-lead VOT to long-lag VOT – than Spanish – from voice-lead to short-lag (e.g., Lisker & Abramson, 1964) – may result in an advantage on the part of the English learners over the Spanish learners. However, this advantage may be limited, since even though English speakers may use all three VOT conditions, voice-lead and short-lag VOT are not used contrastively. Regarding voiceless stops, it is predicted that Spanish learners use less aspiration than English native speakers, whereas English speakers may either use more aspiration than Spanish speakers or simply use their short-lag category. As for voiced stops, it is expected that Spanish speakers will readily use their L1 prevoiced category, as some previous studies have found in speakers of other languages with a similar VOT system to that if Spanish (e.g., Caramazza et al., 1973; Hazan & Boulakia,1993). The use of prevoicing is predicted to decrease as participants gain L2 experience thanks to the exposure to the more common production of voiced stops – i.e., with short lag VOT – but prevoicing is still expected to occur more frequently in the productions of Spanish learners of English than in those of native speakers. English speakers, however, will have to learn to implement a phonetic feature – i.e., prevoicing – contrastively which may as well be challenging. Casillas (2019) found that English beginning learners of Spanish used prevoicing in Spanish voiced stops approximately half the time after a 7-week immersion program in the L2. Thus, experienced learners of Spanish are expected to be able to produce Spanish voiced stops accurately. Overall, the acquisition of L2 voiced stops is expected to be more challenging than that of voiceless stops for both populations (Caramazza et al., 1973; Casillas, 2019).

This section has stated the goal of the current thesis and presented its main research questions. Moreover, the hypotheses have been discussed according to the predictions of the main speech acquisition models and the results of previous research. The following section will thoroughly explain the methodology that has been implemented in order to test the research questions addressed in this study.

4. Methodology

In order to test the perception and production of stops in the L1 and the L2 by L1-Spanish L2-English and L1-English L2-Spanish speakers, two perception tasks and a production task were designed in each language. That is, all participants were tested in their L1 and L2, namely Spanish and English. Studies analyzing VOT perception have traditionally used synthetic stimuli rather than natural speech. This is true of studies evaluating perception in one language (e.g., Abramson & Lisker, 1973), as well as in both the L1 and the L2 (e.g., Williams, 1977). However, synthetic speech is more difficult to encode and requires a greater cognitive effort than natural speech (Pisoni, Manous & Dedina, 1987). More recently, modified natural speech has been used to test bilinguals' categorization of stops (e.g., Casillas & Simonet, 2018; Gonzales & Lotto, 2013; Hazan & Boulakia, 1993; Lev-Ari & Peperkamp, 2013). Thus, the perception experiments reported in this thesis use modified natural speech involving a /p/-/b/ and a /k/-/g/ continuum, as described below (4.2.1). Regarding production, participants had to complete a sentence reading task in which stop-initial target words were embedded.

One of the methodological contributions of this thesis is that it allows the comparison of L2 learners with not only different amounts of L2 experience, but also different L1s. Most previous studies have analyzed the effect of L2 experience on VOT production/perception by contrasting the performance of experienced and inexperienced bilinguals in their L2 and using monolingual speakers' results as a baseline. As mentioned in the literature review (section 2.3.1), Flege (1987) went a step further and analyzed the productions of English-French bilinguals in France and in the United States in both their L1 and their L2, as well as the productions of monolingual speakers of both languages. However, Flege tested the French learners of English in an L2 context, but not in a setting where their L1 is spoken. In addition, Flege (1987) focused on production and did not examine perception. The current study analyzes the perception and production of mirror-image Spanish-English and English-Spanish bilinguals varying in amount of L2 experience and language setting. The following section will provide details on the participants' L1 and L2, their amount of L2 experience, location, gender, age, language dominance, years of L2 instruction, age of L2 learning and amount of weekly L2 use. Next, the method used for the creation of the stimuli and the tasks' design and procedure will be described.

4.1. Participants

As mentioned above, the present thesis studies the effect of L2 experience on VOT production and perception by English-Spanish bilinguals in both an L1 and an L2 context. Accordingly, the data collection took place in two different locations – i.e., in London (United Kingdom) and in Barcelona (Spain) – and included eight mirror-image groups, namely four groups of L1-Spanish speakers and four groups of L1-English speakers differing in amount of L2 experience.

A total of 92 participants were recruited and tested. However, only 80 of them were included in the analysis – 42 females and 38 males –, whereas 11 were discarded for several reasons, such as difficult intelligibility, mixed linguistic background and differences with other participants in the same group in terms of recency of stay in the L2 country. Table 4.1 shows the basic group information, including group name, number of participants, gender, L1, L2, the number of months spent in the L2 setting and the location where they were tested – i.e., Barcelona or London. An extended version of this table with more detailed group information is provided in Appendix A, including additional information about the recency of their stay in the L2 setting, as well as the ranges for each factor. Moreover, participants completed a consent form (see Appendix B) and a linguistic background and use questionnaire based on the Bilingual Language Profile (BLP, Birdsong et al., 2012). The questions can be consulted in Appendix C. The first data collection took place in the United Kingdom (UK) in the Fall of 2017. A total of 47 participants completed the experiment in London, but eight of them were discarded, as explained above. They were tested in the Phonetics Laboratory at Queen Mary University of London (QMUL). A second data collection took place in Barcelona (Spain) in the Fall of 2018. Most subjects were tested in the phonetics laboratory at Universitat Autònoma de Barcelona (UAB) – *Servei del Tractament de la Parla i del So* –, although some participants in the English experienced group were tested at the Universitat de Barcelona (UB) for convenience. Fourty-five participants completed the experiment in Barcelona, of which 41 were included in the analysis.

Group	*N*	Gender	L1	L2	Months in L2 setting	Location	% weekly L2 use	Years of L2 inst.
ENCONT	9	4 F; 5 M	English	NA	NA	London (UK)	-	-
ENINEXPinUK	11	6 F; 5 M	English	Spanish	none / minor	London (UK)	12.3% (8)	3.8 (3.3)
ENEXPinUK	10	4 F; 6 M	English	Spanish	9.4 (4.2)	London (UK)	15% (6.5)	7.5 (5.3)
ENEXPinSP	11	6 F; 5 M	English	Spanish	50.7 (27.3)	Barcelona (Spain)	16.3% (10.5)	2.4 (2.9)
SPEXPinUK	9	5 F; 4 M	Spanish	English	47.9 (23.3)	London (UK)	54.6% (24.6)	13.4 (2.3)
SPEXPinSP	10	7 F; 3 M	Spanish	English	7.4 (6)	Barcelona (Spain)	26.3% (10)	13.9 (1.9)
SPINEXPinSP	10	6 F; 4 M	Spanish	English	none / minor	Barcelona (Spain)	22.7% (9.7)	13.2 (2.2)
SPCONT	10	4 F; 6 M	Spanish	NA	NA	Barcelona (Spain)	-	-

Table 4.1. Characteristics of the eight groups that participated in the study including number of participants (*N*), gender, L1, L2, months in L2 setting, location at the time of testing, % weekly L2 use and years of L2 instruction (inst.). Standard deviations are provided in parentheses.

4.1.1. L1 English groups

The English control group (ENCONT) was made up by nine monolingual speakers of English – five males and four females – and had a mean age of 25.2 (*SD* = 5.4). They were undergraduate and graduate students at QMUL with minimal or no knowledge of a foreign language. Ten participants were tested, but one was excluded because he spoke a language other than English at home.

Fourteen inexperienced English learners of Spanish were tested. Two of them were discarded because they used languages other than English at home and one of them was not included in the analysis because she had spent a few months in Spain at a very early age. Eventually, the English inexperienced group (ENINEXPinUK) consisted of 11 participants – five males and six females – with a mean age of 20.2 (*SD* = 3.5). They were learners of L2 Spanish who had never lived in a Spanish-speaking country, although a few claimed to have visited Spain. They were recruited from the highest levels of Spanish classes at Queen Mary Language Center or were in their first or second year of the BA in Hispanic Studies. However, there was some variability regarding the number of years of L2 instruction that they had completed (*M* = 3.8; *SD* = 3.3). Whereas some started learning Spanish at university, others had already taken some Spanish classes in high school. Their mean age of learning was 13.4 years (*SD* = 3.7) and their level of Spanish was upper-intermediate to advanced. As for mean weekly L2 use, it was calculated based on self-reported answers elicited in the BLP questionnaire, including weekly L2 use with family, friends and at work/school. ENINEXPinUK revealed an average of 12.3% (*SD* = 8) of weekly L2 use. Their language dominance was assessed on the basis of their responses to the BLP questionnaire, which establishes a language dominance score that ranges from -218 to 218. In this case, values close to zero indicate balanced bilingualism, whereas negative values show Spanish dominance and positive values English dominance. ENINEXPinUK obtained a language dominance index of 108.7 (*SD* = 27.3), that is, they were found to be clearly English dominant.

The L1-English L2-Spanish moderately experienced group living in the UK (ENEXPinUK) was made up by six males and four females – a total of 10 participants – with a mean age of 24.2 years (*SD* = 4.8). Initially, four extra participants were tested and discarded for the analysis. Two of them were not analyzed because they spoke a language other than English at home and the other two were discarded because their stay in a Spanish-speaking country took place considerably earlier than that of the other participants in the group. All participants were in their last year of Hispanic Studies and had spent at least one term (*M* = 9.4; *SD* = 4.2) in a Spanish-speaking country during the previous academic year as part of the year abroad program of their studies. Nevertheless, similarly to ENINEXPinUK, participants differed in years of L2 instruction (*M* = 7.5; *SD* = 5.3), as some started to learn Spanish before university, but others did not. Their mean age of learning was 15.5 (*SD* = 3.7). ENEXPinUK obtained a mean score of weekly L2 use of 15% (*SD* = 6.5) and a dominance index of 106 (*SD* = 21.7) – i.e., English-dominant.

The L1-English highly experienced group living in Spain (ENEXPinSP) was made up by 11 teachers of English as a foreign language – five males and six females – with a mean age of 33.2 (*SD* = 6.1). Participants had a heterogeneous dialectal background. Four participants were from Southern England, one from Northern England, one from Scotland, two from Ireland, one from Canada, one from the East Coast of the United States and one from the South of the United States. Even though ideally all participants in this group should be speakers of the same English dialect, participants from different English-speaking areas had to be included in the experiment due to time limitations and the difficulty recruiting volunteers to participate in the study. Participants had been living in Spain for a mean of 50.7 (*SD* = 27.3) months at the time of testing. It should be noted that, although this is the most L2-Spanish experienced English group, they had learnt Spanish formally for a shorter period of time than the groups tested in London, namely for 2.4 years (*SD* = 2.9) and started to learn it later – i.e., with a mean of 18.9 years (*SD* = 2.1). In fact, all but two subjects in the group started to learn Spanish after adolescence and five of them had never learnt Spanish formally. Moreover, their L2 use is more limited than that of their mirror-image group – i.e., SPEXPinUK –, given that they used – almost exclusively – English at work. They obtained a weekly L2 use score of 16.3% (*SD* = 10.5), which is not too different from the ones obtained by the English learners in the UK – 12.3% (*SD* =8) in the case of ENINEXPinUK and 15% (*SD* = 6.5) regarding ENEXPinUK. They obtained a language dominance index of 116.5 (*SD* = 26). Although they were found to be the most highly L1-dominant English group of learners, their score was again very close to that of English learners of Spanish in the UK (ENEXPinUK: 106; ENINEXPinUK: 108.7).

4.1.2. L1 Spanish groups

Ten participants – 4 females and 6 males – with a mean age of 22.7 (*SD* = 4.7) make up the Spanish control group (SPCONT). All of them were also speakers of Catalan. This was not considered a problem given that both Catalan and Spanish have an opposition between voice-lead and short-lag VOT and, in fact, very similar VOT values have been reported for the two languages (Castañeda, 1986; Julià i Muné, 1981; Lisker and Abramson, 1964). Moreover, SPCONT had some basic knowledge of English, as, in Spain, English is a mandatory subject in school, but they rarely used English and self-reported to have a level below intermediate. Furthermore, they had never lived in an English-speaking country.

The Spanish inexperienced group of learners of English (SPINEXPinSP) was formed by 10 participants – 6 females and 4 males –, who were second-year students of English Studies at UAB with a mean age of 21.1 (*SD* = 3.1). According to the level of proficiency required for the second year of the English Studies degree, it is assumed that participants had an upper-intermediate to advanced level of English. Moreover, they had never lived in an English-speaking country. They had been learning English for a mean of 13.2 years (*SD* = 2.2) and started to learn it with a mean age of 5.4 (*SD* = 1.9). Given that English is a mandatory subject in Spanish schools and high schools, we did not expect to find great differences between the Spanish groups considering the amount of years learning English and the age of learning. Moreover, they obtained a weekly L2 use score of 22.7% (*SD* = 9.7) and a dominance index of -81.9 (*SD* = 26.4), showing dominance towards Spanish.

Ten participants – seven females and three males – who had lived in an English-speaking country constitute the Spanish moderately experienced group tested in Spain (SPEXPinSP). Initially, three more participants completed the experiment. Two of them were discarded because their stay abroad took place substantially earlier than that of the other participants. Another participant was not included in the study because she spoke two languages with a different use of VOT to categorize stops – namely German and Spanish – at home. Participants had a mean age of 22.9 (*SD* = 4.6) and were fourth-year students and recent graduates of English Studies who had resided in an English-speaking country during the previous academic year for at least one term. The average number of months they had lived in an English-speaking country was 7.4 (*SD* = 6). According to their studies, their English proficiency was supposed to be advanced to near-native. They had been learning English for a mean of 13.9 years (*SD* = 1.9) and started to learn it at a mean age of 6.1 (*SD* =1.6). As for weekly L2 use, SPEXPinSP obtained a mean of 26.3 % (*SD* = 10) and a dominance index of -75.2 (*SD* = 25.5) – i.e., Spanish dominant.

The highly experienced Spanish participants living in London (SPEXPinUK) were four males and five females – a total of nine participants – with an average age of 24 (*SD* = 3.2). They had been living in the UK for a mean of 47.9 months (*SD* = 23.3) at the time of the experiment. They came from different areas of Spain, including Catalonia, Valencia and Madrid. Thus, most spoke Catalan in addition to Spanish. They had been learning English for 13.4 years (*SD* = 2.3) and started to learn it at a mean age of 5.8 (*SD* = 8.7). As for L2 use, SPEXPinUK is the group that obtained the highest score, namely 54.6% (*SD* = 24.6). Moreover, they presented the most balanced language dominance index, with a score of -41.5 (*SD* = 29.8), although they were still found to be L1 dominant.

All in all, the L1-Spanish and the L1-English mirror-image groups were comparable in terms of amount and recency of L2 experience – the main factors examined in this thesis – as well as in age – all participants were young adults. Moreover, gender was counterbalanced. However, it should be noted that the Spanish and the English groups differed in terms of length of L2 learning, age of L2 learning, L2 use and language dominance. Although, ideally, comparable groups regarding the factors mentioned above should have been tested, this was not possible. Due to the linguistic reality of Spain as opposed to the UK – or any other English-speaking country for that matter – there is a difference in L2 exposure between the L1-Spanish and the L1-English groups. As mentioned above, English is a mandatory subject in the Spanish curriculum, whereas that is not the case of Spanish in the UK. Moreover, it is assumed that Spanish young people consume audiovisual content in English more often than English-speakers do in Spanish. It should also be mentioned that, as stated above, the English experienced group in Spain used their L1 at work, whereas the Spanish group in the UK used the L2. As a result, all

L1-Spanish groups had learnt English for a longer period of time and had started earlier than the L1-English groups regarding Spanish. Furthermore, the Spanish groups used their L2 substantially more often than the English groups. As a matter of fact, the L1-Spanish group that used English the least during the week – i.e., SPINEXPinSP, with 22.7% – made a greater use of their L2 than the L1-English group that used Spanish the most – i.e., ENEXPinSP, with 16.3%. Similarly, although all groups were L1-dominant, all Spanish groups were found to be more linguistically balanced than the English groups, being the most L1-dominant Spanish group – i.e., SPINEXPinSP, with a dominance score of -81.9 – more balanced than the least L1-dominant English group – ENEXPinUK, with a score of 106. Given these differences between the two main populations, when analyzing the results of the experiments, a chapter will be dedicated to exploring individual differences and their effect on the perception and production of L2 and L1 stops (see Chapter 8). The design of the tasks used in the perception and the production experiments is explained next.

4.2. Task design

4.2.1. Perception experiment

The perception experiment consisted of two identification tasks – i.e., one for /p/-/b/ and one for /k/-/g/ – which were presented once in each language. The tests included a number of stimuli that varied along a VOT continuum, ranging from voice-lead to long-lag VOT. The following subsections describe how the stimuli were created and selected (see section 4.2.1.1) and how the task was implemented (see section 4.2.1.2.).

4.2.1.1. Stimuli creation and testing

Initial /p/-/b/ and /k/-/g/ productions of a phonetically trained male bilingual speaker of Spanish and English were recorded in an acoustically treated room at the Phonetics Laboratory at UAB using an Audio-Technica AT 2050 microphone and an Alesis Multimix 8 mixer. The software used for the recording was Audacity. The speaker was asked to deliberately produce instances of aspirated, short-lag unaspirated and prevoiced bilabial and velar stops followed by the high-front vowel /i/, which has been reported to be one of the most perceptually equivalent vowels in Spanish and English (Cebrian, 2019).

One continuum was created on the basis of a selected token for each of the three VOT conditions and place of articulation in order to pilot the resulting stimuli with native listeners of both English and Spanish and then choose the best continuum to test the two populations under study. A total of 30 productions – that is, 10 instances each of prevoiced, short-lag and long-lag VOT productions – for each of the two points of articulation examined were extracted and analyzed. Since the experiment focuses on VOT perception, an attempt was made to neutralize other cues for voicing (see section 2.5 'Voicing cues in stops' in the literature review for a detailed explanation). A total of 30 productions, that is, 10 instances each of type of VOT production – i.e., prevoiced, short-lag and long-lag – for each of the two places of articulation examined were extracted and analyzed. The following secondary cues were measured for each production: burst duration, burst intensity, vowel intensity, burst center of gravity (COG), f0 at burst, f0 transition into the vowel, F1 at burst and F1 transition into the vowel. The mean values of these secondary cues were calculated for each phone – i.e., [p^h], [p], [b], [k^h], [k] and [g] – to determine the prototypical values of each secondary voicing cue in each context.

Given that no consistent pattern among the phones was found regarding COG, this cue was disregarded.

Thus, the mean values for the remaining cues were calculated for each place of articulation – i.e., the mean among prevoiced, short-lag and aspirated /p/-/b/ productions and the mean among prevoiced, short-lag and aspirated /k/-/g/ productions – with an aim to obtain ambiguous values in terms of voicing. A burst for each type of VOT production – i.e., prevoiced, short-lag and long-lag VOT – and a vowel were selected and modified to match the ambiguous values that were obtained (see Table 4.2). The bursts extracted from each context that had the closest duration to the mean calculated for each point of articulation (/p/-/b/: 9.5 ms; /k/-/g/: 16.5 ms) were used to create the stimuli. That is, a total of six bursts – three for each place of articulation – were selected. Then, the intensity of the burst was modified in order to obtain ambiguous values in terms of voicing (/p/-/b/ continua: 59.6 dB; /k/-/g/ continua: 57 dB). Regarding the vowel, the productions that had the closest values to the mean f0 and F1 transitions calculated for each contrast were selected for each place of articulation. In the same fashion as with the burst, the intensity of the vowel was modified so as to match the average calculated between the three types of productions for each point of articulation (/p/-/b/: 68.8 Hz; /k/-/g/: 69.9 Hz). The same vowel was used in all the tokens from the three continua created for each place of articulation (see Table 4.3 for the vowel measurements).

	Duration	Burst intensity	Vowel intensity
Mean for all bilabial tokens	9.5	59.6	69.8
selected [p^h] token	9.8	59.6	69.8
[p] continuum	8.5	59.6	69.8
[b] continuum	8.6	59.6	69.8
Mean velars	16.5	57	68.9
[k^h] continuum	16.6	57	68.9
[k] continuum	17	57	68.9
[g] continuum	16.4	57	68.9

Table 4.2. Mean burst duration (in ms), mean burst intensity and mean vowel intensity (in dB) calculated for all bilabials and all velars and the actual values used in each continuum.

	/p/-/b/	/k/-/g/
f0 at 10 ms	136.4	153.4
f0 at 20 ms	135.6	146.8
f0 at 30 ms	135.3	145.5
f0 at 40 ms	135.4	144.7
f0 at 50 ms	135.9	144.5
f0 mean	114.6	143.4
F1 at vowel onset	378	392.5
F1 at 10 ms	382.1	389.9
F1 at 20 ms	382.9	381.2
F1 at 30 ms	375.2	384
F1 at 40 ms	362.2	393
F1 at 50 ms	355.2	391.4

Table 4.3. Mean F0 and F1 measurements (in Hz) at different points into the vowel for the vowels selected for the creation of the /p/-/b/ and the /k/-/g/ continua.

A total of six continua – i.e., three for place of articulation and one for each type of burst, extracted from a prevoiced, short-lag and aspirated stop – were created. After modifying intensity, the selected bursts and vowels were concatenated in Praat (Boersma & Weenink, 2016) to create the first stimuli, which served as a basis for the creation of the remaining steps. Stimuli with negative VOT values were created by inserting cycles of prevoicing of ca. 5 ms that had been extracted from the same production with voice-lead VOT. Similarly, aspirated stimuli were created by the addition of aspirated portions of speech in steps that increased in about 5 ms.

The resulting continua were tested on six Spanish and six English native speakers with no or minor knowledge of foreign languages in order to determine which set of stimuli were more appropriate for the purpose of this experiment. Two two-alternative forced-choice identification tests – i.e., one including all bilabial tokens and one with all velar tokens – in Spanish and in English were created using the tool FOLERPA (Ferramenta On-Line para ExpeRimentación PerceptivA, Fernández Rei, Aguete Cajiao & Osorio Peláez, n.d.) and were administered online. The /p/-/b/ pilot identification test included a total of 52 stimuli varying in steps of about 5 ms in a VOT range of -30 ms to 38 ms approximately. The VOT range was decided considering the results obtained in Gorba (2016). Regarding the /k/-/g/ pilot test, participants heard a total of 78 stimuli – i.e., 26 for each continuum – with VOT values that ranged from about -50 ms to 92 ms. More stimuli were tested for /k/-/g/ because the experimenter did not have previous experience testing velar stop perception. In both continua, stimuli were presented 3 times in a random order. Task instructions and identification labels were provided in the language that was being tested. Response labels included the sound and an example of a word beginning with the target stop and vowel /i/. Thus, the words 'Peter' and 'beetle' were used as response alternatives in the English identification test, whereas '*pico*' (peak) and '*bicho*' (bug) were used in the Spanish test. Regarding the tests including velar stops, the options provided were 'keeper' and 'geeky' in the English test and '*quita*' (release) and '*guita*' (twine) in the Spanish test.

After participants had completed the pilot tests, one VOT continuum for each place of articulation was selected on the basis of response consistency. Moreover, a VOT range that was appropriate for both Spanish and English speakers was determined based on the participants' responses. The objective was to use a VOT range that went from clearly voiced to clearly voiceless stop that was appropriate to test both English and

Spanish perception, as well as to limit the number of steps to a practical number. The final /p/-/b/ continuum included stimuli that were created using a burst that was extracted from a voiced context and had a VOT range of -30.4 to 57.9 ms. Similarly, the stimuli selected for the /k/-/g/ continuum included a burst extracted from a voiced context and had VOT values ranging from -25 ms to 72 ms. It should be noted that greater VOT values are used in the velar identification tests than in the bilabial tests because /k/ presents longer VOT than /p/ (see section 2.5.1 for more details). In both continua, 17 stimuli were used for the final data collection. Table 4.4 shows the actual VOT values of all steps in the continua[2].

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

Table 4.4. /p/-/b/ and /k/-/g/ continua: actual VOT values in ms.

[2] Even though the two continua have the same number of steps, which vary in intervals of 5 ms, the difference between the two endpoints is greater regarding the /k/-/g/ continuum than in the case of the /p/-/b/ continuum. This difference stems from the fact that the burst used to create the velar continuum is greater than the burst used in the bilabial continuum as a result of place of articulation. It should also be noted that there are no steps with 0 ms of VOT given that all stimuli include a burst.

4.2.1.2. Identification task

A two-alternative forced-choice identification test including the stimuli described above was created for each place of articulation. Then, a version for each language was prepared, that is, a total of four tests were designed – a /p/-/b/ test in English, a /p/-/b/ test in Spanish, a /k/-/g/ test in English and /k/-/g/ test in Spanish. The stimuli presented in the Spanish and English tests were the same, but the instructions and response alternatives differed according to the language being tested. The tests were implemented in Praat, version 5.3.56 (Boersma & Weenik, 2016).

Stimuli were presented four times in a random order (i.e., 17 stimuli X 4 repetitions). Six practice trials preceded each test, so that participants became familiar with the task and the type of stimuli that they were going to be exposed to. Every trial presented a single stimulus and participants had the chance to hear it one more time by pressing on a replay button. Subjects had the option to take a short break after every 20 trials and then clicked on the screen to continue.

The options provided were written in conventional orthography. Just as in the pilot test used to select and validate the continua described above (see section 4.2.1.1), response labels included the sound and an example of a word beginning with the target sounds followed by vowel /i/. However, some of the response alternatives were changed from those used in the stimuli validation test. The response labels included the target sound and an example of a word beginning with that sound followed by vowel /i/ in the language that was being tested. The options used in the English /p/-/b/ tests were ‘‘b’ as in beetle’ and ‘‘p’ as peeler’. ‘Peter’ – which was the word used in the pilot test – was changed for ‘peeler’ because, as it is a proper name, it is sometimes used in Spanish as well and, thus, may not effectively activate an English language mode. ‘‘*b’ como en bicho*’ (‘‘b’ as in bug’) and ‘‘*p’como en pico*’ (‘‘p’ as in beak’) were the response alternatives used in the Spanish version of the /p/-/b/ test. Regarding the English /k/-/g/ test, the response alternatives provided were ''g' as in geeky' and ''k' as in keeper'. In Spanish, the words 'quita' and 'guita' were substituted for more frequent words, namely '*guiso'* (stew) and '*quito'* (I remove). The response alternatives used were ''*g' como en guiso'* (''g' as in stew') and ''*k' como en quito'* (''k' as in I remove'). Screenshots illustrating the way the response alternatives appeared on the screen at every trial in the English and the Spanish versions the identification tests are provided in Figures 4.1 and 4.2, respectively. Participants completed the tests in the Phonetics Laboratories located at UAB, UB or at QMUL. The same high-quality noise-cancelling headsets (Beyerdynamic DT 770 M) were used in all testing places.

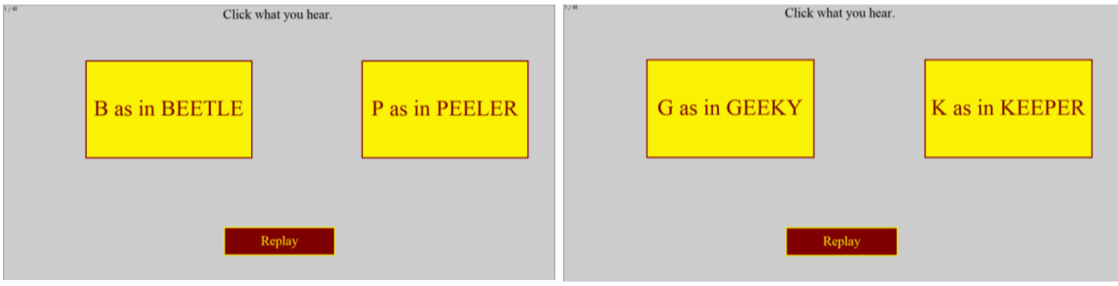

Figure 4.1. Screenshot of the English identification tests.

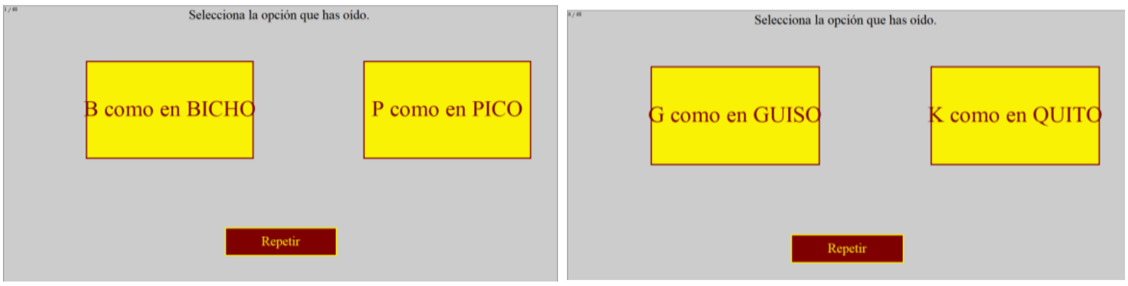

Figure 4.2. Screenshot of the Spanish identification tests.

4.2.2. Production experiment

Production data were elicited using two carrier sentence lists, one in Spanish and one in English. Participants were asked to read a total of 32 sentences in each language, which included 5 words for each target phone plus 12 fillers. Each sentence was repeated twice. The carrier sentences were ‘X is the next word’ in English and ‘X es la siguiente palabra in Spanish’. With a view to controlling for Spanish allophonic variation, the target word appeared in absolute initial position, thus avoiding post-vocalic spirantization of voiced stops, common in Spanish. Regarding the number of syllables, all words were disyllabic and they were stressed on the first syllable. The complete list of sentences for each language are provided in Appendix D.

The production of the participants recruited in London was recorded at the Phonetics Lab at QMUL in a soundproof room. A Neumann TLM 103 microphone and a booth mixer Steinberg UR22 MKII were used. Similarly, the production tasks completed in Barcelona were recorded in an acoustically treated room at the Phonetics Laboratory at UAB (*Servei del Tractament de la Parla i del So*) using an Audio-Technica AT 2050 microphone and an Alesis Multimix 8 mixer. In both cases, recordings were made with the software Audacity and the files were saved in WAV format. A few recordings were made at the GRAL’s (*Grup de Recerca en Adquisió de Llengües*, that is, the Language Acquisition Research Group) laboratory at UB in a soundproof cabin using a Shure-SM58 microphone and a Marantz PMD-660 digital recorder. The sound files were also recorded in WAV format.

4.3. Procedure and task order

The order in which participants completed the tests was carefully designed as well. All participants completed one of the reading tasks at the beginning. This was done in order to minimize the possibility that participants could guess what the target phones under study were, as the perception tasks are more explicit about what is being tested. So as to counterbalance language and place of articulation in the identification tests, bilingual subjects completed the tasks in four different orders which varied in the order of presentation of the language tested – i.e., orders A and B – and place of articulation – orders 1 and 2. The four orders are presented in Table 4.5. Place of articulation in the identification tests was counterbalanced in all groups, including controls (see Tables 4.5 and 4.6). Moreover, bilinguals had to watch a short video between the English and the Spanish experiments – that is, before changing the testing language – in order to activate the desired language mode. The total duration of the experiment was about 25 minutes for bilinguals and 10 minutes for the monolingual control groups, as they had to complete the test only in one language. The BLP questionnaire was sent to participants via e-mail previous to the day of the experiment and a consent form was signed just before the beginning of the experiment.

	A1	**A2**	**B1**	**B2**
1	English production	English production	Spanish production	Spanish production
2	EN ID TEST: /k/-/g/	EN ID TEST: /p/-/b/	SP ID TEST: /k/-/g/	SP ID TEST: /k/-/g/
3	EN ID TEST: /p/-/b/	EN ID TEST: /k/-/g/	SP ID TEST: /p/-/b/	SP ID TEST: /p/-/b/
4	Video in Spanish	Video in Spanish	Video in English	Video in English
5	Spanish production	Spanish production	English production	English production
6	SP ID: /k/-/g/	SP ID: /p/-/b/	EN ID: /k/-/g/	EN ID: /p/-/b/
7	SP ID: /p/-/b/	SP ID: /k/-/g/	EN ID: /p/-/b/	EN ID: /k/-/g/

Table 4.5. Counterbalanced orders of completion of the tasks among L2 speakers.

	A	**B**
1	L1 production	L1 production
2	L1 ID: /k/-/g/	L1 ID: /p/-/b/
3	L1 ID: /p/-/b/	L1 ID: /k/-/g/

Table 4.6. Counterbalanced orders of completion of the tasks among the control groups.

This section has presented the methodology followed in the two experiments conducted in this study, namely in the perception and the production experiment, including the characteristics of the participants, the stimuli creation, the task and the order of completion. The following chapter will explain how the data was analyzed and present the results for each of the experiments.

5. Perception experiment

The current chapter presents the results obtained in the perception experiment. First, the perception data analysis will be explained and the potential order effects will be examined. Then, the results obtained in the between-groups within-language experiment will be reported, first for the English condition and then for the Spanish condition, followed by a comparison of the results obtained for /p/-/b/ and /k/-/g/. After that, the results obtained in the between-languages within-group analysis will be presented. A summary of the results will be provided after presenting each analysis. Finally, the outcome of the two analyses will be discussed at the end of this chapter.

5.1. Data analysis

The perception experiment consisted of two identification tests – one for each point of articulation – and the participants performed two versions of each test, one in English and one in Spanish. The tasks involved the identification of stimuli from a voiced-voiceless stop continuum as either the voiced or the voiceless counterpart in each language. A total of 92 participants completed the experiment – 47 recruited in the UK and 45 in Spain. Given that some participants were found not to match the criteria for any group, 12 were discarded. Moreover, some data were lost and could not be used in the analysis. More specifically, some results for the Spanish perception tests were not properly saved, namely ENEXPinSP10's results for both places of articulation as well as SPINEXPinSP11's results for the Spanish /k/-/g/ test. Therefore, 80 participants were included in the analysis of the results for the perception experiment (9 ENCONT, 11

ENINEXPinUK, 10 ENEXPinUK, 10/11 ENEXPinSP, 9 SPEXPinUK, 10 SPEXPinSP, 9/10 SPINEXPinSP and 10 SPCONT).

Participants' responses in the perception experiment were extracted from the Praat tests and collected and analyzed in an Excel sheet. The dependent measure was percent identification as the voiceless stop at each step in the continuum, which served to obtain an identification function for each participant. The location of the function on the continuum indicates the amount of VOT necessary for participants to start identifying voiceless stops; that is, the later on the continuum the function is located, the more VOT is needed to perceive a voiceless stop. The steepness of the curve shows the degree of hesitation of participants in their identification responses. If very few steps are needed from consistently hearing a voiced stop to consistently hearing a voiceless stop, the categorization curve will be steep. The more steps needed to transition from voiced to voiceless identification, the shallower the slope will be. The steps on the continuum where participants hesitate the most – that is, where voiceless stops obtain close to 50% of identifications – are considered the crossover area between the two categories.

The measures used to quantify perception were category boundary – i.e., the exact point in the continuum where a listener stops categorizing a phone in favour of another one – and slope – which shows consistency in participants' responses –, following Aliaga-García and Mora (2009) and Casillas and Simonet (2018). Category boundary was calculated by converting the identification responses along the continuum of each participant into a logistic function. The resulting constants (b0) and slopes (b1) were applied in the formula -LN(b0)/LN(b1) in order to obtain a numeric value for their boundary. The resulting category boundary was given in ms. Slope was also explored as a measure. However, no consistency was found within groups and almost none of the comparisons between groups yielded a significant result. For this reason, the analysis of the perception results focuses on the boundary measure.

In order to explore the effects of L2 experience on L1 and L2 stop perception, a series of analyses were carried out. First, for each language – i.e., English and Spanish–, the results for all L2 groups sharing the same L1 and the corresponding control group were compared. These between-groups within-language analyses were conducted by means of one-way ANOVAs (dependent variable: category boundary; independent variable: group). Secondly, the performance of the groups sharing the same L1 in the two languages tested was analyzed by means of two-way ANOVAs (dependent variable: category boundary; independent variables: language and group). In the event that a significant effect of language was found, paired-samples t-tests (dependent variables: English boundary, Spanish boundary) were carried out for each of the three groups to see if the difference between languages yielded different degrees of significance for every group. Moreover, Pearson's correlations were used to compare the results for each place of articulation tested – i.e., bilabial and velar – and to investigate the direction of boundary shift in bilingual speakers (section 5.3.3). The effect of order – which was counterbalanced across participants, as explained in the 'Methodology' (Chapter 4) – was also examined by conducting a two-way ANOVA for each language, as will be discussed next.

5.2. Order effects on the perception task performance

Half the participants completed the experiment first in English and then in Spanish, whereas the other half completed the test in the reverse order. Thus, first, an analysis was carried out to determine whether the order in which languages were tested –

i.e., English first or Spanish first – had an impact on category boundary location. Separate two-way ANOVAs were conducted for each language – i.e., English and Spanish – place of articulation – i.e., bilabials and velars – and L1 – i.e., L1-English participants and L1-Spanish participants. The dependent variable was category boundary and the independent variable was test order – i.e., English-Spanish vs. Spanish-English.

Regarding English perception, it was found that all L1-English groups obtained numerically later /p/-/b/ category boundaries when they completed the experiment in Spanish first than when they did so in English, as Table 5.1 shows (English-Spanish order: ENINEXPinUK: 14 ms; ENEXPinUK: 15.4 ms; ENEXPinSP: 14.9 ms; Spanish-English order: ENINEXPinUK: 15.4 ms; ENEXPinUK: 16.4 ms; ENEXPinSP: 17.5 ms). Smaller differences were found regarding /k/-/g/, particularly in the case of ENINEXPinUK, who obtained almost identical category boundaries in the two conditions (English-Spanish: 27.3 ms; Spanish- English: 27.2 ms). A somewhat small difference between the two orders was also found regarding ENEXPinUK (English-Spanish: 26.6 ms; Spanish-English: 27.7 ms), whereas the greatest numerical difference was observed in the case of ENEXPinSP (English-Spanish: 27.5 ms; Spanish-English: 31.8 ms). That is, ENEXPinSP presented a later category boundary in English when they were tested first in Spanish. However, order effects did not reach significance for any of the contrasts [/p/-/b/: $F(1, 23) = 0.36$; $p = .550$; /k/-/g/: $F(1, 23) = 2.032$; $p = .167$].

Group	/p/-/b/		/k/-/g/	
	English-Spanish	Spanish-English	English-Spanish	Spanish-English
ENINEXPinUK	14 (3.2)	15.4 (2.9)	27.3 (10.6)	27.2 (4.1)
ENEXPinUK	15.4 (2)	16.4 (5.5)	26.6 (2.2)	27.7 (6.3)
ENEXPinSP	14.9 (2.3)	17.5 (1.6)	27.5 (2.2)	31.8 (3.9)
SPINEXPinSP	9.2 (2.5)	9.3 (4.4)	21 (2.2)	24.1 (3.6)
SPEXPinSP	8.2 (0.6)	11.3 (5.1)	23.4 (2.5)	25.9 (4.5)
SPEXPinUK	12.4 (2)	11.7 (3.6)	24.3 (2.2)	24.7 (3)

Table 5.1. All groups' mean category boundaries (in ms) for /p/-/b/ and /k/-/g/ in English according to order of testing (English-Spanish vs. Spanish-English). The standard deviation is provided in brackets.

Regarding the Spanish learners, no clear pattern was found in the case of the category boundary for /p/-/b/: SPINEXPinSP obtained almost identical boundaries in both orders (English-Spanish order: 9.2 ms; Spanish-English order: 9.3 ms), SPEXPinSP presented a later boundary when they were tested in Spanish first (11.3 ms) than when they were tested in English in the first place (8.2 ms), whereas the opposite was true regarding SPEXPinUK (English-Spanish order: 12.4 ms; Spanish-English order: 11.7 ms), although in this case the difference between the two orders was relatively small. As for /k/-/g/, SPINEXPinSP and SPEXPinSP obtained somewhat later category boundaries when they completed the experiment in English in the first place than when they did so in Spanish (English-Spanish order: SPINEXPinSP: 21 ms; SPEXPinSP: 23.4 ms; Spanish-English order: SPINEXPinSP: 24.1 ms; SPEXPinSP: 25.9 ms), whereas almost identical boundaries were obtained by SPEXPinUK in the two orders (English-Spanish order : 24.3 ms; English-Spanish order: 24.7 ms). Nevertheless, just as in the case of the English learners, no order effect was found neither for the /p/-/b/ category boundary nor for the /k/-/g/ category boundary [/p/-/b/: $F(1, 26) = 2.142; p = .155$; /k/-/g/: $F(1, 26) = .631$].

As for the perception of Spanish stops, no clear effect of order was found in the English groups. As Table 5.2 illustrates, ENINEXPinUK obtained later boundaries for

/p/-/b/ when they were tested first in English than when they completed the Spanish tasks in the first place (16.9 ms and 13.5 ms, respectively), whereas the opposite was found regarding ENEXPinUK (English-Spanish order: 15.6 ms; Spanish-English order: 17.4 ms) and ENEXPinSP (English-Spanish order: 14.2 ms; Spanish-English order: 16 ms). As for /k/-/g/, ENINEXPinUK had later category boundaries in the Spanish-English order than in the English-Spanish order (24.1 ms and 27.1 ms, respectively), whereas the opposite was found regarding ENEXPinUK (English-Spanish order: 28.3 ms; Spanish-English order: 22.9 ms) and ENEXPinSP obtained similar results in both orders (English-Spanish order: 27.9 ms; Spanish-English order: 28.2 ms). As a result, no significant effect of order was revealed [/p/-/b/: $F(1, 25) = .001$; $p = .99$; /k/-/g/: $F(1, 25) = .095$; $p = .760$].

In the case of the Spanish learners, all groups were consistently found to present later category boundaries in Spanish when they were tested in English first. In the case of /p/-/b/, SPINEXPinSP were the group that presented the greatest numerical difference between the two orders (English-Spanish order: 11.2 ms; Spanish-English order: 6.4 ms), whereas SPEXPinUK obtained the most similar category boundaries (English-Spanish order: 12.1 ms; Spanish-English order: 10 ms). As for /k/-/g/, SPEXPinUK had the most different category boundaries in the two orders (English-Spanish order: 23.6 ms; Spanish-English order: 18.5 ms), whereas SPEXPinSP obtained the most similar values (English-Spanish order: 22.6 ms; Spanish-English order: 19.7 ms). Thus, a significant effect of order was found both regarding /p/-/b/ and /k/-/g/ category boundary [/p/-/b/: $F(1, 23) = 7.717$; $p = .011$; k/-/g/: $F(1, 23) = 12.392$; $p = .002$]. That is, after having completed the English experiment in the first place, the Spanish learners of English tended to have earlier category boundaries in Spanish than when the Spanish test was performed first. It is possible, thus, that these participants' L1 perception may have been influenced by having performed the test in English in the first place – and having had to interpret the stimuli in terms of L2 categories – before having to do so in the L1[3]. Still, considering the lack of consistency of order effects across groups and languages – i.e., order was only found to have a significant effect on the L1 perception of the Spanish groups –, it was decided to conduct the same analyses on all groups, regardless of the order of completion of tasks. It should be noted that the perception results obtained by the English and the Spanish groups – who appear to be influenced differently by order effects – were analyzed separately. Moreover, order effects will be considered in the interpretation of the results. The following sections will present the between-group within language analyses, first for the English identification tests and afterwards for the Spanish tests.

Group	**/p/-/b/**		**/k/-/g/**	
	English-Spanish	**Spanish-English**	**English-Spanish**	**Spanish-English**
ENINEXPinUK	16.9 (3.7)	13.5 (4.2)	24.1 (9.5)	27.1 (6.3)
ENEXPinUK	15.6 (3.6)	17.4 (5.1)	28.3 (4.2)	22.9 (5)
ENEXPinSP	14.2 (3.4)	16 (2.7)	27.9 (2.7)	28.2 (3.3)
SPINEXPinSP	11.2 (1)	6.4 (2.8)	23 (2.6)	19.3 (4.5)
SPEXPinSP	9.3 (1.8)	6.7 (3.6)	22.6 (2.1)	19.7 (1.5)
SPEXPinUK	12.1 (2.4)	10 (4.7)	23.6 (1.8)	18.5 (3.3)

Table 5.2. All groups' mean category boundaries (in ms) for /p/-/b/ and /k/-/g/ in Spanish according to order of testing (English-Spanish vs. Spanish-English). The standard deviation is provided in brackets.

[3] The order in which each point of articulation – i.e., the /p/-/b/ and the /k/-/g/ tests – was presented was also counterbalanced. A two-way ANOVA was conducted considering all possible combinations of language and point of articulation orders – i.e., orders A1, A2, B1 and B2, as explained in the methods section – and replicated the same result as the analysis on language order. Spanish speakers presented greater VOT values in Spanish in both orders A1 and A2, that is, after completing the test first in English, regardless of point of articulation.

5.3. Between-groups within-language perception analyses

This section presents the results obtained in the between-groups within-languages analysis, that is a set of analyses comparing the performance of the groups in each of the languages under study. First, the perception of English stops will be presented for the English groups, followed by the Spanish groups. Then, the results obtained in the Spanish condition will be presented, first for the Spanish speakers and then for the English speakers.

5.3.1. Perception of English stops

5.3.1.1. Perception of English stops: English speakers

This section will present the results in the English identification test by the English-L1 groups – i.e., the raw data – and the category boundaries that were obtained after the conversion of results into a logistic function. Figure 5.1 illustrates the identification functions for /p/ obtained by the L1-English groups' in the English condition. ENCONT, whose results were used as the baseline values for English, stopped perceiving /b/ consistently at the 14 ms step, where identification as /p/ reached 19%. The step at which participants in the ENCONT group hesitated the most was the one presenting 19 ms of VOT (58.3%). No other steps obtained a percent score close to 50%, which resulted in a relatively steep identification curve. With 24 ms of VOT, all participants consistently heard /p/. As mentioned above, the measure used to quantify perception was category boundary. The ENCONT's category boundary for the bilabial contrast was 14 ms (see Table 5.3).

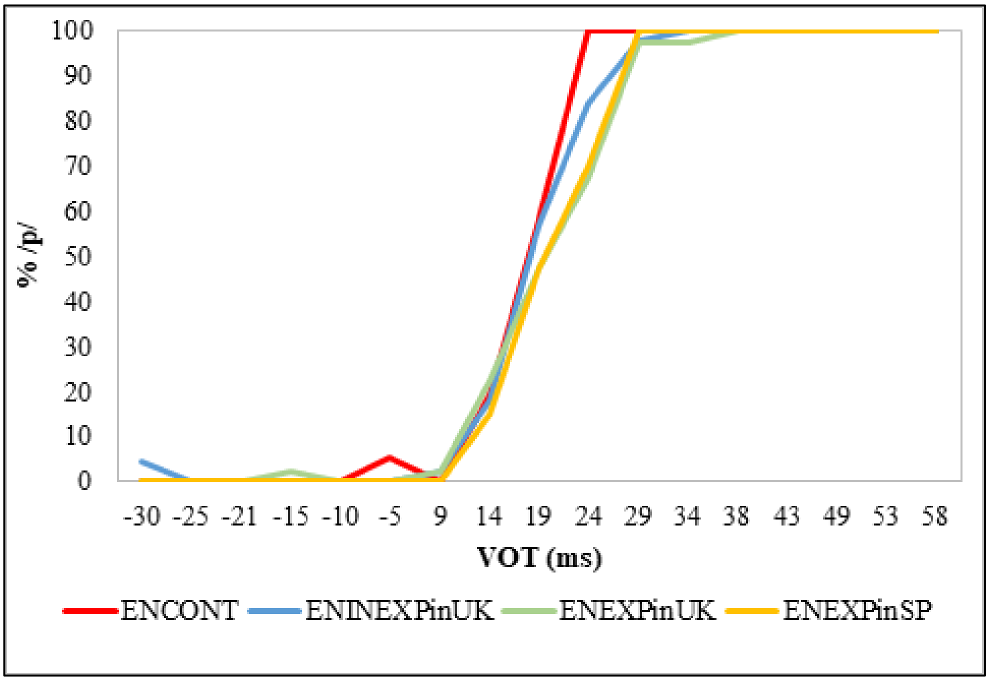

Figure 5.1. L1-English L2-Spanish groups' % identification functions for /p/ in English.

Group	*M*	*SD*	Range
ENCONT	14	2.5	9.1 – 15.8
ENINEXPinUK	14.5	3.1	9.1 – 19.4
ENEXPinUK	15.9	3.9	11.4 – 25.5
ENEXPinSP	16.1	2.4	12 – 19.7

Table 5.3. L1-English L2-Spanish groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English /p/-/b/ contrast in ms.

All learning groups seemed to need somewhat greater VOT values to categorize /p/ than ENCONT, especially the most experienced group (see Figure 5.1). The L1-English group that started hearing /p/ the earliest in the continuum – i.e., the group with the closest categorization curve to ENCONT's, in this case – was ENINEXPinUK. However, the shape of the ENINEXPinUK's slope was shallower than that of ENCONT. ENINEXPinUK's identification responses as /p/ reached 18% at the 14 ms step, 57% at the 19 ms step and 84% with 24 ms. As Table 5.3 illustrates, the category boundary of the ENINEXPinUK group (14.5 ms) was very close to ENCONT's (14 ms).

ENEXPinUK and ENEXPinSP performed similarly, as they presented an overlapping categorization curve along most of the /p/-/b/ continuum. Both groups needed more aspiration than ENCONT and ENINEXPinUK in order to start hearing /p/. With 14 ms of VOT, ENEXPinUK identified /p/ 23% of time, with 19 ms and 24 ms, they obtained the closest response scores to 50 %, namely 48% and 68% and with 29 ms they heard /p/ 98% of time. Their category boundary was set at 15.9 ms of VOT. As for ENEXPinSP, /p/ identification reached a score of 15% with 14 ms, 48% with 19 ms and 70% with 24 ms, reaching a score of 100% with 29 ms. Regarding their category boundary, it was very close to that of ENEXPinUK, as it presented 16.1 ms of VOT. Thus, numerically, learners tended to have later category boundaries than the controls, but the differences were small. A one-way ANOVA was conducted to compare the perception of English controls and L1-English L2-Spanish speakers varying in L2 experience. The statistical analysis failed to find a significant effect of group on category boundary [F (3, 37) = 1.162; p = .337], which indicated that English learners of Spanish did not present L2 influence on L1 bilabial stop perception, as they had native-like L1 categories.

The results for the velar stops are presented in Figure 5.2 and Table 5.4. The English control group needed 32 ms of VOT to hear /k/ in 31% of trials, 37 ms to hear /k/ 81% of time and 47 ms to hear it 92% of time. As Table 5.2 shows, ENCONT's category boundary presented a value of 29.7 ms. The three learner groups patterned similarly and presented a shallower categorization curve than ENCONT. ENEXPinUK and ENINEXPinUK seemed to start hearing /k/ slightly earlier than ENCONT and ENEXPinSP. ENINEXPinUK heard /k/ 18% of trials with 27 ms, 45% with 32 ms, 61% with 37 ms, 82% with 42 ms and 95% with 47 ms. Their category boundary presented a VOT value of 27.3 ms. ENEXPinUK performed very similarly to ENINEXPinUK. They obtained an identification score for /k/ of 23% with 27 ms, of 43% with 32 ms, of 60%

with 37 ms and of 85% with 42 ms, reaching 95% with 47 ms. Their category boundary was also very similar to that of ENINEXPinUK with a value of 27.2 ms of VOT. As for ENEXPinSP, they identified /k/ 16% of time at the 27 ms step, 34% at the 32 ms step, 61% at the 37 ms step, 83% at the 42 ms step and reached a score of 95% with 47 ms. Their category boundary presented 29.4 ms of VOT.

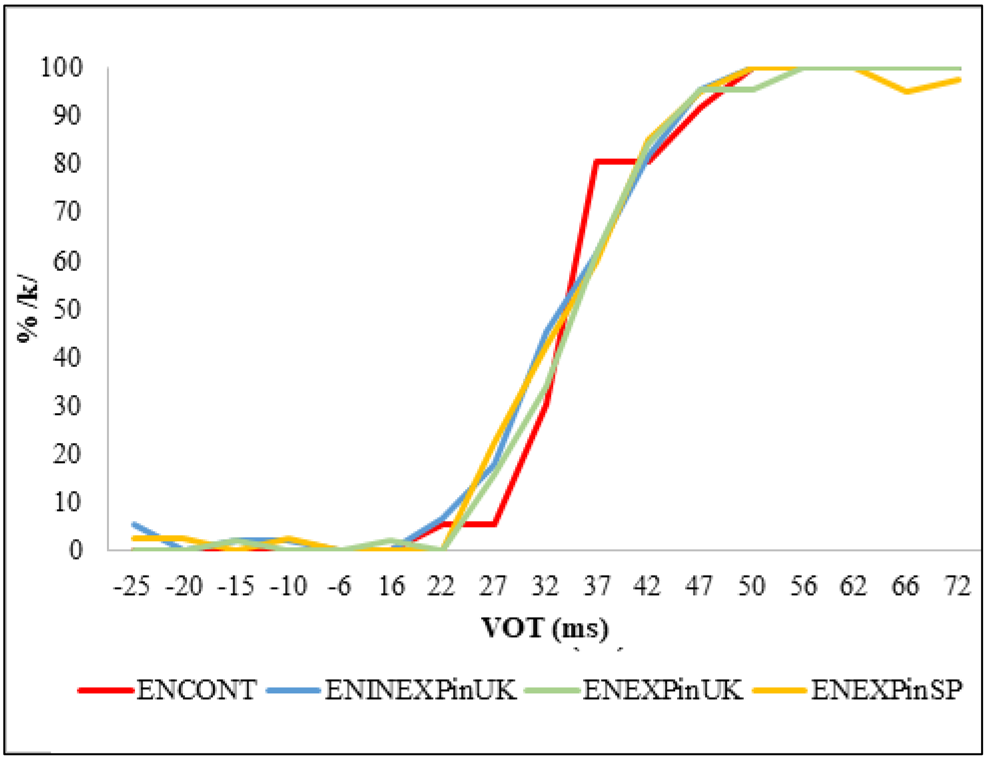

Figure 5.2. L1-English L2-Spanish groups' % identification functions for /k/ in English.

Group	*M*	*SD*	Range
ENCONT	29.7	4.3	22.6 – 35
ENINEXPinUK	27.3	8.5	7.4 – 38.1
ENEXPinUK	27.2	4.5	22 – 38.1
ENEXPinSP	29.4	3.7	24.5 – 38.2

Table 5.4. L1-English L2-Spanish groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English /k/-/g/ contrast in ms.

A one-way ANOVA was conducted to compare the /k/-/g/ category boundaries of bilingual L1-English speakers differing in amount of L2 experience and English controls. No statistically significant effect of group was found [F (3, 37) = 595; p = .622], which indicated that the English learners of Spanish performed in a native-like manner – and, therefore, they did not present L2 influence – in their L1, replicating the result observed for the bilabial stops.

In brief, in the case of the English learners of Spanish, the between groups comparison in English – i.e., their L1 – failed to find an effect of L2 experience on the categorization of the /p/-/b/ and /k/-/g/ contrasts. In both cases, all English learning groups presented similar category boundaries that were close to those of ENCONT, as no significant differences were revealed. The following section will present the results obtained by the Spanish groups in the English identification tests.

5.3.1.2. Perception of English stops: Spanish learners of English

The results of the English version of the bilabial stop identification test obtained by the Spanish speakers are presented in Figure 5.3 and Table 5.5. In this case, the test was examining L2 perception. As expected, L1-Spanish L2-English speakers needed less aspiration than English controls in order to hear /p/. The group that performed most similarly to ENCONT was SPEXPinUK, that is, the most experienced participants. SPEXPinUK identified stimuli as /p/ 17% of time at the 9 ms step, 42% of time at the 14 ms step, reaching 81% with 19 ms of VOT and 100% with 29 ms. Their category boundary was located at 12.1 ms, that is about 2 ms earlier than that of ENCONT (14 ms). SPEXPinSP and SPINEXPinSP performed similarly in the English /p/-/b/ identification task, although SPEXPinSP presented a categorization curve somehow closer to that of controls than SPINEXPinSP. Both groups had a steep curve in the middle steps of the continuum but appeared to hesitate more than ENCONT at the two ends of the continuum, especially in the last third. SPEXPinSP identified /p/ 28% if trials with 9 ms of VOT, 73% with 14 ms and 95% with 24 ms. Their /p/-/b/ category boundary presented a value of 10.1 ms. Regarding SPINEXPinSP, they obtained a /p/ percent identification score of 28% with 9 ms, of 78% with 14 ms, of 83% with 19 ms, reaching 93% with 24 ms. Their category boundary presented the lowest VOT value out of all the L1-Spanish English-L2 groups, namely 9.3 ms.

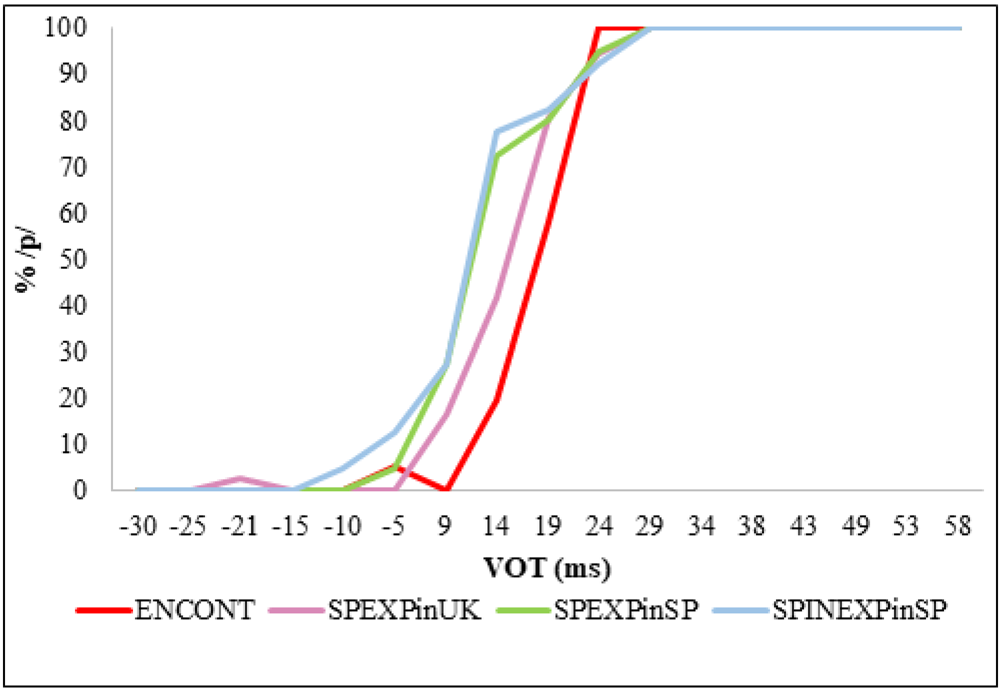

Figure 5.3. L1-Spanish L2-English groups' and English controls' % identification functions for /p/ in English.

Group	*M*	*SD*	Range
ENCONT	14	2.5	9.1 – 15.8
SPEXPinUK	12.1	2.6	9 – 16.8
SPEXPinSP	10.1	4.1	3. 4 – 15.8
SPINEXPinSP	9.3	3.6	4.4 – 17.1

Table 5.5. L1-Spanish L2-English groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English /p/-/b/ contrast in ms.

A one-way ANOVA revealed a significant effect of group on category boundary [F (3, 34) = 3.872; p = .017]. Bonferroni post-hoc tests found a significant difference between SPINEXPinSP and ENCONT (p = .023). None of the other comparisons yielded a significant result ($p > .05$). An effect of L2 experience on L2 stop perception was revealed, given that only the least experienced group differed significantly from English controls, meaning that the more experience groups had, the more likely it was that their category boundary fell within native English speaker values.

In the case of velar stop identification, SPEXPinUK and SPEXPinSP performed similarly to ENCONT and to one another (see Figure 5.4). As for SPEXPinUK, at the 27 ms and the 32 ms step, stimuli were identified as /k/ 42% and 73% of time, respectively, and, at the 42 ms, /k/ obtained an identification score of 95%. SPEXPinUK's category boundary was placed at 24.5 ms. In the case of SPEXPinSP, /k/ was identified 35% of time with 27 ms, 75% with 32 ms and started to identify /k/ 100% of time. As Table 5.6 shows, their category boundary presented 24.9 ms of VOT. The group presenting the earliest categorization curve was SPINEXPinSP, who identified /k/ 13% of trials with 22 ms of VOT, 50% with 27 ms and 95% with 37 ms. When it comes to /k/-/g/ category boundary, they obtained the lowest score (22.8 ms).

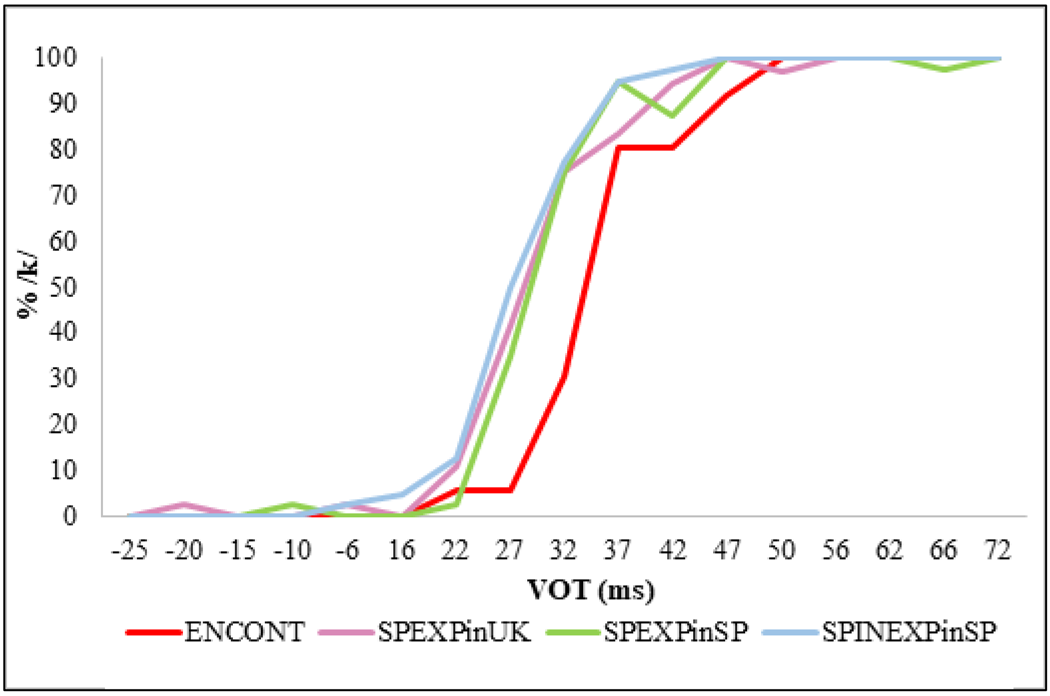

Figure 5.4. L1-Spanish L2-English groups' % identification functions for /k/ in English.

Group	*M*	*SD*	Range (ms)
ENCONT	29.7	4.3	22.6 – 35
SPEXPinUK	24.5	2.4	22 – 28.2
SPEXPinSP	24.9	3.9	20.2 – 33.3
SPINEXPinSP	22.8	4.4	-4.1 – 28.2

Table 5.6. L1-Spanish L2-English groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English /k/-/g/ contrast in ms.

The one-way ANOVA revealed a significant effect of group on category boundary [F (3, 34) = 5.527; p = .003]. Just as with /p/-/b/, Bonferroni post-hoc tests only showed a significant difference between SPINEXPinSP and ENCONT (p = .003), whereas all other comparisons were not statistically different (p > .05), indicating an effect of L2 experience on L2 /k/-/g/ categorization. In short, in the case of the Spanish learners of English, an effect of L2 experience was found on the location of the category boundary for both /p/-/b/ and /k/-/g/, as only SPINEXPinSP were found to present significantly earlier category boundaries from ENCONT, whereas the two experienced groups did not differ significantly from the English monolinguals.

5.3.2. Perception of Spanish stops

5.3.2.1. Perception of Spanish stops: Spanish speakers

SPCONT identified stimuli as /p/ 11% of the trials at the -10 ms step, rising to 43% with -5 ms and 68% with 9 ms and reaching 100% identifications with 19 ms of VOT (see Figure 5.5). Their corresponding category boundary was the earliest out of all Spanish groups, located at 3.4 ms (see Table 5.7). All L1-Spanish learners of English seemed to need greater VOT values than Spanish controls in order to start hearing /p/. SPINEXPinSP and SPEXPinSP obtained similar categorization curves, although SPEXPinSP's was slightly steeper. SPINEXPinSP started hearing /p/ with -5 ms (18%) and obtained an identification score of 48% with 5 ms and of 93% with 19 ms of VOT.

Their category boundary presented a value of 8.4 ms. SPEXPinSP heard /p/ 39% of time at the 9 ms step and 78% of time at the 14 ms step, reaching a score of 100% with 19 ms of VOT. They presented the closest category boundary to SPCONT with 7.8 ms. The group that differed the most from SPCONT was SPEXPinUK, as their categorization slope was located the latest on the continuum and was the shallowest one out of all Spanish groups. SPEXPinUK's /p/-/b/ crossover area was located between the 9 ms (28%), the 14 ms (50%) and the 19 ms steps (67%). With 24 ms of VOT, /p/ reached an identification score of 100%. SPEXPinUK obtained the latest category boundary out of all L1-Spanish speaking groups with 11.2 ms.

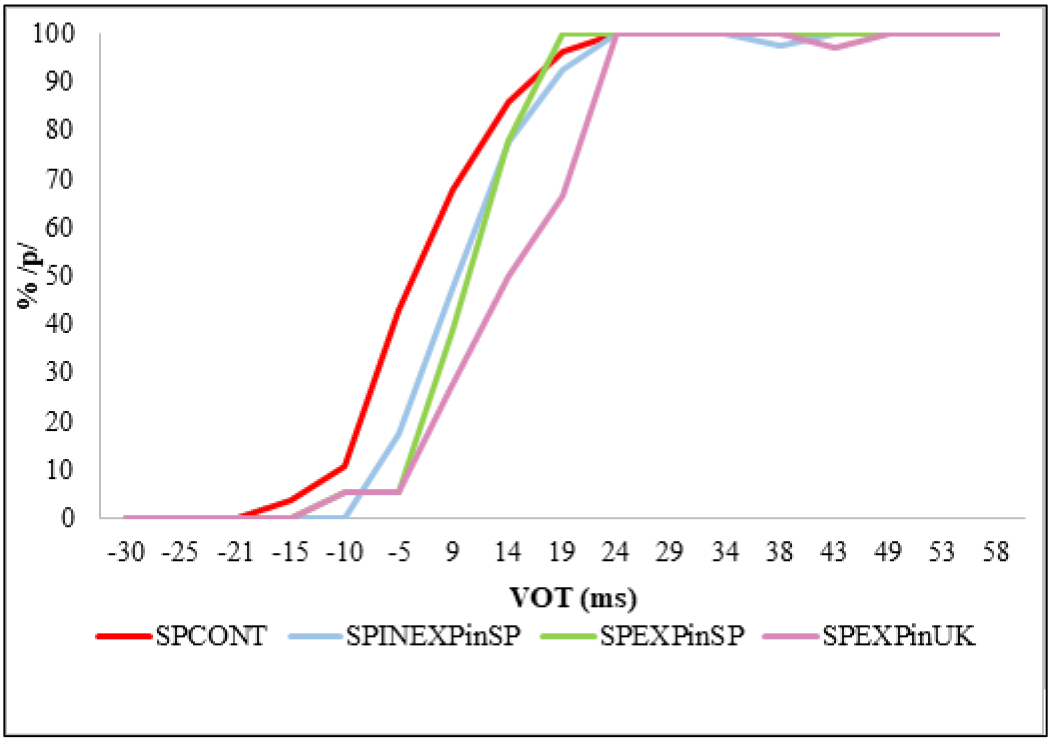

Figure 5.5. L1-Spanish L2-English groups' % /p/ identification functions in Spanish.

Group	*M*	*SD*	Range
SPCONT	3.4	5.4	-7.9 – 9.4
SPINEXPinSP	8.4	3.2	2.5 – 12
SPEXPinSP	7.8	3.2	0.4 – 11.7
SPEXPinUK	11.2	3.5	3.9 – 15.2

Table 5.7. L1-Spanish L2-English groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the Spanish /p/-/b/ contrast in ms.

A one-way ANOVA exploring the effect of group on the location of the /p/-/b/ category boundary was conducted in the same fashion as with the results for English speakers. The analysis revealed a significant effect of group [F (3, 35) = 6.233; p = .002]. The Bonferroni post-hoc tests found a significant difference between SPEXPinUK and SPCONT (p = .001). All other comparisons were not significant, but the SPEXPinSP and SPCONT comparison revealed a marginal difference (p = .052). These findings suggest that, in the case of the Spanish speaking groups, L2 experience had an effect on L1 /p/-/b/ perception, as L2 speakers seemed to present a greater L2 influence on their L1 as they gained experience with the L2.

When it comes to the velar contrast, SPCONT perceived /k/ 15% of time with -15 ms, 35% with 22 ms, 75% with 27 ms and 100% with 37 ms of VOT. Their category boundary was located at 14.8 ms. Just as in the case of bilabial stops, Spanish learners of English seemed to need greater VOT values than Spanish controls in order to start hearing /k/. However, in this case, the difference was smaller. The identification responses of the Spanish learner groups overlapped along part of the continuum (see Figure 5.6). SPINEXPinSP heard /k/ 22% of time with 22 ms of VOT, 64% with 25 ms and reached a score of 100% with 37 ms of VOT. As Table 5.8 shows, SPINEXPinSP's Spanish /k/-/g/ category boundary presented a value of 20.6 ms. As for SPEXPinSP, they heard /k/ 15% of time with 22 ms, 73% with 27 ms and 93% with 32 ms of VOT. Their /k/-/g/ category boundary was located at 20.8 ms. Regarding SPEXPinUK, they identified /k/

19% of trials with 22 ms, 59% with 27 ms, 84% with 32 ms and 94% with 37 ms of VOT. The category boundary calculated for SPEXPinUK was 21.3 ms.

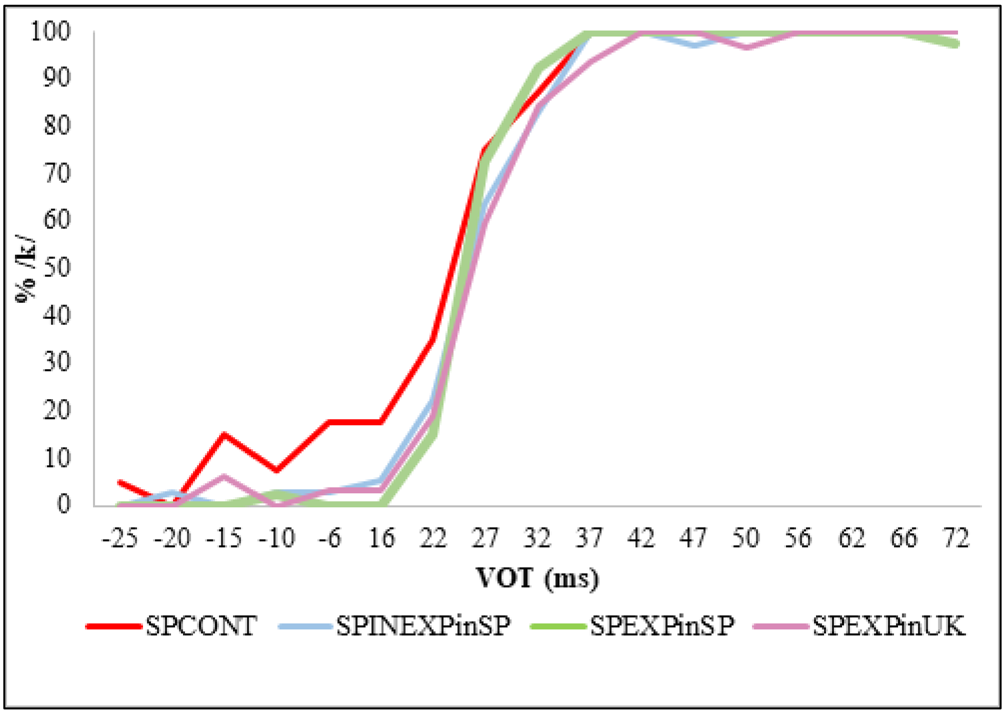

Figure 5.6. L1-Spanish L2-English groups' % identification function for /k/ in Spanish.

Group	*M*	*SD*	Range
SPCONT	14.8	10.3	0 – 27.3
SPINEXPinSP	20.6	4.2	12.4 – 30.1
SPEXPinSP	20.8	2.2	17.8 – 24.5
SPEXPinUK	21.3	3.6	14 – 26.3

Table 5.8. L1-Spanish L2-English groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the Spanish /k/-/g/ contrast in ms.

Just as in the case of /p/-/b/, the one-way ANOVA revealed a significant effect of group on /k/-/g/ category boundary [F (3, 34) = 3.861; p = .018]. The post-hoc Bonferroni tests indicated a significant difference between SPCONT and SPEXPinUK (p = 0.41), whereas all other comparisons did not reach significance (p > .05). It should also be noted that, although the comparison between SPEXPinSP and SPCONT was not significant, the difference reached marginal significance (p = .054). In line with the /p/-/b/ findings – but unlike the English native speakers –, this result suggests that L2 experience may modulate L2 influence on the L1, as only the most experienced group differed significantly from controls and the moderately experienced group nearly did so, whereas that was not the case for the inexperienced group.

In short, SPEXPinUK presented significantly later category boundaries for both contrasts than SPCONT, whereas the SPEXPinSP and SPINEXPinSP did not differ significantly from Spanish controls. Thus, results indicate that L2 experience – and possibly language setting has an influence on the L1 categorization of Spanish learners of English.

5.3.2.2. Perception of Spanish stops: English learners of Spanish

L1-English L2-Spanish speakers needed more aspiration than Spanish monolinguals to start perceiving /p/ (see Figure 5.7). Their category boundary, located at 14.9 ms, was the closest one to that of SPCONT. Regarding ENEXPinSP, their crossover area ranged from the 14 ms to the 19 ms steps, as stimuli were identified as /p/ 31% and 42% of the trials, respectively. /p/ reached a score of 89% with 24 ms and of 100% with 29 ms. As for category boundary, ENEXPinSP presented the closest one to SPCONT with 14.9 ms (see Table 5.9 for the mean category boundaries of all groups). ENEXPinUK identified /p/ 40% at the 19 ms step, 73% at the 24 ms step and 95% at the 29 ms step. Their category boundary was the furthest from that of SPCONT with 16.5 ms. As for ENEXPinUK, they heard /p/ 20% of time with 14 ms, hesitated the most at the 19 ms and 24 ms steps, obtaining a score of 57% and 73%, respectively, and reached a 95% identification score with 29 ms. Their /p/-/b/ category boundary was located at 15.7 ms.

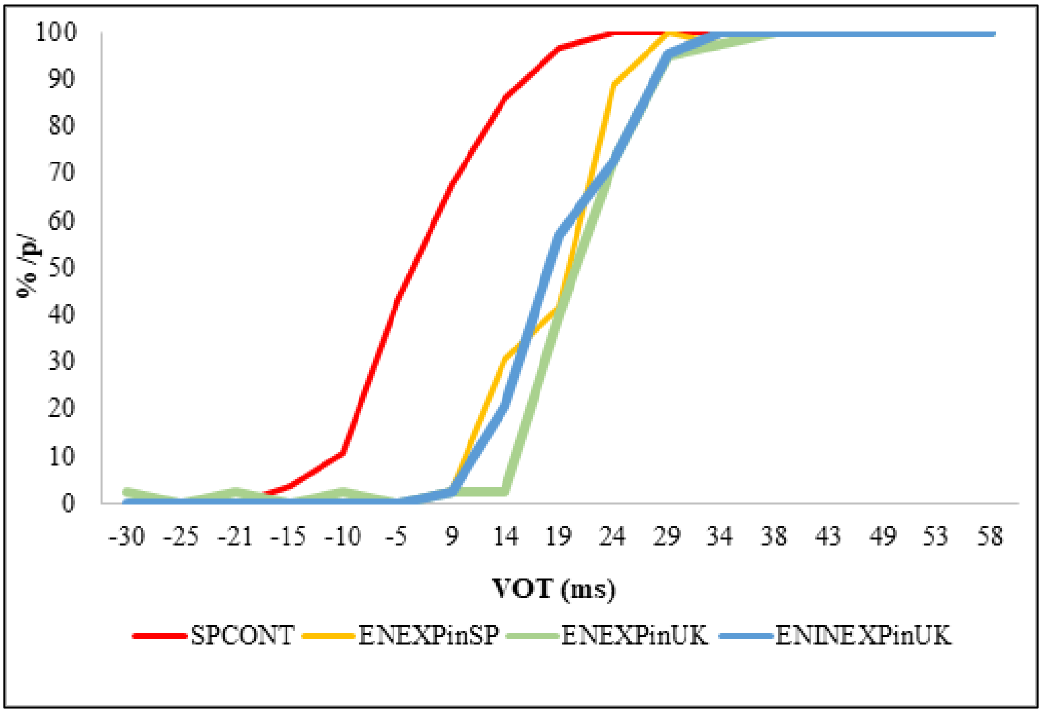

Figure 5.7. L1-English L2-Spanish groups' and SPCONT's % identification functions for /p/ in Spanish.

Group	*M*	*SD*	Range
SPCONT	3.4	5.4	-7.9 – 9.4
ENEXPinSP	14.9	3.1	9.7 – 19.1
ENEXPinUK	16.5	4.7	7.7 – 22.8
ENINEXPinUK	15.7	4.1	9.7 – 23.7

Table 5.9. L1-English L2-Spanish groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the Spanish /p/-/b/ contrast in ms.

The one-way ANOVA revealed a significant effect of group on /p/-/b/ category boundary [F (3, 37) = 19.828; p = .000]. Post-hoc Bonferroni tests showed that all testing groups differed significantly from SPCONT (p = .000 in all cases). In other words, L2 experience did not significantly influence the L2 perception of the /p/-/b/ contrast, as all English groups presented significantly greater category boundaries than SPCONT.

Just as in the case of /p/-/b/ identification, L1-English L2-Spanish speakers needed a greater VOT value to hear /k/ than SPCONT (See Figure 5.8). ENEXPinSP presented the most advanced as well as the steepest category curve out of the English groups. They identified /k/ 19% at the 22 ms step, 59% at the 27 ms step, 84% at the 32 ms step and reached 94% with 37 ms of VOT. As Table 5.10 illustrates, their category boundary was the latest, with 28 ms of VOT. ENEXPinUK heard /k/ 20% of the trials with 22 ms, 33% with 27 ms, 43% with 32 ms and 70% with 37 ms, reaching 95% with 42 ms of VOT. Their category boundary was located at 25.6 ms. ENINEXPinUK was the L1-English group that presented the shallowest categorization curve. They perceived /k/ 20% of time with 22 ms, 45% with 32 ms and 70% with 37 ms, 80% with 42 ms and 92% with 47 ms of VOT. Their category boundary presented 25.2 ms – i.e., numerically the closest one to SPCONT.

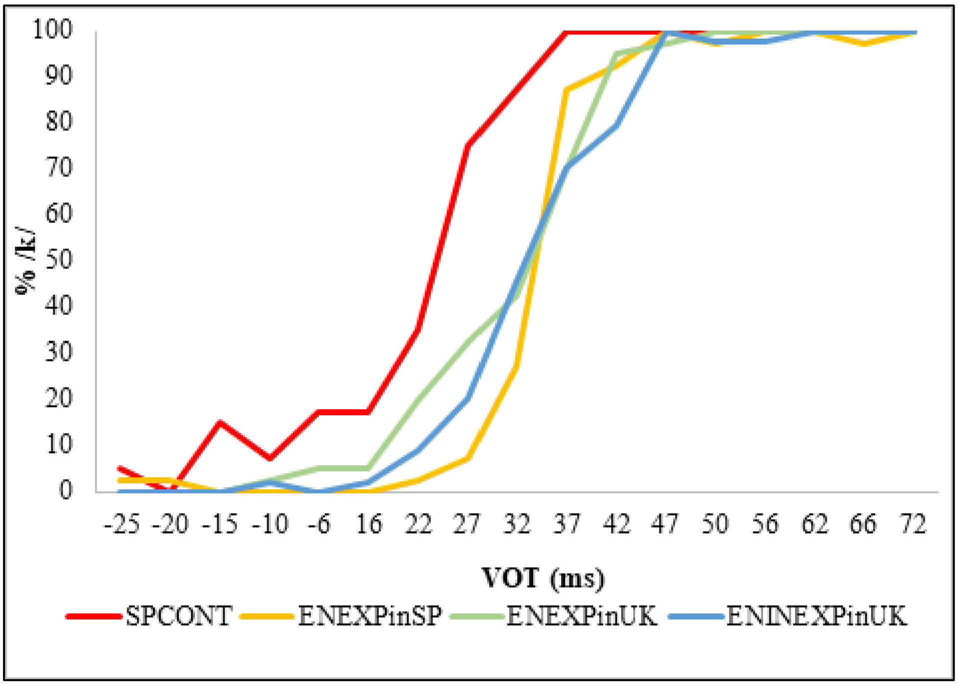

Figure 5.8. L1-English L2-Spanish groups' and SPCONT's % identification curves for /k/ in Spanish.

Group	*M*	*SD*	Range
SPCONT	14.8	10.3	0 – 27.3
ENEXPinSP	28	2.8	22.7 – 33.1
ENEXPinUK	25.6	5.2	17.7 – 33.1
ENINEXPinUK	25.2	8.3	7.7 – 36.2

Table 5.10. L1-English L2-Spanish groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the Spanish /k/-/g/ contrast in ms.

The one-way ANOVA yielded a significant effect of group on /k/-/g/ category boundary [F (3, 37) = 8.282; p = .000]. Bonferroni post-hoc tests found that, again, all L1-English groups differed significantly from SPCONT (SPCONT and ENEXPINSP: p = .004; SPCONT and ENEXPINUK p = .003; ENCONT and ENINEXPINUK p = .000). Thus, just as in the case of /p/-/b/ perception, L2 experience was not found to have an effect on L1 perception, as all groups presented significantly greater category boundaries than Spanish monolinguals. In short, all L1-English L2-Spanish groups – regardless of L2 experience – were found to differ significantly from SPCONT.

5.3.3. The relationship between /p/-/b/ and /k/-/g/ perception

In order to examine the degree to which participants perceived the voicing contrast similarly in bilabial and velar stops, a number of correlation analyses were carried out. Thus, a total of six one-tailed Pearson correlation tests were conducted – i.e., one for each control group in their native language and one for all learner groups sharing the same L1 (L1-English L2-Spanish speakers and L1-Spanish L2-English speakers) in each language – in order to examine the relationship between the two places of articulation. In other words, it was investigated whether participants that presented late /p/-/b/ category boundaries also presented late /k/-/g/ category boundaries and the reverse. Regarding English perception, the test revealed a significant correlation for ENCONT (r = .741;

N = 9; *p* = .011), as well as for the English learners of Spanish (*r* = .498; *N* = 32; *p* = .002). In the case of the Spanish groups, no significant correlation was revealed in English (*r* = .279; *N* = 29; *p* = .071).

As for Spanish perception, SPCONT's, as well as the Spanish testing groups', /p/-/b/ and /k/-/g/ category boundaries were significantly correlated (SPCONT: *r* = .592; *N* = 10; *p* = .036; L1-Spanish L2-English groups: *r* = .436; *N* = 28; *p* = .010). Conversely, no significant correlations were found regarding the English learners of Spanish (*r* = .048; *N* = 31; *p* = .400). All in all, results indicated that functional monolinguals and L2 learners tended to perform similarly across bilabial and velar stop perception in terms of the amount of VOT needed for identification as voiceless in their L1, whereas no clear relationship was found in their L2.

5.3.4. Between groups perception: results summary

The perception of the bilabial and velar stop voicing contrasts in English and in Spanish was examined by means of a series of identification tests – one for each place of articulation and language. The performance of L1-Spanish and L1-English groups differing in L2 experience was compared to that of functional monolinguals of each language. Table 5.11 below summarizes all the significant differences found in both languages. In terms of the identification of English stops, results showed that English learners of Spanish presented slightly later category boundaries than English controls, although this difference did not reach significance in any case. That is, English learners of Spanish presented category boundaries that were comparable to those of English monolinguals. Regarding the Spanish learners of English, a significant effect of group was found in the perception of English stops by Spanish learners of English. Only SPINEXPinSP presented a significantly earlier – i.e., Spanish-like – category boundary than ENCONT, regarding both /p/-/b/ and /k/-/g/ category boundary, whereas the two experienced groups presented comparable category boundaries to English monolinguals – although numerically earlier.

In the case of Spanish, all English learners – regardless of their L2 experience – presented significantly later – i.e., more English-like – category boundaries than Spanish controls. As for the perception of Spanish stops by the Spanish groups, SPEXPinUK, the most L2 experienced group, obtained significantly later /p/-/b/ and /k/-/g/ category boundaries than Spanish controls, whereas the groups residing in an L1 context did not. It is also worth mentioning that, even though this difference only reached significance in the case of SPEXPinUK, all English learners of Spanish needed longer VOT values to perceive voiceless stops than Spanish controls.

Finally, regarding the link between /p/-/b/ and /k/-/g/ category boundaries, a relationship was established in the L1 – i.e., late /p/-/b/ boundaries were correlated to late /k/-/g/– for both L1-English and L1-Spanish learner groups as well as for the Spanish controls, whereas that was not the case in the L2 for any group. The following section will present the results obtained in the analysis carried out to investigate the differences between L1 and L2 stop perception and the effect of L2 experience on that difference.

Group comparisons	Language	/p/-/b/	/k/-/g/
ENCONT vs. ENINEXPinUK	English	14 vs. 14.5	29.7 vs. 27.3
ENCONT vs. ENEXPinUK	English	14 vs. 15.9	29.7 vs. 27.2
ENCONT vs. ENEXPinSP	English	14 vs. 16.1	29.7 vs. 29.4
ENCONT vs. SPINEXPinSP	English	14 vs. 9.3 *	29.7 vs. 22.8 **
ENCONT vs. SPEXPinSP	English	14 vs. 10.1	29.7 vs. 24.5
ENCONT vs. SPEXPinUK	English	14 vs. 12.1	29.7 vs. 24.9
SPCONT vs. SPINEXPinSP	Spanish	3.4 vs. 8.4	14.8 vs. 20.6
SPCONT vs. SPEXPinSP	Spanish	3.4 vs. 7.8	14.8 vs. 20.8
SPCONT vs. SPEXPinUK	Spanish	3.4 vs. 11.2**	14.8 vs. 21.3*
SPCONT vs. ENINEXPinUK	Spanish	3.4 vs. 15.7 **	14.8 vs. 28 **
SPCONT vs. ENEXPinUK	Spanish	3.4 vs. 16.5 **	14.8 vs. 25.6 **
SPCONT vs. ENEXPinSP	Spanish	3.4 vs. 14.9 **	14.8 vs. 25.2 **

Table 5.11. Comparisons between group means for the /p/-/b/ and /k/-/g/ category boundaries (* indicates .05 a significant difference at a .05 level and ** at a .01 level).

5.4. Between-languages perception analysis

This section will present the results obtained in the within-group between-languages analysis, that is, the results of comparing the performance of each of the groups in the two languages – i.e., Spanish vs. English.

5.4.1. L1 and L2 perception by English learners of Spanish

This section compares the performance of the English learners of Spanish in the identification tests in their L1 and L2. All groups performed similarly in the two languages, both in the /p/-/b/ and /k/-/g/ tests, especially ENINEXPinUK, whose categorization curves in Spanish and in English perception overlapped across most of the steps in the continuum (see Figures 5.9, 5.10 and 5.11). English learners presented slightly later categorization curves and category boundaries in English than in Spanish, except in the case of ENINEXPinUK's /p/-/b/ perception, as they obtained a slightly later boundary in the L2 than in the L1 (see Tables 5.12 and 5.13). A two-way ANOVA – with boundary as the dependent variable and language and group as the independent variables – was conducted for each place of articulation and failed to find a significant effect of language [/p/-/b/: F (1) =.029 ; p = .865; /k/-/g/: F (1) = 1.128; p = .293] or group [/p/-/b/: F (2) =.489 ; p = .616; /k/-/g/: F (2) = 1.021; p = 367]. No significant interaction between language and group was found [/p/-/b/: F (2) =.645; p = .528; /k/-/g/: F (2) = .031; p = 970]. These results suggest that English learners of Spanish perceived the /p/-/b/ and /k/-/g/ contrast similarly in their L1 and L2. As a matter of fact, given that none of the L1-English L2-Spanish groups differed from ENCONT in English but all did so from SPCONT in Spanish – as reported in the previous section –, they seemed to be using their L1 categories in both languages.

Group	**English**		**Spanish**		**Sig.**
	M	***SD***	***M***	***SD***	
ENCONT	14	2.5	-	-	-
ENINEXPinUK	14.5	3.1	15.7	4.1	$p > .05$
ENEXPinUK	15.9	3.9	16.5	4.7	$p > .05$
ENEXPinSP	16.1	2.4	14.9	3.1	$p > .05$
SPCONT	-	-	3.4	5.4	-

Table 5.12. L1-English groups' and control groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English and Spanish /p/-/b/ contrast.

Group	**English**		**Spanish**		**Sig.**
	M	***SD***	***M***	***SD***	
ENCONT	29.7	4.3	-	-	-
ENINEXPinUK	27.3	8.5	25.2	8.3	$p > .05$
ENEXPinUK	27.2	4.5	25.6	5.2	$p > .05$
ENEXPinSP	29.4	3.7	28	2.8	$p > .05$
SPCONT	-	-	14.8	10.3	-

Table 5.13 L1-English groups' and control groups' mean perceptual boundary (*M*), standard deviation (*SD*) and range for the English and Spanish /k/-/g/ contrast.

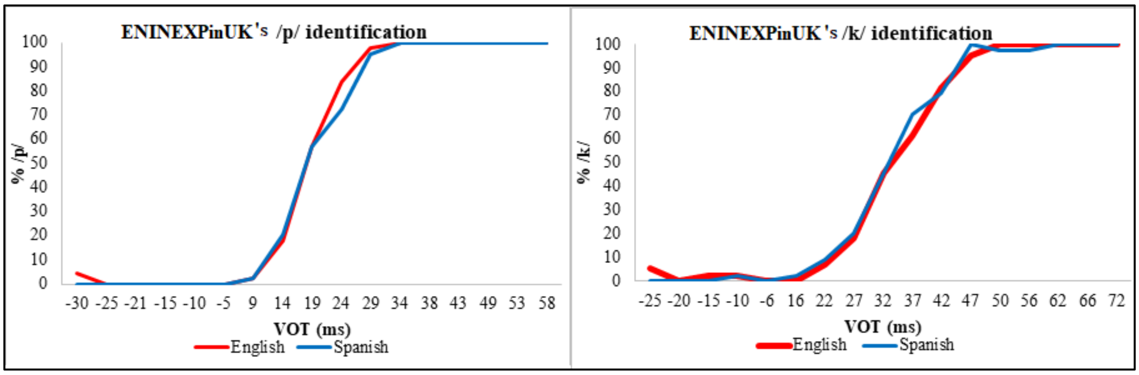

Figure 5.9. ENINEXPinUK's % identification functions for /p/ and /k/ in English (shown in blue) and in Spanish (in red).

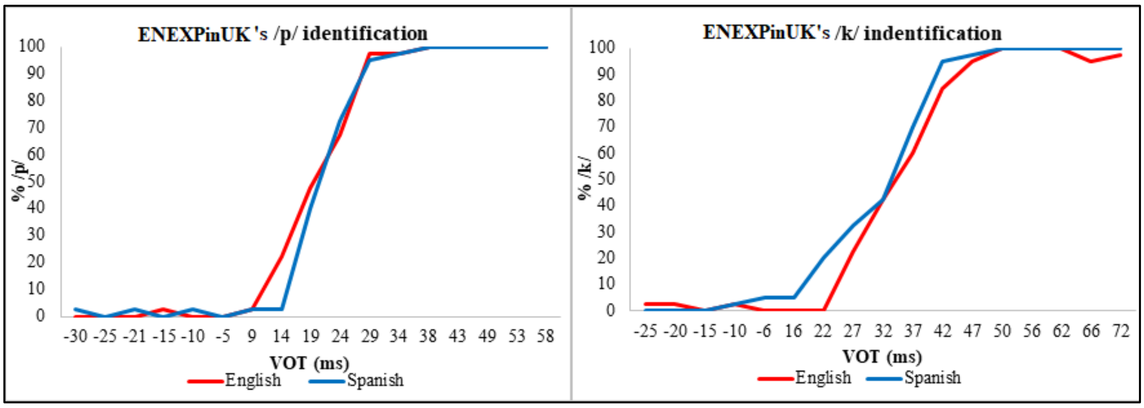

Figure 5.10. ENEXPinUK's % identification functions for /p/ and /k/ in English (shown in blue) and in Spanish (and in red).

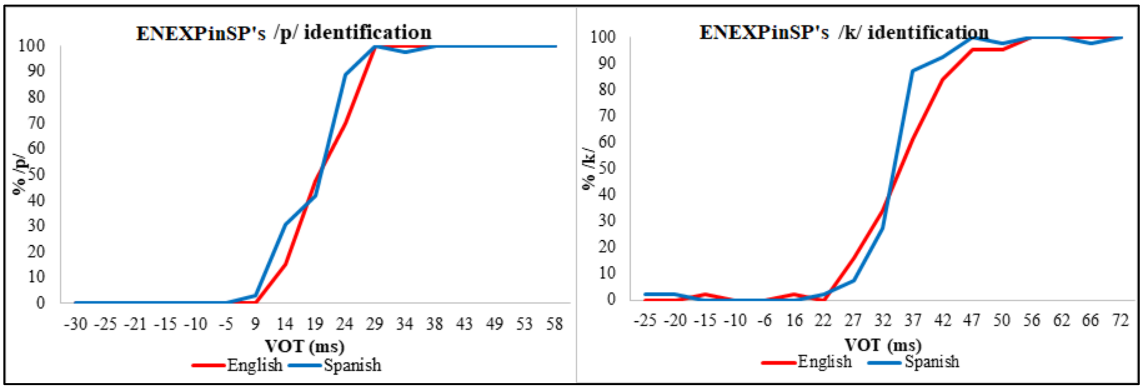

Figure 5.11. ENEXPinSP's identification functions for /p/ and /k/ in English (shown in blue) and in Spanish (in red).

5.4.2. L1 and L2 perception by Spanish learners of English

Figures 5.12, 5.13 and 5.14 illustrate the identification functions obtained by the Spanish learner groups in English and in Spanish. Regarding category boundary, all groups presented values between ENCONT and SPCONT in the two languages (see Table 5.14). Although the differences between the English and the Spanish boundaries were generally numerically small, they were in the expected direction, that is, Spanish learners of English identified voiceless stops in the Spanish tests earlier in the continuum than in the English tests. SPEXPinSP was the group that presented the greatest numerical difference between the two languages, both regarding /p/-/b/ and /k/-/g/.

When it comes to the /p/-/b/ contrast, SPINEXPinSP and SPEXPinSP presented intermediate category boundaries between SPCONT's and ENCONT's values, whereas SPEXPinUK presented values closer to ENCONT in both languages. The two-way ANOVA failed to find a significant effect of language [F (1) = 2.336; p = .132], but there was a significant effect of group [F (2) = 4.047; p = .023]. The group effect was motivated by the later category boundaries obtained by SPEXPinUK. In spite of the fact that the Spanish learning groups were not found to differ significantly in the within-language analysis, the post-hoc Bonferroni tests in the current analysis revealed a significant difference between SPEXPinUK and SPINEXPinSP (p = .044) and a marginal result between SPEXPinUK and SPEXPinSP (p = .054), as the former presented later category boundaries than the two groups tested in Spain. No interaction between language and group was found [F (2) = .267; p = .767].

Group	English boundary		Spanish boundary	
	M	***SD***	***M***	***SD***
SPCONT	-	-	3.4	5.4
SPINEXPinSP	9.3	3.6	8.4	3.2
SPEXPinSP	10.1	4.1	7.8	3.2
SPEXPinUK	12.1	2.6	11.2	3.5
ENCONT	14	2.5	-	-

Table 5.14. L1-Spanish groups' and English control's /p/-/b/ mean category boundaries (*M*) and standard deviations (*SD*) in English and in Spanish.

As for /k/-/g/, all Spanish groups presented values between SPCONT and ENCONT, which, in all cases, were closer to those of SPCONT than to those of ENCONT (see Table 5.15). The two-way ANOVA showed a significant effect of language on /k/-/g/ category boundary [F (1) = 13.648; p = .001] but failed to find a significant effect of group [F (2) = .342; p = .712]. No significant interaction between language and group was found [F (2) = .128; p = .881]. In order to examine which group presented a greater difference between the two languages, a paired-samples t-test was conducted on each group (dependent variables: English /k/-/g/ boundary, Spanish /k/-/g/ boundary). The results showed that the difference was significant only in the case of SPEXPinSP, whereas results were marginally significant for SPINEXPinSP and nonsignificant for SPEXPinUK [SPINEXPinSP: t(8) = 2.175 , p = .061; SPEXPinSP: t(9) = 3.039 , p = .014; SPEXPinUK: t(8) = 1.815 , p = .107]. This finding suggests that the significant result revealed in the analysis including all groups might have been motivated mostly by the performance of SPEXPinSP in each language. It should be noted, however, that sample size was smaller in the paired-samples t-test than in the two-way ANOVA – i.e., 9-10 participants vs. 28 – which may have hindered reaching more significant results.

Overall, it seems that Spanish learners of English have a shared category boundary for English and Spanish /p/-/b/, which may be more Spanish-like or more English-like as a function of their L2 experience – i.e., the more L2 experience, the more L2-like.

Regarding /k/-/g/, Spanish learners had somewhat more distinct L1 and L2 /k/-/g/ category boundaries than in the case of /p/-/b/, particularly regarding SPEXPinSP, who presented a significantly later category boundary in English than in Spanish – i.e., separate L1 and L2 categories – despite not being the most experienced of the three groups.

Group	English boundary		Spanish boundary	
	M	***SD***	***M***	***SD***
SPCONT	-	-	14.8	10.3
SPINEXPinSP	23.6[4]	3.9	20.6	5
SPEXPinSP	24.5	3.9	20.8	2.2
SPEXPinUK	24.9	2.4	21.3	3.6
ENCONT	29.7	4.3	-	-

Table 5.15. L1-Spanish groups' and English control's /k/-/g/ mean category boundaries (*M*) and standard deviations (*SD*) in English and in Spanish.

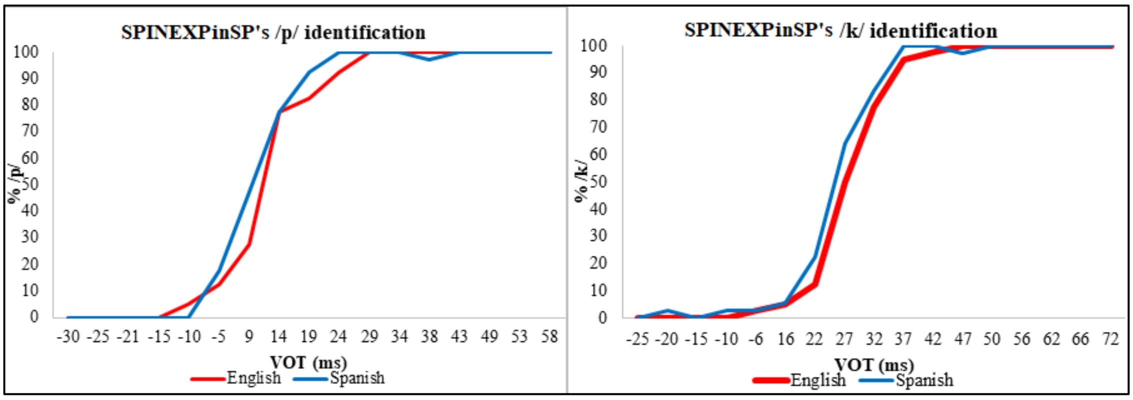

Figure 5.12. SPEXPINinSP's % identification functions for /p/ and /k/ in English (shown in red) and in Spanish (in blue).

[4] Given that SPINEXPinSP11's data in the Spanish /k/-/g/ test was lost, their category boundary in English has also been removed from the between-languages analysis, hence the difference in the mean obtained by SPINEXPinSP in English in Tables 5.11 and 5.15.

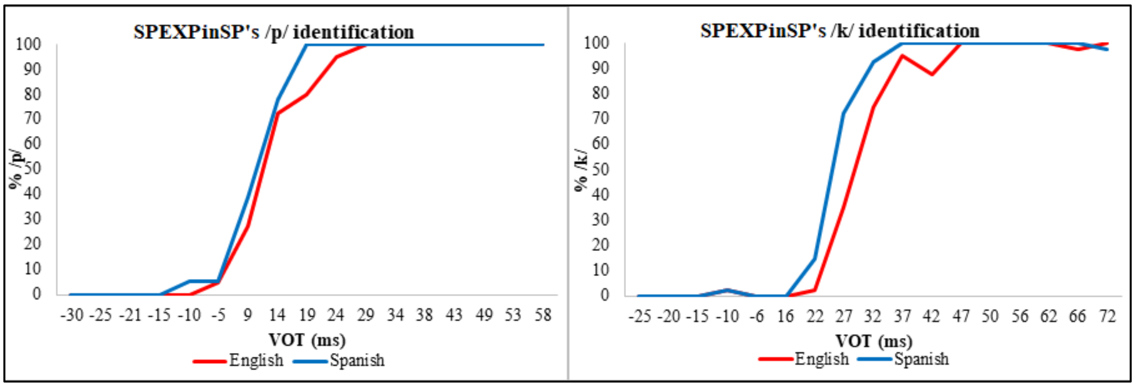

Figure 5.13. SPEXPinSP's % identification functions for /p/ and /k/ in English (shown in red) and in Spanish (in blue).

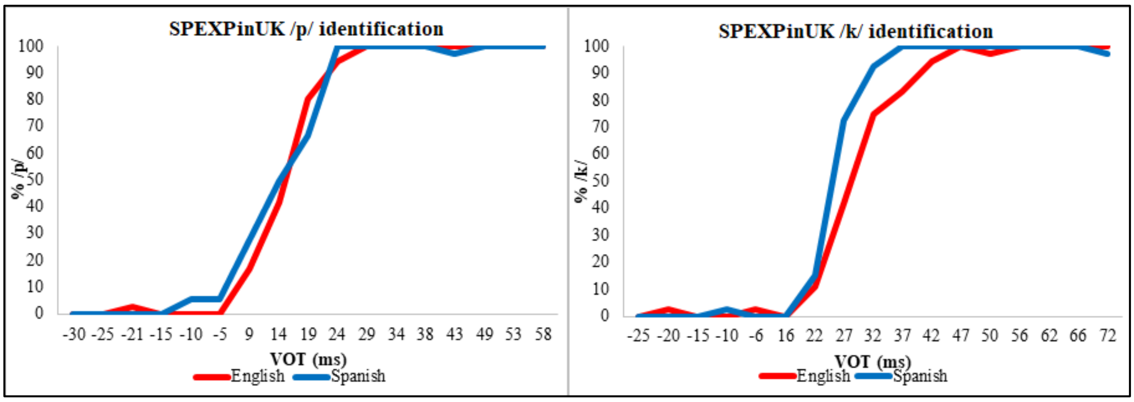

Figure 5.14. SPEXPinUK's identification functions for /p/ and /k/ in English (shown in red) and in Spanish (in blue).

5.4.3. Between languages perception: results summary

The category boundaries obtained by each group in the two languages were compared by conducting a two-way ANOVA for each place of articulation and L1. Results revealed that all L1-English L2-Spanish groups performed similarly – i.e., using English-like values – in the two languages, both regarding /p/-/b/ and /k/-/g/, as no significant differences were found.

Regarding the L1-Spanish participants, they presented category boundaries with numerically intermediate values between Spanish and English controls, which were increasingly more English-like as they gained experience in the case of /p/-/b/– i.e., SPINEXPinSP presented the most Spanish-like boundaries, whereas SPEXPinUK obtained the most English-like ones. However, no significant effect of language was found in the case of/ p/-/b/, indicating that all groups had a single category for both languages. As for /k/-/g/, an effect of language was found, as Spanish speakers presented later category boundaries in English than in Spanish. The paired t-tests revealed a significant difference only regarding SPEXPinSP. Next, the results obtained in the two analyses conducted on the perception experiment – i.e., the between-groups within-language analysis and the between-languages analysis – will be discussed.

5.5. Perception experiment: Discussion

Two analyses have been conducted in order to analyze the perception results. First, separate analyses conducted on each language – English and Spanish – compared the performance of the learner groups with the same L1 with each other and with the control groups. Regarding the categorization of stops in the L1, all English groups were found to perceive English-stops in a native-like manner. By contrast, when it comes to the Spanish learners of English, the most experienced group (SPEXPinUK) presented significantly later category boundaries than Spanish controls, indicating an influence of English on their L1 perception. In the case of L2 perception, all English learner groups were found to present significantly later – i.e., English-like – category boundaries than the Spanish controls. Regarding the Spanish groups, only the inexperienced group was found to present significantly earlier – i.e., Spanish-like – category boundaries than ENCONT, whereas the two experienced groups did not, suggesting a positive effect of L2 experience on L2 stop categorization. The second set of analyses compared the performance of each group on the two languages in order to examine whether L2 learners categorized L1 and L2 stops using the same category boundaries or whether they presented language-specific category boundaries. All English groups performed similarly in the English and the Spanish condition, indicating that they categorized both L1 and L2 stops using the same category boundary. As for the Spanish learners of English, all groups presented somewhat earlier category boundaries in Spanish than in English – as expected – but this difference only reached significance in the case of the /k/-/g/ category boundary. A separate analysis conducted on each group showed that only the moderately experienced learners in Spain (SPEXPinSP) presented significantly different category boundaries for Spanish and English /k/-/g/. Next, the results obtained in the two analyses will be discussed by population, beginning with the L1-English groups, followed by the Spanish learners of English.

Regarding the English learners of Spanish, the fact that there were no differences in their perception of the two languages, along with the fact that the English groups differed from SPCONT in the L2 but not from ENCONT in the L1, supports the assumption that the L2 contrast was mapped onto the L1 contrast, probably resulting in a single shared category for each stop (Flege, 1995, 2002, 2007). In other words, the English groups presented L1-like perceptual categories in both languages, indicating a unidirectional L1 influence on the L2. Similar results were reported by Hazan and Boulakia (1993), as the French-dominant bilingual speakers of English in that study perceived the /p/-/b/ contrast in English in the same manner as French monolinguals perceived French stops. As explained in the methodology section, all L1-English groups were found to be clearly L1 dominant. Given that all groups presented L1-like category boundaries, it appears that no effect of L2 experience was found on the L2 in line with some previous studies (e.g., Bohn & Flege, 1990; Cebrian, 2006). For example, Cebrian (2006) failed to find a positive effect of L2 experience on L2 perception, as Catalan speakers in Toronto did not outperform those living in Catalonia when it comes to the identification of L2 vowels. Regarding L1 perception, no effect of L2 experience was found either, given that, as mentioned above, the boundaries in the two languages were very similar. This result opposes previous findings on L1 perception by L2 speakers (e.g., Cebrian, 2006; Dmitrieva, 2019; Major, 2010), as well as the results obtained by the Spanish groups, as will be discussed below.

The Spanish groups, unlike the English speakers, did not present category boundaries with values that were very similar to those of the L1 controls in either language, but presented intermediate values – between those obtained by the English and the Spanish controls – in both the L1 and the L2. In the case of the L2, all Spanish groups had numerically earlier – more Spanish-like – category boundaries than ENCONT. However, this difference only reached significance in the case of SPINEXPinSP, regarding both /p/-/b/ and /k/-/g/. Given that only the inexperienced group presented significantly earlier category boundaries than the English monolinguals, it appears that L2 experience had a positive effect on L2 stop perception for the L2 English learners. Previous studies have also reported that L2-experience modulated L2 perception accuracy (Flege et al., 1997; Levy, 2009). Flege et al. (1997) found that highly experienced learners of English with different linguistic backgrounds – including L1 Spanish – discriminated /ɛ/-/æ/ and /i/-/ɪ/ more accurately than English learners with the same L1 that had recently arrived in the US and, thus, had less L2 experience. As for the perception of L1 stops, differences between the Spanish groups were also observed. The most L2 experienced group presented greater – and more deviant from SPCONT – category boundaries. Statistical results showed that, whereas SPEXPinUK differed significantly from SPCONT, SPINEXPinSP did not and SPEXPinSP differed only marginally. Similar findings have been reported in the literature (Major, 2010; Cebrian, 2006). For instance, Cebrian (2006) found that L1-Catalan L2-English speakers living in an L1 setting outperformed Catalans living in Toronto in the identification of L1 vowels, showing evidence that prolonged experience with the L2 may result in a deterioration of L1 perceptual abilities.

Regarding the differences between English and Spanish category boundaries, all Spanish groups presented numerically later category boundaries – i.e., more English-like – in their L2 than in their L1, especially in the case of SPEXPinSP, who obtained the most different category boundaries in the two languages. The statistical analysis failed to find a significant effect of language on /p/-/b/ category boundary but did so regarding /k/-/g/. Results showed that L1-Spanish speakers presented significantly later /k/-/g/ category boundaries in English than in Spanish. An analysis conducted on each group indicated that SPEXPinSP was the group that perceived /k/-/g/ the most differently in the two languages. This difference between the two places of articulation may be related to salience. Given that /k/ presents longer VOT values than /p/ (Lisker & Abramson, 1964), Spanish learners of English may have been more successful at perceiving a difference between English and Spanish /k/ than between English and Spanish /p/. The nature of L2 experience may also have had an effect on the ability to distinguish Spanish and English stops, as SPEXPinSP perceived L1 and L2 contrasts more differently than the other Spanish learner groups – i.e., SPEXPinUK and SPINEXPinSP. Having had some experience in an L2 setting may have helped SPEXPinSP adjust the perception of L2 stops to L2-like values. As a matter of fact, as stated above, their category boundaries did not differ significantly from those of ENCONT. However, their amount of L2 experience may not have been sufficient to change their L1 perception since their Spanish category boundaries did not differ significantly from those of SPCONT – as opposed to SPEXPinUK. It should also be noted that they were tested in an L1 setting and that their

L2 immersion finished a few months before the time of testing. As Sancier and Folwer (1997) found for production, recency of L2 experience may have an impact on L1 and L2 performance. However, considering the small sample size – only about 10 data points were used per group –, the fact that SPEXPinSP presented significantly different L1 and L2 category boundaries for /k/-/g/ should be interpreted with caution, especially considering the fact that the same result was not replicated in the case of /p/-/b/.

Overall, results suggest that L2 and L1 stop perception by Spanish learners of English underwent – to a certain extent – a bidirectional crosslinguistic influence given that all groups presented numerically intermediate category boundaries between Spanish controls and English controls in both languages, in line with Williams's (1977) results for Spanish-English bilinguals. However, it should be noted that the amount of influence on stop perception by each language seemed to be modulated by L2 experience. On the one hand, SPEXPinUK presented category boundaries with VOT values that were closer to the L2 – that is, they presented a greater influence from the L2 than from the L1 –, their boundaries did not differ from ENCONT in English but did so from SPCONT in Spanish. On the other hand, SPINEXPinSP's category boundaries were more L1-like, as they did not differ significantly from SPCONT in Spanish but did so from ENCONT in L2 perception. Conversely, SPEXPinSP did not differ from monolinguals either in the L2 or in the L1. Thus, it appears that having no experience in an L2 setting results in a notable influence of the L1 on the L2, whereas a great amount of L2 experience and living in an L2 setting brings about an influence of the L2 on L1 perception. Moreover, a moderate past L2 experience does not considerably influence L1 perception but is sufficient to achieve a target-like L2 perception.

A difference between L1-English L2-Spanish and L1-Spanish L2-English speakers was observed regarding the influence of L2 experience on the L2 and on the L1, as well as on the nature of their categories – i.e., L1-like vs. intermediate, respectively. Whereas the L1-English speakers did not present any effect of L2 experience on their L1 and L2 perception of stops, the opposite was true for the Spanish speakers. That is, in the case of the Spanish learners of English, L2 experience was found to result in more target-like category boundaries in the L2 but less native-like category boundaries in the L1. Moreover, the English learners of Spanish presented a single shared category with L1-like values, whereas all Spanish learners of English presented numerically intermediate categories in both languages, which were found to be separate in the case of the velar contrast for SPEXPinSP. These differences between the two populations may stem from a number of reasons.

First of all, the fact that English learners of Spanish perceived Spanish stops with English-like values may be related to language mode activation. Casillas and Simonet (2018) pointed out the importance of presenting target-like sounding stimuli in order to successfully control for language mode. However, in the present thesis, the same stimuli were used to test both languages, as the aim of the study was to test two populations – English learners of Spanish and Spanish learners of English – whose L1s make a different use of VOT to contrast voicing in their two languages. Thus, the same VOT range, which was appropriate for the two L1s, was used in the two conditions – i.e., in the English and in the Spanish tasks. The presence of aspirated tokens – a feature that is not used in Spanish – in both tests may have favoured the activation of English mode over Spanish mode. As a matter of fact, Bohn and Flege (1993) found that stop identification as voiced-voiceless by Spanish-English bilinguals, as well as by Spanish and English monolinguals, depended on the set of stimuli presented. Short-lag /t/ was identified as /d/ more often when it was presented along with long-lag /t/ than when it was presented with short-lag /d/. Still, the fact that all Spanish groups made a small difference between the two languages in the expected direction – which in fact reached significance for SPEXPinSP regarding /k/-/g/ – suggests that the strategy followed to control language mode (see Chapter 4 on ‘Methodology’) may have been successful – at least to a certain extent.

Another plausible explanation for the difference between the L1-English and the L1-Spanish groups may have to do with the VOT range used in each language. Spanish VOT ranges from voice-lead (voiced stops) to short-lag VOT (voiceless stops). English VOT typically ranges from short-lag (voiced stops) to long-lag VOT (voiceless stops). However, voice-lead VOT is also used to produce voiced stops by some English native speakers (Lisker & Abramson, 1964). Given that Spanish learners of English are acquiring long-lag VOT – i.e., their VOT range is expanding towards greater VOT values –, their category boundaries may be moving forward. Conversely, English covers a wider range of VOT than Spanish, which includes the three VOT conditions – i.e., voice-lead, short-lag and long-lag. Therefore, English learners of Spanish do not need to learn a new feature – even though they do need to learn how the VOT conditions are used in Spanish – and, as a result, they may not move their category boundaries towards Spanish values. Furthermore, the fact that long-lag VOT is a new feature for Spanish learners may facilitate the creation of new L2 categories, as was found for SPEXPinUK, whereas English speakers may simply transfer the use of VOT to contrast voicing from the L1 to the L2.

As mentioned at the beginning of this chapter, order of completion of the experiment seemed to affect the L1 perception of the Spanish groups. However, no order effect was observed in the English groups. The Spanish participants who completed the experiment in English first presented later boundaries in Spanish than those who were tested on their L1 in the first place. That is, it appears that those Spanish learners who completed the experiment in the L2 prior to the L1 presented L2 transfer on their L1 VOT perception. Following up on the discussion in the previous paragraph, given that the same stimuli were used in both tests – including non-Spanish aspirated stops –, it is possible that Spanish participants perceived stops in a more English-like manner in Spanish when the English task preceded the Spanish task due to a combination of task familiarity and a somewhat greater difficulty to control language mode. The fact that order of completion influenced Spanish learners of English, but not English learners of Spanish, may have also contributed to the differences observed in their results.

Finally, as pointed out in the methodology, the L1-English and the L1-Spanish groups differ in certain aspects, such as age of L2 acquisition, years of L2 instruction, amount of weekly L2 use and language dominance. Overall, the Spanish groups started learning their L2 earlier than the English groups and, as a result, had learnt it for a longer period of time. Furthermore, the L1 Spanish groups used their L2 more frequently than the English learners of Spanish and the English groups presented more extreme L1-dominance scores than the Spanish learners, who, even though were also L1-dominant, obtained dominance scores closer to linguistic balance. The fact that Spanish learners of English had overall more contact with their L2 may have provided them with an advantage over the English groups when it comes to L2 performance. These factors will be examined in Chapter 8.

Summary

In summary, the results of the perception experiment show that English learners of Spanish presented an L1-like category boundary that is shared in the two languages. No evidence of L2 experience on their L2 or on their L1 was found. Conversely, Spanish learners of English had category boundaries with intermediate values between ENCONT and SPCONT. There was a small difference in the category boundaries of the two languages in the expected direction – i.e., the English category boundaries had greater values than the Spanish boundaries – which only turned out to be significant in the case of the velar contrast regarding the moderately experienced group (SPEXPinSP). As for the effect of L2 experience, the factor was found to modulate both L2 and L1 perception of stops regarding the Spanish groups, but not in the case of the English groups. Only the least experienced learners were found to present significantly earlier category boundaries in their L2 than the English monolinguals, whereas the most experienced learners had significantly later category boundaries in their L1 than the Spanish monolinguals. The following chapter will present and discuss the results obtained by English learners of Spanish and Spanish learners of English in the production experiment.

6. Production experiment

This chapter presents and analyzes the results obtained in the production experiment. First, the way the production data was analyzed will be explained. Results will be presented first by contrasting the results for different groups within the same language – i.e., the results of the groups sharing the same L1, as well as the corresponding controls will be compared – and then contrasting the performance of each group in the two languages – i.e., the English production of the groups with the same L1 will be compared to their production in Spanish. Finally, all the results obtained in the production experiment will be discussed.

6.1. Data Analysis

The production experiment involved the reading of a total of 32 carrier sentences with 20 target words – i.e., five for each stop – that had to be repeated twice. Just as in the case of perception, 92 participants completed the experiment, but 12 were discarded, as they did not match the criteria for any group. The recordings of the production of one participant (ENEXPinUK12) both in Spanish and in English were not saved correctly and, therefore, could not be included in the data analysis. As a result, the productions of 79 participants were analyzed. Some productions were left out because they were either unclear or produced with a sound other than a stop – i.e., an approximant or a fricative – or because the quality of the recording at the point where the word was produced was not appropriate for measuring VOT – e.g., due to overlapping noise. Table 6.1 below shows the total number of tokens that were analyzed for each group, language and stop category.

Group	*N*	English				Spanish			
		/p/	**/b/**	**/k/**	**/g/**	**/p/**	**/b/**	**/k/**	**/g/**
ENCONT	9	90	90	90	90	-	-	-	-
ENINEXPinUK	11	110	110	110	110	109	109	110	107
ENEXPinUK	9	90	90	90	90	90	90	90	85
ENEXPinSP	11	110	110	110	110	110	110	110	110
SPEXPinUK	9	90	90	90	89	90	90	90	90
SPEXPinSP	10	99	99	100	95	98	100	99	98
SPINEXPinSP	10	100	100	100	98	100	100	100	100
SPCONT	10	-	-	-	-	100	97	100	99

Table 6.1. Total number of productions analyzed per group, language and stop category

The VOT of a total of 5531 target words was measured manually using Praat. Measurements of short-lag and long-lag stops were made from the onset of the burst to the beginning of voicing as indicated by the beginning of periodicity in the signal, taking into account both the spectrogram and the oscillogram (see Figures 6.1., 6.3, 6.5, 6.7, 6.8, 6.11, 6.13, 6.15, 6.16 in section 6.2 for an example). As for voice-lead production, VOT was measured from the onset of voicing until the beginning of the burst (see Figures 6.2, 6.6, 6.10 and 6.14 in section 6.2 for an illustration). Speech rate – measured with tonic vowel length and word length – has been found to influence VOT duration – i.e., a faster speech rate tends to result in shorter VOT values (Kessinger & Blumstein, 1997; Magloire & Green, 1998, Theodore, Miller & de Steno, 2009). In order to determine whether speech rate had an effect on VOT duration – and, thus, raw data could not be used in the analysis – a number of correlations involving the VOT duration of each phone and duration of the tonic vowel and of the whole word were carried out. The data in the English production experiment gathered in the UK – the first data collection – were used, as a decision on the measure used to quantify VOT needed to be made at the early stages of data analysis. The tests failed to establish clear correlations between VOT duration and the other variables and, thus, it was deemed unnecessary to find a leveling measure that controlled for speech rate. Instead, raw data – i.e., VOT duration in ms – were used in the analysis, as has often been done in previous studies (e.g., Antoniou et al., 2010; Caramazza et al., 1973; Casillas, 2019; Flege, 1987; Williams, 1977)[5].

Results were classified according to language, phone – i.e., /p/, /b/, /k/ and /g/ – and the L1 of the participants. A generalized linear mixed model (GLMM) was conducted for each stop category in each language – i.e., one with L1-English groups plus the appropriate control group and one with L1-Spanish groups plus the appropriate control – in order to test the effect of group – the independent variable, which differed in terms of L2 experience – on VOT production – the dependent variable. The model included word and word repetition as repeated measures, group as a fixed effect and participant as a random effect, VOT in ms being the dependent variable. Regarding the differences between English and Spanish production, a GLMM was also conducted. In this case, language was a fixed effect – along with group. Pair-wise comparisons with a Bonferroni correction were also conducted in each model. Moreover, Pearson correlation tests comparing L1 and L2 VOT use were carried out on each phone in order to determine if the amount of VOT used in the L1 was significantly correlated to the amount used in the L2 – i.e., if those participants that used more aspiration/prevoicing in English also did so in Spanish.

Just as in the case of perception, the relationship between the two places of articulation elicited was examined – i.e., bilabials vs. velars. First, it was analyzed whether there was a relationship between /p/ and /k/ VOT production by conducting a correlation test for the testing groups sharing an L1. A separate test was carried out for

---

[5] In the case of voiced stops, an alternative measure was considered, namely a categorical variable regarding the presence or absence of voicing (prevoicing/no prevoicing), resulting in a count measure (number of prevoiced productions out of the total number of target voiced productions). This analysis yielded almost identical results to the analysis of the VOT duration. Thus, for the sake of consistency and following previous studies (e.g., Hazan & Boulakia, 1993; Zampini, 1999), the same measure was used for both voiced and voiceless stops, namely VOT duration in ms.

each control group. It was expected that participants that used longer VOT in their production of /p/ would also produce longer VOT for /k/. The same procedure was followed to analyze the relationship between voiced stops. The production of the two members of each contrast was compared – e.g., /p/ and /b/ and /k/ and /g/. It was expected that those participants that produced voiceless stops with shorter VOT would present voiced stops with lower VOT values – i.e., more prevoicing.

6.2. Between-groups within-language production analysis

This section will present and analyze the production results obtained in each language. Results will be classified according to testing language and the L1 of the groups analyzed, which will be compared to the results obtained by the corresponding control group as well as among the testing groups. First, the results for the English task will be presented. The results of the English-speaking groups will be provided first, followed by the Spanish groups. Then, the results obtained in Spanish by the L1 and the L2 speakers will be presented. In all cases the degree to which VOT production is correlated across the two places of articulation and across voiced-voiceless pairs will be examined. At the end of this section, a summary of the results presented will be provided.

6.2.1. Production of English stops

6.2.1.1. Production of English stops: English speakers

The production experiment consisted of a list of carrier sentences in each language which included the target phones – i.e., /p/, /b/, /k/ and /g/. The VOT of each target word was measured for every participant's production. The results obtained by the English control group have been used as the baseline for English VOT productions. Tables 6.2 and 6.3 show the mean VOT productions in ms for /p/, /b/, /k/ and /g/ by the English groups. In the case of /p/, all L1-English speaking groups presented a mean VOT of around 50 ms (ENCONT and UKINEXPinUK: 52 ms; ENEXPinUK: 50 ms; ENEXPinSP: 46 ms), that is, with long-lag VOT. Figure 6.1 illustrates the production of /p/ with long-lag VOT by a participant in the English control group. Considering the relatively small variability across groups, the GLMM (independent variable: group; dependent variable: VOT) failed to find a significant difference between the L1-English groups in their VOT use for /p/ [$F(3, 396) = 0.562$; $p = .641$].

Group	/p/ VOT mean	/p/ VOT range	b/ VOT mean	/b/ VOT range
ENCONT	52 (20)	16 – 104	1 (25)	-94 – 20
ENINEXPinUK	52 (20)	11 – 131	8 (6)	-47 – 24
ENEXPinUK	50 (18)	16 – 117	-9 (40)	-146 – 23
ENEXPinSP	46 (18)	5 – 94	0 (35)	-160 – 19

Table 6.2. L1-English groups' mean VOT duration (ms), standard deviation (in parentheses) and range for English /p/ and /b/.

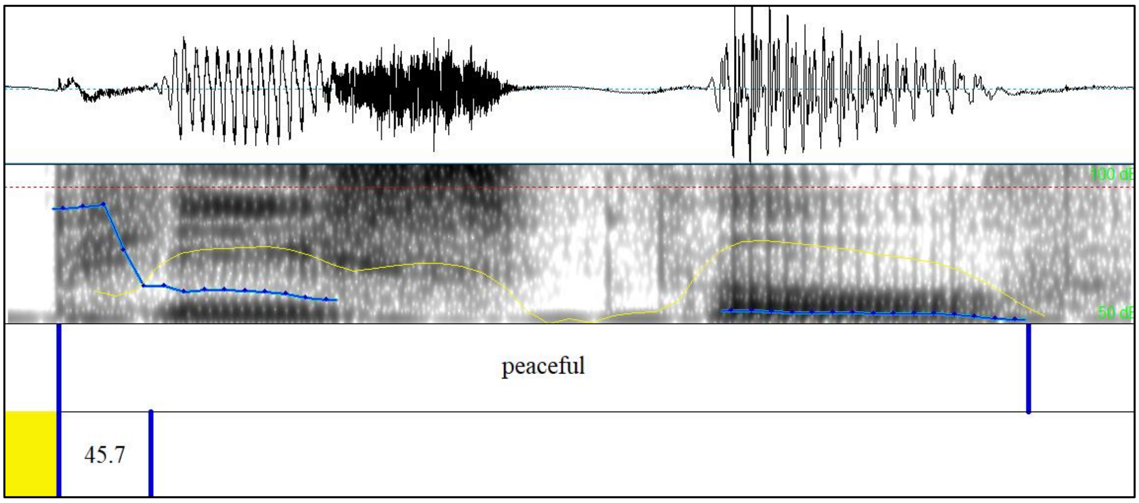

Figure 6.1. Production of the English word 'peaceful' with long-lag VOT by ENCONT03.

Regarding /b/, all L1-English groups presented VOT values around 0 ms, although some mans were positive – i.e., ENCONT's and ENINEXPinUK's – and others were negative – i.e., ENEXPinUK's and ENEXPinSP's. Even though most productions presented short-lag VOT, instances of prevoiced /b/ were also found (Figure 6.2 illustrates a production of /b/ with voice-lead VOT). ENCONT produced /b/ with a mean of 1 ms and 10% of bilabial voiced stops were produced with negative VOT productions, while ENINEXPinUK had the longest mean VOT, namely 8 ms, and presented the lowest percentage of voice-lead productions (1%). ENEXPinUK had a mean VOT value of -9 ms, with 18% of productions realized with voice-lead VOT. ENEXPinSP presented a mean VOT of 0 ms and produced /b/ with a negative VOT 15% of time. Overall, the range of VOT observed with L1-English speakers reflects the variability reported in the literature and the GLMM analysis indicated that differences between groups were not significant [$F(3, 396) = 2.005$; $p = .113$].

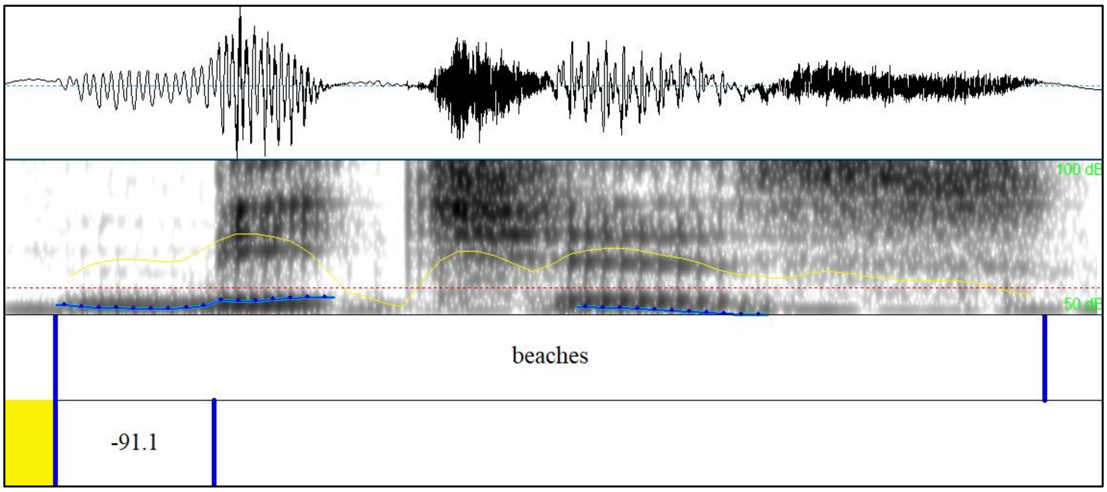

Figure 6.2. Production of the English word 'beaches' with voice-lead VOT by ENEXPinUK07.

Velar stops presented higher VOT means than bilabial stops, as expected. All groups obtained a mean VOT for /k/ between 70 ms and 80 ms (see Table 6.3). The group that produced the longest mean VOT was ENINEXPinUK (79 ms) followed by ENEXPinUK (76 ms). ENEXPinSP and ENCONT obtained a mean of 72 ms. As all groups produced similar VOT values, no significant effect of group was found on /k/ VOT production [$F(3, 396) = .780$; $p = .506$].

Group	/k/ VOT mean	/k/ VOT range	/g/ VOT mean	/g/ VOT range
ENCONT	72 (19)	41 – 134	14 (38)	-255 – 43
ENEXPinUK	79 (18)	41 –129	20 (23)	-109 – 45
ENEXPinUK	76 (19)	40 – 124	1 (48)	-129 – 58
ENEXPinSP	72 (21)	24 – 143	6 (41)	-137 – 49

Table 6.3. L1-English groups' mean VOT duration (ms), standard deviation (in parentheses) and range for English /k/ and /g/.

Just as in the case of /b/, there was variation in the production of /g/. Whereas most productions had short-lag VOT (see Figure 6.3 for an instance of a short-lag production), there were instances of voice-lead in all groups (see Table 6.3 for ranges). In this case, however, all means presented positive values. The group that presented the greatest VOT mean for /g/ was ENINEXPinUK, who produced /g/ with a mean of 20 ms and a percentage of voice-lead productions of 5%, followed by ENCONT, with 14 ms and 8% of short-lag productions. As for ENEXPinSP, their mean VOT for English /g/ was 6 ms and 18% of their productions were prevoiced. ENEXPinUK presented a mean VOT of 1 ms and had the highest percentage of short-lag productions (26%). However, differences between groups were not significant, [$F(3,396)$ 1.348; $p = .258$]. In short, as Figure 6.4 shows, all English groups presented similar VOT values for English /p/, /b/, /k/ and /g/. Thus, no significant differences were found between groups, indicating no influence of Spanish on L1 production.

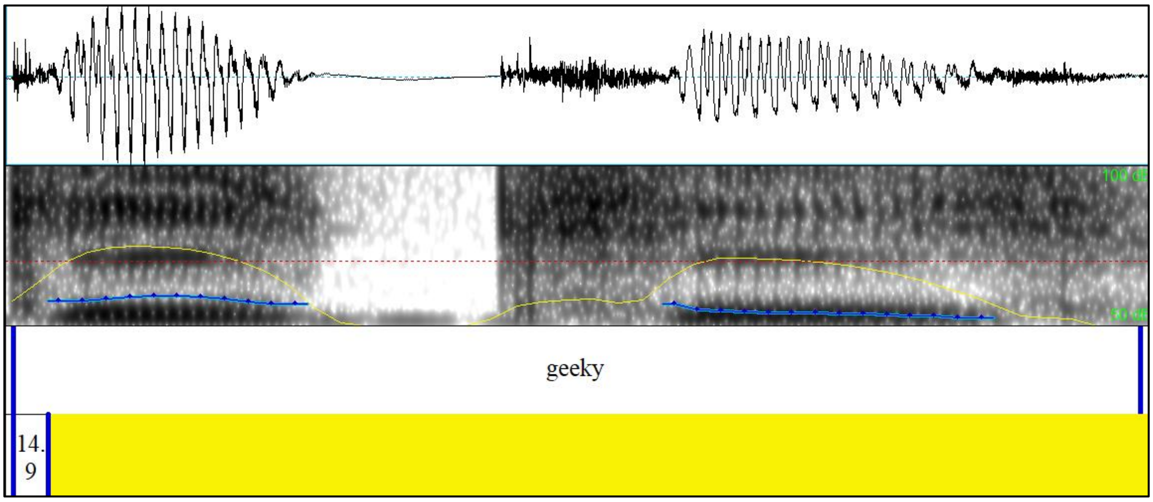

Figure 6.3. Production of the English word 'geeky' with short-lag VOT by ENCONT03.

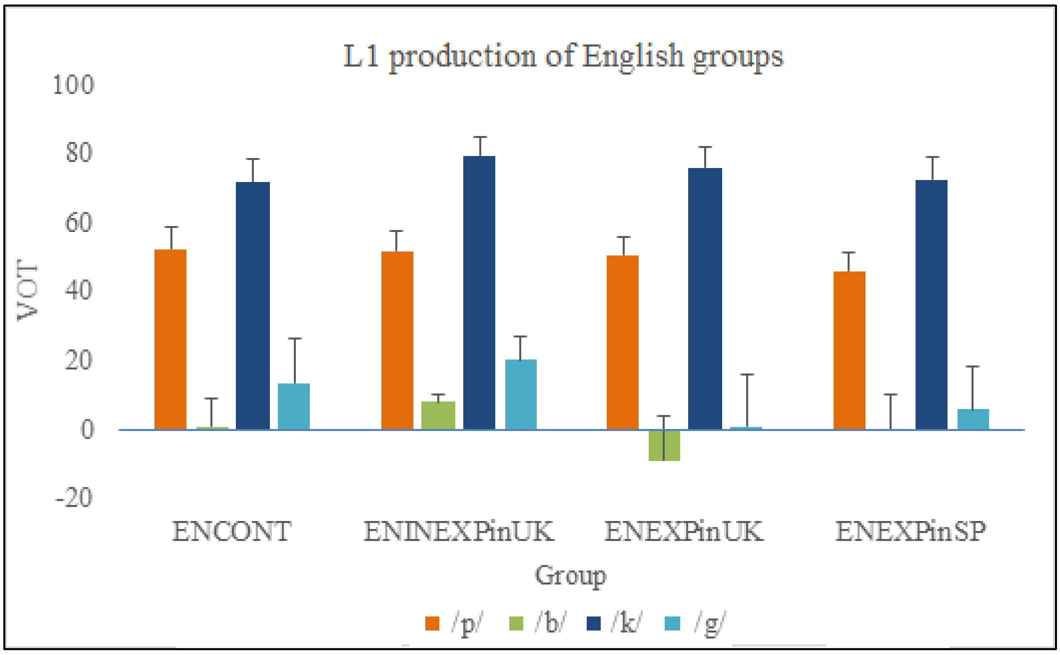

Figure 6.4. L1-English groups' mean VOT for English /p/, /b/, /k/ and /g/ productions in ms.

6.2.1.2. Production of English stops: Spanish learners of English

Spanish learners of English produced /p/ with shorter VOT values than ENCONT (See Table 6.4). Figure 6.5 illustrates the production of /p/ with short-lag VOT values by a Spanish learner of English. The group that produced /p/ with the shortest VOT mean was SPINEXPinSP – with 23 ms. SPEXPinSP and SPEXPinUK presented a similar mean (32 ms and 35 ms, respectively). The GLMM revealed a significant effect of group on /p/

VOT production [$F(3, 375) = 5.706$; $p = .001$]. Bonferroni pair-wise comparisons found that SPINEXPinSP and SPEXPinSP produced significantly shorter VOT values than ENCONT ($p = .000$ and $p = .027$, respectively). Differences between SPEXPinUK and ENCONT and between the L1-Spanish groups were not found to be significant ($p > .05$).

Group	/p/ VOT mean	/p/ VOT range	/b/ VOT mean	/b/ VOT range
ENCONT	52 (20)	16 – 104	1 (25)	-94 – 54
SPEXPinUK	35 (21)	6 – 133	-43 (49)	-182 – 29
SPEXPinSP	32 (25)	0 – 106	-71 (45)	-217 – 16
SPINEXPinSP	23 (22)	2 – 111	-66 (44)	-182 – 27

Table 6.4. L1-Spanish groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for English /p/ and /b/.

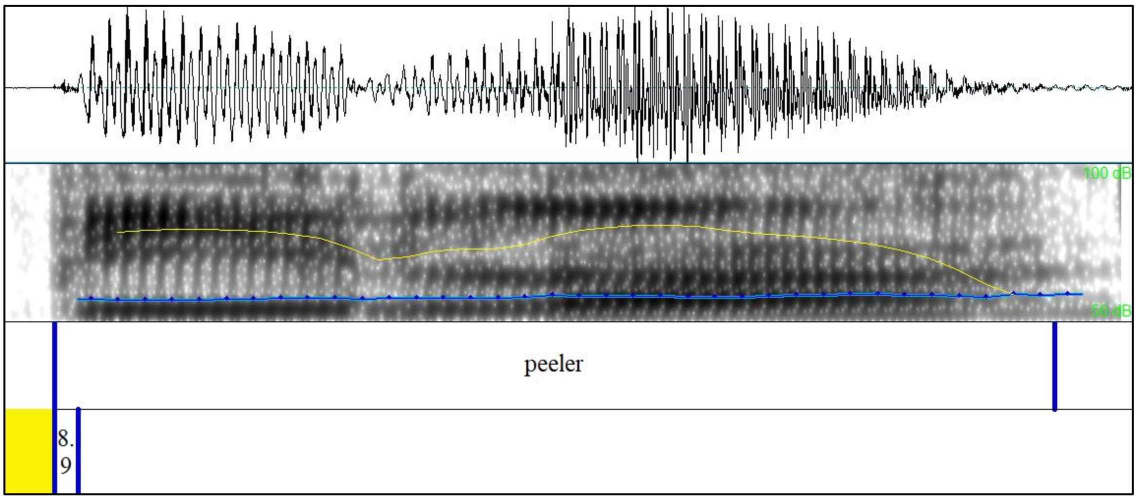

Figure 6.5. Production of the English word 'peeler' with short-lag VOT by SPSPINEXP08.

Regarding /b/, all Spanish L1 groups had a mean VOT with negative values (see Table 6.4). Figure 6.6 shows a production of English /b/ with voice-lead VOT by a Spanish learner of English. However, instances of short-lag VOT were produced by all groups. The group that performed most similarly to ENCONT was SPEXPinUK, with a mean VOT of -43 ms and 59% of prevoiced productions. The two Spanish groups tested in Spain produced /b/ with longer prevoicing than SPEXPinUK. SPEXPinSP had a mean

VOT for /b/ of -71 ms and SPINEXPinSP of -66 ms, with 86% and 88% of prevoiced productions, respectively. A GLMM revealed a significant effect of group on the production of /b/ [$F(3, 375) = 14.190$; $p = .000$]. All Spanish L1 groups presented significantly smaller VOT values – longer prevoicing – than ENCONT (SPINEXPinSP and ENCONT: $p = .000$; SPEXPinSP and ENCONT $p = .000$; SPEXPinUK and ENCONT: $p = .003$) and did not differ among themselves ($p > .05$).

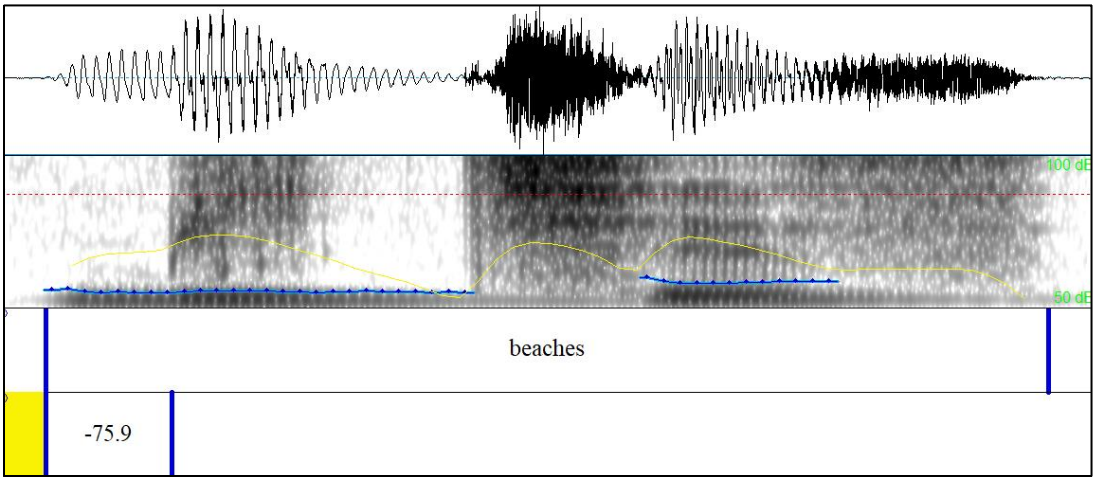

Figure 6.6. Production of the English word ‘beaches’ with voice-lead VOT by SPINEXPinSP07.

Contrary to the results for /p/, as Table 6.5 shows, SPEXPinUK and SPEXPinSP produced /k/ with VOT values that resembled those for ENCONT (72 ms): 73 ms of VOT in the case of SPEXPinUK and 70 ms regarding SPEXPinSP. SPINEXPinSP presented a shorter VOT, namely 53 ms. Figure 6.7 illustrates an instance of a production of /k/ by a Spanish learner using long-lag VOT. A GLMM revealed a significant effect of group on Spanish /k/ VOT production [$F(3, 376) = 2.927$; $p = .034$]. However, Bonferroni pair-wise comparisons failed to establish any significant differences between groups. The effect may have been driven by the fact that SPINEXPinSP’s VOT values were 17-20 ms lower than the other groups’ and the fact that the difference between SPINEXPinSP and SPEXPinUK approximated significance ($p = .070$).

Group	/k/ VOT mean	/k/ VOT range	/g/ VOT mean	/g/ VOT range
ENCONT	72 (19)	41 – 134	14 (38)	-255 – 43
SPEXPinUK	73 (22)	28 – 129	-24 (53)	-132 – 65
SPEXPinSP	70 (22)	21 – 132	-49 (41)	-127 – 59
SPINEXPinSP	53 (30)	15 – 124	-50 (46)	-166 – 46

Table 6.5. L1-Spanish groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for English /k/ and /g/.

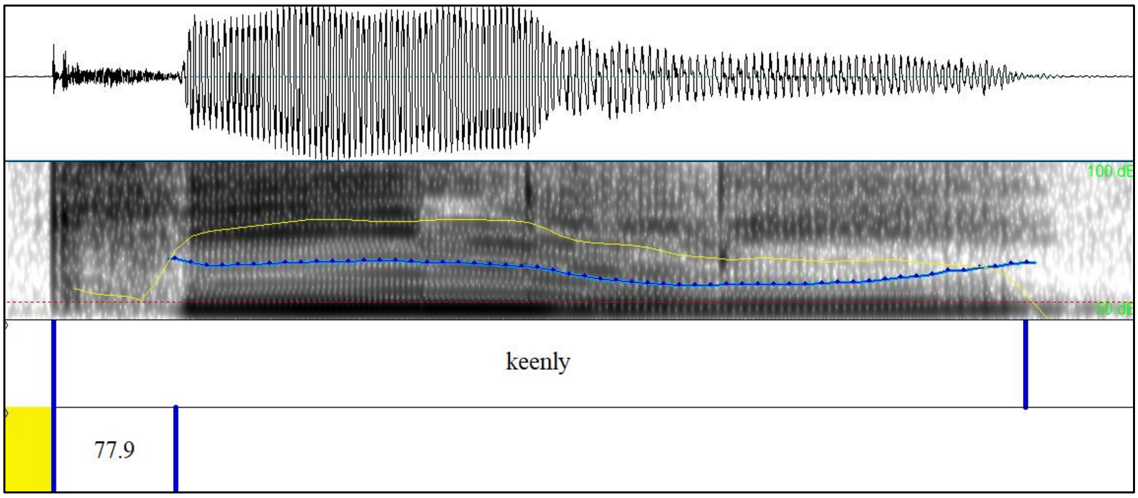

Figure 6.7. Production of the English word 'keenly' with long-lag VOT by SPEXPinSP03.

Just as in the case of /b/, all Spanish learner groups had a mean VOT for /g/ with negative values (see Table 6.5). It should be noted that instances of short lag productions were found in all groups (see Figure 6.8 for an example of a production of /g/ with short-lag VOT by a Spanish learner). SPEXPinUK presented the closest value to ENCONT, namely -24 ms and the lowest percentage of prevoiced tokens out of all L1-Spanish groups (55%). SPEXPinSP and SPINEXPinSP had similar VOT means, -49 ms (with 80% of prevoiced tokens) and -50 ms (83% of prevoiced productions), respectively. A significant effect of group was revealed [$F(3,369) = 18.162$; $p = .000$]. Bonferroni pair-wise comparisons confirmed that all L1-Spanish groups produced significantly lower VOT values – i.e., more prevoicing – than ENCONT (SPINEXPinSP and ENCONT: $p = .000$; SPEXPinSP and ENCONT: $p = .000$; SPEXPinUK and ENCONT: $p = .000$). Moreover, SPEXPinSP differed significantly from SPEXPinUK ($p = .036$).

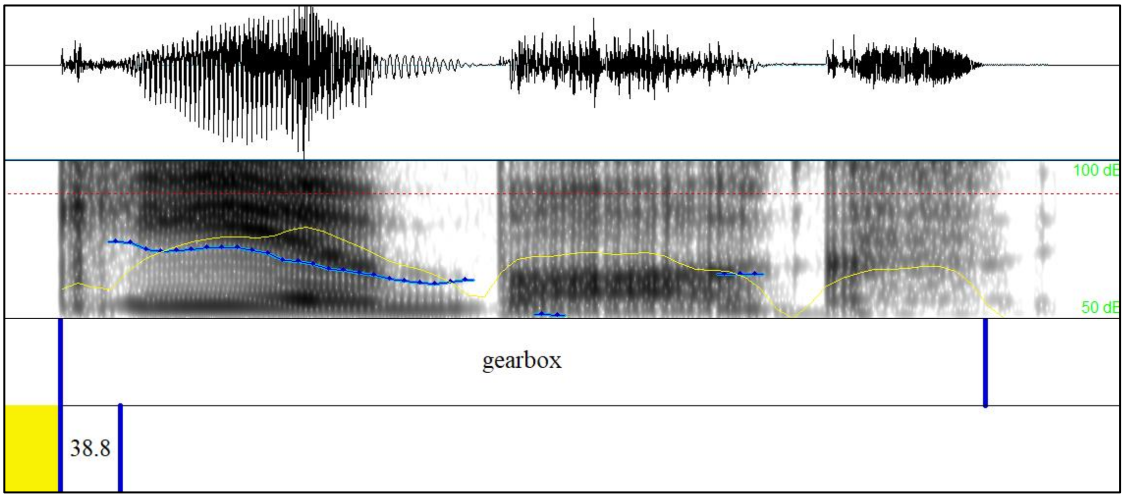

Figure 6.8. Production of the English word 'gearbox' with short-lag VOT by SPEXPinUK02.

In brief, as Figure 6.9 shows, only SPINEXPinSP and SPEXPinSP differed significantly from ENCONT in the production of /p/, as they produced shorter – more Spanish-like –VOT values than English monolinguals. None of the Spanish groups differed from ENCONT producing /k/, whereas, regarding voiced stops, all produced significantly longer prevoicing – i.e., used more Spanish-like VOT values – than ENCONT.

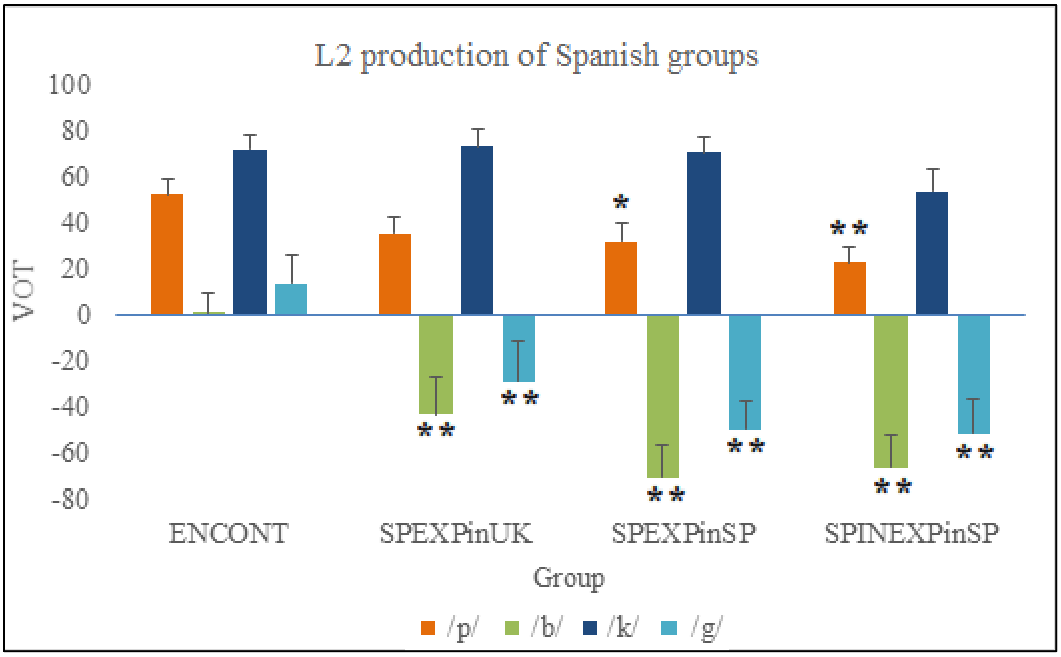

Figure 6.9. L1-Spanish groups' mean VOT for English /p/, /b/, /k/ and /g/ productions in ms. Significant differences between the learning groups and ENCONT are indicated using * (*p* < .05) and ** (*p* < .01).

6.2.2. Production of Spanish stops

6.2.2.1. Production of Spanish stops: Spanish speakers

Regarding stop production in Spanish, L1-Spanish speakers generally performed similarly to SPCONT (see Tables 6.6 and 6.7). SPCONT produced /p/ with a mean VOT of 9 ms, SPINEXPinSP with 8 ms of VOT, whereas SPEXPinSP and SPEXPinUK obtained slightly longer values, namely 10 ms. The GLMM did not yield significant results [$F(3, 384) = 0.918$; $p = .432$].

Group	/p/ VOT mean	/p/ VOT range	/b/ VOT mean	b/ VOT range
SPCONT	9 (4)	2 – 28	-78 (29)	-137 – 7
SPINEXPinSP	8 (5)	1 – 24	-64 (36)	-137 – 13
SPEXPinSP	10 (9)	3 – 84	-80 (26)	-135 – 3
SPEXPinUK	10 (5)	3 – 27	-68 (37)	-126 – 19

Table 6.6. L1-Spanish groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for Spanish /k/ and /g/.

As for /b/, all groups presented a mean VOT with negative values, although there were instances of short-lag productions in all groups (see Table 6.6 for ranges). ENCONT produced /b/ with a mean VOT of -78 ms. Figure 6.10 shows the spectrogram and oscillogram of an instance of a production of /b/ with voice-lead VOT by a participant in the Spanish control group. In spite of the fact that Spanish voiced stops have voice-lead VOT, five productions – three of which were produced by participant SPCONT09 – had positive values. Still, most instances presented voice-lead VOT (95%). Four different participants in the SPINEXPinSP group produced a total of six instances (6%) of short-lag /b/ in Spanish, whereas 94% of their productions had voice-lead VOT. The mean VOT for /b/ of SPINEXPinSP was -64. In the SPEXPinSP group, only one production (1%) presented short-lag VOT, that is, 99% of their productions had prevoicing. Their mean VOT for /b/ was very similar for that of SPCONT, namely -80 ms. SPEXPinUK had the lowest percentage of voice-lead VOT productions, but, still, most of their productions of /b/ in this group presented voice-lead VOT (88%). Short-lag instances were produced by five different participants. The GLMM failed to find a significant effect of group on the production of Spanish /b/ for the Spanish L1-groups [$F(3, 383) = 1.438$; $p = .231$].

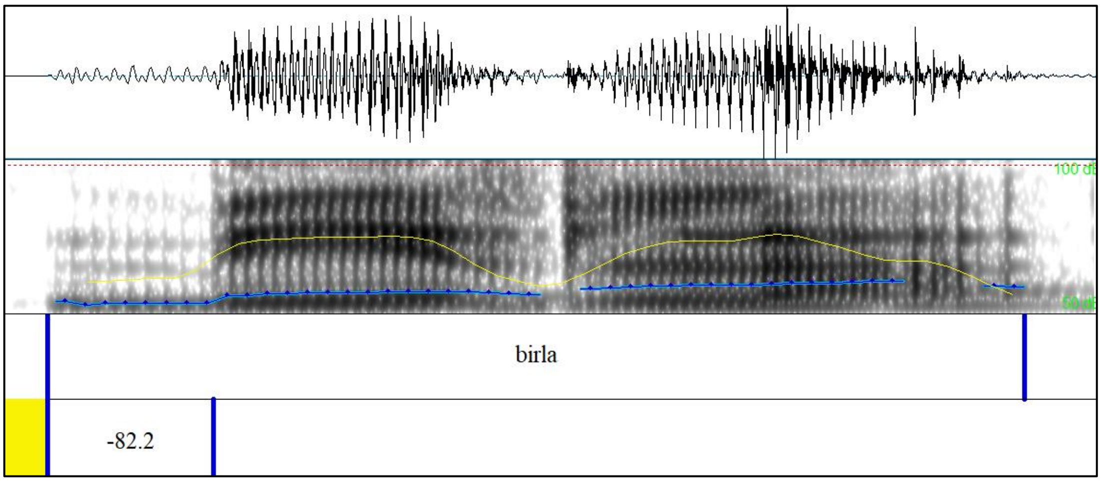

Figure 6.10. Production of the Spanish word 'birla' with voice-lead VOT by SPCONT07.

Spanish participants tended to use short-lag productions of /k/ (see Figure 6.7 for an example by a participant in the control group). Just as in the case of the production of /p/, SPINEXPinSP had a VOT mean for /k/ that was very similar to that of SPCONT – 32 ms and 34 ms, respectively –, whereas the experienced groups had greater VOT values. SPEXPinSP had a mean VOT for /k/ of 37 ms, that is, slightly longer than SPCONT's. The group that presented the greatest VOT mean was SPEXPinUK with 44 ms. The GLMM found a significant effect of group on the production of Spanish /k/ [$F(3, 385) = 3.202$; $p = .023$]. Bonferroni pair-wise comparisons revealed a significant difference between SPINEXPinSP and SPEXPinUK ($p = .022$).

Group	/k/ VOT mean	/k/ VOT range	/g/ VOT mean	/g/ VOT range
SPCONT	34 (11)	10 – 70	-77 (29)	-138 – 22
SPINEXPinSP	32 (11)	10 – 64	-71 (23)	-128 – -28
SPEXPinSP	37 (12)	15 – 77	-80 (51)	-175 – 43
SPEXPinUK	44 (15)	13 – 78	-77 (40)	-239 – 25

Table 6.7. L1-Spanish groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for Spanish /k/ and /g/.

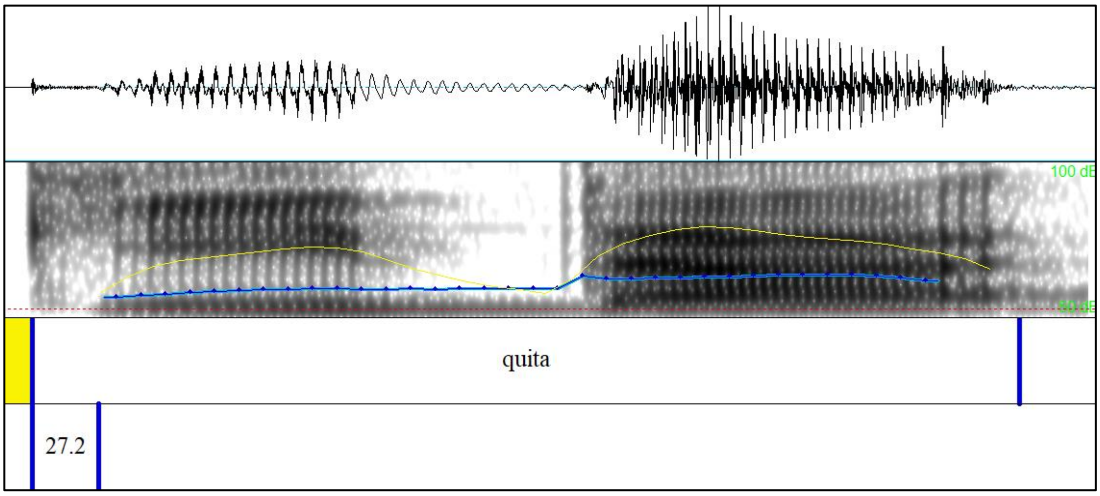

Figure 6.11. Production of the Spanish word 'quita' with short-lag VOT by SPCONT07.

As for /g/, all Spanish L1-groups had a mean VOT with negative values. SPCONT produced a mean VOT of -77 ms and 98% of the productions of /g/ had voice-lead VOT, whereas two instances (produced by SCONT09) had short-lag VOT values. SPEXPinUK also presented a mean VOT of -77 ms. However, they had the smallest percentage of voice-lead VOT productions (93%). It should be noted that all short-lag instances were produced by the same participant, namely SPEXPinUK05. SPINEXPinSP had a mean VOT for /g/ of -71 ms and they did not produce any instance of short-lag VOT. As for SPEXPinSP, they presented the longest prevoicing, namely -80 ms and 99% of their productions had voice-lead VOT. No significant effect of group was revealed [$F(3,383) = 0.330$; $p = .804$].

In short, as Figure 6.12 illustrates, all Spanish learners of English produced comparable VOT values to SPCONT for all stops, as no significant differences between the learner groups and the control groups were found. However, SPEXPinUK produced significantly longer VOT values for /k/ than SPINEXPinSP, suggesting a possible influence of English. This difference should be interpreted with caution, as no differences between SPEXPinUK and SPCONT were found.

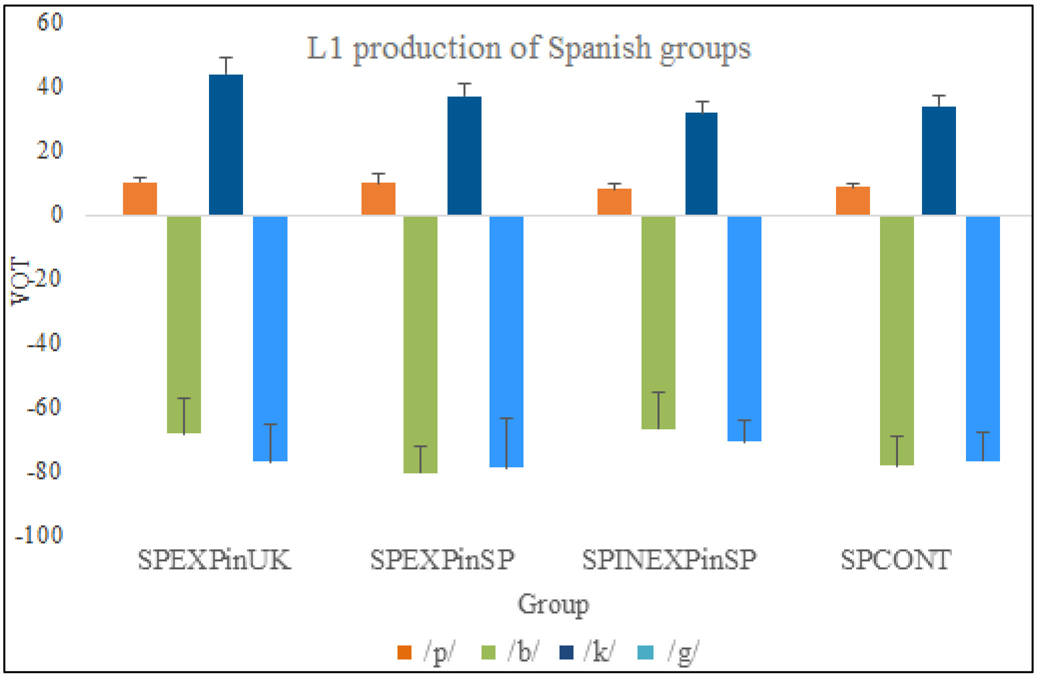

Figure 6.12. L1-Spanish groups' mean VOT for Spanish /p/, /b/, /k/ and /g/ in ms.

6.2.2.2. Production of Spanish stops: English speakers

L1-English L2-Spanish groups used longer VOT values than SPCONT when producing /p/ in Spanish (see Table 6.8). Figure 6.13 illustrates a production of Spanish /p/ with VOT values intermediate between English and Spanish controls produced by an English learner of Spanish. The group that performed the most similarly to SPCONT was ENEXPinSP, with a mean VOT of 14 ms. ENEXPinUK produced a mean VOT of 17 ms, whereas ENINEXPinUK produced the longest VOT values for /p/, that is, 28 ms. As with the previous sets of data, a GLMM was conducted and revealed a significant effect of group on the production of /p/ in Spanish [$F(3, 405) = 5.951$; $p = .001$]. Bonferroni pair-wise comparisons revealed that only ENINEXPinUK produced significantly longer VOT values than SPCONT ($p = .013$). Moreover, ENINEXPinUK were also found to produce Spanish /p/ with significantly longer VOT values than ENEXPinSP ($p = .000$).

Group	/p/ VOT mean	/p/ VOT range	/b/ VOT mean	/b/ VOT range
SPCONT	9 (4)	2 – 28	-78 (29)	-137 – 7
ENEXPinSP	14 (11)	3 – 84	-33 (47)	-118 – 119
ENEXPinUK	17 (11)	2 – 50	-48 (52)	-203 – 49
ENINEXPinUK	28 (23)	6 – 122	-13 (45)	-207 – 22

Table 6.8. L1-English groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for Spanish /p/ and /b/.

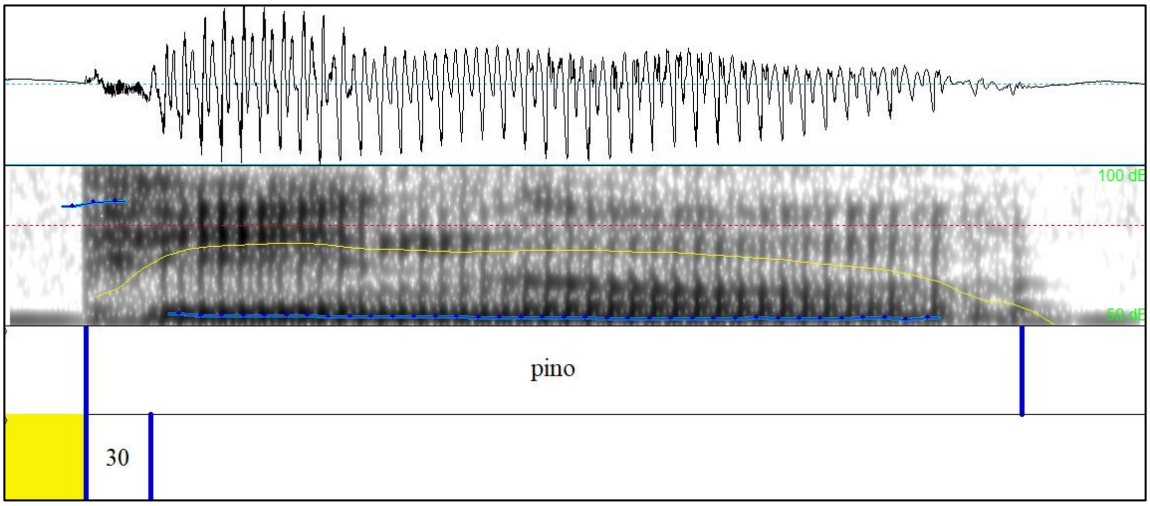

Figure 6.13. Production of the Spanish word 'pino' with long-lag VOT by ENINEXPinUK05.

As for Spanish /b/, all groups presented negative mean VOT values and ENINEXPinUK had the mean that was closest to 0 ms – i.e., -13 ms. Moreover, only 26% of all their productions of /b/ were prevoiced. ENEXPinSP presented a mean VOT value for /b/ of -33 ms. 56% of their productions had negative VOT values. The group that produced the closest values to those of SPCONT's was ENEXPinUK (-48 ms). As a matter of fact, it was the group that produced the greatest percentage of prevoiced tokens (64%). Figure 6.14 shows an example of a production of Spanish /b/ with voice-lead VOT by an English learner in the ENEXPinUK group. The GLMM revealed a significant effect of group on the production of /b/ [$F(3, 402) = 7.990$; $p = .000$]. ENINEXPinUK and ENEXPinSP were found to produce significantly less prevoicing than SPCONT ($p = .000$ and $p = .009$, respectively), whereas ENEXPinUK, the group with the most target-like mean, did not differ from SPCONT. Moreover, ENINEXPinUK was also found to produce significantly greater VOT values than ENEXPinUK ($p = .036$).

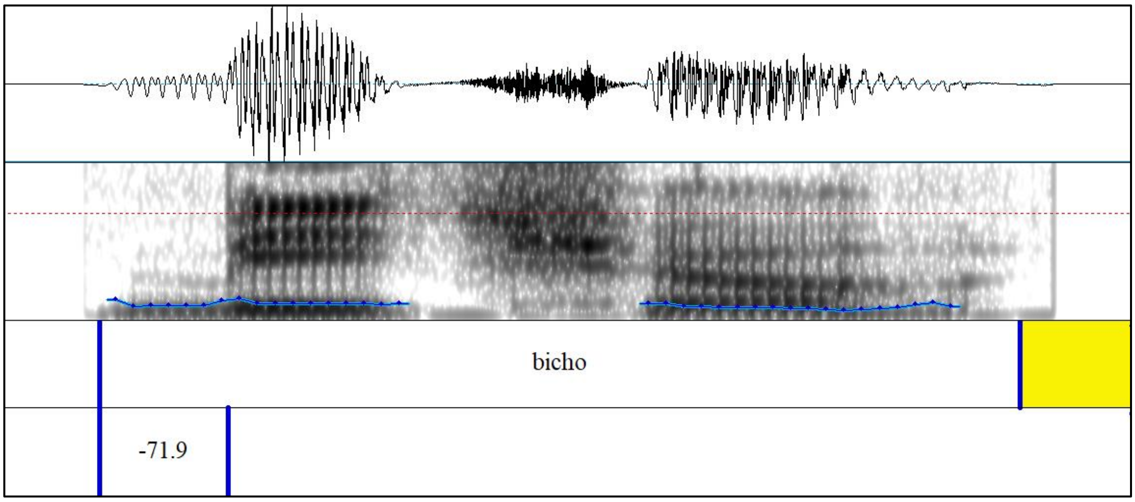

Figure 6.14. Production of the Spanish word 'bicho' with voice-lead VOT by ENEXPinUK02.

In the case of /k/, all groups had longer VOT means than SPCONT (see Table 6.9). Just as with /p/, ENEXPinUK presented the greatest VOT mean (55 ms). ENEXPinUK and ENEXPinSP had very similar means, namely 45 ms and 46 ms, respectively. Figure 6.15 shows an instance of a production of Spanish /k/ with short-lag VOT values by an English learner in the ENEXPinSP group. A GLMM revealed a significant effect of group on the production of Spanish /k/ [$F(3, 406) = 7.041$; $p = .000$]. Bonferroni pair-wise comparisons found that only ENINEXPinUK produced longer VOT values than SPCONT ($p = .000$).

Group	/k/ VOT mean	/k/ VOT range	/g/ VOT mean	/g/ VOT range
SPCONT	34 (11)	10 – 70	-77 (29)	-138 – 22
ENEXPinSP	46 (12)	18 – 79	-29 (48)	-123 – 51
ENEXPinUK	45 (14)	13 – 109	-55 (51)	-175 – 43
ENINEXPinUK	55 (21)	11 – 125	-6 (52)	-198 – 48

Table 6.9. L1-English groups' mean VOT duration (in ms), standard deviation (in parentheses) and range for Spanish /k/ and /g/.

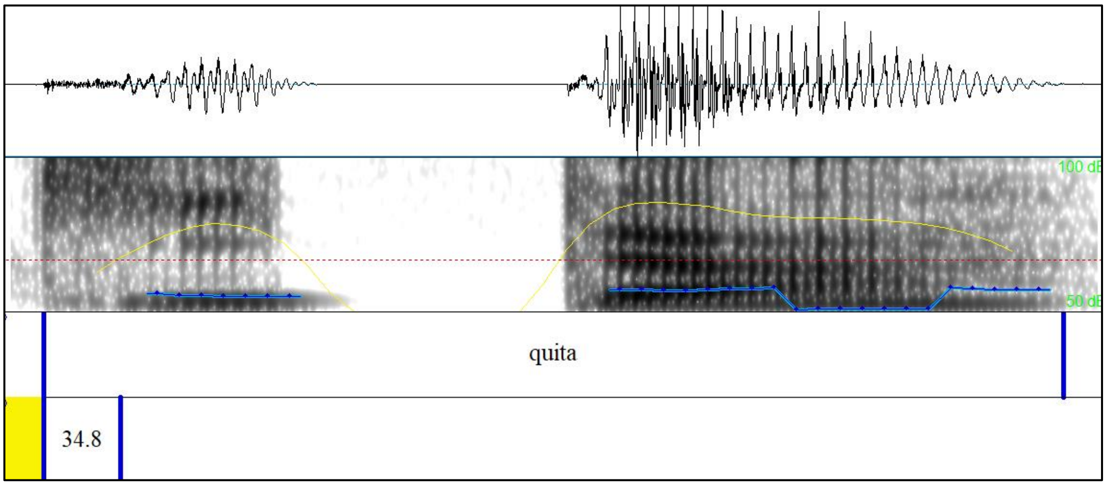

Figure 6.15. Production of the Spanish word 'quita' with short-lag VOT by ENEXPinSP03.

Regarding /g/, instances of both short-lag and voice-lead VOT productions were found, just as in the case of /b/. Figure 6.16 shows an example of a production of /g/ with short-lag VOT by an English learner of Spanish. The group that obtained the closest VOT mean to that of SPCONT – i.e., the lowest value – was ENEXPinUK, with -55 ms. In fact, they produced most /g/ tokens with prevoicing (79%). ENEXPinSP had a mean VOT of -29 ms and produced 62% of tokens with voice-lead VOT. ENINEXPinUK presented a mean VOT for /g/ close to 0 ms – i.e., -6 ms. Only 31% of their productions of /g/ were prevoiced. A significant effect of group on the production of Spanish /g/ was revealed [$F(3, 397) = 8.354$; $p = .000$]. The Bonferroni pair-wise comparisons found that ENINEXPinUK and ENEXPinSP differed significantly from SPCONT ($p = .000$ and $p = .005$, respectively), but failed to establish a significant difference between ENEXPinUK and SPCONT ($p > .05$). Moreover, ENEXPinUK, the group that performed in a more target-like manner, was found to differ significantly from ENINEXPinUK ($p = .009$). In summary, ENINEXPinUK produced longer VOT values than SPCONT for all stops (see Figure 6.17), indicating an influence of English, whereas SPEXPinUK did not differ from Spanish monolinguals. ENEXPinSP produced all Spanish stops with comparable VOT values to SPCONT and, thus, no significant differences from SPCONT were found.

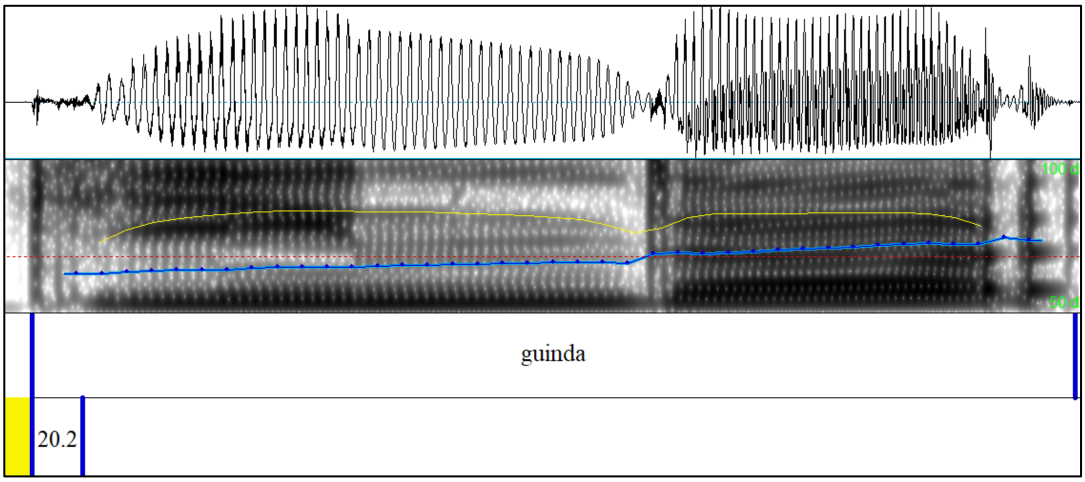

Figure 6.16. Production of the Spanish word 'guinda' with short-lag VOT by ENEXPinSP06.

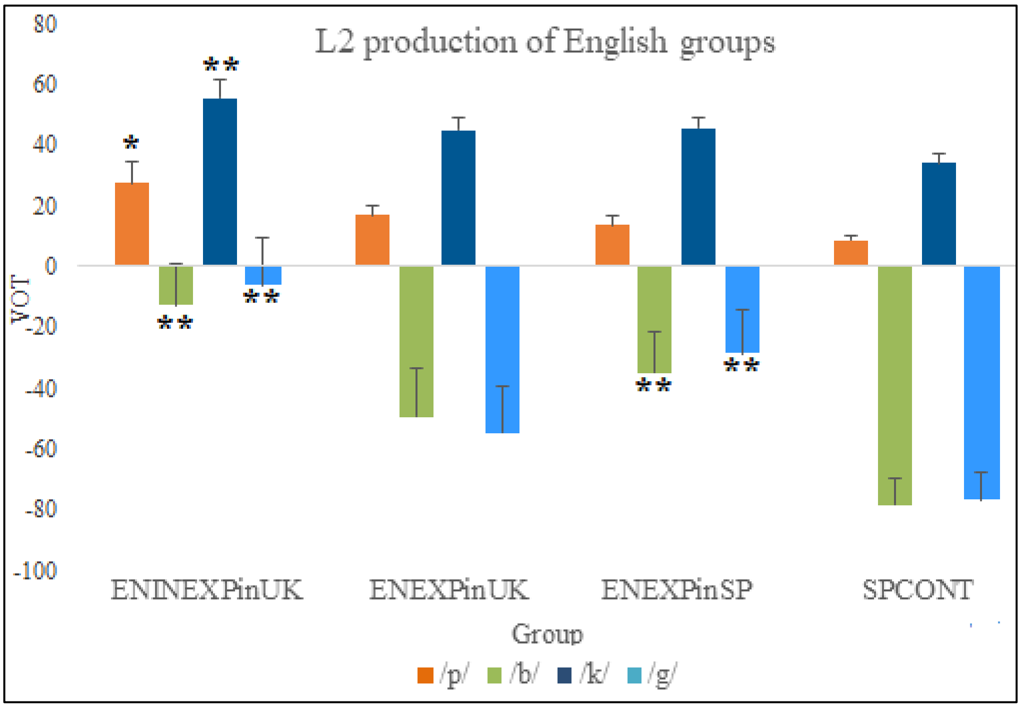

Figure 6.17. L1-English group's mean VOT for Spanish /p/, /b/, /k/ and /g/ in ms. Significant differences between the learning groups and ENCONT are indicated using * (*p* < .05) and ** (*p* < .01).

6.2.3. The relationship between the production of /p/, /b/, /k/ and /g/

Next, it was investigated whether there was a relationship between the VOT production of the different stop categories. Pearson correlation tests were conducted separately for ENCONT, SPCONT, the three L1-English L2-Spanish groups together and the three L1-Spanish L2-English groups as well, so as to determine whether there was a relationship between the production of homorganic stops – i.e., /p/ and /b/ and /k/ and /g/ – and between stops sharing the same voicing quality – i.e., /p/ and /k/ and /b/ and /g/ – in each language. That is, it was examined whether those speakers who used more aspiration also used shorter prevoicing, whether those participants who presented more aspiration in the production of /p/ also did so in the production of /k/ and whether speakers that used longer prevoicing for /b/ also did so in the case of /g/. Regarding English production, only /p/ and /k/ VOT use were found to be significantly correlated in the case of ENCONT. A relationship between the two voiceless stops and the two voiced stops – i.e., between /p/ and /k/ and /b/ and /g/ – was found in the case of the English learners of Spanish and the Spanish learners of English (See Table 6.10). However, no significant correlations were revealed regarding the amount of aspiration and prevoicing with homorganic stops (See Table 6.10).

Group	/p/-/k/	/b/-/g/	/p/-/b/	/k/-/g/
ENCONT	r = .741, N = 9, p = .011*	r = .648; N = 9; p = .030*	r = -.447, N: 9; p = .114	r = -.398, N = 9; p = .114
L1-English groups	r = .778; N = 31; p = .000**	r = .676; N = 31; p = .000**	r = -.152; N = 31; p = .208	r = .226; N = 31; p = .110
L1-Spanish groups	r = .805; N = 29; p = .000**	r = .717; N = 29; p = .000**	r = 139; N = 29; p = .236	r = .091; N = 29; p = .319

Table 6.10. ENCONT's, L2 English groups' and L1-Spanish groups' correlations for /p/-/k/, /b/-/g/, /p/-/b/ and /k/-/g/ productions in English. Significant results are indicated with *(at a .05) and ** (at a .01 level).

As for Spanish production, a significant correlation was found between the production of /p/ and /k/ and /p/ and /b/ in the case of SPCONT, whereas no significant results were revealed regarding /b/ and /g/ and /k/ and /g/ (see Table 6.11). Conversely, in the case of the Spanish learners of English as well as the English learners of Spanish, a significant correlation was found between voiceless and voiced stops, whereas that was not the case regarding homorganic stops (see Table 6.11).

Group	/p/-/k/	/b/-/g/	/p/-/b/	/k/-/g/
SPCONT	*r* = .582, *N* = 10; *p* = .039*	*r* = .358, *N* = 10; *p* = .155	*r* = .711, *N* = 10; *p* = .011*	*r* = .181, *N* = 10; *p* = .308
L1-Spanish groups	*r* = .628; *N* = 29; *p* = .000**	*r* = .680; *N* = 29; *p* = .000**	*r* = .209; *N* = 29; *p* = .139	*r* = -.140; *N* = 29; *p* = .235
L1-English groups	*r* = .768; *N* = 31; *p* = .000**	*r* = .790; *N* = 31; *p* = .000**	*r* = -.083; *N* = 31; *p* = .329	*r* = -.131; *N* = 31; *p* = .241

Table 6.11. SPCONT's, L2 English groups' and L1-Spanish groups' correlations for /p/-/k/, /b/-/g/, /p/-/b/ and /k/-/g/ productions in Spanish. Significant results are indicated with *(at a .05) and ** (at a .01 level).

In short, all groups presented a correlation between the production of /p/ and /k/ and between /b/ and /g/ in both languages – except SPCONT, who only presented a correlation between the latter. In other words, those participants who tended to use greater aspiration for /p/ also did so in the production of /k/, and those participants who used longer prevoicing for /b/ also did in the case of /g/. However, regarding the amount of aspiration and prevoicing, no significant correlation emerged, with the exception of the production of /p/ and /b/ in the case of SPCONT.

6.2.4. Between-groups production: results summary

The production data for each L2 learner group were collected in both the L1 and the L2 and were compared to the native speaker control groups in each case, as well as to the learner groups with the same L1. Both ENCONT and SPCONT performed as expected: ENCONT produced /p/ and /k/ with long-lag values and /b/ and /g/ with values around 0 ms, while SPCONT produced /p/ and /k/ with short-lag VOT and /b/ and /g/ with voice-lead VOT. Regarding the learner groups, they produced L1 stops with native-like values and no significant differences were found when compared to the control groups or among themselves. The only exception was Spanish /k/, as SPEXPinUK were found to produce significantly longer VOT values than SPINEXPinSP (see Table 6.12 for all significant differences between groups).

Regarding the L2, the experienced groups were found to perform in a more target-like manner than the inexperienced groups, especially in the case of voiceless stop production. Some evidence of a positive effect of L2 experience on L2 production was found in the case of the English learners, since only the least experienced group differed significantly from SPCONT in the production of all stops – they used greater, more English-like VOT values. Moreover, ENINEXPinUK used significantly longer VOT values than ENEXPinSP in the production of /p/ and significantly less prevoicing than ENEXPinUK in voiced stop production. Still, a straightforward effect of experience was not found regarding voiced stops: the most experienced group (ENEXPinSP) used significantly less prevoicing than SPCONT, whereas ENEXPinUK – who presented the longest prevoicing for voiced stops both in their L1 and L2 – did not differ from SPCONT for any stop. As for the Spanish learners, the more experienced the group was, the more target-like their VOT values were: SPINEXPinSP and SPEXPinSP produced /p/, /b/ and /g/ – but not /k/ – with significantly different VOT values from ENCONT, whereas

SPEXPinUK differed from ENCONT only in the production of voiced stops. Moreover, SPINEXPinSP produced significantly shorter VOT values in the production of /k/ than SPEXPinUK.

Group comparisons	Language	/p/	/b/	/k/	/g/
ENCONT vs. SPINEXPinSP	(L2) English	52 vs. 23**	1 vs. -66**	72 vs. 53	14 vs. -50**
ENCONT vs. SPEXPinSP	(L2) English	52 vs. 32*	1 vs. -71**	72 vs. 70	14 vs. -49**
ENCONT vs. SPEXPinUK	(L2) English	52 vs. 35	1 vs. -43**	72 vs. 73	14 vs. -24**
SPINEXPinSP vs. SPEXPinUK	(L1) Spanish	8 vs. 10	-64 vs. -68	32 vs. 44 *	-71 vs. -77
SPCONT vs. ENINEXPinUK	(L2) Spanish	9 vs. 28**	-78 vs. -13**	34 vs. 55**	-77 vs. -6**
SPCONT vs. ENEXPinUK	(L2) Spanish	9 vs. 17	-78 vs. -48	34 vs. 45	-77 vs. -55
SPCONT vs. ENEXPinSP	(L2) Spanish	9 vs. 14	-78 vs. -33**	34 vs. 46	-77 vs. -29**
ENEXPinSP vs. ENINEXPinUK	(L2) Spanish	14 vs. 28**	-33 vs. -13	46 vs. 55	-29 vs. -55
ENEXPinUK vs. ENINEXPinUK	(L2) Spanish	17 vs. 28	-48 vs. -13*	45 vs.55	-55 vs. -6**

Table 6.12. Summary of the significant results found between groups in the production of /p/, /b/, /k/ and /g/ in English and in Spanish. Significant results are indicated with * (at a .05 level) and ** (at a .01 level).

It was also investigated whether there was a relationship between the amount of VOT used in the production of homorganic, voiced and voiceless stops. Results indicated that there was a relationship between the production of phones with the same voice quality – between /p/ and /k/ and between /b/ and /g/ –, except in the case of the production of /b/ and /g/ by SPCONT. No clear relationship was found regarding homorganic stops, as only the use of VOT in the production of /p/ and /b/ by SPCONT was found to be significantly correlated.

6.3. Between-languages production analysis

This section compares the results obtained by the learner groups in the English and in the Spanish production tasks. First, the results obtained by the English learners of Spanish will be presented, followed by the results of the Spanish learners of English. Next, the degree to which amount of VOT in the L1 and the L2 are correlated will be examined. The section ends with a summary of the results.

6.3.1. L1 and L2 production by English learners of Spanish

Table 6.13 illustrates the mean productions of all English groups in both languages. As can be observed, all L1-English groups presented different VOT means in their L1 and L2, although in some cases this difference was greater than in others. Regarding /p/, ENINEXPinUK produced English /p/ with long-lag VOT (52 ms) and Spanish /p/ with numerically intermediate values between Spanish and English (28 ms). The two experienced groups, ENEXPinUK and ENEXPinSP, produced both L1 and L2 /p/ with target-like values – i.e., with long-lag VOT in English and with short-lag VOT in Spanish, as indicated by the between groups analysis in section 6.2 –, although they were not numerically identical to those of Spanish controls (see Table 6.13). A GLMM, with VOT of /p/ as the dependent variable, word and word repetition as repeated measures, group as a fixed effect and participant as a random effect was conducted. Language was found to have a significant effect on VOT production [F (1, 613) = 340.548; p = .000]. Bonferroni pairwise comparisons revealed that all learner groups produced /p/ differently in English and in Spanish (p > .01 in all cases). Conversely, group did not yield a significant effect [F (2, 613) = 2.678; p = .070]. An interaction between language and group was found [F (2, 613) = 3.379; p = .035]. This result may stem from the fact that English learners of Spanish performed similarly in their L1, but differed in Spanish, as the results in Section 6.2 indicate.

Group	/p/		/b/		/k/		/g/	
	English	**Spanish**	**English**	**Spanish**	**English**	**Spanish**	**English**	**Spanish**
ENCONT	52	-	1	-	72	-	14	-
ENINEXPinUK	52	28	8	-13	79	55	20	-6
ENEXPinUK	50	17	-9	-48	76	45	1	-55
ENEXPinSP	46	14	0	-33	72	46	6	-29
SPCONT	-	9	-	-78	-	34	-	-77

Table 6.13. Mean VOT for Spanish and English /p/, /b/, /k/ and /g/ (in ms) by the English learners of Spanish and the control groups.

As for /b/, all groups presented mean VOT values closer to 0 ms in English than in Spanish. ENINEXPinUK had a mean VOT in English of 8 ms and presented negative values, though close to 0 ms, in Spanish (-13 ms). As for ENEXPinUK, they presented a negative mean value in both languages. However, their English mean was close to 0 ms (-9 ms), whereas they produced more prevoicing in Spanish (-48 ms). Similarly, ENEXPinSP had a mean VOT in English of 0 ms and presented a mean of -33 ms in Spanish – i.e., intermediate between English and Spanish. The GLMM analysis in this case revealed a significant effect of language on the production of /b/ [F (1, 613) = 67.796; p = .000] and the Bonferroni pairwise comparisons confirmed that all groups produced significantly different VOT values in their L1 and in their L2 (p < .01 in all cases). A significant effect of group was also found [F (2, 613) = 3.373; p = .035], and the Bonferroni pairwise comparisons indicated that ENINEXPinUK – who used prevoicing the least – differed from ENEXPinUK (p = .007) and ENEXPinSP (p = .047) in Spanish, whereas no differences were found in English. In spite of the fact that ENINEXPinUK made a smaller difference between the production of /b/ in English and Spanish than the experienced groups, no significant interaction between group and language was found [*F* (2, 613) = 2.096; *p* = .124].

Regarding /k/, the English learners of Spanish produced greater VOT values in English than in Spanish. Whereas in English they presented a VOT mean between 72 ms and 79 ms, in Spanish they used values between 45 ms and 55 ms. ENINEXPinUK obtained the greatest VOT means for /k/ in both languages, namely 79 ms in English and 55 ms in Spanish. ENEXPinUK and ENEXPinSP had very similar VOT means, 76 ms and 72 ms in English and 45 ms and 46 ms in Spanish, respectively. The GLMM revealed a significant effect of language on the amount of VOT used to produce /k/ [*F* (1, 614) = 251.008; *p* = .000]. Pairwise comparisons showed that all English groups produced /k/ with significantly different values in their L1 and in their L2 (*p* < .01 for all groups). However, the analysis failed to find a significant effect of group [*F* (2, 614) = 2.367; *p* = .256]. No interaction between group and language was found [*F* (2, 614) = 1.335; *p* = .264].

In the case of /g/, it was produced with positive mean values in English and negative mean values in Spanish by all L1-English groups. ENINEXPinUK presented the greatest mean VOT in English (20 ms), as well as the value closest to 0 ms in Spanish (-6 ms), resulting in the smallest difference between the two languages out of all the English groups. ENEXPinUK had the lowest mean VOT in English (1 ms) as well as in Spanish (-55 ms). As a matter of fact, as indicated above, ENEXPinUK was the English group that used prevoicing most often. ENEXPinSP produced /g/ with a mean VOT of 6 ms in English and -29 ms in Spanish. The GLMM found a significant effect of language [*F* (1, 606) = 203.520; *p* = .000] on /g/ VOT. Just as in the case of the other categories, all groups were found to produce /g/ with significantly different VOT in Spanish and in English (*p* < .01 in all cases). No significant effect of group was found

[*F* (2, 606) = 0.758; *p* = .469]. An interaction between group and language was revealed, possibly because differences between languages were greater in the case of ENEXPinUK than in the other two English groups [*F* (2, 606) = 10.862; *p* = .000]. In brief, in the case of the English learners of Spanish, the between languages comparison consistently showed that all groups used significantly greater VOT values – i.e., longer aspiration in voiceless stops and less prevoicing in voiced stops – in English production than in Spanish production.

6.3.2. Spanish learners of English

The L1-Spanish groups produced English and Spanish stops differently in most cases, that is, they tended to use numerically longer VOT values in English than in Spanish, especially regarding the experienced groups (see Table 6.14). SPINEXPinSP produced /p/ with a mean VOT of 8 ms in Spanish and of 23 ms in English. SPEXPinSP and SPEXPinUK produced Spanish /p/ with 10 ms and English /p/ with 32 ms and 35 ms, respectively – i.e., with numerically intermediate values between SPCONT and ENCONT. The GLMM revealed a significant effect of language on the production of /p/ [*F* (1, 571) = 193.749; *p* = .000] and the Bonferroni pair-wise contrasts revealed that all groups produced /p/ with significantly shorter VOT values in Spanish than in English (*p* < .01 in all cases). The analysis failed to find a significant effect of group [*F* (2, 571) = 1.994; *p* = .137], but an interaction between group and language was observed [*F* (2, 571) = 4.782; *p* = .009], probably due to the fact that the experienced groups made a clearer distinction between the production of /p/ in the L1 and in the L2.

Group	/p/		/b/		/k/		/g/	
	Spanish	English	Spanish	English	Spanish	English	Spanish	English
SPCONT	9	-	-78	-	34	-	-77	-
SPINEXPinSP	8	23	-67	-66	32	53	-71	-50
SPEXPinSP	10	32	-80	-71	37	70	-80	-49
SPEXPinUK	10	35	-68	-43	44	73	-77	-24
ENCONT	-	52	-	1	-	72	-	14

Table 6.14. Mean VOT for Spanish and English /p/, /b/, /k/ and /g/ (in ms) by the Spanish learners of English and the control groups.

Regarding /b/, all groups produced this stop with a negative mean VOT. SPINEXPinSP obtained the most similar values in Spanish and in English, -67 ms and -66 ms, respectively. SPEXPinSP presented a slightly smaller VOT mean in Spanish (-80.4 ms) than in English (-70.7 ms). As for SPEXPinUK, they made the greatest difference between Spanish /b/ (-68 ms) and English /b/ (-43.1 ms) production. The GLMM revealed a significant effect of language on the production of /b/ [F (1, 573) = 9.223; p = .002] and pairwise comparisons only found a significant difference between the production of /b/ in Spanish and in English in the case of SPEXPinUK (p = .000). SPINEXPinSP and SPEXPinSP were not found to produce L1 and L2 /b/ differently (p > .05). No significant effect of group was revealed [F (2, 573) = 2.506; p = .082], but an interaction between group and language was found [F (2, 573) = 6.341; p = .002], probably due to the fact that not all groups made a significant difference between L1 and L2 /b/.

All L1-Spanish groups produced Spanish /k/ with shorter VOT values than English /k/. SPINEXPinSP's mean VOT for Spanish /k/ closely approximated that of SPCONT (32 ms), whereas they used intermediate values between Spanish and English for English /k/ (53 ms). SPEXPinSP seemed to produce native-like VOT values in both languages, as they had a mean VOT of 37 ms in Spanish and 70 ms in English. SPEXPinUK produced /k/ with somewhat greater VOT values than SPCONT in Spanish

(44 ms) and with target-like values in English (73 ms). The GLMM revealed a significant effect of language [*F* (1, 573) = 300.159; *p* = .000] and the Bonferroni pairwise comparisons confirmed that all groups performed differently in their L1 and L2 (*p* < .01 for all groups). A significant effect of group [*F* (2, 573) = 3.518; *p* = .030] was also found. The Bonferroni pair-wise contrasts did not find any differences between groups in Spanish, but SPINEXPinSP were found to produce significantly shorter VOT values than SPEXPinSP (*p* = .022) and SPEXPinUK (*p* = .015) in English. The interaction between the two independent variables reached marginal significance [*F* (2, 573) = 2.914; *p* = .055], probably because, although all groups performed differently in their L1 and L2, the magnitude of this difference varied from group to group.

Finally, all Spanish groups produced /g/ with negative mean VOT values in both their L1 and L2. In this case, all groups had more prevoicing in Spanish than in English. As Table 6.14 illustrates, all groups produced Spanish /g/ with mean VOT values between -80 ms and -71 ms, very close to SPCONT's values (-77 ms). As for English /g/, SPINEXPinSP and SPEXPinSP obtained very similar means (-50 ms and -49 ms, respectively), which were intermediate between SPCONT's and ENCONT's values, although closer to SPCONT's. SPEXPinUK produced English /g/ with a mean VOT of -24 ms, that is, intermediate between English and Spanish, but somewhat closer to ENCONT's values. The GLMM found a significant effect of language on the production of /g/ [*F* (1, 563) = 92.119; *p* = .000], but not on group [*F* (2, 563) = 0.932; *p* = .394]. An interaction between group and language was found [*F* (2, 563) = 6.181; *p* = .002], indicating that some groups – probably SPEXPinUK – made a greater difference between the two languages than the others. Still, pair-wise contrasts revealed that all groups produced Spanish and English /g/ with significantly different VOT values (*p* < .01 in all cases). In summary, all Spanish learner groups produced English /p/, /k/ and /g/ with significantly greater VOT values in English than in Spanish, whereas only SPEXPinUK used significantly less prevoicing in the production of English /b/ than for Spanish /b/.

6.3.3. Correlations between L1 and L2 stop production

With a view to find out whether there was a relationship between the use of VOT in the L1 and in the L2 – i.e., whether those participants that used greater VOT values in English also did so in Spanish – one-tailed Pearson correlation tests were conducted for each stop category on groups sharing the same L1, that is, one for the three groups of English learners of Spanish and one for the three groups of Spanish learners of English. Regarding the Spanish speakers, the Pearson correlation test revealed a significant – though moderate – correlation between L1 and L2 /b/, /k/ and /g/ (/b/: r =.363; N = 29; p = .026; /k/: r = .554; ; N = 29; p = .001; /g/: r = .321; N = 29; p =0.45), but not in the case of /p/ (p > .05). Regarding the English speakers, a significant and moderate correlation was found in the case of /k/ (r = .336; N = 31; p = .032), but no significant results were obtained for /p/, /b/ or /g/ (p > .05). Thus, a correlation between the use of VOT for the production of the L1 and L2 counterparts was not found in all cases and, when it was found, it was only moderate.

6.3.4. Between languages production: results summary

The production of L1 and L2 stops was examined and revealed that all English groups produced English stops with significantly longer VOT values than Spanish stops (see Table 6.15 below, which presents all the production results highlighting the significant results). That is, English learners of Spanish adjusted the amount of VOT used in stop production to the language that was being spoken. It was observed that the two experienced groups made a greater difference between the two languages than the inexperienced group in the production of voiceless stops, that is, the greater the amount of L2 experience the greater the difference between L1 and L2 VOT was. As for voiced stops, a clear effect of L2 experience was not found, since the group with intermediate experience was the group that made the greatest difference between the production of L1 and L2 /b/ and /g/. The use of VOT in the L1 and in the L2 was not found to be clearly correlated at an individual level, as only a significant, though moderate correlation was found in the case of /k/, whereas the analyses did not reach significance for /p/, /b/ and /g/.

Regarding the Spanish learners, all groups produced English /p/, /k/ and /g/ with significantly greater VOT values in English than in Spanish, whereas only SPEXPinUK produced English and Spanish /b/ differently. All groups produced stops with native-like VOT values in the L1 and with numerically intermediate values – between ENCONT's and SPCONT's – in the L2, which were more target-like in the case of the experienced groups than for SPINEXPinSP. In short, Spanish learners of English were better able to produce voiceless stops differently in the two languages than voiced stops and the experienced groups were found to make a greater difference between L1 and L2 stop production than the inexperienced learners. Moreover, the correlations between the use of VOT in the L1 and in the L2 by each participant yielded a significant result in all cases, except in the case of /p/, suggesting that those participants who used shorter VOT values in the L1 also did so in the L2, and the other way around.

Group	/p/	/b/	/k/	/g/
ENINEXPinUK	52 vs. 28**	8 vs. -13**	79 vs. 55**	20 vs. -6**
ENEXPinUK	50 vs. 17**	-9 vs. -48**	76 vs. 45**	1 vs. -55**
ENEXPinSP	46 vs. 14**	0 vs. -33**	72 vs. 46**	6 vs. -29**
SPINEXPinSP	8 vs. 23**	-66 vs. -67	32 vs. 53**	-71 vs. -50**
SPEXPinSP	10 vs. 32**	-80 vs. -71	37 vs. 70**	-80 vs. -49**
SPEXPinUK	10 vs. 35**	-68 vs. -43 **	44 vs. 73**	-77 vs. -24**

Table 6.15. Mean VOT for /p/, /b/, /k/ and /g/ obtained by each group in English vs. Spanish. Significant differences between the two languages are indicated with * (at a .05 level) and ** (at a .01 level).

6.4. Production experiment: discussion

The production of Spanish and English voiced and voiceless stops in the L1 and in the L2 has been analyzed by comparing the production of groups sharing the same L1 and differing in amount of L2 experience and the corresponding control groups. Moreover, the performance of the learner groups in English and in Spanish has been compared in order to determine whether they produce L1 and L2 stops differently and whether L2 experience has an effect on the ability to make this difference.

Regarding the ability to distinguish L1 and L2 phones, the results showed that, overall, all learner groups were relatively successful at producing English and Spanish stops differently, although this was not the case for the Spanish L1 groups living in Spain with respect to /b/. More specifically, all the Spanish groups produced /p/, /k/ and /g/ with greater VOT values in English than in Spanish, but only SPEXPinUK distinguished between Spanish and English /b/. Moreover, SPEXPinUK was the group that tended to make the greatest difference between L1 and L2 stop production, whereas SPINEXPinSP made the smallest difference, followed by the moderately experienced group. These group differences support a positive role of L2 experience on the creation of L2 categories.

The different results obtained for /b/ and /g/, that is, the fact that all three L1 Spanish groups produced L1 and L2 /g/ with different VOT values but only SPEXPinUK succeeded in the case of /b/, could be explained by intrinsic differences in VOT between the two places of articulation. Recall that velar stops tend to have longer VOT than bilabial stops (Lisker & Abramson, 1964; Poch, 1985), and thus short-lag /g/ presents longer VOT than short-lag /b/. By contrast, the amount of prevoicing is affected by place of articulation to a lesser extent, given that there is a great variability even within the same speaker (Lisker & Abramson, 1964; Poch, 1985, see section 2.5.1 in the Literature Review). Therefore, the difference between short-lag and voice-lead values is greater with /g/ than in the case of /b/. Consequently, given a similar proportion of short-lag and voice-lead productions, /g/ would have a greater mean VOT value than /b/. As a matter of fact, this is what was found with the Spanish learners of English. This may explain why it was more common to find significant differences between the L1 and the L2 with /g/ than with /b/. As a matter of fact, the percentage of prevoicing used by the Spanish groups to produce each voiced stop supports this assumption. All Spanish groups used a similar proportion of prevoiced stops – as opposed to short-lag VOT – for both voiced stops, and, in fact, English /b/ was produced somewhat more often – between 4% and 6% – with voice-lead VOT in the case of /b/ than in the case of /g/ (see section 6.2)

In the case of the English speakers, all groups seemed to adjust the use of VOT to the language that was being spoken, as they used more aspiration and less prevoicing in English than in Spanish. This finding contrasts with some previous studies, which have reported that early L2 learners – but not late learners, such as it is the case of most of the English participants in this study – are able to produce L1 and L2 stops with significantly different VOT values (e.g., Flege, 1991a). As expected, the least experienced group tended to make the smallest difference between the two languages – especially regarding

/p/. However, it was ENEXPinUK, the group with intermediate L2 experience and the greatest amount of L2 instruction – and not the highly experienced group – the one that made the greatest numerical difference between the two languages, especially in the case of voiced stops. Therefore, it appears that differences regarding amount of L2 instruction may also play a role. Whereas ENEXPinSP was the group with the greatest amount of L2 experience, they had also received less formal education in Spanish. Conversely, the group with intermediate experience (ENEXPinUK) had studied Spanish for the longest period of time. Therefore, it is possible that a combination of a short stay abroad with L2 instruction could help distinguishing L1 and L2 stop production to the same – or even greater – extent than longer periods of immersion in the L2 without explicit instruction. Amount of L2 instruction, as well as other individual factors including amount of weekly L2 use and dominance, will be analyzed in Chapter 7. Another difference between the two English experienced groups that should be taken into consideration has to do with their student status at the time of testing and their relationship to the institution where they completed the experiment. ENEXPinUK were tested in the same institution where they studied Spanish and were, in fact, recruited in class. Conversely, ENEXPinSP were not enrolled in Spanish lessons in the institution where they were tested and, in fact, most of them were not studying Spanish formally at the time of testing. As a result, participants in the ENEXPinUK group – i.e., the students – may have regarded the experiments more as an academic task, and consequently been more careful in the reading task, resulting in more target-like productions of voiced stops.

In spite of the fact that L2 learners tended to make a difference between L1 and L2 stop production, their VOT values were not always comparable to those of the control speakers. Whereas the learner groups tended to produce L1 stops with native-like values, they did not always succeed in producing L2 stops with target-like VOT. Regarding the

English learners, the inexperienced group failed to produce all L2 stops with target-like values, as they differed significantly from those of SPCONT. That is, they presented native-like L1 categories and intermediate – with values between those obtained by ENCONT and SPCONT – in Spanish. ENEXPinSP, the most experienced group, succeeded in producing voiceless stops with target-like VOT values – i.e., that resembled those of SPCONT – but failed to do so in the case of voiced stops, as they used significantly less prevoicing than Spanish controls. In other words, whereas ENEXPinSP presented separate target-like categories for L2 voiceless stops, their L2 categories for voiced stops had intermediate values. ENEXPinUK – the moderately experienced group – produced all L2 stops with VOT values that resembled those of the Spanish controls, that is, they presented target-like separate L2 categories, as no differences with SPCONT were observed. Still, it should be noted that all stops produced by ENEXPinUK presented a mean that was numerically greater than that of SPCONT in all cases. These results support the assumption that L2 experience has a positive effect on L2 stop production, given that the inexperienced learners differed from the Spanish speakers to a greater extent than the experienced groups. Moreover, amount of L2 instruction and student status may have also had an effect on the accuracy of L2 stop production, given that the moderately experienced group outperformed the highly experienced group in voiced stop production.

As for the L1 Spanish speakers, the groups tested in Spain (SPINEXPinSP and SPEXPinSP) failed to produce English /p/, /b/ and /g/ stops with target-like VOT, as they used shorter – more Spanish-like – values than ENCONT. In the case of /p/ and /g/, given that the two groups tested in Spain did not differ from ENCONT in their English production – but used significantly greater VOT values in their English production than in Spanish production –, it could be claimed that they presented a separate intermediate category in the L2. As for /b/, since both SPINEXPinSP and SPEXPinSP differed from ENCONT and did not make a difference between L1 and L2 production, it appears that the two Spanish groups in Spain presented a single shared category with Spanish-like VOT values. The most experienced group (SPEXPinUK) produced /p/ with VOT values that resembled those of controls but used more prevoicing for /b/ and /g/ – i.e., more Spanish-like VOT values – than ENCONT. Therefore, it seems that SPEXPinUK presented a separate target-like L2 category for /p/ and a separate category for /b/ and /g/ with intermediate values between ENCONT's and SPCONT's values. As for k/, all three Spanish groups produced /k/ with VOT values that resembled those of English controls, indicating that they had a separate target-like category for English /k/. Considering the differences observed between groups, it appears that L2 experience has a positive effect on L2 stop production regarding the Spanish learners, as the most experienced group obtained the most target-like productions. However, all groups, including the highly experienced learners, presented greater difficulty in producing voiced stops accurately than they did regarding voiceless stops. As a matter of fact, none of the Spanish groups produced voiced stops in a target-like manner – i.e., they all differed from ENCONT.

In short, all groups produced native-like VOT values in their L1 (cf., Flege, 1987, as will be discussed below), but failed to use target-like VOT in the L2 in some cases. This result suggests that the stop production of inexperienced learners – and of moderately experienced and highly experienced learners in some cases – is an instance of unidirectional CLI – from the L1 on the L2 – which was modulated by L2 experience and, possibly, to other factors like L2 instruction. Thus, even though all groups seemed to be able to distinguish L2 categories from the L1 – except in the case of /b/ on the part of the Spanish learners in Spain –, L2 learners did not always succeed in producing L2 stops 'authentically' – i.e., like native speakers. Still, it should be noted that, in some cases, no differences between the learner groups and the L2 controls were found. More specifically, ENEXPinUK produced all Spanish stops with target-like values and ENEXPinSP did so in the case of Spanish voiceless stops. Furthermore, all Spanish groups produced English /k/ with native-like values and SPEXPinUK also produced /p/ similarly to ENCONT.

The result that experience plays a role in the accuracy of L2 voiceless stop production is in line with previous studies that have found that L2 experienced learners outperformed inexperienced learners in the production of similar L2 phones (e.g., Flege, 1987; Flege et al., 1997; Lev-Ari & Peperkamp, 2013; Levy & Law, 2010). The lack of an influence of the L2 on the L1 is not in accordance with the results reported by Flege (1987), as both experienced French learners of English and experienced English learners of French presented L2 influence on their L1 productions of /t/. This difference may stem from the fact that Flege's L1-English and L1-French experienced groups had lived in an L2 setting for 11.7 and 12.2 years, respectively, whereas the most experienced groups tested in this study had spent about 4 years in an immersion setting. Therefore, it might be possible that the experienced learners in this study did not have enough L2 experience – and did not use the L2 sufficiently, in the case of the English groups – for the L2 to influence L1 VOT production.

One issue that emerges both in the analysis of the effect of L2 experience on L2 production and when comparing the different results obtained for L1 and L2 production is the fact that different results are obtained for voiceless and voiced stops. As has been mentioned above, in the case of the English learners, the inexperienced group failed to produce both L2 voiced and voiceless stops with target-like VOT, whereas the most experienced group only failed to produce Spanish voiced stops accurately. As for the moderately experienced group, they were able to produce both voiced and voiceless stops with target-like VOT. Regarding the Spanish learners, all groups failed to produce English /b/ and /g/ with target-like VOT, regardless of L2 experience, although, numerically, the experienced groups tended to approximate their VOT use to that of L1 English speakers to a greater extent than SPINEXPinSP. Moreover, all Spanish groups tended to use prevoicing in voiced stop production in both languages and, in fact, no significant differences between English and Spanish /b/ were found regarding SPEXPinSP and SPINEXPinSP – although they were found in the case of /g/. Therefore, the accuracy of production of voiced stops in the L2 does not seem to depend solely on L2 experience. This finding is in line with previous studies, which have reported the difficulty that L1 speakers of languages that contrast voice-lead and short-lag VOT have in the acquisition of short-lag voiced stops (Caramazza et al., 1973; Hazan & Boulakia, 1993). For example, Hazan and Boulakia (1993) found that early French-English bilinguals produced both L1 and L2 /b/ with prevoicing.

A plausible explanation for the fact that voiceless stops are acquired earlier than voiced stops – and that the acquisition of the latter is not directly linked to L2 experience, in the case of the English groups – may lie in the nature of the L2 phones from an L1 perspective. Spanish and English voiceless and voiced stops have been considered similar phones, as both adhere to Flege's (1987) definition – i.e., 'an L2 phone which is realized in an acoustically different manner than an easily identifiable counterpart in L1'. However, English voiced stops may be realized using voice-lead VOT – just like Spanish voiced stops – in free or individual variation, as it has been reported in the literature (Lisker & Abramson, 1964) and the results of this study show. Therefore, the feature *prevoicing* may already exist in the L1 of an English speaker. As a result, English learners of Spanish may not need to learn a new feature but how to use it contrastively in the L2. Group averages indicate that the English group that used voice-lead VOT the most in the

L1 – i.e., ENEXPinUK – also presented the longest prevoicing in the L2. Thus, it is possible that the use of voice-lead VOT in the L1 was transferred to the L2. However, this hypothesis is not supported by the correlation analysis exploring the relationship between the use of VOT in English and in Spanish voiced stops – i.e., English /b/ and Spanish /b/ and English /g/ and Spanish /g/ –, as no significant results were revealed. This finding leads us back to the hypotheses stated above regarding both voiced and voiceless stops: L2 instruction and being students at the place of testing may have helped ENEXPinUK produce voiced stops more accurately than the most experienced participants, which had received less formal instruction in Spanish – sometimes none – and were not students in the institution were the experiment took place.

As for the Spanish learners of English, the fact that none of the groups succeeded in presenting target-like VOT values in L2 voiced stop production – all differed significantly from ENCONT – may also be explained by the fact that voice-lead VOT is also used in English – although short-lag VOT is more commonly used. That is, English contrasts aspiration – used in voiceless stops – and the lack of it – in voiced stops –, which includes both short-lag and voice-lead VOT. As a result, Spanish learners of English may use their L1 categories – which present prevoicing – to produce English voiced stops and may not need to create a new category, given that their L1 category does already qualify as a possible production for voiced stops in English.

Summary

It appears that L2 experience has a greater effect on the production of L2 stops than on the production of L1 stops. In the case of the Spanish groups, L2 experience appears to modulate L2 production accuracy, as the experienced groups outperformed the inexperienced group. As for the English learners of Spanish, L2 experience – understood in this study as amount of time spent in the target language country and setting at time of testing – alone does not seem to fully predict accuracy in the production of voiced stops. In spite of the fact that the inexperienced learners differed the most from Spanish controls, the group with intermediate experience – and not the highly experienced group – presented the VOT values that were the closest to those of controls, especially regarding voiced stops. Other factors, such as amount of L2 instruction, L2 use and their relationship with the institution where the experiment was conducted may have contributed to the fact that ENEXPinUK obtained the most target-like VOT values. It should also be noted that voiced stops posed a greater difficulty to be produced with target like VOT values than voiceless stops, particularly in the case of the Spanish learners of English. Regarding L2 category formation, all English groups were able to distinguish English and Spanish categories – i.e., they used greater VOT values in English than in Spanish – even though, as has been mentioned, they were not always target-like.

7. The relationship between perception and production of stops

7.1. Data analysis

Perception and production results were compared so as to determine whether there was a relationship between the two dimensions. In order to do so, each participant's perceptual category boundaries were compared to the mean VOT that they produced. Given that for each point of articulation there was one measure for perception – i.e., /p/-/b/ and /k/-/g/ category boundary – and two for production – i.e., mean /p/ VOT and mean /b/ VOT; mean /k/ VOT and mean /g/ VOT – the perceptual category boundary was contrasted with both the voiced and the voiceless productions sharing the same place of articulation. If there was a relationship, we would expect that those participants and groups that needed greater VOT values to start perceiving a voiceless stop – i.e., presented a later category boundary – would also have longer VOT values in the production of voiceless stops. Similarly, an earlier VOT boundary should be related to smaller VOT values in voiced sounds. Separate two-tailed Pearson correlation tests were conducted for each control group. Regarding the learner groups, the relationship between perception and production was examined considering all three learner groups together and separately. All three groups with same L1 were first tested together so as to assess the relationship between the two modalities with a sample that was sufficiently large. Then separate tests were carried out on each group to explore if groups differed in this respect, even if the sample size was small

As mentioned previously, the data of some participants were lost. The perception results in Spanish of one participant in the ENEXPinSP group were not properly saved, resulting in 10 cases in Spanish and 11 in English for this group, as well as 30 cases in Spanish and 31 in English for all English learners. The production data of one participant in the ENEXPinUK group were also lost. In order to be able to compare the results obtained in the two experiments, the results from those participants which did not present values for the two dimensions were discarded. For this reason, there may be small numerical differences between the group means reported in the current chapter and those provided in Chapters 5 and 6.

7.2. The relationship between perception and production of stops in the L1

In order to examine whether there was a relationship between the perception and production of stops in the L1, the category boundary and the VOT mean obtained by each group and participant were compared for each place of articulation. Following the SLM's assumptions that perception and production are related, it was expected that the two measures would be positively correlated, that is, that participants with a late category boundary would also present a large VOT mean. First, the relationship between the two measures was compared for the control groups. Table 7.1 presents ENCONT's mean perceptual category boundaries for each pair and the mean VOT values for each category. The individual results obtained by each English participant in their L1 are provided in Table AE.1 in Appendix E.

The expected results were not found, as, with a few exceptions, an inspection of the individual data suggested no relationship between the perception and production values. For example, ENCONT1, who obtained the second earliest category boundary (11 ms), produced /p/ with one of the longest mean VOT observed (53 ms). Similar results were found with the other stops. Accordingly, correlation tests revealed no significant results for any category (*p* > .05, see Table 7.3, which present the results obtained for all the tests conducted on the L1).

Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
ENCONT	14 (2.5)	52 (20)	1 (25)	29.7 (4.3)	72 (19)	14 (38)

Table 7.1. ENCONT's category boundaries and VOT means. Standard deviations are provided in parentheses.

Regarding Spanish controls, just as in the case of ENCONT, no clear relationship between the two variables was observed in any of the stop categories (see Table 7.2 for the group means). For instance, SPCONT2 and SPCONT11 obtained very similar /k/-/g/ category boundaries – namely 0 ms and 2.1 ms, respectively –, which were the two earliest boundaries in the group. However, their production of /k/ differed to a greater extent. SPCONT2 produced /k/ with one of the smallest VOT means (15 ms), whereas SPCONT11 had one of the greatest VOT means for /k/ (35 ms). Correlation tests were conducted to compare each stop's category with their corresponding category boundary. None of the tests revealed a significant result (*p* > .05, see Table 7.3).

Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
SPCONT	3.4 (5.4)	8.6 (4.3)	-78 (29)	14.8 (10.3)	34 (11)	-77 (29)

Table 7.2. SPCONT's category boundaries and VOT means in ms. Standard deviations are provided in parentheses.

Groups	*N*	/p/-/b/ and /p/	/p/-/b/ and /b/	/k/-/g/ and /k/	/k/-/g/ and /g/
ENCONT	9	*r* = .524, *p* = .147	*r* = -.387, *p* = .311	*r* = .197, *p* = .644	*r* = -.480, *p* = .191
L1-English L2-Spanish groups together	31	*r* = -.121, *p* = 518	*r* = -.300, *p* = .101	*r* = -.053, *p* = .776	*r* = .161, *p* = .388
ENINEXPinUK	11	*r* = -.428, *p* = .189	*r* = .438, *p* = .178	*r* = -.425, *p* = .193	*ρ* = -.182, *p* = .593
ENEXPinUK	9	*r* = .017, *p* = 966	*r* = -.633, *p* = .067	*r* = .341, *p* = .369	*r* = .240, *p* = .534
ENEXPinSP	10	*r* = .210, *p* = .535	*r* = -.162, *p* = .634	*r* = .341, *p* = .369	*r* = .240, *p* = .534
SPEXPinUK	9	*r* = -.669, *p* = .049 *	*r* = -.394, *p* = .294	*r* = -.453, *p* = .220	*r* = -.037, *p* = .925
SPEXPinSP	10	*r* = .218, *p* = .545	*r* = -.418, *p* = .229	*r* = .282, *p* = .430	*r* = -.174, *p* = .63
SPINEXPinSP	10	*r* = .096, *p* = .794	*r* = -.514, *p* = .128	*r* = -.264, *p* = .462	*r* = -.244, *p* = .497
L1-Spanish L2-English groups together	29	*r* = -.019, *p* = .921	*r* = -.305, *p* = .108	*r* = -.201, *p* = .296	*r* = .058, *p* = .764
SPCONT	10	*r* = -.373, *p* = .289	*r* = .154, *p* = .671	*r* = -.567, *p* = .088	*r* = -.344, *p* = .330

Table 7.3. Summary table reporting the results obtained in the correlation tests comparing the mean category boundaries and VOT productions for each stop in the L1 for the control groups, all learner groups sharing the same L1 (together) and each learner group (separately). Significant results are shown with an *.

Next, the relationship between category boundary and VOT production in the L1 was examined for the L2-learner groups. When it comes to the English learners, it was observed that the groups with more L2 experience seemed to need increasingly more aspiration to identify voiceless stops – i.e., presented later category boundaries –, although the differences between groups in ms were relatively small, namely less than 2 ms (see Table 7.4 for the mean category boundaries and mean VOT productions for each L1-English group and Table AE.1 in Appendix E for individual results). However, the opposite trend was observed in the production of /p/: ENINEXPinUK had the greatest VOT mean, whereas ENEXPinSP had the lowest value. Still, just as in the case of the /p/-/b/ boundary, the differences between groups were very small. Regarding /b/, the group that produced the lowest VOT mean (ENEXPinUK, with -9 ms) was not the group that obtained the earliest category boundary for /p/-/b/. On the contrary, ENEXPinUK had a /p/-/b/ category boundary of 15.8 ms, which was very close to that of ENEXPinSP (16.1 ms), the latest one observed in the English testing groups in their L1. Contrary to expectations, ENINEXPinUK, who obtained the earliest category boundary (14.5 ms), produced the greatest VOT mean for /b/ (8 ms).

In the case of velars, the group with the highest category boundary (ENEXPinSP, with 29.4 ms) was the group with the lowest VOT mean for /k/ (72 ms). ENINEXPinUK and ENEXPinUK obtained very similar category boundaries (27.3 ms and 27 ms, respectively). ENINEXPinUK also produced /k/ with slightly greater VOT values (79 ms) than ENEXPinUK (76 ms). Regarding /g/, some consistency was found. ENEXPinUK presented the lowest VOT value for /g/ production (1 ms), as well as the lowest category boundary among the L1-English groups (27 ms). However, the group that obtained the latest category boundary (ENEXPinSP, with 29.4 ms) was not the group that obtained the greatest VOT mean for /g/. Instead, it was ENINEXPinUK (20 ms), while ENEXPinSP obtained a mean VOT for /g/ of 6 ms. Hence, when exploring the group data, no clear pattern of a relationship between the two modalities emerged. Correlation tests including all English learners of Spanish, which compared the mean VOT with the corresponding category boundary, were carried out for each stop category. None of the tests revealed a significant correlation ($p > .05$, see Table 7.3).

Consistency in the relationship between the perceptual and the production variables within each group was also analyzed (see Table AE.1 in Appendix E for individual results). No clear relationship between boundary and stop production was observed for any category within any L1-English group. As a matter of fact, when it comes to the production of /k/ and the perception of /k/-/g/, some participants in the

ENINEXPinUK group were found to perform contrary to expectations, as, in most cases, participants with the earliest category boundaries were not the ones that produced the shortest VOT means. A correlation test for each category was conducted on each group and no significant correlations were found (*p* > .05 in all cases, see Table 7.3).

Group	/p/-/b/ boundary	/p/ VOT	/b/ VOT	/k/-/g/ boundary	/k/ VOT	/g/ VOT
ENINEXPinUK	14.5 (3.1)	52 (20)	8 (6)	27.3 (8.5)	79 (18)	20 (23)
ENEXPinUK	15.8 (4.1)	50 (18)	-9 (40)	27 (4.7)	76 (19)	1 (48)
ENEXPinSP	16.1 (2.4)	46 (18)	0 (35)	29.4 (3.7)	72 (21)	6 (41)

Table 7.4. L1-English groups' English category boundaries and VOT means in ms. Standard deviations are provided in parentheses.

The relationship between category boundary and VOT production in the L1 was also analyzed in the Spanish learner groups (see Table 7.5 for means and Table AE.2 in Appendix E for individual results). Regarding the general results obtained by each group, a straightforward relationship between /p/-/b/ boundary and stop production was only observed in two cases. For instance, in the case of /p/, SPEXPinUK was the group that obtained the latest category boundary (11.2 ms), as well as the greatest VOT mean (10 ms). Moreover, SPEXPinSP, who had the earliest category boundary /p/-/b/ (7.8 ms), also produced the lowest VOT mean for /b/ (-80 ms). No other relationship between category boundary and stop production was found regarding bilabial stops, as those groups that presented late category boundaries did not use great VOT values. As for velar stops, all Spanish testing groups obtained very similar /k/-/g/ boundaries (see Table 7.5) but produced /k/ differently. SPEXPinUK had the greatest VOT mean for /k/ (44 ms), whereas SPINEXPinSP presented the smallest VOT mean (32 ms). Although all groups obtained a negative VOT mean for /g/, they differed numerically. SPEXPinUK and SPEXPinSP presented somewhat longer prevoicings (-77 ms and -80 ms, respectively) than SPINEXPinSP (-68 ms), but, as mentioned above, their perceptual category boundaries for velar stops were very close. Not surprisingly, the two-tailed correlation tests including all L1-Spanish L2-English groups failed to find a significant correlation between the two measures (see Table 7.2 for results).

An individual analysis was conducted also for each group (see Table AE.2 in Appendix E for individual results). In most cases, a relationship between the two measures was not observed at an individual level. Within the SPINEXPinSP group, no clear pattern was found in any of the stop categories. As for SPEXPinSP, participant SPEXPinSP06 obtained an early /p/-/b/ category boundary (7.7 ms) and produced /p/ with a short VOT mean (11 ms). However, SPEXPinSP11 obtained the earliest category mean for the /p/-/b/ contrast (3.4 ms) and produced /p/ with the greatest VOT mean (70 ms). No clear patterns were observed in this group regarding the other stop categories. In the case of SPEXPinUK, most participants with an early /p/-/b/ category boundary presented short VOT productions for /p/ (SPEXPinUK5, SPEXPinUK6 and SPEXPinUK8), but no clear pattern was found regarding participants with greater VOT means. No straightforward relationship was observed for the other categories. The correlation tests conducted on each group revealed only one significant result: SPEXPinUK's VOT production for /p/ and their /p/-/b/ boundary were found to be negatively correlated, although barely reaching significance. That is, contrary to predictions, the later the category boundaries were, the less VOT was used in production. None of the other correlations reached significance ($p > .05$).

Group	/p/-/b/ boundary	/p/ VOT	/b/ VOT	/k/-/g/ boundary	/k/ VOT	/g/ VOT
SPINEXPinSP	8.4 (3.2)	8 (5)	-67 (36)	21.5 (5)	32 (11)	-68(19)
SPEXPinSP	7.8 (3.2)	10 9)	-80 (26)	20.8 (2.2)	37 (12)	-80 (51)
SPEXPinUK	11.2 (3.5)	10 (5)	-68 (34)	21.3 (3.6)	44 (15)	-77 (40)

Table 7.5. L1-Spanish groups' Spanish category boundaries and VOT means in ms. Standard deviations are provided in parentheses.

7.3. The relationship between perception and production in the L2

The relationship between perception and production in the L2 was also investigated in the same manner that it was examined for the L1 – i.e., at a group level and at an individual level (see Tables AE.2 and AE.3 in Appendix E for individual results). Regarding the L1-English groups, no straightforward relationship between the two measures was generally observed, except in the case of ENEXPinSP's /p/-/b/ boundary (14.9 ms) and mean VOT for /p/ (14 ms), which were both the smallest. In all other cases, the means obtained by the testing groups did not match the expected pattern, given that the groups that obtained the earliest category boundaries were not the ones that had the lowest VOT means (see Table 7.6). As a matter of fact, in some occasions, the opposite trend was observed. For instance, in the case of /b/, the group with the lowest mean was ENEXPinUK (-49 ms), who, contrary to expectations, also presented the latest category boundary. Only the correlation tests conducted on /k/-/g/ boundary and /g/ production yielded a significant result, although the correlation was moderate, whereas no other significant results were revealed for the English learners of Spanish (See Table 7.7).

Group	/p/-/b/ boundary	/p/ VOT	/b/ VOT	/k/-/g/ boundary	/k/ VOT	/g/ VOT
ENEXPinSP	14.9 (3.1)	14 (10.7)	-35 (45)	28 (2.8)	42 (13)	-32 (35)
ENEXPinUK	16.5 (4.7)	17 (10.6)	-49 (51)	26.5 (4.7)	45 (14)	-55 (51)
ENINEXPinUK	15.7 (4.1)	28 (22.9)	-13 (45)	25.2 (8.3)	55 (21)	-6 (52)

Table 7.6. L1-English groups' Spanish category boundaries and VOT means in ms. Standard deviations are provided in parentheses.

Groups	*N*	/p/-/b/ and /p/	/p/-/b/ and /b/	/k/-/g/ and /k/	/k/-/g/ and /g/
All L1-English L2-Spanish groups	30	*r* = -.049, *p* = .793	*r* = -.240, *p* = .193	*r* = -.105, *p* = .582	*r* = .362, *p* = .049*.
ENINEXPinUK	11	*r* = -.018, *p* = .958	*r* = -.603, *p* = .050	*r* = .596, *p* = .053	*ρ* = .227, *p* = .502
ENEXPinUK	9	*r* = -.452, *p* = .222	*r* = .041, *p* = .917	*r* = .635, *p* = .066	*r* = -.150, *p* = .699
ENEXPinSP	10	*r* = .549, *p* = .100	*r* = -.011, *p* = .976	*r* = -277, *p* = .439	*r* = .607, *p* = .048
SPEXPinUK	9	*r* = .228, *p* = .555	*r* = -.074, *p* = .850	*r* = -.140, *p* = .236	*r* = -.157, *p* = .687
SPEXPinSP	10	*r* = -.440, *p* = .203	*r* = .696, *p* = .025 *	*r* = -.529, *p* = .116	*r* = .232, *p* = .519
SPINEXPinSP	10	*r* = -.061, *p* = .868	*r* = -.317, *p* = .371	*r* = -.070, *p* = .847	*r* = -.634, *p* = .049 *
All L1-Spanish L2-English groups	29	*r* = -.047, *p* = .810	*r* = .242, *p* = .206	*r* = -.079, *p* = .684	*r* = -.104, *p* = .591

Table 7.7. Summary table reporting the results obtained in the correlation tests comparing the mean category boundaries and VOT productions for each stop in the L2 for, all learner groups sharing the same L1 (together) and each learner group (separately). Significant results are shown with an *.

Individual results were also analyzed within each group (see Table AE.3 in Appendix E for individual results). In most cases, the relationship between the two variables appeared to be arbitrary. Still, regarding ENEXPinSP, there seemed to be some consistency in the relationship between /p/-/b/ category boundary and /p/ production. For instance, ENEXPinSP01 and ENEXPinSP06, who obtained some of the earliest category boundaries within their group (11.4 ms and 13.5 ms, respectively), also produced the shortest VOT means for /p/ (9 ms and 9.3 ms, respectively). No other straightforward relationships were observed within this group for other categories and no significant correlations were found (See Table 7.7). As for ENEXPinUK, the relationship between category boundary and VOT production seemed to be arbitrary in the case /p/, /b/ and /g/. Regarding /k/, a trend was observed: most participants with early /k/-/g/ category boundaries also presented short VOT means for /k/, whereas those presenting late category boundaries also tended to present long VOT values. Finally, ENINEXPinUK appeared to be the group with the most consistent relationship between the two dimensions, except in the case of /p/, where no straightforward relationship was observed. However, it should be noted that the relationship between /p/-/b/ boundary and /b/ production was the opposite to the expected one. Participants with a later category boundary were the ones that produced the longest prevoicings (i.e., ENINEXPinUK6 and ENINEXPinUK13 had a /p/-/b/ category boundary of 23.7 ms and 19.7 ms, respectively, as well as a VOT mean for /b/ of -62.7 ms and -100.1 ms, respectively). A similar pattern was observed for /k/. Conversely, regarding /g/, the participant with the earliest /k/-/g/ category boundary (ENINEXPinUK, with 7.7 ms) also produced, the longest prevoicing (-119 ms) and one of the participants with the latest boundary (ENINEXPinUK7, with 34.2 ms) presented one of the greatest VOT means (23 ms). That is, as the positive result of the correlation indicated (see Table 7.7), the later the category boundary of a learner was, the more VOT was used in production. In spite of the fact that there were a few apparent relationships between the two measures, the correlation tests did not yield any significant result for any of the L1-English groups, although a marginal result was found for ENINEXPinUK's and ENEXPinUK's /k/-/g/ boundary and /k/ VOT production (see Table 7.7).

Table 7.8 illustrates the mean category boundaries and VOT productions obtained by the Spanish speakers in English. The group that obtained the greatest /p/-/b/ category boundary (12.1 ms), as well as the highest VOT mean for /p/ (35 ms), was SPEXPinUK.

Similarly, SPINEXPinSP obtained the lowest /p/-/b/ category boundary (9.3 ms) and the shortest VOT mean for /p/ (23 ms). However, the opposite was true regarding /b/. In other words, the group that obtained the earliest category boundary (SPINEXPinSP, with 9.3 ms) did not produce the longest prevoicings (-66 ms). Instead, it was produced by SPEXPinSP, who presented a VOT mean for /b/ of -71 ms. SPEXPinUK presented both the latest /p/-/b/ boundary (12.1 ms) and the greatest VOT mean (-43 ms).

Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
SPEXPinUK	12.1 (2.6)	35 (21)	-43 (49)	24.5 (2.4)	73 (22)	-24 (53)
SPEXPinSP	10.1 (4.1)	32 (25)	-71 (45)	24.9 (3.9)	70 (22)	-49 (41)
SPINEXPinSP	9.3 (3.6)	23 (22)	-66 (44)	22.8 (4.4)	53 (30)	-50 (46)

Table 7.8. L1-Spanish groups' English category boundaries and VOT means in ms. Standard deviations are provided in parentheses.

There seems to be a somewhat more straightforward relationship between /k/-/g/ boundary and /k/ production. SPINEXPinSP obtained the earliest category boundary (22.8 ms) as well as the lowest VOT mean for /k/ (53 ms). SPEXPinUK and SPEXPinSP obtained very similar /k/-/g/ category boundaries (24.5 ms and 24.9 ms, respectively), as well as very close mean VOT values for /k/ (73 ms and 70 ms, respectively). Regarding /g/, the earliest category boundary (20 ms) and the smallest mean VOT for /g/ (-50 ms) were both obtained by SPINEXPinSP. The greatest mean VOT value (-24 ms) was produced by SPEXPinUK. Although SPEXPinUK did not obtain the latest category boundary (24.5 ms), it was very close to that of SPEXPinSP's (24.9 ms). A correlation test for each stop category was conducted considering all Spanish groups and, in spite of the apparent numerical relationship between the two measures at a group level, none of the tests revealed a significant result (p > .05 see Table 7.7).

It was also explored whether there was a more consistent relationship between the two variables at an individual level within each group (see Table AE.2 in Appendix E for all individual results). Just a few straightforward relationships – according to expectations – were observed. For example, within the inexperienced group, SPINEXPinSP10 was the only participant that performed as expected: they presented one of the latest /p/-/b/ category boundaries in the group (11.7 ms), as well as some of the greatest VOT means for /p/ and /b/ (46.8 ms and -14.4 ms, respectively). In some cases, participants performed contrary to expectations. For instance, SPEXPinSP10 and SPEXPinSP11 presented the lowest /k/-/g/ category boundaries (20.2 ms and 22 ms, respectively), but produced /k/ with the greatest VOT values in their groups (79 ms and 106.4 ms, respectively). A somewhat more straightforward relationship was observed in the SPEXPinSP group, although only regarding /p/-/b/ boundary and /b/ production. Participants with earlier category boundaries tended to present smaller VOT means, whereas those with a later category boundary often produced /b/ with greater VOT means. As for SPEXPinUK, no clear patterns were found. The correlation tests conducted on each group revealed a significant positive correlation between the VOT production for /b/ and /p/-/b/ category boundary in the case of SPEXPinSP (r = .696, n = 10, p = .025). That is, a later boundary for the bilabial contrast was correlated to greater VOT values – i.e., less prevoicing. However, SPINEXPinSP's /g/ production and /k/-/g/ boundaries were, contrary to predictions, found to be negatively correlated at a significant level – i.e., later boundaries were correlated to more prevoicing (r = -.634, n = 10, p = .049). All other comparisons were not significant (see Table 7.7). It should be noted, though, that sample size in these tests was very small and, therefore, results need to be interpreted with caution.

7.4. Summary of results

The relationship between the perception and production of L1 and L2 stops was investigated by comparing the category boundaries and mean VOT values of each participant for each category – i.e., /p/, /b/, /k/ and /g/. A series of two-tailed Pearson correlations comparing the two measures were conducted on each control group. An analysis was also carried out on the L1-English and the L1-Spanish participants, considering all groups with the same L1 together. Moreover, separate correlations were conducted on each group in order to assess whether L2 experience modulated the alignment between perception and production. Generally, no link between the two dimensions was found in either language: only four correlations – two positive, in agreement with expectations, and two negative, contrary to predictions – were found. Regarding the L1, no relationship between category boundary and mean VOT use was observed in the control groups for any category. The fact that no link was established for the control groups already suggested that the relationship between perception and production would not be straightforward in the testing groups. In fact, only one correlation was observed in the L1 of the learner groups and it was in the opposite direction to predictions. As for the L2, no apparent relationship was revealed in most cases. Only three significant correlations were found and they were in opposite directions – i.e., in two cases late category boundaries were correlated to longer VOT values in production, whereas, in the other case, late category boundaries were correlated to shorter VOT values. Therefore, there does not seem to be a straightforward relationship between category boundary and VOT production, neither in the L1 nor in the L2. Furthermore, L2 experience does not appear to modulate the relationship between the two modalities, as no clear pattern was observed within groups.

7.5. Discussion

The relationship between the results obtained in the perception and production experiments was examined by comparing the category boundaries and the mean VOT values obtained by each participant. It was expected that those participants that presented late category boundaries would also use great VOT values in production. As has already been mentioned, only a total of four significant correlations – out of which two were in the opposite direction to predictions – were found, considering both the L1 and the L2. That is, no clear link between the two measures was found for any group.

The failure to establish a straightforward relationship between the two dimensions is in agreement with previous findings. With only a few exceptions (e.g., Casillas, 2019; Newman, 2003), most studies examining consonants and, particularly, stops have failed at establishing a clear link between the two dimensions, both in the L1 and in the L2 of adult speakers (e.g., Bailey & Haggard, 1973; Shultz et al., 2012; Williams, 1977). Regarding the L1, Bailey and Haggard (1973) and Shultz et al. (2012) failed to find a relationship between VOT weighting in perception and production. To my knowledge, only one study examining VOT found a straightforward correlation between perception and production at an individual level in the L1 (Newman, 2003) and one in the L2 (Casillas, 2019). In this case, VOT production was correlated to the amount of VOT in the stimuli used in a goodness rating task. In the case of the L2, an early study by Williams (1977) showed that Spanish-English bilinguals who produced target-like VOT values in the L2 did not always succeed in perceiving English stops accurately, as shown by an identification and a discrimination task. Even though the outcome of the present study replicates the most common result regarding the relationship between VOT production and perception, it should be noted that the fact that a variety of different tasks, measures and populations have been tested in previous research does not facilitate valid comparisons between studies.

In spite of the fact that this thesis did not show a clear alignment between the perception and production measures used, this result may not necessarily invalidate a possible link between the two modalities. Methodological issues may have also resulted in an apparent unrelatedness of the two dimensions. One plausible explanation that applies to both the control and testing groups might be that the perception and the production measures are not comparable. The fact that the perception results were based on a limited number of steps on a continuum and the production results on actual productions resulted in greater variation in the latter than in the former. Moreover, the production measure was a mean of all the productions of the same speaker, which may as well differ from one another, that is, the same speaker may realize the same stop differently, even if it is in the same context.

Another aspect that should be considered, especially regarding L2 learners, is the context in which the target phones were embedded in each task. In the case of perception, the stimuli used were single syllables presented in isolation, whereas the production task involved words in the context of a complete meaningful sentence. Thus, the two tasks may not be comparable when it comes to lexical load. It should also be acknowledged that the perceptual stimuli were more meaningful in English and in Spanish. Regarding /pi/, the sequence carries meaning in both languages, but is more commonly used in English – “pea”/ “pee” – than in Spanish – the mathematical number π, “pi”. As for /bi/, both languages have a lexical representation for the sequence – “bee” in English and “vi”, “I saw”, or “bi”, the abbreviation for bisexual, in Spanish. However, the first Spanish word is spelled with <v>, which may hinder the activation of this lexical item, as the instructions of the task specified that the stimuli included the sound represented by “b” –

or “p”, for that matter – being the response alternative provided ‘*’b’ como en bicho*’ (‘’b’ as in bug’). Moreover, “bi” may not be as frequently used in Spanish as the word “bee” is in English. As for /ki/ and /gi/, no lexical representation for these sequences is found in Spanish, whereas it is in English regarding /ki/ – i.e., “key”. As a result of the greater meaningfulness of the stimuli in English than in Spanish, English may have been activated in the Spanish perception task since, as Grosjean (2001) claims, the inclusion of lexical elements of the language that is not being used – or simply the belief on the part of the listener that they may be exposed to it – would change the language mode from monolingual to bilingual. Another methodological aspect that should be acknowledged is that, due to the two populations and languages that were tested, the Spanish task included aspirated tokens – a feature that is not present in Spanish. Just as in the case of the lexical load, the inclusion of a phonetic feature that is specific to English when testing Spanish perception may have favoured the activation of an English mode and, thus, an English-like categorization.

Considering the methodological issues mentioned above, the apparent mismatch between perception and production can be interpreted in SLM terms or, alternatively, following the L2LP. The SLM’s merger hypothesis acknowledges the possibility of having one single merged phonological category for the two languages which may show traits of both the L1 and the L2. However, the category can be implemented differently at a phonetic level according to the nature of the input received. That is, if the input resembles the L1 to greater extent, the phonetic realization of the merged category will be more L1-like, whereas the opposite will be true if the input received is more L2-like. Thus, a somehow English-like input in both the English and Spanish perception tasks may have influenced their categorization towards an English-like perception. As a matter of fact, as was shown in sections 5.4 and 6.3, L2 learners were more successful at separating the L1 from the L2 in the production task than in the perception task. Furthermore, the English groups categorized L1 stops according to English VOT values in both languages, whereas the Spanish learners perceived both languages with numerically intermediate values, which turned out to be significantly different from both English and Spanish controls in the case of the most experienced group. In short, different phonetic realizations of the same category were implemented in perception and in production according to the nature of the input received in each experiment, as the production experiment may have set the correct language more appropriately than the perception task.

A similar explanation can be inferred from the L2LP. Even though Escudero's (2005, 2009) model considers that the L1 and the L2 phonological systems are separate, it acknowledges the possibility of an intermediate L1-L2 perception. In this case, the intermediate performance is explained following Grosjean's (2001) Language Mode theory: given that the stimuli used in the perception experiment were more meaningful in English – and, in fact, some presented exclusively English-like phonetic features – English may have been activated in the Spanish perception task to a greater extent than it was in the production experiment. That is, language mode activation might have been better controlled in the production task than in the perception task, yielding non-comparable results for the two experiments. However, it should be noted that the potential explanations related to input and language mode activation would only apply to the L2 learners, not to control groups.

Another important issue that needs to be considered is the order in which each modality is acquired in the L2 – i.e., whether perception or production develop earlier or whether both develop simultaneously. Even though no significant relationship between the two measures was revealed by the statistical analysis, the results obtained in each experiment separately – i.e., the perception and the production experiment – indicate that

L2 learners were more successful at producing L2 stops than at perceiving them. This was true especially in the case of the English groups, who presented L1-like categories in the two languages in perception, but separate L1 and L2 categories – which were either intermediate between the two languages or target-like – in production. This finding challenges the SLM's and PAM's assumption that L2 perception precedes production. A somewhat less clear pattern was observed in the Spanish groups. In both experiments, the experienced groups outperformed the inexperienced one but the ENEXPinSP were more accurate in perception – they did not differ from ENCONT – than in production – they differed from ENCONT regarding /p/, /b/ and /g/. However, the fact that the Spanish groups were generally better able to distinguish L1 and L2 stops in production than in perception suggests that production accuracy may precede perception. All in all, it appears that the English learners were more successful at producing L2 stops accurately – and differently from the L1 – than at perceiving them, whereas a pattern regarding what modality is mastered first was not so clearly observed for the Spanish groups. However, as has been mentioned above, the comparisons of the results obtained in each experiment need to be made with caution due to the possible methodological limitations related to language mode control.

Summary

In short, no clear link between perception and production has been found in either language. This result is in agreement with previous studies that have failed to establish a relationship between the perception and the production of stops (e.g., Bailey & Haggard, 1973; Lev-Ari & Peperkamp, 2013; Zampini, 1999). However, it is possible that methodological limitations related to task design and the measures used may have obscured a possible relationship between the two dimensions. Furthermore, the outcome of this study regarding the L2 can be interpreted in terms of L2 speech acquisition models that consider that the performance of an L2 learner may differ according to the input received (the SLM) and the degree of activation of each language (the Language Mode theory and the L2LP). As for the order in which the modalities develop, it appears that a more target-like performance was achieved in production, particularly, in the case of the English groups, challenging the SLM's claim that perception precedes production.

8. The effect of individual factors on stop perception and production

8.1. Data analysis

Previous studies have shown that individual factors in L2 learners may influence their perception and production (e.g., Piske et al., 2001). Information about participants' linguistic background and language use was collected by means of a questionnaire, which was based on the Bilingual Language Profile (BLP, Birdsong et al., 2012). This information will be explored in this chapter in light of the results obtained in the experiments. The questionnaires revealed some variability within groups when it comes to the amount of time spent in an immersion setting. A continuous variable – i.e., number of months spent in an L2 country – was used to measure L2 experience in order to assess whether these individual differences had an impact on VOT perception and production. However, it should be noted that the number of months spent in an L2 country was not the only difference between the groups, as setting – i.e., immersion or instructional – at the time of testing also differed between groups.

As was described in the Methodology (Chapter 4) , participants did not only differ in terms of L2 experience but also regarding other factors, such as language dominance, years of formal L2 instruction, age of L2 learning – henceforth L2 AOL – and weekly L2 use (see Appendix F for individual results). This section will examine the effect of these factors on L1 and L2 stop perception and production. Regarding perception, a regression analysis (dependent variable: category boundary; independent variables: language dominance, years of L2 instruction, L2 AOL, percent weekly L2 use and number of months spent in an L2 country) was conducted for each point of articulation in each language. Similarly, a regression analysis was carried out to assess the effect of these variables on production (dependent variable: mean VOT; independent variables: language dominance, years of L2 instruction, L2 AOL, percent weekly L2 use and number of months spent in an L2 country). The general effect of individual differences was tested by simultaneously assessing all coefficients – i.e., independent variables. The significance of each independent variable was analyzed by conducting individual regression coefficients. All learner groups sharing the same L1 were included in the same analysis, that is, the variable group was not considered in the analysis.

The dominance score obtained in the BLP questionnaire ranges from -218 to 218. A score of 0 indicates perfectly balanced bilingualism, whereas positive values show English dominance and negative values Spanish dominance. In other words, the greater the score, the more English dominant and the smaller the score, the more Spanish dominant. It was expected that a greater English dominance would result in later category boundaries in perception and greater VOT values in production, whereas the opposite was expected in the case of Spanish dominance. Similarly, it was predicted that a greater amount of years of L2 instruction, weekly L2 use and number of months in an immersion setting would result in more-target like L2 perception and production, but in a less native-like performance in the L1. Finally, an earlier age of learning was expected to be related to a more target-like perception and production in the L2, but in a less native-like performance in the L1.

8.2. The effect of individual differences on stop perception

The following sections present the results obtained in the analysis of the effect of individual factors on the stop perception of English learners of Spanish and Spanish learners of English. The results obtained for the L1 will be presented first, followed by the results for the L2.

8.2.1. The effect of individual differences on L1 perception

The effect of individual factors on the L1 perception of the English learners was explored in the first place. Table 8.1 illustrates the mean dominance score, years of L2 instruction, L2 AOL and % weekly L2 use per group, (see section 4.1 for a description and Table AF.1 in the Appendix F for all participants' values). In short, the most experienced group (ENEXPinSP) was the group that had received the least amount of L2 instruction (*M* = 2.4 years; range = 0 – 7) and had started learning Spanish at the most advanced age (*M* = 18.9; range = 14 – 20). As a matter of fact, they were found to be the most English-dominant group (*M* = 116.5; range = 79.9 – 165.5). ENEXPinSP's percent L2 use was the greatest (*M* = 16.3; range = 10 – 40), but it was still relatively low and similar to that of the learners living in the UK. As for ENEXPinUK, they had learnt Spanish for the longest period of time (*M* = 7.5 years; range = 3 – 18) and were found to be the least strongly L1 dominant group (*M* = 106; range = 74.3 – 129.9), tightly followed by the inexperienced group, although both groups were still found to be clearly L1-dominant. ENINEXPinUK used Spanish the least (*M* = 12.3%; range = 3 – 27%) and started to learn Spanish at the youngest age (*M* = 13.4; range = 10 – 20). It should be noted that group was not considered a factor in the analysis and all L1-English speakers were included in the same test (see Appendix F for individual results). Still, the different characteristics of each group are relevant to interpret the results obtained in the perception and production experiments, especially considering those cases in which L2 experience does not seem to fully account for group differences.

A multiple regression analysis was conducted in order to assess the effect of language dominance, years of L2 instruction, L2 AOL, L2 use and number of months spent in an L2 setting (independent variables) on L1 category boundary (dependent variable) on the English groups. Individual factors were not found to significantly predict /p/-/b/ category boundary [F (5, 18) = .755; p = .549; R^2 = .173] or /k/-/g/ category boundary [F (5, 18) = .428; p = .823; R^2 = .106]. Moreover, none of the individual independent variables were revealed to have a significant effect on L1 category boundary (p > .05 in all cases).

Group	Dom.	L2 inst.	L2 AOL	% L2 use	Months in L2 setting	L1 /p/-/b/ boundary	L1 /k/-/g/ boundary
ENINEXPinUK	108.7 (27.3)	3.8 (3.3)	13.4 (3.7)	12.3% (8)	-	14.5 (3.1)	27.3 (8.5)
ENEXPinUK	106 (21.7)	7.5 (5.3)	15.5 (3.7)	15% (6.5)	9.4 (4.2)	15.9 (3.9)	27.2 (4.5)
ENEXPinSP	116.5 (26)	2.4 (2.9)	18.9 (2.1)	16.3% (10.5)	50.7 (27.3)	16.1 (2.4)	29.4 (3.7)

Table 8.1. L1-English groups mean dominance score (dom.), years of L2 instruction (L2 inst.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and English category boundaries. Standard deviations are provided in parentheses.

As explained in the Methodology (see section 4.1), the Spanish learners of English were more homogeneous when it comes to individual factors. All groups started learning English at a mean comparable age (between 5.4 and 6.1) and had learned it for a similar amount of time (between 13.2 and 13.9 years), probably because English is a mandatory subject in the curriculum of Spanish schools. As expected, SPEXPinUK were the most balanced group in terms of language dominance (M = -41.5; range = -92.6 – 0.6), although still Spanish-dominant, and used English the most during the week (M = 54.6%; range = 20 – 93%). Table AF.1 in Appendix F shows individual results. The effect of individual factors on the L1 category boundaries of the Spanish learners was investigated by means of a multiple regression analysis, as was done for the L1-English groups. The analysis in this case also failed to reveal a significant effect of any variable on /p/-/b/ category boundary [*F* (5, 21) = 1.455; *p* = .246; R^2 = .257] and on /k/-/g/ boundary [*F* (5, 20) = .461; *p* = .800; R^2 = .103]. Similarly, none of the individual regression coefficients conducted on each factor was found to have any significant effect on L1 category boundary when tested separately (*p* > .05 in all cases).

Group	Dom.	L2 inst.	L2 AOL	% L2 use	Months in L2 setting	L1 / p/-/b/ boundary	L1 /k/-/g/ boundary
SPINEXPinSP	-81.9 (26.4)	13.2 (2.2)	5.4 (1.9)	22.7% (9.7)	-	8.4 (3.2)	20.6 (4.2)
SPEXPinSP	-75.2 (25.5)	13.9 (1.9)	6.1 (1.6)	26.3% (10)	7.4 (6)	7.8 (3.2)	20.8 (2.2)
SPEXPinUK	-41.5 (29.8)	13.4 (2.3)	5.8 (1.7)	54.6% (24.6)	47.9 (23.3)	11.2 (3.5)	21.3 (3.6)

Table 8.2. L1-Spanish groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and Spanish category boundaries. Standard deviations and provided in parentheses.

8.2.2. The effect of individual differences on L2 perception

Next, the effect of the individual factors on L2 perception was investigated. Regarding the English groups (see Table 8.3 for group means and Table AF.1 in Appendix F for individual results), no significant effect of the independent variables on /p/-/b/ category boundary was found [*F* (5, 18) = .774; *p* = .581; R^2 = .177]. As for /k/-/g/, a significant effect was revealed [*F* (5, 18) = 6.373; *p* = .001; R^2 = .639]. Dominance [*t* (18) = -3.302; *p* = .004], years of L2 formal instruction [*t* (18) = -4.155; *p* = .001] and amount of weekly L2 use [*t* (18) = -4.051; *p* = .001] were found to have a significant impact, as the individual coefficient conducted on each variable indicated, whereas L2 AOL and months in the L2 setting did not (*p* > .05). The expected relationship was found between L2 years of instruction and L2 use and category boundary, given that, as the former increased, the value of the latter decreased – i.e., category boundaries were earlier and, thus, more native-like. The opposite was found in the case of dominance, since a greater dominance score – i.e., greater English dominance – appeared to result in an earlier category – i.e., more target-like – boundary.

Group	Dom.	L2 inst.	L2 AOL	% L2 use	Months in L2 setting	L2 /p/-/b/ boundary	L2 /k/-/g/ boundary
ENINEXPinUK	108.7 (27.3)	3.8 (3.3)	13.4 (3.7)	12.3% (8)	-	15.7 (4.1)	25.2 (8.3)
ENEXPinUK	106 (21.7)	7.5 (5.3)	15.5 (3.7)	15% (6.5)	9.4 (4.2)	16.5 (4.7)	25.6 (5.2)
ENEXPinSP	116.5 (26)	2.4 (2.9)	18.9 (2.1)	16.3% (10.5)	50.7 (27.3)	14.9 (3.1)	28 (2.8)

Table 8.3. L1-English groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and Spanish category boundaries. Standard deviations and provided in parentheses.

The same analysis was carried out on the L2 category boundaries of the Spanish groups (see Table 8.4 for group means). In the case of /p/-/b/, no significant effects were revealed [F (5, 21) = .368; p = .864; R^2 = .081]. As for /k/-/g/ boundary, the general analysis of the multiple regression test did not find a significant effect of individual differences as a whole [F (5, 21) = 1.530; p = .223; R^2 = .267], but the individual analysis showed that years of L2 instruction was found to have a significant effect on category boundary [t(21) = -2.412; p = .025]. Nevertheless, the relationship found between L2 instruction and /k/-/g/ category boundary was not in the expected direction. Whereas greater amount of L2 instruction was predicted to result in a more native-like perception – i.e., in a later category boundary –, the opposite was found.

Group	Dom.	L2 instr.	L2 AOL	% L2 use	Months in L2 setting	L2 /p/-/b/ boundary	L2 /k/-/g/ boundary
SPINEXPinSP	-81.9 (26.4)	13.2 (2.2)	5.4 (1.9)	22.7% (9.7)	-	9.3 (3.6)	22.8 (4.4)
SPEXPinSP	-75.2 (25.5)	13.9 (1.9)	6.1 (1.6)	26.3% (10)	7.4 (6)	10.1 (4.1)	24.9 (3.9)
SPEXPinUK	-41.5 (29.8)	13.4 (2.3)	5.8 (1.7)	54.6% (24.6)	47.9 (23.3)	12.1 (2.6)	24.5 (2.4)

Table 8.4. L1-Spanish groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and English category boundaries. Standard deviations and provided in parentheses.

8.3. The effect of individual differences on stop production

This section presents the results of the analyses conducted on the effect of the individual variables considered in this study (i.e., language dominance, years of L2 instruction, L2 AOL and weekly L2 use) on L1 and L2 production.

8.3.1. The effect of individual factors on L1 stop production

Just as in the case of perception, an individual analysis was conducted on groups sharing the same L1 in order to investigate the effect of the individual factors on L1 VOT stop production. The measure used was mean VOT production per participant (see Table 8.5 for group means and Table AG.1 in Appendix G for individual results). A multiple regression analysis was carried out for each stop category. In the case of the L1-English speakers, a significant effect of individual factors was found regarding voiceless stop production [/p/: F (5, 18) = 10.085; p = .000; R^2 = .737; /k/: F (5, 18) = 3.252; p = .029; R^2 = .475], whereas the tests conducted on voiced stops did not yield a significant result [/b/: F (5, 18) = 305; p = .903; R^2 = .078 ; /g/: F (5, 18) = .979; p = .457; R^2 = .214]. In the case of /p/, the individual coefficients showed a significant effect of language dominance [*t* (18) = -5.741; *p* = .000], L2 AOL [*t* (18) = 3.048; *p* = .007], L2 use [*t* (18) = -3.716; *p* = .002] and number of months spent in L2 country [*t* (18) = -2.609; *p* = .018], but not of L2 instruction and L1 VOT production ($p > .05$). The impact of AOL, L2 use and months spent in an L2 country on L1 production followed the expected pattern. In the case of AOL, a later age was related to a greater VOT production. A greater percentage of L2 use and a greater number of months spent in an L2 setting appeared to result in a smaller – more Spanish-like – /p/ VOT value. Conversely, contrary to predictions, the opposite was found regarding dominance, as a smaller score – i.e., less English-dominant – was related to a greater /p/ VOT mean. As for /k/, only dominance [*t* (18) = -2.561; *p* = .020] and number of months spent in an L2 country [*t* (18) = -2.178; *p* = .043] were found to have an effect on VOT production, again, contrary to predictions: less English-dominance resulted in greater VOT values. A shorter period of time spent in an L2 country resulted, as expected, in greater VOT values.

Group	Domi.	L2 instr.	L2 AOL	% L2 use	Months in L2 setting	L1 /p/	L1 /b/	L1 /k/	L1 /g/
ENINEXPinUK	108.7 (27.3)	3.8 (3.3)	13.4 (3.7)	12.3% (8)	-	52 (20)	8 (6)	79 (18)	20 (23)
ENEXPinUK	106 (21.7)	7.5 (5.3)	15.5 (3.7)	15% (6.5)	9.4 (4.2)	50 (18)	-9 (40)	76 (19)	1 (48)
ENEXPinSP	116.5 (26)	2.4 (2.9)	18.9 (2.1)	16.3% (10.5)	50.7 (27.3)	46 (18)	0 (35)	72 (21)	6 (41)

Table 8.5. L1-English groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL); % (weekly) L2 use, months spent in an L2 setting and mean English VOT productions. Standard deviations and provided in parentheses.

The same analysis was conducted on the Spanish groups (see Table 8.6 for group means and Appendix G for individual results). None of the tests considering all variables revealed a significant result, although marginal results were found in the case of /k/ [/p/: F (5, 21) = 2.101; p = .105; R^2 = .333; /b/: F (5, 21) = .294; p = .911; R^2 = .065; /k/:

F (5, 21) = 2.574; *p* = .057; R^2 = .380; /g/: *F* (5, 21) = .124; *p* = .985; R^2 = .029]. The individual analysis showed that number of years of L2 instruction had a significant effect on /p/, production [*t* (21) = 2.728; *p* = .013]. The relationship between the two variables follows the prediction that a longer period of L2 learning would result in a greater – less Spanish-like and more English-like – VOT production. Similarly, years of L2 instruction seemed to influence /k/ production in a similar manner that it was found to affect /p/, but the result was not statistically significant [*t* (21) = 1.864; *p* = .076]. All other tests failed to find a significant result (*p* > .05).

Group	Dom.	L2 instr.	L2 AOL	% L2 use	Months in L2 setting	L1 /p/	L1 /b/	L1 /k/	L1 /g/
SPINEXPinSP	-81.9 (26.4)	13.2 (2.2)	5.4 (1.9)	22.7% (9.7)	-	8 (5)	-64 (36)	32 (11)	-71 (23)
SPEXPinSP	-75.2 (25.5)	13.9 (1.9)	6.1 (1.6)	26.3% (10)	7.4 (6)	10 (9)	-80 (26)	37 (12)	-80 (51)
SPEXPinUK	-41.5 (29.8)	13.4 (2.3)	5.8 (1.7)	54.6% (24.6)	47.9 (23.3)	10 (5)	-68 (37)	44 (15)	-77 (40)

Table 8.6. L1-Spanish groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and mean Spanish VOT productions. Standard deviations and provided in parentheses.

8.3.2. The effect of individual differences on L2 stop production

The effect of the individual factors considered in this thesis on L2 production was investigated following the same procedure as in the case of L1 production (see Table 8.7 for group means and Table AG.1 in Appendix G for individual results). As for the English speakers, the multiple regression analyses failed to reveal any significant results [/p/: *F* (5, 18) = 1.652; *p* = .197; R^2 = .315; /b/: *F* (5, 18) = 1.827; *p* = .158; R^2 = .337; /k/: *F* (5, 18) = 1.053; *p* = .418; R^2 = .226; /g/: *F* (5, 18) = .485; *p* = .783; R^2 = .119].

Furthermore, none of the individual coefficients revealed a significant effect of any of the individual factors on L2 VOT production (*p* > .05 in all cases).

Group	Dom.	L2 instr.	L2 AOL	% L2 use	Months in L2 setting	L2 /p/	L2 /b/	L2 /k/	L2 /g/
ENINEXPinUK	108.7 (27.3)	3.8 (3.3)	13.4 (3.7)	12.3% (8)	-	28 (23)	-13 (45)	55 (21)	-6 (52)
ENEXPinUK	106 (21.7)	7.5 (5.3)	15.5 (3.7)	15% (6.5)	9.4 (4.2)	17 (11)	-48 (52)	45 (14)	-55 (51)
ENEXPinSP	116.5 (26)	2.4 (2.9)	18.9 (2.1)	16.3% (10.5)	50.7 (27.3)	14 (11)	-33 (47)	46 (12)	-29 (48)

Table 8.7. L1-English groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and mean Spanish VOT productions. Standard deviations and provided in parentheses.

Regarding the Spanish speakers (see Table 8.8 for group means and Table AG.1 in Appendix G for individual results), none of the multiple regression tests revealed a significant effect of the individual factors on VOT production in the L2 [/p/: F (5, 21) = .1.045; p = .418; R^2 = .199; /b/: F (5, 21) = .800; p = .562; R^2 = .160; /k/: F (5, 21) = .739; p = .603; R^2 = .150; /g/: F (5, 21) = 1.570; p = .212; R^2 = .272]. Still, in the case of /g/, the individual tests conducted on each variable showed that the number of months spent in an L2 setting had a significant effect on /g/ production [t (21) = 2.418; p = .025]. The relationship between this factor and /g/ VOT production was in the expected direction, since a greater number of months spent in an immersion setting resulted in greater VOT values in English production.

Group	Dom.	L2 instr.	L2 AOL	% L2 use	Months in L2 country	L2 /p/	L2 /b/	L2 /k/	L2 /g/
SPINEXPinSP	-81.9 (26.4)	13.2 (2.2)	5.4 (1.9)	22.7% (9.7)	-	23 (22)	-66 (44)	53 (30)	-50 (46)
SPEXPinSP	-75.2 (25.5)	13.9 (1.9)	6.1 (1.6)	26.3% (10)	7.4 (6)	32 (25)	-71 (45)	70 (22)	-49 (41)
SPEXPinUK	-41.5 (29.8)	13.4 (2.3)	5.8 (1.7)	54.6% (24.6)	47.9 (23.3)	35 (21)	-43 (49)	73 (22)	-24 (53)

Table 8.8. L1-Spanish groups mean dominance score (dom.), years of L2 instruction (L2 instr.), L2 age of learning (AOL), % (weekly) L2 use, months spent in an L2 setting and mean English VOT productions. Standard deviations and provided in parentheses.

8.4. Summary of results

The effect of individual differences concerning the factors examined – i.e., language dominance, years of L2 instruction, L2 age of learning, percent of weekly L2 use and number of months spent in an L2 setting – on L1 and L2 stop perception and production was analyzed. A regression analysis was carried out for each set of groups sharing the same L1 – i.e., the English-L1 speakers and the Spanish-L1 speakers. Only a few significant results were observed and some of them were found to be in the opposite direction to what was expected, particularly concerning language dominance.

In the case of perception, the individual differences were not found to influence the L1 perception of either the English or the Spanish groups (See Tables 8.9 and 8.10 for a summary of results). As for the L2, the L1-English groups presented some effect of the individual factors on their /k/-/g/ category boundary, but not on /p/-/b/. The amount of years of L2 instruction and weekly L2 use was found to affect the boundary of the velar contrast in the expected direction: the greater the number of years having learnt Spanish and the greater the L2 use, the more target like their perception of Spanish /k/-/g/ was. Nonetheless, an effect of dominance that contradicted predictions was also found on /k/-/g/ boundary. Regarding the Spanish learners, only years of L2 instruction seemed to influence /k/-/g/ perception, although it was in the opposite direction to what was expected – i.e., the greater the amount of L2 instruction, the earlier the category boundary.

Contrary to what was found in perception, some instances of a significant effect of these factors was observed in the L1 production of the English speakers, although only for the voiceless stops (see Table 8.9). Age of learning, amount of L2 use and months spent in an L2 country were found to influence the VOT production of /p/ following expectations. Still, language dominance seemed to predict the VOT used in /p/ production in the opposite direction to what was expected – i.e., the more English dominant, the shorter the VOT. In the case of /k/, a significant effect of number of months spent in an L2 setting was found following expectations. Nevertheless, linguistic dominance was revealed, one more time, to have the opposite effect to what was predicted on English /k/ production. As for the L2, no significant relationships were found. Regarding the Spanish participants, only years of L2 instruction had a significant effect on /k/-/g/ perception, although it was in the opposite direction to what was expected – i.e., the greater the amount of L2 instruction, the earlier the category boundary (see Table 8.10).

The individual factors assessed were not found to influence the L2 production of the English learners for any stop category. Similarly, none of the regression analysis conducted on the Spanish learners revealed a significant effect of individual factors on L2 VOT production. However, the individual analysis conducted on each variable revealed some significant effects. Years of L2 instruction appeared to influence the production of /p/ in the L1 – i.e., the greater the amount of L2 instruction, the less native-like their VOT was. Moreover, the longer participants had lived in an L2 setting, the more target-like their production of English /g/ – in terms of VOT – appeared to be.

All in all, it seems that the factors analyzed are not good predictors of L1 and L2 stop perception and production in the current study, as no clear patterns were observed.

Although in a few occasions some of the factors were found to have a significant effect, none of them seemed to have a consistent effect across languages – L1 and L2 – categories – /p/, /b/, /k/ and /g/ – or modalities – perception and production. Moreover, some of the relationships were found to be in the opposite direction to predictions.

	Perception				**Production**							
Independent variable	**L1**		**L2**		**L1**				**L2**			
	/p/-/b/	**/k/-/g/**	**/p/-/b/**	**/k/-/g/**	**/p/**	**/b/**	**/k/**	**/g/**	**/p/**	**/b/**	**/k/**	**/g/**
General result (all coefficients)	-	-	-	**	**	-	*	-	-	-	-	-
Dominance score	-	-	-	** !	**!	-	*!	-	-	-	-	-
Years of L2 instruction	-	-	-	**	-	-	-	-	-	-	-	-
L2 AOL	-	-	-	-	**	-	-	-	-	-	-	-
Months in L2 country	-	-	-	-	*	-	*	-	-	-	-	-
% L2 use	-	-	-	**	**	-	-	-	-	-	-	-

Table 8.9. Summary of the results obtained by the L1-English groups in the multiple regression analysis. '-' indicates no significant result, '*'/'' significant in the expected direction,'*!'/'**!' significant result in the opposite direction to expectations and '(*)' marginal result.**

	Perception				**Production**							
Independent variable	**L1**		**L2**		**L1**				**L2**			
	/p/-/b/	**/k/-/g/**	**/p/-/b/**	**/k/-/g/**	**/p/**	**/b/**	**/k/**	**/g/**	**/p/**	**/b/**	**/k/**	**/g/**
General result (all coefficients)	-	-	-	-	-	-	(*)	-	-	-	-	-
Dominance score	-	-	-	-	-	-	-	-	-	-	-	-
Years of L2 instruction	-	-	-	*!	**	-	(*)	-	-	-	-	-
L2 AOL	-	-	-	-	-	-	-	-	-	-	-	-
Months in L2 country	-	-	-	-	-	-	-	-	-	-	-	*
% L2 use	-	-	-	-	-	-	-	-	-	-	-	-

Table 8.10. Summary of the results obtained by the L1-Spanish groups in the multiple regression analysis. '-' indicates no significant result, '*'/'' significant in the expected direction,'*!'/'**!' significant result in the opposite direction to expectations and '(*)' marginal result**.

8.5. Discussion

The effect of individual factors – i.e., number of months spent in al L2 setting, language dominance, years of formal L2 instruction, L2 AOL and percent weekly L2 use – on L1 and L2 stop perception and production was investigated. Overall, no clear effect of the factors analyzed was found. Only a few correlations were revealed and half of them were in the opposite direction to predictions. Moreover, no consistency within language – L1 or L2 – or within groups was observed. The lack of an apparent effect of individual factors on L2 perception and production in the present thesis opposes the outcome of previous studies (e.g., Flege, 1991a; Flege & Fletcher, 1992; Hazan & Boulakia, 1993; Piske et al., 2001). This disagreement with the literature will be discussed addressing each factor separately.

It was expected that the number of months spent in an L2 setting would be found to have a positive effect on the L1 and L2 stop perception and production of L2 learners, especially considering the fact that an effect of group – which differed in amount of L2 experience – was revealed in both modalities, particularly in the L2. However, the number of months spent in an immersion setting was only found to have a significant impact on a few occasions. More specifically, this variable had an effect on the L1 production of voiceless stops by the English speakers, as well as on the production of L2 /g/ by the Spanish learners of English. In the case of the English learners, the expected outcome was found in the case of the production of English /p/ and /k/, since a greater amount of time spent in an L2 setting resulted in shorter VOT values, possibly due to an influence of Spanish on their L1 production. As for the Spanish groups, the effect of number of months spent in an L2 setting on English /g/ production was also the expected one: a greater number of months spent in an L2 setting resulted in more target-like – greater – VOT values. No other effects of number of months spent in al L2 setting were found on the L2, whereas the group analysis revealed an effect of L2 experience on the perception of the Spanish learners and on the production of all stops for both populations (see sections 5.3 and 6.2).

The outcome of the analysis conducted on the number of months in L2 setting may seem to oppose the results obtained in the analysis conducted on groups, as no effect of L2 experience was found on L1 production (See section 6.2). Nevertheless, recall that groups differed not only in amount of L2 experience but also in the setting where they were tested – L1 vs. L2 setting. Therefore, a potential explanation to this apparent mismatch in the results of the two analyses could be the difference in language setting, which has been found to have an effect on L2 and L1 performance in previous studies (e.g., Sancier & Fowler, 1997). More specifically, the most experienced groups were living in an immersion setting, whereas the moderately experienced groups and the inexperienced groups were tested in their home country. In short, not only the number of months in an immersion setting distinguishes groups and, thus, the results obtained in the group analysis may not be completely comparable to the results obtained for the variable months in L2 setting.

Another factor that was investigated was amount of L2 instruction. Just as in the case of months in an L2 setting, L2 instruction did not have a strong and consistent effect on L2 learners, as it was only found to significantly influence L1 and L2 performance on a few occasions and, on one of them, the effect was contrary to the predicted one – i.e., a greater amount of L2 instruction resulted in a less accurate categorization of the velar contrast in Spanish learners. This finding is not in accordance with some previous studies (e.g., Flege & Fletcher, 1992), but is in agreement with others (e.g. Elliott, 1995; Flege et al., 1999; Thomson, 1991). This mismatch in the literature is attributed by Elliot (1995) and Piske et al. (2001) to the fact that L2 instruction on L2 pronunciation has traditionally received scant attention in the L2 classrooms. In the case of the Spanish learners, it should also be considered that most participants had received formal instruction in English since primary school, as it is a mandatory subject in the Spanish curriculum, and most participants continued to study English outside school. Still, whereas some studied instrumental English in language schools, others – particularly the Spanish participants tested in the Spain – were studying a university degree in English studies. Therefore, it is possible that it was not strictly the number of years that participants had received L2 instruction but the type of instruction that had an impact on L2 performance. More specifically, the Spanish learners in Spain, who were majoring in English, had received a more thorough instruction of English, even if the total number of years studying the L2 was very similar between groups (SPINEXPinSP: 13.2; SPEXPinSP: 13.9; SPEXPinUK: 13.4). Thus, the small variability in the actual number of years having studied English along with the difference in the type of training received between the groups tested in Spain and the one tested in the UK would explain the small impact that this factor had on the Spanish participants. Regarding the English groups, as explained in section 4.1, the most experienced group living in an immersion setting was the one that had received the least formal instruction in the L2. Therefore, it is possible that amount of L2 experience and setting were confounding variables in the analysis of the effect of L2 instruction. That is, a clear effect of L2 instruction on the performance of the English groups may not have been found because L2 experience was greater in the experienced group in Spain than in the moderately experienced group in the UK – who, nevertheless, had received a significantly greater amount of L2 instruction (ENEXPinUK: 7.5 vs. ENEXPinSP: 2.4).

Age of learning was not found to have an effect on L1 and L2 accuracy either (c.f., Flege et al., 1995; Flege et al., 1999), as only one significant result was revealed. As has already been mentioned, there is little variability regarding L2 AOL in the Spanish groups, since English is a mandatory course in the Spanish primary and secondary education curriculum. This little variability would explain the lack of an influence of AOL on the Spanish learners. As for the English learners, the group that started learning Spanish the latest was the highly L2 experienced group tested in Spain. Just as in the case of amount of L2 learning, L2 experience – as analyzed in the between-groups analysis – may have confounded the effect of L2 AOL.

L2 use was found to influence the perception and production of L2 learners in two occasions and only regarding the English learners. Previous studies have found that, although L2 use had an effect on L2 performance, it only accounted for a small percentage – i.e., 9% – of the variability in the degree of foreign accentedness of L2 learners (Tahta et al., 1981). Similarly, Purcell and Suter (1976, 1980) found that a combination of L2 use with length of residence in an L2 setting was a good predictor of L2 accuracy in production. In short, even though L2 use has been found to have some effect on bilinguals' performance, it does not seem to be a crucial factor by itself. Purcell and Suter's (1976, 1980) finding that a combination of both L2 use and L2 experience is a determining factor on L2 production can be applied to the present study. Whereas in the case of the Spanish learners, amount of L2 use and L2 experience seemed to go hand in hand – i.e., the most experienced group also used the L2 the most –, only small differences were observed between the English groups. The English group living in Spain presented a very similar score for percent weekly L2 use to those of the groups living in the UK (ENINEXPinUK: 12.3%; ENEXPinUK: 15%; ENEXPinSP: 16.3%). Therefore, it is possible that the small difference in L2 use and a greater amount of L2 instruction had resulted in the moderately experienced group outperforming the highly experienced group in L2 voiced stop production.

Finally, dominance was not found to have a clear effect on L1 and L2 perception and production. In fact, it only appeared to influence the English learners' performance on a few occasions and showed the opposite results to what was predicted – i.e., a more balanced score resulted in less native-like productions in the L1, as well as in a less target-like perception in the L2. Recall that the most English-dominant group was the English group living in Spain (ENEXPinSP), although the dominance scores between groups were similar (ENINEXPinUK: 108.7; ENEXPinUK 106; ENEXPinSP: 116.5). Therefore, it is possible that the small variability between groups along with the fact that groups did differ more clearly regarding other factors – such as L2 experience and L2 instruction, which do not go hand in hand – have influenced this result. Methodological limitations should also be considered. In the case of L2 perception, as has been argued, an English mode might have been activated in the Spanish identification task, resulting in later – more English-like – category boundaries than it would be expected for Spanish. As for production, an alternative explanation can be provided in line with Flege's (1995) category dissimilation. The mechanism of category dissimilation takes place, according to Flege, in order to make L1 and L2 stops more distinct. As a result, participants with lower English-dominant scores may have used greater – i.e. more extreme – VOT values in English /p/ and /k/ production to make them more acoustically different from the Spanish categories. Still, no effect of dominance was revealed for any other comparison, suggesting that dominance had, in any case, a small influence on L1 and L2 perception and production. Nevertheless, some previous studies (e.g., Hazan & Boulakia, 1993; Flege et al., 2002) found that language dominance was a good predictor of L2 performance in production – i.e., L2 speakers were less accented in their dominant languages. It should be noted that all the groups tested in the present thesis – both L1-English and L1-Spanish – were found to be L1-dominant to a greater or lesser extent.

Thus, even though the degree of language dominance varied between groups – especially in the case of the Spanish speakers –, all groups would be expected to be more accurate in L1 production than in L2 production, as was the case in this study. Another issue that should be considered is the measure used to quantify dominance in the current study as opposed to previous ones. Hazan and Boulakia (1993) determined linguistic dominance according to the accentedness ratings – provided by monolinguals speakers – that the bilinguals obtained in each language, whereas Flege et al. (2002) quantified dominance on the basis of self-reported proficiency. The present study used the BLP questionnaire to assign a dominance score to each participant. The score was calculated by considering a series of variables concerning language history, use, proficiency and attitudes. Therefore, the conflicting findings between this study and previous ones may also have to do with the measure used to determine dominance.

The fact that the combination of individual differences as a whole – i.e., the general result of the regression analysis considering all variables – also failed to reveal straightforward results is also worth discussing. No influence was found in the Spanish groups, whereas some significant results were revealed in the case of the English groups, especially regarding the production of voiceless stops. Still, as a whole, individual differences did not prove to have a strong effect on the L1 and L2 perception and production of L2 learners. In the case of the Spanish learners, it is possible that the small variability between participants regarding some factors – such as amount of L2 instruction or AOL – resulted in a lack of significant results. As for the English groups, there was also little variation regarding some factors like L2 use and language dominance. Moreover, the participants that had spent the greatest amount of time in an L2 setting – mostly in the ENEXPinSP group – also tended to have received the smallest amount of

L2 instruction and had started learning Spanish the latest. This opposition may have neutralized the effect of individual differences.

Summary

This chapter has investigated the effect of individual factors – including language dominance, L2 instruction, L2 AOL, weekly L2 use and number of months spent in an L2 setting – on the L1 and L2 perception of stops by the two populations under study. Even though the goal of this thesis is not to investigate the effect of these factors on the L1 and on the L2, their potential influence was explored given the findings of previous research and the fact that individual information had been collected. All in all, individual factors were not found to have a determining effect on L2 and L1 perception and production, in opposition with previous studies. Methodological differences may account for this disagreement with the literature. Furthermore, the little variation between participants concerning some factors, along with the fact that some participants presented high values in factors that were expected to positively influence L2 performance but small values in others may have resulted in a neutralization of their effect. For example, regarding the English learners of Spanish, there was a greater difference between participants in the number of months spent in an L2 setting than in amount of L2 use and index of language dominance. Moreover, the participants in the group that had spent a longer period of time in an L2 setting had also received less L2 instruction and had started learning Spanish later in life than the other two groups. As for the Spanish groups, even though some variation was found in the amount of L2 instruction, participants living in Spain had received a more specialized instruction in English, as they were studying a university degree in English Studies. In short, individual factors were not found to have a determining effect on L1 and L2 stop perception and production for neither of the populations tested. Further research specially designed to assess the effect of individual factors on the L1 and L2 may help shed some light on the inconclusive results reported in this chapter.

9. General discussion and conclusions

The current chapter discusses the findings reported in Chapters 5, 6, 7 and 8 by providing a comprehensive global perspective to the study. The discussion will be organized around the research questions presented in Chapter 3, which will be confirmed or falsified based on the outcome of the experiments and their relation to the literature. Next, the limitations of this study will be acknowledged and the directions for possible further research will be pointed out. The chapter ends with the final conclusions and implications of this study.

9.1. General discussion

The goal of the present thesis is threefold. First, it investigates the effect of L2 experience – i.e., amount of time resided in an L2 setting – on the L1 and L2 stop perception and production of initial stops. Second, it aims to determine whether L2 learners present a single category or a shared category for similar phones, particularly stops, and whether L2 experience has an effect on category formation. Finally, the relationship between perception and production in the L1 and in the L2 is explored. Two populations were examined, namely L1-English learners of Spanish and L1-Spanish learners of English differing in amount and type of L2 experience. The target phones analyzed are /p/, /b/, /k/ and /g/. Coronal stops were not investigated in this study as they are produced in a different place of articulation in the two languages under study.

In order to investigate these issues, two experiments were designed, namely a perception and a production experiment. The data of a total of 80 participants, which were divided into eight different groups according to their L1, amount of L2 experience and setting, were considered for the experiment. The groups tested included two control groups – one group of English functional monolinguals and one group of Spanish functional monolinguals – as well as six groups of mirror-image L1-English L2-Spanish/ L1-Spanish L2-English speakers varying in amount of L2 experience. The mirror-image testing groups were made up of two groups of inexperienced learners living in their homeland who had received some formal education in the L2 – one group for each L1 –, two groups of moderately experienced L2 learners with formal instruction in the L2 living in an L1 setting and two groups of highly experienced L2 learners living in an immersion setting. The control groups were only tested in their language, whereas the testing groups had to complete the experiment in both their L1 and L2. The perception experiment consisted of two separate identification tests for /p/-/b/ and /k/-/g/ that were presented under two conditions – i.e., in a task presented in English and in a task presented in Spanish. Production was examined using a carrier sentence reading task in each language that included the target phones in absolute initial position.

The perception experiment (see Chapter 5) revealed that L2 experience had an effect on both L1 and L2 perception by the Spanish learners of English – i.e., a greater amount of L2 experience resulted in a more target-like L2 perception but in a less native-like perception in the L1. All groups obtained intermediate categories – with values that were numerically between the L1 and the L2 – that were increasingly more L2-like as participants gained experience and that were not different from one another – i.e. there was no significant difference between the L1 and the L2 perceptual categories – except in the case of the velar stops on the part of the moderately experienced group. However, no effect of experience was found regarding the English groups, who perceived both L1 and L2 stops according to English VOT values. Regarding production (see Chapter 6), both the L1-English and the L1-Spanish participants had more target-like voiceless stop categories as they gained experience in the L2. A difference between voiced and voiceless stop production was observed, as voiced stops were found to be generally more challenging to produce authentically than voiceless stops. A clear effect of L2 experience on L2 voiced stop production was not observed in the English groups, as the moderately experienced group outperformed both the inexperienced and the highly experienced learners of Spanish. Moreover, L2 experience was found to have a very small impact on L1 production. Only one difference between groups sharing the same L1 was observed. More specifically, the Spanish experienced group in the UK produced significantly greater VOT values than the Spanish inexperienced group – but they did not differ from Spanish controls. Contrary to perception, L2 learners were able to make a difference between L1 and L2 stop production, as they used significantly greater VOT values in English than in Spanish – with the exception of /b/ regarding the Spanish groups living in Spain. Having analyzed the perception and production results, it was investigated whether there was a relationship between the two dimensions at an individual level (Chapter 7). A straightforward relationship between the two measures was not observed. In the case of the English groups, the general results of the two experiments suggested that the English learners of Spanish were more successful in the production of L2 stops than in their perception, but a similar trend was not observed with the Spanish learners. Considering the variability among participants and the findings of previous studies regarding the influence of other individual factors, Chapter 8 explored the effect of a number of such factors following information elicited in the linguistic background questionnaire. The factors examined were amount of time spent in an L2 setting – as a continuous variable –, language dominance, L2 AOL, L2 instruction and weekly L2 use. No clear effect of these factors on accuracy of L1 and L2 perception and production was revealed. The discussion will address the results that have just been summarized, which will be related to the main literature in the field and will be organized around the research questions presented in section 3.2.

9.1.1. RQ.1: The effect of L2 experience on L2 perception and production

The first research question addressed the effect of L2 experience on L2 perception (RQ.1.1) and production (RQ.1.2). More specifically, the perception and production of the /p/-/b/ and the /k/-/g/ contrast by Spanish learners of English and English learners of Spanish in their corresponding L2 were analyzed. It was hypothesized that L2 experience would have a positive effect on both L2 perception and production – i.e., the greater the amount of L2 experience, the more-target-like their L2 perception and production would be. Overall, a clearer effect of L2 experience was found in production than in perception, particularly when it comes to voiced stops.

Regarding perception, differences between the L1-English and the L1-Spanish groups were found. On the one hand, all English groups, regardless of amount of L2 experience, had perceptual category boundaries with English-like VOT values that differed significantly from those of the Spanish control group. That is, no effect of L2 experience was found with the English groups. On the other hand, a positive effect of L2 experience was observed in the Spanish groups: the greater their amount of L2 experience, the more target-like their English category boundaries were. As a matter of fact, only the Spanish inexperienced learners of English differed significantly from the English controls in both /p/-/b/ and /k/-/g/ categorization, whereas the moderately experienced group and the highly experienced group did not differ from the English controls in any case. It should be noted, however, that there was a numerical difference between the two experienced groups and the English controls, as the former presented somewhat earlier category values than the latter.

The lack of an effect of experience on the English groups does not only oppose the results obtained for the Spanish groups but also results reported in previous studies (e.g., Flege et al., 1997; Levy, 2009). For instance, Flege et al. (1997) reported that highly experienced learners of English – with heterogeneous linguistic background – were more accurate at discriminating /ɛ/-/æ/ and /i/-/ɪ/ than inexperienced learners of English with the same L1. This disagreement with the literature and with the findings for the Spanish groups might be due to two different reasons. First, it is possible that no apparent effect of L2 experience was found in the L2 perception of the English groups due to their relatively limited L2 use and amount of L2 instruction. Although no clear effect of individual differences was revealed for the English groups (as shown in Chapter 8) – probably due to the small variation in terms of L2 use and the possible confound of L2 instruction with L2 experience –, it became evident that the Spanish groups had received considerably more L2 instruction and used their L2 to a greater extent than the English learners. Another potential explanation is related to methodological issues, particularly regarding the stimuli used. Bohn and Flege (1993) reported that stop voicing identification by Spanish-English speakers depended on the set of stimuli presented – i.e., on the range of VOT tested. Thus, they found that short-lag /t/ was more frequently identified as voiced than as voiceless when presented with long-lag tokens in the same task than when presented alongside short-lag /d/. Recall that for the purposes of consistency and balance, the same stimuli were used in the two languages and with the two populations – L1-English and L1-Spanish speakers. Therefore, both the Spanish and the English identification tests included tokens with aspirated VOT, a feature that is used in English but not in Spanish. An attempt was made to control language mode in perception by providing the instructions in the language of the test and by completing the production task in the same language – a reading task that may as well contribute to language mode activation – prior to each perception task. Nevertheless, in the case of the Spanish perception task, the input received by the participants – which included English-like stimuli – might have triggered the activation of an English language mode and resulted in English-like categorizations (Escudero, 2005, Grosjean, 2001). This potential bias towards an English perception could have been controlled by embedding the stimuli in a carrier sentence in the target language or in a word that presented phonetically distinct features that were specific to each language (Gonzales & Lotto, 2013; Casillas & Simonet, 2018). This issue will be discussed below.

The production experiment revealed comparable results between the two populations. In both cases, amount of L2 experience was found to modulate L2 production, given that the groups with a moderate or an extensive L2 experience were found to outperform those with no L2 experience. Nevertheless, L2 experience was found to have a clearer effect on voiceless stop production than on voiced stops. Regarding voiceless stops, in the case of the English learners, only the inexperienced group produced significantly longer – more English-like – VOT values than the Spanish controls, whereas the moderately experienced group and the highly experienced group were found to produce L2 voiceless stops accurately. Moreover, the inexperienced learners were also found to produce /p/ with significantly longer VOT values than the most experienced group. As for the Spanish speakers, only the highly experienced group was found to present VOT values for /p/ and /k/ that were comparable to those of the English controls. The inexperienced group and the moderately experienced group presented shorter – more Spanish-like – VOT values than the English controls regarding the production of /p/, whereas all groups appeared to produce /k/ accurately. In short, it appears that L2 experience had a positive effect on the production of voiceless stops for both the English learners of Spanish and the Spanish learners of English. This finding is in agreement with previous studies (e.g., Flege, 1987; Flege et al., 1997; Lev-Ari & Peperkamp, 2013; Levy & Law, 2010). For example, Flege (1987) reported that highly experienced English learners of French and highly experienced French learners of English produced L2 /t/ with VOT values that resembled those of native speakers to a greater extent than those of inexperienced learners. Another factor that may have played an important role on L2 voiceless stop production is setting. In the case of the English learners, the experienced group living in Spain produced Spanish /p/ with significantly more target-like VOT values than the inexperienced learners, whereas the moderately experienced group (living in the UK) did not differ from the inexperienced learners, but also presented comparable VOT values to those of Spanish controls. As for the Spanish groups, only the highly experienced learners were able to produce both voiceless stops accurately, whereas the groups tested in Spain only used target-like VOT values in the case of /k/. Previous research has pointed out that the likelihood that L2 speakers may experience changes in their L1 and L2 perception is related to the setting where they have been staying. Sancier and Fowler (1997) found that a stay in the L1 or the L2 setting had an effect of the perception of L2 accentedness, as the L1 was found to affect the L2 to a greater extent when production was evaluated right after a stay in an L1 setting, whereas less L1 influence was observed right after a stay in an L2. Even though language setting and amount of L2 experience are confounded in this study – participants in the L2 setting had considerably more L2 experience than those tested in the L1 setting –, individual differences in the number of months spent in an L2 setting did not prove to have a determining effect on L2 production (see Chapter 8 on individual differences). Therefore, it is plausible to claim that language setting combined with L2 experience had a positive effect on the production of voiceless stops, as, overall, the group in the L2 setting – the highly experienced group – outperformed the moderately experienced group.

As for voiced stops, L2 experience was found to exert a smaller effect on /b/ and /g/ production than on /p/ and /k/. In the case of the Spanish learners, even though the most experienced group yielded numerically more similar values to those of the English controls, all groups produced voiced stops with significantly more prevoicing – i.e., more Spanish-like values – than the English controls. Previous studies have also found that L1 speakers of languages that use voice-lead VOT distinctively in voiced stop production – e.g., Spanish, French and Italian – tend to use prevoicing in English (Caramazza et al., 1973; Hazan & Boulakia, 1993; Mackay et al., 2001). The fact that L2 voiced stops seem to be more difficult to acquire than voiceless stops can be explained by the phonetic and phonological status of prevoicing as opposed to aspiration in the two languages under study. Whereas aspiration is used distinctively in English and it is not a feature that is present in Spanish, prevoicing is used distinctively in Spanish and may also occur in English subject to free variation. As a matter of fact, the present study, in agreement with previous research (e.g., Lisker & Abramson, 1964), showed that a number of speakers in every L1-English group alternated the use of prevoicing and short-lag VOT in their productions of L1 voiced stops, even though short-lag productions were more common (see section 6.2.1). Therefore, in the case of the Spanish speakers, it cannot be claimed that the use of prevoicing in English is non-native. As a result, the Spanish categories for voiced-stops – i.e., with voice-lead VOT values – may have been directly transferred to the L2 and may be experiencing no changes or smaller changes than the voiceless stops.

Regarding the English learners, the group with intermediate experience outperformed both the inexperienced and the highly experienced group, as they were the only group that produced voiced stops with an amount of prevoicing comparable to that of Spanish bilinguals, even if numerically smaller. Initially, it was considered that the use of prevoicing in the L1 – which, as has already been explained, is subject to individual and free variation in English – might have been transferred from the L1 to the L2, given that the moderately experienced English learners of Spanish were the group that presented the longest prevoicing in both languages. However, the correlations conducted between the means obtained in the two languages did not support this idea, as the amount of VOT used in /b/ and /g/ production in the L1 was not found to be correlated to its use in the L2. Another interpretation for the fact that the moderately experienced learners outperformed the highly experienced English group is related to the amount of L2 use and L2 instruction received. As is has been mentioned above even though the English speakers living in Spain were the most experienced group, living in an immersion setting did not result in a greater use of the L2, as all L1-English groups presented comparable values. Moreover, the group that had received the smallest amount of L2 instruction was the most experienced group, whereas the moderately experienced learners had received the greatest amount of L2 instruction. It was considered that the relatively scarce L2 use was not sufficient to acquire L2 voiced stops and that a more explicit instruction might be needed in order to adjust the use of voice-lead VOT from a phonetic feature – in English – into a phonological feature – in Spanish. Nevertheless, L2 instruction did not prove to have a significant effect on L2 voiced stop production for the English learners (see section 8.3.2). The moderately experienced English group and the highly experienced English group also differed in their occupation. The moderately experienced group were students of Spanish at the time of testing in the same institution were the experiments took place, which may have resulted in a different disposition to do the experiment – i.e., as if it was a test –, and, thus, in more accurate productions of Spanish voiced stops. Conversely, the highly experienced speakers were teachers of English in Spain who were not studying Spanish at university at the time of testing and, thus, they may not have been as careful as the moderately experienced learners in their L2 productions. Further research on the acquisition of voice-lead VOT as a distinctive feature is required to better understand what factors contribute to change their status from phonetic to phonological.

In sum, it appears that L2 experience has a positive effect on L2 perception and production. A more straightforward relationship was found in the case of the Spanish groups, as a positive effect of experience was found in both perception and production – i.e., a greater amount of L2 experience resulted in a more target-like perception and production of L2 stops. In the case of the English learners, no effect of experience was found in perception and, regarding production, it only had a determining effect in the case of voiceless stops. A less straightforward effect of L2 experience was revealed for voiced stop production as neither the highly experienced English group nor the highly experienced Spanish group were found to produce L2 stops in a target-like manner. In fact, all groups presented VOT means that were numerically – and also significantly, except in the case of the moderately experienced English group – different from those of their corresponding L2 controls. As regards as perception, the lack of a clear effect of L2 experience with the English groups might be due to the undesired activation of an English language mode as a result of the inclusion of aspirated tokens in the Spanish task. Moreover, overall, both populations appeared to have greater difficulty in producing VOT values that resembled those of native speakers in the case of voiced stops than regarding voiceless stops and L2 experience appeared to have a more limited effect on the former than on the latter. This difference between voiced and voiceless stop production may be related to the different status that voice-lead VOT has in the two languages under study, namely, a phonological status in Spanish (voiced stops are always produced with voice-lead VOT) and a phonetic status in English (voiced stops range from voice-lead to short-lag VOT).

9.1.2. RQ.2. The effect of L2 experience on L1 perception and production

The second research question considers the effect of L2 experience on L1 perception (RQ.2.1) and production (RQ.2.2). The phones and population under study were the same as those investigated in the case of the L2. It was hypothesized that the greater the amount of L2 experience, the more L2 influence L2 learners would present on their L1. However, no clear effect of L2 experience was observed on the L1. As a matter of fact, experience was found to have a smaller effect on the L1 than on the L2, particularly when it comes to production (see previous section).

Regarding perception, no effect of L2 experience on the L1 was observed in the English learners, as all groups perceived the voicing contrast in bilabial and velar stops in their L1 in a similar way to the English control group. As for the Spanish learners, numerically, all groups presented somewhat later perceptual boundaries than the Spanish controls, but only the most experienced learners differed significantly from controls. This finding is in agreement with some previous studies that have reported that L1 perception may become less accurate as a result of increased L2 experience (e.g., Cebrian, 2006; Dmitrieva, 2019; Major, 2010). For instance, Cebrian (2006) found that L1-Catalan speakers living in Catalonia outperformed L1-Catalan speakers living in Canada in their identification of L1 front vowels. As discussed in the case of the L2, the fact that Spanish learners had moved their L1 category forward – i.e., they needed greater VOT values to perceive a voiceless stop than Spanish controls – but the English learners had not modified their L1 perception can be attributed, among other things, to the VOT range used by each language. The fact that Spanish uses a more limited VOT range than English – which, in fact, also covers the Spanish VOT range – could have resulted in the modification of the category boundary of the Spanish learners towards more English-like values. However, the L1-English speakers may not have modified their category boundary because their L1 already covered the Spanish VOT range.

In addition, the methodological limitations acknowledged in the previous section should also be considered. The fact that no L2 influence on the L1 was found for the English groups but that it was revealed for the Spanish speakers might be related to language mode activation. As has already been mentioned, the inclusion of aspirated tokens in the Spanish test might have activated English – to a greater or lesser extent –, resulting in a later – more English-like – categorization of Spanish stops by the Spanish learners. To my knowledge, no previous studies – except a study previous to this one (Gorba, 2016) – have investigated the effect of L2 experience on the perception of stops in the L1. Further research is required in order to better understand the effect of L2 experience on L1 perception. Differences in L2 use and L2 instruction between the two populations could also explain the fact that the Spanish groups – particularly regarding the Spanish participants living in the UK – presented an influence of the L2 on the L1 whereas that was not the case in the English groups. That is, a frequent use of the L2 on the part of the Spanish learners (54.6% weekly in the case of the highly experienced group) may have resulted in an influence of the L2 on L1 perception, whereas a relatively scarce use of Spanish by the English learners (even in an immersion setting, with a 16.3% of weekly use) may not have been enough for L2-on-L1 CLI to take place. Moreover, the Spanish groups had received more L2 instruction and started to learn the L2 at an earlier age than the English learners. For instance, the Spanish experienced group in the UK had learnt English for 13.4 years, whereas the English learners in Spain – the group with the smallest L2 instruction – had learnt Spanish for a mean of 2.4 years. This difference may have also resulted in a greater influence of the L2 on the L1 on the part of the Spanish groups than regarding the English learners. Still, it should be noted that, at an individual level, L2 use and L2 instruction alone did not prove to have a determining effect on L1 perception (see Chapter 8). Finally, differences in terms of order effects between the two populations should be acknowledged. Whereas no effect of order was found on the English groups, the Spanish learners were found to present later category boundaries on their L1 when they had previously completed the perception task in English. It is possible that a combination of task familiarity and a potential greater difficulty to control language mode when completing the Spanish test in the second place resulted in a somewhat more L2-like categorization in their L1. Thus, the order effects on the Spanish groups and the lack of them on the English groups may also account for the differences they presented in L1 stop categorization.

In the case of production, no clear effect of L2 experience was found on the L1 of the L2 learners. Neither the English learners of Spanish nor the Spanish learners of English differed from the control groups of their L1. However, one significant difference was revealed between two learner groups sharing the same L1: the highly experienced Spanish group presented significantly greater VOT values than the inexperienced Spanish group in their production of Spanish /k/. Therefore, it seems possible that their L1 production is slowly deviating from the L1 towards more L2-like VOT values. Nevertheless, overall, a straightforward effect of L2 experience was not observed. This finding is in opposition to Flege (1987), who reported that experienced English learners of French and experienced French learners of English living in an L2 setting produced /t/ with values that differed significantly from those of the control groups of their L1 – i.e., significantly shorter in the case of the L1-English speakers and significantly longer when it comes to the L1-French speakers. This discrepancy could be explained by the fact the experienced learners tested in Flege (1987) had spent a considerably longer period of time living in an immersion setting than the participants tested in this study – i.e., about 12 years as opposed to about four years, respectively. Thus, it appears that a longer period of immersion in an L2 setting might be necessary for L1 attrition to take place in production. As a matter of fact, a tendency to use greater – more English-like – VOT values in Spanish is already observed in the highly experienced Spanish group, which, as has been mentioned, produced significantly longer VOT values than the inexperienced Spanish learners.

All in all, it seems that L2 experience has a limited effect on the L1: only the most experienced Spanish group was found to deviate from the Spanish controls in perception and from the inexperienced group – but not from the Spanish controls – in production. No effect of L2 experience was observed in the English groups. It is possible that a greater amount of L2 experience is needed for the L2 to exert a clear influence on the L1, as previous studies reflecting an influence of the L2 on the L1 generally involved participants that had been living in an L2 country for a longer period of time than the participants in this study. Moreover, the fact that the Spanish participants used the L2 to a greater extent and had received more formal instruction in the L2 than the English participants may have also contributed to the former – particularly when it comes to the participants in an immersion setting – presenting a greater influence of the L2 on the L1 than the latter.

9.1.3. RQ.3. Differences between languages and L2 category formation

The third research question was concerned with the differences that L2 learners made between the perception and production of stops in the L1 and in the L2. Moreover, it addressed whether L2 experience had an effect on this difference. According to the SLM's predictions for L1 and L2 similar phones, it was hypothesized that L2 learners may present a shared category for the two languages with phonetic features intermediate to the L1 and the L2. As participants gained L2 experience, their categories were expected to be more L2-like and, possibly, eventually distinct in the two languages. The results obtained in the English and in the Spanish tasks were compared and interpreted in terms of category formation following the predictions of two of the main current L2 speech learning models (SLM: Flege, 1995, 2002, 2007; L2LP: Escudero, 2005, 2009). A different outcome was revealed for perception and production. More specifically, L2 learners made a greater difference between the L1 and the L2 in production than in perception.

Overall, the analysis of the results obtained in the perception experiment indicated that L2 learners perceived L1 and L2 stops similarly. Regarding the English participants, they obtained very similar category boundaries in English and in Spanish. Contrary to predictions, ENEXPinUK's and ENEXPinSP's /p/-/b/ boundaries were slightly later in Spanish than in English, although the difference was very small. This finding further indicates that English learners of Spanish did not tend to use an earlier boundary in Spanish and, in fact, they categorized voiced and voiceless stops in terms of the English VOT contrast in both languages. As has already been discussed, a potential activation of English in the two testing conditions – i.e., in the English and in the Spanish test – due to the nature of the stimuli used, which included from prevoiced to aspirated tokens, could have resulted in a similar – English-like – performance in the two task, especially in the case of the English learners of Spanish.

Whereas the English groups presented L1-like perceptual category boundaries in both languages, all Spanish groups appeared to have 'compromise' (Flege, 1991c) VOT values – i.e., intermediate between those obtained by the English and the Spanish controls. The intermediate values tended to be more L2-like with greater experience, as the highly experienced group obtained the most English-like category boundaries in both languages, whereas the inexperienced group tended to present the most Spanish-like values. Just as the English groups, the Spanish speakers did not generally make a great difference between L1 and L2 stop categorization. In their case, small numerical differences in the expected direction – i.e., they obtained earlier category boundaries in Spanish than in English – were observed, but only in the case of the moderately experienced group's /k/-/g/ boundary was the difference between the two languages significant (English: 24.5 ms; Spanish: 21.3 ms). Moreover, the moderately experienced group also made the greatest numerical – though nonsignificant – difference between the English and Spanish /p/-/b/ boundaries. The moderately experienced learners might be more successful at making a difference between the two languages due to the fact that, even though they had some experience in an immersion setting, they were living in their home country at the time of testing. That is, their past experience with English may have helped them improve their L2 perception, but, given that the experience was past and relatively short (*M* = 8.6 months) it may not have significantly affected the L1. As a matter of fact, their category boundaries did not differ from those of Spanish controls, whereas those of the highly experienced group did. However, given the reduced size of the sample – 10 moderately experienced learners were included in the analysis –, this one significant difference between the two languages should be interpreted with caution.

As has been mentioned, the Spanish groups tended to perform somehow differently in the two conditions – as opposed to the English groups – and their boundaries for English were always numerically later than their boundaries for Spanish, as expected. One possibility is that the presence of English-sounding stimuli – which may activate the language that is not being tested (Grosjean, 2001) – had a greater influence on the L1-English groups than on the Spanish groups due to language dominance. Grosjean's (2001, 2012) Language Mode Theory claims that bilingual speakers who are clearly dominant in one language – such as it is the case of the participants in this thesis – have more difficulty in deactivating the dominant language when the weaker language is being used than the other way around. The stronger language is only expected to be completely deactivated when the context is fully monolingual – i.e., instances or features of the dominant language are not present. That is, the inclusion of an English sounding feature (aspiration) may have triggered the activation of an English mode to a greater extent in the English groups (English-dominant) than in the Spanish groups (Spanish-dominant). Alternatively, or additionally, this difference between the two populations may be related to L2 use. That is, the Spanish participants might be slightly better at perceiving the L1 and the L2 differently because they used English to a greater extent than the English learners used Spanish (see Chapter 8). Another possibility is that differences between the two populations may indicate processes of L2 categorization that are specific to each L1. The process of acquisition of L2 stops may be different for a Spanish learner of English and for an English learner of Spanish because the L1 VOT system of the latter involves a greater range than that of the former. As a result, the English speakers perceived L1 and L2 stops according to English VOT values because their L1 already covers the Spanish VOT range. Conversely, the Spanish groups presented intermediate categories – that were more or less L2-like according to L2 experience – because, by learning English, their VOT range increasingly extended towards greater values.

The production experiment revealed that L2 learners were better at producing L1 and L2 stops differently than they were at perceiving them. More specifically, all English groups were found to use significantly greater VOT values in English than in Spanish in the production of both voiceless and voiced stops. This result is in line with previous studies that have found that L2 speakers are able to produce L1 and L2 stops differently (e.g., Antoniou et al., 2010; Flege, 1990). All L1-English groups produced L1 stops with native-like VOT values, whereas some differences were found in the case of L2 stops. Even though their productions in the two languages were different, the inexperienced group produced significantly longer VOT values than the Spanish monolinguals for all stops. That is, they used numerically intermediate values between the L1 and the L2 in Spanish production that were significantly different from those of the Spanish controls (in line with Flege, 1990). The same was true for the highly experienced group regarding voiced stop production. Conversely, the English group with moderate L2 experience produced significantly different L1 and L2 categories that were comparable to those of native speakers of both languages (in agreement with the heritage speakers in Antoniou et al., 2010). In sum, the group that was the most successful at making a difference between English and Spanish stops and using target-like VOT values in both languages was the moderately experienced group, although all English groups produced L1 and L2 stops with significantly different values. Therefore, it does not seem that L2 experience alone has a straightforward influence on producing L1 and L2 stops differently, particularly when it comes to voiced stops. In line with what has been discussed in sections 9.1.1. and 9.1.2, it is possible that the moderately experienced learners were more successful at producing L1 and L2 voiced stops differently and with target-like values than the highly experienced group due to the amount of L2 instruction received or the carefulness in which they completed the reading task given their occupation – i.e., students of Spanish might have been more careful than the English teachers. As for the Spanish groups, they were successful at producing /p/, /k/ and /g/ differently in the two languages, but only the highly experienced Spanish group did so regarding /b/. In this case, L2 experience did seem to have a clear positive effect on making a difference between L1 and L2 production, as the most experienced group was the most successful. It should also be noted that, in the case of the Spanish groups, L2 experience and L2 use go hand in hand – i.e., the most experienced group used the L2 the most, whereas the least experienced group used the L2 the least. Moreover, there were greater differences in terms of L2 use between the Spanish participants living in an L1 and in an L2 setting than in the case of the English groups. Thus, it is possible that the fact that the Spanish experienced group in the UK produced L1 and L2 stops more differently than the other groups was not only motivated by L2 experience but also by L2 use.

In sum, it appears that L2 learners were more successful at producing L1 and L2 stops differently than they were at perceiving them. This outcome could be interpreted differently according to different L2 acquisition models. It should be noted that the models considered address different dimensions of speech: whereas the L2LP discusses category formation in perception (Escudero, 2005, 2009), the SLM considers production, which is believed to rely on perception (Flege, 1995, 2002, 2007) – i.e., perception leads production[6]. First, category formation in perception will be discussed considering the L2LP's premises, then category formation in production will be interpreted in SLM terms, followed by a more comprehensive view including the results obtained for both dimensions.

The L2LP claims that the L1 and the L2 are two separate systems. When acquiring a similar L2 phone, the equivalent L1 category will initially be copied into the L2 system and gradually evolve towards more target-like values. At an end-state, the L1 and L2 categories will present completely different phonetic values. It should also be noted that the L2LP does not consider regressive CLI – from the L2 into the L1, in this case. Any intermediate perception between the two languages is explained by the parallel activation

[6] The results have not been discussed in terms of the PAM given the nature of the tests conducted. Only an identification test – and not a discrimination test – was carried out, hindering the application of the PAM's premises, which consider how an L2 contrast is mapped onto the L1 according to how accurately the two phones are discriminated.

of the two systems due to the nature of the input. The results obtained in this study can be interpreted in two different ways. On the one hand, it can be considered that the final stage of the L2 learners has not been reached and, as a consequence, their L1 categories still present L1-like values. Particularly, this assumption could explain the performance of the English groups, who categorized both L1 and L2 stops according to English values. However, if we consider the results in the production task, it appears that the English learners are in fact able to make a difference between the L1 and the L2. Language activation would more satisfactorily explain the differences in the perception results obtained by the English and the Spanish groups. Particularly, a bias towards an English-like perception due to the inclusion of aspirated tokens – which are also more meaningful in English than in Spanish – in the Spanish task may explain the fact that the English learners categorized stops according to English VOT values in both the English and the Spanish task. As for the Spanish groups, a combination of both premises may explain the outcome of the experiment. Their intermediate perception in both languages may reflect their English categories, which are still developing and present L1-like traits – to a greater or lesser extent – and which were used in both tests due to the activation of English. The more experienced the learners were, the more L2-like the values they used to categorize stops were, indicating that their L2 categories were developing towards target-like values.

Next, the results of the English and Spanish experiments will be discussed considering the SLM's premises. The production experiment revealed that, in most cases, the L1 and the L2 categories were found to differ significantly from one another, particularly when it comes to voiceless stops. That is, it appears that, generally, the L2 learners used different VOT values in production according to the language that they were speaking. When it comes to voiced stop production, the Spanish learners seemed to be using their L1 category in the L2 production of /b/, as their L1 and L2 VOT values did not differ from one another – except in the case of the most experienced group – and their L2 production differed from that of English native speakers. This finding could be interpreted in terms of category assimilation, that is, their L2 category was assimilated to their existing L1 category and category formation was blocked (Flege, 1995). As has been previously discussed, the acquisition of English voiced and voiceless stops may be different due to the phonetic status of prevoicing and the phonological use of aspiration. Given that voice-lead VOT is sometimes used in English, Spanish learners may have simply assimilated English voiced stops to their existing Spanish categories, which are produced with prevoicing. Still, the SLM does not consider category assimilation an end state, as new categories may still be created with adequate input. As a matter of fact, the highly experienced learners of Spanish appear to be more successful at making a difference between L1 and L2 voiced stops and present somehow more target-like values than the less experienced groups – although they still differed significantly from English controls in /b/ production.

However, a comprehensive analysis of the results – considering both the perception and the production experiment – gave rise to an alternative interpretation. Before moving on with the discussion, a few premises of the SLM need to be reminded. First of all, contrary to the L2LP, the SLM claims that the L1 and the L2 coexist in a common phonological space, which may result in bidirectional CLI – i.e., from the L1 onto the L2 and from the L2 onto the L1. Moreover, L2 perception is believed to precede L2 production. Thus, even though the model explicitly addresses production, it is taken for granted that the perceptual dimension will have developed – at least – to the same extent. However, the results obtained in the perception and the production experiments challenge this assumption, given that the L2 learners were generally more accurate and made a greater distinction between the two languages in production than in perception.

Nevertheless, this mismatch between the outcome of the two experiments can be explained in terms of the ‘merger hypothesis’ (MH, e.g., Flege, 2007), especially in the case of the Spanish learners. The MH posits that L2 learners may have a single shared category in the two languages that presents intermediate values between the L1 and the L2. Depending on the input received, the category will present features that are more L1-like or L2-like. The MH is similar to Escudero’s understanding of language mode activation in that it allows for an intermediate performance between L1 and L2 values that may vary according to the nature of the input received. However, the two models differ in that the SLM assumes that intermediate realizations are realizations of the same (shared) category, whereas the L2LP considers that the L2 system is separate from the L1’s and that intermediate perception stems from a parallel activation of the two systems.

As has already been discussed, the input presented in the perception task in Spanish included English-like features, whereas the production experiment only involved language-specific information in context. Therefore, it is plausible that the merged category of the Spanish learners was implemented with somewhat more English-like features in both the English and the Spanish task, whereas more language-specific VOT values were used in production. Moreover, the merged category of the Spanish learners seems to be evolving towards more target-like values, as experienced L2 learners produced more target-like VOT values in the L2. The English groups may as well present a merged category, which may have used more extreme English-like values in perception than the Spanish groups – who used L1-L2 intermediate values – due to English dominance. As mentioned above, given that English is their dominant language, the English learners may have activated English to a greater extent in the Spanish task – i.e., in their weaker language – than the Spanish groups, who were Spanish dominant. Just as in the case of the Spanish learners, the merged category of the English learners seems to be produced with phonetically more target-like values as they gain L2 experience, particularly regarding voiceless stops. In sum, both the English and the Spanish learners presented a shared phonological category for the two languages that was implemented differently in the perception experiment and in the production experiment, possibly due to issues related to the nature of the input and the resulting language mode activation, as has been mentioned. Whereas in perception similar values were obtained for the two languages, the L2 learners were more successful at producing the L1 and the L2 phones with phonetically distinct values. Other factors, such as L2 instruction, may help the English groups develop target-like values in voiced stops, as the moderately experienced group outperformed the highly experienced group.

In short, the differences obtained in the Spanish vs. the English condition have first been discussed for the perception and the production results separately and, then, considering both modalities globally. Regarding perception, all groups – with the only exception of /k/-/g/ perception by the moderately experienced Spanish group – perceived L1 and L2 stops similarly, which suggests that L2 learners present a single shared category for the two languages. Alternatively, it is possible that English was activated in both the English and the Spanish tests – particularly in the case of the English groups, who were English-dominant. This may have resulted in an English-like perception in the two conditions by the English groups and in an intermediate perception – between the L1 and the L2 – by the Spanish learners. Moreover, the greater numerical difference between the two languages on the part of the Spanish learners can also be explained by the fact that English was activated to a greater extent by the English groups due to English dominance. In the case of production, results suggest that all participants presented separate categories for voiceless stops. As for voiced stops, all English groups as well as the Spanish learners in the UK, perceived L1 and L2 stops differently. The two Spanish groups in Spain only made a difference between the L1 and the L2 regarding /g/ and used their L1 category for /b/ production in both languages. A global examination of the results suggests that L2 learners may present a merged category that shows L1 or L2-like phonetic characteristics to a greater or lesser extent depending on the linguistic nature of the input. Given that the production task only included language specific features, participants were able to make a greater distinction between the two languages than in the perception task, which included English-like features in both conditions.

9.1.4. RQ.4. The relationship between perception and production

The fourth research question addressed the relationship between perception and production both in the L1 (RQ.4.1.) and in the L2 (RQ.4.2.). Moreover, it was investigated whether L2 experience had an effect on the relationship between the two dimensions (RQ.4.3.). In spite of the conflicting results reported in the literature (e.g., Newman, 2003; as opposed to Shultz et al., 2012), this thesis follows the assumptions of the Motor Theory (Liberman et al., 1962; Liberman et al., 1967; Liberman and Mattingly, 1985) and the Direct Realist Approach's (Fowler, 1986, 1990) and it was hypothesized that there would be a relationship between perception and production in the L1 at an individual level. In other words, it was expected that those participants that presented late category boundaries would also present great VOT values and the other way around. As for the L2, a less straightforward link was expected, since the L2 categories may still be developing. It was also assumed that a greater amount of L2 experience would result in a more straightforward alignment between the two dimensions in the L2 and a somewhat less clear relationship in the L1, as the L2 might have influenced the two modalities in the L1 to different extents. The means obtained by each participant for the production of each category under study – /p/, /b/, /k/, /g/ – were inspected individually and then correlated to the category boundaries they obtained for each voiced-voiceless homorganic pair. Separate analyses were conducted for the learners with different L1s – i.e., one test for the L1-English participants and one test for the L1-Spanish participants – and for each language – English and Spanish. Control groups were analyzed separately. Furthermore, in order to investigate the effect of experience, a separate analysis was also conducted for each group.

Regarding the L1, no clear link and, thus, no significant correlations between category boundary and VOT production were found in any group for any of the stop categories analyzed – with the only exception of /p/ and /p/-/b/ in the case of the highly experienced Spanish group. In fact, no significant correlations were observed for the functional monolingual (control) groups: those participants presenting late category boundaries – in comparison with other participants in the same group – did not also present great VOT values. Regarding the learner groups, only one significant correlation was found, but it was in the opposite direction to expectations. Those participants in the most experienced Spanish group that had later category boundaries presented shorter VOT values, as compared to other participants in the same group. This lack of a clear relationship between the two dimensions is in agreement with Shultz et al. (2012), which failed to find a relationship between VOT and f0 weighting in perception and its use in production. The fact that no significant correlations were observed for the control groups already suggested that no clear correlations were going to be found for the other groups, as it was the case. This finding challenges the Motor Theory's and the Direct Realist Approach's (e.g., Fowler, 1986, 1990; Liberman et al., 1962) view that phonological categories are based on the speakers' own speech, as well as some studies that have found a clear link between the two dimensions (e.g., Perkell et al., 2004; Newman, 2003). However, it is in agreement with some previous studies that have failed to find a straightforward relationship in the L1 (e.g., Bailey & Haggard 1973; Shultz et al. 2012). For instance, Bailey and Haggard (1973) failed to find a significant correlation between the perception and the production of VOT and of f0 to contrast voiced and voiceless stops.

Similar results were observed in the case of the L2. Only two significant correlations – both regarding voiced stops – were found. More specifically, SPEXPinSP's /p/-/b/ boundary and /b/ production and SPINEXPinSP's /k/-/g/ boundary and /g/ production were significantly correlated. Whereas the former was in the expected direction – i.e., longer prevoicing was correlated with an earlier category boundary –, the opposite was true for the latter – i.e., longer prevoicing values were correlated to later category boundaries. Thus, the small number and the direction of the significant correlations indicate that there was not a straightforward relationship between the two measures in the L2. This finding challenges the SLM's assumption that there is a link between the two dimensions and that perception leads production and is in line with previous studies that have failed to find a clear relationship with consonants in the L2 (e.g., Bailey & Haggard, 1973; Gorba, 2016; Lev-Ari & Peperkamp, 2013; Sheldon & Strange, 1982; Williams, 1977). For instance, a study prior to this thesis also failed to find a link between category boundary location and amount of VOT, as the participants that obtained the latest category boundaries did not always produce the greatest VOT values and the other way around (Gorba, 2016).

However, the fact that a straightforward correlation between category boundary and VOT production was not found does not necessarily mean that a link between the two dimensions does not exist. The lack of an alignment between the results obtained in the two experiments may stem from the fact that the measures used to quantify perception – category boundary – and production – mean VOT – were not comparable. One issue that should be taken into consideration is variability. Whereas the production measure for each participant was calculated using all the VOT productions of each phone – which may vary from one another –, the measure used for perception – category boundary – was obtained based on the responses of each participant on a VOT continuum with two specific endpoints. That is, the production measure allowed for more variability than the perception measure.

Another difference between the perception and production experiment that should be considered is the lexical characteristics of the stimuli involved in each task. Grosjean (2001) claims that the inclusion of stimuli – both perceptual and lexical – that resemble the language that is not being used, or simply the belief that they may be included, will bring about the activation of this language. Whereas in the production task the target phones were embedded in a real word and in a sentence in the target language, the stimuli in the perception task were isolated syllables that had to be identified as the initial syllable of a word in the language that was being tested. Thus, the language-specific lexical information of the production task was greater than that of the perception task, favouring the activation of the desired language mode in the production task to a greater extent than in perception. Moreover, some of the stimuli presented in the perception task carried meaning in English but not in Spanish. The sequences /pi/, /bi/ and /ki/ have a meaning by themselves in English – 'pea'/'pee', 'bee' and 'key', respectively. In the case of Spanish, /bi/ – *vi*, 'I saw'– or *bi*, the abbreviation for bisexual, and /pi/ – 'pi', the mathematical number pi (π) – are also meaningful in Spanish. However, the fact that *vi* is spelled with a <v> and the low frequency of the abbreviation *bi* could have limited the lexical access to these words, as the response alternatives provided in the test referred to 'b'. It should also be noted that 'pi' is not a frequent word in Spanish, as it is exclusively used in the field of mathematics. As a result, it is also possible that the English lexicon was accessed in the Spanish task, resulting in an activation of the English language mode.

As a matter of fact, a possible bias towards an English-like identification might have been observed in the two populations tested: the English groups performed according to English-like values in both conditions and the Spanish learners used intermediate values – which were more English-like as participants gained experience – in the two languages. Still, in the case of the Spanish groups, numerical differences between the L1 and the L2 – and a significant difference in the case of /k/-/g/, particularly regarding the moderately experience group – were found in the expected direction. In other words, the Spanish learners obtained later category boundaries in English than in Spanish, which may indicate a greater effect of the English bias on the English participants.

The apparent different degree of activation of English between the two populations might be related to language dominance, that is, an English language mode might have been triggered to a greater extent in the English participants because their dominant language was English, whereas the opposite was true in the case of the Spanish learners. Grosjean's (2001, 2012) Language Mode theory supports this idea. As mentioned above, Grosjean (2001, 2012) claims that highly language dominant bilinguals – such as it is the case of the learners in this study – tend to mix codes to a greater extent than more balanced bilinguals when using their non-dominant language. However, they may be able to deactivate the weaker language to a greater extent in a completely monolingual environment. The English learners' weak language (Spanish) was influenced by their stronger language (English) in the Spanish perception task, which included stimuli that resembled their stronger language. However, they were able to better control language mode in the Spanish production task – and, therefore, avoid English influence to a greater extent than in perception – given that the task was presented entirely in one language. In the case of the Spanish groups, the Spanish task – which included tokens that resembled English – was set in their dominant language and, thus, their weaker language (English) may have exerted smaller influence on their performance. Similar studies also support this claim. For instance, Elman et al. (1977), reported that balanced English-Spanish bilinguals were more successful than highly-language dominant bilinguals when it comes to making a difference between the identification of stops in English and in Spanish. It should be noted, however, that language mode activation would only explain the performance of the L2 learners in the perception task but could not have affected the control groups, which were functional monolinguals.

The order in which each modality is acquired in the L2 is also worth discussing. Contrary to the SLM's and PAM's assumption that accurate perception is a prerequisite for accurate production, the general results obtained in the two experiments suggest that L2 learners were more successful at producing L2 stops than at perceiving them, particularly in the case of the English learners. All English groups presented English-like category boundaries in the two languages in perception but were able to make a distinction between the L1 and the L2 in production. In the case of the Spanish learners, however, the order in which the two modalities developed is not clear. On the one hand, all Spanish groups were more successful at producing L1 and L2 stops differently than they were at perceiving them. On the other hand, the moderately experienced group was more target-like in perception than in production, whereas the other two groups performed similarly in the two tasks – i.e., the inexperienced groups tended to differ from English controls whereas the highly experienced group tended to perform in a target-like manner. Therefore, the order in which the two modalities were acquired cannot be established for the Spanish groups. Moreover, the fact that a relationship between the perceptual and the production measures was not found also suggests that the two dimensions were not acquired in parallel.

All in all, no link between perception and production was found neither in the L1 nor in the L2, challenging the assumptions made by some theoretical models such as the Motor Theory and the Direct Realist Approach – regarding the L1 – or the SLM – regarding the L2. Consequently, no effect of L2 experience could be observed. Moreover, the SLM's and PAM's belief that perception precedes production was challenged, particularly considering the results obtained for the English learners. Nevertheless, these result needs to be interpreted with caution, as methodological limitations such as the comparability of the two measures and the linguistic nature – both phonetic and lexical – of the stimuli used in each of the tasks should be taken into account. Further research using comparable measures to quantify perception and production needs to be conducted in order to cogently determine the relationship between the two dimensions.

9.2. Limitations and further research

The present study has a number of methodological limitations that could be improved in further research. One of the main limitations is that the factor examined – i.e., L2 experience – has two major components that were confounded – namely amount of L2 experience and language setting. That is, the participants that presented the greatest amount of L2 experience were living in an L2 setting at the time of testing, whereas the learners with intermediate and no experience were tested in an L1 setting. This decision was made for two reasons. First, in an attempt to expand the design of Flege (1987), the current study aimed at a mirror image group design with the same number of groups and levels of experience – two groups in an L1 setting and one in an L2 setting – for each L1. Flege's (1987) design was more limited, as it included three different levels of L2 experience – two in an L1 setting and one in an immersion setting – for one L1 (English) but only one group – living in an immersion setting – was tested for the other L1 (French). Moreover, the profile of the participants in each of the groups implied differences in the amount of time spent in an L2 setting. That is, the moderately experienced groups were university students who had spent a few months in an L2 setting as part of their year/term-abroad, whereas the highly experienced groups were made of professionals who had moved to an L2 setting to work and, as a result, had been living there for a longer period of time. Therefore, further research examining L2 learners with a comparable amount of L2 experience living in the two settings needs to be carried out in order to determine the effect of these two factors on L1 and L2 perception and production separately.

Another issue concerning the participants tested is the difference between the English and the Spanish groups in terms of amount of L2 instruction, L2 AOL, L2 use and language dominance. The Spanish participants had learnt English for a longer period of time than the English groups and had also started to learn it at a considerably earlier age. This difference stems from the fact that English is a mandatory course in the Spanish curriculum in primary and secondary school, whereas the opposite is not the case in the UK. Therefore, recruiting comparable L1-Spanish and L1-English groups in terms of AOL and L2 instruction was very difficult. In addition, the Spanish learners were found to use their L2 to a greater extent than the English learners. A greater presence of English in the media and in entertainment may have played a role in this difference. Furthermore, the English group living in Spain used their L1 at work – as they were mostly English teachers –, whereas the Spanish participants living in London spoke mainly English. Therefore, comparisons between the groups with different L1s and claims about the effect of directionality on L2 acquisition need to be made with caution. Even though the two populations were comparable in terms of L2 experience, they differed in amount of L2 instruction, degree of language-dominance and amount of L2 use, which may have also influenced their L2 and L1 perception and production, as some previous studies have reported (e.g., Flege & Fletcher, 1992 and Tahta et al., 1981, respectively). As a matter of fact, different results were found for the two populations, especially when it comes to perception. An effect of L2 experience on both the L1 and the L2 was revealed for the Spanish learners – i.e., a greater amount of L2 experience resulted in more target-like L2 perception, as well as in less native-like L1 perception. Nevertheless, no effect was found for the English groups, who perceived both L1 and L2 stops according to English VOT values.

The relatively reduced sample size also hinders the conclusions drawn from this thesis. Even though a total 81 participants were included in the analysis, each group only included between nine and 11 participants. Therefore, the results obtained in the group comparisons and, especially, in the statistical analyses conducted on each group separately – e.g., the correlations conducted between category boundary and mean VOT values – should be interpreted with caution. A bigger sample size was not analyzed for a number of reasons. In the first place, although 92 participants were initially tested, 12 had to be discarded, mainly because they presented a complex linguistic background that differed from the rest of the participants. There were other factors that made recruiting a larger number of participants difficult, such as time limitations – the experimenter was in London only for a limited period of time – and the fact that participation was mostly on a voluntary basis given the lack of funds to pay the participants

There were methodological limitations regarding the nature of the stimuli used to test perception that may have hindered the activation of a complete monolingual mode in the Spanish condition, as discussed above. Recall that, in order to test the mirror-image populations in both their L1 and L2, the same stimuli – which included aspirated tokens, a feature present only in English – were used in the English and Spanish conditions.

Moreover, the sequences used (/pi/, /bi/, /ki/ and /gi/) were more meaningful – or more frequently used – in English than in Spanish. As a result, English may have been activated during the Spanish perception task, possibly limiting the conclusions reached regarding CLI and category formation in perception, as well as the relationship between perception and production. Further research should consider language mode control in the creation of the stimuli – as it has been done in some previous studies, such as Gonzales and Lotto, (2013) and Casillas and Simonet, (2018) – in order to assess L2 category formation. Moreover, ideally, more tasks could have been included to test perception, particularly a discrimination task, in order to obtain more comprehensive data on the categorization of L1 and L2 stops by L2 learners. Nevertheless, given the duration of the experiment and the number of tasks that participants had to complete, it was decided to only include an identification task, which is the most frequently used task to test VOT categorization in the literature (e.g., Casillas & Simonet 2018; Caramazza et al. 1973; Elman et al., 1977; Gonzales & Lotto, 2013; Hazan & Boulakia 1993). Another methodological factor that complicated the comparison between perception and production was the variability of each of the measures, being the measure used for production more subject to variability than the one for perception. Comparable tasks in terms of language mode activation as well as comparable measures in terms of variability are needed to draw conclusions on the relationship between the two dimensions.

Another aspect that needs to be acknowledged is the fact that the production task was a carrier sentence reading task and was, thus, very controlled. It was decided to use this type of task in order to control for the context in which the target phones appeared, since voiced stops in Spanish are spirantized in all context but at the beginning of the utterance or after a nasal stop. It is possible, thus, that different results might have emerged with spontaneous speech. Further research should consider the inclusion of a more naturalistic task in order to assess the L2 production of L2 speakers.

Finally, it should also be noted that the only voicing cue that was investigated in the current thesis is VOT, whereas other secondary cues, such as f0, have been found to influence the perception of stops as voiced or voiceless (e.g., Idemaru & Holt, 2011; Whalen et al., 1993). This is particularly true in those cases where VOT values are ambiguous – i.e., intermediate between those found in a voiced and voiceless category. Still, secondary cues for voicing have not been neglected in this study, as they were controlled in the creation of the stimuli (see section 4.2.1.1). More specifically, f0 and F1 transition, burst duration and intensity and vowel intensity were modified in order to obtain ambiguous values – i.e., intermediate between voiced and voiceless stops. As a result, participants were expected to rely solely on VOT in the identification of stops as voiced and voiceless.

In short, methodological issues concerning possible confounding factors in the participants and the potential limitations of the tasks – especially regarding language mode control – hinder the interpretation of the comparisons between the two populations and between the results obtained in the production and perception experiments. Further research should consider these limitations in order to be able to provide more straight-forward answers to questions that remain unanswered – in this study in particular and in the field of L2 speech acquisition in general –, including L2 category formation and the relationship between perception and production.

9.3. Summary, implications and conclusions

The present thesis has investigated the perception and production of L1 and L2 initial stops in two populations, namely in L1-English learners of Spanish and in L1-Spanish learners of English who varied in amount and type – i.e., present or past – of L2 experience. Instances of bidirectional CLI were observed in the Spanish learners, especially regarding perception, whereas, only the L1 was found to influence the L2 in the case of the English learners. Previous research examining L2 speakers with a greater amount of L2 experience than participants in this study did find clear instances of L2 influence on the L1 (e.g., Flege 1987). Thus, it seems that a smaller amount of L2 experience may be sufficient for it to modulate the influence of the L1 on L2 than the other way around. Moreover, a greater effect of L2 experience – on both the L2 and the L1 – was found for the Spanish learners than for the English learners. This difference could be explained by the fact that, even though the Spanish and the English groups were comparable in terms of L2 experience, the Spanish groups had received more formal education in the L2 and used the L2 to a greater extent than the English groups. In addition, the potential bias towards an English language mode activation in the perception task might have influenced the results in a way that the English speakers performed in an English-like manner in both languages, whereas the Spanish participants presented intermediate values – which may actually represent their English categories.

The results in the production task highlighted differences between the acquisition of L2 voiceless and L2 voiced stops. More specifically, the Spanish learners had more difficulty in producing voiced stops differently in English and Spanish than they did regarding voiceless stops. The Spanish learners produced longer VOTs for English /p/ and /k/ – and also /g/ – than for the Spanish counterparts, but the production of /b/ in the L1 and in the L2 of the groups tested in Spain did not differ significantly. It should be noted, though, that the use of voice-lead VOT is also found in speakers of English, although the most common production is with short-lag VOT. Given that prevoicing is a possibility in English, Spanish speakers may have assimilated L2 voiced stops to the L1 category, blocking the creation of a new L2 category. As for the English speakers, L2 experience was not found to have a determining effect on the accuracy of L2 voiced stop production, as the moderately experienced learners outperformed the highly experienced learners. L2 experience may not suffice to learn the phonological status of prevoicing in Spanish – as opposed to its phonetic status in English – and other factors such as the frequency of use of prevoicing in the L1 and amount of L2 experience and L2 use may play a role. However, none of these factors proved to have a significant effect on the production of L2 voiced stops of English learners of Spanish. The use of VOT for voiced stops in the L1 was not found to be significantly correlated to its use in the L2 and, only a few significant results of the effect of other factors on L1 and L2 performance were revealed, being some of them in the opposite direction to what was predicted. Thus, further research on the acquisition of prevoicing as a phonological feature is required in order to determine which factors help its acquisition, as only very few studies have investigated it (Baese-Berk, 2019; Casillas, 2019; Zampini, 1998).

The results obtained in the two languages were also discussed in terms of the current L2 speech acquisition models, more specifically following the L2LP's (Escudero, 2005, 2009) and the SLM's (Flege 1995, 2007) premises. The perception results indicated that all groups tended to perceive L1 and L2 stops similarly, whereas a greater distinction was made in production. However, the Spanish groups assimilated the L2 categories for voiced stops to the existing L1 category, which presented voice-lead VOT values. The fact that L2 learners were generally more successful at producing L1 and L2 stops differently than at perceiving them was interpreted as a result of the activation of English in the Spanish perception task – following Grosjean's Language Mode theory and Escudero's interpretation –, since the presence of English-like features in the Spanish task might have resulted in a bias towards an English-like categorization. In SLM's terms, the nature of the input received in the perception task may have resulted in an English-like perception in the two languages – particularly in the case of the English groups, who were highly dominant in English – and in more target-like realizations in the production tasks. That is, the same merged category – which may present more L1-like or more L2-like features – was implemented differently in the perception and in the production tasks according to the linguistic information they presented – which was language-specific in production and more English-like in perception. In short, the findings of this study can be interpreted both in terms of the SLM and the L2LP, because, even though they differ in their understanding of how the L1 and the L2 interact – i.e., in a common vs. a separate phonological space, respectively – both models allow for intermediate realizations as well as target-like realizations of the same phones depending on the nature of the input received. Moreover, Grosjean's Language Mode theory is key to understanding the different performance of the L2 learners in the production and in the perception tasks.

The present thesis has also attempted to shed some light on the relationship between perception and production, which has obtained opposing results in the literature. In this case, no clear relationship could be established between perception and production, both in the L1 and the L2. This finding challenges theoretical assumptions of some L1 acquisition models, such as the Motor Theory (Liberman et al. 1967; Liberman and Mattingly 1985) and the Direct Realist Approach (Fowler 1986, 1990), and L2 acquisition models, like the SLM (Flege 1995, 2007). However, methodological differences, such as the variability of the measures and the nature of the tasks – could have resulted in a difficult evaluation of the relationship between the two modalities. More specifically, the production measure allowed for greater variability than the perception measure, since the former was based on a number of productions and the latter on responses on a given continuum with two endpoints. Moreover, the production task – which included full monolingual sentences – may have set language mode more accurately than the perception task – which presented syllables in isolation, some of which sounded particularly English-like.

Regarding its contribution, the present thesis is relevant to the field of L2 speech in that it investigates issues that are understudied – such as the effect of L2 experience on the L1 or the acquisition of prevoicing – and that have resulted in conflicting findings in the literature – i.e., the relationship between perception and production. The main contribution of this study is that it has investigated the effect of L2 experience on the L1, which has received scarce attention in previous research – particularly in the case of perception. As a matter of fact, to my knowledge, no previous studies (except Gorba, 2016, 2018) have investigated the effect of L2 experience – understood as amount of time spent in an L2 setting – on the perception of L1 stops. In spite of the different linguistic reality of the two languages tested, the current study is one of the few that included two mirror image populations – in terms of L1 and L2 –, which permitted the comparison of the results obtained for English learners of Spanish and Spanish learners of English. As a matter of fact, the two populations have been found to encounter different difficulties in the acquisition of L2 stops. For instance, English learners of Spanish presented greater difficulty in perceiving L2 stops accurately than Spanish learners of English. Furthermore, the current thesis did not only examine voiceless stops but also voiced stops, which have been understudied in the literature, especially when it comes to the acquisition of prevoicing. As a matter of fact, to my knowledge, only a few studies have investigated the acquisition of prevoicing in English speakers (Baese-Berk, 2019; Casillas, 2019;

Zampini, 1998). In addition, despite the methodological limitations, the testing of both perception and production has allowed the comparison of the two modalities both in the L1 and in the L2, a matter in which previous studies have found opposing results (e.g., Newman, 2003 vs. Shultz et al., 2012).

The outcome of this thesis may have some implications related to the role of L2 experience and other factors on L2 acquisition, the current L2 speech acquisition models and the teaching of L2 stops. The main factor investigated was L2 experience, which has been defined in different ways in the literature and was understood mainly as length of residence in the present thesis, following Flege's research (e.g., 1987). L2 experience has been found to have an effect on L2 and – to a lesser extent – L1 perception and production. Results indicate, regarding the English learners, that a greater amount of L2 experience results in a more target-like production of the L2 and, in the case of the Spanish learners, in a more target-like production and perception of the L2 and a less native-like L1 perception. Still, differences in the results between the two populations under study cannot be accounted solely on L2 experience, but may also be related to other factors, such as L2 use, L2 instruction and the different ranges of VOT used to contrast voicing in the two languages. Thus, even though length of residence has a considerable effect on CLI, other factors need to be considered.

One of the goals of the current thesis was to provide evidence in favour or against the main L2 speech acquisition models (Escudero, 2005, 2009; Flege, 1995, 2002) and, thus, this study has been interpreted according to their premises. The outcome of the experiments – along with the results obtained in previous studies (e.g., Bailey & Haggard, 1973; Lev-Ari & Peperkamp, 2013; Sheldon & Strange, 1982; Williams, 1977) – suggest that the SLM's and the PAM's assumption that perception leads production in L2 acquisition may need to be revisited. As a matter of fact, the relationship between the two modalities appears to be more complex than simply one preceding the other and requires further research to be understood – especially using comparable tasks and measures. Moreover, the relationship between the two modalities and how they interact should be thoroughly discussed in the L2 speech models, as their understanding will be a great milestone in the field of L2 acquisition. In fact, this knowledge will serve as guidance to teach L2 pronunciation more efficiently – e.g., knowing whether the perceptual or production abilities (or both, simultaneously) should be trained first.

A methodological factor that may account for the different results obtained in the perception and production experiment is language mode activation. As has been mentioned, the undesired potential activation of English in the Spanish tasks may have resulted in a somehow more English-like identification of stops in the Spanish task, especially in the case of the English learners. Thus, results suggest, as Grosjean (2001, 2012) claims, that different results may be obtained according to the degree of activation of each language. This outcome points to interesting future research directions that may help gain knowledge in the process of CLI and what linguistic scenarios may trigger it – to a greater or lesser extent. Understanding how CLI is caused is of great importance to explain the processes of L2 acquisition and language attrition and, thus, language activation should be attached greater importance by the speech acquisition models.

Finally, the different results obtained for voiced and voiceless stops – i.e., voiced stops appear to be more difficult to acquire than voiceless stops – implies that a greater effort may need to be made in the teaching of voiced stops. Whereas the production of voiceless stops is generally addressed in the L2 classroom – i.e., Spanish learners of English are instructed to aspirate and English learners of Spanish are instructed not to aspirate –, prevoicing has traditionally received less attention. Therefore, just as in the case of the use of aspiration, L2 learners may need to be explicitly instructed on how to use this feature. Particularly, English learners of Spanish should learn that their L1 voiced stop category needs to be split into two in their L2, since Spanish contrasts short-lag (voiceless stops) and voice-lead (voiced stops) VOT. As for the Spanish learners, they should become aware that prevoiced voiced stops are less common than those produced with short-lag VOT in English. That is, even though Spanish learners of English may not need to learn to produce voiced stops with short-lag VOT – given that prevoiced stops are allowed in free variation in English –, they should learn that voiced stops in English may be produced – in fact more frequently – with a similar VOT to Spanish voiceless stops.

In conclusion, this study has analyzed the effect of L2 experience on L1 and L2 perception and production and has investigated the relationship between the two dimensions. Instances of bidirectional CLI have been found, although the L2 appears to affect the L1 to a lesser extent than the other way around and a greater amount of L2 experience is required for regressive CLI to take place than for progressive CLI. Moreover, L2 experience was found to exert a greater effect on the L1 and L2 of Spanish learners of English than on English learners of Spanish, possibly due to the difference in VOT range between the two languages under study and differences concerning other factors, such as L2 use and amount of L2 instruction. Furthermore, L2 voiced stops were found to be more difficult to acquire than voiceless stops. The phonological use of prevoicing in Spanish – as opposed to its phonetic status in English – may have resulted in a greater difficulty in producing voiced stops accurately than voiceless stops. Finally, no straightforward link between perception and production has been established neither in the L1 nor in the L2, indicating a complex relationship between the two modalities and demonstrating the difficulty of finding comparable methodologies and measures that allow valid comparisons. Further research is required to better understand the processes that take place in L2 speech acquisition and to validate the current L2 speech models. A careful control of language mode – which should be purposely set to both a monolingual mode and to a bilingual mode – is essential to understand how the L1 and the L2 interact and to disentangle the nature of the relationship between perception and production.

References

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

Abramson, Arthur. S. & Leigh Lisker (1970). Discriminability along the Voicing Continuum: Cross-Language Tests. *Proceedings of the Sixth International Congress of Phonetic Sciences* 196 (7): 569–573.

Ainsworth, William. A. & Kukdip. K. Paliwal (1984). Correlation between the Production and Perception of the English Glides /w, r, l, j/. *Journal of Phonetics* 12 (3), 237.

Aliaga-García, Cristina & Joan Carles Mora (2009). Assessing the Effects of Phonetic Training on L2 Sound Perception and Production. *Recent Research in Second Language Phonetics/Phonology: Perception and Production* 231.

Antoniou, Mark, Catherine T. Best, Michael. D. Tyler & Christian Kroos (2010). Language Context Elicits Native-like Stop Voicing in Early Bilinguals' Productions in both L1 and L2. *Journal of Phonetics* 38 (4): 640–653. https://doi.org/10.1016/j.wocn.2010.09.005

Aoyama, Katsura, James E. Flege, Susan G. Guion, Reiko Akahane-Yamada & Tsuneo Yamada (2004). Perceived Phonetic Dissimilarity and L2 Speech Learning: The Case of Japanese /r/ and English /l/ and /r/. *Journal of Phonetics* 32 (2): 233–250.

Baese-Berk, Melissa. M. (2019). Interactions between Speech Perception and Production during Learning of Novel Phonemic Categories. *Attention perception* 81(4), 981.

Bailey, Peter. J. & Mark P. Haggard (1973). Perception and Production: Some Correlations on Voicing of an Initial Stop. *Language and Speech* 16 (3): 189–195.

Baker, Wendy & Pavel Trofimovich (2006). Perceptual Paths to Accurate Production of L2 Vowels: The Role of Individual Differences. *IRAL–International Review of Applied Linguistics in Language Teaching* 44 (3): 231–250.

Benkí, José R. (2005). Perception of VOT and First Formant Onset by Spanish and English Speakers. In James Cohen, Kara T. McAlister, Kellie Rolstad & Jeff MacSwan (Eds.), *Proceedings of the 4th International Symposium on Bilingualism* (pp .240–248). Somerville, MA: Cascadilla Press.

Bergmann, Christopher, Amber Nota, Simone A. Sprenger & Monika S. Schmid (2016). L2 Immersion Causes Non-native-like L1 Pronunciation in German Attriters. *Journal of Phonetics* 58: 71–86. https://doi.org/https://doi.org/10.1016/j.wocn.2016.07.001

Best, Catherine T. (1994). The Emergence of Native-Language Phonological Influences in Infants: A Perceptual Assimilation Model. In Judith C. Goodman & Howard C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words.* (pp. 167–224). The MIT Press.

Best, Catherine. T. (1995). A Direct Realist View of Cross-Language speech perception. *Speech Perception and Linguistic Experience. Issues in Cross-Language Research*, April: 167–200. https://doi.org/10.1016/0378-4266(91)90103-S

Best, Catherine T., Gerald McRoberts & Nomathemba Sithole (1988). Examination of Perceptual Reorganization for Nonnative Speech Contrasts: Zulu Click Discrimination by English-Speaking Adults and Infants. *Journal of Experimental Psychology. Human Perception and Performance* 14: 345–360. https://doi.org/10.1037//0096-1523.14.3.345

Best, Catherine T., & Michael D. Tyler (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In Ocke-Schwen Bohn &

Murray J. Munro (Eds). *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, 10389 (2): 13–34. https://doi.org/10.1121/1.1332378

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

Boersma, Paul & David Weenink (2016). *Praat* (Version 5.3.56)[Computer software].

Bohn, Ocke-Schwen, & Catherine T. Best (2010). Testing PAM and SLM: Perception of American English approximants by native German listeners. In Katarzyna Dziubalska-Kołaczyk, Magdalena Wrembel & Małgorzata Kul (Eds.), *New Sounds 2010: Proceedings of the Sixth International Symposium on the Acquisition of Second Language Speech* (pp. 43–48).

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

Bohn, Ocke-Schwen, & James. E. Flege (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition* 14 (2): 131-158.

Bohn, Ocke-Schwen & James E. Flege (1993). Perceptual switching in Spanish/English bilinguals. *Journal of Phonetics* 21: 267–290. https://doi.org/10.1016/S0095-4470(19)31339-7

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

Borden, Gloria, Adele Gerber & Gary Milsark (1983). Production and perception of the

/r/-/l/ contrast in Korean adults learning English. *Language Learning* 33 (4): 499–526.

Brunner, Jana (2011). The Influence of Auditory Acuity on Acoustic Variability and the Use of Motor Equivalence During Adaptation to a Perturbation. *Journal of Speech, Language, and Hearing Research* 3 (727).

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

Caramazza, Alfonso, Grace H. Yeni-Komshian, Edgar B. Zurif & Evelyn Carbone (1973). The acquisition of a new phonological contrast: The case of stop consonants in French-English bilinguals. *The Journal of the Acoustical Society of America* 54 (2): 421-428.

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

Casillas, Joseph V. (2019). Phonetic Category Formation is Perceptually Driven During the Early Stages of Adult L2 Development. *Language and Speech* 28. https://doi.org/10.1177/0023830919866225

Casillas, Joseph. V., Yamile Díaz & Miquel Simonet (2015). Acoustics of Spanish and English coronal stops. In The Scottish Consortium for ICPhS 2015 (Eds.), *ICPhS 2015: Proceedings of the 18th International Congress of Phonetic Sciences*.

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

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

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

Cebrian, Juli (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, 145 (1): EL52–EL58.

Chodroff, Eleanor. (2014). Burst spectrum as a cue for the stop voicing contrast in American English. *Journal of the Acoustical Society of America* 136 (5): 2762.

Chomsky, Noam. (1957). *Syntactic structures.* Mouton & Co.

Cooper, William E., & Marc R. Lauritsen (1974). Feature processing in the perception and production of speech. *Nature*, 252 (5479): 121–123.

Darcy, Isabelle, & Franziska Krüger (2012). Vowel perception and production in Turkish children acquiring L2 German. *Journal of Phonetics* 40 (4): 568–581.

Darcy, Isabelle, Joan Carles Mora & Danielle Daidone (2014). *Attention control and inhibition influence phonological development in a second language, Concordia Working Papers in Applied Linguistics* 5*:* 115–129.

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

Docherty, Gerard J. (1992). *The timing of voicing in British English obstruents* (9). Walter de Gruyter.

Docherty, Gerard J., Dominic Watt, Carmen Llamas, Damien Hall & Jennifer Nycz (2011). Variation in Voice Onset Time along the Scottish-English Border. In Wai

Sum Lee & Eric Zee (Eds.), *ICPhS 2011: Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 591–594).

Elliott, A. Raymond (1995). Foreign language phonology: Field independence, attitude, and the success of formal instruction in Spanish pronunciation. *The Modern Language Journal* 79 (4): 530–542.

Elman, Jeffrey L., Randy L. Diehl & Susan E. Buchwald (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America* 62 (4): 971–974. https://doi.org/10.1121/1.381591

Escudero, Paola (2000). *Developmental Patterns in the Adult L2 Acquisition of New Contrasts: The Acoustic Cue Weighting in the Perception of Scottish Tense/Lax Vowels by Spanish Speakers*. (Unpublished M.Sc. dissertation). University of Edinburgh.

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

Escudero, Paola (2009). The Linguistic Perception of Similar L2 sounds. *Journal of Neuroscience*: 151–190.

Fathman, Ann (1975). The Relationship between Age and Second Language Productive Ability. *Language Learning* 25 (2): 245–253.

Fernández Rei, Elisa, Alba Aguete Cajiao, & César Osorio Peláez (n.d.). *FOLERPA: Herramienta On-line para el Desarrollo de Experimentos Perceptivos*.

Flege, James E. (1981). The Phonological Basis of Foreign Accent: A Hypothesis. *TESOL Quarterly* 15. https://doi.org/10.2307/3586485

Flege, James E. (1984). The Effect of Linguistic Experience on Arabs' Perception of the English /s/ vs. /z/ Contrast. *Folia Linguistica* 18 (1–2): 117–138.

https://doi.org/10.1515/flin.1984.18.1-2.117

Flege, James E. (1987). The Production of “New” and “Similar” Phones in a Foreign Language: Evidence for the Effect of Equivalence Classification. *Journal of Phonetics, 15*(1), 47–65. https://doi.org/10.1016/S0095-4470(19)30537-6

Flege, James E. (1988a). Effects of Equivalence Classification on the Production of Foreign Language Speech Sounds. In Allan James & Jonathan Leather (Eds.), *Sound Patterns in Second Language Acquisition* (pp. 9–39). Dordrecht: Foris.

Flege, James E. (1988b). The Production and Perception of Foreign Language Speech Sounds. *Human Communication and Its Disorders: A Review – 1988*, 224–401.

Flege, James E. (1990). Production of the English word-final /t/-/d/ contrast by native speakers of Mandarin and Spanish. *Acoustical Society of America Journal* 88 (81). https://doi.org/10.1121/1.2029176

Flege, James E. (1991a). Age of Learning Affects the Authenticity of Voice-Onset Time (VOT) in Stop Consonants Produced in a Second Language. *Journal of the Acoustical Society of America* 89 (1): 395–411. https://doi.org/10.1121/1.400473

Flege, James E.. (1991b). The Interlingual Identification of Spanish and English Vowels: Orthographic Evidence. *The Quarterly Journal of Experimental Psychology Section A* 43 (3): 701–731. https://doi.org/10.1080/14640749108400993

Flege, James E. (1991c). Perception and Production: The Relevance of Phonetic Input to L2 Phonological Learning. Crosscurrents in Second Language Acquisition and Linguistic Theories, 2: 249-289.

Flege, James E. (1995). Second Language Speech Learning: Theory, Findings, and Problems. *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*: 233–277. https://doi.org/10.1111/j.1600-0404.1995.tb01710.x

Flege, James E. (2002). Interactions between the Native and Second-Language Phonetic

Systems. *An Integrated View of Language Development: Papers in Honor of Henning Wode*: 217–244.

Flege, James E. (2003). Assessing Constraints on Second-Language Segmental Production and Perception. In Niels O. Schiller & Antje S. Meyer (Eds.), *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*: 319–355.. Berlin, de Gruyter Mouton.

Flege, James E. (2007). Language Contact in Bilingualism: Phonetic System Interactions. *Laboratory Phonology* 9.

Flege, James E., Ocke-Schwen Bohn & Sunyoung Jang (1997). Effects of Experience on Non-native Speakers' Production and Perception of English Vowels. *Journal of Phonetics* 25 (4): 437–470. https://doi.org/10.1006/JPHO.1997.0052

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

Flege, James E. & Wieke Eefting (1987). Production and Perception of English stops by Native Spanish Speakers. *Journal of Phonetics* 15 (1): 67–83.

Flege, James E. & Kathryn Fletcher (1992). Talker and Listener Effects on the Perception of Degree of Foreign Accent. *The Journal of the Acoustical Society of America* 91: 370–389. https://doi.org/10.1121/1.402780

Flege, James E., Ian R. A. MacKay & Diane Meador (1999). Native Italian Speakers' Perception and Production of English Vowels. *The Journal of the Acoustical Society of America* 106 (5): 2973–2987. https://doi.org/10.1121/1.428116

Flege, James E., Ian R. A. MacKay & Thorsten Piske (2002). Assessing Bilingual Dominance. *Applied Psycholinguistics* 23 (4): 567–598. https://doi.org/10.1017/S0142716402004046

Flege, James E., Murray J. Munro & Ian R. A. Mackay (1995). Effects of Age of Second-Language Learning on the Production of English Consonants. *Speech Communication*, *16*(1), 1–26. https://doi.org/10.1016/0167-6393(94)00044-B

Flege, James E., Grace Yeni-Komshian & Serena Liu (1999). Age Constraints on Second-Language Acquisition. *Journal of Memory and Language* 41: 78–104. https://doi.org/10.1006/jmla.1999.2638

Fowler, Carol A. (1986). An Event Approach to the Study of Speech Perception from a Direct–Realist Perspective. *Journal of Phonetics* 14 (1), 3–28.

Fowler, Carol A. (1990). Calling a Mirage a Mirage: Direct Perception of Speech Produced without a Tongue. *Journal of Phonetics* 18 (4): 529–541.

Fowler, Carol A., Valery Sramko, David J. Ostry, Sarah A. Rowland & Piere Hallé (2008). Cross Language Phonetic Influences on the Speech of French-English Bilinguals. *Journal of Phonetics* 36 (4): 649–663. https://doi.org/10.1016/j.wocn.2008.04.001

Fox, Robert A. (1982). Individual Variation in the Perception of Vowels: Implications for a Perception-Production Link. *Phonetica: International Journal of Speech Science* 39 (1): 1.

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

Gorba, Celia (2016). *The Effect of Foreign Language Experience on the Categorization and Production of Native and Non-native Stops by Spanish Learners of English*. (Unpublished M.A. dissertation). Universitat Autònoma de Barcelona.

Gorba, Celia (2018). The Effect of L2 Experience on the Categorization of Native and Non-native stops by Spanish Learners of English. In Sara Martin, David Owen, and

Elisabet Pladevall-Ballester (Eds.), *Persistence and Resistance in English Studies. New Research*: 163–173. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Grosjean, François (1997). The Bilingual Individual. *Interpreting* 2 (1–2): 163–187.

Grosjean, François (1998). Studying Bilinguals: Methodological and Conceptual Issues. *Bilingualism: Language and Cognition* 1 (2): 131–149. https://doi.org/DOI: 10.1017/S136672899800025X

Grosjean, François (2001). The Bilingual's Language Modes. In Janet Nicole (Eds.), *One Mind, Two Languages: Bilingual Language Processing* (pp. 1–22). Wiley-Blackwell.

Grosjean, François (2012). *Bilingual and Monolingual Language Modes*. Oxford: John Wiley and Sons.

Guion, Susan G., James E. Flege, Reiko Akahane-Yamada & Jessica C. Pruitt (2000). An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults' Perception of English Consonants. *The Journal of the Acoustical Society of America* 107 (5): 2711–2724.

Harada, Tetsuo (2003). L2 Influence on L1 Speech in the Production of VOT. *Proceedings of the 15th International Congress of Phonetic Sciences*, 1085–1088.

Hazan, Valerie L., & Georges Boulakia (1993). Perception and Production of a Voicing Contrast by French-English Bilinguals. *Language and Speech* 36 (1): 17–38. https://doi.org/10.1177/002383099303600102

Herd, Wendy, Devan Torrence & Joy Cariño (2016). Prevoicing Differences in Southern English: Gender and Ethnicity Effects. *The Journal of the Acoustical Society of America* 139 (4): 2217.

Hunnicutt, Leigh & Paul A. Morris (2016). Prevoicing and Aspiration in Southern

American English. *University of Pennsylvania Working Papers in Linguistics*, *22*(1), 24.

Idemaru, Kaori & Lori L. Holt (2011). Production and Perception of English /l/ and /r/ by 4-, 5-, and 8-Year-Old Children. In Wai Sum Lee & Eric Zee (Eds.), *ICPhS 2011: Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 926–929).

Iverson, Paul & Patricia K. Kuhl (1995). Mapping the Perceptual Magnet Effect for Speech Using Signal Detection Theory and Multidimensional Scaling. *The Journal of the Acoustical Society of America* 97: 553–562. https://doi.org/10.1121/1.412280

Jia, Gisela, Winifred Strange, Yanhong Wu, Julissa Collado & Qi Guan (2006). Perception and Production of English Vowels by Mandarin Speakers: Age-related Differences Vary with Amount of L2 Exposure. *The Journal of the Acoustical Society of America* 119 (2): 1118. https://doi.org/10.1121/1.2151806

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

Jun, Sun-Ah, & Islay Cowie (1994). Interference for 'New'and 'Similar' Vowels in Korean Speakers of English. *Ohio State University*. *Working Papers* 43: 117–130.

Keating, Patricia A. (1984) Phonetic and phonological representation of stop consonant voicing, *Language* 60: 286–319 .

Kessinger, Rachel H. & Sheila E. Blumstein (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* 25 (2): 143–168. https://doi.org/10.1006/jpho.1996.0039

Klatt, Dennis H. (1975). Voice Onset Time, Frication, and Aspiration in Eord-initial Consonant Clusters. *Journal of Speech and Hearing Research* 18 (4): 686–706.

Koenig, Laura L. (2000). Laryngeal Factors in Voiceless Consonant Production in Men, Women, and 5-year-olds. *Journal of Speech, Language, and Hearing Research* 43 (5): 1211–1228.

Kuhl, Patricia K. (1991). Human Adults and Human Infants Show a "Perceptual Magnet Effect" for the Prototypes of Speech Categories, Monkeys Do Not. *Perception* 50 (2): 93.

Kuhl, Patricia K. (1993). Innate Predispositions and the Effects of Experience in Speech Perception: The Native Language Magnet Theory. In Bénédicte De Boysson-Bardies, Scania de Schonen, Peter Jusczyk, Peter MacNeilage, John Morton (Eds.), *Developmental neurocognition: Speech and face processing in the first year of life.* (pp. 259–274). Kluwer Academic/Plenum Publishers. https://doi.org/10.1007/978-94-015-8234-6_22

Kuhl, Patricia K. (2000). A New View of Language Acquisition. *Proceedings of the National Academy of Sciences of the United States of America*, 97 (22), 11850.

Kuhl, Patricia K., Conboy, B., Coffey-Corina, S., Padden, D., Rivera Gaxiola, M., & Nelson, T. (2008). Phonetic Learning as a Pathway to Language: New Data and Native Language Magnet Theory Expanded (NLM-e). *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences* 363: 979–1000. https://doi.org/10.1098/rstb.2007.2154

Lenneberg, Eric H. (1967). Biological Foundations of Language. *Hospital Practice* 2 (12): 59-67.

Lev-Ari, Shiri & Sharon Peperkamp (2013). Low Inhibitory Skill Leads to Non-native Perception and Production in Bilinguals' Native Language. *Journal of Phonetics*, 41 (5), 320–331. https://doi.org/10.1016/j.wocn.2013.06.002

Levy, Erika S. (2009). Language Experience and Consonantal Context Effects on

Perceptual Assimilation of French Vowels by American-English Learners of French. *The Journal of the Acoustical Society of America* 125 (2): 1138-1152.

Levy, Erika S. & Franzo F. Law (2010). Production of French Vowels by American-English Learners of French: Language Experience, Consonantal Context, and the Perception-Production Relationship. *The Journal of the Acoustical Society of America* 128 (3), 1290. https://doi.org/10.1121/1.3466879

Liberman, Alvin M, Franklin S. Cooper, Katherine S. Harris & Peter Macneilage (1962). A Motor Theory of Speech Perception. *Proceedings of Speech Communication Seminar*. Royal Institute of Technology Stockholm.

Liberman, Alvin M., Franklin S. Cooper, Donald P. Shankweiler & Michael Studdert-Kennedy (1967). Perception of the Speech Code. *Psychological Review* 74 (6): 431.

Liberman, Alvin M., Pierre C. Delattre & Franklin S. Cooper (1958). Some Cues for the Distinction between Voiced and Voiceless Stops in Initial Position. *Language and Speech* 1 (3): 153–167.

Liberman, Alvin M. & Ignatius G. Mattingly (1985). The Motor Theory of Speech Perception Revised. *Cognition* 21 (1): 1–36.

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

Lisker, Leigh, & Arthur Abramson (1967). Some Effects of Context on Voice Onset Time in English Stops. *Language and Speech* 10 (1): 1–28.

Llanos, Fernando, Olga Dmitrieva, Amanda Shultz & Alexander L. Francis (2013). Auditory Enhancement and Second Language Experience in Spanish and English Weighting of Secondary Voicing Cues. *The Journal of the Acoustical Society of America* 134 (3): 2213–2224. https://doi.org/10.1121/1.4817845

Llisterri, Joaquim (1995). Relationships between Speech Production and Speech Perception in a Second Language. *Proceedings of the 13th International Congress of Phonetic Sciences* 4: 92–99.

Long, Michael H. (1990). Maturational Constraints on Language Development. *Studies in Second Language Acquisition* 12 (3): 251–285. https://doi.org/DOI: 10.1017/S0272263100009165

MacKain, Kristine, Catherine T. Best & Winifred Strange (1981). Categorical Perception of English /r/ and /I/ in Japanese Bilinguals. *Applied Psycholinguistics* 2: 369–390. https://doi.org/10.1017/S0142716400009796

MacKay, Ian RA, James Emil Flege, Thorsten Piske &Carlo Schirru (2001). Category restructuring during second-language speech acquisition. *The Journal of the Acoustical Society of America* 110 (1): 516-528.

Magloire, Joel G. (1999). A Cross-language Comparison of Speaking Rate Effects on the Production of Voice Onset Time in English and Spanish. *Phonetica* 56 (3–4): 158.

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

Major, Roy C. (2010). First Language Attrition in Foreign Accent Perception. *International Journal of Bilingualism* 14 (2): 163–183. https://doi.org/10.1177/1367006910363063

Miller, Joanne L., Kerry P. Green & Adam Reeves (1986). Speaking Rate and Segments: A Look at the Relation between Speech Production and Speech Perception for the Voicing Contrast. *Phonetica*: 43 (1–3): 106–115. https://doi.org/10.1159/000261764

Martínez Celdrán, Eugenio (1991). Sobre la Naturaleza Fonética de los Alófonos de" b, d, g" en Español y sus Distintas Denominaciones. *Verba: Anuario Galego de*

Filoloxia (18): 235-253.

Machuca, María J. (1997). *Las Obstruyentes No Continuas del Español: Relación entre las Categorías Fonéticas y Fonológicas en Habla Espontánea*. (Unpublished doctoral dissertation). Departamento de Filología Española. Facultad de Filosofía y Letras. Universidad Autónoma de Barcelona.

Miyawaki, Kuniko, James J. Jenkins, Winifred Strange, Alvin M. Liberman, Robert Verbrugge & Osamu Fujimura (1975). An Effect of Linguistic Experience: The Discrimination of (r) and (l) by Native Speakers of Japanese and English. *Perception & Psychophysics* 18 (5): 331–340. https://doi.org/10.3758/BF03211209

Mora, Joan Carles & Marianna Nadeu (2012). L2 Effects on the Perception and Production of a Native Vowel Contrast in Early Bilinguals. *International Journal of Bilingualism* 16 (4): 484–500. https://doi.org/10.1177/1367006911429518

Moyer, Alene (1999). Ultimate Attainment in L2 Phonology: The Critical Factors of Age, Motivation, and Instruction. *Studies in Second Language Acquisition* 21: 81–108.

Murray J. Munro (1993). Productions of English Vowels by Native Speakers of Arabic: Acoustic Measurements and Accentedness Ratings. *Language and Speech* 36 (1): 39.

Newman, Rochelle S. (2003). Using Links between Speech Perception and Speech Production to Evaluate Different Acoustic Metrics: A Preliminary Report. *The Journal of the Acoustical Society of America* 113 (5): 2850–2860. https://doi.org/10.1121/1.1567280

Oyama, Susan (1976). A Sensitive Period for the Acquisition of a Nonnative Phonological System. *Journal of Psycholinguistic Research* 5 (3): 261–283. https://doi.org/10.1007/BF01067377

Oyama, Susan (1978). The Sensitive Period and Comprehension of Speech. *NABE*

Journal, 3 (1): 25–40. https://doi.org/10.1080/08855072.1978.10668342

Parikh, Gaurang (2005). The Influence of Noise on Vowel and Consonant Cues. *Journal of the Acoustical Society of America* 118 (6): 3874.

Patkowsky, Mark S. (1990). Age and Accent in a Second Language: A Reply to James Emil Flege. *Applied Linguistics* 11 (1): 73–89. https://doi.org/10.1093/applin/11.1.73

Peperkamp, Sharon, & Camilla Bouchon (2011). The Relation between Perception and Production in L2 Phonological Processing. *Twelfth Annual Conference of the International Speech Communication Association*.

Perkell, Joseph S., Frank H. Guenther, Harlan Lane, Melanie L. Matthies, Ellen Stockmann, Mark Tiede & Majid Zandipour (2004). The Distinctness of Speakers' Productions of Vowel Contrasts is Related to their Discrimination of the Contrasts. *The Journal of the Acoustical Society of America* 116 (4): 2338–2344. https://doi.org/10.1121/1.1787524

Piper, Terry & Dilek Cansin (1988). Factors Influencing the Foreign Accent. *Canadian Modern Language Review* 44 (2): 334–342.

Piske, Thorsten, Ian R. A. MacKay & James E. Flege (2001). Factors Affecting Degree of Foreign Accent in an L2: A Review. *Journal of Phonetics* 29 (2): 191–215. https://doi.org/10.1006/jpho.2001.0134

Pisoni, David, Laura Manous & Michael Dedina (1987). Comprehension of Natural and Synthetic Speech: Effects of Predictability on the Verification of Sentences Controlled for Intelligibility. *Computer Speech & Language*, 2: 303–320. https://doi.org/10.1016/0885-2308(87)90014-3

Poch, María Dolores (1984). *Las Oclusivas Sordas del Español. Estudio Acústico Espectrográfico de las Realizaciones Intervocálicas en Contacto con Vocal*

Acentuada. (Doctoral dissertation). Departamento de Filología Hispánica, Universitat Autònoma de Barcelona.

Poch, María Dolores (1985). Caractérisation Acoustique des Occlusives de l'Espagnol: le Problème du VOT. *Revue de Phonétique Appliquée* 77: 477–490.

Purcell, Edward T. & Richard W. Suter (1980). Predictors of Pronunciation Accuracy: A Reexamination. *Language Learning* 30 (2): 271–287.

Rallo Fabra, Lucrecia & Joaquín Romero (2012). Native Catalan Learners' Perception and Production of English Vowels. *Journal of Phonetics* 40 (3): 491–508. https://doi.org/10.1016/j.wocn.2012.01.001

Rangel, Heriberto, Asbleidy Torres & Brayan Mojica (2016). Medidas de VOT para Oclusivas Españolas en Estudiantes Universitarios: Nuevas Aproximaciones de Análisis. *Revista Científica Signos Fónicos* 2: 95–106. https://doi.org/10.24054/01204211.v1.n1.2016.1843

Rosner, Burton S., Luis E. López-Bascuas, José E. García-Albea & Richard P. Fahey (2000). Voice-Onset Times for Castilian Spanish Initial Stops. *Journal of Phonetics* 28 (2): 217–224. https://doi.org/10.1006/jpho.2000.0113

Sancier, Michele L. & Carol A. Fowler (1997). Gestural Drift in a Bilingual Speaker of Brazilian Portuguese and English. *Journal of Phonetics*, *25*(4), 421–436. https://doi.org/10.1006/jpho.1997.0051

Schertz, Jessamyn, Taehong Cho, Andrew Lotto & Natasha 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.

Scobbie, James M. (2006). Flexibility in the Face of Incompatible English VOT Systems. *Laboratory Phonology 8: Varieties of Phonological Competence* 675: 367-392

Scovel, Thomas (1988). *A Time to Speak: A Psycholinguistic Onquiry into the Critical*

Period for Human Speech. Newbury House Publishers.

Seliger, Herbert W. (1975). Maturational Constraints in the Acquisition of Second Language Accent. *Language Sciences* 36: 20–22.

Sheldon, Amy & Strange, Winifred (1982). The Acquisition of /r/ and /l/ by Japanese Learners of English: Evidence that Speech Production can Precede Speech Perception. *Applied Psycholinguistics* 3 (3): 243–261. https://doi.org/10.1017/S0142716400001417

Shultz, Amanda A., Alexander L. Francis & Fernando Llanos (2012). Differential Cue Weighting in Perception and Production of Consonant Voicing. *The Journal of the Acoustical Society of America* 132 (2), EL95–EL101. https://doi.org/10.1121/1.4736711

Suter, Richard W. (1976). Predictors of Pronunciation Accuracy in Second Language Learning. *Language Learning* 26 (2): 233–253.

Swartz, Bradford L. (1992). Gender Difference in Voice Onset Time. *Perceptual and Motor Skills* 75 (3): 983–992.

Tahta, Sonia, Margaret Wood & Kate Loewenthal (1981). Foreign Accents: Factors Relating to Transfer of Accent from the First Language to a Second Language. *Language and Speech* 24 (3): 265–272.

Theodore, Rachel M., Joanne L. Miller & David DeSteno (2009). Individual Talker Differences in Voice-Onset-Time: Contextual Influences. *The Journal of the Acoustical Society of America* 125 (6): 3974–3982. https://doi.org/10.1121/1.3106131

Thompson, Irene (1991). Foreign Accents Revisited: The English Pronunciation of Russian Immigrants. *Language Learning* 41 (2): 177–204.

Trofimovich, Pavel & Paul John (2011). When Three Equals Tree. *Applying Priming*

Methods to L2 Learning, Teaching and Research: Insights from Psycholinguistics: 105–129.

Trubetskoi Nikolai S. (1969). *Principles of phonology.* Translated by Christiane A. M. Baltaxe. University of California Press.

Villamizar, Tania (2002). Caracterización Acústica de las Consonantes Oclusivas en el Español de Venezuela. El Parámetro Duración (Silencio-Barra de Explosión-VOT). *Lengua y Habla* 7 (0): 140–157.

Volaitis, Lydia E. & Joane L. Miller. (1992). Phonetic Prototypes: Influence of Place of Articulation and Speaking Rate on the Internal Structure of Voicing Categories. *The Journal of the Acoustical Society of America* 92 (2): 723–735. https://doi.org/10.1121/1.403997

Whalen, Douglas H., Arthur. S. Abramson, Leigh Lisker, & Maria Mody (1993). F0 Gives Voicing Information Even with Unambiguous Voice Onset Times. *Journal of the Acoustical Society of America* 93 (4): 2152-2159.

Whiteside, Sandra P. & Caroline J. Irving (1997). Speakers' Sex Differences in Voice Onset Time: Some Preliminary Findings. *Perceptual and Motor Skills* 85 (2): 459-463.

Williams, Lee (1977). The Perception of Stop Consonant Voicing by Spanish-English Bilinguals. *Perception & Psychophysics* 21 (4): 289–297. https://doi.org/10.3758/BF03199477

Wode, Hennig (1978). The Beginnings of Non-School Room L2 Phonological Acquisition: A Survey of Problems and Issues Based on Data from English as L2 with German as L1. *International Review of Applied Linguistics in Language Teaching* 16, 109.

Zampini, Mary L. (1998). The Relationship between the Production and Perception of L2

Spanish Stops. *Texas Papers in Foreign Language Education* 3 (3): 85.

Appendix A. Detailed background information regarding the participants in the perception and production experiments.

Group label	Group description	*N*	Gender	Age	Age range	L1	L2	Location	Lang. setting	L2 exp.	Range of L2 exp.	Recency of L2 stay	Range of recency of L2 stay	Lang. dom.	Range of dom.
ENCONT	EN control group	9	4 F; 5 M	25.2 (5.4)	18–31	EN	NA	London (UK)	L1	NA	NA	NA	NA	NA	NA
ENINEXPinUK	EN inexperienced speakers in the UK	11	6 F; 5M	20.2 (2.8)	18–28	EN	SP	London (UK)	L1	0	NA	NA	NA	108.7 (27.3)	81.3 – 146.3
ENEXPinUK	EN L2 moderately experienced speakers in the UK	10	4 F; 6 M	24.2 (4.8)	19–32	EN	SP	London (UK)	L1	9.4 (4.2)	4-18	6.8 (4.4)	3-14	106 (21.7)	74.3 – 129.9
ENEXPinSP	EN L2 highly experienced speakers in Spain	11	6 F; 5 M	33.2 (6.1)	23–45	EN	SP	Barcelona (Spain)	L2	48 (27.3)	12-84	current	NA	116.5 (26)	79.9 – 165.5
SPEXPinUK	SP highly experienced speakers in the UK	9	5 F; 4 M	25.0 (3.2)	22–31	SP	EN	London (UK)	L2	47.9 (23.8)	16-96	current	NA	-41.5 (29.8)	-92.6 – 0.6
SPEXPinSP	SP L2 moderately experienced speakers in Spain	10	7 F; 3 M	22.9 (4.6)	21–36	SP	EN	Barcelona (Spain)	L1	6.2 (2.5)	3-12	9.5 (3)	5-14	-75.2 (25.5)	-110.6 – -41
SPINEXPinSP	SP L2 inexperienced speakers in Spain	10	6 F; 4 M	21.1 (3.1)	19–29	SP	EN	Barcelona (Spain)	L1	0	NA	NA	NA	-81.9 (26.4)	-120 – -23.1
SPMONO	SP control group	10	4 F; 6 M	22.7 (4.7)	18–31	SP	NA	Barcelona (Spain)	L1	NA	NA	NA	NA	NA	NA

Table AA.1. Detailed information about the learning groups, including group description, number of participants (*N*), gender, age, L1, L2, location, language setting (lang. setting), L2 experience (exp.), recency of L2 stay and language dominance (lang. dom., based on the BLP dominance score, which ranges from -218, Spanish dominant, to 218, English dominant, being 0 perfect bilingual balance). Standard deviations are provided in parentheses.

Appendix B. Information sheet and consent form. Spanish and English versions

Appendix B1. English version of the information sheet and consent form

Participant's information

Participant's code (to be completed by the experimenter):_____________

Order of completion (to be completed by the experimenter): ___________

Full name: ____

Date of birth: _________________________

Place of birth: _________________________

Place of residence: _____________________

First language(s): ____

Other languages: ____

How long have you lived in an English-speaking country? _______________

How long have you lived in a Spanish-speaking country? _______________

Information sheet

The present experiment is part of the data collection for a PhD thesis. It consists of two main tasks, a production task and an identification task. The production task involves the reading of 30 sentences in English and/or Spanish. The perception task is an identification test in which participants will be asked to select the sound they believe they have been exposed to. The experiment should last between 20 to 40 minutes, depending on group assignment as well as on the participant's own pace to complete each task.

All the information regarding the participants, including name, background and performance in the tests is anonymous, and will only be used for scientific and statistical purposes. The participant may stop the recording and may decide to withdraw from the study at any time.

Consent form

I ________________________________ (full name) understand what the present study involves and agree to participate.

Date: Signature:

B2. Spanish version of the information sheet and consent form

<u>Información del participante</u>

Código del participante:_____________

Orden: ___________

Nombre completo:____

Fecha de nacimiento: ________________________

Lugar de nacimiento: ________________________

Lugar de residencia: ____________________

Primera lengua(s):____

Otras lenguas: ____

¿Has vivido alguna vez en un país angloparlante? _______________

¿Cuánto tiempo has vivido en un país donde se hable español?__________

Información

Este experimento es parte de la recogida de datos para una tesis doctoral. Consiste en dos tareas principales: un ejercicio de producción y otro de identificación. El ejercicio de producción incluye la lectura de 32 frases cortas. La tarea de percepción es un test de identificación en el que los participantes deben responder qué sonido creen que han escuchado. El experimento dura entre dura entra 15 y 20 minutos, dependiendo del ritmo de cada participante.

Toda la información relativa a los participantes y el resultado de su test son anónimos y serán usados únicamente con fines científicos. El participante puede retirarse del estudio en cualquier momento.

Yo _________________________________ (nombre completo) entiendo en qué consiste el presente experimento y estoy de acuerdo en participar.

Data: Firma:

Appendix C. Bilingual Language Profile questionnaire (English version)

Appendix C1. Original BLP questions

Bilingual Language Profile: English-Spanish

We would like to ask you to help us by answering the following questions concerning your language history, use, attitudes, and proficiency. This survey was created with support from the Center for Open Educational Resources and Language Learning at the University of Texas at Austin to better understand the profiles of bilingual speakers in diverse settings with diverse backgrounds. The survey consists of 19 questions and will take less than 10 minutes to complete. This is not a test, so there are no right or wrong answers. Please answer every question and give your answers sincerely. Thank you very much for your help.

I. Biographical Information

Name ________________________________	Today's Date ____/____/______

Age______	☐ Male / ☐ Female	Current place of residence: city/state________________ country__________

Highest level of formal education:	☐ Less than high school	☐ High school	☐ Some college
	☐ College (B.A., B.S.)	☐ Some graduate school	☐ Masters
	☐ PhD/MD/JD	☐ Other: ____________	

II. Language history

In this section, we would like you to answer some factual questions about your language history by placing a check in the appropriate box.

1. At what age did you **start learning** the following languages?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Since birth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Since birth	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

2. At what age did you **start to feel comfortable** using the following languages?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
As early as I can remember	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	not yet

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
As early as I can remember	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+	not yet

3. How many years of **classes (grammar, history, math, etc.)** have you had in the following languages (primary school through university)?

English

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

Spanish

☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+

4. How many years have you spent in a **country/region** where the following languages are spoken?

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
English	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

5. How many years have you spent in a **family** where the following languages are spoken?

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
English	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

6. How many years have you spent in a **work environment** where the following languages are spoken?

	0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20+
English	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

III. Language use

In this section, we would like you to answer some questions about your language use by placing a check in the appropriate box. Total use for all languages in a given question should equal 100%.

7. In an average week, what percentage of the time do you use the following languages **with friends**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
English	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other languages	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

8. In an average week, what percentage of the time do you use the following languages **with family**?

	0%	10%	20%	30%	40%	50%	60%	70%	80%	90%	100%
English	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Spanish	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐
Other languages	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐	☐

9. In an average week, what percentage of the time do you use the following languages **at school/work**?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

10. When you talk to yourself, how often do you **talk to yourself** in the following languages?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

11. When you count, how often do you **count** in the following languages?

English	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Spanish	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%
Other languages	☐ 0%	☐ 10%	☐ 20%	☐ 30%	☐ 40%	☐ 50%	☐ 60%	☐ 70%	☐ 80%	☐ 90%	☐ 100%

IV. Language proficiency

In this section, we would like you to rate your language proficiency by giving marks from 0 to 6.

	0=not well at all						*6=very well*
12. a. How well do you speak **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you speak **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
13. a. How well do you understand **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you understand **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
14. a. How well do you read **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you read **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
15. a. How well do you write **English**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. How well do you write **Spanish**?	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

V. Language attitudes

In this section, we would like you to respond to statements about language attitudes by giving marks from 0-6.

	0=disagree						*6=agree*
16. a. I feel like myself when I speak **English**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I feel like myself when I speak **Spanish**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
17. a. I identify with an **English-speaking** culture.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I identify with a **Spanish-speaking** culture.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
18. a. It is important to me to use (or eventually use) **English** like a native speaker.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. It is important to me to use (or eventually use) **Spanish** like a native speaker.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
19. a. I want others to think I am a native speaker of **English**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6
b. I want others to think I am a native speaker of **Spanish**.	☐ 0	☐ 1	☐ 2	☐ 3	☐ 4	☐ 5	☐ 6

Appendix C2. Additional questions to the BLP questionnaire

When did you last live in an English-speaking country?

When did you last live in a Spanish-speaking country?

Appendix D. Production elicitation lists.

Appendix D1. English production task

Read each of the following sentences twice:

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

29. Ancient is the next word
30. Peacock is the next word
31. Gearshift is the next word
32. Keyhole is the next word

Appendix D2. Spanish production task

Lee cada una de las siguientes frases dos veces:

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

31. Guita es la siguiente palabra

32. Pino es la siguiente palabra

Appendix E. Perceptual boundaries and mean VOT values for all participants in all groups.

Appendix E1. L1 perception and production results

L1-English participants' perception and production results in English

Participant	Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
ENCONT1	ENCONT	11.4	53	14	26.6	75	31
ENCONT2	ENCONT	9.1	34	7	35.0	57	20
ENCONT3	ENCONT	15.8	44	4	31.1	72	13
ENCONT4	ENCONT	15.8	44	10	26.3	53	13
ENCONT5	ENCONT	12.0	40	11	26.6	59	24
ENCONT6	ENCONT	15.5	77	7	22.6	78	24
ENCONT7	ENCONT	15.8	47	5	33.1	67	-9
ENCONT8	ENCONT	15.5	65	-35	32.7	104	-15
ENCONT10	ENCONT	15.5	65	-14	33.6	83	22
ENINEXPinUK4	ENINEXPinUK	13.2	54	10.	22.7	70	27
ENINEXPinUK5	ENINEXPinUK	13.2	53	7	32.6	73	20
ENINEXPinUK6	ENINEXPinUK	11.7	43	8	38.1	67	22
ENINEXPinUK7	ENINEXPinUK	11.7	68	6	30.8	89	22
ENINEXPinUK8	ENINEXPinUK	15.8	59	8	26.6	86	23
ENINEXPinUK9	ENINEXPinUK	17.4	63	8	32.0	80	21
ENINEXPinUK10	ENINEXPinUK	15.2	48	0	20.2	78	-19
ENINEXPinUK11	ENINEXPinUK	15.2	36	9	35.7	61	25
ENINEXPinUK12	ENINEXPinUK	19.4	30	12	27.0	89	33
ENINEXPinUK13	ENINEXPinUK	9.1	62	6	7.4	86	23
ENINEXPinUK14	ENINEXPinUK	17.7	53	13	27.3	93	26
ENEXPinUK2	ENEXPinUK	14.5	40	8	25.6	55	23
ENEXPinUK3	ENEXPinUK	11.4	55	-10	22.0	75	-20
ENEXPinUK4	ENEXPinUK	13.9	54	-2	24.8	82	17
ENEXPinUK6	ENEXPinUK	14.0	61	10	25.0	86	23
ENEXPinUK7	ENEXPinUK	13.9	49	-20	24.8	78	-58
ENEXPinUK8	ENEXPinUK	25.5	48	-29	38.1	87	22
ENEXPinUK10	ENEXPinUK	17.4	37	-23	28.8	64	-42
ENEXPinUK13	ENEXPinUK	17.7	74	-21	29.2	94	24
ENEXPinUK14	ENEXPinUK	13.5	36	7	24.5	61	17
ENEXinSP01	ENEXPinSP	17.4	36	6	28.2	72	24
ENEXPinSP02	ENEXPinSP	15.8	47	-1	25.1	77	14
ENEXPinSP03	ENEXPinSP	12.0	44	3	24.5	58	19
ENEXPinSP04	ENEXPinSP	13.2	47	6	29.6	72	9

Participant	Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
ENEXPinSP05	ENEXPinSP	17.7	74	39	28.5	94	-12
ENEXPinSP06	ENEXPinSP	13.6	41	-15	28.8	62	-6
ENEXPinSP07	ENEXPinSP	19.7	51	-33	28.9	67	-5
ENEXPinSP08	ENEXPinSP	17.7	33	7	38.2	84	30
ENEXPinSP09	ENEXPinSP	17.4	38	-37	28.8	60	-46
ENEXPinSP10	ENEXPinSP	15.2	37	18	31.4	62	20
ENEXPinSP11	ENEXPinSP	17.4	58	6	31.9	88	17

Table AE.1. L1-English participants' individual English category boundaries and VOT means in ms. Standard deviations are provided in parentheses

L1-Spanish participants' perception and production results in Spanish

Participant	Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
SSCONT1	SPCONT	5.3	8	-60	23.8	34	18
SPCONT2	SPCONT	-7.9	7	-106	0.0	39	15
SPCONT3	SPCONT	9.4	7	-78	20.2	26	33
SPCONT4	SPCONT	2.5	7	-89	24.5	29	16
SPCONT5	SPCONT	6.6	7	-95	24.8	29	28
SPCONT6	SPCONT	9.4	9	-68	11.6	33	19
SPCONT7	SPCONT	7.7	6	-90	7.7	27	37
SPCONT9	SPCONT	-0.5	11	-39	2.5	38	36
SPCONT10	SPCONT	2.5	10	-76	18.2	41	12
SPCONT11	SPCONT	-0.6	16	-75	2.1	45	35
SPINEXPinSP01	SPINEXPinSP	5.3	7	-52	22.0	22	-53
SPINEXPinSP02	SPINEXPinSP	6.6	3	-71	22.0	19	-80
SPINEXPinSP04	SPINEXPinSP	11.8	8	-41	24.8	34	-58
SPINEXPinSP06	SPINEXPinSP	2.5	9	-40	17.0	30	-32
SPINEXPinSP07	SPINEXPinSP	5.3	10	-47	12.4	38	-62
SPINEXPinSP08	SPINEXPinSP	11.3	7	-79	24.1	32	-67
SPINEXPinSP09	SPINEXPinSP	9.0	11	-53	17.7	36	-60
SPINEXPinSP10	SPINEXPinSP	9.7	9	-79	20.2	45	-86
SPINEXPinSP11	SPINEXPinSP	12.0	11	-76		38	-85
SPINEXPinSP13	SPINEXPinSP	10.2	6	-104	24.8	24	-93
SPEXPinSP01	SPEXPinSP	7.7	7	-83	20.2	30	-109
SPEXPinSP02	SPEXPinSP	0.4	7	-47	21.9	27	-37
SPEXPinSP04	SPEXPinSP	10.0	10	-79	20.2	35	-65
SPEXPinSP05	SPEXPinSP	11.7	8	-88	21.6	29	-100
SPEXPinSP06	SPEXPinSP	7.7	5	-102	24.5	42	-103
SPEXPinSP07	SPEXPinSP	7.4	7	-79	18.4	40	-49
SPEXPinSP08	SPEXPinSP	10.0	11	-72	17.8	34	-73
SPEXPinSP09	SPEXPinSP	8.0	20	-57	24.1	53	-73
SPEXPinSP10	SPEXPinSP	10.0	13	-95	20.2	33	-71
SPEXPinSP11	SPEXPinSP	4.9	11	-102	19.6	48	-109
SPEXPinUK1	SPEXPinUK	9.7	7	-52	21.8	40	-88
SPEXPinUK2	SPEXPinUK	14.2	7	-70	26.3	26	-73
SPEXPinUK5	SPEXPinUK	9.6	14	-46	21.5	44	-28
SPEXPinUK6	SPEXPinUK	3.9	15	-58	14.0	62	-84
SPEXPinUK7	SPEXPinUK	9.3	10	-81	22.3	64	-67
SPEXPinUK8	SPEXPinUK	13.5	12	-72	23.8	38	-92
SPEXPinUK9	SPEXPinUK	13.9	10	-78	24.1	48	-93
SPEXPinUK10	SPEXPinUK	11.4	9	-90	20.2	43	-48
SPEXPinUK11	SPEXPinUK	15.2	8	-68	18.0	32	-77

Table AE.2. L1-Spanish participants' individual Spanish category boundaries and VOT means in ms. Standard deviations are provided in parentheses

Appendix E2. L2 perception and production results

L1-English participants' perception and production results in Spanish

Participant	Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
ENEXPinSP01	ENEXPinSP	11.4	9	-68	22.7	42	-44
ENEXPinSP02	ENEXPinSP	15.8	16	-77	28.8	48	-66
ENEXPinSP03	ENEXPinSP	9.7	14	3	30.7	38	-11
ENEXPinSP04	ENEXPinSP	15.8	13	6	27.6	54	27
ENEXPinSP05	ENEXPinSP	17.4	23	-51	26.6	62	-43
ENEXPinSP06	ENEXPinSP	13.5	9	-42	28.8	46	-67
ENEXPinSP07	ENEXPinSP	13.8	12	-40	26.2	37	-33
ENEXPinSP08	ENEXPinSP	19.1	12	6	27.0	45	17
ENEXPinSP09	ENEXPinSP	19.1	20	-47	28.8	13	-73
ENEXPinSP11	ENEXPinSP	13.5	9	-29	33.1	35	-31
ENEXPinUK2	ENEXPinUK	9.8	26	-67	23.8	57	-47
ENEXPinUK3	ENEXPinUK	17.7	8	-88	0.0	37	-65
ENEXPinUK4	ENEXPinUK	7.7	25	-8	20.2	49	-18
ENEXPinUK6	ENEXPinUK	17.1	18	-83	24.5	52	-126
ENEXPinUK7	ENEXPinUK	21.2	14	-14	24.8	45	-57
ENEXPinUK8	ENEXPinUK	22.8	24	2	11.6	51	-9
ENEXPinUK10	ENEXPinUK	15.5	11	-68	7.7	45	-69
ENEXPinUK13	ENEXPinUK	19.7	11	-87	2.5	34	-48
ENEXPinUK14	ENEXPinUK	15.8	15	-20	18.2	35	-49
ENINEXPinUK4	ENINEXPinUK	11.4	24	-7	28.9	44	-24
ENINEXPinUK5	ENINEXPinUK	13.6	26	0	32.7	62	13
ENINEXPinUK6	ENINEXPinUK	23.7	10	-63	36.2	27	-11
ENINEXPinUK7	ENINEXPinUK	13.5	13	9	34.2	55	23
ENINEXPinUK8	ENINEXPinUK	9.7	47	1	28.8	75	26
ENINEXPinUK9	ENINEXPinUK	15.8	24	6	24.5	50	-43
ENINEXPinUK10	ENINEXPinUK	13.5	18	6	20.2	53	-7
ENINEXPinUK11	ENINEXPinUK	16.8	16	8	23.7	46	21
ENINEXPinUK12	ENINEXPinUK	19.4	24	10	18.1	55	18
ENINEXPinUK13	ENINEXPinUK	19.7	77	-100	7.7	89	-119
ENINEXPinUK14	ENINEXPinUK	15.5	26	-10	22.1	54	12

Table AE.3. L1-English participants' individual Spanish category boundaries and VOT means in ms. Standard deviations are provided in parentheses

L1-Spanish participants' perception and production results in English

Participant	Group	/p/-/b/ boundary	/p/ mean VOT	/b/ mean VOT	/k/-/g/ boundary	/k/ mean VOT	/g/ mean VOT
SPEXPinUK1	SPEXPinUK	13.9	10	-93	24.3	38	-63
SPEXPinUK2	SPEXPinUK	15.1	61	-16	28.2	84	-25
SPEXPinUK5	SPEXPinUK	9.3	14	-15	26.6	65	18
SPEXPinUK6	SPEXPinUK	9.0	35	-34	22.0	80	-30
SPEXPinUK7	SPEXPinUK	10.3	52	-84	26.3	69	-23
SPEXPinUK8	SPEXPinUK	11.3	30	-60	26.3	76	-70
SPEXPinUK9	SPEXPinUK	11.6	42	1	22.3	100	15
SPEXPinUK10	SPEXPinUK	11.7	36	-45	22.3	67	-10
SPEXPinUK12	SPEXPinUK	16.8	38	-43	22.0	78	-25
SPEXPinSP01	SPEXPinSP	15.2	31	-36	33.3	54	-45
SPEXPinSP02	SPEXPinSP	15.8	38	-50	28.8	72	-37
SPEXPinSP04	SPEXPinSP	9.7	26	-92	22.4	72	-57
SPEXPinSP05	SPEXPinSP	9.0	28	-96	22.7	63	-35
SPEXPinSP06	SPEXPinSP	7.7	11	-87	24.8	72	-79
SPEXPinSP07	SPEXPinSP	8.0	17	-72	22.6	64	-29
SPEXPinSP08	SPEXPinSP	15.8	13	-50	26.3	61	-21
SPEXPinSP09	SPEXPinSP	8.0	38	-37	25.8	60	-46
SPEXPinSP10	SPEXPinSP	8.0	46	-60	20.2	79	-46
SPEXPinSP11	SPEXPinSP	3.4	70	-126	22.0	106	-91
SPINEXPinSP01	SPINEXPinSP	7.7	10	-71	24.5	37	-54
SPINEXPinSP02	SPINEXPinSP	8.0	9	-69	28.2	28	-61
SPINEXPinSP04	SPINEXPinSP	9.8	37	-59	28	68	-44
SPINEXPinSP06	SPINEXPinSP	7.4	40	-74	24.8	82	-54
SPINEXPinSP07	SPINEXPinSP	4.4	30	-83	22.0	79	-50
SPINEXPinSP08	SPINEXPinSP	5.7	12	-53	21.9	32	-33
SPINEXPinSP09	SPINEXPinSP	11.4	25	-59	18.0	55	-36
SPINEXPinSP10	SPINEXPinSP	11.7	47	-14	17.8	69	-53
SPINEXPinSP11	SPINEXPinSP	9.6	7	-48	15.9	40	-21
SPINEXPinSP13	SPINEXPinSP	17.1	10	-132	27.0	46	-90

Table AE.4. L1-Spanish participants' individual English category boundaries and VOT means in ms. Standard deviations are provided in parentheses

Appendix F. Individual factors and mean category boundaries for all participants

Participant	Group	L1	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/-/b/ boundary	English /k/-/g/ boundary	Spanish /p/-/b/ boundary	Spanish /k/-/g/ boundary
ENEXPinSP01	ENEXPinSP	English	99.0	0	20	37%	60	17.4	28.2	11.4	22.7
ENEXPinSP02	ENEXPinSP	English	ND	ND	ND	ND	60	15.8	25.1	15.8	28.8
ENEXPinSP03	ENEXPinSP	English	165.5	0	20	7%	12	12.0	24.5	9.7	30.7
ENEXPinSP04	ENEXPinSP	English	105.2	6	20	17%	72	13.2	29.6	15.8	27.6
ENEXPinSP05	ENEXPinSP	English	79.9	6	16	13%	36	17.7	28.5	17.4	26.6
ENEXPinSP06	ENEXPinSP	English	125.1	3	20	7%	72	13.6	28.8	13.5	28.8
ENEXPinSP07	ENEXPinSP	English	94.3	7	14	10%	12	19.7	28.9	13.8	26.2
ENEXPinSP08	ENEXPinSP	English	147.9	2	19	10%	36	17.7	38.2	19.1	27.0
ENEXPinSP09	ENEXPinSP	English	104.4	0	20	30%	72	17.4	28.8	19.1	28.8
ENEXPinSP10	ENEXPinSP	English	113.0	0	20	ND	84	15.2	31.4	ND	ND
ENEXPinSP11	ENEXPinSP	English	130.7	0	20	17%	12	17.4	31.9	13.5	33.1
ENEXPinUK02	ENEXPinUK	English		15	20	30%	12	14.5	25.6	9.8	21.9
ENEXPinUK03	ENEXPinUK	English	90.1	18	14	10%	12	11.4	22.0	17.7	18.4
ENEXPinUK04	ENEXPinUK	English	118.1	3	18	13%	9	13.9	24.8	7.7	23.6
ENEXPinUK06	ENEXPinUK	English	74.3	7	13	20%	6	14.0	25.0	17.1	27.0
ENEXPinUK07	ENEXPinUK	English	120.0	4	20	13%	9	13.9	24.8	21.2	25.0
ENEXPinUK08	ENEXPinUK	English	105.3	8	12	10%	6	25.5	38.1	22.8	29.8
ENEXPinUK10	ENEXPinUK	English	128.8	5	15	10%	18	17.4	28.8	15.5	28.8
ENEXPinUK12	ENEXPinUK	English	ND	9	12	13%	4	17.4	28.8	17.4	17.7
ENEXPinUK13	ENEXPinUK	English	81.5	3	20	20%	12	17.7	29.2	19.7	33.1

Participant	Group	L1	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/-/b/ boundary	English /k/-/g/ boundary	Spanish /p/-/b/ boundary	Spanish /k/-/g/ boundary
ENEXPinUK14	ENEXPinUK	English	129.9	3	11	10%	6	13.5	24.5	15.8	30.7
ENINEXPinUK04	ENINEXPinUK	English	ND	1	11	10%	0	13.2	22.7	11.4	28.9
ENINEXPinUK05	ENINEXPinUK	English	134.2	1	20	7%	0	13.2	32.6	13.6	32.7
ENINEXPinUK06	ENINEXPinUK	English	ND	1	20	3%	0	11.7	38.1	23.7	36.2
ENINEXPinUK07	ENINEXPinUK	English	81.3	8	13	7%	0	11.7	30.8	13.5	34.2
ENINEXPinUK08	ENINEXPinUK	English	99.2	1	15	20%	0	15.8	26.6	9.7	28.8
ENINEXPinUK09	ENINEXPinUK	English	ND	1	11	10%	0	17.4	32.0	15.8	24.5
ENINEXPinUK10	ENINEXPinUK	English	146.3	8	10	3%	0	15.2	20.2	13.5	20.2
ENINEXPinUK11	ENINEXPinUK	English	129.4	3	12	17%	0	15.2	35.7	16.8	23.7
ENINEXPinUK12	ENINEXPinUK	English	88.4	7	11	27%	0	19.4	27.0	19.4	18.1
ENINEXPinUK13	ENINEXPinUK	English	ND	ND	ND	ND	0	9.1	7.4	15.5	7.7
ENINEXPinUK14	ENINEXPinUK	English	82.0	7	11	20%	0	17.7	27.3	19.7	22.1
SPEXPinUK01	SPEXPinUK	Spanish	-44.9	13	8	53%	25	13.9	24.3	9.7	21.8
SPEXPinUK02	SPEXPinUK	Spanish	-37.6	12	3	57%	16	15.1	22.3	14.2	26.3
SPEXPinUK05	SPEXPinUK	Spanish	-44.9	14	7	47%	38	9.3	22.0	9.6	21.5
SPEXPinUK06	SPEXPinUK	Spanish	-26.2	12	5	40%	38	9.0	28.2	3.9	14.0
SPEXPinUK07	SPEXPinUK	Spanish	0.6	10	6	93%	96	10.3	26.6	9.3	22.3
SPEXPinUK08	SPEXPinUK	Spanish	-15.1	17	6	87%	50	11.3	22.0	13.5	23.8
SPEXPinUK09	SPEXPinUK	Spanish	ND	ND	ND	ND	60	11.6	26.3	13.9	24.1
SPEXPinUK10	SPEXPinUK	Spanish	-70.9	13	7	20%	48	11.7	26.3	11.4	20.2
SPEXPinUK12	SPEXPinUK	Spanish	-92.6	16	4	40%	60	16.8	22.3	15.2	18.0
SPEXPinSP01	SPEXPinSP	Spanish	-95.1	13	6	27%	7	15.2	33.3	7.7	20.2
SPEXPinSP02	SPEXPinSP	Spanish	-109.3	13	6	10%	3	15.8	28.8	0.4	21.9
SPEXPinSP04	SPEXPinSP	Spanish	-48.1	13	5	40%	6	9.7	22.4	10.0	20.2
SPEXPinSP05	SPEXPinSP	Spanish	-110.6	13	8	17%	5	9.0	22.7	11.7	21.6

Participant	Group	L1	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/-/b/ boundary	English /k/-/g/ boundary	Spanish /p/-/b/ boundary	Spanish /k/-/g/ boundary
SPEXPinSP06	SPEXPinSP	Spanish	ND	13	8	13%	6	7.7	24.8	7.7	24.5
SPEXPinSP07	SPEXPinSP	Spanish	-74.9	13	8	27%	6	8.0	22.6	7.4	18.4
SPEXPinSP08	SPEXPinSP	Spanish	-54.7	13	6	37%	8	15.8	26.3	10.0	17.8
SPEXPinSP09	SPEXPinSP	Spanish	-76.1	18	3	30%	5	8.0	25.8	8.0	24.1
SPEXPinSP10	SPEXPinSP	Spanish	-41.0	17	5	30%	4	8.0	20.2	10.0	20.2
SPEXPinSP11	SPEXPinSP	Spanish	-67.4	13	6	33%	12	3.4	22.0	4.9	19.6
SPINEXPinSP01	SPINEXPinSP	Spanish	-81.4	14	6	10%	0	7.7	24.5	5.3	22.0
SPINEXPinSP02	SPINEXPinSP	Spanish	-110.2	11	6	20%	0	8.0	28.2	6.6	22.0
SPINEXPinSP04	SPINEXPinSP	Spanish	-70.5	12	5	27%	0	9.8	28.1	11.8	24.8
SPINEXPinSP06	SPINEXPinSP	Spanish	-80.9	11	4	20%	0	7.4	24.8	2.5	17.0
SPINEXPinSP07	SPINEXPinSP	Spanish	-84.7	13	6	43%	0	4.4	22.0	5.3	12.4
SPINEXPinSP08	SPINEXPinSP	Spanish	-97.1	16	5	10%	0	5.7	21.9	11.3	24.1
SPINEXPinSP09	SPINEXPinSP	Spanish	-82.6	11	10	23%	0	11.4	18.0	9.0	17.7
SPINEXPinSP10	SPINEXPinSP	Spanish	-68.7	15	4	20%	0	11.7	17.8	9.7	20.2
SPINEXPinSP11	SPINEXPinSP	Spanish	-120.0	17	5	23%	0	9.6	15.9	12.0	ND
SPINEXPinSP13	SPINEXPinSP	Spanish	-23.1	12	3	30%	0	17.1	27.0	10.2	24.8

Table AF.1. All participants' language dominance score, years of L2 instruction, L2 AOL, % weekly L2 use and category boundaries.

Appendix G. Individual factors and mean VOT productions for all participants

Participant	Group	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/	English /b/	English /k/	English /g/	Spanish /p/	Spanish /b/	Spanish /k/	Spanish /g/
ENEXPinSP01	ENEXPinSP	99.0	0	20	37%	60	36.4	6.1	72.4	23.9	9.0	-68.0	42.0	-43.9
ENEXPinSP02	ENEXPinSP	ND	ND	ND	ND	60	46.7	-0.7	76.6	14.4	15.8	-77.1	48.0	-66.2
ENEXPinSP03	ENEXPinSP	165.5	0	20	7%	12	44.4	3.1	58.1	18.9	14.1	2.6	38.2	-10.5
ENEXPinSP04	ENEXPinSP	105.2	6	20	17%	72	46.9	5.6	71.5	8.6	13.3	5.7	54.2	27.0
ENEXPinSP05	ENEXPinSP	79.9	6	16	13%	36	73.7	38.6	93.5	-11.8	22.8	-51.2	61.9	-42.7
ENEXPinSP06	ENEXPinSP	125.1	3	20	7%	72	40.8	-15.1	61.7	-6.0	9.3	-41.6	46.1	-67.2
ENEXPinSP07	ENEXPinSP	94.3	7	14	10%	12	50.5	-32.9	67.2	-4.6	11.6	-40.1	37.3	-33.1
ENEXPinSP08	ENEXPinSP	147.9	2	19	10%	36	33.0	7.3	84.2	29.7	11.9	5.7	45.4	17.1
ENEXPinSP09	ENEXPinSP	104.4	0	20	30%	72	37.7	-36.6	60.4	-45.5	19.9	-47.1	13.4	-73.3
ENEXPinSP10	ENEXPinSP	113.0	0	20	ND	84	37.0	17.6	61.9	19.8				
ENEXPinSP11	ENEXPinSP	130.7	0	20	17%	12	57.7	5.8	87.9	16.5	9.4	-29.2	34.7	-31.5
ENEXPinUK02	ENEXPinUK	ND	15	20	30%	12	40.3	8.3	54.8	22.6	26.1	-67.2	56.6	-46.9
ENEXPinUK03	ENEXPinUK	90.1	18	14	10%	12	54.8	-10.4	75.1	-19.8	8.2	-88.3	37.0	-65.4
ENEXPinUK04	ENEXPinUK	118.1	3	18	13%	9	54.3	-1.9	82.3	17.4	25.4	-8.4	49.1	-18.3
ENEXPinUK06	ENEXPinUK	74.3	7	13	20%	6	60.5	10.3	86.1	23.0	18.2	-83.4	51.8	-126.3
ENEXPinUK07	ENEXPinUK	120.0	4	20	13%	9	48.9	-20.2	77.5	-57.9	13.9	-14.0	44.8	-56.8
ENEXPinUK08	ENEXPinUK	105.3	8	12	10%	6	48.2	-29.3	86.7	22.3	24.4	-7.7	51.4	-8.9
ENEXPinUK10	ENEXPinUK	128.8	5	15	10%	18	36.9	-22.8	63.8	-42.3	10.7	-68.1	44.5	-69.4
ENEXPinUK12	ENEXPinUK	ND	9	12	13%	4	ND	ND	ND	ND	ND	ND	ND	ND

Participant	Group	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/	English /b/	English /k/	English /g/	Spanish /p/	Spanish /b/	Spanish /k/	Spanish /g/
ENEXPinUK13	ENEXPinUK	81.5	3	20	20%	12	74.4	-21.2	94.4	24.3	10.7	-87.4	33.9	-48.1
ENEXPinUK14	ENEXPinUK	129.9	3	11	10%	6	35.6	6.9	61.2	16.5	14.6	-20.3	35.4	-49.4
ENINEXPinUK04	ENINEXPinUK	ND	1	11	10%	0	54.1	10.1	70.1	26.9	23.6	-6.6	44.2	-24.4
ENINEXPinUK05	ENINEXPinUK	134.2	1	20	7%	0	52.7	7.2	72.9	19.8	25.5	0.4	61.7	12.5
ENINEXPinUK06	ENINEXPinUK	ND	1	20	3%	0	43.2	7.6	66.9	21.6	9.5	-62.7	26.5	-10.8
ENINEXPinUK07	ENINEXPinUK	81.3	8	13	7%	0	67.6	6.0	88.5	22.1	12.9	8.5	55.1	22.7
ENINEXPinUK08	ENINEXPinUK	99.2	1	15	20%	0	59.2	8.0	85.7	23.4	46.7	1.4	75.1	25.5
ENINEXPinUK09	ENINEXPinUK	ND	1	11	10%	0	63.1	7.8	80.4	20.7	24.1	5.7	49.6	-43.2
ENINEXPinUK10	ENINEXPinUK	146.3	8	10	3%	0	48.1	0.4	78.4	-19.2	17.5	5.9	53.3	-6.6
ENINEXPinUK11	ENINEXPinUK	129.4	3	12	17%	0	35.6	9.2	61.0	25.3	16.0	8.0	46.3	20.6
ENINEXPinUK12	ENINEXPinUK	88.4	7	11	27%	0	29.9	12.2	89.3	33.1	24.4	10.4	55.0	18.1
ENINEXPinUK13	ENINEXPinUK					0	62.1	6.4	85.9	23.1	76.7	-100.1	89.4	-118.7
ENINEXPinUK14	ENINEXPinUK	82.0	7	11	20%	0	53.1	12.9	93.1	25.6	26.3	-10.0	53.5	12.0
SPEXPinUK01	SPEXPinUK	-44.9	13	8	53%	25	9.9	-92.9	38.2	-62.7	7.2	-52.3	39.9	-87.5
SPEXPinUK02	SPEXPinUK	-37.6	12	3	57%	16	61.4	-15.7	84.3	-25.4	6.5	-70.2	25.8	-72.5
SPEXPinUK05	SPEXPinUK	-44.9	14	7	47%	38	14.1	-14.5	65.3	18.2	13.5	-46.2	43.9	-27.8
SPEXPinUK06	SPEXPinUK	-26.2	12	5	40%	38	34.8	-33.7	80.3	-29.6	15.2	-57.6	61.9	-84.3
SPEXPinUK07	SPEXPinUK	0.6	10	6	93%	96	52.2	-83.7	69.0	-22.9	10.0	-80.9	63.8	-67.0
SPEXPinUK08	SPEXPinUK	-15.1	17	6	87%	50	29.8	-60.4	76.1	-70.4	11.6	-71.7	37.7	-92.1
SPEXPinUK09	SPEXPinUK	ND	ND	ND	ND	60	41.9	0.7	99.9	15.4	9.7	-78.2	48.0	-92.7
SPEXPinUK10	SPEXPinUK	-70.9	13	7	20%	48	36.0	-44.7	66.5	-10.2	8.6	-90.4	43.2	-95.9

Participant	Group	Dom. score	Years of L2 instr.	L2 AOL	% weekly L2 use	Months in L2 setting	English /p/	English /b/	English /k/	English /g/	Spanish /p/	Spanish /b/	Spanish /k/	Spanish /g/
SPEXPinUK12	SPEXPinUK	-92.6	16	4	40%	60	38.0	-43.4	77.5	-25.4	8.2	-68.1	31.5	-76.6
SPEXPinSP01	SPEXPinSP	-95.1	13	6	27%	7	30.8	-36.4	54.2	-45.0	7.1	-83.3	30.3	-108.5
SPEXPinSP02	SPEXPinSP	-109.3	13	6	10%	3	37.7	-49.7	71.7	-37.0	6.7	-47.1	26.6	-36.5
SPEXPinSP04	SPEXPinSP	-48.1	13	5	40%	6	26.2	-91.7	72.4	-56.5	9.8	-79.4	35.3	-64.7
SPEXPinSP05	SPEXPinSP	-110.6	13	8	17%	5	27.7	-96.1	62.7	-35.3	8.4	-87.5	28.6	-99.8
SPEXPinSP06	SPEXPinSP	ND	13	8	13%	6	11.0	-87.4	71.6	-79.4	5.4	-101.9	42.0	-103.0
SPEXPinSP07	SPEXPinSP	-74.9	13	8	27%	6	17.1	-71.6	63.8	-29.4	7.0	-79.0	40.1	-48.6
SPEXPinSP08	SPEXPinSP	-54.7	13	6	37%	8	13.1	-50.2	61.3	-21.0	10.8	-72.3	33.5	-72.7
SPEXPinSP09	SPEXPinSP	-76.1	18	3	30%	5	37.7	-36.6	60.4	-45.5	19.9	-57.0	52.6	-73.3
SPEXPinSP10	SPEXPinSP	-41.0	17	5	30%	4	46.4	-59.5	79.0	-46.0	12.8	-94.7	33.0	-70.9
SPEXPinSP11	SPEXPinSP	-67.4	13	6	33%	12	70.0	-125.8	106.4	-91.4	11.2	-101.5	48.3	-108.7
SPINEXPinSP01	SPINEXPinSP	-81.4	14	6	10%	0	10.1	-71.0	37.4	-53.5	7.1	-51.7	21.8	-53.4
SPINEXPinSP02	SPINEXPinSP	-110.2	11	6	20%	0	9.1	-68.7	28.2	-60.9	2.8	-70.9	18.5	-80.1
SPINEXPinSP04	SPINEXPinSP	-70.5	12	5	27%	0	37.0	-59.0	67.6	-43.5	8.2	-55.0	34.0	-57.8
SPINEXPinSP06	SPINEXPinSP	-80.9	11	4	20%	0	39.8	-74.0	81.6	-54.3	9.0	-40.2	29.9	-54.0
SPINEXPinSP07	SPINEXPinSP	-84.7	13	6	43%	0	29.6	-82.5	78.5	-50.2	10.0	-59.2	37.5	-71.7
SPINEXPinSP08	SPINEXPinSP	-97.1	16	5	10%	0	12.4	-53.4	31.7	-33.3	6.5	-79.2	31.7	-66.9
SPINEXPinSP09	SPINEXPinSP	-82.6	11	10	23%	0	24.5	-59.4	54.8	-36.2	11.0	-53.3	36.0	-59.8
SPINEXPinSP10	SPINEXPinSP	-68.7	15	4	20%	0	46.8	-14.4	68.5	-52.8	9.2	-79.4	45.2	-86.4
SPINEXPinSP11	SPINEXPinSP	-120.0	17	5	23%	0	6.5	-47.8	40.3	-21.0	10.7	-75.8	37.9	-84.6
SPINEXPinSP13	SPINEXPinSP	-23.1	12	3	30%	0	9.7	-131.5	45.6	-90.2	6.4	-103.6	24.2	-93.1

Table AG.1. All participants' language dominance score, years of L2 instruction, L2 AOL, % weekly L2 use and mean VOT productions.