Universitat Autònoma de Barcelona

L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study

Angélica Foresti Carlet

Universitat Autònoma de Barcelona

L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study

Angélica Foresti Carlet

PhD Dissertation

Supervised by : Dr. Juli Cebrian

Programa de Doctorat en Filologia Anglesa

Departament de Filologia Anglesa i de Germanística

Universitat Autònoma de Barcelona

February 2017

To Ivone Antônia Foresti

Acknowledgements

There are several people I would like to thank for helping me in the completion of my PhD. First and foremost, I want to thank my supervisor, *Dr. Juli Cebrian*. His invaluable feedback, support, attention to detail and drive for excellence have been crucial to the completion of this piece of work. I am deeply appreciative of his time and dedication, which enriched not only this dissertation, but my entire academic experience. Thank you so much Juli, I really couldn't have done it without you.

I am also thankful to all the professors of the English department at the UAB who have helped me to improve my knowledge, but especially to Professor *Dr. Maria Josep Solé*, who has collaborated intensively in my professional development as a phonetics and phonology instructor and researcher throughout the years. Moreover, I would like to say thanks to *Dr. Hortensia Currell* for her friendship, support and academic guidance and also to *Dr. Montserrat Capdevilla* and *Dr. Mireia Llinàs* for providing me with professional opportunities. Also, a special thanks to *Arnau Cunill* and *Dr. Susagna Tubau* for their assistance in the collection of the post-test data for this dissertation.

A special thanks to *Dr. Joan Carles Mora*, for all the advice and feedback throughout the years and for providing me with the opportunity to participate in the *Barcelona L2 speech group* meetings. In these meetings, I found the opportunity to meet like-minded peers, present my work, acquire more knowledge and grow as a researcher.

Very importantly, I would like to say a huge thanks to *Dr. Ocke Bohn*, who accepted me on a research stay abroad, allowed me to attend his wonderful lectures and provided me with invaluable feedback on my research design. The experience broadened my views of academia and provided me with additional necessary skills to complete this piece of work.

A special thanks to *Dr.Denise Cristina Kluge* for accepting me on my second research stay abroad, sharing her knowledge and expertise with me, and providing me with great academic opportunities while abroad whilst becoming a special friend.

I would also like to thank all the participants (and former students) and talkers of this study, as it would have not been possible without them. Moreover, I would like to say a huge thanks to all of the raters, who are also my dear friends from Today School Barcelona (you know who you are!). I would like to especially thank *Amanda Maddock*, who was always keen to help whenever native ears were required. Also, thanks to *Carles Salse*, who helped me with the technical aspects of the data collection at UAB.

I am thankful to *Dr. Paul Iverson* for making me feel welcome on my visit to UCL and for his advice and support during my stay. Also to *Emma Brint*, who assisted me with the recruitment and recording of the English native talkers for this study. Their help was essential in making the creation of original materials for this project possible.

A very special thanks to my husband, *Niall Richardson*, who has been my pillar throughout these years. He has always believed in me, especially encouraging me in moments of doubt. Moreover, apart from all of the invaluable emotional support, he has also contributed to different stages of this thesis, from the initial stimuli validation to the very final proofing.

My family and friends deserve my warmest thanks, especially my parents (*Ivone Antônia Foresti* and *Josué Carlet*), my step father (*Vanor Giacomello*), my brothers (*Ismael Foresti Carlet, Giovani Foresti Carlet and João Paulo Giacomello*), my sister-in-law (*Kelly Romagna Carlet*) and my mother-in-law (*Elisebeth Anne Richardson*), who even from a distance, supported and encouraged me at all times.

The "journey" of writing a PhD brought many special people into my life and I do not have enough words to thank all the friends who supported me and somehow helped me to conclude this work. *Carmen Andrés*, *Mariana Larentis* and *Maria Rosa Garrido*, for always being there throughout the entire process. I thank them for the fun moments and also for helping and supporting me during the tough ones. Thanks for becoming true friends and for being there for me. *Hanna Kivistö-de Souza* and *Anabela Rato*, with whom I became very close to from the day we met. I thank them for their friendship, support, for being academic role models for me and for the joint study ideas that we have embarked on so far. Let these fruitful collaborations and amazing overseas friendships carry on. To *Anna Marczyk*, *Andreia Rauber*, *Izabella Julkowska*, *Fernanda Gláucia* and *Camilla Søballe Horslund* for becoming dear friends, providing me with academic support and especially for making this journey much more enjoyable. Moreover, I would like to thank all of the participants of the *"BCN academic writing group"* who believed in the idea of making PhD writing a less lonely process. A special thanks to *Robbie*, *Sapna* and *Mar*, who became good friends along the way.

Moreover, during this last year of my PhD, I have had the challenge, but also the great opportunity, of working in two different institutions, namely UPF and UIC. I would very much like to thank several people whom I met along the way, especially *Dr. Carmen Pérez Vidal* for believing in me and for providing me with academic opportunities, *Sonia López* and *Helena Roquet* for the peer support and motivation and to *Victoria Monje Sangüesa* for helping me out with parts of the editing of this thesis. Also, a warm thanks to all of my current colleagues at the education department at the UIC, who have emotionally supported me in these final stages.

Last, but undoubtedly not least, I am sincerely thankful for the financial support I have received from the English department at UAB through the PhD grant ***PIF (429-02-1/2011)*** and to the research grant from the Spanish Ministry of Economy and Competitiveness ***(FFI2013-46354-P)***.

Barcelona, February 2017

Abstract

Learning a foreign language (L2) in an instructional setting is characterized by limited exposure to the target language. This scenario might be problematic for accurate L2 language learning, since authentic input is necessary to enhance L2 learning. Against this background, a possible source of target language experience and immediate corrective feedback can be found in L2 phonetic training, as it provides learners with native input and may focus on particularly challenging L2 structures. This study compares the effect of two high variability phonetic training (HVPT) methods on specifically attended sounds and on implicitly exposed but unattended sounds. Several training regimes are implemented aimed at improving the perception and production of a subset of English vowels (/i ɪ æ ʌ ɜː/) and initial and final stops by Spanish/Catalan bilingual learners of English. Thus this study addresses the following questions: (a) whether training can improve the perception and production of trained as well as untrained segments, (b) whether improvement generalizes to novel stimuli and talkers, (c) if improvement is retained over time, (d) which training method (Identification (ID) or categorical Discrimination (DIS)) is more effective, and (e) what are the participants' impressions of phonetic training as a L2 training tool. A total of 100 bilingual Catalan/Spanish learners of English were divided into four experimental groups and a control group and were tested on their identification of English sounds presented in CVC non-words before and after a five-week training period, and two months later. L2 production was assessed before and immediately after training through a picture naming task and analysed by means of native speaker judgments. The trained groups differed either in terms of training method (ID, DIS) or focus of training (consonants, vowels), resulting in four different groups. Crucially, all four groups were trained with the same sets of CVC non-words (e.g. *zat*, *zut*, *zad*, *zud*), exposing learners to attended contrasts within trials and to unattended contrasts across trials.

The results reveal that all experimental groups significantly outperform the controls in their identification of trained sounds (vowels and initial stops), showing the efficacy of both phonetic training methodologies (ID and categorical AX DIS). However, while both experimental groups perform similarly when modifying initial stop perception, the ID trainees outperform the DIS trainees on trained vowel perception. These results suggest that modifying the perception of different types of segments might require different training procedures and amounts of training time. Interestingly, only the DIS trainees show a significant improvement in the perception of untrained/unattended L2 sounds, indicating that this training method may be more suited to enhancing learners' perception of attended as well as unattended target sounds. Regarding generalization and retention, the results point to the superiority of the ID task over a categorical DIS task when training vowel sounds. Moreover, the results indicate that both methods are well suited to training initial consonants to the same extent. With respect to production, only the vowel ID trainees significantly improve their production of trained sounds, which shows that pronunciation improvement might take place as a result of an identification perceptual training regime, even in the absence of production training. Finally, students' opinions of phonetic training as an EFL tool are positive overall and ID is favoured over DIS as a training method. Globally, these findings suggest that while both methods are effective for training L2 perception, ID and DIS methods may promote improvement, generalization and retention for vowels and for consonants to different degrees. The better results obtained with ID training, particularly for vowels, and the fact that only DIS promoted improvement with untrained sounds (cross-training effects) may be related to the nature and focus of the tasks and/or to the acoustic characteristics of the target sounds. These results may have implications for future research on phonetic training and practical applications in the teaching of L2 pronunciation.

Keywords: Phonetic training, L2 speech learning, attention, L2 perception and production

Resumen

El aprendizaje de un idioma extranjero en un entorno formal se caracteriza por una exposición limitada a la lengua meta (L2), lo que puede ser problemático debido a la falta de “input” auténtico, necesario para mejorar el aprendizaje de una L2. En este contexto, una alternativa a la escasez de "input" puede ser el entrenamiento fonético, que proporciona a los alumnos un aporte nativo y puede centrarse en estructuras de la L2 particularmente difíciles para los aprendices. Este estudio compara el efecto de dos métodos de entrenamiento de alta variabilidad fonética o ‘high variability phonetic training’ (HVPT) sobre sonidos específicamente entrenados y sobre sonidos no entrenados pero implícitamente presentados. Con este fin, se implementan diversos regímenes de entrenamiento fonético cuyo objetivo es mejorar la percepción y la producción de 5 vocales del inglés británico (/i ɪ æ ʌ ɜː/) y de las consonantes oclusivas en posición inicial y final de palabra por hablantes bilingües de catalán y castellano. Así, este estudio investiga: (a) si el entrenamiento fonético recibido puede mejorar la percepción y la producción de segmentos entrenados y no entrenados, (b) si la mejora se generaliza a nuevos estímulos y hablantes, (c) si la mejora se mantiene un tiempo después, (d) qué método de entrenamiento, identificación (ID) o discriminación categórica (DIS) es más eficaz, y (e) cuáles son las impresiones de los participantes sobre el entrenamiento fonético como una herramienta de instrucción fonética. Cien estudiantes de inglés como lengua extranjera fueron divididos en cuatro grupos experimentales y un grupo de control. Su identificación de los sonidos del inglés fue evaluada inmediatamente antes y después de un período de entrenamiento fonético de cinco semanas y también dos meses después de la finalización del mismo. Los estímulos fueron presentados en palabras sin sentido en el formato de consonante-vocal-consonante (CVC). La producción de los sonidos de la L2 fue evaluada a través de una tarea de denominación de imágenes antes e inmediatamente después del entrenamiento y analizados mediante juicios de hablantes nativos de inglés. Los grupos entrenados diferían tanto en términos de método de entrenamiento (ID, DIS) como en el enfoque del entrenamiento (consonantes, vocales) dando lugar a cuatro grupos diferentes. Crucialmente, los cuatro grupos fueron entrenados con los mismos estímulos CVC (por ejemplo, zat, zut, zad, zud), exponiendo a los participantes a los contrastes fonéticos entrenados y a los contrastes fonéticos no entrenados.

Los resultados revelan que todos los grupos experimentales superaron significativamente al grupo de control en su identificación de sonidos entrenados (vocales y consonantes oclusivas en posición inicial), mostrando la eficacia de ambas metodologías de entrenamiento fonético (ID y AX DIS). Sin embargo, mientras que ambos grupos experimentales mejoran su percepción de las oclusivas iniciales de manera similar, los aprendices de ID superan a los aprendices de DIS en la percepción de vocales específicamente entrenadas después del entrenamiento fonético. Estos resultados sugieren que la modificación de la percepción de los diferentes tipos de segmentos (vocales, consonantes) puede requerir diferentes procedimientos y duraciones de entrenamiento distintas. Curiosamente, sólo los aprendices de DIS mostraron una mejora significativa en la percepción de los sonidos no específicamente entrenados, lo que indica que este método de entrenamiento puede proporcionar mejoras en la percepción de sonidos entrenados y sonidos no entrenados pero implícitamente presentados. En cuanto a la generalización y a la retención de los efectos del entrenamiento, los resultados con sonidos vocálicos apuntan a la superioridad de la tarea de ID sobre la tarea categórica de DIS. Además, ambos métodos son adecuados para entrenar consonantes iniciales de manera similar. Con respecto a la producción, sólo los aprendices de ID entrenados en vocales fueron capaces de mejorar significativamente su producción de los sonidos vocálicos. Este resultado demuestra que la mejora en la producción de sonidos de una L2 puede tener lugar a través de un régimen de entrenamiento de percepción, incluso en la ausencia de entrenamiento específico de producción. Por último, las opiniones de los estudiantes acerca del entrenamiento fonético como una herramienta de enseñanza de L2 fueron en general positivas, e ID fue más valorado que DIS como un método de formación. Globalmente, estos resultados sugieren que ambos métodos son efectivos para entrenar la percepción de una L2. Sin embargo, los métodos pueden promover mejoras, generalización y retención de los distintos segmentos en diferentes grados. Los mejores resultados obtenidos con el método ID, en particular con las vocales, y el hecho de que sólo el método DIS proporcione la mejora de sonidos no entrenados pueden estar relacionados con la naturaleza y el fin de cada metodología y/o con las propiedades acústicas de cada segmento. Las consecuencias teóricas y prácticas de estos resultados pueden ser de utilidad para futuros trabajos de investigación y aplicaciones prácticas de aprendizaje de la pronunciación.

Palabras clave: Entrenamiento fonético, aprendizaje de L2, atención, percepción y producción de una L2

Table of contents

List of figures

List of tables

1. Introduction

Research in adult second language (L2) acquisition has revealed that L2 sounds that are non-existent in the learners' first language (L1) tend to be challenging due to the interplay of several factors, including the effect of existing L1 phonetic categories[1]. This difficulty in non-native sound categorization can be related to a failure in perceiving and consequently producing some phonetic categories of the target language accurately (Bohn & Munro, 2007).

The influence of existing L1 phonetic categories on the development of L2 categories has been the object of study of a number of L2 speech models. The most closely connected to the present study are subsequently described. The Speech Learning Model (SLM, Flege, 1995, 2003), one of the most influential L2 speech models, states that a non-native sound that is dissimilar to a pre-existing L1 sound will eventually be acquired with more ease than an L2 sound that has a similar counterpart in the L1 language inventory. According to the SLM, given enough input and experience, learners might succeed in establishing separate long-term memory representations from the pre-existing L1 categories and, as a consequence, both L1 and L2 categories will coexist in the same phonological space. A second relevant speech perception model, the Native Language Magnet model (NLM e.g., Kuhl & Iverson, 1995), also postulates that non-native sound perception occurs within the native language's boundaries. In fact, the authors explain that good exemplars of L1 speech sounds act as perceptual magnets and thus attract perceptually similar target language sounds, which impedes the formation of new categories. Similarly, the Perceptual Assimilation Model (PAM-L2, Best & Tyler,

[1] Phonetic categories are defined as "the distribution of acoustic tokens which together are perceived as mapping to a phoneme in the listener's inventory" (Earle & Myers, 2014, p. 1192).

2007), explores the possible outcomes of the interference from L1 categories on L2 sound discrimination accuracy by establishing and describing possible cross-language category assimilation patterns.

Looking beyond the influence of the L1 phonological system, there are numerous additional factors that might contribute to poor L2 perception and/or accented speech, such as the incoherent relationship between spelling and pronunciation (Pierce, 2014, Thomson & Isaac, 2009; Thomson & Campagna, 2010) and other non-linguistic factors such as motivation, onset age of learning (AOL), length of residence (LOR) in the target-language country, the amount of L2 language experience and amount of L1 and L2 use, amongst others (Calvo Benzies, 2014; Darcy, Ewert & Lidster, 2012; Piske, MacKay & Flege, 2001).

Previous studies have revealed that increased L2 experience (greater amount of native input and L2 use) might promote L2 category learning. In fact, exposure to the L2 is viewed as a crucial aspect in order to develop the ability to distinguish native from non-native sounds (Flege, 1991; Ingram & Park, 1997) and yield better perception and consequently better production of L2 sounds in naturalistic settings (e.g., Flege, Bohn & Jang, 1997).

The present thesis, however, is set in an instructional English as a foreign language (EFL) context, where native input is not always present in the classroom and is limited outside the classroom (Muñoz, 2008; Saito, 2012, 2015). Research investigating the effect of formal instruction (FI) on oral skills and/or pronunciation has failed to provide evidence that such input can alter these domains (Fullana & Mora, 2009; García-Lecumberri & Gallardo del Puerto, 2003; Monje-Sangüesa, 2016), unless a large amount of exposure is undertaken (Saito, 2015) and is accompanied by corrective feedback (Saito, 2012) and/or explicit pronunciation instruction (Kissling, 2013; Thomson & Derwin, 2014; Gordon & Darcy, 2016). One noteworthy reason for this lack of success lies in the fact that English pronunciation has been described as one of the most challenging skills to be learned (Calvo

Benzies, 2016) and taught in the FL classroom (Celce-Murcia, Brinton & Goodwin, 1996; Darcy et.al., 2012). Due to the scarcity of good quality input in the FL context described above, specialized phonetic training emerges as a potentially effective source of foreign language experience. Phonetic training can provide leaners with the amount of native input and of corrective feedback needed to enhance L2 learning. According to Iverson, Pinet and Evans (2012), phonetic training "provides a useful addition to real-world L2 experience" (p.15).

A common type of phonetic training is "perceptual training", which comprises specialized perceptual techniques that aim at directing second language learners' attention to specific phonetic cues and challenging contrasts present in the target language by means of feedback. In doing so, perceptual training has been shown to be efficient in improving learners' ability to better perceive and consequently produce L2 sounds. Several phonetic training studies have been shown to have a positive effect on the perception/production of L2 consonant sounds (Flege, 1995b; Hardison, 2003; Hazan, Sennema, Iba & Faulkener, 2005; Iverson & Evans, 2007; Pisoni, Lively & Logan, 1994; Strange & Dittmann, 1984, Yamada, Tohkura, Bradlow & Pisoni, 1996; among others) and L2 vowel sounds (Aliaga-García, 2010; Aliaga-García & Mora, 2009; Cenoz & García-Lecumberri, 1999; Iverson & Evans, 2007; Lacabex, García-Lecumberri & Cooke, 2008; Rato, 2014; among others).

Most of the success reported in phonetic training research has been attributed to the exposure of learners to highly variable input, known as High Variability Phonetic Training (HVPT, Logan, Lively & Pisoni, 1991). This method exposes trainees to a variety of talkers and contexts and thus forces participants to pay attention to distinctive characteristics of the target sounds rather than specific talker-related characteristics. Moreover, HVPT training has been found to promote the generalizability of learning to multiple dimensions (Logan & Pruitt, 1995; Lively, Logan & Pisoni, 1993; Bradlow, Pisoni, Akahane-Yamada & Tohkura,

1997) and to promote retention of learning over time, thus providing support for the robustness of the training method. According to Logan and Pruitt (1995), the choice of training task and generalization task is critical in order to understand the processes that underlie successful perceptual learning. Although both identification (ID) and discrimination (DIS) training tasks may be useful in training different perceptual skills (Logan & Pruitt, 1995), the effectiveness of ID training has been said to be superior to discrimination training (Jamieson & Moroson, 1986). A possible reason might be the fact that ID procedures force listeners to attend to relevant between-category variability, while DIS tasks focus on within-category variability (Jamieson & Moroson, 1986). However, Flege (1995b) directly compared these two types of methodologies (AX discrimination and forced-choice identification) in a study aimed at training Mandarin leaners of English to perceive English unreleased final /t/ and /d/. The results revealed that both types of training improved learners' perception and generalized their learning equally, challenging previous results and views on each methodology. Two recent studies have also shown the effectiveness of the two methods when training tone contrasts (Wayland & Li, 2008) and coda nasals (Nozawa, 2015).

The current study follows up on Flege's (1995b) study and expands it by looking at the effect of the two types of HVPT methods on the perception and production of vowels and of word-initial and word-final stop consonants. Specifically, this study explores the effect of two high variability phonetic training (HVPT) methods – identification (ID) and categorical[2] discrimination (DIS) - and investigates whether they promote improvement of perception and production of standard Southern British English (SBE) sounds by native speakers of Catalan. Furthermore, this study addresses the issue of whether the knowledge obtained through training generalizes to novel items and voices and whether this knowledge is retained over time.

[2] According to Flege (1995b), both terms "categorical" and/or "categorial" can be used to describe a categorical/categorial discrimination task (CDT or DIS). Thus, in the present thesis they will be used interchangeably.

Moreover, this study is innovative as it also investigates wether there is an effect of implicit exposure on phonetic training through a "cross-training methodology"; that is, if training affects the untrained sounds that are present in the stimuli, and, if so, which of the two training method facilitates this learning the most. The above mentioned issues are answered by looking at the results obtained by 100 EFL undergraduate students in a longitudinal study which made use of a pre-test/post-test/delayed post-test design. The main aim of this research is to better understand the learning mechanisms that are crucial to the enhancement of L2 acquisition. Ultimately, findings from this study intend to contribute to a better understanding of the role of perceptual training on L2 learning and the development of non-native categories.

1.1. Organization of the dissertation

The current dissertation is organized into 7 chapters/sections. **Chapter 1** consists of a brief introduction to the field and to the current study. **Chapter 2** presents an overview of the literature important to the present study, concerning speech perception, production, and phonetic training, and giving emphasis to high variability phonetic training (HVPT). **Chapter 3** describes the sound inventories of the languages investigated in the present study, namely Spanish/Catalan (the participants' L1(s)) and standard Southern British English (the participants' target L2), briefly describes relevant L2 speech studies involving Catalan/Spanish learners of English on both vowel and consonants, and finally presents the hypotheses following the L2 speech perception models. **Chapter 4** describes the aims of the present investigation by stating the research questions and consequently formulates hypotheses in line with the previous literature described. **Chapter 5** explains the methodology used, describing the experimental design, the participants, the training and testing tasks and the procedure that formed part of the present study. **Chapter 6** presents the results and is divided into four different parts. *Part I* describes the perceptual data results obtained and discusses the main findings; *part II* describes the results found for generalization and retention effects and explains these results; *part III* presents the production results found in the study and its corresponding discussion, and the last part (*part IV*) describes and explains the qualitative results obtained through the post-training survey. Lastly, **Chapter 7** provides the reader with a global discussion of the outcomes, discusses some pedagogical implications and study limitations and additionally makes suggestions for further research. Finally, it summarizes the conclusions of the study.

2. Literature review

This chapter presents an overview of the main issues related to the effects of phonetic training on the perception and production of vowels and consonants by EFL learners. The present chapter is organized in three sections. The first one provides a contextualisation of L2 speech learning in EFL contexts and a brief description of the obstacles encountered by learners throughout the process (2.1). In the second one (2.2), three relevant L2 speech perception models are briefly explained, namely the NLM, the SLM and PAM/PAM-L2. Lastly, the third one deals with phonetic training and provides an overview of the main research findings in the field (2.3).

2.1. L2 speech learning in an EFL context

It is notorious that the foreign language (FL) setting is characterized by limited target language exposure (Celce-Murcia et al., 1996; Muñoz, 2008; Saito, 2015). According to Larson-Hall (2008), the FL instruction comprises minimal input conditions and is usually delivered in no more than four hours of instruction per week. Therefore, the quantity of language input received by learners acquiring a language in these contexts is usually restricted to the teachers' instruction, which might partly be delivered in the learners' native language (Muñoz, 2008). Another crucial issue is the quality of the input received (Derwing & Munro, 2015), since native input is particularly scarce in FL contexts and contact with foreign accented English produced by the peers in class ends up being a major source of target language exposure. Muñoz (2011) and Derwing and Munro (2015) highlight the need to provide learners not simply with additional exposure but also with greater quality of exposure inside and outside the classroom. Carlet and Rato (2015) found that Portuguese EFL learners outperformed Catalan learners of English when identifying and discriminating aspirated and unaspirated English voiceless stops, despite the comparable amount of years of formal instruction (FI) between the two groups and the similarity of the L1s. The authors attributed these outcomes to the greater quantity and possibly quality of input received by the Portuguese learners outside the classroom. This is because Portuguese learners are exposed to native English input through original version TV programs and films on a regular basis rather than being generally exposed to foreign shows and films dubbed in the local language, as is the case in Spain. Thus, the Spanish/Catalan FL scenario described here is problematic, since authentic input has been found to be a compulsory condition for enhanced L2 knowledge (Long & Larsen-Freeman, 1991; Celce-Murcia et al., 1996; Flege, 1991).

Another drawback of formal instruction (FI) is the lack of focus on pronunciation teaching. In fact, pronunciation has often been described as the neglected skill in the ESL classroom (Celce-Murcia et al., 1996; Morley, 1991) despite its importance and interconnection with the four basic linguistic skills: reading, writing, listening and speaking (Darcy et al., 2012). Morley (1991) claims that ignoring pronunciation teaching entails a lack of responsibility on the part of the teacher, as pronunciation empowers one's communicative skills and also increases one's intelligibility and self-confidence.

Historically, the amount of time dedicated to pronunciation in the EFL classroom has varied over the years due to different reasons, including the teachers' lack of training and/or confidence (Baker, 2014; Derwing, 2010; Murphy, 2014), the lack of appropriate teaching materials (Busà, 2008; Derwing, 2008; Darcy et al., 2012) and especially the fact that other skills, such as grammar and vocabulary, are prioritized (Morley, 1991; Levis, 2005). Moreover, English pronunciation has been labelled as one of the most difficult skills to be acquired and developed by Spanish L2 learners (Martínez-Flor & Usó-Juan, 2006; Aliaga-García, 2007; Calvo Benzies, 2016) and to be taught in the EFL classroom (Celce-Murcia et al., 1996; Darcy et al., 2012). Besides, despite the recognized importance of pronunciation teaching and the fact that explicit pronunciation instruction can be quite effective (Kissling, 2013; Thomson & Derwin, 2014), teachers often remain uncertain about how to incorporate it into the curriculum (Celce-Murcia et al., 1996; Levis & Grant, 2003).

As a consequence of the lack of focus on pronunciation in the EFL context, studies have reported mixed results when investigating pronunciation attainment as a result of FI. For instance, in a study assessing the impact of formal instruction on the perception and production of English final stops by Spanish/Catalan learners of English in an FL context, Fullana and Mora (2009) failed to find a significant improvement. Along the same lines, García-Lecumberri and Gallardo del Puerto (2003) have reported pronunciation attainment in a FI setting to be feeble when looking at Spanish learners differing in age and proficiency.

Conversely, Pérez-Vidal, Juan-Garau and Mora (2011) found a significant improvement in the perception and production of English vowels and stop consonants as a result of English formal instruction. Similarly, Saito and Hanzawa (2015) found that large amounts of FL instruction (> 875hr) resulted in improved L2 oral ability by Japanese learners of English; however, the large inter-subject variability found in their data suggests that success was somewhat dependent on the extra-curricular language experience that some learners were exposed to. In a previous study, Saito and Lyster (2012) pointed to the need of corrective feedback in order for L2 formal instruction to become effective. Thus, the little success in L2 phonological development reported in these studies may be associated with the lack of attention pronunciation receives in the L2 classroom (Celce-Murcia et al., 1996; Calvo Benzies, 2016; Aliaga-Garcia, 2007), alongside the quantity and quality of target language input available (Muñoz, 2011; Derwing & Munro, 2015; Carlet & Rato, 2015).

Against this background, a possible source of target language experience and immediate corrective feedback can be found in L2 phonetic training, as it provides learners with an opportunity to receive the necessary native input to acquire distinctive phonetic differences. In fact, Pereira (2014) reported that Chilean learners of English were able to improve their L2 vowel perception during a six-week training regime to a similar degree to the one observed after a period of three years of formal instruction. Phonetic training is the focus of the present study and will be further explored in section (2.3).

2.2. L2 speech models

The current subsection aims at describing three of the current speech perception models that explain the L2 phonological acquisition phenomenon from the interaction between the L1 and L2 systems: the Native Language Magnet Model (Kuhl & Iverson, 1995; Iverson & Kuhl, 1995, 1996; Kuhl, 2000), the Perceptual Assimilation Model (Best, 1995; Best & McRoberts, 2003; Best & Tyler, 2007) and the Speech Learning Model (Flege, 1995, 2003). Additionally, the SLM addresses the issue of the relationship between perception and production. Moreover, these models predict that language experience and enough exposure to good quality input (as the one received through phonetic training) are indispensable for L2 successful attainment. Even though the present study is not directly testing the ability to discern L1 from L2 sounds, it is concerned with assessing if perception of difficult contrasts improves as a result of language experience as provided by a laboratory training regime, and thus the description of these models is part of the scope of this thesis.

2.2.1. Native Language Magnet model

According to the Native language magnet model (NLM), language and perceptual abilities are innate and are later affected by the speaker's experience with the L1. Thus, this theory is based on the findings that speech perception for adults and infants is altered and/or constrained as a function of previous language exposure and not by lack of brain plasticity. Interestingly, according to the model, infants in their first year of life are non-language-specific perceivers, as they are able to discriminate any given sound, native or non-native. After this first year of life, discrimination abilities focus on the L1 sounds as the speaker tunes into the sounds of the ambient language and becomes a language-specific perceiver (Kuhl, 1993).

The NLM postulates that unfamiliar sounds are perceived within the native language frame. This is so because adult listeners of a certain language are adept at identifying the best exemplars of phonetic categories (prototypes) in their L1 and these prototypes play a special role in speech perception. Perceptually, the prototypes pull other exemplars of the category towards them, acting as native language magnets (Kuhl, 1993). Infants exhibit the perceptual magnet effect in their native language, which aids infants acquire the perceptual categories of their L1.

Regarding L2 sound acquisition, L2 sounds that are perceptually comparable to L1 sounds may be perceived as poor exemplars of L1 sounds at first and thus mapped onto pre-existing L1 categories. Different levels of L1-L2 mapping difficulty are predicted by the model. According to the NLM, the more similar an L2 sound is to the L1 system, the greater the native language magnet effect and therefore the greater the difficulty to distinguish these sounds from L1 sounds. Conversely, the more dissimilar an L2 sound is, the easier discrimination becomes as this sound is not bound to be highly affected by the L1 magnet. As regards to L2 production, the NLM postulates that speech perception precedes and alters production (Kuhl & Iverson, 1995; Iverson & Kuhl, 1995).

2.2.2. Perceptual Assimilation Model

The Perceptual Assimilation Model (PAM) is an extension of the “Direct-realist approach” to perception (Fowler, 1986; Gibson & Gibson, 1955) in that its core principle considers that perception is directly created from the environment, and it is not filtered through mental constructs. Therefore, mental representations of phonetic categories are not required in order to yield perceptual learning according to PAM. Conversely to representationalistic views, PAM maintains a direct realist approach and claims that perception is not always 100% accurate. Therefore, at the infant stage, sound perception is attuned to the L1 inventory and, consequently, identifying familiar sounds becomes natural and ecological. Preliminarily, sounds are conceived in terms of grouped speech gestures and not individual gestures being presented simultaneously (Best, 1995). However, having established this finely attuned perception of grouped speech gestures, it becomes more difficult to perceive speech sounds that differ from those in terms of gestures and combinations, such as non-native sounds. In Best's words (1995) non-native segments are the ones whose "gestural elements (…) do not match precisely any native constellations"(p. 193).

Best and Tyler (2007) extend the original model to L2 perception (PAM-L2) and describe different perceptual assimilation patterns of L2 phonemes into the native phonemic inventory. PAM-L2 postulates that the L1 does not necessarily hinder L2 perception. In fact, this model makes predictions about how successful the perception of a certain non-native contrast should be for the L2 learner according to its degree of similarity or dissimilarity with the native phones. Perception of non-native sounds is hence connected to that of native sounds. More specifically, PAM-L2 postulates that "categorisation" takes place when a non-native sound is perceived as existing in the native language, whereas it is "uncategorised" if it fails to be ascribed to it. In other words, when learners incorporate new L2 sounds to their phonological system, they assimilate that new sound according to the acoustic cues present in their L1, perceiving the new sound as pre-existing one(s) in their phonological space, instead of creating a new category for that sound. A last scenario anticipated by PAM-L2 involves the sound being perceived as a non-speech sound, and thus patterning as "non-assimilated" in terms of the pre-existing L1 categories. The different types of assimilation patterns for unfamiliar L2 sounds into the L1 phonological space predicted by PAM/PAM-L2[3] are the following:

1. **Two-Category Assimilation (TC),** when the listener perceives two L2 sounds in terms of two different L1 categories. Discrimination is predicted to be excellent.
2. **Category-Goodness Assimilation (CG),** when two L2 sounds map onto the same category in the native sound system, but one is a better fit than the other. Discrimination is predicted to be moderate to good.
3. **Single-Category Assimilation (SC),** when two L2 sounds are heard as the same L1 phoneme, but neither is really a good fit, and neither is particularly 'better' than the other. Discrimination is predicted to be poor.
4. **Uncategorized-Categorized assimilation (UC),** involves two L2 sounds, one is assimilated to a native category and the other one is uncategorized. Discrimination of contrasts is expected to be good to very good.
5. **Uncategorized-Uncategorized (UU),** when two non-native sounds are uncategorized. The level of difficulty in the discrimination of these sounds will depend on how perceptually close the two non-native sounds are from each other and from the L1 phones. Thus, discrimination can vary from poor to excellent.

According to Best and Tyler (2007) a speaker's L1 experience can systematically constrain perception of L2 speech contrasts. In their analysis, the authors also explore an assumption that adult learners experience difficulties perceiving non-native sounds, but nevertheless improvement is possible as perceptual learning continues in adulthood and increasing experience with the L2 triggers perceptual readjustments. To uncover new insights

[3] PAM-L2 (Best & Tyler, 2007) differs slightly from PAM (Best, 1995), as it posits four different types of assimilation patterns rather than five. The TC and UC assimilation patterns are two possible PAM patterns corresponding to PAM-L2 case 1(TC).

on non-native speech perception, Best and Tyler (2007) make a critical comparison of Flege's (1995) SLM and PAM. The most striking and relevant distinction concerns the notion of phonetic category. While in the SLM the mechanisms that are used to acquire the L1 sound system, including the formation of long-term memory representations (phonetic categories), tend to remain intact over the life span and are thus available to L2 learners, PAM's ecological approach argues that rather than creating a new category, L1 phonetic categories can evolve over time to incorporate the additional properties of the L2 categories. However, it is also postulated that L2 speakers struggle to establish a contrast between first language and second language phonetic categories that co-exist in a common phonological space (Best & Tyler, 2007). No decisive conclusions are reached regarding perception necessarily having to take place before production by the authors. However, it is not discarded that failing to correctly perceive a new L2 sound can prevent its subsequent production.

2.2.3. Speech Learning Model

One of the most influential models in L2 speech perception and production research is Flege's Speech Learning Model (SLM, Flege, 1995, 2003). The SLM bears differences from other second language acquisition approaches that focus primarily on naive listeners such as the Perceptual Assimilation Model (PAM). Conversely, the SLM's focus shifts to experienced listeners (Flege, 1995). In fact, a crucial assumption made by SLM is that the L2 speaker's phonological system remains malleable over the life span. In this sense, the SLM is clearly pertinent to the population of this study and it challenges the validity of the canonical Critical Period Hypothesis (CPH) for L2 acquisition, by predicting different degrees of success in L2 speech perception based on the phonetic distance between L1 and L2 phonemic categories and the effect of experience with the L2 (Flege, 1995).

The SLM does not question the fact that early learners are able to form new L2 categories with more ease than late learners (Flege, 1995; Flege & Mackay, 2004). However, it suggests that young learners acquiring their first language (L1) are more likely to receive greater quantity and higher quality of input than the average late L2 learner (Flege, 1995) and both quality and quantity of input play important roles in L2 category formation (Flege, 2016). Another important difference between children learning their L1 and L2 adult learners is the fact that "when L1 learning children misarticulate sounds, they get feedback. L2 learners, especially adults, rarely get this kind of feedback" (Flege, 2016, p. 19).

According to the SLM, phonetic categories are long-term mental representations of the distinctive features of speech sounds necessary to process speech. Moreover, the model posits that L1 phonetic categories established in the early years are continuously modified over the life span in order to reflect the properties of all L1 or L2 phones identified as a realization of each category (Flege, 1995). Furthermore, it is suggested that the L1 and L2 categories cohabit in a shared phonological space, unavoidably influencing each other. The model posits that learners perceive the sounds of the L2 in relation to the nearest sound of the L1, though new category creation may be possible. In fact, L2 sounds might be perceived as "identical", "new" or "similar". The further an L2 sound is from an L1 phoneme perceptually, the more likely it is perceived as a "new" sound and consequently a new category may be established, given enough exposure to the target language (e.g. L2 English /ɜː/ having no clear match in L1 Catalan)[4]."Identical" categories, that is, sounds that are indistinguishable from the L1 counterparts are predicted not to pose any problems to L2 learners, as no category formation is required (e.g. L2 English /iː/ mapped onto L1 Catalan /iː/). Conversely, the more "similar" an L2 sound is perceived to be from an L1 sound, the less likely it is to be successfully acquired (e.g. L2 English /ʌ/ mapped onto L1 Catalan /a/).

[4] Cross-linguistic mappings are further explained in section 3.1.

The failure to establish a distinct category for "similar" phones is predicted to result in L1-based foreign-accented productions. Importantly, detecting differences between L1 and L2 sounds is crucial in order to ultimately establish accurate target-like categories for L2 sounds. In Flege's words, "the greater the perceived phonetic dissimilarity between an L2 sound and the closest L1 sound, the more probable it is that phonetic differences between the sounds will be discerned" (Flege, 1995, p. 239). Moreover, unlike the articulatory-driven basis of the PAM model, the SLM suggests that the development of mental representations is triggered auditorily, which is the scope of the present study.

As far as production is concerned, the SLM attributes L2 phonological errors mostly to incorrect perception. Nonetheless, other possible causes for inaccurate production are not completely discarded. Furthermore, the SLM predicts that even when a new phonetic category is formed for an L2 sound, it may not necessarily be produced in a native-like manner. However, this outcome might be changed and consequently improved so long as enough authentic input is provided (Flege, 1991, 1995).

Among the specific predictions made by each model, there are underlying similarities to be found. The three models postulate that early experience with a native language outlines the perceptual organization of the listener and this organization perceptually filters the way L2 sounds are perceived and consequently categorized. Moreover, the three models agree that L2 language experience is key to improving phonological attainment as it might enhance the capacity to establish new categories of a target language, or re-organize the L1 categories, in the case of PAM-L2. Hence, phonetic training might be an efficient tool to provide the necessary input to enhance L2 categorization (Thomson & Derwin, 2014).

2.3. Perceptual training

Perceptual training comprises specialized perceptual techniques that aim at directing second language learners' attention to specific phonetic cues present in challenging target language contrasts by means of feedback (Cebrian & Carlet, 2014). In doing so, perceptual training has been shown to be efficient in improving learners' ability to better perceive (Iverson & Ewans, 2009; Lacabex et al., 2008; Logan et al., 1991; Strange & Dittmann, 1984; Thomson, 2012, among others) and to a lesser extent consequently produce L2 sounds (e.g. Bradlow et al., 1997; Hazan et al., 2005; Aliaga-García & Mora, 2009 for consonants; Thomson, 2011; Rato & Rauber, 2015; Iverson et al., 2012 for vowels). Moreover, as previously mentioned, in the absence of sufficient target language input, phonetic training can be an alternative way of exposing EFL learners to selected target language structures in order to trigger the necessary processes required for L2 category learning. Perceptual training emerged in the 1960s as a therapy procedure in the area of speech and hearing sciences. It originally aimed at aiding the performance of patients with speech disorders and/or hearing impaired individuals (Pierce, 2014). Over the last few decades, several researchers have transferred this therapy procedure to the field of second language acquisition and the efficacy of perceptual phonetic training with L2 learners has been investigated since.

2.3.1. Methodological issues in phonetic training

Logan and Pruitt (1995) list several factors that predict the success of phonetic training regimes, including the type of training task, the nature of the stimuli, the type of instruction and the presence or absence of feedback. A combination of the correct characteristics can generate an "optimal training formula". According to Ellis (2005), exposing learners to highly variable natural stimuli and immediate feedback on the identification of sounds leads to fast and efficient learning. These methodological components of training are described next since they are essential to the understanding of the present study.

2.3.1.1. Highly variable natural stimuli

Logan et al. (1991) incorporated the therapeutic procedure of auditory bombardment (Hodson & Paden, 1983) into training studies. As part of L2 speech learning research, it was coined as high variability phonetic training (HVPT). In both fields, the related procedures consist in providing repeated, systematic exposure to multiple exemplars of phonological targets and contrasts (Bowen, 1998). Thus, the stimuli usually involve contrastive minimal pairs produced by multiple talkers in multiple phonetic contexts. In their study, Logan et al. (1991) trained Japanese speakers to identify English /r/ and /l/ by means of fifteen identification training sessions of approximately 40 minutes in the course of three weeks. By exposing learners to natural stimuli from several voices (6 different talkers), the authors aimed at preventing the participants from relying on talker specific characteristics present in the stimuli. Post training identification scores confirmed that the learners were better able to perceive the target sounds after training, and they had also generalized their perceptual learning to novel words and voices. Lively et al. (1993) extended these findings by exposing one group of learners to only one talker and one group of learners to several talkers. The group exposed to the more variable stimuli outperformed the group trained by a single talker in both novel word and novel talker generalization results. More recent studies reproduced these findings (Pierce, 2014; Wong, 2014). For instance, Wong (2014) trained Cantonese ESL learners to better perceive and produce English /e/ and /æ/ and additionally assessed whether learners differing in English proficiency level benefited more from HVPT or from low variability phonetic training (LVPT). Both training methods enhanced learners' vowel perception and production abilities as well as promoted generalization; however, the HVPT groups outperformed the LVPT groups in all tasks, independently of their proficiency level. These two studies confirm that HVPT can be more effective than other procedures that use a single speaker for training late learners (but cf. Evans & Martín-Alvarez (2016) for young learners (9-12 years)). While the ability to use HVPT might improve with age (Evans & Martín-Alvarez, 2016), the variable input might allow mature trainees to establish more robust phonetic categories and thus generalize learning with more ease. According to Thomson (2012), this is so as it provides trainees with more natural L2 variability, similar to the variability that learners would face in a natural foreign language setting. Besides, it stops learners from relying on certain talker-specific characteristics (Logan et al., 1991).

Evaluating further the potential differences between HVPT and LVPT, Perrachionne, Lee, Ha and Wong (2011) assessed whether different students learn better from different types of stimuli variability in training procedures. Through some pre-training evaluation consisting of a battery of tests, they identified a group of high-aptitude learners (HAL) and a group of low-aptitude learners (LAL). Their findings revealed that HVPT improved learning only for learners with strong perceptual abilities (HAL). Participants with less robust perceptual abilities (LAL) were impaired by HVPT and improved significantly more when exposed to the LVPT regime. The authors mention that LAL may become overwhelmed by highly variable stimuli, which hinders their capacity to attend to the target features or contrasts to be learned. These results point to the importance of considering individual differences in pre-training aptitudes when evaluating the efficacy of any speech training paradigm.

In brief, the training studies reviewed in this subsection have shown that although both high and low variability methods might be efficient in improving learners' perception and/or production of L2 phones, HVPT has been found to obtain better results with adult learners. Still, the importance of accounting for individual differences (different levels of aptitude) has been raised. The present study tried to account for possible individual differences by equally distributing learners to the different groups, so that groups were comparable in terms of the participants' results at pre-test. Even though aptitude level is not controlled for in this measure, each experimental group presumably contained a similar number of "high perceivers" and "low perceivers" (see section 5.1.3 for more details).

2.3.1.2. Feedback

Feedback is defined as the "information provided to learners about their performance" during or after training (Logan & Pruitt, 1995, p. 361). Empirical evidence form early training studies reveal that feedback training is an efficient way of promoting rapid learning (Ellis, 2005), even though feedback might not be indispensable for learning to take place (Gibson & Gibson, 1955). The two most common types of feedback found in the bulk of training literature are "immediate or trial by trial feedback" and "cumulative feedback". The former refers to the information given to the participant immediately after each trial and the latter refers to the information provided to the learner after several trials or at the end of the training block/session. According to Logan and Pruitt (1995), while "trial by trial feedback" is crucial so that learners are able to carry on with or modify the way they are performing the training tasks accordingly, "cumulative feedback" probably does not influence L2 leaning directly, as it is mainly motivational. Perception training studies that use feedback have revealed fairly reliable results, as speakers' perceptual abilities are usually positively influenced by the training regime (Bradlow, 2008). The current study adopts a trial by trial feedback approach followed by a cumulative feedback approach at the end of each session.

The following section elaborates on the training tasks previously used in the literature and the results associated with each type of task.

2.3.1.3. Training tasks

According to Logan and Pruitt (1995), “the choice of training task and generalization task is critical in understanding the processes that underlie successful perceptual learning” (p. 354). In the training literature, most studies make use of discrimination and/or identification tasks to train learners’ perception and/or production of target language distinctions. Discrimination tasks entail discerning two or more target language (TL) stimuli separated by an interstimulus interval (ISI). In this task, two stimuli (produced by two different speakers) are heard by the listener, whose job is to decide if the two stimuli belong to the same L2 phonetic category or to two different L2 categories by answering “same” or “different” (Beddor & Gottfried, 1995). The discrimination task used in the current study is an AX categorical discrimination task with a long (1.15ms) ISI (More details on ISI can be found in section 5.4.3.2). On the other hand, an identification task, which “is the most widely used method of stimulus presentation in speech perception research” (Logan & Pruitt, 1995, p. 357), consists of identifying a single stimulus according to a set of given options representing different categories. The task used in the present study is a Forced-choice (FC) task, which forces participants to label the stimuli heard as one of the different target language phonetic categories provided.

Even though early findings with stop consonants reveal the efficacy of discrimination (DIS) tasks in modifying learners’ categorical perception of these sounds (Carney, Widin & Viemeister, 1977; Pisoni, Aslin, Perey & Hennessy, 1982; McClaskey, Pisoni & Carrell, 1983; see section 2.3.3.2 for a more detailed review of training studies on consonants), the efficacy of identification (ID) training has been said to be superior to discrimination training as an L2 training tool (Jamieson & Moroson, 1986; Logan & Pruitt, 1995; Ellis, 2005, among others). The idea that ID is better than DIS first arose from the comparison of two studies, one that made use of an auditory discrimination task (Strange & Dittmann, 1984) and one that made use of an identification task (Jamieson & Moroson, 1986). However, these two canonical training studies consisted of low variability phonetic training (LVPT) regimes, as Strange and Dittmann (1984) used an auditory DIS task (as opposed to categorical multiple-talker DIS tasks). Therefore, concluding that ID is superior to DIS from these studies may not be applicable to HVPT. Moreover, these comparisons were made across studies, with different learner populations and on different speech contrasts (voicing contrast vs. place of articulation contrast). It is relevant to note that previous research has shown that a novel voicing contrast is learned with more ease than a novel place of articulation contrast (Werker & Tees, 1984; Strange, 1992), as there are phonological, phonetic and acoustic differences between these types of contrasts (Strange & Dittmann, 1984).

As discussed previously, few prior studies (Flege,1995b; Wayland & Li, 2008; Nozawa, 2015) and the current thesis actually address the comparison of the effects of ID and categorical DIS tasks incorporating highly variable stimuli in the same study, thus in a controlled manner. The characteristics and reported effectiveness of these two main types of perceptual tasks are reviewed in the remaining of this section.

Strange and Dittmann (1984) investigated the effects of an AX auditory discrimination training regime on the perception of /r/ and /l/ by Japanese learners of English, who were exposed to synthesised minimal pair contrasts (i.e. rock and lock) during 14-18 training sessions. The results showed that four out of seven participants improved their perception of the target sounds after training and partially generalized their learning to another synthesised minimal pair contrast. Moreover, an AX training regime generalized to both an ID task and a DIS oddity task. However, no generalization effects to natural stimuli were observed. Due to this lack of generalization, the authors concluded that the effect of training with an AX DIS method was limited.

Two years later, Jamieson and Moroson (1986) aimed at training Franco-Canadian speakers to better identify the distinction between the English voiced and voiceless dental fricatives in synthetic stimuli by means of an adaptive identification task with immediate feedback. The training regime consisted of twelve sessions and lasted ninety minutes in total. The results revealed that the participants not only improved their perceptual abilities, but also generalized their improvement to the identification of /ð /- /θ/ in natural stimuli.

Thus, identification was said to be more effective than AX auditory discrimination as it forces listeners to incorporate between-category variability in the formation of the novel phonetic categories (Jamieson & Moroson, 1986; Logan & Pruitt, 1995; Ellis, 2005). Logan et al. (1991) explained that identification tasks require learners to label stimuli based on their long-term memory representations and encourage them to classify stimuli into categories, while discrimination tasks simply involve comparing one stimulus with another. Additionally, studies have found positive generalization results with ID tasks and it has been concluded that identification tasks promote robust learning as it leads to generalization to novel stimuli (Strange & Dittmann, 1984; Logan & Pruitt, 1995). On the other hand, discrimination tasks are said to enhance sensitivity to within-category acoustic differences that are irrelevant to category formation (Jamieson & Moroson, 1986) and learners might rely on information stored in the sensory memory in order to perform an AX auditory discrimination task. However, it has been suggested that discrimination tasks could be more suitable early in learning when the basic dimensions of variability are being discovered (Logan & Pruitt, 1995). Moreover, Pisoni and Lively (1995) explain that both types of training can be used in order to improve different perceptual skills. While identification training improves an "acquired equivalence", discrimination training improves an "acquired distinctiveness" (p.445).

Strange (1992) explained that contrary to the auditory (AX) discrimination, a categorical discrimination task (CDT or DIS) might force listeners to ignore within-category acoustic variation while focusing on phonetically relevant distinctions that define category boundaries. Polka (1992) explained that a categorical discrimination procedure involves perceptual consistency, since the highly variable stimuli for each category might include a wide range of acoustic variants. Strange (1992) added that this perceptual consistency may encourage L2 categorization. In fact, categorical discrimination procedures have been shown to measure several of the same perceptual abilities as identification tasks (e.g., Flege, 2003; Højen & Flege, 2006). Despite measuring similar abilities, a few differences between a categorical DIS task and an ID task were pointed out by Iverson et al. (2012). According to the investigators, categorical DIS tasks might "tap into relatively low levels of phonetic processing that change little due to experience during adulthood (e.g., Iverson et al., 2003; Iverson, Ekanayake, Hamann, Sennema & Evans, 2008), rather than tapping into higher levels of phonological encoding that are akin to identification" (p.14).

In 1995, Flege assessed the efficacy of both procedures (ID and categorical DIS) in a single study, incorporating highly variable stimuli in both procedures. The study, which was the first one to directly compare both tasks in a controlled manner, aimed at training Mandarin learners of English to better identify the final /d/ and /t/ during seven sessions. One experimental group was trained by means of a two forced-choice identification task, another group received the input through a categorical (or categorial) discrimination task and a control group was untrained. Post-training identification scores confirmed that both experimental groups outperformed the controls on the performance with the trained target sounds, they both showed generalization of knowledge and similar long term effects (see section 2.3.2 for more details on generalization and retention of training effects). These findings confirmed the efficacy and robustness of both training methods and challenged the general claim that ID training is superior to same/different discrimination training (Flege, 1995b).

Further, Wayland and Li (2008) trained native Chinese and native English listeners to discriminate Thai contrasts by means of ID and DIS tasks. The findings revealed that both ID and DIS training procedures were similarly effective in enhancing listeners' discrimination of Thai tone contrasts and that the native Chinese listeners outperformed the native English listeners. Thus, the authors concluded that both methods were equally effective in improving tone perception and that the prior experience with a tone language explained the Chinese participants' advantage over the NES when learning Thai tones.

A more recent small scale study (Nozawa, 2015) has provided evidence to the same effect when training Japanese leaners on coda nasals, as both ID and categorical ABX DIS methods promoted significant gains to a similar extent. Conversely, the investigator found that the same methods were not equally effective when training the same participants on vowel sounds, as the findings revealed that an ID method was superior to ABX DIS when training vowels. This study will be further described in section 2.3.3.1, since it sheds some light on the issue of cross-training also explored in this thesis.

Having argued that categorical DIS also promotes L2 learning (Flege, 1995b; Wayland & Li, 2008; Nozawa, 2015), it is important to acknowledge the effectiveness of the identification task, as it has been reported as a successful tool in several studies (e.g., Jamieson & Morosan, 1986; Bradlow et al., 1997; Hardison, 2003, 2005; Hazan et al., 2005; Nobre-Oliveira, 2007; among several others). Moreover, identification tasks usually promote robust learning (Logan & Pruitt, 1995), which can be evidenced by different measures of training efficacy, such as generalization, retention and/or transfer of perceptual learning to production. These measures are explained next.

2.3.2. Evaluating the effectiveness of training

The main goal of a training procedure is to enhance L2 learners' ability to perceive and/or produce the target sounds accurately. Successful gains in performance oscillate around 10%-20% (Bradlow et al., 1997; Flege, 1995b; Iverson et al., 2012) and the most common procedure to measure these gains is a pre-test/post-test design (Logan & Pruitt, 1995). Additionally, two other measures have commonly been used to evaluate the robustness of the improvement that comes from training: generalization and retention of knowledge over time. The former is usually assessed immediately after the post-test and it usually contains novel items, voices and/or contexts. Retention testing, on the other hand, usually takes place from one to several months after the completion of the training regime. These two additional training measures are further explained next.

2.3.2.1. Generalization: Robust learning

Generalization of learning is used to understand what the subject has learned during training, and, according to Logan and Pruitt (1995), generalization to novel stimuli is essential for a training regime to be considered effective. In their own words, "generalization is the ability to transfer the acquired knowledge to multiple dimensions (…) and if generalization occurs we can be more confident that robust learning has occurred" (p.371). Generalization often occurs when learners are exposed to highly variable stimuli rather than exposed to a single speaker (Lively, et al., 1991, 1993, see section 2.3.1.1 for further details).

Moreover, identification tasks have been said to promote more robust learning than discrimination tasks (Jamieson & Moroson, 1986; but c.f. Flege, 1995b; Wayland & Li, 2008).

Different types of generalization have been investigated and demonstrated in the existent training literature. The two most widely investigated types are transfer to novel talkers, i.e. new stimuli produced by talkers that are different from the ones used in training (e.g. Flege, 1995b; Iverson & Evans, 2009; Wang & Munro, 2004; Rato, 2014), and to new items, i.e. type of stimuli that are different from the ones used in training. The latter commonly consist of: a) new words and new sounds (e.g. Yamada et al., 1996; Wang, 2002; Wang & Munro, 2004; Hazan et al., 2005); b) generalization from natural to synthetic stimuli or vice-versa (e.g. Wang & Munro, 2004; Nobre-Oliveira, 2007; Iverson, Hazan & Bannister, 2005); and c) generalization from non-words to real words or vice versa (e.g. Nishi & Kewley-Port, 2007; Pierce, 2014, Hazan et al. 2005). In addition, generalization to new phonetic contexts (e.g. different place of articulation) is another dimension investigated in the training literature (Rochet, 1995, Thomson, 2012).

For instance, in a vowel training study, Rato (2014) tested generalization to physically different items from the ones used in training produced by novel talkers. The training regime aimed at improving the perception and production of three pairs of English vowel sounds (/ɪ-iː/, /ʊ-uː/, /e-æ/) by Portuguese advanced learners of English. Post-training identification tasks yielded significant perceptual and production gains in vowel learning of all target sounds and robust learning, tested by means of a generalization test, of two of the target vowel contrasts, namely (/ɪ-iː/, /e-æ/), for which generalization to new words produced by new talkers was found. Additionally, long term effects of learning were still observed two months after the post-test.

In a different study assessing stimuli generalization effects, Pierce (2014) trained Chinese and Korean advanced ESL learners on the perception of L2 vowels through different paradigms. The training stimuli were made of non-words and testing assessed whether the training effects generalized to the perception of real word stimuli. The Chinese and Korean participants who were trained by multiple-speaker paradigms (HVPT) transferred their ability to identify the target vowel contrasts in non-words to real-word minimal pairs. Conversely, the participants trained by means of a single speaker (LVPT) were not found to generalize the learning outcomes to real words. The author concluded that the use of identification training with multiple speakers promoted this generalization.

Generalization to new phonetic contexts has also been reported. For instance, after training Mandarin learners on the perception of French stops with synthesized tokens of /bu/ and /pu/, Rochet (1995) observed that learners were better able to identify the stops /b/ and /p/ followed by new vowels (/a/ and /iː/) and the novel voiceless stops (/t/ and /k/) followed by /u/. Thomson (2012) also assessed transfer to a new phonetic context in a study in which 26 Mandarin learners were trained to improve their perception of 10 English vowels produced in a post labial stop context. After eight short sessions with modified stimuli, leaners' ability to identify the English vowels improved significantly for the trained phonetic context and also generalized to a new one (velar stop context). Moreover, the improvement obtained after training was retained for a period of a month after training completion.

Another commonly assessed type of generalization is task generalization (e.g. production training procedures that generalize to perception and/or vice versa). A number of previous studies have shown that production procedures positively influence both perception and production of non-native sounds (Catford & Pisoni, 1970; Mathews, 1997; Lacabex, García-Lecumberri & Cooke, 2008; Aliaga-García, 2009; Kartushina, Hervais-Adelman, Frauenfelder & Golestani, 2015). However, Hattori and Iverson (2008) and Hattori and

Iverson (2009) found that production training was only effective at the level of production and not at the level of perception when training Japanese learners of English on the /r-l/ distinction. As regards to the effect of perceptual training procedures that generalize to production, several recent studies investigated this effect, since "perception training is easier to administer" than production training (Rochet, 1995, p. 396) and found different degrees of success. Some perceptual training regimes did not result in transfer to production (e.g., Hwang & Lee, 2015), while others showed either partial or full generalization to production (Hazan et al., 2005; Aliaga-García & Mora, 2009; Lacabex & García-Lecumberri, 2010; Lambacher et al, 2005; Lengeris, 2008; Iverson et al, 2012; Thomson, 2011; Thomson & Derwin, 2014), showing that perceptual training might enhance production without specific production training. This type of generalization, which is one of the foci of the present study, will be further discussed in section 2.3.4.

2.3.2.2. Retention of learning: Long-term effects of training

Similarly to generalization results, testing the long-term effects of training might shed some light on whether robust L2 categories have been established in the L2 learners' perceptual space (Flege, 1995b). Some studies provide empirical evidence to that effect. For instance, Lively et al. (1994) assessed the long-term effects of a HVPT training approach that aimed at training Japanese learners of English to perceive the /r-l/ distinction. The results revealed that the high-variability training regime promoted long-term modification of speakers' phonetic perception that remained for six months after the completion of the training. Bradlow, Yamada, Pisoni and Tohkura (1999) succeeded in improving the perception and production of English /r/ and /l/ by Japanese L2 learners of English after a HVPT identification procedure that lasted 20-30 minutes in each session, and totalled 15 hours. A delayed test illustrated that the effect of training persisted three months after training.

Several more recent studies have also reported successful retention of learning after periods of time ranging from one to twelve months. For instance, retention of L2 vowel learning has already been observed one month (Thomson, 2012) and two months after training ended (Rato, 2014). Other studies reported that the degree of perceptual accuracy was maintained for over three months (Wang, 2002; Wang & Munro, 2004; Nishi & Kewley-Port, 2007), six months (Wang et al., 1999) and over a year (Cenoz & García-Lecumberri, 1999). The outcomes of the studies described in this section provided empirical evidence that perceptual training is efficient in promoting long term effects, which confirms the robustness of the training procedure and the relevance of phonetic training as an L2 teaching tool (Logan & Pruitt, 1995).

2.3.3. Perceptual training studies

A fair amount of perceptual training studies have been carried out looking at different L1-L2 language combinations, different target structures and methodologies, and reporting different degrees of success. Training studies have focused on different aspects of the L2, i.e. consonants and/or vowels (see a detailed review below), stress and intonation (Schwab & Llisterri, 2011, 2014; Wang, Spence, Jongman & Sereno, 1999; Wang, Jongman & Sereno, 2003; Wayland & Li, 2008, Hardison, 2004) and rhythmic and gestural patterns (Hirata & Kelly, 2010; Kushch, Gluhareva, Llanes & Prieto, 2016). In addition, studies have investigated various methodological issues, such as the use of synthetic vs. natural stimuli (Nobre-Oliveira, 2007, Iverson et al., 2005), the inclusion of real vs. non-word stimuli (Nishi & Kewley-Port, 2007; Pierce, 2014, Hazan et al., 2005; Thomson & Derwin, 2016), the number of target sounds to include in training (Nishi & Kewley-Port, 2007, 2008;

Lerdpaisalwong, 2015), the contribution of visual and audio-visual cues (Hardison, 1999; 2003; Hazan et al., 2005; Wang, Behne & Jiang, 2008; Aliaga-García, 2009; Pereira, 2014) and the role of attention and explicit instruction (Pederson & Guion-Anderson, 2010; Nozawa, 2015, Alves & Luchini, 2016).

Taken together, all studies cited above (among many others) provide empirical evidence that L2 learners' perception and production might be modified as a result of laboratory training, provided highly variable stimuli and appropriate tasks are applied. As the present study investigates the cross training effects (i.e. attended vs. unattended stimuli) on L2 consonant and vowel perception by Catalan learners of English, perceptual training studies dealing with the role of attention will be presented first, followed by a review of training studies on consonants and/or vowels.

2.3.3.1. Perceptual training studies on cross-training effect: Training attention

One of the goals of this thesis is to assess if training benefits generalize to other sounds present in the stimuli but not specifically trained. In order to investigate this, two groups of learners were trained on vowels and two other groups were trained on consonants. All groups were trained using the exact same stimuli and were tested on both consonants and vowels (see section 5.2 for experimental design details). This possible training effect on untrained or unattended segments present in the stimuli will be referred to as "cross-training" throughout this thesis. There is little previous work on this issue in the phonological domain and this effect has previously been referred to in the literature as "training attention" (Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010).

In the fields of cognitive science and SLA, several studies have addressed the issue of attention. Attention is the cognitive process that allows learners to select and focus on specific stimuli in the input, while ignoring others. According to James (1890, as cited in

Schmidt, 2001), attention is "the taking possession by the mind, in clear and vivid form, of one out of what seems several simultaneously possible objects or trains of thought" (p. 403). Schmidt (2001) adds that attention's core characteristics are "focalization and concentration of consciousness" (p. 13). In second language acquisition, investigators evaluate what kind of information learners pay attention to in the L2 input, as it has been reported that not all stimuli are attended to (Broadbent, 1958). In fact, according to the *filter model of attention* proposed by Broadbent (1958), stimuli undergo a pre-selection before being attended to. Van Patten (1996) explains that the leaners' focal attention is responsible for selecting solely the input which is considered important for further processing. This selection occurs since the attentional system is said to be limited in its nature (Mc Laughin, 1987; Petersen & Posner, 2012; Wager, Jonides & Smith, 2006; Christensen & Humes, 1997). Importantly, paying attention has been found to be a pre-requisite for learning to take place (Schmidt, 2001; Leow, 2012), as "unattended stimuli persist in immediate short-term memory for only a few seconds at best, and attention is the necessary and sufficient condition for long-term memory storage to occur" (Schmidt, 2001:20). Moreover, attention is said to occur either voluntarily and/or passively (Neumann, 1996; Schmidt, 1990, 2001). In fact, Schmidt (2001) explains that while incidental learning might take place, paying attention is facilitative and indispensable. In the author's own words: "people learn about the things that they attend to and do not learn much about the things they do not attend to (p. 30)".

Regarding the role of attention in L2 sound perception, Pederson and Guion-Anderson (2010, p. 54) suggest that "accurate perception of at least some non-native phonemes requires reorienting of attention" and thus attention plays an important role in phonetic attainment. Attention orienting is described as "the aligning of attention with a source of sensory input or an internal semantic structure stored in memory" (Posner, 1980). In the SLA literature, researchers have attempted to manipulate attention orienting through three different methods:

a) input enhancement (Smith, 1993; Van Patten, 1990, 2002), b) cue enhancement (Jamieson & Moroson, 1986; Iverson et al., 2003; Iverson et al., 2005; Yamada & Tohkura, 1992), and c) instruction manipulation (Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010; Nozawa, 2015; Alves & Luchini, 2016). As the latter case is the most closely related with the present investigation, a few studies that orient attention through instruction will be reviewed next.

In a training study, Pederson and Guion-Anderson (2010) tested if directed attention promoted learning of phonetic information. The authors controlled participants' attention by training two groups (N=21) of English monolingual speakers on Hindi vowels and initial stop consonants with the exact same stimuli (i.e. 27 monosyllabic words beginning with one of the target stops and containing one of the target vowels). However, while one group was told to attend to vowels present in the stimuli, the other group was instructed to attend to consonants present in the same stimuli. Both groups of learners were trained by means of an identification training regime with immediate feedback on the attended segment. Testing consisted of a discrimination task, which assessed learners' perception of the attended and unattended segments, that is, vowels and consonants for both groups. Their findings showed that identification training was successful in improving learners' perceptual abilities on the trained/attended sounds; however, it did not promote improvement with the untrained/unattended segments. The authors concluded that "orienting attention during phonetic training facilitates learning of the specific class of stimuli to which the participants are instructed to attend" (p. 58).

Expanding on the topic of training attention, a recent small-scale study (Nozawa, 2015) trained four different groups (N=7) of Japanese learners of English on vowels and nasal sounds in coda position with the exact same stimuli. Testing and training stimuli involved words containing seven American English (GA) vowel sounds embedded in /bVb/,

/bVd/, bVg/, /bVm/, /bVn/ and /bVŋ/ contexts. Two groups were trained on vowel perception, one by means of an ID task and the other by means of a categorical AXB task. The other two groups were trained on nasal consonant perception, also by means of an ID task and a categorical AXB discrimination task. Testing assessed the participants' perception of both vowels and nasals, independently of the focus of their training. In other words, the study assessed if vowel training affected the identification of the nasal consonants present in the stimuli, and if training on nasals affected the identification of the vowels present in the stimuli. The stops /b,d,g/ were not the focus of the study and were simply used to test vowel perception embedded in CVC words. Interestingly, results patterned differently for vowels and consonants. The identification-trained group improved on vowel perception, outperforming the other experimental group and suggesting that ID training is superior to ABX DIS training when training L2 vowels. However, DIS was as effective as ID when training word final nasal sounds. Furthermore, the two vowel-oriented groups showed a tendency of improvement with the untrained sounds, as the results yielded a non-significant but enhanced nasal identification after training, suggesting that "exposure to stimuli might be able to enhance listeners' sensitivity even when their attention is not on the target sound" (p. 4).

Another recent HVPT study assessed whether explicit instruction had an impact on the effectiveness of a HVPT procedure (Alves & Luchini, 2016). The authors tested the perception and production of English word-initial voiceless stops by 24 Spanish learners of English (L1=Argentine Spanish). Participants were divided into two experimental groups and a control group. Both experimental groups received the exact same perceptual ID training making use of the exact same stimuli; however, one group was explicitly informed about the target sounds under investigation, whereas the other group was not aware of which phonological aspect they should attend to during training. Both groups showed enhanced stop consonant perception after the training regime, but partial improvement in production was observed only with the group in which learners were told to attend to the stop consonants under investigation (/p/ improved but not /t/ and /k/).

The findings in Alves and Luchini (2016) suggest that instruction, awareness of trained sounds and attention might play an important role in phonetic training studies and corroborate previous findings which suggested that "training attention" is crucial in phonetic learning, at least when training involves only ID tasks (Pederson & Guion-Anderson, 2010). However, Nozawa (2015) presented some evidence that so long as sufficient and balanced exposure to the target sounds was provided, both attended and unattended conditions could be numerically enhanced through ID and DIS training. The numerical, albeit non-significant, improvement found by Nozawa (2015) was attributed to the small sample size of the study. This thesis explores this issue further by assessing the effects of two different training methods (Categorical DIS and ID) on the perception of attended and unattended segments (including vowels and stop consonants in initial and final position) by Spanish/Catalan learners of English. Having reviewed the role of attention and orienting attention through explicit instruction in phonetic training studies and the impact they might have in the cross-trained segments, the next section aims at providing the reader with an overview of previous training studies carried out on consonant sounds.

2.3.3.2. Perceptual training studies on consonants

A significant part of the L2 training literature has been devoted to training Japanese speakers to perceive the challenging English contrast /r/ and /l/ (Bradlow, 2008, Bradlow et al., 1997; Iverson et al., 2005; Lively et al., 1993, 1994; McCandliss, Fiez, Protopapas, Conway & McClelland, 2002; Pisoni et al., 1994; Protopapas & Calhoun, 2000, among many others). The need to train this distinction arises since both non-native categories are assimilated to a single L1 category (Strange, 1992; Best, 1995; Best & Tyler, 2007) and, according to Iverson et al. (2005), the acquisition of /r/ and /l/ by Japanese learners of English exemplifies the canonical example of difficult L2 contrast learning.

The current study examines the perception and production of the English stop voicing contrast in initial and final position and previous studies on initial voicing contrasts report beneficial effects event after short training periods, showing that not all novel phoneme contrasts require the same level of effort from L2 learners. In fact, the learnability and trainability of initial voicing contrasts have been said to be relatively easy in comparison to other distinctions (Strange & Dittmann, 1984), as they are psychoacoustically robust (Burnham, 1986) and most L2 learners have some experience with VOT categorization from their L1 (Pisoni et al., 1982). Strange and Dittmann (1984) add that “voicing distinctions (…) are very common”, and that “there may be a correlation between ease of perceptual differentiability and usage in languages (p. 142)”.

According to Flege (1989) learning success as a result of training studies greatly depends on the auditory nature of the feature which is being trained, as some non-native contrasts are auditorily easier to detect than others, such as initial voicing distinctions. The author also adds that contrasts that are not existent in the native language, but which are easily distinguishable, are acquired easily. On the same matter, Munro and Derwing (2008) explain that some English sounds appear to be effortlessly acquired by L2 learners as they may simply be easy to perceive and produce. These premises go in line with the early training results found by Carney et al. (1977), who report that English native listeners could modify initial stop VOT discrimination and identification after a simple discrimination training regime (12 sessions). Moreover, Pisoni et al. (1982) reported rapid improvement, up to 85% accuracy, after a single training session (1 hour) on differentiating non-native initial stop consonants differing in VOT. McClaskey et al. (1983) trained English speakers to discriminate synthesized stops differing in VOT (ranging from – 70 to +70ms). Discrimination training proved effective to enhance learners' discrimination of voicing distinctions and to generalize learning to places of articulation that were not present in the training. The trainability of stop consonants was also investigated by Werker and Tees (1984), who trained thirty English speakers to perceive two non-native contrasts (the Hindi dental vs. retroflex stops, and the voiceless vs. voiced dental aspirated stops) by means of natural stimuli and continuous feedback. The researchers found that perceptual training enhanced only the perception of the voicing contrast, as the perception of the place contrast remained unaltered. These findings might suggest that voicing contrasts are more easily altered than place contrasts, and thus the latter might require a different length of training and/or different techniques. A few more recent studies also confirmed that VOT perception can be modified with ease (Aliaga-García & Mora, 2009; Collet et al., 2013; Everitt, 2015; Alves & Luchini, 2016). For instance, Collet et al. (2013) observed changes in the perception of English stops by French leaners of English after a short-term identification training regime. The results revealed enhanced L2 voicing identification and discrimination, as well as categorical perception.

The effects of training in final voicing distinctions have also been investigated. For instance, Trapp and Bohn (2000) observed enhanced perception of English word-final /s/-/z/ by Danish learners of English and transfer of learning to novel phonetic environments and unfamiliar voices as a result of four sessions of identification training that included highly variable stimuli. Specifically looking at the trainability of final stops, previous findings reveal that modifying final consonant perception tends to be more challenging than modifying initial consonants, in line with the increased difficulty in perceiving coda consonants rather than consonants in the onset position (Jang, 2014). Flege (1989) investigated the learning of

English word final unreleased /t/ and /d/ by three groups of Chinese learners of English as an L2. The participants were grouped according to their L1 (Mandarin, Taiwanese and Shangainese) and were assigned to either of the two training experiments that will be explained next. In the first one, the participants were trained to identify four different minimal pairs, which resulted in a non-significant learning gain from pre-test to post-test (9%). In the follow-up experiment, 16 different Chinese L2 learners of English were exposed to only two different minimal pairs with more repetitions, which yielded a slightly larger and significant amount of learning gain from pre-test to post-test (11%) and generalization of learning to the other two untrained minimal pairs. The author concluded that exposure to fewer L2 contrasts with an increased amount of trials is more beneficial to promote learning. That study provided empirical evidence for another important finding in L2 speech training research, namely the influence of the L1 on L2 performance. Flege (1989) observed that the Chinese learners whose L1 allows for word-final obstruents (Taiwanese and Shangainese) were able to benefit more from training than those learners whose native language (Mandarin) does not permit final stops. The author hypothesized that perceptual training in cross-language speech perception might be more effective for L2 features that are already present in the leaners' L1.

Along the same lines, the learnability of the voicing cue might be connected with the existence of this feature in the L1, even if in a different position of the word (Flege, Munro & Skelton, 1992; Flege, Munro & MacKay, 1995; Flege, 1998). Empirical evidence from these studies shows that advanced learners of English whose L1 has obstruent voicing as a cue in positions other than final (Spanish and Italian) performed better in final consonant identification than learners whose L1 does not have obstruent voicing as a cue (Korean and Mandarin).

In a subsequent training study, Flege (1995b) demonstrated the effectiveness of two HVPT tasks, a categorical discrimination task and a 2FC (forced-choice) identification task, on training the same unreleased /d/-/t/ final stop contrast for 7 sessions over 3 months. Both groups significantly improved their identification performance after training. The length of the training used here (7 sessions) and the lack of success in the first experiment carried out with a more extensive number of stimuli (Flege, 1989) suggest that training final consonants might differ from training initial consonants in that it may require a longer training period.

This section has reviewed a few previous perceptual training studies carried on consonants, in different word positions. Findings with initial consonants point to the fact that the perception of VOT distinctions might be easily modified with perceptual training (Carney et al., 1977; Aliaga-García & Mora, 2009; Collet et al., 2013; Everitt, 2015; Alves & Luchini, 2016). Conversely, training final stop consonant perception might require more time and effort (Flege, 1995b), especially if the voicing contrast does not occur in the learners' native language (Flege, 1989) or it is not prevalent in the L1 (Flege et al., 1992; Flege et al.,1995; Flege, 1998). The following section will introduce perceptual training studies carried out on vowels, shedding some light on the different methods that have been found to be successful in attaining vowel learning.

2.3.3.3. Perceptual training studies on vowels

Vowels are said to be generally perceived less categorically than consonants (Fry, Abramson, Eimas & Liberman, 1962; Pisoni, 1973) and thus pose more challenges to L2 learners (Pereira, 2014). The learnability of vowels through laboratory training has been investigated extensively in the last few decades (Aliaga-García & Mora, 2009; Cebrian & Carlet, 2014; Cenoz & García-Lecumberri, 1999; Iverson & Evans, 2007; Iverson & Evans, 2009, 2011; Lacabex et al., 2008; Lengeris, 2008; Nishi & Kewley-Port, 2007, 2008; Nobre-Oliveira, 2007; Rato, 2014; Thomson, 2011, 2012; Wang & Munro, 2004; among many others) and according to Nishi and Kewley-Port (2007, p. 1497), the "differences between vowels and consonants indicate that vowel training may require somewhat different protocols from those frequently used in consonant training". The studies that are most relevant and connected to the current investigation are reviewed next, aiming at shedding some light on different types of factors that may affect the efficacy of a training regime and main findings found in the extant literature. Studies that reflect on stimuli considerations (functional load and stimuli presentation) will be reviewed first, followed by studies that elaborate on L1 differences (vowel inventory size/complexity). The section ends with a review of studies that investigate learner characteristics that may play an important role in the outcome of the training studies (individual differences and proficiency level).

In a L2 vowel training study that took into account the notion of functional load of specific sounds, Wang and Munro (1999) trained Mandarin L2 learners of English to identify English /i-ɪ/ and /uː-ʊ/ sounds in a single training session with synthesized stimuli and immediate feedback. The investigators found better performance with the high front vowel pair /i-ɪ/, which was accounted for by the result of the instruction received but also attributed to the greater functional load of /i-ɪ/ when compared to /uː-ʊ/ (i.e., many words are distinguished by /i-ɪ/, while there are very few pairs that are distinguished by /uː-ʊ/).

Training efficacy has also been related to the number of target sounds included in the training. Nishi and Kewley-Port (2007) explored the effects of training set sizes by dividing their Japanese participants into two experimental groups. A "full set" training group was exposed to nine English vowels during training, whereas a "subset" training group was exposed to three English vowels that pose challenges to these learners. Despite the positive results found for both groups in identification performance and generalization to new items and talkers, the group trained on fewer vowels did not show transferability to untrained sounds. Thus, the use of full set vowel training proved more effective as trainees were exposed to various acoustic cues within the training set and were provided with a richer learning opportunity. In a subsequent study (Nishi & Kewley-Port, 2008), the investigators tested whether exposing learners to a combination of “full set” and “subset” training would be more beneficial than the optimal exposure to “full set” training only that had previously been observed. Post training scores revealed enhanced L2 perception for all participants, independent of the training set size they were exposed to and no significant difference between the performances of the two given groups.

Turning our attention to studies that look at how L1-L2 differences in vowel inventory relate to L2 training, it is relevant to recall that previous research in second language perception has shown that a more complex L1 vowel inventory enhances learners’ ability to attend to cues in a native-like way when perceiving English L2 vowels (Hacquard, Walter & Marantz, 2007; Bohn, 1995; Fox, Flege & Munro, 1995; Kivistö-de Souza & Carlet, 2014). This effect was supported by the findings of two vowel training studies (Iverson & Evans, 2007; 2009). The former (Iverson & Evans, 2007) examined if L2 learners with different first-language vowel systems (Spanish, French, German and Norwegian) learn the English vowel system similarly as a result of training. Results showed that learners with a more complex L1 vowel inventory (German and Norwegian) outperformed the learners with a simpler L1 vowel space (Spanish and French), suggesting that having a larger and more complex L1 vowel inventory might be beneficial in promoting more accurate L2 vowel perception. Nonetheless, the investigators observed a remarkable degree of consistency in the way L1 vowels were perceived across groups of learners, as “all groups used formant movement and duration to recognize vowels, and learned new aspects of the English vowel system rather than simply assimilating vowels into first-language categories” (p. 2842). In a subsequent study, Iverson and Evans (2009) exposed two groups of learners differing in their L1s (Spanish vs. German) to five high-variability training sessions on English vowels. In line with the previous study, German learners outperformed the Spanish learners in the identification of non-native vowel sounds due to their more complex L1 vowel space. In a follow-up experiment, the Spanish learners were exposed to ten additional training sessions, which afforded similar training gains for the Spanish listeners as the German learners. This study showed that different learners (with different L1s) require different amounts of training sessions in order to attain L2 vowel learning.

Further, individual differences such as differences in level of proficiency or amount of training obtained have also been found to affect the effects of vowel training. For instance, Wang and Munro (2004) trained native speakers of Mandarin and Cantonese on the three English vowel contrasts (/i-ɪ/, /e-æ/, /uː-ʊ/). In general, positive evidence for L2 perceptual learning, generalization and long-term retention was observed. However, findings revealed that different learners required different amounts of time to reach perceptual attainment, in line with the results reported by Iverson and Evans (2009). Moreover, the investigators concluded that computer-based training is an efficient means to provide learners with a more flexible and personalized schedule during the training period. Expanding on the topic of individual differences, Aliaga-García, Mora and Cerviño-Povedano (2011) investigated the relationship between phonological short term memory (PSTM) and L2 perceptual abilities during an L2 vowel training regime that lasted 10 sessions. The 60 Catalan/Spanish participants were allocated to either low or high PSTM capacity groups. Findings revealed that the high PSTM capacity learners outperformed the low PSTM capacity learners, suggesting that PSTM might play a role in L2 speech learning and categorization.

In another study that looked into proficiency differences, Iverson et al. (2012) explored whether training was equally effective for French learners of English differing in proficiency level and amount of exposure. A group of beginner learners in their home country and a group of mid-intermediate learners living in the UK underwent a vowel identification training regime and were tested on the identification, discrimination and production of 14 English vowels and diphthongs. Both groups showed a slight effect of training on discrimination ability, as well as significantly improved their identification and production as a result of training, indicating that amount of experience is not an underlying factor for improvement to take place. Previous research carried out in naturalistic settings defined the notion of threshold level (Collentine, 2009), which suggests that learners at a lower level might make greater progress than their higher level counterparts.

The studies and findings reviewed above discussed the trainability of vowels, which is one of the objectives of the current study. Different types of factors that may influence the outcome of HVPT have been identified, including methodological factors, L1 vowel inventory characteristics and individual learner differences. Next section presents an overview of perceptual training studies that deal with vowels and consonants in a controlled manner (i.e. in the same study). The section starts by reviewing previous studies that involved Spanish/Catalan learners, since they are the target population of the present study.

2.3.3.4. Perceptual training studies on vowels and consonants

Previous research points to the fact that vowels and consonants are perceived differently (Fry et al., 1962; Pisoni, 1973; Strange, 2007), involve different levels of processing in order to be learned (Pisoni, 1973), activate different neuronal patterns when processed (Carreiras & Price, 2008; Engineer, Rahebi, Buell, Fink & Kilgard, 2015) and thus might require different types of perceptual training (Nishi & Kewley-Port, 2007, Nozawa, 2015). According to Pereira (2014), the learning of consonants and vowels require different degrees of effort from L2 learners. In the author's words, this is because:

> "Consonants tend to be more stable than vowels in their acoustical properties (Liberman, Cooper, Shankweiler & Studdert-Kennedy, 1967) and therefore they may be easier to perceive. Vowels, on the other hand, tend to have higher intensity and longer duration than consonants, less vocal tract constriction and they are all voiced sounds (Best, Halle, Bohn & Faber, 2003; Knight, 2011). Vowels may present more within-stimulus variability as they may differ in pitch, loudness and quality (Ladefoged, 2006) and due to their central place in the syllable, they tend be articulated between consonants and influenced by these neighbouring sounds (p. 30)".

A HVPT study that looked at both vowels and consonants was carried out by Aliaga-García and Mora in 2009. This study aimed at investigating the effect of six two-hour mixed-methods phonetic training sessions on the perception and production of the initial English stops (/p-b/ and /t-d/) and four English vowels (/æ-ʌ/ and /ɪ-iː/) by a group Catalan/Spanish native speakers. The tasks included articulatory explanations, perception tasks and practice with production based software. As for the consonants, training was effective in modifying learners' degree of categoriality in the identification of two of the target stops, namely /p/ and /b/, significantly. Moreover, a significant modification towards longer VOT production for the voiced bilabial stop /p/ and a marginally significant effect for the alveolar voiced stop /t/ were found. Feedback identification training on vowel perception resulted in significantly enhanced discrimination scores for all target sounds; however, no improvement on vowel production accuracy, as measured through F1 and F2 values, was observed. The findings of this study suggest that the phonetic training regime promoted pronunciation gains differently for vowels and consonants. The authors speculated that the effect of phonetic training may "differ greatly according to phonetic dimension and sound contrast" (p. 24).

In a different study that dealt with vowels and consonants, Cebrian and Carlet (2014) assessed the effect of a three-week HVPT regime on the perception of four English consonant sounds (/v/-/b/ and /d/-/ð/) and two vowel pairs (/iː/-/ɪ/and /æ/-/ʌ/) by Catalan-Spanish advanced learners of English. The results pointed to a significant positive effect of a relatively short-term phonetic training method for a subset of the target consonants and vowels, namely /v/, /d/, /iː/, /ʌ/ and marginally for /b/. Interestingly, L1-L2 perceptual mapping was not the main predictor of perceptual success in this study, as the English /iː/, which has been reported to have a near-identical counterpart in Catalan (Cebrian, 2006, 2015; Cebrian, Mora & Aliaga-García, 2011), was the least successfully identified vowel at the outset of the study. On the other hand, a non-native sound such as /v/ was already relatively successfully identified at pre-test and improved around 10% due to training. The investigators suggested that different factors might have influenced the studies' outcome: the advanced learners' phonetic and metalinguistic knowledge in the case of /v/-/b/ and /æ/-/ʌ/, word frequency differences in the case of /d/-/ð/, and the influence of vowel duration in the /iː/-/ɪ/ distinction. Specifically, the tense vowel was more successfully identified when preceding a voiced consonant than a voiceless consonant, and the opposite held true for the lax vowel /ɪ/, showing an over-reliance on temporal cues, in line with previous perceptual studies (Aliaga-Garcia & Mora, 2009; Cebrian, 2006, 2007; Kivistö-de Souza & Carlet, 2014). Thus, the findings in Cebrian and Carlet (2014) add to the existing HVPT literature by providing empirical evidence that the learnability of different target sounds is not affected homogeneously.

Still, some similarities between training vowels and training consonants have been reported. For instance, Lerdpaisalwong (2015) explored if training set size had an effect on consonant training. Expanding on Nishi and Kewley-Port's (2007, 2008) findings that training on a full set of vowels is more beneficial than training on a subset of vowels, Lerdpaisalwong trained two groups of Thai EFL learners on vowels and consonants with different set training sizes. The results corroborated the previous findings on vowels and provided empirical evidence that a full set training is also more beneficial in order to train consonant sounds. Lerdpaisalwong suggests that while vowels and consonants differ in many respects, as far as set training size exposure is concerned, there is a relationship between the acquisition of L2 vowels and consonants.

Further exploring the relationship between vowel and consonant acquisition, Lee and Hwang (2016) investigated the effects of perceptual training on Korean early learners of English (aged 11-12). The stimuli were made of highly variable large sets of English vowel and consonant distinctions (covering the entire vowel and consonant systems of English). Results revealed that consonants were initially better perceived and obtained greater improvement than vowels. Interestingly, greater improvements were observed with the poorly identified sounds at pre-test than with the well identified ones, since there was more room for improvement. This goes in line with previous findings in the training literature; for instance, Pereira (2014, p. 128) explains that "having more room for improvement plays a role in the amount of benefit the learners can obtain from a perceptual training programme".

Further studies contrasting vowel and consonant training are needed to understand better which methods are more effective with each type of sound. For instance, the heated discussion about the effectiveness of training tasks (ID vs. DIS) may not possibly be settled without considering that the learning of consonants and vowels does not entail the same degree of effort from L2 learners (Pereira, 2014). Previous findings (Nozawa, 2015) suggest that while an ID task is superior for training L2 vowels, both ID and DIS promote consonant learning to the same extent. All in all, the results of the studies described in this section provide empirical evidence to the fact that despite some similarities between vowel and consonant learning (Lerdpaisalwong, 2015), these two segments might require different training procedures (Nishi & Kewley-Port, 2007; Nozawa, 2015). That is because vowels and consonants appear to be learned to different degrees, consonants being acquired with more ease than vowels overall in a few previous studies (Cebrian & Carlet, 2014; Lee & Hwang, 2016; Aliaga-García & Mora, 2009; Pisoni, 1973; Pereira, 2014).

2.3.4. Effects of perceptual training on L2 production

The link between perception and production in L2 speech learning has been investigated in a number of studies (Bohn & Flege, 1990, 1992, 1997; Flege et al., 1997; Sheldon & Strange, 1982) and this relationship is said to be complex in nature (Sheldon & Strange, 1982; Llisterri, 1995; Kluge, Rauber, Reis & Bion, 2007; Baker & Trofimovich, 2005). Three divergent views are found in the literature: some studies indicate that perceptual learning is a pre-requisite to production learning (Flege, 1984; Flege & Hillenbrand, 1984; Flege, 1995), some provide evidence for a simultaneous development of the two skills (Best, 1995; Fowler, 1996; Rochet, 1995; Bradlow et al., 1997; Hazan et al., 2005; Thomson, 2008), and other studies propose that production changes might take place prior to perceptual changes (Flege & Eefting, 1987; Sheldon & Strange, 1982; Llisterri, 1995; Kluge et al. 2007; Hattori & Iverson, 2010). Since the current thesis investigates the effect of perceptual training on production, it explores the assumption that perception precedes and aids the acquisition of L2 sound production (Flege, 1995) and it investigates further the link between the two skills.

Numerous researchers have investigated whether perception training is effective in improving L2 production (e.g. Yamada et al. 1996; Flege, 1989; Bradlow et al., 1997; Hardison, 2003; Rochet, 1995; Hazan et al. 2005; Lambacher et al., 2005; Lengeris, 2008; Iverson et al., 2012; Thomson, 2011; Pereira, 2014; Rato, 2014, among others) and have found different degrees of success when training different segments. For example, as regards to production of the canonical /r-l/ distinction, Bradlow et al. (1997) observed that teaching perceptive skills to learners using multiple talkers, natural stimuli and identification tests improved both the perception and the production of the English /r-l/ distinction by Japanese learners of English. In a subsequent study, following the same testing and high-variability training procedures used in Bradlow et al. (1997), Bradlow et al. (1999) explored the long-term effects of training on the perception and production of the /r-l/ contrast by Japanese learners of English. The results revealed that learning persisted three months after the completion of the training both in production and in perception.

Perceptual training studies dealing with stop consonant learning, however, have reported partial generalization effects to production (Rochet, 1995; Reis & Nobre-Oliveira, 2007; Alves & Luchini, 2016; Aliaga-García & Mora, 2009). For instance, Rochet (1995) observed production improvement for English voiced and voiceless stops as a result of perceptual training. However, the production improvement for voiceless stops was more robust than for the voiced counterparts. Reis and Nobre-Oliveira (2007) also reported partial perceptual training effects in the production of stops by their Brazilian Portuguese learners of English. Only two of the three trained English voiceless stops improved, namely /p/ and /t/. Similarly, Aliaga-García and Mora (2009) found partial generalization to production in a study which aimed at training Catalan learners of English to perceive initial stop consonants /p-b/ and /t-d/ by using a variety of training methods, including articulatory descriptions and perceptual tasks. Interestingly, production gains were only found for the sound that obtained the lowest scores at pre-test (/p/) and, despite the improvement, learners still did not reach native speaker values.

Turning our attention to vowel segments, studies also report mixed results as regards to generalization from perceptual training to production gains. For instance, Aliaga-García and Mora (2009) also found that perceptual training only enhanced the discriminability of the target vowel sounds (/æ/ - /ʌ/, /ɪ/ - /iː/). Production improvement, measured through F1 and F2 frequency values, was not observed after training. In another vowel study, Iverson et al. (2012) perceptually trained two groups of L2 learners on 14 English vowel sounds and assessed both their perceptual and production abilities. Partial production improvement was observed (only three vowel sounds) for both groups of learners, regardless of their proficiency level.

Further empirical evidence of partial transfer to production was observed by Rato and Rauber (2015), who examined whether high variability perceptual training can modify L2 vowel production abilities of European Portuguese (EP) native speakers. Learners were trained on three English vowel contrasts (/ɪ-iː/, /ʊ-uː/, /e-æ/) by means of a combination of identification and discrimination tasks. Production data was elicited from participants by means of a sentence reading task at three different times, pre-test, post-test and delayed post-test. Participants' vowel quality, as measured through spectral Euclidean distances and duration ratios, was significantly modified for a subset of the trained vowel sounds after training, confirming a significant effect on pronunciation accuracy for some target sounds as a result of perceptual training.

In a vowel HVPT study with stimuli manipulation, Nobre-Oliveira (2007) assessed the effect of training with cue enhancement on the perception and production of six L2 English vowels sounds by two groups of native speakers of Brazilian Portuguese. One group was trained with synthetic stimuli (i.e., isolated vowels with F1 and F2 enhancement) and the other group with natural stimuli. A combination of identification and discrimination tasks with immediate feedback were used to expose participants to the target language contrasts. Results yielded production improvement of the target structures only for the group trained with synthetic stimuli, showing that cue enhancement is beneficial when promoting generalization to production.

A more recent study investigating the relationship between perception and production abilities and the type of stimuli used in training was conducted by Thomson and Derwin (2016). The authors assessed whether English L2 learners obtain better results in production when exposed to perceptual vowel training that made use of non-words (i.e. vowels in isolated open syllables) or real words. Mixed L1 learners of English as a second language were divided into two experimental groups (non-word group vs. real word group) and an untrained control group. All learners were tested on their production of English vowel sounds before and after the 40 sessions of perceptual training and recordings were judged by native English speakers. The results revealed that the group trained with nonsense words outperformed the group exposed to real words in L2 vowel production, even though the words used for testing and training were the same for the latter group (real words). The investigators concluded that in order to improve learners' production skills, L2 vowel training should contain some focus on phonetic level information, as opposed to focusing exclusively on vowel sounds in real words.

Some recent training studies have contrasted perception and production training in the same study (Lacabex et al., 2009; Aliaga-García, 2010, Aliaga-García, 2016) revealing that both types of training produce positive learning effects on perceptual and production abilities to similar extents. For instance, Aliaga-García (2016) carried out a study which aimed at assessing the efficacy of two HVPT methods (perceptual auditory vs. imitation production-based) on the identification and articulation of the full set of English RP monophthongal vowels (/iː ɪ e ɜː a ʌ ɑː ʊ uː/) by Spanish and Catalan learners of English. Two groups of learners were trained by means of natural CVC words produced by 10 different talkers for five weeks. Both methods resulted in significant perceptual gains to a similar extent and improved the production of a subset of the target vowels towards a more native-like standard. However, production training promoted spectral shift to a greater extent than the perceptual paradigm. Thus, while both methods might benefit perception to the same extent, production training may be better than perceptual training for production improvement. Let us recall that the present investigation looks at the effects of perceptual training on both perception and production.

Summary

The instructional setting is characterized by little native input and consequently little improvement in pronunciation is generally reported. Current L2 speech models (NLM, SLM and PAM) emphasize the role of experience in accurate L2 speech perception and production. Thus, phonetic training becomes an importat tool in this context, as it offers an alternative source of specialized input and specific focus on challenging L2 sounds. The efficacy of phonetic training is typically evaluated by assessing L2 learners' changes in perceptual and or production abilities as a consequence of training. Moreover, robustness of learning is generally associated with the extent of the generalization of learning to untrained structures and the retention of learning beyond the immediate training effect. The studies reviewed in this chapter show that a High Variability Phonetic Training (HVPT) is more effective than a Low Variability Phonetic Training (LVPT) for adult L2 learners, independent of their proficiency level. HVPT methods have generally been found to be efficient as they may promote: a) generalization to new stimuli, new voices and new contexts, b) retention over different amounts of time and c) production gains to variable degrees.

Few previous studies have compared the efficacy of the two methods here investigated, namely ID vs. Categorical DIS tasks. Earlier findings indicate a greater effect of identification tasks over auditory discrimination tasks, as only the former enhanced between category distinction. However, later studies point to the potential benefit of categorical discrimination tasks as an L2 training tool. Importantly, previous research indicates that different methods may be effective in training different types of sounds (e.g. consonants vs. vowels). Moreover, orienting attention through instructions has been found to play a role in the outcomes of phonetic training studies. However, a tendency to improvement on unattended stimuli has also been reported. The next chapter introduces the sound inventories of both languages involved in the present study: standard Southern British English (the target language) and the participants' L1s (Catalan and Spanish).

3. Catalan and Standard British English inventories

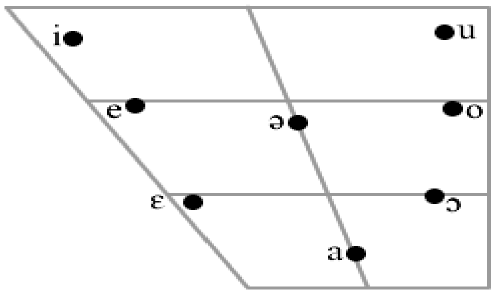
The present study investigates the acquisition of five English vowels, namely /æ/ - /ʌ/, /ɪ/ - /iː/, /ɜː/ and the English stops both word-initially and word-finally. Given the fact that L2 learners have a fully developed L1 system and L2-L1 mapping is a common source of L2 errors, it becomes important to describe the differences and similarities between the L1 (Spanish/Catalan) and the L2 (SBE) sounds that constitute the target sounds of the present study. Moreover, the most common pronunciation mistakes by this specific population of English learners will be described in light of previous research. All participants of the present study are bilingual Spanish and Catalan speakers; however, the focus of the present study is Catalan as the two languages have the same stop consonants and Spanish has a subset of the Catalan vowels. A few relevant features of the Spanish inventory will be described, whenever appropriate and divergent from the Catalan inventory.

3.1. Vowel inventories and cross linguistic difficulties

3.1.1. Catalan vowel inventory

Catalan is a Romance language spoken in Catalonia, the Balearic Islands and parts of the Valencia region (eastern and north eastern Spain), Andorra, neighbouring parts of the south of France and the town of L'Alguer in Sardinia. Eastern Catalan is the variety spoken in the Barcelona metropolitan area, where the present study took place. Its vowel inventory is constituted of eight simple vowels that are solely spectrally distinct, seven stressed peripheral phonemes /a ɛ e i ɔ o u/ and a reduced vowel in unstressed position [ə] (Recasens, 1993). According to Carbonell and Llisterri (1999), if the syllable becomes unstressed, the vowel sounds /ɛ e a / are reduced to the schwa-like vowel [ə] and the vowel sounds /ɔ o/ are raised to /u/, whereas the two high front vowels /i u/ do not undergo any vowel reduction process and maintain their quality in unstressed positions. Figure 3.1 illustrates the vowel inventory of eastern Catalan. In addition, Catalan has a variety of rising and falling diphthongs basically including all combinations of the vowels described and the high front or high back glides (Recasens, 1991).

Figure 3.1. Eastern Catalan vowel inventory (Adapted from Carbonell & Llisterri (1999))



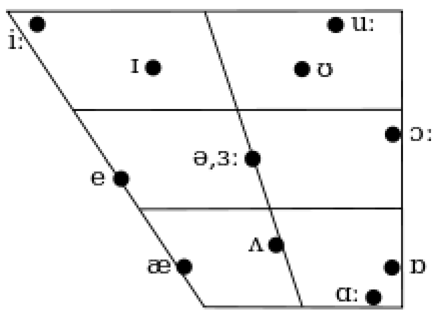
3.1.2. SBE vowel inventory

The target English variety investigated in the present study is Standard Southern British English (SBE)[5], which is the one that is usually described in English phonetics and pronunciation courses by the university where the participants were majoring their English studies. There are 11 stressed phonemes /iː ɪ e ɜː æ ʌ ɑː ɒ ɔː uː ʊ/ and one unstressed phoneme, schwa [ə] (Cruttenden, 2014). According to Solé (1991), English vowels have their full vowel quality when stressed. However, the same vowels tend to be reduced to /ə ɪ ʊ/ when unstressed. SBE is often described as having five falling diphthongs /eɪ əʊ aɪ aʊ ɔɪ/ and three centering diphthongs /ɪə eə ʊə/ (Cruttenden, 2014). Furthermore, the standard southern British English (SBE) system is characterized by the fact that apart from spectral differences

[5] The target contrasts investigated in this study are common to most varieties of English, though some vowels (e.g. /æ/) may differ somewhat in spectral and temporal cues across varieties.

among the phonemes, duration differences between the lax and the tense vowels are also relevant. There are five tense vowels in British English /iː ɜː ɑː ɔː uː/, and six lax vowels /ɪ e æ ʌ ɒ ʊ/. It is important to note that spectral differences are the main cue that differentiates articulatory close tense and lax vowel sounds in English and that duration is a secondary cue, at least in some varieties. Native speakers of (American) English have been found to rely mostly on spectral cues when discriminating tense and lax vowel pairs, predominantly in the case of the high vowels (Hillenbrand, Clark & Houde, 2000). Likewise, Escudero (2000) and Escudero and Boersma (2004) found that despite their awareness of temporal cues, L1 (Scottish and SBE) English speakers mainly used the spectral information as the primary cue when perceiving the tense/lax distinction present in the high front vowels. Figure 3.2 illustrates the SBE vowel inventory.

Figure 3.2. Southern British Vowel Inventory (Adapted from Roach (2004))



3.1.3. Cross linguistic difficulties with vowels

The expected cross-linguistic segmental difficulties regarding the five target vowel sounds of the present study are described next, in light of previous research. The SBE vowel pair /iː/-/ɪ/ poses perception and production problems for Spanish and Catalan learners of English since in their L1 there is only one vowel in the high front acoustic region, and neither

Spanish nor Catalan has a comparable tense-lax distinction. In a cross-linguistic perceptual assimilation study with native Catalan speakers differing in the amount of L2 experience, Cebrian (2006) found that the Canadian English /iː/ and the Eastern Catalan /i/ were highly comparable acoustically. Moreover, the acoustic similarity was supported by the results of a perceptual assimilation task, in which Catalan native speakers identified English vowels in terms of Catalan categories: English /iː/ was consistently identified as Catalan /i / (99% of the time) and obtained high goodness of fit ratings (a mean of 6 out of 7). Similar results were reported in two subsequent studies testing the perceptual similarity between Catalan and SBE (Cebrian et al., 2011; Cebrian, 2015; also Rallo-Fabra & Romero, 2012, for American English).

Regarding the lax vowel /ɪ/, Cebrian (2006) reported that the spectrally closest Catalan vowel was /e/. Perceptually, English /ɪ/ seems to have a poorer match in Catalan, as it has been found to assimilate to both Catalan /e/ and Catalan /i/ to varying degrees in different studies, with consistently lower goodness of fit ratings (Cebrian, 2006, 2015; Cebrian et al. 2011; Rallo-Fabra, 2005). In sum, Catalan learners of English have been found to assimilate English /iː/ to their existing L1 /i/ category, while /ɪ/ has been less consistently classified, being mapped onto Catalan /e/, and to a lesser extent to /i/. Regarding the role of acoustic cues, Catalan learners of English have often been shown to fail to make use of spectral cues when discerning the English /iː/-/ɪ/ pair and consequently to rely mostly on temporal cues to distinguish the two sounds in both their perception and production (Cebrian, 2006, 2007; Cerviño & Mora, 2009; Mora & Fullana, 2007; Morrison, 2009; Kivistö-de Souza & Carlet, 2014). As duration is not a distinctive feature in Spanish and Catalan, this over-reliance in duration most probably cannot be attributed to a negative transfer from the L1. According to Bohn's *Desensitization Hypothesis* (1995), the lack of experience with spectral differences in this specific vowel region in the L1 might result in the need to resort to other cues to the tense-lax distinction, namely the more readily available temporal cues. This may take effect possibly due to acoustic salience, the effect of orthography or even the effect of instructional biases (Cebrian, 2006; Wang & Munro, 1999).

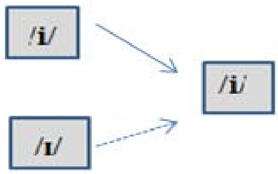
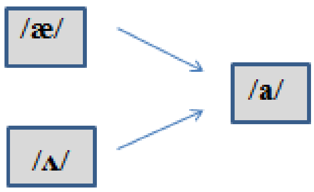
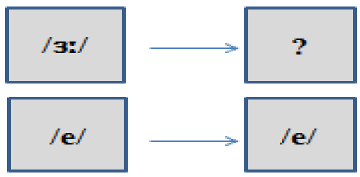
Regarding the SBE low vowels, the /æ/-/ʌ/ contrast also poses a problem for these learners as there is only one vowel in that acoustic region in Spanish and Catalan, the low central vowel /a/ (Cebrian, 2015; Rallo-Fabra & Romero, 2012). Rallo-Fabra and Romero (2012) found that due to the existence of this single central category in the low region, Catalan learners experienced great difficulty when discriminating between the Catalan and American English vowel pairs /a/-/æ/ and /a/-/ʌ/. Similar results were found by Cebrian and colleagues (Cebrian et al., 2011; Cebrian, 2015), who explored the perceptual similarity between Catalan and British English vowels. The results pointed to a high degree of perceptual similarity between the Catalan central vowel /a/ and the British English sound /æ/. As regards the SBE low (or low mid) central vowel /ʌ/, Cebrian et al. (2011) explain that it is also frequently assimilated to Catalan /a/ but with a lower degree of perceptual similarity than its counterpart /æ/. In fact, Catalan/Spanish bilinguals have been found not only to perceive, but also to produce both English /æ/ and /ʌ/ in terms of the one exemplar present in their L1 (Aliaga-García & Mora, 2009; Fullana & MacKay, 2003).

Concerning the central vowel /ɜː/, Cebrian et al. (2011) report no clear match for this sound in Catalan, being assimilated to the Catalan sounds /ɛ e o ɔ/ with low degrees of perceptual assimilation. Moreover, Cebrian (2015) found the SBE central vowel /ɜː/ to be very dissimilar from L1 vowels and stated that "such vowel is a good candidate for accurate L2 categorization given enough exposure to the target language" (p. 4) in accordance with the SLM (Flege, 1995). Moreover, the perception of this specific vowel sound by monolingual Spanish speakers has been investigated by Iverson and Evans (2007) in a perceptual assimilation study that looked at four different groups of learners of English as an L2. Their findings revealed that the British English sound /ɜː/ was assimilated to L1 Spanish /o/ by the native Spanish participants. In a study that assessed the trainability of visual, audio-visual and audio-only perception, Pereira (2014) found that Chilean Spanish speakers obtained very high scores from the outset of the study when visually identifying the sound /ɜː/ and improved their production of this specific sound with the aid of training. The author attributes this positive result to the fact that new "visemes" may pattern similarly to new phonemes, so that the more dissimilar ones are potentially acquired with more ease.

The production of the central target vowel /ɜː/ by Catalan learners has been under investigated, as to the best of our knowledge there is no study that reports production results for this sound by this population of learners. Evidence from classroom observations suggests that the target vowel /ɜː/ tends to be commonly mispronounced by Spanish/Catalan learners of English either as the sounds /e/ or /ɛ/ and its pronunciation tends to be misguided by the orthography. Following the SLM and the perceptual results found by Pereira (2014), the lack of equivalence of the non-native sound to L1 sounds might benefit the trainability of this sound and consequently its production by Spanish/Catalan L2 learners.

Even though PAM-L2 makes predictions that reflect the discriminability of L2 sounds by describing possible ways in which the L2 sounds may map onto L1 sounds and the present study only looks at sound identification, it is interesting to look into the predictions made by the model in relation to the target sounds of the current thesis, since discrimination and identification skills are usually found to be correlated (Iverson et al., 2012). Recall that PAM-L2 model (Best & Tyler, 2007) predicts that accurate perception of non-native sounds depends on how closely each sound maps on to the existing categories of L1. Therefore, a non-native sound can either be assimilated to the existing categories (categorized), perceived as a non-native sound (uncategorized) or perceived as a non-speech sound (non-assimilated). PAM-L2 establishes different predictions about discrimination accuracy depending on the way target sounds assimilate to, or map onto, L1 categories (see section 2.2.2). Table 3.1 illustrates the target vowel sounds of the present study, the predicted assimilation pattern in accordance with PAM-L2 and the expected discrimination level.

Table 3.1. Target vowel sounds and predicted assimilation pattern in accordance with PAM-L2

Target vowel sounds	Assimilation pattern (PAM-L2)	L2 – L1 Prediction	Discrimination level
/iː/ -/ɪ/[6]	Category goodness difference assimilation (CG)		Moderate to very good
/æ/-/ʌ/	Single-category assimilation (SC)		Poor
/ɜː/-/e/[7]	Uncategorized-Categorized (UC)		Good

Regarding the vowel sounds investigated in the present study, the high front vowel pair /iː/-/ɪ/ would be assimilated in the frame of the "category goodness difference assimilation", since the English vowel /iː/ has been found to be highly comparable to the

[6] Different category goodness of fit are represented in Table 3.1 by the use of the solid and dashed lines. The former indicates a high similarity, and the latter indicates a moderate to low similarity.
[7] Despite not being one of the target sounds of the present study, /e/ has been used in order to make discrimination predictions in accordance with PAM.

Catalan sound /i/ and on the other hand, the sound /ɪ/ has been found to be perceived as a poorer fit of the same L1 sound. Thus, discrimination between the non-native categories /iː/ and /ɪ/ is good and, providing sufficient exposure, learners may perceive the existing spectral and durational contrasts between the L1 and L2 sounds, which in turn might promote the formation of a new phonological category.

The low-central vowel pair /æ/-/ʌ/ is predicted to be the most challenging target sound pair to be discriminated and consequently acquired, as both non-native sounds as perceived as the same L1 sound (/a/) with a similar category goodness of fit[8]. Consequently, a poor discrimination level would be estimated and words such as "cat" and "cut" could be perceived as homophonous. Previously reported findings from Rallo-Fabra and Romero (2012) support this prediction.

As regards to the SBE central vowel /ɜː/, which has been described as a "dissimilar" sound and would be considered "uncategorized" following PAM's classification, it constitutes of a good candidate for accurate L2 categorization given sufficient exposure (Flege, 1995). In this case, the distance between the L1 and L2 categories in the phonological space determines the learnability of the specific sound (poor to very good). The further apart the L2 category is from the pre-existing categories, the better the predicted discriminability is. As this specific central sound would be perceptually isolated if successfully mapped onto the L1 vowel space, it is a good candidate to accurate perception (Flege, 1995; Cebrian, 2015).

Knowing that discrimination and identification of L2 sounds is often correlated, identification predictions of each individual sound are subsequently attempted. In the same lines as previously argued for the discriminability of TL sounds, the most dissimilar sound /ɜː/ will be the best identified sound. /iː/ will also be successfully identified, being the vowel

[8] SBE /æ/ is perceived as slightly closer to L1 Catalan vowel /a/ than /ʌ/ is (Cebrian et al., 2011).

most consistently assimilated to an L1 vowel, followed by its lax counterpart /ɪ/. The sounds that will pose more difficulty for the leaners will be /æ/-/ʌ/, since they will both be confused with the L1 sound (/a/). It is relevant to note that although individual sounds are explained and specific predictions are made for each sound, the target sounds of the present study will be treated globally; that is, vowel perception and production and consonant perception and production.

Having described the vowel inventory of the two languages under investigation in the present study (SBE and Catalan) and the expected cross linguistic difficulties for vowels, the stop consonant inventories of both languages and the corresponding predicted cross linguistic problems are described next. This study looks at both initial and final stops; however, they are treated as different measures. The initial stop consonants and the corresponding predictions will be introduced first, followed by the description and expectations for final stops.

3.2. Initial stop consonant sound inventories and cross linguistic difficulties

Most languages in the world distinguish consonants by voicing, and voice onset time (VOT), which is "the time with respect to release for the onset of voicing" (Mackay, 1987, p. 93), is an acoustic cue associated with this distinction. Ferrand (2007, p. 267) explains that most "speakers in all languages rely on this acoustic feature to categorize a segment either as voiced or voiceless". Therefore, the acquisition of VOT is a crucial feature in second language learning.

Even though stops are among the most frequent sounds across human languages (Yavaş, 2007; Ferrand, 2007; Yavaş & Wildermuth, 2006), they are not always phonetically realized similarly. In a study examining the production of stops in 11 different languages, Lisker and Abramson (1964) identified three main VOT modes: (a) a *voice lead* mode in which voicing onset precedes the release from stop closure (interval of time between -125 and 0), (b) a *short-lag* mode in which voicing onset is more or less simultaneous with release from stop closure (interval of time between 0 ms. to +35ms), and (c) a *long-lag* mode in which voicing onset occurs substantially after the release (interval of time between 35ms and 100ms). The stop consonant inventories of Catalan and English will be further described in line with the concepts here described.

3.2.1. Catalan initial stop consonant inventory

In Spanish and Catalan, initially positioned voiced stops are realized with a voice lead VOT mode, which results in negative VOT values and prevoicing (Lisker & Abramson, 1964; Cho & Ladefoged, 1999). Prevoicing is an important feature in order to distinguish between voiced stops /b d g/ and voiceless unaspirated stops /p t k/ in Spanish and Catalan, as aspiration is a non-existent phonetic property in Romance languages (Ladefoged, 1962). Thus, Spanish and Catalan are characterized by the production of unaspirated rather than aspirated initial stops. The Catalan voiceless stops are produced with a short-lag mode, in which the onset of voicing coincides with the release of the stop closure, resulting in VOT values that are close to zero, between zero and 10 ms according to Lisker and Abramson (1964). Even though in this study stops are not tested in medial position, it is relevant to note that Spanish and Catalan voiced stops /b, d, g/ undergo spirantization when found between continuant sounds. However, they remain stops following a pause or a non-continuant sound.

3.2.2. SBE initial stop consonant inventory

In standard Southern British English (SBE), the initial voiced stops /b d g/ are mainly unaspirated and tend to be produced simultaneously with the onset of voicing, which results in a VOT value near zero (Docherty, 1992; Berry, 2004). According to Berry (2004), the exact short lag values in English range from zero to 30 milliseconds. However, short lag is not the only possible mode in which voiced stops can be produced in utterance initial position. In fact, previous studies have reported a bimodal pattern for English voiced stops, as some speakers produced them with short lag VOT and some with voice lead (Lisker & Abramson, 1964; Smith, 1975; Berry, 2004). Prevoicing tends to occur when voiced stops are preceded by voiced sounds, especially when preceding high vowels and in labial or alveolar places of articulation (Smith, 1975). On the other hand, English voiceless stops /p t k/ in the onset of a stressed syllable are aspirated. These aspirated stops are produced with a long-lag VOT mode, in which the voicing onset occurs substantially after the release. This voicing delay results in a VOT of 30 ms or longer, corresponding to the aspiration interval (Benkí, 2005; Cho & Ladefoged, 1999; Lisker & Abramson, 1964). According to Kent, Read and Kent (1992, p. 106) "Aspiration is a breathy noise generated as air passes through the partially closed vocal folds and into the pharynx" and according to Lisker and Abramson (1964), aspiration is the cue that reliably distinguishes initial voiced and voiceless stops at the beginning of a stressed syllable in English.

According to previous research, there are three essential variables that have been found to affect English VOT realization: place of articulation (POA), speech rate and the height of the following vowel segment. According to Lisker and Abramson (1964), POA can affect the amount of aspiration produced: velar stops have the longest lag, followed by the alveolar stops and the bilabials. Thus, generally the further back the POA, the higher the VOT values (Lisker & Abramson, 1964; Docherty, 1992; Cho & Ladefoged, 1999). As for vowel height, stops tend to have longer lag preceding higher vowels and shorter lag preceding low vowels (Flege, Frieda, Walle & Randazza, 1998; Yavaş, 2002, 2007; Yavaş & Wildermuth, 2006). According to Yavaş (2007), "since the high tongue position that is assumed during the stop closure in anticipation of a subsequent high vowel would result in a less abrupt pressure drop, a stop produced as such will have longer lag than the one produced before a low vowel" (p.493).

Importantly, speech rate has been described as possibly "the most influential factor in VOT variation" (Reis & Nobre-Oliveira, 2007), as the slower the speech rate, the longer the VOT measurements tend to be. This important effect of speaking style and consequently speaking rate on the VOT of English stops has been reported in several previous studies (e.g. Labov, 1972; Solé & Estebas, 2000; Reis & Nobre-Oliveira, 2007; Mora, 2008; Bach, 2012). Solé and Estebas (2000) explain that "aspiration of English voiceless stops has the status of a phonological rule and it adjusts to varying speech rate" (p. 443). Conversely, the authors add that the same is not true for languages such as Catalan, Spanish and French, as VOT is independent of speaking rate in these Romance languages. Thus, speech rate might affect L2 pronunciation, since a cross-language comparison of this feature is not straightforward (Bach, 2012; Solé & Estebas, 2000).

In view of these findings, the three influencing factors described above have been controlled for in the present study, as the stimuli contained a balanced number of high and low vowels, an equal number of alveolar, bilabial and velar stops and the production task was administered in a way that speaking rate was comparable among speakers (see section 5.3 for details on the stimuli, and section 5.5 for details on the procedure).

3.2.3. Cross linguistic difficulties with initial stop consonants

As previously mentioned, the VOT mode for Catalan voiceless stops (short lag) might correspond to the VOT mode for English voiced stops in utterance initial position.[9] According to the SLM, this possible similarity may enhance L2 misperception and consequently misproduction, as L2 learners have to override this "similarity" in order to learn the phonetic details of the target language (see section 2.2.3 for more details on the SLM). Moreover, English voiceless aspirated stops have no counterpart in the Catalan consonant inventory. Thus, learning the English VOT patterns is challenging for these learners (Alves & Zimmer, 2015; Fullana & MacKay, 2008) and English voiceless stops are usually produced without the required aspiration. This possible lack of aspiration in learners' production may result in intelligibility issues, given the fact that native English speakers rely on this cue when distinguishing voiceless and voiced stops in word-initial position (Schwartzhaupt, Alves & Fontes, 2015). Alves and Zimmer (2015) point out that aspiration might be seen by L2 learners as a feature that helps reducing foreign accent rather than an essential cue to establish a phonological distinction.

Regarding the POA of the stops, both English and Spanish have coronal stops. However, the coronal stops in these two languages differ not only in terms of VOT, but also in place of articulation (Casillas, Díaz & Simonet, 2015), as /d/ and /t/ are alveolar in English while they are dental (i.e., [d̪ t̪]) in both Spanish and Catalan. Therefore, the English /d t/ involve an additional source of difficulty for Spanish/Catalan EFL learners (Cortés 2003, 2007).

As regards to predictions for the initial stops, it is probable that the English voiceless stops will not pose much of a problem to the L2 learners due to the strength of the

[9] This is not true for all speakers, as a bimodal pattern has previously being described.

aspiration cue. Conversely, voiced stops may be problematic since the short lag VOT may be interpreted as voiceless stops by Catalan/Spanish listeners. Importantly, the coronal stops may pose an added difficulty due to the difference in POA in the L1 and the L2.

3.3. Final stop consonant sound inventories and cross-linguistic difficulties

3.3.1. Spanish/Catalan final stops inventory

Voicing is not contrastive in final position in either Spanish or Catalan. In Spanish, the few finally positioned /b d g/ undergo either spirantization, in which the stops are weakened to voiceless fricatives, or are omitted (Harris, 1969; Flege & Davidian, 1984; Mascaró, 1991). As for Catalan, the non-occurrence of voiced stops word finally is due to the phonological process of final obstruent devoicing **(**Recasens, 1993; Cebrian, 2000). According to Recasens (1993), Catalan syllable final obstruents are necessarily voiceless, since voicing neutralization takes place in prepausal contexts or when preceding a voiceless consonant. Conversely, Catalan final obstruents are voiced when preceding a voiced consonant. It is relevant to note that in the present study stops are solely tested in prepausal position, where the result of L1 neutralization is always a voiceless stop. Thus, the (prepausal) L1 final stop consonant inventory of the learners of the present study is constituted of three voiceless stops, /p t k/.

3.3.2. SBE final stop consonant inventory

In SBE there are six final stop consonants /p t k b d g/. The voiced stops /b d g/ are usually devoiced to some extent in final position, especially when phrase-final or when followed by a voiceless consonant (Raphael, 1972; Roach, 2000). Thus, the authors explain that vocal fold vibration is not the main cue that distinguishes the final voiced-voiceless contrast in English, as this feature is not consistently present. Due to this voicing inconsistency, the duration of the preceding vowel becomes an important cue to final voicing distinction. According to Raphael (1972), "the presence of voicing during the closure period of a final consonant does have some cue value, although it is minor when compared to that of vowel duration", as the author also affirms that "preceding vowel duration is a sufficient cue to the perception of the voicing characteristic of a word final stop, fricative or cluster" (p. 1301). Flege (1989) provided evidence to that effect. He tested native English speakers' perception of modified final coronal stops /t d/ preceded by the English high front vowels /iː ɪ/. The results showed that native speakers of English did not rely on the voicing cue, as a high percentage of correct identification was observed even when the voicing and the release burst cues had been omitted.

3.3.3. Cross linguistic difficulties with final stop consonants

Due to the similar but yet different patterns of voicing in the final consonants of the participants' L1 and L2 described above, this feature constitutes of a prime candidate for negative phonetic transfer. According to the SLM, sounds classified as "similar" are harder to acquire, as they continue to be perceived according to the pre-existing L1 phonetic category. Consequently, previous research has found that Spanish and Catalan speakers usually resort to L1-based strategies such as voicing assimilation and devoicing, spirantization or deletion (Cebrian, 2000; Flege & Davidian, 1984) when perceiving and producing final voiced obstruents. Cebrian (2000, p. 2) explains that these learners "have to overcome their L1 mechanisms" in order to learn this L2 feature. Moreover, according to Gonet & Święciński (2012), "the implementation of the voicing of English word-final obstruents constitutes a serious problem for foreign speakers of English whose native languages organize the phonological ‘voicing’ contrast differently” (p. 7). This difficulty with perception and/ or production of coda stops is enhanced by the fact that coda position is known to be a less stable position in terms of the strength of consonant cues, which makes final consonant perception more challenging than initial consonant perception (Jang, 2014).

Evidence of the difficulty experienced by Catalan learners of English with English final voiced stops is reported by Fullana and Mora (2007) in a study comparing the perception and production of final stops by 48 native speakers of Catalan/Spanish to that of English native speakers. The results of an AXB task revealed very high discrimination scores for the final stops /b d t d/ by the L2 learners. By contrast, the authors found that the production patterns of both learners and native speakers significantly differed in two characteristics: a) the duration of the vowels preceding the stop consonants, as the Catalan learners produced shorter vowels than the NES, and b) the voicing used in the final stop closure, since the L2 learners devoiced the stimuli more often than the NES. The authors concluded that even though the word final voicing contrasts examined in their study were accurately perceived by the participants, they were less robustly realized by the Catalan L2 learners than by NES.

Regarding the acquisition of final stops, the SLM claims that learners who lack L2 consonant sounds in certain allophonic positions in their native perceptual space (e.g. final voiced stops) should be able to create novel phonetic categories for those specific sounds. However, empirical evidence shows that English final stop perception and production by learners whose L1 does not allow for final stops voicing is challenging (Flege et al., 1992; Flege et al., 1995; Flege, 1998; Flege, 1989; Flege, 1995b). Thus, it can be predicted that the Catalan trainees, in the same fashion as other speakers whose L1 does not contain voicing contrast in final stops, might be faced with a challenging task at pre-test, but may be able to improve their perception and consequently production of the English final voiced stops as a result of training, in line with previous research (Flege, 1998; Flege, 1989; Flege, 1995b).

Summary

The goal of the current chapter was twofold.It aimed at reviewing the vowel and stop consonant inventories of the languages that are the focus of this study and at discussing the main difficulties expected for the Catalan learners when perceiving and producing the L2 target sounds in the light of previous research and the predictions by the current L2 speech learning models. In the next chapter, the main objectives and research questions will be presented and specific hypotheses will be formulated.

4. The current study

The aim of this thesis is to further our understanding of the process of acquiring consonant and vowel sounds of a second language (L2) and to investigate specific training techniques that facilitate learning. This research project explores the effect of two high variability phonetic training (HVPT) methods and investigates whether they promote improvement of perception and production of standard Southern British English (SBE) sounds by native speakers of Catalan. Specifically, this study aims to investigate if the predicted improvement is greater following an identification training methodology (ID) or an AX categorical discrimination training (DIS) methodology. It also aims at evaluating the effectiveness of each method for untrained sound contrasts by means of a cross-training design, that is, by means of exposing trainees to both consonant and vowel contrasts throughout training, but having learners focus specifically on only one type of sound, either consonants or vowels.

Moreover, this research project aims at investigating if the knowledge acquired through training generalizes to novel non-words and real words, to novel talkers that subjects have not been exposed to during training, and to production. Generalization allows us to see if the potentially positive effect of training goes beyond the specific training stimuli, indicating that robust learning has taken place (Logan & Pruitt, 1995). Another important aspect addressed by this study is the retention of the acquired knowledge over time, which provides a way to measure the mid or even long-term effect of the changes that we observed during training. This thesis specifically looks at retention over a period of two months.

In light of the issues presented in the literature review and the aims of the current study, the research questions addressed by this project are presented below, and subsequently the main hypotheses are formulated.

4.1. Research questions

RQ.1 Does high variability phonetic training on vowels positively affect the perception of trained as well as untrained SBE segments by Spanish/Catalan bilinguals?

> **Sub-RQ.1.1** Which type of training (identification (ID) or categorical discrimination (DIS) is more effective in promoting improvement on trained segments (vowels) for the vowel trainees?
>
> **Sub-RQ.1.2** Which type of training (identification (ID) or categorical discrimination (DIS) is more effective in promoting improvement on untrained segments (consonants) for the vowel trainees?

RQ.2 Does high variability phonetic training on consonants positively affect the perception of trained as well as untrained SBE segments by Spanish/Catalan bilinguals?

> **Sub-RQ.2.1** Which type of training (identification (ID) or categorical discrimination (DIS) is more effective in promoting improvement on trained segments for the consonant trainees?
>
> **Sub-RQ.2.2** Which type of training (identification (ID) or categorical discrimination (DIS) is more effective in promoting improvement on untrained segments for the consonant trainees?

RQ.3 Do the improvements obtained through training generalize to real words (1), novel non-words (2) and novel talkers (3)?

> **Sub-RQ.3.1** Which type of training (ID, DIS) is more efficient in promoting each type of generalization (novel real words (1), novel non-words (2) and novel talkers (3))?

RQ.4 Does perceptual phonetic training promote retention of acquired perceptual knowledge beyond the training phase?

> **Sub-RQ.4.1** Which type of training (ID, DIS) is more efficient in promoting retention of learning two months after training?

RQ.5 Does perceptual phonetic training lead to improvement in production (as perceived by native English speakers) of trained sounds and untrained sounds?

> **Sub-RQ.5.1** Which type of training (ID, DIS) is more effective in improving subjects' production of trained and untrained sounds for the vowel trainees?

> **Sub-RQ.5.2** Which type of training (ID, DIS) is more effective in improving subjects' production of trained and untrained sounds for the consonant trainees?

RQ.6 What are the trainees' impressions on phonetic training as an EFL tool to improve learners' pronunciation?

> **Sub-RQ.6.1** Which training regime (ID, DIS) is preferred by the learners?

4.2. Hypotheses

In relation to **RQ.1** and **RQ.2**, improvement from pre-test to post-test is predicted to occur with trained sounds as have been found in previous studies involving HVPT (see section 2.3.3). However, the extent of the improvement may vary across groups and/or type of segments. Besides, improvement on trained sounds is expected to be greater than in untrained but implicitly exposed ones (Pederson & Guion-Anderson, 2010; Nozawa, 2015) due to the role of feedback (Logan & Pruitt, 1995) and explicit instruction (Alves & Luchini, 2016). However, improvement of untrained sounds might occur as a consequence of the balanced exposure received throughout the training regime, supporting previous tendencies of improvement (Nozawa, 2015).

Despite the predicted overall improvement, the performance with vowels and consonants may differ, as vowels and consonants have been said to be perceived differently (Fry et al., 1962; Strange, 2007), involve different levels of processing when learned (Pisoni, 1973) and thus might require different types of perceptual training (Nishi & Kewley-Port, 2007; Pereira, 2014, Nozawa, 2015). Thus, learners' consonant perception is predicted to be modified with more ease than their vowel perception, especially in the case of initial stops (Collet et al., 2013; Everitt, 2015; Alves & Luchini, 2016), since aspiration is a robust cue which might facilitate perception (Burnham, 1986).

Regarding the final stops, since their acquisition is challenging for L2 learners' (Cebrian, 2000; Flege et al., 1992; Flege et al., 1995; Flege, 1998; Flege, 1989; Flege, 1995b) and their perception is predominantly cued by the duration of the preceding vowel (Raphael, 1972; Flege, 1989), it can be hypothesised that the trainability of these segments will follow a pattern similar to that of the vowels' rather than that of the initial stops (i.e. with less ease than initial stops (Jang, 2014)). Nonetheless, the expectation is that L2 learners will be able to improve their perception and consequently their production of the English final stops to some extent as a result of the training received (Flege, 1998; Flege, 1989; Flege, 1995b).

Regarding the effectiveness of the training tasks (**Sub-RQ.1.1** and **Sub-RQ.2.1**), since both training methods tap into similar levels of processing (Flege, 2003; Højen & Flege, 2006) and also promote L2 categorization (Polka, 1992; Strange, 1992), it is possible that both training methods will be equally effective in improving learners' performance after training, in accordance with Flege (1995b) and Wayland and Li (2008). However, due to the phonological and phonetic differences between vowels and consonants (Pisoni, 1973), different training tasks might be more beneficial for different segments (Nishi & Kewley-

Port, 2007; Pereira, 2014, Nozawa, 2015). It is possible that ID trainees will obtain better results than DIS when identifying vowels (Nozawa, 2015) and both groups will improve similarly with consonants (Flege, 1995b; Nozawa, 2015). This might be related to the fact that the different tasks (ID and DIS) enhance different abilities (Jamieson & Moroson, 1986; Pisoni & Lively, 1995; Iverson et al., 2003, Iverson et al., 2012), which in turn might be more suitable for identifying different segments.

Additionally, for **RQ.3,** if the training methods used promote robust learning, perceptual improvement obtained during training will transfer, at least partially, to novel dimensions such as real words (Nishi & Kewley-Port, 2007; Pierce, 2014), new voices (Flege, 1995b; Iverson & Evans, 2009, Wang & Munro, 2004) and to novel tokens (Hazan et al., 2005; Lacabex et al., 2008; Rato, 2014, Pereira, 2014). Thus, both methods (**Sub-RQ.3-1**) may be able to promote generalization to new talkers and voices to a similar extent, showing robust learning as a result of the identification training (ID) and the categorical discrimination training (DIS).

Furthermore, if the training methods used are found to be effective in enhancing learners' perception of L2 sounds, it is likely that retention of the perceptual abilities (**RQ.4**) is observed. All groups (**Sub-RQ.4.1**), independently of the training method they were exposed to, might retain the L2 learning two months after the completion of the training regime, which will shed some light on the long-term effects of HVPT training (e.g., Bradlow et.al, 1997; 1999; Nishi & Kewley-Port, 2007; Rato, 2014; Wang & Munro, 2004). Retention effects will thus provide further support for the efficacy of the methods, especially for consonants (Flege, 1995b).

Assuming there is improvement from pre-test to post-test, the expectation is that at least some positive changes in learners' production of directly trained L2 sounds (**RQ.5**) will be found, providing further support that perceptual training with no focus on production may lead to production gains, even if to a lesser extent than the perceptual gains (Rochet, 1995; Yamada et al., 1996; Bradlow et al., 1997, 1999; Hardison, 2003, 2004; Hazan et al., 2005; Iverson & Ewans, 2009; Thomson, 2011; Pereira, 2014). These findings are predicted to corroborate the SLM's predictions, which postulate that there is a link between perception and production. In fact, according to the SLM, perception precedes production and segments are produced "only as accurately as they are perceived" (Flege, 2003, p. 344). Regarding the effectiveness of each method, to our knowledge the current study pioneers in assessing the training effects on production by contrasting ID tasks with categorical DIS tasks (early work made use of auditory discrimination tasks). Therefore, the prediction proposed here for sub-RQ.5.1 is based on previous results involving measures of training robustness other than production skills. Taking into consideration Flege's (1995b) results that both perceptual training methods appeared to be equally effective in improving perception, it is hypothesized that both training tasks (ID and DIS) will promote robust learning to the same extent and thus generalization to production will occur for all trainees, even if to different degrees across segment types.

Lastly, regarding **RQ.6**, learners are expected to react positively towards the use of phonetic training as a pronunciation tool, and especially since this method was incorporated in the subject "English Phonetics and Phonology I" providing them with extracurricular practice. As regards to the last sub question (**Sub-RQ 6.1**), it is hypothesised that the identification training regime will be favoured by students over the discrimination training. This is because an ID task might be seen as a more meaningful task by the participants, due to the presence of the labels and the consequent focus on form that it entails.

5. Methods

The experiments described in this chapter were conducted to investigate the effects of two different high variability perceptual training (HVPT) methods, namely categorical discrimination and identification, on the perception and production of trained and untrained sounds by Spanish-Catalan bilinguals learning English as an L2. The experimental design adopted consisted of a pre-test/training/post-test/delayed post-test regime that lasted ten weeks in total. The L2 data for the present study was collected from September 2014 until February 2015 at Universitat Autònoma de Barcelona (UAB), Spain. This chapter first presents the L2 participants' background and their characteristics, and then it describes the data collection instruments, the design and procedure of the testing and training tasks used at each stage of the study and ends by introducing the methods employed for data analysis.

5.1. L2 participant recruitment and selection

A total of 166 undergraduate students at *Universitat Autònoma de Barcelona* were recruited to participate in the experiments. All participants were second-year English majors and agreed to participate in the study by signing a consent form (Appendix A). These L2 learners of English were enrolled in an introductory phonetics and phonology course, namely *English Phonetics and Phonology I,* and received a total of eight percent course credit for their participation and completion of all tasks they were required to perform. At the time of testing, participants were also enrolled in their third semester of instrumental English courses, corresponding to a level of English between upper-intermediate and advanced (approximately a B2/C1 level in the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFRL, Council of Europe, 2001)). In order to assess the individuals' proficiency further, a grammar and vocabulary test (the Cambridge online test)[10] was administered. The test consisted of 25 questions and was administered through a Moodle platform students had access to. The results of this level test results are presented in Table 5.1 below (section 5.1.3), which shows several measures for each experimental group.

A pre-selection phase took place based on learners' self-reported language background information, which was obtained by means of a language background questionnaire (see Appendix B). This instrument contained 21 questions and aimed at gathering some general information about the participants, their native language(s) and further details of their English L2 usage and exposure. Forty six students did not have the desired homogeneous L1 background for this study, namely being Spanish and Catalan bilinguals, and thus were directed to an alternative activity, an online transcription platform for phonetic transcription practice (more details are provided in section 5.4.3.4). These students declared to belong to one of the following four scenarios: a) monolingual Spanish speakers (36.9%), b) Spanish and Catalan bilinguals who also spoke English, Dutch or German at home (26.1%), c) speakers of a different L1 altogether (30.4%) and d) speakers who learned Catalan after the age of six (6.5%). The age of six was chosen, as previous research has found a link between early bilingualism and L2 perception (Polka, 1992). Finally, 20 randomly selected learners were used to pilot an additional perceptual training method that was later found not to fit the general purpose of the present project. Therefore, these participants were discarded from the current study. The remaining 100 learners underwent a ten-week-long testing/training regime and will be considered the participants of the present study. They are described in more detail in the next section.

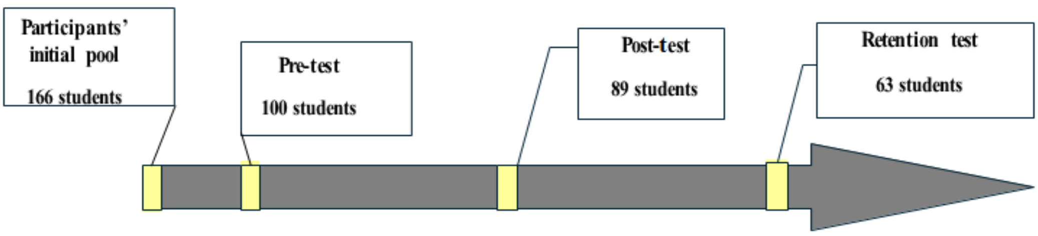
[10] http://www.cambridgeenglish.org/test-your-english/adult-learners/

5.1.1. Participants

One hundred bilingual Catalan/Spanish speakers participated in the present study and were divided into five groups (*N*=20). There were four experimental groups and a control group (henceforth **CG**). The experimental groups were two discrimination training groups and two identification training groups trained on either consonants or vowels. They will be referred to as **ID_V, ID_C, DIS_V,** and **DIS_C**, where ID/DIS corresponds to the training method (ID= identification training; DIS = categorical discrimination training) and V/C indicates the segment they were trained on (V = vowels and C = consonants).

Eleven out of the 100 initial participants did not complete all the training sessions and/or the post-test and thus were discarded from the study. Losing participants in the course of a longitudinal training study is not unusual and it has in fact been referred to as participant attrition or participant mortality (Mackey & Gass, 2005). In addition, the number was further reduced at the last stage of testing, the retention test. Sixty three participants returned to this stage, which took place 2 months after the post-test phase. The decrease in participants at this latter phase can be explained by the fact that the retention test was administered in a different semester from the rest of the study and it became more difficult to recruit all the students back to perform this last task. The number of participants at the different stages of the study (pre-test, post-test and retention test) is presented schematically in Figure 5.1.

Figure 5.1. Number of participants at each stage of the study



5.1.2. Participants' characteristics

The participants were 81 female and 19 male students and their mean age was 19.9, ranging from 18 to 34. They were all Catalan/Spanish bilinguals; however, they differed in terms of their self-reported language dominance. Language dominance in the current study was defined by the participants themselves in accordance to their proficiency, language used at home and/or general language usage (see the language background questionnaire in Appendix B). Fifty-five participants reported to be Catalan dominant and forty-five claimed to be Spanish dominant. The participants' mean age of first exposure to English was six (ranging from 2 to 12) and the mean number of years of formal instruction was 13 (ranging from 5 to 27). Moreover, none of the selected participants reported having spent longer than four months in an English-speaking country or reported having any vision and/or hearing difficulties. Most of these learners had no previous knowledge of English phonetics and phonology at the beginning of the study, with the exception of the students retaking the subject (33% of the initial number of participants). The decision to keep the repeaters in the study was based on the fact that a training study follows a repeated measures design and thus crucial comparisons involved the performance of each individual participant at different stages in the study. In addition, it was decided not to deprive repeaters from the extra practice that these sessions entailed.

5.1.3. Distribution into experimental groups

The distribution of participants into groups was based on the participants' global perceptual scores obtained at pre-test, and aimed at forming homogenous groups at the outset of the study. The global pre-test scores were the participants' combined vowel and consonant perception scores across two types of words: real and nonsense words (see section 5.4.2.2 for the design of the pre-test). Moreover, a similar distribution of course repeaters, overachievers[11] (scores higher than 75%) and underachievers (scores lower than 55%) was carefully applied to each group. Table 5.1 shows the initial descriptive statistics and scores for each of the five groups.

Table 5.1. Language background information and initial scores by group

	Group	CG	ID_V	ID_C	DIS_V	DIS_C	Total
Pre-test score	Total n	(n=20)	(n=20)	(n=20)	(n=20)	(n=20)	(n=100)
	Repeaters	n= 7	n= 7	n= 6	n= 6	n= 7	n= 33
	Mean	66.73	67.08	66.89	66.25	67.29	66.85
	SD	7.04	6.76	6.98	5.52	4.98	6.20
	Range	54.8-84.1	54.3-81.1	52.6-75.9	54.3-80.5	55.3-75.1	52.6-84.1
Age	Mean	19.6	19.7	20.5	19.9	20.2	19.9
	SD	1.4	1.6	2.1	1.5	3.6	2.2
	Range	18-23	18-25	18-26	19-25	18-34	18-34
AOL	Mean	5.8	5.9	6.5	5.6	6.4	6.1
	SD	2.2	2.2	2.3	1.7	2.1	2.1
	Range	03-12	03-12	02-10	03-10	02-08	02-12
Time abroad (in months)	Mean	0.5	0.4	0.5	0.5	0.4	0.5
	SD	1.1	0.9	0.8	0.9	0.7	0.9
	Range	0-3.50	0-4	0-3	0-4	0-2.5	0-4
Level test grade	Mean	7.6	7.5	7.1	7.2	7.4	7.4
	SD	2.2	1.6	1.5	1.8	1.8	1.7
	Range	3.7-10	4.5-9.5	4.7-10	3.5-9.2	4.5-10	3.5-10

Note. AOL= Age of English learning onset

As can be seen in Table 5.1, the five groups were very comparable at the beginning of the study with a mean percentage accuracy of 66.8, varying from 52.6 to 84.1. Statistical analyses confirmed that the five groups did not differ from each other significantly (further details of the statistical analysis are presented in the results section 6.1). Moreover, the groups had a 7.4 mean score on the level test, ranging from 3.5 to 10.

[11] 75% correct identification was set as the over-achieving score as very few students obtained a global initial score higher than this. Recall that the global initial scores were composed by the identification of real as well as non-words.

5.2. General design and timeline

The four experimental groups and the control group were tested before (pre-test), after (post-test and generalization) and two months after (retention) a five-week training period. Testing included perception tasks[12] (vowel and consonant identification) and a production task (details are given in section 5.4.2 below). Generally, students' production was assessed before they performed the perception tasks in order to avoid any possible bias in the production data resulting from previous exposure to the perceptual test stimuli. Each participant was assigned a weekly slot of 35 minutes, which was part of the coursework for the subject they were enrolled in, and were expected to complete all first nine weeks in order to be granted course credit. The task order and the approximate study timeline[13] can be seen in Table 5.2.

Table 5.2. Approximate study timeline

WEEK 1	Production pre-test
WEEK 2	Perception pre-test
WEEK 3	Training session 1 (Perception only)
WEEK 4	Training session 2 (Perception only)
WEEK 5	Training session 3 (Perception only)
WEEK 6	Training session 4 (Perception only)
WEEK 7	Training session 5 (Perception only)
WEEK 8	Production post-test + Perception post-test (Generalization 1+3a)
WEEK 9	Separate generalization tests (Generalization 2+3b)
WEEK 10 (2 months later)	Retention test

[12] The original study design included ID and DIS tasks. However, due to the large number of participants at pre-test (166 learners) and due to practical reasons, namely students' schedule and computer availability, the testing time had to be reduced. Thus, the ID task was chosen over the DIS task, as it is the most commonly documented testing task in the phonetic training literature (e.g., Flege 1995b; Logan & Pruitt, 1995; Beddor & Gottfried, 1995).

[13] The original study design included 11 weeks and six training sessions. However, due to a student strike on week six, the number of sessions had to be adjusted. Moreover, this timeline differed for some individuals in cases when students missed a session and had to make up for it on a different day.

The perceptual pre-test, post-test and retention test were identical and consisted of vowel and consonant identification in non-words produced by different talkers from those used in the training phase, in order to assess generalization to new talkers (generalization 3a below). Additionally, at pre-test and post-test, consonant and vowel identification was also tested in real words spoken by novel talkers (generalization 1). Generalization 1 and 3a required the inclusion of the non-words and real words at pre-test and post-test phases, in order to possibilitate comparison of performance before and after training. Following the post-test, an additional separate generalization test was designed to test generalization to new non-words produced by familiar training talkers (generalization 2) and to familiar non-words produced by absolute new talkers (generalization 3b). Absolute new talkers refer to talkers that have not been previously heard, neither at testing nor at training stages. Gen 1 and 3a were meant to test the efficiency of training providing pre and post training measures for comparison, and Gen 2 and 3b further investigated the robustness of learning, that is, if it extended beyond trained structures and testing stimuli. The different types of generalization are illustrated and summarized in Table 5.3.

Table 5.3. Generalization design

Gen	Generalization type	Talkers	Time of assessment
Gen 1	Real words produced by novel talkers	T1+ T2	Pre-test + Post-test
Gen 2	Novel non-words produced by familiar (training) talkers	T3 +T4	Separate gen. test
Gen 3a	Familiar (training) non-words produced by novel talkers	T1 + T2	Pre-test + Post-test
Gen 3b	Familiar non-words produced by absolute novel talkers	T7 + T8	Separate gen. test

Note: Gen = Generalization.

The production phase assessed the target vowel and consonant sounds embedded in real word stimuli. The decision to use real words for the production assessment was driven by practical reasons, namely the difficulty of eliciting sounds embedded in non-word stimuli accurately, and the large number of students taking part in the study. Hence, the production phase tested generalization of perceptual training carried out with non-words to production of real word stimuli. As regards to the training, two experimental groups received five weeks of direct training on vowel target sounds while being implicitly exposed to stop consonant target sounds (embedded in the CVC stimuli), whereas the other two experimental groups received five weeks of direct training on stop consonant target sounds while being implicitly exposed to the vowel target sounds (embedded in the CVC stimuli). The control group was asked to transcribe texts phonemically using an online platform (see section 5.4.3.4 for further details). The training for the control group was intended to provide a similar amount of target language instruction received by the other groups without specific perceptual training. Table 5.4 provides a schematic summary of the type of the stimuli used in training and testing in terms of the use of real and non-sense words, the talkers that provided the stimuli for each task and the nature of the generalization tests.

Table 5.4. Stimuli type and talker design

<table>
<tr><th></th><th>Pretest, Post-test and Retention test</th><th>Training phase</th><th colspan="2">Generalization to novel talkers and tokens</th><th>Production test</th></tr>
<tr><td rowspan="2">Type of segment
and
Stimuli</td><td rowspan="2">V + C
Non-words
(Generalization 3a)
+
(Generalization 1)
V + C
Real words*</td><td>Vowel groups
Non- words
Focus and feedback on V contrasts
Cross-trial exposure to C contrasts</td><td rowspan="2">(Generalization 2)
V + C
Novel non-words</td><td rowspan="2">(Generalization 3b)
V + C
Non-words
used in training
produced by absolute novel talkers</td><td rowspan="2">V + C
Real words</td></tr>
<tr><td>Consonant groups
Non-words
Focus and feedback on C contrasts
Cross-trial exposure to V contrasts</td></tr>
<tr><td>Talkers</td><td>T1, T2</td><td>T3, T4, T5, T6</td><td>T3, T4</td><td>T7, T8</td><td></td></tr>
</table>

Note. C= consonants, V= vowels, T= talker. * Real words were not included at the retention phase, only at pre-test and post-test.

5.2.1. Target sounds

The target sounds were the five standard Southern British English (SBE) vowels /iː ɪ æ ʌ ɜː/ and the six SBE stop consonants /p t k b d g/, placed either word initially or finally. The reason to focus on these specific English sounds was motivated by predictions of theoretical models of L2 speech acquisition such as Flege's SLM (1995) and Best and Tyler's PAM-L2 (2007), which take into account the influence of the L1 inventory when predicting difficulties in the acquisition of a second language, as well as on the basis of previous research on commonly mispronounced L2 English sounds by native speakers of Spanish/Catalan (Cebrian, 2000, 2006; Cebrian et al., 2011; Rallo-Fabra & Romero, 2012; Mora & Fullana, 2007).

5.3. Stimuli

The stimuli used in the present study consisted of both real and nonsense CVC words produced naturally (i.e., non-modified). Real words and non-words were used at the testing phase and only non-words were used at the training phase. The stimuli elicitation, preparation and validation as well as the details of the stimuli used at testing and/or at training phases will be described next.

5.3.1. Stimuli elicitation and preparation

A total of ten standard southern British English (SBE) native speakers (five females and five males) were recorded to obtain the training and testing stimuli. The recordings were carried out by the author of this thesis during a brief stay in London, UK. In order to recruit participants, an online research database that was accessed through the *Speech, Hearing and Phonetic Science Department* at *University College London (UCL)*. Talkers were asked to fill in a language background questionnaire previous to being recorded (See Appendix C). All talkers were from and had spent most of their lives in the south of England and thus they spoke a homogeneous and standard variety of British English, fulfilling the requirement for selection. None of the ten talkers reported speaking any other languages fluently and/or having any knowledge or contact with Spanish/Catalan in their daily routines. Talkers were recorded within a period of a week and were paid for their participation. All participants reported having normal vision and hearing. A summary of the talkers' biographical information can be seen in Table 5.5.

Table 5.5. Biographical information of the SBE Talkers

Talker	Age	Occupation	Gender	Place of birth	Place(s) of residence of the past 10 years
T1	23	Artist	F	Jersey, UK	Jersey/London, UK
T2	33	IT analyst	M	London, UK	London, UK
T3	24	Student	F	London, UK	London, UK
T4	28	Student	M	London, UK	London, UK
T5	24	Sales assistant	F	London, UK	London, UK
T6	25	Student	M	Croydon, UK	Croydon/ London, UK
T7	39	Unemployed	F	London, UK	London, UK
T8	32	Student	M	London, UK	London, UK
T9	26	Student	M	London, UK	London, UK
T10	24	Student	F	Jersey, UK	Jersey/ London, UK

Recordings took place in a soundproof chamber at the speech laboratory at UCL and each word was recorded three times, with additional repetitions whenever necessary. The recordings were carried out using *Cool edit 2000* software, a *Rode NT-1AX* microphone, an *Edirol UA25* audio interface and were digitized at a 44.1 kHz sampling rate and 16 bit quantification. Stimuli were embedded in carrier sentences and displayed to talkers by means of a PowerPoint presentation.

Nonsense words were elicited first. In order to ensure the desired pronunciation of the nonsense words a rhyming carrier sentence was used, that is, the carrier sentence contained a real word that rhymed with the target nonsense word, e.g., *It rhymes with "real word", "non-sense word". I say "non-sense word" now; I say "non-sense word" again*. A couple of sentences illustrating the elicitation procedure of the non-words can be seen in Figure 5.2. The real word stimuli were elicited by means of the following carrier sentence: *I say "real word", I say "real word" now, I say "real word" again*. Recordings took around 45-60 minutes per talker and all instances were closely monitored by the researcher in order to guarantee the desired pronunciation of each target sound. In a few instances, talkers were asked to repeat mispronounced words and non-words.

Figure 5.2. Sample elicitation carrier phrase for non-words

In order to obtain the actual word stimuli, each recording was annotated using *Praat* (Boersma, 2001). The CVC words and non-words were segmented and each individual token was extracted and saved as a wav file. All resulting stimuli were intensity-normalized to 70 dB and placed in between 75 milliseconds of silence intervals with the aid of Praat scripts.

5.3.2. Stimuli selection

Having recorded more instances and talkers than the necessary to fit the design (three instances of each item were recorded by ten talkers), a stimuli selection took place. The aim of this selection was twofold; it first aimed at selecting the best item for each word/non-word per talker, and, secondly, it aimed at selecting eight talkers only. In order to select the best stimuli, a 7-alternative forced-choice identification task was built for vowel sounds and two 6-alternative forced-choice identification tasks were built for consonants, one for initial and one for final consonants. The tests contained the three instances of each word and non-words produced by the ten talkers and were administered using the software *TP* v. 3.1. (Rauber, Rato, Kluge & Santos, 2012). Two researchers (the author of the study and an experienced phonetician) performed these tasks over the course of several days. They first labelled the sounds by using the response options /æ ʌ ɪ iː ɜː e ɑː/ and /p t k b d g/ for vowels and consonants respectively, and then rated how good the tokens were in a 3-point Likert scale (1 being "good", 2 "average" and 3 "poor"). All misidentified tokens were considered poor. Researchers agreed on most of the stimuli, and in cases of disagreement, the specific tokens were re-assessed. Given equal rating, the first instance of the target word was selected. If this was not possible, the second recorded instance of the target word was then given preference and the last elicited word would be chosen only in case both previous tokens were not good exemplars. Moreover, at this stage, four males and four female talkers were selected to fit the study design. The selection was made based on the researchers' impressions on each talker's voice quality and intelligibility[14]. See Table 5.4 above for a schematic summary of the distribution of talkers across testing and training tasks.

[14] The most clear and intelligible talkers (according to the researchers' opinion) were chosen for training, then for testing and subsequently for generalization.

5.3.3. Stimuli validation

Three native British English speakers performed the perception tasks (pre-test/post-test and generalization test) in order to validate the selected stimuli and provide a native speaker score as baseline data. Table 5.6 lists the biographical information of the three English native speakers (NES) who validated the stimuli. Moreover, the same NES performed two further perception tests, which made use of the training non-word stimuli that did not appear in the testing phase. In each case, there was a test for consonant stimuli and a test for vowel stimuli.

Table 5.6. NES Biographical information – Stimuli validation

Speaker	Age	Occupation	Gender	Place of birth	Place(s) of residence of the past 10 years
NES-1	54	EFL teacher	F	Oxford, UK	Barcelona, Spain
NES-2	53	Programmer	M	London, UK	Barcelona, ES
NES-3	34	EFL teacher	M	Irvine, UK	Glasgow, UK/Barcelona, ES

Note. M= male, F=female, EFL= English as a foreign language.

Two of the English native speakers were from the south of England and therefore, speakers of the SBE target variety. The third NES, who was originally from Scotland, had been largely exposed to SBE through teaching English as a foreign language (TEFL) materials and was instructed to consider this variety of English when identifying the target sounds. As the results for the three NES were comparable, it was assumed that the third native English speaker did not encounter any difficulties while labelling the stimuli. All NES had been living in Barcelona for at least 6 years at the time of the data collection and performed the tests in a quiet room by making use of good quality headphones (SONY ZX310) connected to a laptop computer. The NES' results for each test are presented in Table 5.7.

Table 5.7. Average correct identification scores of testing and training stimuli by NES

NES	Pre-test + Generalization	Training stimuli Vowel test	Training stimuli Consonant test
NES-1	99.37	97.20	98.35
NES-2	97.47	96.50	95.40
NES-3	96.20	95.37	95.20
Total	**98.42**	**96.35**	**96.32**

As can be seen in Table 5.7, the initial perceptual performance of the three native speakers in the identification tasks was higher than 95% for all tests, indicating that the testing and training stimuli were appropriately representative of each category tested. The misidentified stimuli were checked, replaced with other tokens whenever necessary and presented to the same native speakers on a following perception test until 100% of the stimuli were correctly identified. Then the whole set of materials were assessed again by the two researchers who had selected the first set of stimuli.

5.3.4. Training stimuli

Training material consisted of 576 trials in total (72 nonsense words x 4 talkers x 2 repetitions). Unmodified CVC nonsense words produced by four Southern British native talkers (two males and two females) were used in order to provide stimuli variability, which is characteristic of HVPT (see Table 5.4 in section 5.2 for the distribution of talkers among testing and training tasks). The decision to use nonsense words for training was driven by the need to eliminate any word knowledge bias, as nonsense words require the listener to rely solely on their auditory skills, as well as the need to create a balanced corpus of stimuli with same number of word pairs for each vowel and each consonant contrast, as explained below.

The objective of the training was to train participants explicitly on either the consonant contrasts or the vowel contrasts, but to expose all participants to both vowel and consonant contrasts across trials. To that effect every non-word contained one of the seven selected English vowels /æ ʌ ɪ iː ɜː e ɑː/ and one of the six English stop consonants /p t k b d g/ either initially or finally. In order to do so, quadruplets of non-words were created to obtain a balanced "cross-design"[15], so that all groups were presented with the exact same set of non-words during training. For instance, the quadruplet *vap*, *vup*, *vab*, *vub* was used to train the vowel trainees on the /æ/-/ʌ/ distinction within trials (*vap* vs. *vup*, *vab* vs. *vub*), while implicitly exposing them to the final consonant voicing distinction across trials (*vap* vs. *vab*, *vup* vs. *vub*). Moreover, the same quadruplet of non-words was used to train consonant trainees on the final consonant voicing distinction /p/-/b/ (*vap* vs. *vab*, *vup* vs. *vub*) while indirectly exposing them to the vowel /æ/-/ʌ/ distinction across trials (*vap* vs. *vup*, *vab* vs. *vub*). To provide another example, the quadruplet *teedge, tidge, deedge, didge* was used the same way to train either the high front vowel distinction or the initial consonant voicing counterparts /t/-/d/, while exposing the participants to the other feature across trials. Non-sense minimal pairs of the form CVC were generated so that the six target consonants (/p t k b d g/), were placed both on onsets and/or codas of the four vowel pairs /æ/-/ʌ/, /ɪ/- /iː/ and /ɜː/-/ɑː/ or /ɜː/-/e/. A total of 24 quadruplets were designed, containing 88 different non-words (some of the same /ɜː/ items were contrasted with /e/ and /ɑː/ words). The complete list of non-words is presented in Table 5.8. Hence, vowel training focused on vowel distinctions and consonant training focused on consonant distinctions, while the exact same set of stimuli were used in each case. Clearly, obtaining a sufficient number of "quadruplets" with real words would not have been possible, hence one of the motivations for the use of non-words.

The vowel /ɜː/, which is one of the target sounds of the present study, was trained by being contrasted with both /ɑː/ and/or /e/. This is so because minimal pairs generated with the combination of both these sounds might be possible sources of confusion to Spanish/Catalan

[15] Name attributed to the design used in this study, as learners were exposed to the different segments across trials.

L2 learners, especially in production (i.e. hurt-heart, bird-bed). Thus, 50% of the /ɜː/ stimuli contrasted this target vowel sound with /ɑː/ and the other 50% contrasted it with the vowel sound /e/. All words listed below formed part of the test; however, only the /ɑː/ stimuli produced by talkers T3 and T4 and the /e/ stimuli produced by talkers T5 and T6 were included. Table 5.8 presents the 88 different words forming the balanced quadruplets created and used to train vowels and consonants through a balanced cross-design. As can be seen, each quadruplet contrasted one of the target vowel pairs (/æ/-/ʌ/, /ɪ/-/iː/, /ɑː/-/ɜː/, /e/-/ɜː/), and at the same time, one of the target stops (/t/-/d/, /b/-/d/, /k/-/g/) either in initial or final position.

Table 5.8. Training stimuli organized by vowel contrast (columns) and consonant contrast (rows)

	æ - ʌ	**ɪ - iː**	**ɑː- ɜː**	**e - ɜː**
t/d_#	dadge-tadge	deege-teege	darge-targe	dedge-tedge
	dudge-tudge	didge-tidge	derge-terge*	derge-terge*
p/b_#	pav- bav	peedge-beedge	parsh-barsh	perf-berf
	puv -buv	pidge-bidge	persh-bersh	peff-beff
k/g_#	kak-gak	keedge-geedge	karch-garch	ketch-getch
	kuk-guk	kidge-gidge	kerch-gerch*	kerch-gerch*
#_t/d	zat -zad	jeet-jeed	zart-zard	chert-cherd
	zut - zud	jit-jid	zert-zerd	chet-ched
#_p/b	vap -vab	veep-veeb	jarp- jarb	jep-jeb
	vup - vub	vip-vib	jerp-jerb*	jerp-jerb*
#_k/g	vak -vag	veek-veeg	vark-varg	vek-veg
	vuk -vug	vik-vig	verk-verg*	verk-verg*

Note. _# - initial; #_ - final, * Items that appear twice on the list. Recall that some /e/ items appear twice as /e/ was contrasted with /ɑː/ in some pairs and with /ɜː/ in others.

5.3.5. Testing stimuli

Testing materials consisted of unmodified CVC nonsense words produced by a male speaker and a female speaker of southern British English (T1 and T2). Stimuli from these speakers were not used in the training corpus, so that testing already measured generalization to new talkers. The pre-test, post-test and the retention test were identical. Stimuli consisted of a subset of the non-words used at the training phase and involved 30 words (i.e. 5 target vowels x 6 words) of CVC nonsense words to test the perception of five southern British English vowels /iː ɪ æ ʌ ɜː/ and 24 words (i.e. 6 target consonants x 2 contexts (word initial and word final) x 2 words) of CVC nonsense words to test the perception of stop consonants /p t k b d g/, six word initially and six word finally. In addition, 19 non-words (12 for consonants and 7 vowels) were included as practice tokens in order to guarantee that the task procedure was understood and eight non-words involving the vowels /e/ and /ɑː/ were included as testing fillers. Real words, which are testing a type of generalization in this study (generalization 1), were included in the testing stimuli as there was the need to obtain an initial identification score for real word stimuli. Generalization to real words will be explained in more detail in section 5.4.2.2.2. The number of non-word stimuli used in the consonant and vowel pre-test, post-test and retention tests is presented in Table 5.9 and a complete list of testing stimuli can be found in Appendix D.

Table 5.9. Testing stimuli breakdown

	Words	Talkers (T1, T2)	Repetitions	Total number of trials
Practice items (V+C)	19	1 (T1)	1	19
Initial consonant non-words	12	2	2	48
Final consonant non-words	12	2	2	48
Vowels non-words	30	2	2	120
Fillers non-words	4	2	2	16

5.3.6. Generalization stimuli

As previously mentioned, this study looked at three different types of generalization: generalization to real words (generalization 1), generalization to novel non-words (generalization 2) and generalization to novel speakers with familiar non-words and novel non-words (generalization 3a and generalization 3b). Some studies consider the use of different talkers in testing and training already an example of generalization to new talkers (Iverson & Evans, 2009; Nobre-Oliveira, 2007; Pereira, 2014). This study assessed generalization to new talkers further by including "absolute new talkers" in order to avoid familiarization with talkers at pre-test. A complete list of the generalization stimuli can be found in Appendix E and the details are described next.

5.3.6.1. Stimuli for testing generalization to real words

The stimuli for testing generalization to real words constituted of 17 minimal pairs of CVC real words (five testing vowel perception and 12 testing consonant perception). In addition, eight real words containing the sounds /e/ and /ɑː/ were included as testing fillers. A stimuli breakdown for generalization to real words can be seen in Table 5.10.

Table 5.10. Generalization to real word stimuli breakdown

	Words	Talkers (T1, T2)	Repetitions	Total number of trials
Final consonant Real	12	2	2	48
Initial Consonants Real	12	2	2	48
Vowels Real	10	2	2	40
Fillers (Real)	4	2	2	16

5.3.6.2. Novel non-words generalization stimuli

Novel non-words were created in order to test if knowledge acquired during training could be generalized beyond the training tokens (generalization 2). The generalization stimuli were made of 68 non-words in total (34 novel non-words x 2 talkers). There were 24 new consonant words and 10 new vowel words produced by two familiar talkers from training (T3, T4) as shown in Table 5.11.

Table 5.11. Generalization to novel non-word stimuli breakdown

	Novel words	Familiar Talkers (T3, T4)	Repetitions	Total number of trials
Final consonant	12	2	2	48
Initial Consonants	12	2	2	48
Vowels	10	2	2	40

5.3.6.3. Novel talkers generalization stimuli

As previously mentioned, the testing stimuli (pre-test and post-test) were produced by two talkers (T1, T2) not present/used during training and this procedure aimed at testing generalization to new talkers (see section 5.3.5 for information on the testing stimuli). Additionally, two absolute new talkers (talkers that were not present at testing or at the training phases) were used in order to test if the improvement obtained during training could be generalized to stimuli produced by completely unfamiliar voices. The generalization stimuli were made of 68 non-words in total (34 familiar non-words x 2 talkers). There were 24 familiar consonant non-words and 10 familiar vowel non-words produced by two novel talkers as evidenced in Table 5.12.

Table 5.12. Generalization to novel talkers' stimuli breakdown

	Familiar words	Novel talkers (T7, T8)	Repetitions	Total number of trials
Final consonant	12	2	2	48
Initial Consonants	12	2	2	48
Vowels	10	2	2	40

5.4. Testing and training

This subsection describes the instruments, the individual tasks, and the procedure used for testing and training.

5.4.1. Perceptual testing and training tool

The perceptual testing and training were delivered by means of the software ***TP 3.1*** (Rauber et al., 2012). TP is a training freeware that provides feedback, which is a crucial feature in training studies. Feedback has been shown to be important for training as it enables subjects to determine whether what they are doing is appropriate or not, and decide to continue responding in the same way as before or change the way they are responding (Logan & Pruitt, 1995). In the current study, the trainees received immediate feedback on the directly trained segments after each trial indicating if their perceptual answer was correct or incorrect, and global feedback at the end of each session indicating the total number of hits and errors.

5.4.2. Testing tasks

Several tasks were used to test and train the L2 participants of this study. Perception was assessed by means of a forced choice identification task and production was assessed by means of a picture naming task. Two tasks were used in order to train L2 perception, namely a categorical AX discrimination task and a forced choice identification task. Additionally, the control group and the learners who were excluded at the pre-selection phase were engaged in a non-perceptual training regime, which involved the completion of phonemic transcription exercises as described below. Table 5.13 summarizes the general group/task design of the study.

Table 5.13. Group/task study design.

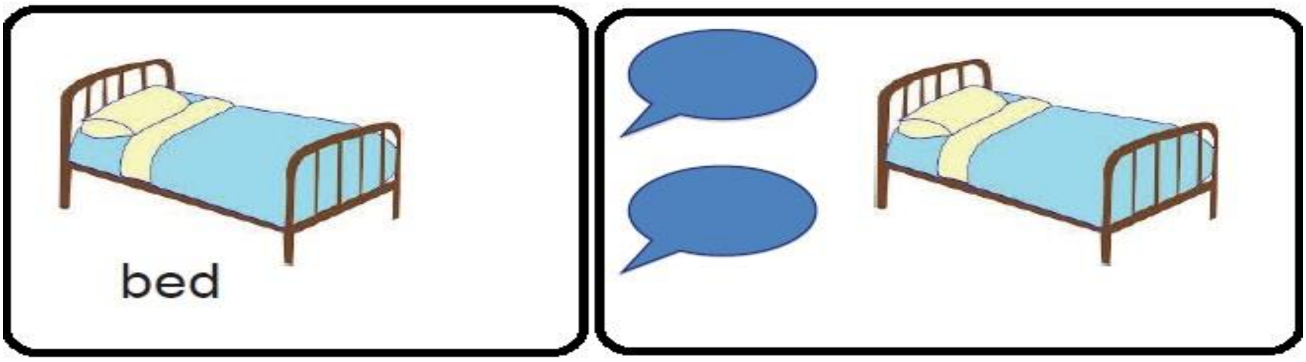
Group	Perceptual testing*	Training	Production testing
ID_V	FC_ID task (V + C)	FC_ID (V only)	Picture naming task (V + C)
DIS_V		AX_DIS (V only)	
ID_C		FC_ID (C only)	
DIS_C		AX_DIS (C only)	
Control Group**		Unrelated transcription exercises	

Note. V=vowels, C= consonants, FC_ID = Forced choice identification task, AX_DIS = same or different discrimination task.
* Testing refers to pre, post, generalization and retention tests.
**Control group and 46 students initially excluded at the pre-selection phase.

5.4.2.1. Production task

The learners' L2 production skills were assessed by means of a picture naming task before and after training (pre-test and post-test). Participants were asked to name 27 different pictures from a PowerPoint presentation (the complete set of pictures is presented in Appendix F). They were first presented with the visual (picture) and the written representation (word) of each target production. After one second, the written representation automatically disappeared and two speech balloons signalled and prompted the production of each token twice. This method was implemented in order to diminish the orthographic interference as much as possible (so that students would not simply read the words out), but at the same time, to ensure that the intended word was produced by the participants. Figure 5.3 illustrates the elicitation procedure of the practice word "bed".

Figure 5.3. Picture naming task used for production



Even though training was administered with nonsense words, learners' production skills were assessed by means of real words only. The decision of testing production by means of real words was motivated by the difficulty of eliciting the correct sounds embedded in non-words and the large number of students recorded. Out of the 27 different words elicited, three were practice words (*boat*, *bed* and *bird*) that were used to ensure that the task was understood and to adjust the recording volume. Some of the actual 24 test words were used to test the production of both vowels and final consonants. These words are marked with an (*) in Table 5.14, which presents all the production test words.

So that the phonetic context could not affect the production of the target sounds, each target vowel and consonant was presented in two phonetic environments equally represented in the elicitation list. Thus, the five target vowels were elicited in words involved in minimal pairs differing in the voicing of the final consonant, that is, each vowel appeared before a voiced consonant and before a voiceless consonant in the course of the production task. Similarly, final consonants were elicited following tense vowels and lax vowels equally, and initial consonants were elicited before low vowels and before mid and high vowels as can be seen in Table 5.14. The initial consonants were elicited before different vowels because voice onset time has been reported to be influenced by the height of vowels following the stop consonant (Higgins, Netsell & Schulte, 1998; Yavaş, 2007). According to Yavaş (2007), aspiration before high vowels is longer than before low vowels.

Table 5.14. Production elicitation list

Vowel words		Consonant words (Initial)		Consonant words (Final)	
_Vless	**_Vced**	**_Low**	**_High**	**Lax**	**Tense**
cap*	cab*	park*	pet	cap	carp
buck*	bug*	tart	tip	bit	feet
bit*	bid*	cap*	kilt	buck	park
feet*	feed*	bark	bet	cab	carb
hurt	heard	dart	dip	bid	feed
		gap	guilt	bug	league

Note. _Vless = preceding voiceless sound, _Vced = Preceding a voiced sound, _Low = preceding a low vowel, _High = preceding a non-low vowel.

Recordings took place at a sound attenuated chamber at the speech laboratory (*Servei de Tractament de la Parla*) at *Universitat Autònoma de Barcelona*. The recordings were carried out using *Adobe Audition CS5.5* software, a *Rode NT1-A* microphone, a *Cakewalk UA-25EX* audio interface and were digitized at a 44.1 kHz sampling rate and 24 bit quantification.

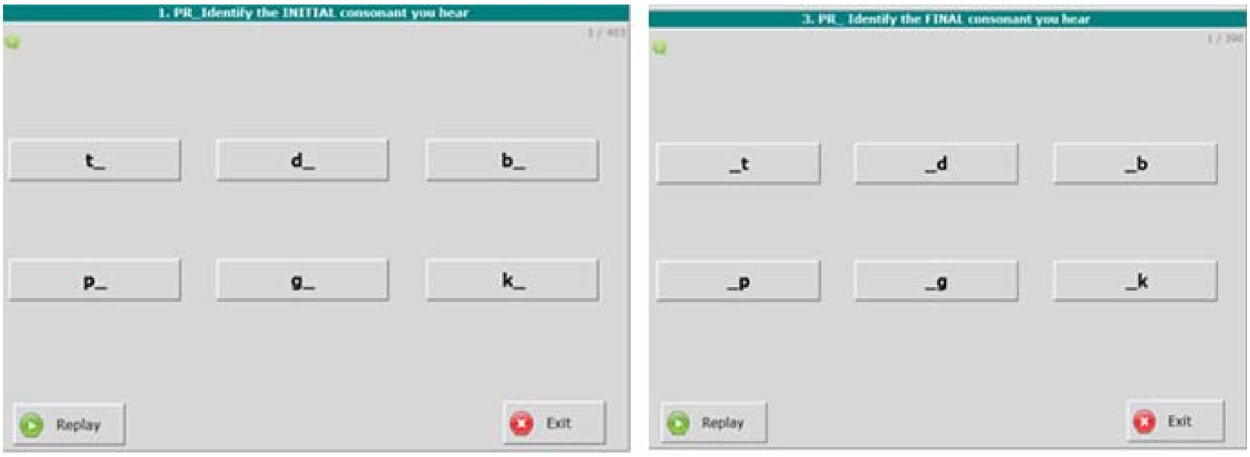
5.4.2.2. Perception tests

5.4.2.2.1. Pre-test, post-test and retention test

The pre-tests, the post-tests and the retention tests were identical and consisted of a six-forced choice consonant identification task (divided into two sections: one for initial and one for final consonants) and a seven-forced choice vowel identification task. The total number of trials including practice trials, test trials and fillers for the vowel and the consonant tests involving real and non-sense words was 403 trials (refer back to section 5.3.5 for information on the number and distribution of stimuli for each condition).

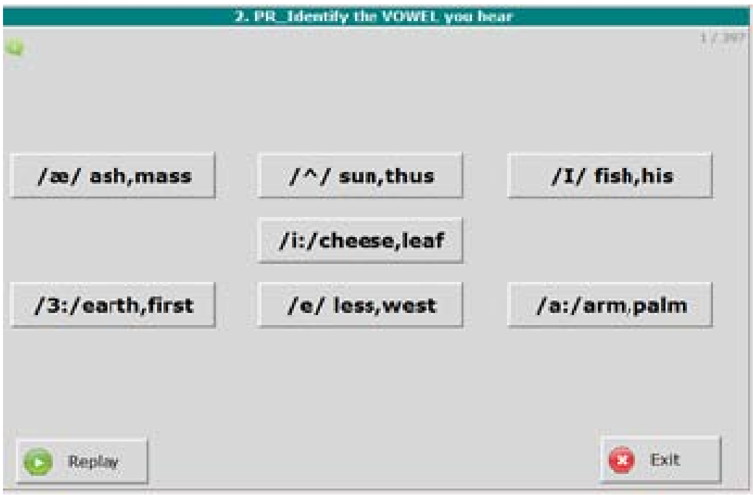
The different response options varied according to the test. The following labels *(p_, t_, k_, b_, d_, g_)* were used in order to assess initial consonant perception and *(_p, _t, _k, _b, _d, _g)* were used in order to assess final consonant perception. The underscore was used to signal the position of the remaining part of the word and therefore to signal if the target sounds were word-initial or word-final, as evidenced in Figure 5.4.

Figure 5.4. Consonant response options



As regards to the vowel test, students were provided with phonetic-like symbols (TP software does not accept phonetic fonts) and two exemplar real words for each sound tested. Hence, the seven response options used were: */æ/ ash, mass, /ʌ/ sun, thus, /i:/ fish, his, /i:/cheese, leaf, /3:/earth, first, /e/ less, west and /a:/arm, palm,* as shown in Figure 5.5. Students were provided with the correct phonetic symbols during the instruction and before they started the tests. The reason for using exemplar real words for the vowels and not for the consonants lies in the fact that there is a relatively clear grapheme-phoneme correspondence for stop consonant sounds (with “c” being the only exception), whereas the correspondence is less evident for vowel sounds. The words used to illustrate each sound did not appear in the test and were carefully selected by trying to minimize the use of stops as CVC onsets or codas, so they would be as dissimilar to the stimuli as possible. All words, except for *palm*, fit the desired criteria[16]. The participants were informed that the response words did not appear in the test and they should only be used as reference to each individual target vowel sound.

Figure 5.5. Vowel test and response options



Each test lasted approximately 25-35 minutes. Participants were provided with a 19-trial practice session (seven trials for vowels, six trials each for initial and final consonants), before the task itself began. The order of the trials of the test and the practice trials were randomized. Moreover, the order of the tasks was counterbalanced across participants: half of them performed the initial consonant identification first, followed by the vowel identification test and the final consonant identification test, while the other half started with the final consonant test and ended with the initial consonant test. The vowel test was always kept in between the two consonant tests. Students were asked to read the instructions before starting

[16] Palm was kept in the attempt to have a commonly known monosyllabic word to represent this sound. Moreover, as the /ɑ:/ sound was a filler, this token was not expected to influence the research outcome.

each individual task (the precise instructions are presented in Appendix G). Participants could take a short break at any time during the test and were actually encouraged to do so to avoid fatigue effects. No feedback was given throughout the test; however, learners were provided with their global number of hits and/or errors at the end of each session.

5.4.2.2.2. Generalization tests

Generalization of learning was examined by testing the participants' identification of the target sounds embedded in real words (Generalization 1), new non-words (Generalization 2), and non-words produced by new and absolute new talkers (Generalization 3a and 3b, respectively). Generalization to new non-word stimuli and to novel talkers were presented to participants in a separate test, which took place on week nine and contained 272 trials in total (136 for novel non-words and 136 for "absolute new talkers". Refer back to section 5.3.6 for details on the number of stimuli for each condition). The reason for having the generalization test as a separate test was motivated by the fact that this study aimed at testing novel non-words and absolute novel talkers that had not been previously heard at any other stage of the study. All types of generalization were assessed by means of the same forced choice (FC) identification tasks described in the previous section: a six-forced choice identification task for consonants and a seven-forced choice identification task for vowels.

5.4.3. Training

As mentioned before, there were four different experimental groups and a control group and all of them went through a different type of training. For the experimental groups, training consisted of five sessions of perceptual training administered using TP software. All subjects were aware that they were being trained on either vowel or consonant sounds. For the control group, a non-perceptual training, which consisted of five sessions of transcription exercises performed on an online platform, was designed. All training methods are described in more detail next.

5.4.3.1. Identification training

The ID groups were trained by means of FC identification tasks involving 288 stimuli (72 words x 4 speakers) and two repetitions of each stimulus, totalling 576 trials. Similarly to Flege (1995b), the motivation for including two repetitions was to guarantee that both training methods (ID and DIS) exposed trainees to the exact same number of stimuli, as two words are heard at each trial in a discrimination task. The task used and the response options were the same as in the pre-test, post-test and retention (see section 5.4.2.2.1 for more details). Participants listened to one stimulus at a time and were asked to label it using one of the response options provided. There were two identification-trained groups, one trained on consonants and one on vowels. The exact same words were used; however, the consonant group was asked to attend to and identify the initial and final consonants that were part of the stimuli, whereas the vowel groups were asked to attend to and identify the vowel sounds present in the stimuli. During training, participants received immediate feedback on the segment they were attending to, and thus, being directly trained on.

5.4.3.2. Discrimination training

The DIS groups were trained by means of AX categorical discrimination tasks (e.g., Flege, Munro & Fox, 1994) consisting of 288 trials (2 words presented every trial). Participants listened to two stimuli (A and X) and had to indicate whether A and X belonged to the same vowel or consonant category (e.g., tadge (A), tudge (X) – different; or tadge (A), tadge (X) – same). Each trial contained words from two different talkers and the members of each contrast alternated in the two positions (A and X). This procedure resulted in four possible combinations for each pair (e.g., T1/æ/-T2/ʌ/; T2/ʌ/-T1/æ/; T1/ʌ/-T2/æ/ and T2/æ/-T1/ʌ/). The interstimulus interval (ISI) was 1.15 seconds. The ISI is a very important feature in discrimination tasks, as the shorter it is, the more reliance on the sensory memory capacity may take place (Pisoni, 1973; Logan & Pruitt, 1995). Conversely, a longer ISI might force perception at the phonemic rather than at the acoustic level (Højen & Flege, 2006), which is said to promote L2 categorization (Strange, 1992). Thus, a high ISI value was set with the intent of assuring perception at the phonemic rather than at the acoustic level (Højen & Flege, 2006) and in order to make the task comparable to the ID task. In order to perform this task, it becomes essential that listeners make use of their internalized categories rather than relying on their sensory memory.

As previously mentioned, there were four different talkers in each training session (T3, T4, T5 and T6), which is a characteristic of high variability phonetic training. Thus, the total number of talker combinations for the AX DIS task was 12 (six talker pairings and two orders), resulting in 1728 (864 "same" and 864 "different") stimuli for vowels (3 vowel pairs[17] (/æ/-/ʌ/, /ɪ/-/iː/, /ɑː/-/ɜ/, /e/-/ɜ/) x 12 words x 12 talker combinations) and 1728 (864 "same" and 864 "different") stimuli for consonants (6 stops (p t k b d g) x 2 word position (initial, final) x 12 tokens x 12 talker combinations). As the total number of stimuli was too large to include in one session, three different discrimination tests were created by dividing the twelve talker combinations into three groups (each containing all four talkers), as shown in Table 5.15. In order to maintain the same number of stimuli presented at each session comparable to the identification training, the tokens were halved and therefore 144

[17] Items for /ɜː/ were halved, so the combination of (/ɑː/-/ɜː/, /e/-/ɜː/) is considered as one vowel pair only.

"different" and 144 "same" trials were used in each training session. As previously mentioned, the present study was designed to have six training sessions, and thus, expose the learners to each talker combination group twice. However, as only five sessions took place due to unforeseen reasons (i.e. a student strike), talker combination 3 (AX session 3) was performed only once by the participants, whereas the other two were performed twice.

Table 5.15. Distribution of talker combinations per session in the AX discrimination task

Talker combination 1 AX session 1	Talker combination 2 AX session 2	Talker combination 3 AX session 3
T5 +T6	T5+T3	T5+T4
T6+T5	T3+T5	T4+T5
T3+T4	T4+T6	T6+T3
T4+T3	T6+T4	T3+T6

5.4.3.3. Post training survey

Immediately after the last training session, the four experimental groups completed a post training survey[18] administered via Google forms, which aimed at gathering some information regarding students' impressions on the training methodology and their opinions about phonetic training as an EFL tool in general. This study is innovative in the sense that it directly compares two training methods (ID vs. categorical DIS) to train different segments (vowels and consonants) in a single study. Therefore, it was deemed relevant to compare students' opinions about the two training methods and at the same time assess students' impressions about consonant training and vowel training. Previous research involving the voicing contrast in English final stops reported that identification training was better

[18] The survey idea was inspired by previous work (Rato, 2014), and some of the questions were adapted from the same source.

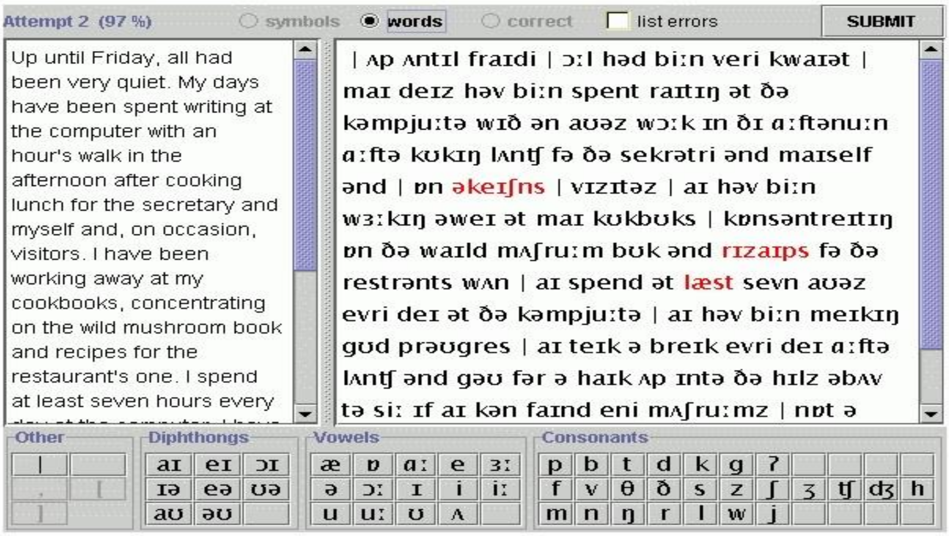
perceived by students than AX discrimination training (Flege, 1995b). This study seeks to evaluate whether this finding is replicated and if it extends to initial consonants and vowels.

5.4.3.4. Transcription training

Two groups of students, namely the control group and the 46 students whose L1 was not Catalan or Spanish, performed phonetic transcription exercises on an online platform, *The web transcription tool* (Cooke, García-Lecumberri, Maidment & Ericsson, 2005), in order to obtain course credit for their participation. After being pretested, the controls performed five online sessions (the same number of sessions as the experimental groups) and were asked to return to the lab in order to perform the post-test and the separate generalization test. The other group of students performed a total of seven online sessions in order to complete the extra-curriculum course work for this subject, as they were not required to return to the lab at the post-test phase. The first session required participants to transcribe words, which did not contain any of the target stop consonants or vowels and was carefully created by the researcher. The subsequent sessions (2-7) required participants to transcribe texts chosen from a database provided by *The web transcription tool*. Phonetic transcription exercises were chosen as the activity for the control group to complete in order to ensure that students were not exposed to extra oral practice, which could interfere with the perceptual outcomes of the present study.

The online platform provided students with automatic feedback on their submitted attempts by highlighting mistaken symbols and words directly in the transcription output (Figure 5.6) and then providing them with lists of frequent errors from each session. Being the registered tutor on the platform, the researcher had access to students' task completion, grade obtained and mistakes made at each session performed. This data will not be analysed as part of the present study.

Figure 5.6. Sample feedback provided by *The web transcription tool* (Cooke et al., 2005)



5.5. Procedure

All perceptual testing and training tasks took place in the computer room at the speech laboratory (*Servei de Tractament de la Parla*) at UAB, and were always closely monitored by the researcher. There were only eight computers available, so students were performing the sessions in small groups. Each participant signed up for a weekly slot of 35 minutes and was expected to attend these regular sessions as part of the coursework for the subject they were enrolled in. Due to the large number of students, communication between the researcher and students was facilitated by a Moodle platform created especially for this training study. Several procedures and/or any changes on the schedule (i.e. due to strikes or class cancelations) were posted online and students were immediately notified via e-mail. A personalized lab file, in which students wrote down their global scores obtained in each session, was used in order to keep track of students' attendance. At the end of each day, all absent students were contacted and a make-up session was arranged.

5.6. Analysis

5.6.1. Analysis of perceptual data

Overall correct identification means for each participant and then per group at each testing phase (pre-test, post-test, generalization and retention test) were first calculated. Subsequently, gain scores were calculated by subtracting the pre-test score value from the test being analysed, resulting in the following criterion: *X test score minus pre-test score*, where X = gain. Thus, the main gain score upon completion of training was calculated as *post-test score minus pre-test score*, the generalization gain score was calculated as *generalization test score minus pre-test score* and the retention gain scores resulted from subtracting the pre-test scores from the retention scores, that is *retention scores minus pre-test scores*. The decision of having the pre-test as the baseline data is motivated by the fact that at pre-test all groups were homogeneous in term of their global scores, thus allowing the comparison of the different groups' post training scores to their pre-test scores, where all groups were comparable. All gain scores were submitted to statistical analyses and will be reported in chapter six, part I.

5.6.2. Analysis of production data

The L2 production data was analysed by means of native English speaker judgments. As having access to native speakers living in an English speaking country was not possible at the time, a decision was made to recruit a uniform population of English-speaking raters in

Spain. Twelve native speakers of British English residing in Barcelona at the time of the experiment assessed the learners' productions by means of identification and goodness ratings tasks.

5.6.2.1. Raters

The raters were English teachers living in Spain at the time of testing and who had been teaching English for a number of years. Due to the fact that all raters were working in the same language centre in Barcelona at the time of testing and had previously received a similar teaching training, their rating abilities were expected to be comparable. Table 5.16 displays the biographical information of the English-speaking raters.

Table 5.16. Production data raters

Speaker	Age	Occupation	Place of birth	Place(s) of residence of the past 10 years
RT1	33	EFL teacher	Derby, UK	Derby, UK/ Barcelona, ES
RT2	34	EFL teacher	Colchester, UK	Colchester, UK/ Barcelona, ES
RT3	29	EFL teacher	London, UK	London, UK/ Barcelona, ES
RT4	45	EFL teacher	Ashington, UK	Barcelona, ES
RT5	43	EFL teacher	Manchester, UK	Manchester, UK/ Barcelona, ES
RT6	45	EFL teacher	Croydon, UK	Croydon, UK/ Barcelona, ES
RT7	55	EFL teacher	Oxford, UK	Barcelona, ES
RT8	36	EFL teacher	Newcastle-Upon-Tyne, UK	Newcastle-Upon-Tyne, UK /Barcelona, ES
RT9	45	EFL teacher	York, UK	York, UK/Barcelona, ES
RT10	34	EFL teacher	Bath, UK	Bath, UK/ Barcelona, ES
RT11	36	EFL teacher	London, UK	Barcelona, ES
RT12	45	EFL teacher	Derby, UK	Derby, UK/ Barcelona, ES

Note, RT = Rater; EFL = English as a foreign language

5.6.2.2. Identification and goodness rating task

As explained in section 5.4.2.1, the production stimuli list contained 34 tokens, two per target sound (5 vowels, 6 initial consonants, 6 final consonants), and it was elicited twice (pre-test and post-test) for each of the 89 participants, totalling 6052 items to analyse. Given the great amount of L2 productions, a decision was made to limit the analysis of consonants to the sounds that have previously been identified as possible sources of mispronunciation for Spanish/Catalan learners of English (Cebrian, 2000; Aliaga-García & Mora, 2009; Mora & Fullana, 2007), that is, the initial voiceless stops and the final voiced stops. This resulted in 22 tokens, two per target sound (5 vowels, 3 initial consonants, 3 final consonants), per participant, totalling 3916 tokens (89 participants x 2 test times x 22 words). As this number was still too large for a single rating test, the 3916 words were distributed among 15 different perception tests. There were seven different perception tests for vowels containing 266-268 tokens and eight perception tests for consonants (4 for initial consonants and 4 for final consonants) containing 254-256 tokens each. In order to evaluate all the productions while limiting the amount of tests per rater to a reasonable number, each rater was assigned 1/3 of the speech samples and each individual production was evaluated by four different raters. This is a number of raters that is comparable to that of previous training studies (e.g., Iverson et al., 2012; Hardison, 2014).

The native listeners thus performed five forced choice identification tasks with goodness of fit ratings, created with Praat. A 6-alternative forced-choice task was implemented for both initial and final consonant tests and a 7-alternative forced-choice task was implemented for vowel sounds. Raters were asked first to identify the sound they heard and then to rate it on a 9-point Likert scale, where 1 meant “hard to identify as the selected sound” and 9 “easy to identify as the selected sound”. The response options used for the consonants involved the corresponding phonemes /p t k b d g/, whereas the response options used in the vowel tests contained two sample words representing each sound. The choice of words was motivated by the need to have each sound followed by a voiced and a voiceless consonant, in line with the context in which the target vowels were elicited. Thus, the response labels were: *"cheese/leaf", "dig/lick", "egg/less", "mud/nut", "lad/rat", "serve/earth" and "psalm/parse"*. The instructions given to the raters for the consonant and the vowel identification and rating tasks are given in Appendix I, which also illustrates the visual display of response alternatives for these tasks.

The tests were presented to the raters individually in a quiet room, using good quality headphones (SONY ZX 310) connected to a laptop. Raters were asked to adjust the volume of the headphones at a comfortable listening level before performing the tests. A short practice session containing six stimuli was performed before the rating task in order to guarantee task comprehension and to clarify any questions. Short breaks were included after every 80 stimuli and raters were notified about the break by means of an automated message. In order to calculate the scores for each token for each participant, the correctly identified tokens were attributed the value [1] and the misidentified tokens were attributed the value [0]. Subsequently, these values were multiplied by the rating score received for each production token. For instance, if a correctly identified token had received a rating score of [7], the calculation would be as follows: [1 * 7 = 7]. However, all misidentified tokens received a value of [0] as the result of the calculation [0 * X = 0]. This calculation was done for every token and applied to the ratings of every different rater. Subsequently, the median values across the four raters for each token for each participant were calculated. Inter-rater reliability analysis was assessed by using an intra-class correlation coefficient (ICC) test with an "absolute agreement" level. Having checked that the inter-rater reliability tests resulted in high values of consistency and absolute agreement among the four raters (see results section

6.7 for more details), the median value for each token at pre-test and post-test was calculated and subsequently the production gain scores were obtained by subtracting the pre-test scores from the post-test scores.

5.6.3. Analysis of qualitative data (post-training survey)

A qualitative analysis was carried out on the post training survey that was administered with the experimental groups immediately after the training regime. The 73 trainees completed the survey, and the results are presented and discussed regarding the difficulty, likeability and usefulness of each training method as reported by the students' opinions.

6. Results

The results will be divided into four main parts, part I, II, III and IV. Part I begins with the results obtained in the perception tasks (pre-test and post-test), followed by the generalization and delayed test results presented in Part II. Then, Part III will display the results for the production task and a comparison of perception and production results. Part IV will end with the qualitative results of the post-training survey. For the sake of clarity, a discussion section will follow each set of results (discussion I, discussion II, discussion III and discussion IV) and a separate chapter (chapter 7) will present a global discussion of the results, tying together the main findings of the current study with the pre-existing findings from the theoretical framework and the subsequent predictions.

In each case, the performance with trained sounds and untrained sounds is reported for the vowel-trained groups first, followed by the performance with trained and untrained sounds for the consonant-trained groups. The performance of the trained groups is compared to that of the control group in all cases. The data presented in this thesis will be analysed by means of generalized linear mixed effect models (GLMM)[19], which were chosen due to the fact that they combine fixed and random factors in the linear predictor. The linear analyses were conducted on the average scores obtained by each participant for each segment type (initial consonant, final consonant, vowels) and therefore participant was the random factor entered in the model consistently throughout the analyses presented in this chapter. It is important to note that in order to fit the aim of the present study, vowels and consonants will generally be treated as global measures. Originally, initial and final consonants were considered to fit a single global measure for consonants and this explains the fact that the

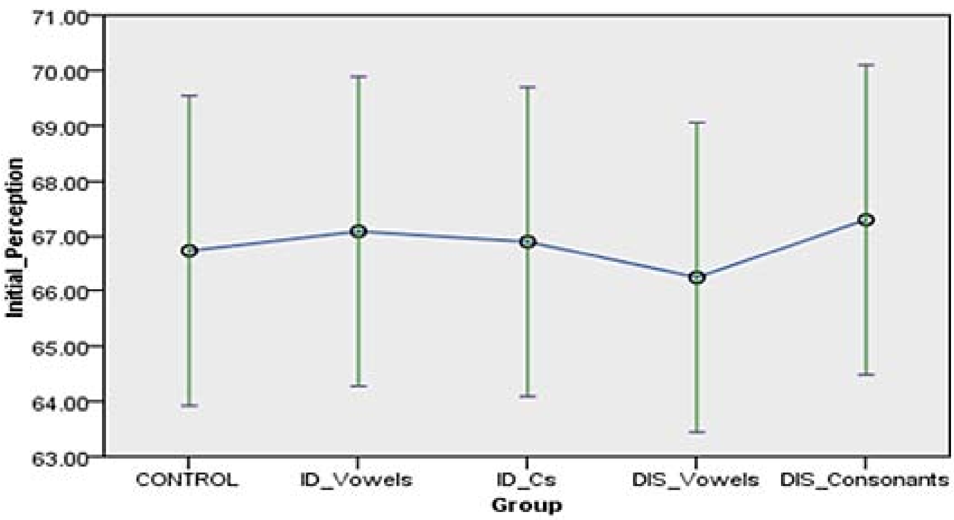
[19]The generalized linear model expands the general linear model so that the dependent variable is linearly related to the factors and covariates via a specified link function. Moreover, the model allows for the dependent variable to have a non-normal distribution. It covers widely used statistical models, such as linear regression for normally distributed responses, logistic models for binary data, log linear models for count data, complementary log-log models for interval-censored survival data, plus many other statistical models through its very general model formulation" (IBM Corp, 2013: 57).

consonant subcategories (initial and final) involved half the number of trials than the vowel categories. However, as consonants patterned differently depending on their position in the word, each subcategory was subsequently analysed as a separate measure.

6.1. Pre-test scores and group division

The global pre-test perception scores were used in order to divide participants into five homogeneous groups, N=20 each. In order to confirm if the performance of the groups was similar before training, a GLMM (Generalized linear mixed model) was conducted on the mean identification scores obtained by each group, which is the mean across all segment types for both real and non-words. The results confirmed that the five groups were homogeneous from the outset as no statistically significant difference was found among them ($F(4, 95) = .078$, $p > .05$). Figure 6.1 illustrates the performance of all five groups before training and their homogeneity.

Figure 6.1. % Global correct identification scores (across sounds) for each group before training



Part I – Perception results

6.2. Identification test results for trained and untrained sounds

The first research question was concerned with the direct and indirect effect of the HVPT methods here investigated, that is, the effect obtained on the trained and untrained segments by each training group. The direct and indirect effect of two HVPT methods (Categorical DIS and ID) will be reported by contrasting the improvement obtained by the trainees from pre-test to post-test on trained and untrained sounds embedded in nonsense words with the control group's improvement on the same segments. Results for the vowel training groups will be shown first, followed by the results for the consonant training groups. In all cases, percentage correct identification for each test will be shown first as a way of displaying descriptive statistics, and subsequently gain scores (i.e., difference between post-test and pre-test scores) will be presented and submitted to statistical analyses.

6.2.1. Vowel training groups

Two groups were trained on vowels, one by means of an identification task (ID_V) and one by means of a discrimination task (DIS_V). The results obtained on vowel and consonant perception by these two groups at pre-test and post-test will be described next.

6.2.1.1. Post-test results

Correct identification scores at pre-test and post-test and gain scores (difference between post-test and pre-test) were calculated for the two groups trained on vowels (ID_V, DIS_V) and the control group, and are shown in Table 6.1.

Table 6.1. Pre-test, post-test and gain scores for trained and untrained sounds in non-words obtained by the vowel trained groups and the control group

	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.1	9.9	55.5	6.5	52.9	9.5
POST	57.8	10.2	65.3	9.7	79.1	13.3
GAIN	3.7	4.2	9.8	7.6	26.3	7.6
Untrained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	74.3	9.7	77.6	11.9
POST	79	13.4	76.7	11.4	74.2	11.9
GAIN	0.9	5.6	2.4	10.3	-3.4	9.4
Untrained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	64.7	14.5	71	6.4
POST	69.6	6.29	72.4	6.9	69.4	10
GAIN	0.5	4.5	7.7	13.2	-1.6	8.2

As observed in Table 6.1, the three groups performed numerically better at post-test than a pre-test on the identification of the trained segment, vowels. This is particularly evident in the case of the ID_V group (26% gain), and also observed with the DIS_V group (9.8%). The numerical improvement obtained by the control group (3.7%) may reflect the learning outcomes of the formal phonetics learning, or of the general exposure to English in this and other courses, that took place between the pre-test and the post-test phases. The results for the untrained sounds are less notable and less consistent, as only the DIS_V group appears to show some improvement (7.7% with final consonants). It is relevant to note that, overall, learners appear to identify consonants better than vowels, as all groups present higher scores on initial and final consonant perception than on vowel perception from the outset of the study.

6.2.1.2. Gain scores on trained and untrained sounds

As the groups did not differ statistically at pre-test (refer back to section 6.1), the effect of training was further explored by comparing the amount of gain (i.e., the difference between post-test and pre-test scores), as shown in Table 6.1 above. The amount of gain in the L2 perception of trained sounds (vowels) for the two experimental groups falls within the range of improvement typically reported in the phonetic training literature, that is, around 10% or more (Jamieson & Moroson, 1986; Flege, 1989; Logan & Pruitt, 1995; Flege, 1995b; Iverson & Evans, 2009).

The gain scores for trained and untrained sounds were submitted to a generalized linear mixed-effects model (GLMM), in which the fixed effects were group (ID, DIS and CG), segment (vowels, initial consonant and final consonant) and a group by segment interaction. The analysis revealed a significant main effect of group ($F(2, 153) = 6.32$, $p < .01$), and segment ($F(2, 153) = 46.83$, $p <.01$) and a significant group by segment interaction ($F(4, 153) = 22.11$, $p <.01$). The group effect found by the model is related to the fact that the control group performed differently from the experimental groups. In fact, sequential Bonferroni pairwise comparisons confirmed that the two experimental groups outperformed the controls on the overall perception of L2 segments ($p <.01$ in both cases). This is illustrated in Figure 6.2 below.

The effect of segment can be explained by the fact that the trained segments (vowels) obtained higher gains than the untrained segments (initial and final consonants). This was confirmed by sequential Bonferroni pairwise comparisons (p <.01 in both cases). No significant difference was found between the initial and final consonant performance (p >.05). Figure 6.3 illustrates the overall segment perception across all groups.

Figure 6.2. Overall gain for non-words obtained by vowel trained groups and control group

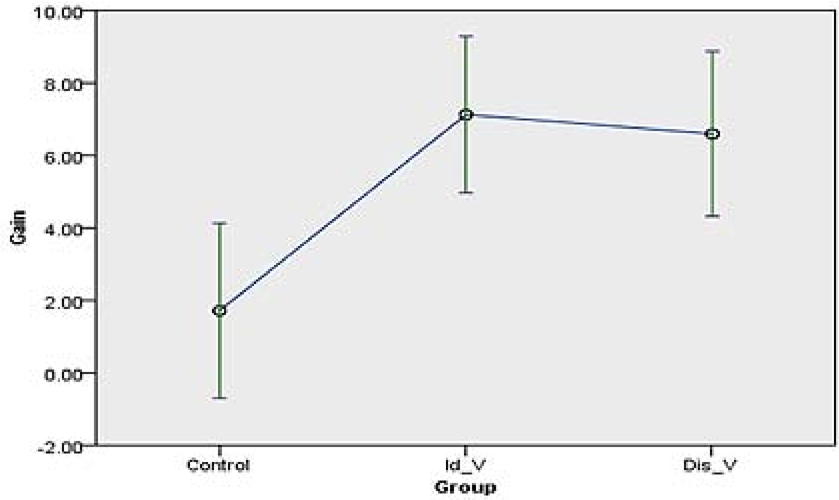
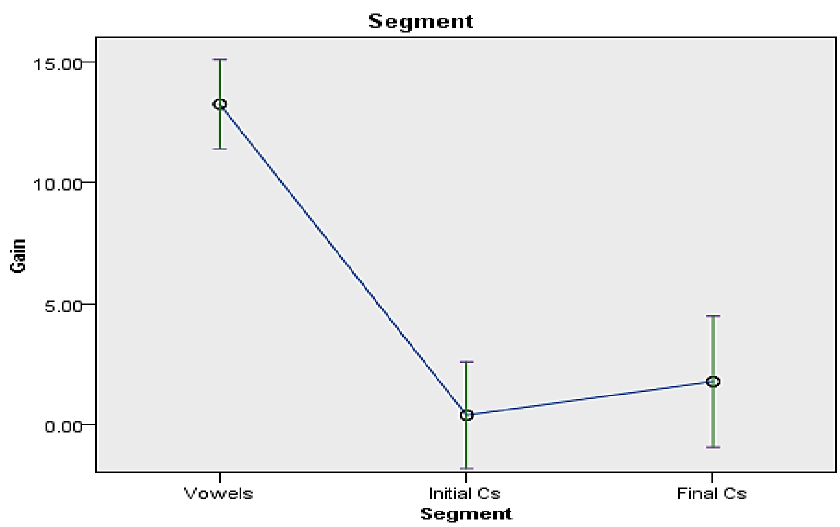
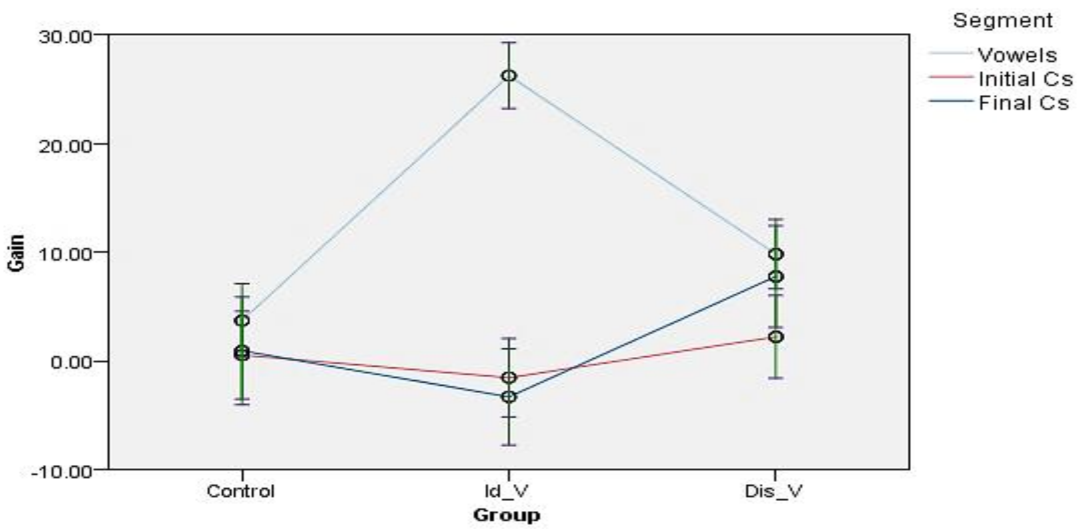


Figure 6.3. Overall gain for non-words for each segment type across vowel trained groups and control group



As for the group by segment interaction, it can be observed that training affected sounds differently for each training group. Specifically, the interaction is explained by the identification group's much greater difference in gain between vowels and consonants in comparison to the other groups. This is illustrated in Figure 6.4, which compares the perception on trained and untrained sounds by each of the three groups. Given the significant interaction, further simple effects analyses were carried out on the results for each segment type separately and are reported next.

Figure 6.4. Gain score for trained and untrained sounds in non-words obtained by vowel trained groups and control group



6.2.1.3. Gain scores on trained sounds (vowels)

The results for the trained segments (vowels) yielded a significant main effect of group ($F(2, 153) = 53.29$, $p < .01$), and pairwise comparisons with a sequential Bonferroni correction revealed that both groups (DIS_V and ID_V) significantly outperformed the controls ($p < .01$ in both cases) in their improved identification of vowels and that the ID_V significantly outperformed the DIS_V ($p < .01$). These results suggest that whilst both training methods are effective for training vowel perception, identification training may be superior to discrimination training in directing the learners' attention to specific vowel sounds.

This study evaluates the effect of training on trained and untrained sounds, represented by global measures of improvement across a set of vowels and a set of consonants. Thus, a detailed analysis of the effect of training on each individual vowel and consonant lies beyond the objective of this investigation. Nevertheless, the data allow us to explore if individual vowels followed a general similar pattern. Therefore, for the sake of completion, descriptive statistics were calculated for each individual sound, and will be presented at the end of each section. The mean identification scores at pre and post-test and the respective gain scores for each individual vowel for the vowel training groups and the control group are shown in Table 6.2.

Table 6.2. Gain scores for individual vowel sounds (trained segment) in non-words obtained by the vowel trained groups and the control group.

	CONTROL			DIS_V			ID_V		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
æ	39.8 (17.6)	40.8 (19.9)	1.0 (20.6)	42.8 (18.7)	50.6 (25.0)	7.8 (20.0)	31.0 (21.4)	77.7 (20.1)	46.6 (20.3)
ɜː	39.1 (21.0)	41.4 (19.6)	2.3 (12.8)	47.9 (23.7)	68.1 (20.8)	20.1 (21.8)	39.0 (17.8)	81.5 (17.0)	42.5 (17.4)
ɪ	69.0 (15.7)	80.2 (14.2)	11.2 (17.1)	72.0 (15.7)	81.0 (14.4)	9.0 (21.2)	75.0 (14.3)	82.9 (13.6)	7.9 (18.7)
iː	67.4 (15.6)	64.5 (19.2)	-2.9 (13.3)	61.1 (13.3)	60.4 (16.1)	-0.7 (18.9)	65.8 (12.9)	77.9 (13.5)	12.1 (17.7)
ʌ	55.2 (22.7)	62.0 (21.0)	6.8 (12.5)	53.4 (12.2)	66.2 (18.6)	12.7 (16.3)	53.5 (20.6)	75.6 (18.3)	22.1 (14.3)

It is interesting to note that some vowels seemed to improve more than others. At pre-test, on the whole all groups had the greatest difficulty identifying /æ/ and /ɜː/, followed by /ʌ/, while /iː/ and /ɪ/ were better identified. The ID_V group is the group that improved the most, and also the one that obtained more comparable results across vowels at post-test (76-83%, compared to 51-81% for DIS). Misidentification errors generally involved /iː/-/ɪ/ and /æ /- /ʌ/ confusions, while /ɜː/ was most often misheard as /ʌ/ or /e/. The improvement seen with the control group, mostly for the sounds /ɪ/ and /ʌ/ might be the result of the formal phonetics instruction and the phonological awareness of the new English sounds. Generally, the ID trainees obtained numerically higher scores than the DIS trainees, who in turn also seemed to outperform the CG group, in line with the global results across vowels previously described.

6.2.1.4. Gain scores on untrained sounds (initial and final consonants)

Regarding the untrained sounds, whilst the results of the follow-up simple effects tests revealed no significant effect of group for initial consonant segments ($F(2, 153) = 1.01$, $p > .05$), the results for final consonants yielded a significant effect of group ($F(2, 153) = 5.82$, $p < .01$). Pairwise comparisons revealed that the DIS_V had a significantly greater gain than both the controls ($p < .05$) and the ID_V ($p < .01$) in their perception of final consonant stimuli. It is relevant to mention, however, that the DIS_V group had a lower score for final consonants at pre-test, so the improvement here observed is partly connected to this fact. Interestingly, the scores for the three groups are numerically similar after training (69-72%). Tables 6.3 and 6.4 present the individual consonant sound performance at pre-test, post-test and the corresponding gain scores obtained by the vowel training groups and the control group.

It is interesting to note (Table 6.3) that all three groups were at ceiling in their perception of initial voiceless stops from the outset of the study. Hence basically no improvement was observed for these sounds after training. Identification of initial voiced stops posed a greater challenge, as scores ranged from 31%-70% at pretest and remained between 38%-70% at post-test. Further, some voiced stops seemed to be harder to identify than others, with the alveolar voiced stop /d/ being the sound most frequently misidentified by all three groups at both testing times.

Table 6.3. Gain scores for individual initial consonant sounds (untrained segment) in non-words obtained by the vowel trained groups and control group

	CONTROL			DIS_V			ID_V		
	PRE	POST	GAIN	PRE	POST	GAIN	PRE	POST	GAIN
SOUND	% (SD)	% (SD)	% (SD)	% (SD)	% (SD)	% (SD)	% (SD)	% (SD)	%(SD)
b_	70.3	69.5	-0.8	70.8	64.6	-6.3	64.4	50	-14.4
	(27.0)	(25.4)	(14.4)	(25.0)	(25.5)	(29.1)	(29.3)	(32.7)	(22.6)
d_	46.9	43.8	-3.1	31.3	38.2	6.9	47.5	40.6	-6.9
	(29.8)	(31.6)	(16.7)	(33.0)	(36.8)	(27.1)	(31.8)	(35.1)	(33.8)
g_	56.3	62.5	6.3	45.8	58.3	12.5	56.9	58.8	1.9
	(27.8)	(31.0)	(15.1)	(21.9)	(26.8)	(18.6)	(22.4)	(22.3)	(17.3)
p_	98.4	99.2	0.8	100	99.3	-0.7	98.8	99.4	0.6
	(4.3)	(3.1)	(5.5)	0.0	(2.9)	(2.9)	(5.6)	(2.8)	(6.3)
t_	98.4	99.2	0.8	100	99.3	-0.7	98.8	96.9	-1.9
	(4.3)	(3.1)	(5.5)	0.0	(2.9)	(2.9)	(3.8)	(14.0)	(14.7)
k_	98.4	100	1.6	97.9	99.3	1.4	99.4	100	0.6
	(6.3)	(0.0)	(6.2)	(6.4)	(2.9)	(7.2)	(2.8)	(0.0)	(2.7)

As regards to the final consonant sounds (Table 6.4), the results were more varied, as different sounds patterned differently. The voiceless alveolar stop was the best identified by the three groups at both testing times, followed by the bilabial and the velar voiceless stops. The voiced stops obtained lower identification scores than the voiceless stops and relatively similar scores across places of articulation, particularly at post-test. The DIS group obtained the greatest gains, in line with the global results across consonants reported above.

Table 6.4. Gain scores for individual final consonant sounds (untrained segment) in non-words obtained by the vowel trained groups and control group

	CONTROL			DIS_V			ID_V		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
_b	50 (22.5)	52.3 (18.9)	1.5 (11.9)	43.7 (22.3)	64.5 (22.3)	20.8 (25.3)	55.0 (21.6)	60.6 (27.5)	5.6 (18.3)
_d	64.8 (22.9)	60.1 (21.9)	-4.7 (15.0)	63.1 (25.6)	65.2 (22.8)	2.1 (25.0)	65.0 (22.7)	60.0 (22.7)	-5 (23.0)
_g	58.5 (15.6)	60.1 (17.2)	1.5 (18.1)	58.3 (19.8)	69.4 (18.6)	11.1 (26.0)	68.7 (19.6)	63.1 (23.8)	-5.6 (18.3)
_p	78.1 (9.6)	78.1 (11.6)	0 (11.1)	75.6 (30.2)	67.3 (20.6)	-8.3 (32.9)	70.6 (15.8)	64.3 (22.6)	-6.2 (25.8)
_t	91.4 14.9	93.7 12.0	2.3 (8.1)	87.5 (19.1)	96.5 (9.3)	9.0 (20.9)	93.7 (11.8)	95.0 (13.0)	1.2 (8.9)
_k	71.0 (19.2)	73.4 (18.1)	2.3 (18.3)	59.7 (26.2)	71.5 (16.3)	11.8 (27.9)	73.1 (23.0)	73.7 (19.4)	0.6 (29.6)

6.2.2. Consonant training groups

Two groups were trained on consonants, one by means of an ID task (ID_C) and one by means of a DIS task (DIS_C). The groups' performance on trained (initial and final consonants) and untrained sounds (vowels) is shown next and contrasted with the performance of the control group (CG).

6.2.2.1. Post-test results

Correct identification scores obtained at pre-test, post-test and gain scores were calculated for the two consonant trained groups (ID_C, DIS_C) and the control group and are shown in Table 6.5.

Table 6.5. Pre-test, post-test and gain scores for trained and untrained sounds in non-words obtained by consonant trained groups and control group

	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	69.5	9.6	72.8	9.5
POST	79	13.4	82.3	9.8	88.7	7.1
GAIN	0.9	5.6	12.8	11	15.9	10.6
Trained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	70	16.8	69.9	8.8
POST	69.6	6.2	74.3	7.3	75.4	7.0
GAIN	0.5	4.5	4.3	13.9	5.5	8.9
Untrained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.1	9.9	55	7.5	56.2	10.1
POST	57.8	10.2	64.7	8.4	55.3	12.2
GAIN	3.7	4.2	9.7	7.4	-0.9	7.9

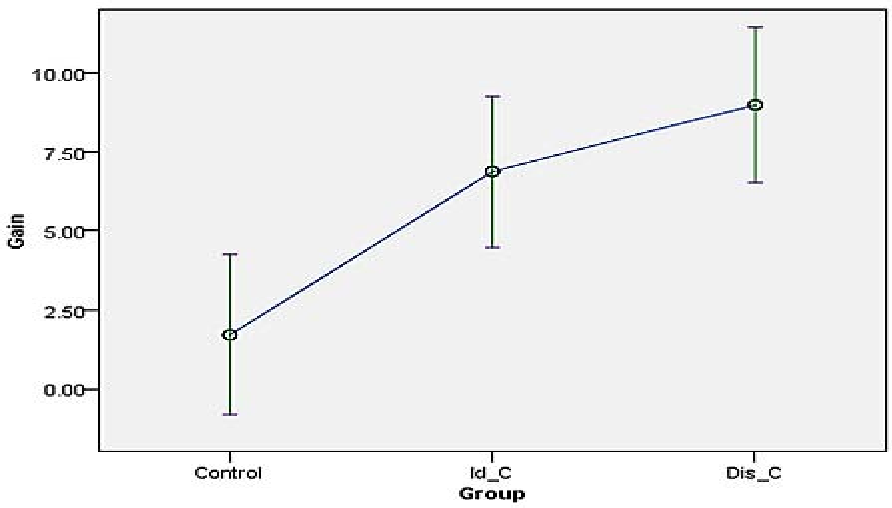
The results of the consonant training groups at pre-test are on the whole comparable to those reported for the vowel training groups, and show that in general all groups obtained the lowest correct identification scores for vowels (53-56%), followed by the scores for final stops (65-71%) and initial stops (70-78%), see Tables 6.5 and 6.1 above. The consonant training groups obtained numerically higher scores at post-test than at pre-test for the trained segments (initial and final consonants). But only the DIS group shows a notable numerical difference with the untrained sounds (from 55% at pre-test to 64.7% at post-test).

6.2.2.2. Gain scores on trained and untrained sounds

Gain scores for trained and untrained sounds were calculated for each group and are shown in Table 6.5 above. As can be observed, the DIS_C group improved 12.8% on initial consonant and 4.3% on final consonant performance, whereas the ID_C group improved 15.9% on initial consonant and 5.5% on final consonant identification. The amount of gain evidenced by the control group in this case was comparatively small (0.9% for initial consonants and 0.5% for final consonants).

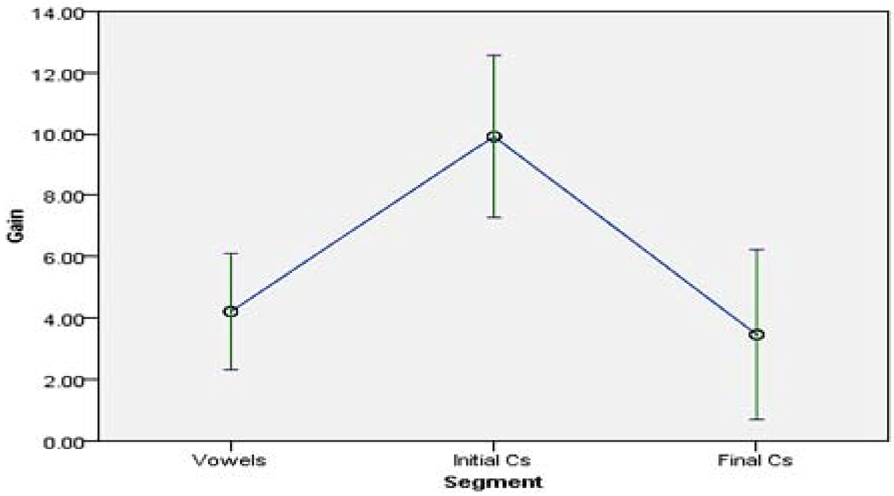
A generalized linear mixed-effect model was conducted on the gain scores. The fixed effects were group (ID, DIS and CG), segment (vowels, initial consonant and final consonant) and a group by segment interaction. The analysis revealed a significant effect of group ($F(2, 144) = 8.71$, $p < .01$), and segment ($F(2, 144) = 7.40$, $p < .01$) and a group by segment interaction ($F(4, 144) = 7.06$, $p < .01$). Pairwise contrasts (sequential Bonferroni) revealed that both groups significantly outperformed the controls ($p < .01$) on the overall performance (trained and untrained segments) and that both groups did not differ significantly from each other ($p > .05$). The performance of the three different groups is illustrated in Figure 6.5.

Figure 6.5. Overall gain for non-words obtained by consonant trained groups and control group



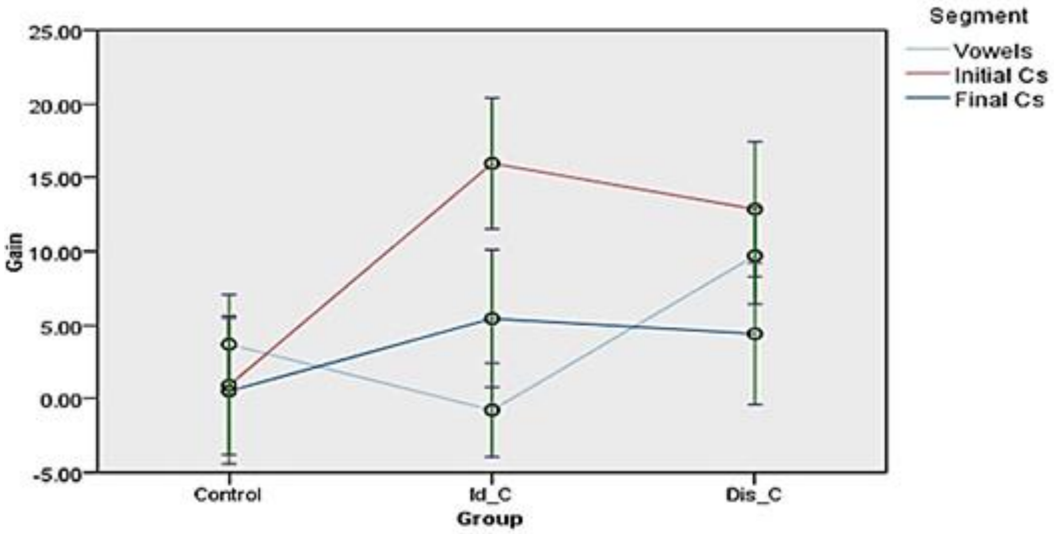
As regards to the segment effect, there was more improvement with initial stops than with the other two segment types (final consonants and vowels). This difference reached statistical significance as confirmed by sequential Bonferroni pairwise comparisons (p <.01). However, the amount of gain on the perception of vowels and the amount of gain on final consonants did not differ significantly, as can be observed in Figure 6.6. It is important to recall that the two experimental groups were trained (i.e. received feedback) on initial and final consonants but no feedback and/or training was received on the vowel segments apart from formal phonetics instruction as part of their degree.

Figure 6.6. Overall gain for non-words for each segment type across consonant training groups



Regarding the group by segment type interaction, similarly to the result previously reported for the vowel trained groups, the type of segment moderated the effects of the training received (ID, DIS or CONTROL) on the gain scores. The interaction is explained by the fact that while differences between segments were small for the control group, the trained groups showed greater differences between segment types. In particular, the ID groups gain scores for initial consonants, and to lesser extent final consonants, were notably higher than their gain scores for vowels, unlike the other groups (see Figure 6.7).

Figure 6.7. Group by segment gain on trained and untrained sounds in non-words obtained by consonant training groups and control group



As in the previous section, further simple effects analyses were carried out to explore each segment type separately and to compare the results for trained and untrained sounds. The results for the trained segments are explored first, followed by those for the untrained segments.

6.2.2.3. Gain scores on trained sounds (initial and final consonants)

Regarding the trained segments, simple effects tests results revealed a significant effect for initial consonants ($F(2, 144) = 11.52$, $p < .01$) and pairwise contrasts (sequential Bonferroni) revealed that both experimental groups significantly outperformed the controls ($p < .01$ in both cases) and that the trained groups did not differ significantly from each other ($p > .05$). No effect emerged for the final consonants ($F(2, 144) = 1.12$, $p > .05$), despite the numerical advantage of the experimental groups (4.3-5.5% gain) over the controls (0.5%). Table 6.6 and 6.7 show the results for each group for the six initial stops and six final stops, respectively. These values are shown in order to provide some additional descriptive statistics, but the aim of the present study is to assess the effect of training on trained and untrained sounds by global measures of improvement across individual consonant and individual vowel sounds.

Table 6.6. Gain scores for individual initial consonant sounds (trained segment) in non-words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN %(SD)
b_	70.3 (27.0)	69.5 (25.4)	-0.8 (14.4)	53.7 (27.1)	64.7 (24.3)	11 (26.9)	56.9 (23.6)	78.5 (19.5)	21.6 (29.9)
d_	46.9 (29.8)	43.8 (31.6)	-3.1 (16.7)	25.7 (25.2)	55.9 (29.7)	30.2 (32.2)	31.2 (24.3)	73.6 (24.6)	42.4 (31.8)
g_	56.3 (27.8)	62.5 (31.0)	6.3 (15.1)	41.2 (16.9)	77.9 (15.6)	36.7 (19.9)	50.6 (22.0)	83.4 (11.3)	32.8 (20.6)
p_	98.4 (4.3)	99.2 (3.1)	0.8 (5.5)	99.2 (3.0)	100 (0)	0.8 (3.0)	100 (0)	99.3 (2.9)	-0.7 (2.9)
t_	98.4 (4.3)	99.2 (3.1)	0.8 (5.5)	100 (0)	95.5 (18.2)	4.5 (18.2)	100 (0)	99.3 (2.9)	-0.7 (2.9)
k_	98.4 (6.3)	100 (0.0)	1.6 (6.2)	97.0 (8.3)	100 (0)	3.0 (8.3)	97.9 (6.4)	98.6 (5.9)	0.7 (9.1)

Similarly to the results shown previously for the vowel trained groups, the consonant trained groups obtained very high scores on the identification of initial voiceless stops and hence little or no improvement (gain) was observed from pre-test to post-test for these specific stops. As regards to the initial voiced stops, both consonant training groups showed notable numerical gain for the three stops, particularly with the alveolar and velar stops, followed by the bilabial stop. This is in contrast with the results for the vowel training groups.

The ID_C group obtained the highest identification scores for all initial voiced stops at post-test.

Table 6.7. Gain scores for individual final consonant sounds (trained segment) in non-words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
_b	50 (22.5)	52.3 (18.9)	1.5 (11.9)	61.8 (17.9)	67.6 (19.8)	5.9 (20.3)	54.2 (16.7)	72.92 (18.3)	18.75 (28.2)
_d	64.8 (22.9)	60.1 (21.9)	-4.7 (15.0)	67.6 (25.8)	72.0 (24.4)	4.4 (22.1)	68.1 (23.9)	84.03 (14.1)	15.97 (25.3)
_g	58.5 (15.6)	60.1 (17.2)	1.5 (18.1)	65.4 (25.6)	59.5 (21.4)	-5.9 (27.3)	61.8 (22.5)	58.33 (16.7)	-3.47 (19.6)
_p	78.1 (9.6)	78.1 (11.6)	0 (11.1)	69.8 (25.8)	76.5 (21.6)	6.7 (33.1)	70.1 (14.3)	68.75 (22.4)	-1.39 (22.2)
_t	91.4 14.9	93.7 12.0	2.3 (8.1)	91.9 (15.9)	92.6 (18.8)	0.7 (22.7)	96.5 (9.4)	86.81 (20.8)	-9.72 (22.1)
_k	71.0 (19.2)	73.4 (18.1)	2.3 (18.3)	63.2 (20.9)	77.9 (15.0)	14.7 (23.1)	69.4 (22.4)	81.94 (10.7)	12.5 (24.6)

Regarding the final consonants, it can be said that the identification of the alveolar stops was the least difficult for the participants, in line also with the results for the vowel trained groups. The identification scores obtained for the voiceless alveolar stop /t/ were already numerically high from the outset for all groups and subsequently little gain was observed. However, the trained groups obtained varied results when identifying the other final stop sounds: the ID_C improved the most with final /b/, followed by /d/ and /k/, and the DIS_C improved the most with final /k/, followed by /p/, /b/ and /d/. In brief, only the trainees showed improvement from pre-test to post-test, the control group's results revealed little variation, as indicated above.

6.2.2.4. Gain scores on untrained sounds (vowels)

Interestingly, when looking at the untrained segments (vowels) the pattern observed earlier with the vowel training groups is replicated: the DIS_C trainees outperformed the other groups with untrained sounds. This is confirmed by the simple effects tests results for the untrained segment (vowels), which yielded a significant effect of group ($F(2, 144) = 10.33$, $p < .01$) and the pairwise comparisons with a Bonferroni correction that revealed that the DIS_C group significantly outperformed both the CG ($p < .05$) and the ID_C group ($p < .01$). The individual sound performance on vowel segments (untrained segment) obtained by the consonant trained groups can be seen in Table 6.8.

Table 6.8. Gain scores for individual vowel sounds (untrained segment) in non-words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
æ	39.8 (17.6)	40.8 (19.9)	1.0 (20.6)	38.0 (17.7)	44.6 (23.2)	6.6 (25.4)	35.9 (15.5)	33.3 (13.5)	-2.5 (14.9)
ɜː	39.1 (21.0)	41.4 (19.6)	2.3 (12.8)	48.5 (21.4)	69.4 (23.2)	20.8 (22.1)	45.4 (20.0)	52.8 (21.1)	7.4 (18.8)
ɪ	69.0 (15.7)	80.2 (14.2)	11.2 (17.1)	74.3 (17.6)	77.4 (10.7)	3.2 (16.0)	77.8 (12.0)	79.9 (19.1)	2.1 (16.9)
iː	67.4 (15.6)	64.5 (19.2)	-2.9 (13.3)	60.0 (15.3)	68.6 (13.0)	8.6 (13.5)	66.2 (13.4)	58.1 (22.2)	-8.1 (19.3)
ʌ	55.2 (22.7)	62.0 (21.0)	6.8 (12.5)	54.2 (17.6)	63.5 (22.2)	9.3 (17.9)	54.4 (21.2)	55.5 (19.6)	1.1 (21.9)

Turning the attention to the results for individual vowels, it is interesting to note that the high front lax vowel /ɪ/ was the most accurately identified by the three groups. The control group improved the most with this vowel, and the results for all three groups at post-test were comparable (77-80%). The sound /æ/ was the most difficult sound for all groups and little or no improvement was found across groups, having the DIS_C improving the most with this sound (6.6%). The sound /ɜː/ was the sound that showed the highest improvement for both experimental groups (20.8% for the DIS_C, and 7.4% for the ID_C). The DIS_C showed a positive numerical improvement in all vowel sounds from pre-test to post-test, even though they did not receive any training on vowel sounds. Interestingly, the results for individual sounds are similar to those obtained for the vowel trained groups in that the vowels that were more accurately identified were /ɪ/, /iː/ followed by /ʌ/, with /æ/ and /ɜː/ being the most challenging at the outset. However, after training the DIS_C group showed the most consistent improvement, most notable with the vowel sound /ɜː/. The results reported so far will be discussed next.

6.3. Discussion I - Effects of training on trained and untrained sounds

This part of the study explored the effects of two different perceptual training methods (Categorical DIS and ID) on the perception of L2 vowels and consonants by Spanish/Catalan learners of English. More specifically, it addressed the first and second research questions, which were concerned with the direct and indirect effect of the HVPT methods here investigated. The effect of DIS and ID training will be discussed following the order in which the results were presented, that is, starting with the results for the vowel training groups, followed by the discussion of the results obtained by the consonant training groups. Recall that as the stimuli used in testing and training were elicited from different talkers (refer back to sections 5.2 and 5.3 for more details), the results reported here already reflect generalization to two novel talkers.

6.3.1. Vowel training groups

The results obtained by vowel training groups showed that both experimental groups (DIS_V and ID_V) significantly improved their perception of vowel sounds (as measured by an identification task) and outperformed the control group. However, the group trained by means of an ID task obtained higher gains than the group trained on DIS (26.3% vs. 9.8% improvement from pre to post-test). Even though this superiority might be partly connected to a task familiarity effect[20], the amount of the difference between the ID_V and the DIS_V's results seems to indicate that an ID method is more effective than a categorical DIS task for improving learners' identification of L2 vowel sounds. This finding goes in line with previous work involving the identification of vowels and nasals (Nozawa, 2015) and some previous views and findings on the comparison of auditory ID vs. DIS tasks in the literature (Strange & Dittmann, 1984; Jamieson & Moroson, 1986; Logan & Pruitt, 1995; Logan et al., 1991; Pisoni et al., 1993).

Nonetheless, the fact that the categorical DIS training regime was also effective in improving vowel perception, even if to a lesser extent than the ID method, cannot be overlooked. In fact, the DIS_V group improved (9.8%) from pre-test to post-test, which verges on the 10% expected training improvement reported in previous studies (Logan & Pruitt, 1995; Flege, 1995b; Iverson et al., 2012). Thus, the results found here confirm the efficacy and robustness (evidenced by the generalization to novel voices) of both training methods (ID and categorical DIS) on training L2 vowel perception and add to the existent literature that had previously contrasted ID tasks to auditory DIS tasks. According to Strange and Dittmann (1984), auditory discrimination training was viewed as "slow and effortful" and Jamieson and Moroson (1986) explain that:

[20] Recall that both testing and training tasks involved the identification of the target sounds.

> "Discrimination training tends to increase intra phonemic sensitivities, which in normal phonetic development may necessarily be processed as perceptually irrelevant. Such increased within-category sensitivity could diminish attempts to form a new phonetic category" (p. 207).

By contrast, the success and robustness of the categorical DIS task as a training tool found in this study may be linked to the fact that, unlike its auditory counterpart, this task exposes learners to multiple tokens of each category. This exposure is likely to contain a wide range of acoustic variants (Polka, 1992), which may encourage formation of "new equivalence classes" for L2 phonetic segments (Strange, 1992).

Turning our attention to the results obtained on the untrained sounds (initial and final consonants), it was observed that only the DIS training group improved the perception of untrained but implicitly exposed sounds. This group's improvement was 2.4% on initial consonants and 7.7% on final consonants. The overall small and non-significant improvement in the perception of initial consonants is partly accounted for by the ceiling effect observed at pre-test on the identification of voiceless stops. This ceiling effect may be due to aspiration being a strong cue when differentiating voiced and voiceless English stops (Burnham, 1986). Even though aspiration is a non-existent phonetic property in Romance languages (Ladefoged, 1971, 2001) such as Spanish and Catalan, according to Flege (1989), an acoustic cue which is non-existent in the L1 but is very relevantly salient can be learned with ease. Similarly, the lack of clear voicing through closure, which signals initial voicing in the L1 might account for the lower results obtained for the initial voiced stops. Catalan learners possibly had more difficulty to discern L2 voiced vs. voiceless stops apart as the L1 voiceless stops have short lag VOT.

Regarding the results on final consonants (untrained sounds), while the ID_V group did not differ from the controls, the DIS_V group significantly outperformed the other two groups. This outcome points to an advantage of discrimination training in the perception of untrained sounds and it suggests that while identification training may be an overall more effective method for L2 vowel training, discrimination training may promote generalization to the final stop consonant present in the stimuli to a greater extent. Three potential reasons for this "cross-training effect" are discussed next.

The first potential reason for this cross-training effect observed might be connected to the stimuli presentation. Since categorical DIS training involves a direct comparison between two physically present stimuli and the same pairs of words were always presented in two possible orders (e.g., *tidge-teedge* and *teedge-tidge*), this exposure may have made listeners more alert to differences across trials and enhanced sensitivity to the untrained segments as well as the trained segments. Conversely, an ID task exposes learners to one stimulus at a time and thus such cross-trial comparison is less evident. Another possible account may be connected to the absence vs. presence of labels. The absence of labels present in the DIS procedure may have allowed DIS trainees to focus on the word as a whole, hence noticing not only the vowel present in the stimuli, but the final consonant following the target sound. By contrast, the forced-choice ID training might have directed learners' attention solely to the target sounds embedded in the words, as listeners were forced to choose one of the forced choice labels provided. In fact, according to Logan and Pruitt (1995) and Beddor and Gottfried (1995), the presence of labels in an identification task may be problematic, since they may have an impact on listeners' performance. Polka (1992) explains that a forced choice identification task limits the learners' perception to the given labels.

A third potential explanation for the effect observed lies in the fact that consonants affect vowels and vice-versa (Recasens & Mira, 2015; Steinlen, 2005). For example, vowel duration is an important cue to postvocalic final obstruent voicing (Raphael, 1972; Roach, 2000). Thus, the results found here may indicate that the discrimination trainees were able to learn this feature from the vowel training regime received and apply it accurately when perceiving the final consonant stimuli. Since this systematic variation in vowel duration of the preceding vowel is a within-category characteristic (i.e. vowels becoming shorter when preceding a voiceless sound), it would make sense that only the discrimination trainees would become sensitive to such cue. Let us recall that most of the criticism against the use of auditory DIS tasks as a L2 training task was connected to the increase of within-category variability awareness (Strange & Dittmann, 1984; Jamieson & Moroson, 1986; Logan & Pruitt, 1995). Thus, by increasing their within-category sensitivity of the target vowel sounds, the DIS trainees might have become aware of the knock-on effect of the consecutive consonant on the preceding vowel. Hence, the within-category differences present in vowels may not be necessary or strictly relevant for the identification of the vowel target sounds (Jamieson & Moroson, 1986), but they might be an applicable cue for the identification of the final stop that follows this given vowel, explaining the cross-training effect here observed.

6.3.2. Consonant training groups

The results obtained by the consonant trainees reveal that both groups (ID_C and DIS_C) outperformed the controls in their perception of initial stops (trained segments), but unlike the vowel trained groups, the two experimental groups did not differ from each other. The slight numerical difference between the two experimental groups in the perceptual gain with initial (DIS_C =12.8% and ID_C =15.9%) and final consonant identification (DIS_C =4.3% and ID_C =5.5%) could be accounted for by a task effect, since one of the groups was tested and trained with a similar identification test. However, the non-significant numerical difference observed between the two experimental groups suggests that task familiarity had little effect. This result indicates that both training methods are efficient in changing L2 leaners' perception of initial stop consonants to the same extent. Even though this study is the first to compare these two specific training methods in order to train the perception of initial stops, these findings go in line with Flege's (1995b) findings for final stop modification and Nozawa's (2015) findings for final nasals.

However, no significant improvement was found with the final consonant stimuli in the present study. The gains obtained by both experimental groups were small and thus, non-significant. This may indicate that changing L2 learners' perception of initial and final consonants require different amounts of time. It is relevant to recall that the 5 sessions of 30-40 minutes that the participants received were divided into two equal parts, so that both initial and final consonant could be trained. Therefore, the total amount of training received for each consonant position (over 5 sessions) totaled about 75-100 minutes of training.

Taken together, the different results for initial and final consonants suggest that such a short amount of training is sufficient to positively modify L2 VOT perception; however, it is not sufficient to modify final consonant perception. Such an outcome is in line with some previous findings that demonstrated that the perception of initial stops differing in VOT can be easily modified through training (Carney et al., 1977; Pisoni et al., 1982; Collet et al., 2013; Aliaga-García & Mora, 2009). On the other hand, learning to distinguish stops in final position does not seem to take place with the same ease as initial stop learning, as evidenced by Flege's work (1989 and 1995b). Recall that in the first study (1989), Flege concluded that final stop consonant training was only successful when training included a large number of trials and a small set of words. That was so, as only the second experiment, which included more trials and a smaller number of words, succeeded in improving learners' perception of final stops. Similarly to Flege's (1989) first experiment, the current study may not have exposed participants to enough repetitions of each word, due to the short length of the training here applied being divided among the 6 final stop target sounds (/p t k b d g/). In Flege's second study (1995b), participants succeeded in modifying their perception of two word-final stops (/t d/) after the exposure and feedback of seven training sessions. Possibly, instead of the training here applied, a longer training regime would have been successful in modifying learners' perception of coda stops (Flege, 1989).

One can speculate that the reason for final consonant modification being more effortful than VOT perception may lie in the fact that initial stop voicing contrasts, which are often cued by differences in VOT, are usually perceived in a categorical mode, whereas final consonant perception, which is partly cued by the preceding vowel duration in English, is usually perceived in a continuous mode rather than in a categorical one (Raphael, 1971, Pisoni, 1973). Strange and Dittmann (1984) suggest that the "perception of voicing distinction may be less difficult to modify" (p.142) than other distinctions, since voicing distinctions are very common across languages and there may be a relation between ease of perceptual differentiability of a cue and its usage in languages.

Another possible reason was suggested by Flege (1989), who argued that the success of training also depends on the auditory nature of the feature which is being learned. In this sense, contrasts which are not present in the L1 but which are easily distinguishable (e.g. VOT), since they are psychoacoustically robust (Burnham, 1986), should be acquired more easily than less distinguishable contrasts (e.g. final consonant voicing contrasts).

Another motive that might explain the greater success with initial consonants rather than with final consonants is connected with the fact that identifying English consonants in onset position might be less challenging than identifying English consonants in coda position. Such effect has already been reported in L2 perceptual studies (e.g. Jang, 2014) as well as in L1 perceptual studies (e.g. Redford & Diehl, 1999; Swingley, 2009). This is because coda position is known to be a less stable position in terms of the strength of consonant cues, since consonants are more likely to be weakened or lost in final position and consonant contrasts to be neutralized (Cebrian, 2000).

Interestingly, when looking at the untrained segments (vowels) the reverse pattern observed earlier with the vowel training groups is replicated with the consonant training groups: only the DIS_C trainees outperformed the other groups when identifying the untrained sounds. Taken together, the results may indicate that while both ID and DIS are effective to train L2 initial stops, categorical DIS does not only promote learning of trained sounds, but also of untrained but implicitly exposed vowel sounds across trials, in line with the non-significant tendency found by Nozawa (2015). The fact that the ID training does not have an effect on the untrained segments might be accounted for by the fact that by specifically directing learners' attention to trained segments, identification training might have abstracted learners' attention away from any other characteristics present in the stimuli. This result is in line with previous research on training attention (Pederson & Guion-Anderson, 2010) that showed that an identification training regime improved learners' perception of the attended segments (i.e. trained segments) only, as no gain on the unattended segments (untrained but implicitly exposed segments) was found (but cf. Nozawa, 2015, who found a tendency to improve the perception of final nasals both with ID and DIS vowel trainees).

The discrimination procedure, on the other hand, promoted learning to the untrained segments (vowels) as well as the trained segments. Along similar lines to what was argued before, the stimuli presentation and/or the absence of labels might have triggered such cross-training effect. The other possible motive is the increased within-category awareness promoted by the DIS training method, which might have drawn learners' attention to the systematic variation in vowel duration that is present before the target stop consonant sounds. Most probably, in the attempt to discriminate the final consonant stimuli, the consonant trainees (DIS_C) might have realized that the final consonant voicing cue was not sufficient to predict the correct perception and were forced to pay attention to the preceding vowel duration as a more consistent and reliable cue (Raphael, 1972; Roach, 2000, Flege, 1989). Consequently, by paying attention to this cue and due to the metalinguistic knowledge from the phonetics course received, vowel category learning might have been facilitated.

6.3.3. Perception of individual sounds

The overall vowel perception across all five groups ranged between 53-56% at pre-test and it increased to 55-79% at post-test, showing differences among groups in the amount of improvement obtained. Even though a detailed analysis of individual sounds is not one of the goals of this thesis, the overall general patterns observed with each individual sound will be briefly discussed in line with the initial predictions that would derive from the models reviewed in the introduction (PAM-L2 and SLM) and/or the comparison between the L1 and the L2.

Looking at vowel perception first, all five groups experienced more difficulty when identifying /æ/ and /ɜː/, followed by /ʌ/ at the outset of the study (pre-test), while the high front vowels /iː/ and /ɪ/ were better identified at this initial stage. The range of the overall perception across all five groups for each individual sounds at pre and post-test can be seen in Table 6.9.

Table 6.9. Range of the overall perception of individual vowel sounds across groups

SOUND	PRE	POST
æ	31%-42%	33%-78%
ɜː	39%-48%	41%-81%
ɪ	69%-78%	77%-83%
iː	60%-67%	60%-78%
ʌ	53%-55%	55%-75%

The sound identification at pre-test partly follows the initial predictions of the current study, as according to PAM-L2 the two sounds /æ/ and /ʌ/ were expected to pose a lot of difficulties to learners and the high front vowels /iː/ and /ɪ/ were predicted to be perceived with more ease. This is because the latter pair was classified as a “category-goodness difference” assimilation pattern and the predicted discriminability would be moderate to very good, and the former pair was classified as a “single-category assimilation” pattern and the predicted discriminability would be poor. However, the central sound /ɜː/ was predicted to be successfully perceived from the outset. This claim was based on previous training research findings (Pereira, 2014) and on the transfer of PAM-L2 predictions on L2 sound discrimination to L2 sound identification. According to PAM-L2, since the target sound /ɜː/ would fall into a “categorized-uncategorized” assimilation pattern when contrasted with the L2 sound /e/, the predicted discriminability would be “good”. While this prediction was not met regarding the initial sound perception (as observed in Table 6.9), it was met when looking at the amount of improvement across groups, as all four groups improved greatly on their perception of this sound as a result of the training regime. While this result is partly accounted for by the lower scores obtained at pre-test for this specific sound, it also provides evidence that non-native sounds that are dissimilar to the pre-existing L1 sounds in the learners’ inventory (Flege, 1995) and have no clear match in the L1 sound inventory (Cebrian et al., 2011) can be learned with ease through perceptual training, as they “constitute good candidates for correct L2 categorization given enough exposure to the target language” (Cebrian, 2015:4) in accordance with the SLM (Flege, 1995).

Looking at the amounts of improvement for each individual sound per group, the highest amount of vowel improvement was observed by the ID_V group on the sounds /æ/ and /ɜː/ (46.6% and 42.5% respectively), followed by 22.1% improvement observed on the sound /ʌ/. The DIS_V group showed the greatest improvement on the sounds /ɜː/ and the sound /ʌ/ (20.1% and 12.7% respectively), followed by 9% on the sound /ɪ/ and 7.8% on the sound /æ/. Interestingly, the DIS_C followed a very similar pattern as regards to the amount of improvement, as the sound which showed the greatest improvement was /ɜː/ (20.8%), followed by the sounds /ʌ/ (9.3%), /iː/ (8.6%) and /æ/ (6.6%). Considering the DIS_C group did not have any type of feedback or oriented attention on vowel sounds, the amounts of improvement observed by this group provide evidence that exposure itself had an impact on the perception of individual target sounds (Nozawa, 2015; Gibson & Gibson, 1955) and that a cross-training effect took place (refer back to section 6.3.1 for a discussion on the cross-training effect observed).

Turning our attention to consonants, it is relevant to note that overall learners identified consonants better than vowels, as all groups presented higher scores on initial and final consonant perception than on vowel perception from the outset of the study. Such result is yet another indication that vowels and consonants are perceived differently (Fry et al., 1962; Strange, 2007) and might activate different levels of processing (Pisoni, 1973) when they are being learned.

Regarding the initial stops, the scores at pre-test were notably higher than for the other two segment types (70%-79%). Nonetheless, improvement as a result of training was still observed, as at post-test the results ranged from (74%-89%). The main patterns observed as regards to the initial consonants were that participants performed at ceiling with initial voiceless stops (97%-100%) from the outset, leaving limited room for improvement. This is probably accounted for by the fact that aspiration is a salient and detectable cue in perception (Burnham, 1986) and some sounds are learned with more ease, since they are more easily perceived and produced (Munro & Derwing, 2008). On the other hand, the perception of voiced initial stops was more challenging. This might be explained by the L1-L2 VOT differences involving initial voiced stops (L1 = voice lead and L2 = short lag), and by the fact that learners commonly perceive L2 sounds through the frame of their L1 language inventory (Flege, 1995; Best & Tyler, 2007).

For the initial voiced stops, the results were lower when identifying initial /d/ at pre-test (26%-47%), followed by the initial velar sound /g/ (41%-57%) and the initial /b/ (53%-71%). The higher scores at the outset for the bilabial stop are most probably connected to the fact that this stop was not always produced with short lag VOT. Several instances of this stop contained prevoicing for both testing talkers (T1 and T2). Let us recall that according to Smith (1975), prevoicing can be a cue present in native English speech, taking place most frequently for labial and dental stops and preceding a high vowel. Moreover, the author adds that voice lead is more frequent and of greater prevoicing duration for male speakers. Thus, learners performed better when identifying the labial stops, since some instances of this sound were quite similar to the L1 sound and thus posed fewer perceptual problems to the learners.

As regards improvement, all initial voiced stops improved as a result of the consonant training received, particularly /g/ and /d/, which were the sounds that improved also with the DIS_V training. The bilabial stop /b/ improved the least, probably because it started out with higher scores. The greatest improvement was found with /d/, particularly with the ID_C group. These results indicate that similarly to the result observed for vowel sounds, the most “dissimilar” sound in the learners’ inventory, /d/[21], was poorly identified at the onset of the study; however, it was the sound which showed the greatest improvement after feedback training, in line with the SLM predictions (Flege, 1995) and with previous training findings (Pereira, 2014).

Regarding the final consonants, the observed difference between pre-test (65%-71%) and post-test (69%-75%) was much smaller. Results were lower in general for final voiced

[21] The English voiced alveolar sound /d/ differs in place of articulation in the learners L1 and the L2, as this sound is dental in Spanish/Catalan.

stops (44%-69% at pre-test and 52%-84% at post-test) than for final voiceless stops (59%-96% at pre-test and 64%-96% at post-test). This might be accounted for by the fact that English final voiceless stops are “identical” sounds to the L1 final voiceless stops (Flege, 1995) and, according to the SLM, they are not predicted to pose much difficulty to Catalan learners. The voiced stops, on the other hand, were predicted to pose more difficulties to these learners since sounds classified as “similar” are difficult to acquire (Flege, 1995). Further, previous research has found that Spanish and Catalan speakers usually resort to L1-based strategies such as devoicing when perceiving final voiced obstruents (Cebrian, 2000; Flege & Davidian, 1984). Moreover, the little improvement observed might be explained by the fact that a voicing contrast in final position may require a longer training period than other distinctions (Flege, 1989; Flege, 1995b), especially since the voicing contrast in word-final stops is not an existing feature in the L1 (Recasens, 1993; Cebrian, 2000) and, according to Flege (1989), features that are not present in the L1 pose more of a challenge to learners.

6.3.4. Control group

Interestingly, the control group of the current study showed some improvement from pre-test to post-test. Let us recall that the control group in the present study was not completely untrained, as these participants underwent a semester of formal phonetics instruction and performed additional transcription exercises. While hardly any improvement took place for the consonant sounds, most probably due to the fact that the identification of these sounds started out higher than vowel scores, the controls enhanced their perception of vowel sounds from pre to post-test, mostly of two sounds: /ɪ/ and /ʌ/. This improvement is probably accounted for by two main factors: the positive impact of formal phonetics instruction (Darcy et al., 2012; Kissling, 2013; Thomson & Derwin, 2014), since it entails focus on form (Long, 1991), and the consequent increase in phonological awareness (metalinguistic knowledge) of the two novel English sounds by the controls. According to Kivistö-de Souza (2015), explicit phonetics instruction might enhance the awareness of the target sounds by making them more noticeable, and as a consequence better perception and/or production of the specific target sounds might occur. Nonetheless, the vowel trainees and the consonant trainees improved significantly more than the controls in vowel and consonant identification (and even on untrained sounds on two occasions), hence showing the efficacy of the training received.

Part II – Tests of robustness

6.4. Generalization results

As previously explained, this study looked at three different types of generalization: generalization to real words (generalization 1), generalization to novel non-words (generalization 2) and generalization to novel speakers in two different instances, pre- and post-test talkers and absolute new talkers (generalization 3a and generalization 3b respectively). As generalization 3a has already been described, the results will be described in the following order: generalization 1 followed by generalization 2 and then generalization 3b. At the end of each result section, a corresponding summary is included in in order to guide the reader. Results for the vowel trained groups will be presented first by being contrasted with the results for the control group, followed by the results obtained by the consonant trained groups contrasted with the control group's performance.

6.4.1. Generalization 1. Real words

6.4.1.1. Vowel training groups

The results obtained for vowel (trained segment) and consonant (untrained segment) perception in real words by the two vowel trained groups (ID_V, DIS_V) will be presented next. Perception of real words was assessed at pre-test and at post-test in order to have a measure for real word identification comparable to that of non-sense word identification. Non-words were the primary measure as they eliminate the word familiarity effect, which has been found to affect the accuracy and speed of word processing (Grosjean, 1980).

Correct identification percentages for vowel and consonant sounds embedded in real word at pre-test and post-test, and the corresponding gain scores, were calculated for each group. Statistical analyses were carried out on the amount of gain obtained by each group, as previously done for the nonsense words. Table 6.10 displays the percentage correct identification of vowels in real words for each of the vowel trained groups (ID_V, DIS_V) and the control group at pre-test and post-test.

Table 6.10. Pre-test, post-test and gain scores for trained and untrained sounds in real words obtained by the vowel trained groups and the control group

	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	72.2	11	78.2	9.7	73.1	11.2
POST	79.5	10.3	79.7	11.1	88.5	9.5
GAIN	7.3	9.2	1.5	11.7	15.4	8.8
Untrained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	80.8	11.0	77.5	11.3	81.1	10.6
POST	83.2	10.9	81.2	7.6	77.1	11.1
GAIN	2.4	8.2	3.7	7.5	-4	7.1
Untrained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	65.5	8.3	63.8	10.5	68.3	5.6
POST	68.3	8.7	68.2	9.5	66.4	7.1
GAIN	3	4.5	4.4	10.9	-1.9	6.8

Interestingly, if we compare the results in Table 6.10 with those obtained for non-words (see Table 6.1), we can observe that vowel identification scores are higher in real words than in nonsense words (72-78% vs. 54-56% at pre-test), but consonant identification is comparable across the two word conditions (76-81% vs. 74-78% for initial stops and 66-

68% vs. 65-71% for final stops). Regarding the amount of vowel gain observed from pre-test to post-test, the ID_V, the group that improved the most with non-words (26%) was also the group that obtained the greatest gains with real words (15%). The DIS_V training regime, on the other hand, did not seem to enhance the ability to identify sounds in real words, as its trainees only improved 1.5% with the training received. While this little improvement is connected to the considerably high scores at pre-test, it also shows that the improvement observed with non-words (9.8%) may not generalize to real word stimuli. In the case of the controls, the learners appeared to improve more in real word identification than when identifying non-words (7.3% vs. 3.7%).

Turning our attention to the gains observed on the untrained segments, the DIS_V trainees and the controls showed some improvement when identifying initial and final consonants (untrained segment) embedded in real words, even if small (2.4-3% and 3.7-4.4%, respectively). No improvement can be observed for the ID_V group, replicating the results obtained with non-words. The statistical relevance of the group comparisons is examined next.

A generalized linear mixed model (GLMM) with group (ID, DIS, Control) and segment (vowels, initial Cs and final Cs) as fixed effects was conducted on the gain scores for real words. The results yielded a significant effect of segment ($F(2, 153) = 9.97$, $p < .001$), confirming that vowels obtained a greater gain than consonant sounds (Figure 6.8). Pairwise comparisons confirmed that the effect of sound is explained by the greater gain obtained by the vowels, as vowel improvement significantly differed from the improvement found for both consonant segments (p <.001 in both cases).

No significant effect of group emerged, but there was a group by segment interaction ($F(4, 153) = 8.28$, $p < .001$), meaning that different groups performed differently on different segments. Figure 6.9 shows the overall group performance and Figure 6.10 illustrates the performance obtained for the different groups and for each type of segment. Again, the larger difference between the vowel and consonant perception for the identification trainees in comparison with the performance of the other groups possibly accounts for the interaction. The interaction was further explored by means of follow-up simple effects tests on each segment type separately as it was previously done for the non-words.

Figure 6.8. Overall gain for real words obtained by vowel training groups and control group

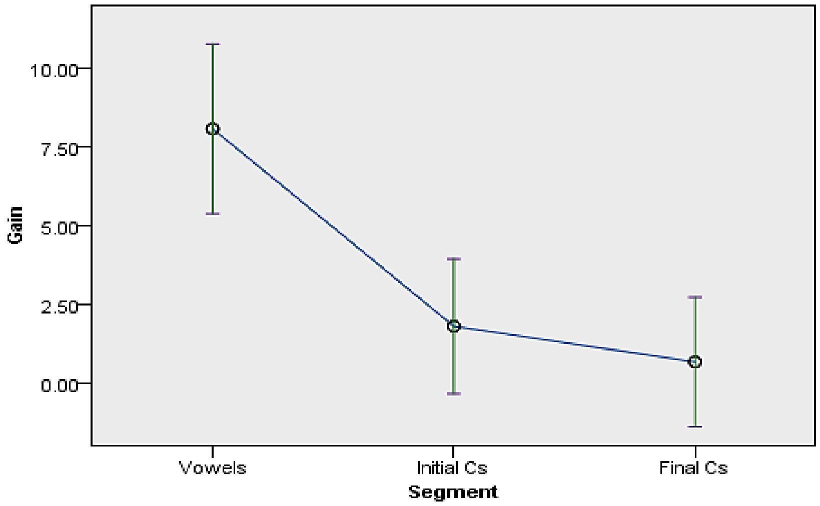


Figure 6.9. Overall gain for real words for each segment type across vowel trained groups and control group

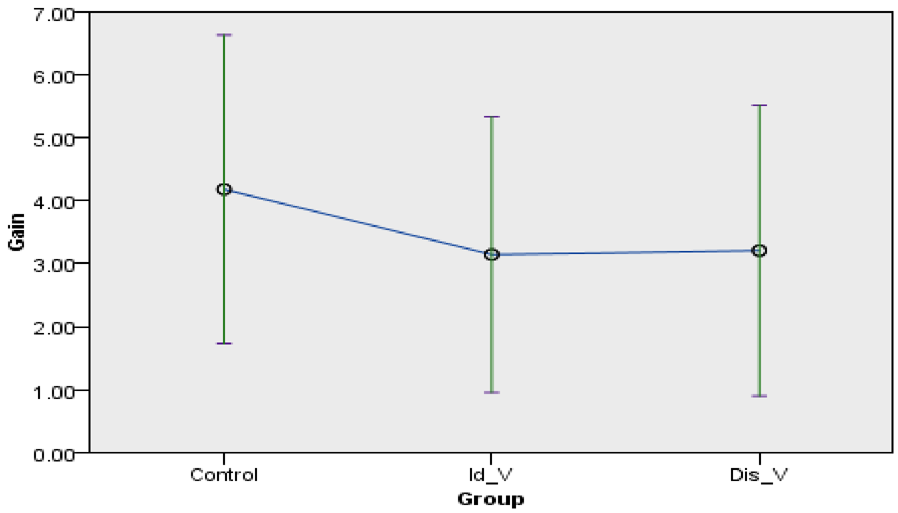
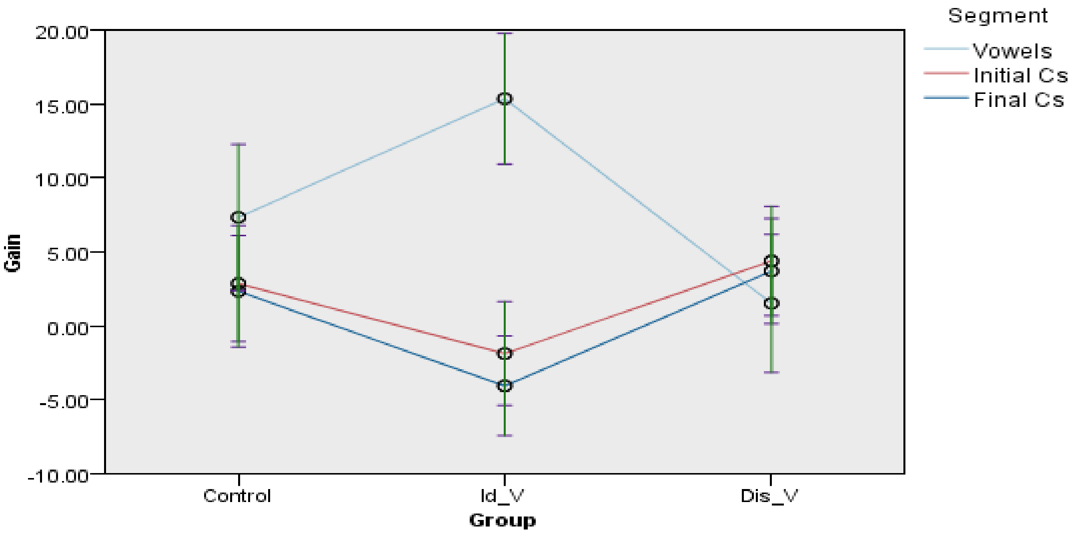


Figure 6.10. Group by segment gain perception on trained and untrained sounds in real words obtained by vowel trained groups and control group



6.4.1.1.1. Generalization gain scores on trained sounds (vowels)

The results for vowel sounds yielded a significant main effect of group, $F(2,153) = 9.16$, $p < .001$. Sequential Bonferroni pairwise comparisons confirmed that only the identification group outperformed the control group, $p < .05$. Moreover, the ID_V group outperformed the DIS_V, indicating that generalization to real words for the trained sounds only occurred after receiving identification training ($p < .01$). The identification scores for each individual vowel in the real word condition by each group can be seen in Table 6.11.

When looking at the individual vowel sounds, the control group seems to have improved more than the DIS_V for three out of the five sounds, namely /æ/, /ɜː/ and /ʌ/, which is probably explained by the higher scores obtained by the DIS_V at the onset of the study. At post-test, however, the scores for the specific sounds mentioned do not seem to differ much across the two groups. The ID_V, however, obtained numerically higher identification scores than the controls and the DIS_V group on the identification of /æ/, /ʌ/ and /ɜː/ after training, and both experimental groups improved numerically more than the controls for the sound /iː/, although all three groups reached similar identification scores with both /iː/ and /ɪ/ at post-test .

Interestingly, the pattern of difficulty on the whole matches the one found for non-words previously. The vowel /æ/ obtained the lowest scores, while /iː/, /ʌ/ and particularly /ɪ/ were better perceived. Moreover, overall scores were higher with real word identification than with non-words, in particular regarding the sound /ɜː/.

Table 6.11. Gain scores for individual vowel sounds (trained segment) in real words obtained by the vowel trained groups and the control group

	CONTROL			DIS_V			ID_V		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
æ	50.8 (30.8)	57.0 (34.4)	6.2 (33.8)	63.2 (24.9)	63.9 (27.7)	0.7 (26.2)	47.5 (30.2)	83.1 (25.1)	35.6 (29.6)
ɜː	67.9 (27)	85.9 (21.8)	17.9 (22.8)	79.2 (23.5)	81.2 (27.9)	2.1 (24.7)	71.9 (25.9)	95 (10.2)	23.1 (25.7)
ɪ	89.0 (22.5)	93.7 (22.1)	4.7 (12.6)	88.9 (13.5)	95.1 (8.7)	6.2 (16.7)	90 (9.6)	94.3 (9.5)	4.3 (15.8)
iː	82.0 (16.4)	79.7 (17.6)	-2.3 (13.1)	72.9 (18.3)	78.4 (14.1)	5.5 (23.2)	73.1 (17.8)	77.5 (17.5)	4.4 (22.3)
ʌ	71.1 (30.8)	81.2 (20.9)	10.1 (34.8)	86.8 (24.0)	79.8 (23.9)	-6.9 (28.2)	83.1 (9.6)	92.5 (9.5)	9.4 (15.8)

6.4.1.1.2. Generalization gain scores on untrained sounds (initial and final consonants)

When considering the untrained sounds, the results of the simple effects test for initial stops yielded a main effect of group, ($F(2, 153) = 3.23$, $p < .05$). Pairwise contrasts with a Bonferroni correction confirmed that the discrimination trainees' gain scores for initial consonants (real words) were significantly higher than the identification trainees' gain scores, $p < .05$, but did not differ from the controls', $p > .05$. This result shows a tendency for discrimination training to have a greater effect on untrained initial consonants sounds than identification training. However, this tendency is partly connected to the fact that the ID trainees' obtained a negative gain, not only to the gain obtained by the DIS group, which was relatively small (3.7%).

Regarding the results of the simple effects test for the final consonant stimuli, a main effect of group was also found, $F(2,153) = 5.629$, $p < .01$. Pairwise comparisons with a Bonferroni correction established that the discrimination trainees' and the control groups' gains scores were significantly greater than the identification trainees' gain scores ($p < .01$ and $p < .05$, respectively). These results are somewhat connected to the drop in gain for the ID_V group and the low scores at the outset for the DIS_V trainees, and they replicate the results obtained with non-sense words. Tables 6.12 and 6.13 show the individual sound performance for the initial and final consonant sounds, respectively, embedded in real words.

Table 6.12. Gain scores for individual initial consonant sounds (untrained segment) in real words obtained by the vowel trained groups and control group

	CONTROL			DIS_V			ID_V		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN %(SD)
b_	53.9 (30.2)	58.6 (29.5)	4.7 (23.7)	41.7 (35.1)	45.1 (27.8)	3.5 (28.1)	45.0 (31.8)	38.8 (34.9)	-6.3 (25.2)
d_	68.8 (21.9)	72.7 (21.5)	3.9 (20.3)	67.4 (22.7)	72.9 (17.3)	5.6 (25.1)	76.9 (21.2)	60.0 (25.5)	-16.9 (21.2)
g_	69.5 (22.8)	74.2 (20.1)	4.7 (16.4)	61.1 (29.4)	76.4 (25.3)	15.3 (22.9)	69.4 (26.1)	70.0 (23.4)	0.6 (18.8)
p_	99.2 (3.1)	100.0 (0.0)	0.8 (3.1)	99.3 (2.9)	97.9 (6.4)	-1.4 (5.9)	98.1 (6.1)	97.5 (7.7)	-0.6 (10.3)
t_	98.4 (4.3)	98.4 (4.3)	0.0 (4.6)	99.3 (2.9)	98.6 (4.0)	-0.7 (5.2)	99.4 (2.8)	100.0 (0.0)	0.6 (2.8)
k_	95.3 (12.8)	95.3 (12.8)	0.0 (0.0)	96.5 (5.8)	96.5 (5.8)	0.0 (9.6)	98.1 (8.4)	96.3 (8.2)	-1.9 (10.2)

Little or no improvement can be observed for the initial voiceless stops due to the high scores already obtained at pre-test, in line with the results reported for non-words. Regarding the voiced initial stops, similarly to the improvement observed with non-words, the velar stop showed the highest improvement for the DIS_V (15.3% for real words and 12.5% for non-words). However, the bilabial sound that had been considerably well identified from the outset in non-words, was poorly identified at both testing times when embedded in real words (42-53% at pre-test, 39-59% at post-test). In turn, the alveolar stop was more poorly identified embedded in non-words than in real words. The ID_V group showed none or very little improvement on the untrained sounds.

Table 6.13. Gain scores for individual final consonant sounds (untrained segment) in real words obtained by the vowel trained groups and control group

	CONTROL			DIS_V			ID_V		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
_b	36.7 (28.7)	51.6 (32.9)	14.8 (16.0)	54.9 (28.2)	56.9 (30.7)	2.1 (21.1)	52.5 (28.0)	58.1 (28.8)	5.6 (24.5)
_d	65.6 (26.8)	58.6 (28.0)	-7.0 (25.0)	46.5 (25.7)	49.3 (28.6)	2.8 (22.5)	59.4 (26.2)	54.4 (17.3)	-5.0 (30.2)
_g	50.0 (15.1)	50.8 (22.6)	0.8 (19.1)	52.1 (18.8)	56.3 (23.2)	4.2 (31.5)	50.0 (22.6)	48.8 (21.0)	-1.3 (23.6)
_p	80.5 (16.4)	78.1 (20.7)	-2.3 (20.5)	75.7 (21.2)	80.6 (21.1)	4.9 (26.5)	86.3 (13.4)	72.5 (26.5)	-13.8 (28.1)
_t	79.7 (26.0)	83.6 (31.0)	3.9 (12.7)	77.8 (24.1)	81.3 (22.0)	3.5 (23.4)	76.9 (24.8)	80.0 (18.3)	3.1 (26.9)
_k	80.5 (12.0)	87.5 (13.7)	7.0 (14.4)	75.7 (19.9)	84.7 (18.5)	9.0 (23.4)	85.0 (26.2)	85.0 (17.3)	0.0 (30.2)

As regards to the final consonant sounds, the voiceless stops were better identified than the voiced stops by the three groups at both testing times. The identification of the voiced stops resulted in lower values than the voiceless stops, ranging from (36.7%) to (65.6%), and the highest improvement was observed for the bilabial sound (14.8% by the CG), which is probably due to the low scores obtained by the controls at pre-test (37%). For the final voiceless stops, the velar stop showed the greatest improvement for DIS_V and the CG (9% and 7% respectively), followed by the bilabial and then the alveolar stop. Overall, final voiced stops continue to be a challenge, as was previously shown with non-words.

6.4.1.2. Consonant training groups

The results concerning the perception of initial consonants and final consonants (trained segments) and vowels (untrained segment) in real word stimuli by the two consonant trained groups (ID_C, DIS_C) at pre-test and post-test and the corresponding gain scores will be described next.

Correct identification scores at pre-test and post-test and gain scores for consonant and vowel identification in real words were calculated for each consonant trained group (see Table 6.14). Statistical analyses were carried out on the amount of gain obtained by each group, as previously done for the non-words. As observed with the vowel training groups, vowel identification in real words was more accurate than in non-words, already at pre-test, while for the consonants the type of word did not seem to influence the results as much.

Even though differences from pre-test to post-test are not large, probably due to high scores at pre-test in some cases, the ID_C is the group that seems to improve the most, particularly on initial consonants (13.1%). The improvement in final consonant identification is relatively small for all groups, having the ID_C improving the most, 5.1%. Regarding the untrained sounds, both the ID_C and the CG show a greater gain with vowel sounds, though the results for all three groups are comparable at post-test. Similarly to what was observed previously for the vowel training groups, the DIS training regime did not seem to improve learners' perception of the L2 sounds embedded in real word stimuli.

Table 6.14. Pre-test, post-test and gain scores for trained and untrained sounds in real words obtained by consonant trained groups and control group

	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	80.9	11	77.5	10.3	75.2	20.8
POST	83.2	10.9	81.9	11.4	88.3	5.9
GAIN	2.3	8.2	4.4	8.3	13.1	19.9
Trained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	65.5	8.2	69.3	11.4	66.7	8.6
POST	68.3	8.6	72.3	7.8	71.7	8.0
GAIN	2.9	4.4	3.1	7.8	5.1	9.9
Untrained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	72.2	11	80.9	9.7	73.7	11.2
POST	79.5	10.3	82.2	6.4	78.2	10.6
GAIN	7.3	8.2	1.3	8.3	4.4	19.9

In this case, the analysis (GLMM) with group (ID_C, DIS_C, Control) and segment (vowels, initial consonants and final consonants) as fixed effects yielded no significant effect of segment ($F(2, 144) = 0.89$, $p > .05$), no effect of group ($F(2, 144) = 2.73$, $p = .069$) and no group by segment interaction ($F(2, 144) = 1.68$, $p > .05$), indicating that consonant training did not result in generalization to real words for neither of the two groups trained on consonants. Thus, despite the numerically greater improvement obtained by the identification training group (Figure 6.11), the groups did not differ significantly.

Regarding the absence of a main effect of segment, it is interesting to note that even though the two experimental groups here described were trained by means of different feedback training methods on initial and final consonants, the real word perception of the trained segments was not superior to the real word perception of the untrained segments, that is, vowels (see Figure 6.12).

Figure 6.11. Overall gain perception for real words obtained by consonant training groups and control group

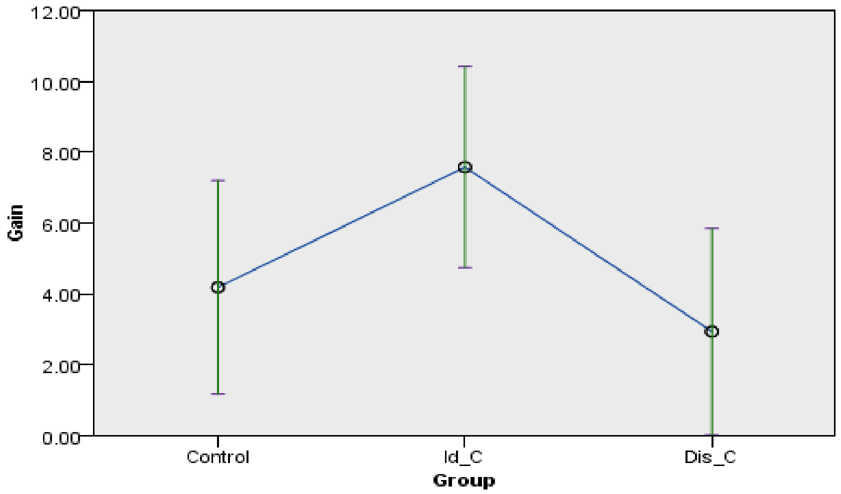
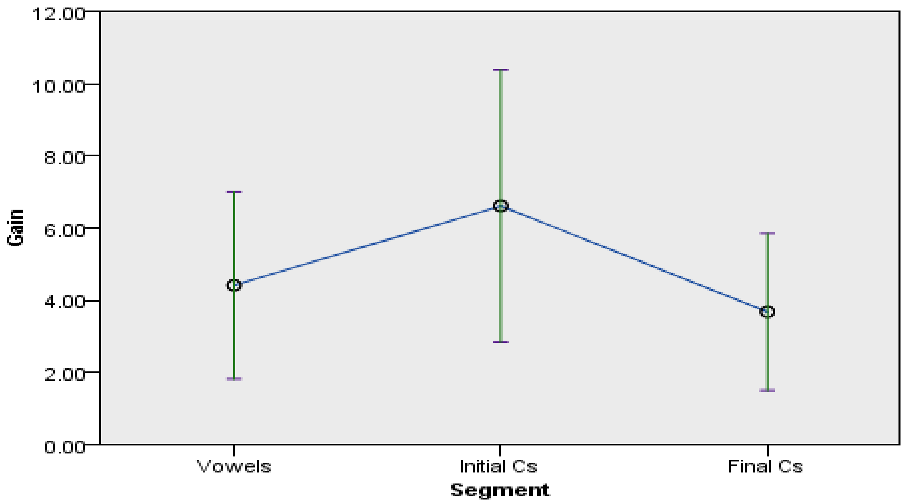
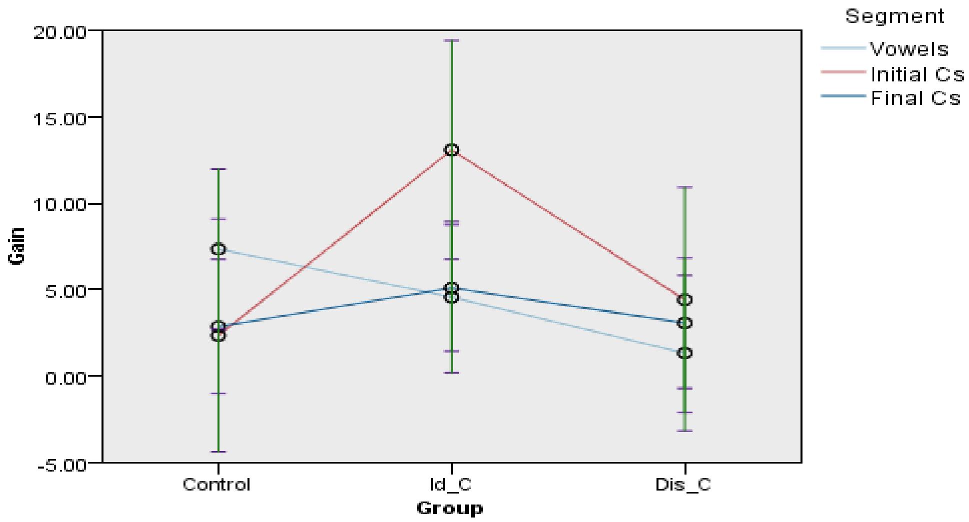


Figure 6.12. Overall gain perception for real words for each segment type across consonant trained groups and control group



The lack of a group by segment type interaction indicates that, despite the numerically greater results for initial consonants obtained by the ID group, the three groups performed similarly in their identification of the three different segments when embedded in real words. This supports the idea that the effects obtained during the consonant training regime with non-words (where both trained groups outperformed the controls with initial Consonants) did not generalize to real word stimuli. Figure 6.13 illustrates the segment by group perception of consonants (trained segments) and vowels (untrained segments) in real word stimuli.

Figure 6.13. Group by segment gain perception on trained and untrained sounds in real words obtained by consonant training groups and control group



6.4.1.2.1. Generalization gain scores on trained sounds (initial and final consonants)

Even though no interaction was found, it is interesting to point out the high numerical improvement obtained by the ID_C group for the initial stops (13.1%), as it falls within the range of reported amounts of improvement in earlier training studies. It is worth noting too that this improvement may be driven by the voiced stops, as there was hardly any improvement with the already high scores for initial voiceless stops. Table 6.15 displays the individual sound values for initial consonant segments and Table 6.16 shows the individual sound values for final consonant segments. Similarly to the results reported previously all groups identified initial voiceless stops very successfully from the beginning of the study.

The improvement with initial stops thus basically results from the amount of gain obtained for the voiced stops by the experimental groups, particularly the ID_C (16-22% gain), followed by the DIS_C group (9-16%), although this improvement was not significant either in the general analysis or when voiced stops only were taken into account[22]. The high variability found in the data accounts for this lack of significance, since it indicates that the improvement was driven by a few participants only and it was not a general group trend.

Table 6.15. Gain scores for individual initial consonant sounds (trained segment) in real words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN %(SD)
b_	53.9 (30.2)	58.6 (29.5)	4.7 (23.7)	37.5 33.9	50.0 36.4	12.5 38.8	45.1 33.9	66.7 20.1	21.5 33.5
d_	68.8 (21.9)	72.7 (21.5)	3.9 (20.3)	64.7 21.3	71.3 19.6	6.6 16.0	68.8 26.5	84.7 15.3	16.0 26.0
g_	69.5 (22.8)	74.2 (20.1)	4.7 (16.4)	65.4 23.2	74.3 21.4	8.8 19.6	59.0 32.1	79.9 22.3	20.8 34.6
p_	99.2 (3.1)	100.0 (0.0)	0.8 (3.1)	99.3 3.0	99.3 3.0	0.0 4.4	91.7 23.5	99.3 2.9	7.6 23.9
t_	98.4 (4.3)	98.4 (4.3)	0.0 (4.6)	99.3 3.0	98.5 4.2	-0.7 3.0	93.1 15.2	99.3 3.5	6.3 15.4
k_	95.3 (12.8)	95.3 (12.8)	0.0 (0.0)	99.3 3.0	98.5 6.1	-0.7 6.9	93.8 23.5	100.0 0.0	6.3 23.5

As for the final stops (Table 6.16), participants performed better with voiceless stops than with voiced ones, the latter ranging from (36.7%) before training to (67.4%) after training and the former ranging from (68.8%) before training to (86%) after training. As regards to the improvement observed for each individual final voiced stop, the highest

[22] Additional analyses were conducted for voiced stops only and the improvement was also found not to be significant.

improvement was observed for the bilabial sound (16.7% by the ID_C, 14.8% by the CG and 1.5% by the DIS_C). As previously mentioned, the CG's improvement is probably accounted for by low scores initially obtained. Both experimental groups obtained some improvement on the perception of the alveolar stop (DIS_C: 8.8% and ID_C: 7.6%). Little or no improvement was observed for the final voiced velar stop.

Table 6.16. Gain scores for individual final consonant sounds (trained segment) in real words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
_b	36.7 (28.7)	51.6 (32.9)	14.8 (16.0)	63.2 (27.8)	64.7 (27.7)	1.5 (17.6)	50.7 (29.2)	67.4 (23.5)	16.7 (36.4)
_d	65.6 (26.8)	58.6 (28.0)	-7.0 (25.0)	51.5 (30.6)	60.3 (23.5)	8.8 (30.5)	54.9 (25.1)	62.5 (22.7)	7.6 (31.5)
_g	50.0 (15.1)	50.8 (22.6)	0.8 (19.1)	68.4 (23.0)	62.5 (25.4)	-5.9 (29.0)	59.7 (22.5)	50.7 (19.9)	-9.0 (19.6)
_p	80.5 (16.4)	78.1 (20.7)	-2.3 (20.5)	76.5 (22.0)	71.3 (25.3)	-5.1 (27.6)	77.8 (18.0)	75.7 (21.6)	-2.1 (25.1)
_t	79.7 (26.0)	83.6 (31.0)	3.9 (12.7)	81.6 (16.0)	86.0 (18.7)	4.4 (18.2)	68.8 (26.8)	81.9 (18.8)	13.2 (31.0)
_k	80.5 (12.0)	87.5 (13.7)	7.0 (14.4)	74.3 (18.5)	89.0 (13.2)	14.7 (17.3)	88.2 (10.0)	92.4 (13.7)	4.2 (12.1)

6.4.1.2.2. Generalization gain scores on untrained sounds (vowels)

Turning our attention to the perception of the untrained sounds, vowels, it can be observed that the control group's improvement on vowels embedded in real words (7.3 %) was numerically higher than the performance of the two experimental groups (ID = 4.5%, DIS =1.3%). However, this difference can be explained by the control group's lower scores at pre-test, and in fact the three groups obtained comparable scores at post-test.

Table 6.17 shows the individual vowel sound performance for each group. As can be observed, the identification of different vowels varied across groups and presumably among individuals, as the standard deviation values were markedly high in some cases (especially for the sounds /æ/ and /ʌ/). The sound /æ/ continued to be the most challenging vowel, while /ɪ/, /iː/ and /ʌ/ obtained better results overall, particularly /ɪ/. As previously found with the vowel training groups, the central vowel /ɜː/ is better perceived in real words than in non-words. Interestingly, the control group obtained the lowest score for this vowel at pre-test but its scores are comparable to those of the experimental groups at post-test, which may indicate a parallel effect of the specialized instruction (phonetics course).

Table 6.17. Gain scores for individual vowel sounds (untrained segment) in real words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)	PRE % (SD)	POST % (SD)	GAIN % (SD)
æ	50.8 (30.8)	57.0 (34.4)	6.2 (33.8)	56.6 (34.6)	63.2 (29.5)	6.6 (45.1)	44.4 (28.8)	53.5 (34.8)	9.0 (38.8)
ɜː	67.9 (27)	85.9 (21.8)	17.9 (22.8)	89.0 (19.7)	89.7 (16.1)	0.7 (24.0)	84.0 (17.6)	93.1 (10.7)	9.0 (17.6)
ɪ	89.0 (22.5)	93.7 (22.1)	4.7 (12.6)	97.1 (7.0)	95.6 (7.6)	-1.5 (10.7)	89.6 (13.7)	98.6 (4.0)	9.0 (14.7)
iː	82.0 (16.4)	79.7 (17.6)	-2.3 (13.1)	76.5 (16.5)	79.4 (15.3)	2.9 (15.0)	74.3 (19.4)	66.0 (19.1)	-8.3 (23.1)
ʌ	71.1 (30.8)	81.2 (20.9)	10.1 (34.8)	85.3 (19.4)	83.1 (20.2)	-2.2 (20.4)	76.4 (26.0)	79.9 (31.5)	3.5 (18.1)

Summary

In summary, learners identified L2 vowel sounds more successfully when embedded in real words rather than in non-words. However, the same was not true for the identification of *stops, as both word conditions were identified with a similar degree of success. Regarding the training effects, while all trained groups outperformed the controls in the identification of trained sounds in non-words, the results for generalization 1 revealed that only the ID_V training promoted generalization effects for vowel sounds in real words. Despite the lack of improvement in vowel identification, the DIS_V showed a tendency to obtain higher gains with the untrained sounds (consonants), in line with the results obtained for the DIS groups with non-words. However, in the case of real words, group differences in the perception of untrained sounds were not significant. Regarding the consonant training groups, no evidence of generalization to real words was observed other than the numerically higher results for ID_C with initial consonants. Having reported the results for generalization 1 (real words), the outcome of the tests of generalization 2 (novel non-words) will be presented next.*

6.4.2. Generalization 2. Novel non-words

In order to assess the degree to which the effects of training generalized to novel items (i.e. novel CVC non-words) produced by familiar talkers (talkers heard at the training phase), a further test was administered in week nine, a week after the post-test took place (refer back to section 5.2 for further information). The generalization test assessed both trained and untrained sounds and the test layout and response options were exactly the same as the ones used for the post-test. Generalization should only be expected for the segments that improved with training, since generalization works as a measure of training robustness. According to Flege (1995b) "a high degree of generalization suggests that a training procedure has engendered the formation of a long-term memory representation that is more abstract than the sum total of the physical properties encountered in the training stimuli" (p. 435). Therefore, generalization will be assessed only for those segments that improved from pre-test to post-test.

Evaluating if the effect of training generalizes to a new dimension raises the question of what constitutes evidence of generalization. Different measures can be found in the training literature. While some studies compare the generalization results to the pre-test scores (Nobre-Oliveira, 2007), other studies assume that generalization occurs when the generalization results are comparable to the post-test results, and not significantly lower (Hazan et al, 1995; Flege, 1995b; Cebrian & Carlet, 2014; Rato, 2014). Thus, a lack of significant difference between the post-test and the generalization test is interpreted as evidence of generalization.

For the present study, both generalization measures previously mentioned will be taken into account, in an attempt to provide a complete picture of the generalization effects. Moreover, generalization will only be considered when there is a significant gain from pre-test to post-test (pre < post). Therefore, *strong generalization* will be considered to take place when the generalization results are as high as or higher than the post-test scores and differ from pre-test results [(*post-test = or < generalization)* + (*generalization > pre-test*)]. *Weak generalization*, on the other hand, will include the cases in which the generalization results are somewhat lower than (but not significantly different from) the post-test, but they are also not significantly different from pre-test. The generalization scores for trained and untrained sounds obtained by the vowel training groups will be reported first, followed by the generalization scores for trained and untrained sounds obtained by the consonant training groups. For the sake of brevity and simplicity, the focus here is on generalization globally, thus the results for individual sounds will not be reported.

6.4.2.1. Vowel training groups

The percentage correct identification obtained for trained and untrained sounds at pre-test, post-test and generalization test by the two vowel training groups (ID_V, DIS_V) and the control group (CG) were calculated and are shown in Table 6.18. The cases in which, training resulted in significant gain from pretest to post-test are highlighted in the Table. These were the improvement on vowel perception by both the DIS_V and ID_V groups (56% to 65% and 53% to 79%, for pre-test and post-test, respectively) and the improvement on final consonant perception by the DIS_V group (64% to 72%). In fact, it can be observed from Table 6.18 that the two experimental groups also maintained, or even increased, their vowel identification scores from post-test to generalization test (from 79% to 80% for ID_V and from 65% to 76% for DIS). However, the DIS_V generalization scores for final consonants were considerably lower than at post-test (72% vs. 61%). Basically in all other cases (CG and ID_V results for final consonants, and all groups' results for initial consonants), the difference between all three tests was numerically small. The exception was the results for the CG group with regards to vowels, as the group obtained higher identification scores at novel word generalization test than at pre and post-test (68% vs. 54% and 58%, respectively). This may be related to the formal instruction received, or perhaps another possibility is that these words and/or talkers posed fewer problems to the learners. Still, the CG group's scores were lower than those obtained by the trainees.

Table 6.18. Pre-test, post-test and generalization scores (novel words) for trained and untrained sounds in non-words obtained by the vowel trained groups and the control group

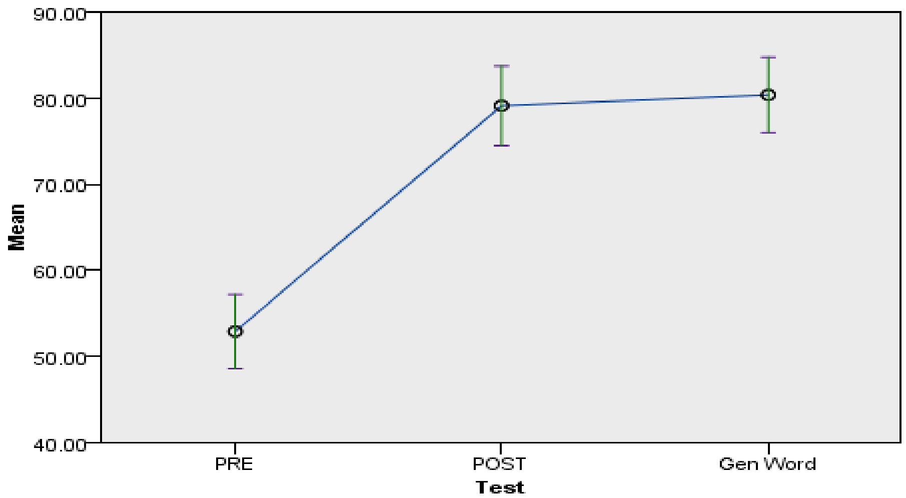
	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.1	9.9	**55.5**	6.5	**52.9**	9.5
POST	57.8	10.2	**65.3**	9.7	**79.1**	13.3
GEN WORDS	68.4	12.4	75.9	8.3	80.4	9.8
Untrained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	74.3	9.7	77.6	11.9
POST	79	13.4	76.7	11.4	74.2	11.9
GEN WORDS	79.9	8.7	78.8	7.4	75.8	11.0
Untrained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	**64.7**	14.5	71	6.4
POST	69.6	6.29	**72.4**	6.9	69.4	10
GEN WORDS	64.9	11.4	60.8	9.2	64.2	8.2

In order to confirm if the experimental groups generalized the knowledge acquired during training, generalized linear mixed models (GLMM) were conducted on the percentage scores obtained by the L2 learners at the three different tests for each experimental group individually[23]. The results found for the ID_V trainees are explored first, followed by the results found for the DIS_V trainees.

[23] Generalization of learning is only reported for the groups that underwent training; however, since the CG vowel scores were also higher for novel non-words generalization than at post-test, analyses contrasting the CG scores at the three testing phases (pre-test, post-test and non-word generalization test) with the three experimental groups which improved on vowels (ID_V, DIS_V, DIS_C) were carried out. A significant group difference was found ($F(3,200) = 9.525$, $p < .001$) and Bonferroni pairwise comparisons confirmed that these three groups outperformed the controls on the generalization of target vowel sounds in novel non-words ($p < .001$ for the ID_V group and $p < .05$ for the DIS_V and DIS_C groups).

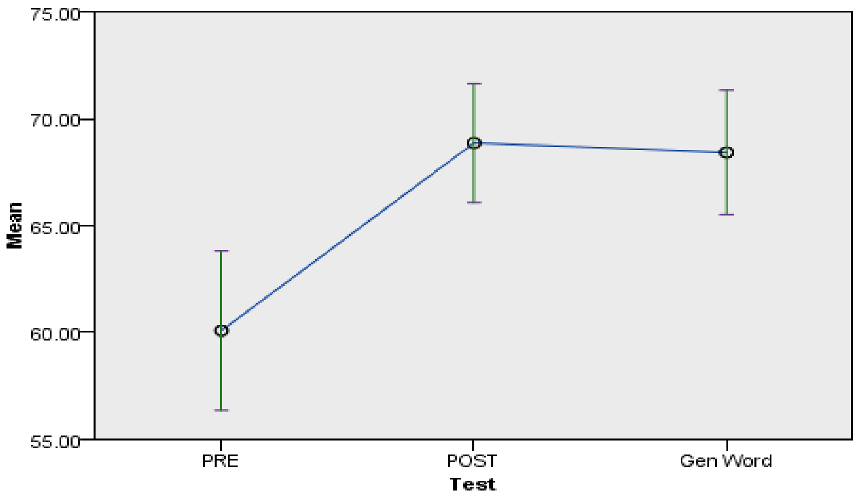
Given that the ID_V improved only on vowels (trained segments), the percentage scores obtained on vowel perception at the three different tests (pre-test, post-test and generalization 2) were entered in a generalized linear model (GLMM). The fixed effect entered in the model was test (pre-test, post-test and generalization 2) and participants was the random effect. The results yielded a significant effect of test ($F(2, 57) = 50.42$, $p < .001$), confirming that the ID group performed significantly better after the training than at pre-test, as evidenced in Figure 6.14. Furthermore, pairwise comparisons with a Bonferroni adjustment confirmed that the pre-test performance differed from both the post-test and the generalization test ($p < .001$). Conversely, post-test and generalization test results did not differ significantly. Thus, strong generalization to novel words was observed for the ID_V trainees.

Figure 6.14. Mean scores obtained by the ID_V group at pre-test, post-test and generalization to novel non-words tests



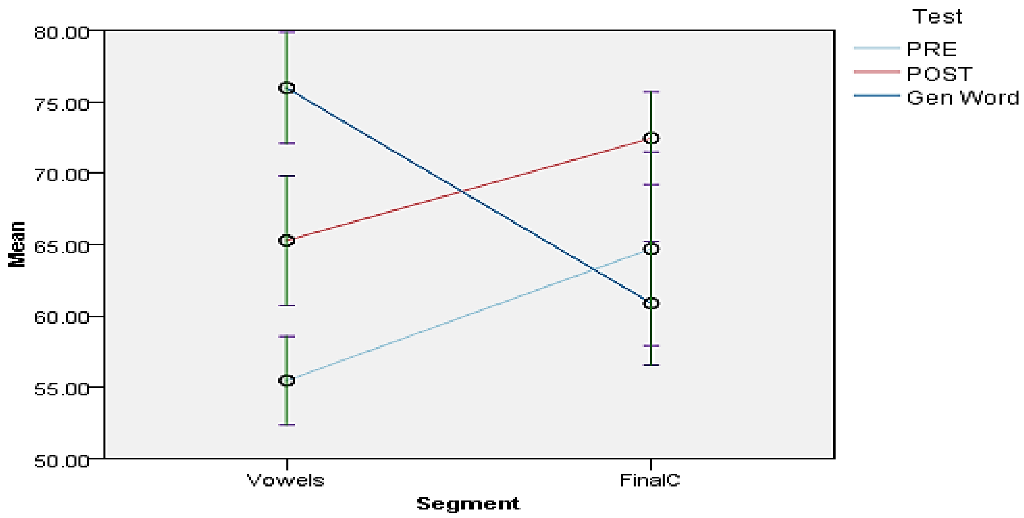
Since the DIS_V group presented positive training results on both vowels and final consonants (untrained segments), the percentage scores obtained for these two segments by the DIS_V trainees on the three tests were entered into a generalized linear mixed model (GLMM), in which the fixed effects were segment (vowels and final consonants), test (pre-test, post-test and generalization to novel non-words) and a segment by test interaction. The results yielded no significant effect of segment ($F(1, 102) = 0.06$, $p > .05$), a significant effect of test ($F(2, 102) = 8.06$, $p < .001$) and a significant segment by test interaction ($F(2, 102) = 19.44$, $p < .001$). Despite being trained on vowels, no significant effect of segment was observed by the model, indicating that the DIS_V group's overall mean scores for vowels (trained segments) and for consonants (untrained segments) were similar collapsing across the three testing phases. Regarding the significant effect of test, pairwise comparisons revealed that the pre-test scores significantly differed from both the post-test and the novel non-word generalization scores ($p < .01$ in both cases), whereas the two post-training assessment tests did not differ from one another ($p > .05$), as evidenced in Figure 6.15.

Figure 6.15. Mean scores obtained by the DIS_V group at pre-test, post-test and generalization to novel non-words tests



The significant segment by test interaction might be due to the fact that the scores for vowels increased from post-test to the novel non-words generalization test, but the scores decreased from post-test to the generalization test for final stops. Figure 6.16 illustrates this interaction for the two segments here analysed.

Figure 6.16. Mean scores by segment (vowels and final consonants) obtained by DIS_V group at pre-test, post-test and generalization 2



When looking at the results for each segment separately, simple test effects revealed a significant test effect for vowels ($F(2, 102) = 33.69$, $p < .001$) and a significant test effect for final consonants ($F(2, 102) = 9.48$, $p < .001$). Pairwise comparison confirmed that in the case of vowels, the pre-test scores significantly differed from both the post-test and the generalization scores ($p < .001$), confirming both the main results previously found and the fact that *strong generalization* took place for this segment. Interestingly, the generalization scores were significantly higher than the post-test results for the DIS_V trainees, suggesting that these tokens (either triggered by the novel words or by the familiarity with the training talkers) might have posed less difficulty for the L2 learners. Regarding the final consonant generalization, the pairwise comparisons pointed to a significant difference among the pre-test and post-test scores, confirming the previously reported main findings and no significant difference between the pre-test and the generalization scores. Moreover, the DIS_V group's performance on the generalization test was significantly lower than their performance obtained at post-test. Thus, the DIS_V group failed to generalize the learning to untrained sounds (final consonant performance) observed at post-test.

6.4.2.2. Consonant training groups

The percentage correct identification obtained for both trained (consonants) and untrained sounds (vowels) at pre-test, post-test and generalization test 2 by the two consonant training groups (ID_C, DIS_C) and the control group (CG) were calculated and are shown in Table 6.19. Generalization effects were assessed for the cases that reached significance from pre-test to post-test, that is, initial consonants (trained segment) for the ID_C group and both initial consonants (trained segments) and vowels (untrained segment) for the DIS_C. In the same fashion as previously done for vowels, these cases are indicated in boldface.

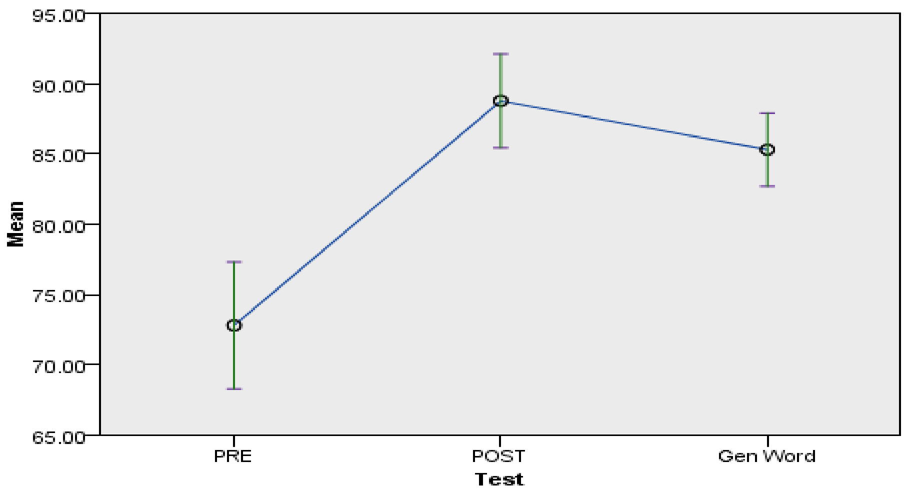
For the initial consonant identification, scores at the generalization test seem to be numerically similar to the ones at post-test. In the case of the vowels, higher scores were observed at the generalization test than at the post-test phase across groups (including the CG). For the remaining cases, little difference among the three tests is observed. As described for the vowel trained groups above, generalized linear mixed models (GLMM) were conducted on the percentage scores obtained by the consonant trained groups and CG at the three different tests. The results found for the ID_C trainees are explored first, followed by the results found for the DIS_C trainees.

The percentage scores obtained by the ID_C group on the identification of initial stops (trained segments) at the three different tests (pre-test, post-test and generalization 2) were entered in a generalized linear model (GLMM). The fixed effect was test (pre-test, post-test and generalization 2) and participants were entered as the random effect. The results yielded a significant effect of test ($F(2, 51) = 17.06$, $p < .001$), and Bonferroni adjusted pairwise comparisons confirmed that that the ID_C group performed significantly better at post-test and at generalization test than at pre-test, as evidenced in Figure 6.17. Furthermore, the post-test and the generalization performance did not differ significantly, confirming that a strong generalization to the trained segments took place.

Table 6.19. Pre-test, post-test and generalization scores (novel words) for trained and untrained sounds in non-words obtained by consonant trained groups and control group

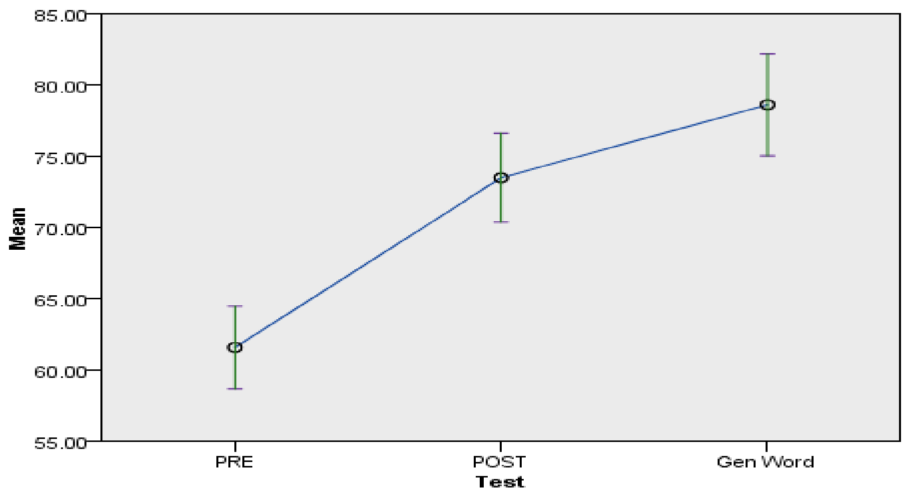
	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	**69.5**	9.6	**72.8**	9.5
POST	79	13.4	**82.3**	9.8	**88.7**	7.1
GEN_W	79.9	11.5	81.5	8.7	85.3	7.8
Trained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	70	16.8	69.9	8.8
POST	69.6	6.2	74.3	7.3	75.4	7.0
GEN_W	65.0	7.9	70.2	6.8	72.9	6.5
Untrained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.1	9.9	**55**	7.5	56.2	10.1
POST	57.8	10.2	**64.7**	8.4	55.3	12.2
GEN_W	68.4	11.3	75.8	8.2	62.9	8.9

Figure 6.17. Mean scores (initial consonants) obtained by the ID_C group at pre-test, post-test and generalization to novel non-words tests



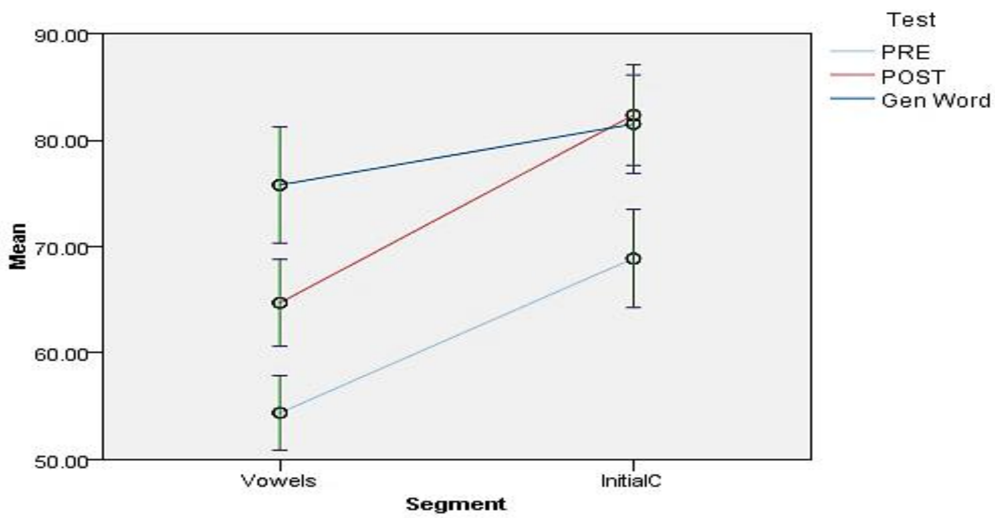
Turning our attention to the DIS_C group's performance, let us recall that this group had shown a significant gain (12.8%) when identifying initial consonants (trained segment) and a significant gain (9.7%) when identifying vowels (untrained segments). Thus, the percentage scores obtained by the DIS_C for these segments at the three given tests (pre-test, post-test and generalization to novel words) were introduced into a GLMM in the same fashion previously done for the other experimental group. The model pointed to a significant test effect ($F(2, 94) = 30.53$, $p < .001$), a significant segment effect ($F(1, 94) = 45.77$, $p < .001$) and a significant test by segment interaction ($F(4, 94) = 3.254$, $p < .05$). The test effect can be accounted for by the fact that the trainees performed significantly better after training than at pre-test, as evidenced in Figure 6.18. Bonferroni adjusted pairwise comparisons yielded a significant difference between the pre-test and the post-test performance (p.> 001), between the pre-test and the generalization test ($p > .001$) and between the post-test and the generalization test ($p > .05$).

Figure 6.18. Mean scores obtained by the DIS_C group at pre-test, post-test and generalization to novel non-words tests



The significant segment effect is due to the higher scores obtained when identifying initial consonants than when identifying vowels. Regarding the test by segment significant interaction, it might be accounted for by the fact that the improvement found at post-test was maintained in the case of initial consonant identification, whereas it increased in the case of the vowels (untrained segments). The interaction and the mean score for both segments here investigated obtained by the DIS_C group at pre-test, post-test and at generalization to novel non-words can be seen in Figure 6.19.

Figure 6.19. Mean scores by segment (vowels and final consonants) obtained by DIS_C group at pre-test, post-test and generalization 2



Having found a significant interaction, follow-up simple tests were carried out for each segment individually. The results yielded a significant main effect of test both for vowels ($F(2, 94) = 22.82$, $p < .001$) and for initial consonants ($F(2, 94) = 10.45$, $p < .001$). Regarding the initial consonant performance, Bonferroni adjusted pairwise comparisons confirmed that the pre-test identification scores significantly differed from both the post-test scores and the generalization scores ($p < .001$), and the latter two did not differ. Thus, strong generalization to novel initial consonant non-words was observed for the DIS_C group. In the case of vowels (untrained segment), pairwise comparisons with a Bonferroni adjustment confirmed that the three tests significantly differed from each other; pre-test vs. post-test/generalization test (p >.001) and post-test vs. generalization test (p <.01). These results confirm that the perception on the untrained segment (vowels) improved with training and generalized to novel stimuli (strong generalization).

Interestingly, these results point to the fact that the tokens used to test vowel perception at the generalization test posed less difficulty to the participants, since the CG also obtained higher scores at this phase of the study on vowel perception. However, as previously reported (refer back to section 6.4.2.1), a GLMM carried out on the vowel percentage scores obtained by the three groups that significantly improved on vowel perception from pre-test to post-test (ID_V, DIS_V and DIS_C) and the CG confirmed that the three experimental groups outperformed the controls. Thus, even though the stimuli seemed to be easier for all groups (even the CG), the three experimental groups generalized vowel learning.

Summary

Overall, the results for generalization 2 (novel non-words) revealed that both vowel training regimes (ID_V; DIS_V) were able to promote strong generalization to novel non-words produced by familiar talkers. However, the cross-training effect observed by the DIS_V from pre to post-test did not generalize to novel non-words. Similarly, the two consonant training groups (ID_C; DIS_C) also promoted strong generalization to initial consonants (trained segment), which had originally improved from pre to post-test. Additionally, the cross-training effect on vowels observed from pre-test to post-test for the DIS_C also generalized (strong generalization) to novel non-words. The following section will introduce the results found for generalization to novel talkers.

6.4.3. Generalization 3. Novel talkers

Another goal of this thesis was to assess whether the improvement obtained during the training regime was specific to the talkers heard during this phase, or if such improvement would generalize to novel talkers. As previously stated (Chapter 5, section 5.2), the training talkers and the testing talkers were not the same. Thus, an improvement from the pre-test to the post-test indicates a positive generalization to novel talkers. Figure 6.2 and the results reported in Section 6.2.1 confirm the significant improvement from pre-test to post-test for both vowel training groups and Figure 6.5 and the results in section 6.2.2 confirm the significant improvement from pre-test to post-test for both consonant trained groups.

Having established that learning generalized to novel talkers, that is, talkers not present in the training phase, this thesis aimed at investigating if the learning would generalize to *absolute new talkers*, that is, talkers present neither in the testing nor in the training phase. Talkers present in the testing phase can already be considered familiar talkers to some extent at post-test, as participants have already been previously exposed to them. Thus, this issue was further explored with the inclusion of two completely different talkers. Familiar tokens produced by two absolute novel talkers (T7 and T8) were presented to the learners at week 9 (in the generalization test, see section 5.2 for more info). The generalization test included a subset of the testing and training stimuli involving both trained and untrained sounds. The generalization scores for trained and untrained sounds obtained by the vowel training groups will be reported first, followed by the generalization scores for trained and untrained sounds obtained by the consonant training groups.

6.4.3.1. Vowel training groups

The percentage correct identification obtained for trained and untrained sounds at pre-test, post-test and generalization 3 test (*absolute new talkers*) by the two vowel training groups (ID_V, DIS_V) and the control group (CG) are shown in Table 6.20. As large differences were found for the results for each of the two talkers tested at this phase (T7 (female) and T8 (male)), separate results are shown for each talker. Moreover, due to a technical problem, no final consonant items were included for talker 8. Thus, final consonant generalization to absolute new talker will be based on the performance obtained for talker 7 only. As previously mentioned, generalization effects are only expected for the segments that improved with training (in boldface in Table 6.20), namely vowels for both experimental groups (ID_V and DIS_V) and final consonants for the DIS_V.

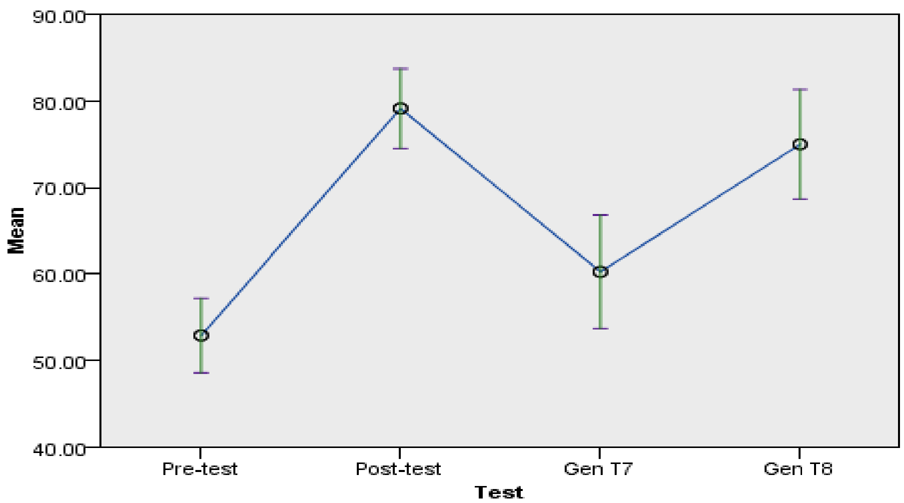
Table 6.20. Pre-test, post-test and generalization scores (Novel talkers) for trained and untrained sounds in non-words obtained by the vowel trained groups and the control group

	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.2	10	**55.6**	6.5	**53**	9.5
POST	57.8	10.2	**65.4**	9.7	**79.2**	10.2
GEN T7	49.1	16.7	48.8	16.0	60.2	14.7
GEN T8	57.5	15.3	61.1	16.1	75	14.2
Untrained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	74.3	9.7	77.6	11.9
POST	79.0	13.4	76.7	11.4	74.2	11.9
GEN T7	82.2	9.9	77.3	9.8	73.9	9.2
GEN T8	76.1	12.3	70.8	11.6	67	10.8
Untrained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	**64.7**	14.5	71	6.4
POST	69.6	6.29	**72.4**	6.9	69.4	10
GEN T7	71.2	11.9	66.2	8.6	68.1	8.3

As evidenced in Table 6.20, all participants seemed to identify vowels produced by talker T8 better than those produced by T7 and initial consonants produced by T7 better than those produced by T8. As the stimuli were successfully validated by native speakers before the beginning of this study, this might indicate that some specific characteristic present in the talkers' speech posed problems for the L2 learners. These talker specific characteristics are investigated and discussed further in section 6.6.3.

In order to assess if the ID_V group generalized their vowel learning to tokens produced by the absolute novel talkers, a GLMM was conducted on the vowel percentage scores obtained by these L2 learners at pre-test, post-test and for the two different speakers at generalization 3 test. The fixed effects entered in the model were tests (pre-test, post-test, generalization T7 and generalization T8). The results revealed a significant effect of test ($F(3, 76) = 26.77$, p >.001). Bonferroni adjusted pairwise comparison results are shown in Table 6.21. Figure 6.20 illustrates the performance obtained by the ID_V group at pre-test, post-test and at the generalization test for each absolute new talker separately.

Figure 6.20. Mean vowel scores obtained by the ID_V group at pre-test, post-test and generalization to two absolute novel talker tests (T7, T8)



As observed in Table 6.21, only the performance with vowel tokens produced by talker T8 significantly differed from the initial vowel performance at pre-test. Moreover, such performance did not differ from the performance obtained at post-test, indicating that a strong generalization to an absolute new talker (T8) took place. T7's results did not reach post-test values, probably due to the particularities of this talker's production of some vowels, as explained above and further discussed in section 6.6.3.

Table 6.21. Results of Bonferroni pairwise comparisons for pre-test, post-test and generalization 3 obtained by the ID_V group on vowels

Comparison	Bonferroni adjusted significance
Pre-test vs. Post-test	p<.001***
Gen T7 vs. Pre-test	p>.05
Gen T8 vs. Pre-test	p<.001***
Gen T7 vs. Post-test	p<.001***
Gen T8 vs. Post-test	p>.05
Gen T7 vs. Gen T8	p<.01**

Regarding the DIS_V group, let us recall that there was an initial improvement for both vowels (trained segments) and final consonants (untrained segments). The results for the trained segments will be reported first, followed by the results for the untrained segments. Looking at vowel generalization performance for the DIS_V, percentage correct vowel identification scores at the four tests (pre-test, post-test, gen T7 and gen T8) were entered in a GLMM as fixed effects. The model revealed a significant effect of test ($F(3, 56) = 5.25$, $p < .01$) and the results of the pairwise comparisons with a Bonferroni correction can be seen in Table 6.22. Figure 6.21 illustrates the vowel performance obtained by the DIS_V group at the four different tasks.

Figure 6.21. Mean vowel scores obtained by the DIS_V group at pre-test, post-test and generalization to two absolute novel talker tests (T7, T8)

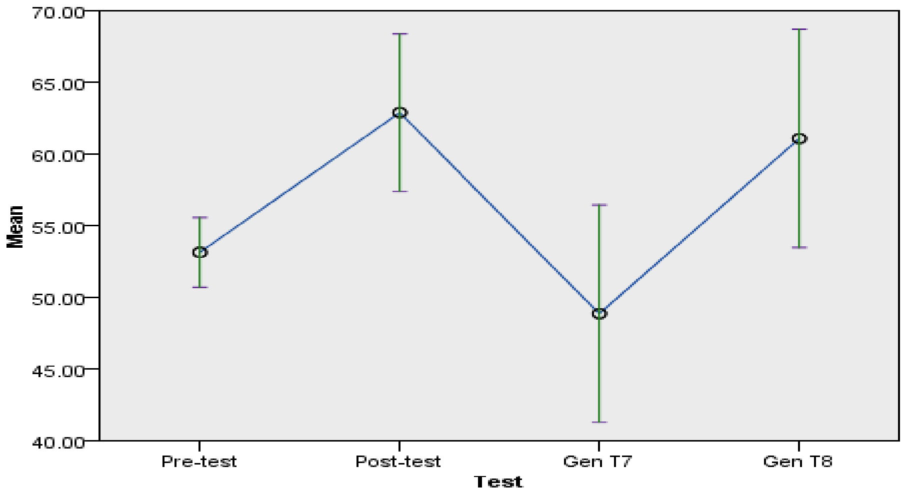


Table 6.22. Results of Bonferroni pairwise comparisons for pre-test, post-test and generalization 3 obtained by the DIS_V group on vowels

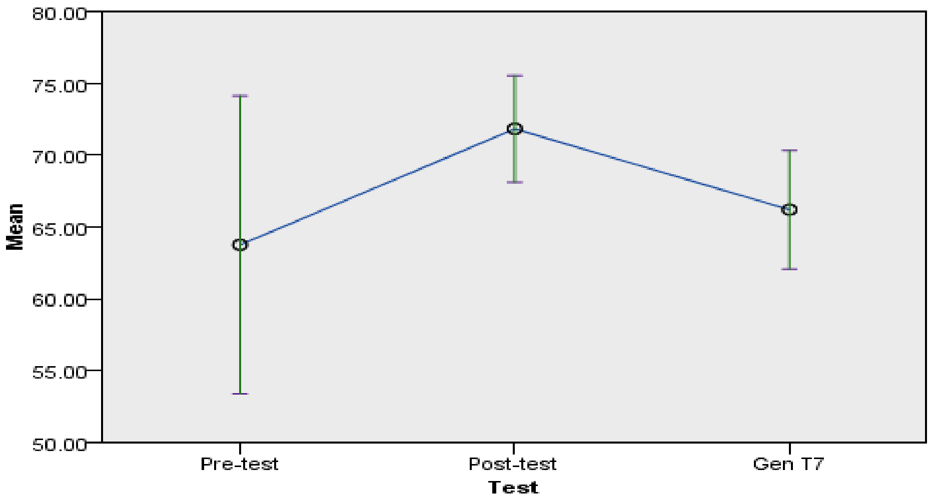
Comparison	Bonferroni adjusted significance
Pre-test vs. Post-test	p<.05*
Gen T7 vs. Pre-test	p>.05
Gen T8 vs. Pre-test	p>.05
Gen T7 vs. Post-test	p<.05*
Gen T8 vs. Post-test	p>.05
Gen T7 vs. Gen T8	p>.05

The results obtained for the DIS_V group revealed that the identification performance with the items produced by talkers T7 and T8 did not differ from the pre-test performance, indicating that strong generalization to novel talkers did not take place for this experimental group. However, post-test scores and generalization to T8 scores did not differ significantly either, illustrating a weak generalization effect. By contrast, there was no generalization to words produced by T7 as the results in this case were significantly lower than the post-test results. Moreover, possibly due to the large standard deviation found in the data (see Figure

6.20), the identification scores obtained for both talkers did not differ statistically from one another.

The DIS_V group also improved their perception on final consonant sounds as a result of the vowel training regime received. In order to assess the robustness of this "cross-training" improvement, a GLMM on the final consonant scores obtained by the DIS_V at the three different tests (pre-test, post-test and generalization T7) was conducted. The model yielded no effect of test ($F(2, 39) = 2.65$, $p > .05$), as pairwise comparisons revealed that the three tests did not differ among themselves. Possibly, the significant result previously found in the analysis of gain was not replicated here due to large standard deviation from the data at pre-test (see Figure 6.22) and the fact that more variables were entered in the model (group and segment) rather than a single effect as done here. The absence of a significant difference between the post-test and the generalization test suggests that the post-test scores were generalized to novel words (weak generalization). However, the generalization scores were not robust enough to differ from the pre-test scores. The final consonant performance at the three tests here investigated can be seen in Figure 6.22.

Figure 6.22. Mean final consonant scores obtained by the DIS_V group at pre-test, post-test and generalization to an absolute novel talker test (T7)



6.4.3.2. Consonant training groups

Table 6.23 shows the scores obtained for both trained (consonants) and untrained sounds (vowels) at pre-test, post-test and *novel talker generalization test* by the two consonant training groups (ID_C, DIS_C) and the control group (CG). The effects which were found to be significant from pre-test to post-test are highlighted.

Table 6.23. Pre-test, post-test and generalization scores (Novel talkers) for trained and untrained sounds in real words obtained by consonant trained groups and control group

	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	78.1	12.4	**69.5**	9.6	**72.8**	9.5
POST	79	13.4	**82.3**	9.8	**88.7**	7.1
GEN T7	82.2	9.9	83.8	10.5	86.3	7.1
GEN T8	76.1	12.3	73.4	9.7	78.9	9.2
Trained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	69.1	7.8	70	16.8	69.9	8.8
POST	69.6	6.2	74.3	7.3	75.4	7.0
GEN T7	71.2	11.9	73.2	11.1	75.5	11.2
Untrained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	54.1	9.9	**55**	7.5	56.2	10.1
POST	57.8	10.2	**64.7**	8.4	55.3	12.2
GEN T7	49.1	16.7	53.7	11.6	46.4	20.9
GEN T8	57.5	15.3	59	12.7	42.8	17.4

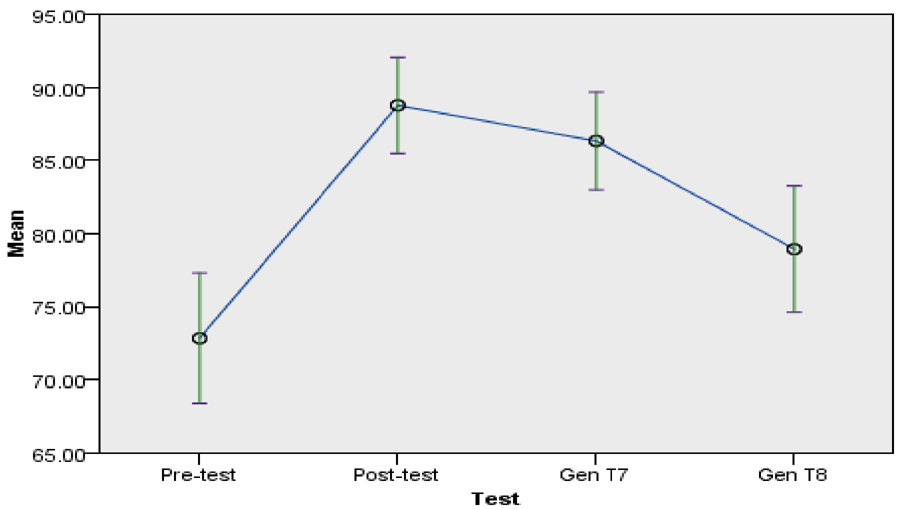
As previously observed, there was a difference in performance depending on the talker (T7 or T8). In the case of initial consonant sounds, participants seemed to identify T7's productions with more ease than T8's productions. The reverse pattern is observed for vowel perception. Analyses were conducted on the scores obtained by the consonant training groups. The results for the ID_C will be reported first, followed by the results for the DIS_C.

The GLMM on the initial consonant scores for the ID_C revealed a significant effect of test ($F(3, 68) = 13.45$, $p < .001$) and Bonferroni pairwise comparisons are shown in Table 6.24. Only the performance with initial consonant tokens produced by talker T7 significantly differed from the initial performance at pre-test. Consistently, such performance did not differ from the performance obtained at post-test, indicating that a strong generalization to the absolute new talker (T7) took place. By contrast, no generalization to talker 8's stimuli was observed, as the results in this case were comparable to pre-test and significantly lower than at post-test. Moreover, the scores obtained with tokens produced by the two talkers significantly diverged from one another. These results confirm that the ID_C group generalized the initial consonant learning only for one talker, talker 7. Figure 6.23 illustrates the performance obtained by the ID_C group at pre-test, post-test and at the generalization test for each absolute new talker separately.

Table 6.24. Results of Bonferroni pairwise comparisons for pre-test, post-test and generalization 3 obtained by the ID_C group on initial consonants

Comparison	Bonferroni adjusted significance
Pre-test vs. Post-test	p<.001***
Gen T7 vs. Pre-test	p<.001***
Gen T8 vs. Pre-test	p>.05
Gen T7 vs. Post-test	p>.05
Gen T8 vs. Post-test	p<.01**
Gen T7 vs. Gen T8	p<.05*

Figure 6.23. Mean initial consonant scores obtained by the ID_C group at pre-test, post-test and generalization to two absolute novel talker tests (T7, T8)



The analysis of the DIS_C's performance with initial consonants replicated the results found with the ID_C group (see Figure 6.24). The GLMM yielded a significant effect of test ($F(3, 62) = 8.72$, $p <.001$) and Bonferroni adjusted pairwise comparisons revealed that DIS_C generalized initial consonant learning only for talker 7 (see Table 6.25). As explained above, the fact that generalization to new talker is only found with T7 in the case of initial stops is likely related to some specific characteristic found in T8's speech, which hindered accurate perception. This issue is further explored in result section 6.6.3.

Figure 6.24. Mean initial consonant scores obtained by the DIS_C group at pre-test, post-test and generalization to two absolute novel talker tests (T7, T8)

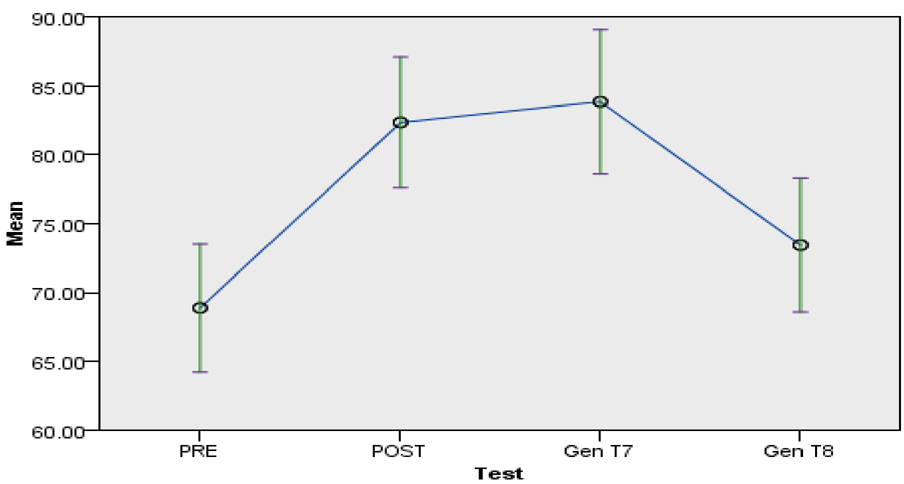


Table 6.25. Results of Bonferroni pairwise comparisons for pre-test, post-test and generalization to two absolute novel talker tests (T7, T8) obtained by the DIS_C group on initial consonants

Comparison	Bonferroni adjusted significance
Pre-test vs. Post-test	p<.001***
Gen T7 vs. Pre-test	p<.001***
Gen T8 vs. Pre-test	p>.05
Gen T7 vs. Post-test	p>.05
Gen T8 vs. Post-test	p<.05*
Gen T7 vs. Gen T8	p<.05*

With respect to the untrained sounds (vowels), recall that an improvement from pre-test to post-test was found for the DIS_C group only. Figure 6.25 illustrates the vowel performance obtained by this group at pre-test, post-test and at the generalization test for each absolute new talker separately. The GLMM yielded a significant effect of test ($F(3, 62) = 5.72$, $p < .01$). Bonferroni adjusted pairwise comparisons are reported in Table 6.26.

Figure 6.25. Mean vowel scores obtained by the DIS_C group at pre-test, post-test and generalization to two absolute novel talker tests (T7, T8)

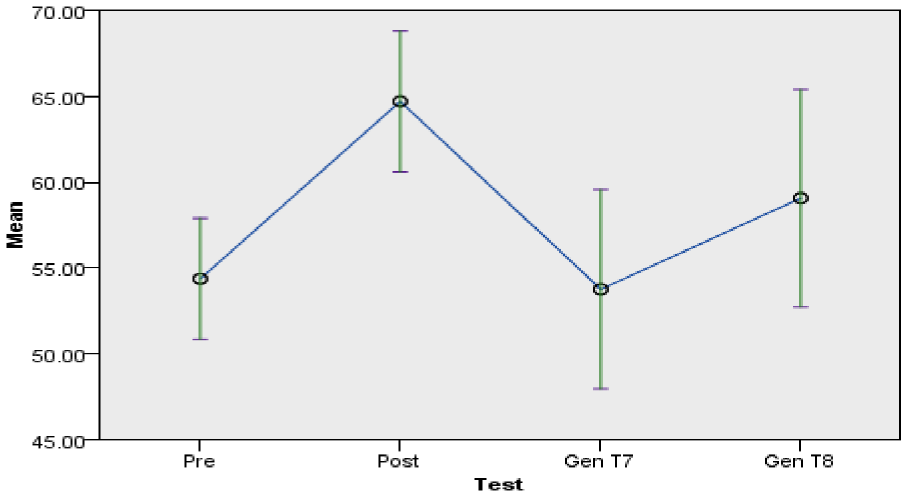


Table 6.26. Results of Bonferroni pairwise comparisons for pre-test, post-test and generalization to two absolute novel talker tests (T7, T8) obtained by the DIS_C group on vowels

Comparison	Bonferroni adjusted significance
Pre-test vs. Post-test	p<.01**
Gen T7 vs. Pre-test	p>.05
Gen T8 vs. Pre-test	p>.05
Gen T7 vs. Post-test	p<.05*
Gen T8 vs. Post-test	p>.05
Gen T7 vs. Gen T8	p<.05*

As observed in Table 6.26, the initial performance at pre-test did not differ from the vowel performance with tokens produced either by talker T7 or by T8. However, only the performance with T8's stimuli was comparable to the performance at post-test, indicating that weak generalization to the absolute new talker (T8) took place. Generalization to T7 did not take place for the DIS_C, in line with the vowel perception results for T7 obtained by the vowel training groups. Moreover, the scores obtained with tokens produced by the two talkers significantly diverged from one another. Such results confirm that the DIS_C group generalized (weak generalization) the vowel learning through consonant training (cross-training effect) for one of the tested talkers, T8.

Summary

Generalization to novel talkers was already assessed at the testing phase (pre-test and post-test), and thus the results reported in section 6.2 and discussed in section 6.3 reflect on each group's ability to generalize their learning to unfamiliar voices. This section reported the results of generalization to "absolute novel talkers" and revealed that, overall, generalization to absolute novel voices was talker-specific. It was observed that some specific characteristics present in the talkers' speech posed problems for the Catalan learners of

English. This issue is discussed further in section 6.6.3. Both vowel training groups failed to generalize learning for stimuli produced by talker T7, whereas generalization to T8 took place for both groups. However, while the ID_V showed a strong generalization effect, the DIS_V showed a weak generalization effect on vowels. Moreover, the cross-training effect found for DIS_V on final consonants was generalized (weak generalization) in the case of T7. Turning our attention to the consonant training groups (ID_C and DIS_C), both experimental groups failed to generalize initial consonant learning in the case of T8. However, both consonant groups generalized learning to the same extent for T7, as in both cases "strong generalization" was observed. Moreover, the cross-training effect observed on vowels was also generalized (weak generalization) when identifying stimuli produced by T8.

6.5. Delayed post-test results

Two months after the post-test completion, a delayed post-test was administered. The aim of this test was to assess the long-term effects of the training procedures here applied. Given that fewer participants took part in this last phase of the study, the subsequent analyses are only taking into account the trainees that completed the three testing phases, which explains the different values presented here and in previous sections. The total number of participants at this phase was less homogeneous among groups, as there were 9 controls, 17 ID_V trainees, 12 DIS_V trainees, 12 ID_C participants and 13 DIS_C participants. The results for the subset of participants from the vowel training groups (ID_V, DIS_V) will be presented first, followed by the results for the subset of participants from the consonant training groups.

6.5.1. Vowel training groups

Correct identification percentages for the trained and untrained segments obtained by the vowel trained groups and the controls at pre-test, post-test and delayed post-test were calculated and can be seen in Table 6.27. The three groups performed either numerically higher or similarly to the post-test on the delayed post-test, including the CG[24]. Similarly to the generalization rationale argued before, some studies consider positive results on the delayed post-test by comparing it to the post-test scores (Hazan et al, 1995; Flege, 1995b; Cebrian & Carlet, 2014; Rato, 2014). No significant difference when comparing these two times reveals a positive retention of the knowledge acquired during training. The present study will use the same measure with the addition of the comparison to the scores obtained at pre-test. As was done in the previous section, cases where significant gain from pre-test to post-test was found are highlighted in boldface in the Table.

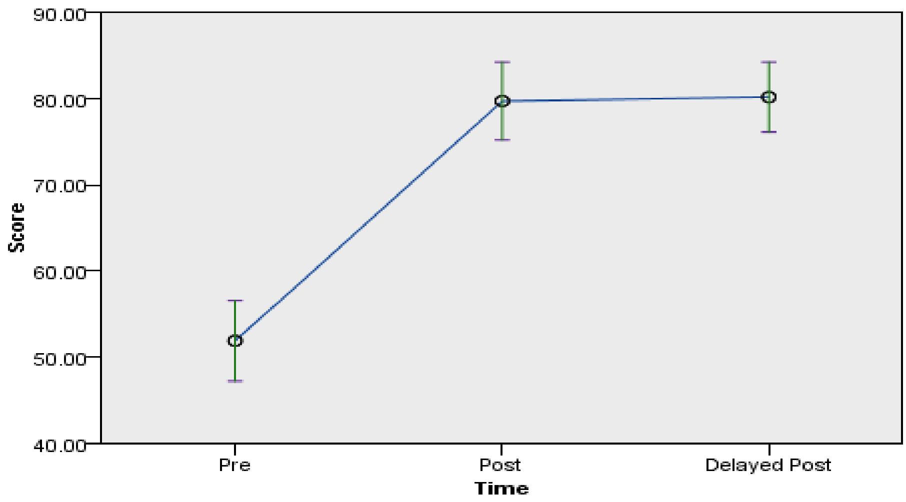
[24] Analyses were conducted on the scores for the CG and no significant effect of time was found (F (2, 72) = 1.84, p >.05).

Table 6.27. Pre-test, post-test and delayed post-test scores for trained and untrained sounds in non-words obtained by the vowel trained groups and the control group

	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	56.7	11.3	**53.0**	4.2	**51.8**	9.7
POST	61.9	11.1	**62.8**	9.4	**79.7**	9.3
DELAYED POST	63.3	14.0	60.4	8.2	80.1	8.3
Untrained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	79.2	15.0	72.9	9.5	80.0	11.1
POST	82.6	14.7	74.3	11.7	75.4	12.2
DELAYED POST	80.2	13.9	76.7	10.5	78.3	12.4
Untrained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	72.9	5.9	**63.7**	17.9	71.7	6.3
POST	71.5	6.4	**71.7**	6.6	70.6	9.0
DELAYED POST	74.1	7.6	70.8	6.9	74.1	6.5

Each group's performance is analysed individually, and the ID_V's results are reported first, followed by the results found by the DIS_V. In order to verify if the ID_V group retained the improvement on trained sounds two months later, the identification scores obtained at the three different times (pre-test, post-test and delayed post-test) were entered into a GLMM with time as the fixed effect. The model yielded a significant effect of time ($F(2, 48) = 51.35$, $p < .001$) and Bonferroni adjusted pairwise comparisons confirmed that the performance obtained at pre-test significantly differed from the performance at post-test and delayed post-test ($p < .001$ in both cases). Importantly, the delayed post-test results did not differ from the post-test results, confirming that learning was retained for a period of two months. Figure 6.26 illustrates the pre-test, post-test and delayed post-test training performance for the ID_V group.

Figure 6.26. Mean scores on vowels obtained by the ID_V group at pre-test, post-test and delayed post-test



Turning our attention to the DIS_V group and first assessing the direct effect obtained with vowels (see Figure 6.27), the scores obtained at the three testing times (pre-test, post-test and delayed post-test) were introduced in a GLMM, which yielded a significant effect of time ($F(2, 33) = 7.62$, $p < .002$). Pairwise comparisons with a Bonferroni correction confirmed that knowledge was retained over time, as the delayed post-test performance differed from the pre-test performance and did not differ from the post-test performance.

Regarding the long term effects of the improvement obtained by the DIS_V with the untrained segments (final consonants), the GLMM analysis in this case revealed no significant effect of time ($F(2, 33) = 1.10$, $p > .05$). The lack of significance from pre-test to post-test found here is probably accounted for by the large standard deviation in the current pre-test data (17.9, see Figure 6.28), since fewer participants participated at this stage and therefore only the scores for the participants who took part on the three stages are analysed. Recall that the main result found with the full set of participants reached significance (refer back to section 6.2.1) and that is the reason to investigate whether the knowledge was retained. Despite the lack of significance, numerically the results follow the same pattern as with vowels, with a notable increase from pre-test to post-test (64-72%) and the delayed post-test score (71%) being closer to the post-test score than to the pre-test.

Figure 6.27. Mean scores on vowels obtained by the DIS_V group at pre-test, post-test and delayed post-test

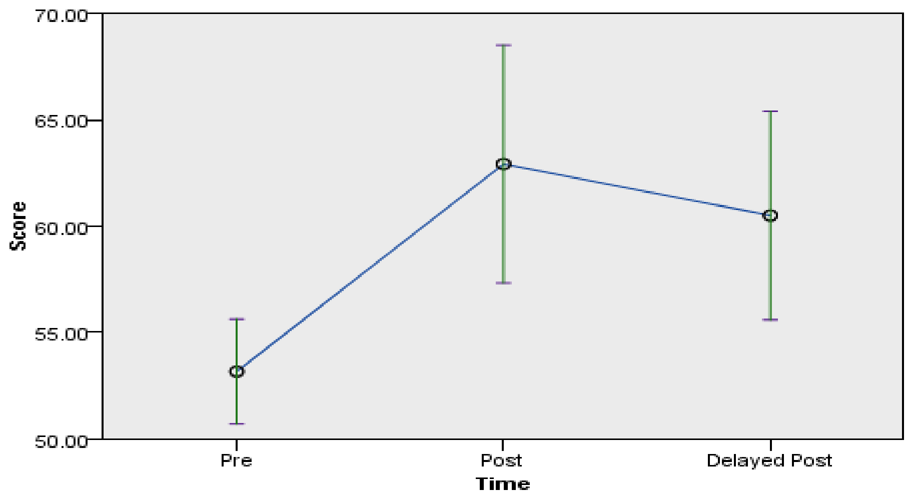
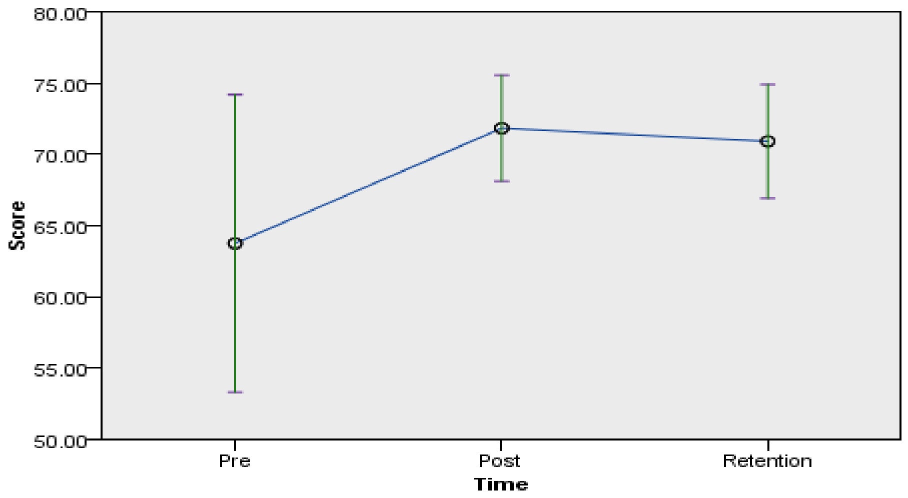


Figure 6.28. Mean scores on final consonants obtained by the DIS_V group at pre-test, post-test and delayed post-test



6.5.2. Consonant training groups

Moving our attention to the consonant training groups (ID_C and DIS_C), correct identification percentages for trained and untrained segments obtained by the subset of participants who participated at the three testing times (pre-test, post-test and delayed post-test) were calculated and can be seen in Table 6.28. As the training regime received only resulted in identification improvement of initial consonants (trained segment) for the ID_C group and of initial consonants (trained segments) and vowels (untrained segment) for the DIS_C, generalization effects are only assessed for these segments for each training group. In the same fashion as previously done for vowels, these cases are shown in boldface in the Table.

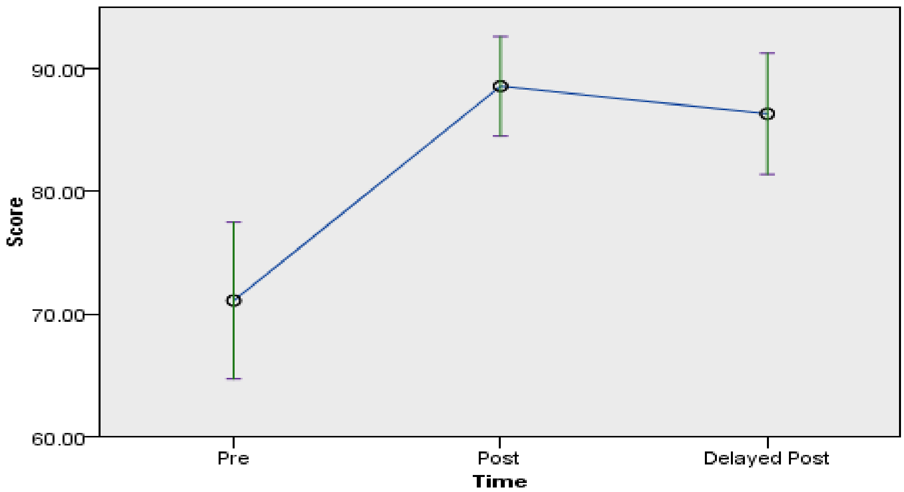
Table 6.28. Pre-test, post-test and delayed post-test scores for trained and untrained sounds in real words obtained by consonant trained groups and control group

	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	79.2	15.0	**71.8**	9.4	**71.0**	10.9
POST	82.6	14.7	**84.6**	9.3	**88.4**	6.8
DELAYED POST	80.2	13.9	86.4	9.7	84.3	14.4
Trained sounds Final consonants	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	72.9	5.9	73.2	9.0	71.2	8.6
POST	71.5	6.4	75.6	6.3	75.2	6.0
DELAYED POST	74.1	7.6	76.3	7.4	75.2	7.5
Untrained sounds Vowels	**%**	**SD**	**%**	**SD**	**%**	**SD**
PRE	56.7	11.3	**53.8**	7.6	57.4	10.5
POST	61.9	11.1	**64**	9.3	57.3	10.6
DELAYED POST	63.3	14.0	65.2	17.5	60.0	8.3

Comparing pre-test and post-test results, there was the improvement on initial consonant perception by both the DIS_C (12.8%) and ID_C (15.9%) and on vowel perception by the DIS_C (9.7%). Crucially, as evidenced in Table 6.28, all groups performed numerically similarly or better at the delayed post-test than at the post-test, showing that no group experienced a large decline on the identification scores after two months. In order to assess if the experimental groups retained the knowledge acquired during training over a period of two months, a generalized linear mixed model (GLMM) was again conducted on the percentage scores obtained by the trainees at the three different times. As previously done, the results found for the ID_C group are shown first, followed by the results found for the DIS_C group.

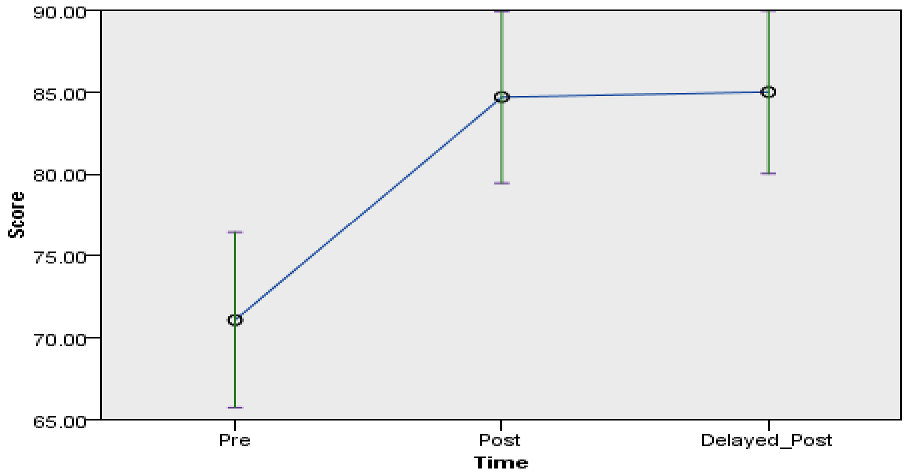
The percentage scores obtained for initial consonant perception by the ID_C at the three different times were entered in a GLMM, which yielded a significant effect of time ($F(2, 33) = 11.46$, $p < .001$), and Bonferroni adjusted pairwise comparisons confirmed that the ID_C group performed significantly better at post-test and at delayed post-test than at pre-test ($p < .001$ and $p < .01$ respectively). Additionally, the post-test and the delayed post-test performance did not differ significantly, confirming that initial consonant learning was retained over two months. Figure 6.29 illustrates the ID_C's performance at the three testing times.

Figure 6.29. Mean scores on initial consonant obtained by the ID_C group at pre-test, post-test and delayed post-test



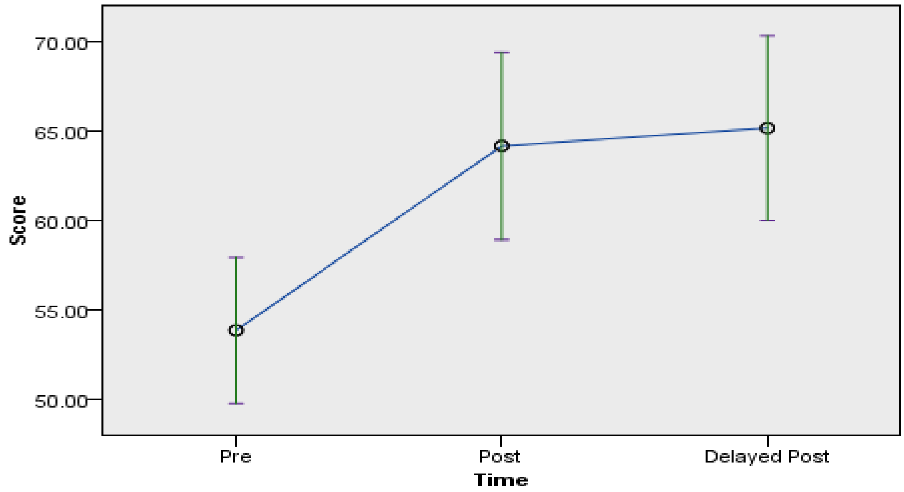
With the aim of examining the DIS_C group's performance with initial consonants, a GLMM was conducted on the scores obtained for initial consonant perception. The model revealed a significant effect of time, $F(2, 36) = 9.369$, $p < .01$. Bonferroni pairwise comparisons established that the pre-test differed from the post-test and the delayed post-test, confirming that long term effects of training took place for the DIS_C group. Figure 6.30 illustrates the identification performance of the DIS_C group.

Figure 6.30. Mean scores on initial consonant obtained by the DIS_C group at pre-test, post-test and delayed post-test



Finally, looking at the delayed post-test performance of the DIS_C with the untrained segments (vowels), the GLMM in this case also yielded a significant effect of time ($F(2, 36) = 7.98$, $p < .001$), and pairwise comparisons confirmed that the pre-test differed from the other two tests ($p < .01$ in both cases), which did not differ between themselves. Figure 6.31 shows the performance obtained by the DIS_C with the untrained sounds at the three testing times: pre-test, post-test and delayed post-test.

Figure 6.31. Mean scores on vowels obtained by the DIS_C group at pre-test, post-test and delayed post-test



Summary

This section looked at the long term effects of the two training regimes under investigation (ID vs. DIS) on trained and untrained segments. It is important to recall that the analysis presented here focused solely on those participants who completed all three tasks. Results revealed that all experimental groups retained their knowledge over a period of two months on directly trained sounds, that is, on vowel sounds for the vowel training groups (ID_V and DIS_V) and for initial consonants for the consonant experimental groups (DIS_C and ID_C). Regarding the cross-training effect, the DIS_C retained the vowel knowledge gained as a result of the consonant training. In turn, the DIS_V failed to retain the final consonant knowledge over two months, although numerical differences were still observed. All in all, these findings emphasize the positive impact of the perceptual training regime received by the experimental groups, as all groups retained the knowledge obtained during training over a period of two months in almost all instances.

6.6. Discussion II – Generalization and retention effects

This part of the study explored whether or not the two different perceptual training methods (Categorical DIS and ID) promoted generalization and/or retention effects. More specifically, it addressed the third and fourth research questions, which were concerned with measuring the robustness of the training methods by assessing generalization to real words, novel non-words and novel talkers (RQ.3) and retention of learning over a period of two months (RQ.4). The generalization effects will be discussed in the same order the results have been presented (real words, novel non-words and novel talker). Subsequently, retention effects, which were measured through a delayed post-test two months after training, will be discussed.

6.6.1. Generalization 1 – Real words

The first type of generalization dealt with the effect of the two HVPT methods (ID vs. categorical DIS) on the perception of real word stimuli, even though training strictly made use of non-words. Using non-word stimuli in training was chosen with the intention of avoiding the influence of lexical knowledge in the outcomes of the study (Grosjean, 1980). Knowing that sound perception in real words and in non-words triggers dissimilar sources of linguistic information (Ziegler, Besson, Jacobs, Nazir & Carr, 1997), comparing leaners' perception of sounds embedded in these two types of stimulus is not ideal (Hazan et al., 2005). Thus, real word stimuli were tested at two instances (pre-test and post-test) and were analysed in terms of the amount of gain from pre-test to post-test.

Recall that the analysis of gains on non-words from pre-test to post-test showed that both vowel training groups outperformed the CG and that the ID_V group outperformed the DIS_V in vowel perception. Moreover, the DIS_V group improved on the perception of final consonant stimuli, as a result of the vowel training regime received. Regarding the consonant training groups (ID_C and DIS_C), both experimental groups improved their perception of initial consonant stimuli to a similar extent and the DIS_C group additionally improved on the perception of the untrained sounds, vowels.

The results obtained through real word identification by the two vowel training groups (ID_V, DIS_V) show a different picture. The ID_V was the only group that outperformed the controls and, thus, the only group that generalized the learning acquired through training to real words. The DIS_V group's performance with real words at pre-test was numerically higher than the ID_V's (78% vs. 73%), and there was no change after training (79% vs. ID_V's 86%). This result points to two important facts. First of all, it provides further evidence of the robustness of the ID training method, in line with previous studies (Nishi & Kewley-Port, 2007; Pierce, 2014), and secondly, it indicates that an ID training method is superior to a categorical DIS method when promoting generalization to real words stimuli, in line with results found in early training studies that used auditory rather than categorical tasks (Strange & Dittmann, 1984; Jamieson & Moroson, 1986). A possible explanation for this superiority might be connected to the presence of labels in the ID task, which provided learners with focus on phonetic form (i.e. phonetic symbols and/or orthography), which is said to impact speech perception (Saito, 2015).

It is relevant to recall that the DIS_V group was asked to differentiate two stimuli heard consecutively, in terms of their categorical similarity and/or dissimilarity. These trainees did not have the information as to which sound categories they were hearing, as all sounds were presented in a randomized order. Moreover, during feedback, the information provided to the learners simply informed them about whether or not the two stimuli previously heard belonged to the same category. However, it did not explicitly tell the learners which category each sound belonged to. In fact, some learners' comments in the post-training survey (Appendix H) illustrate their desire to know which sound categories they were attending to during the DIS training. In turn, the vowel ID trainees were asked to label each sound they heard into a given L2 category, and through feedback they were informed as to the accuracy of their labelling. Identification is a type of *covert* task (Bohn, 2002), in which the single category presented at each trial is directly compared with a pre-existing memory representation, in the absence of another physically present category. Moreover, alongside the phonetic symbol, each label or response alternative in the vowel ID training task also contained two written word examples of each sound category (e.g., /iː/ - cheese/leaf; /ɜː/- earth/first). Thus, during each trial, the identification group was forced to relate the sound they heard to a given phonetic symbol and/or a familiar spelling. There may be a link between the use of phonetic symbols and orthographic representation and the generalization to real words.

In addition to the positive impact of focus on form on speech perception, the presence or absence of spelling has been found to affect the results of L2 word recognition studies, since orthography has a positive impact on L2 lexical representations (Escudero, Hayes-Harb & Mitterer, 2008). Even though this is a different area of investigation, similar reasons may apply to the identification of L2 sounds. For instance, Escudero et al. (2008) exposed two groups of Dutch learners of English to novel L2 lexical items containing a confusable L2 sound distinction for these learners (/æ/- /ɛ/). One group was provided with the auditory form only and the other group was exposed to the orthographic form in addition to the auditory form. Word recognition abilities were measured in an eye-tracking paradigm, which provided evidence that the information of the spelled forms during training facilitated accurate lexical activation. The authors concluded that "auditorily confusable L2 words which are learned only on the basis of their auditory forms will not result in a lexical contrast" (p. 358). Thus, it is possible that the presence of both elements in the ID labels (phonemic symbol and orthography) might account for the superiority of the ID method over the DIS method found in the present study when generalizing learning to real word stimuli.

Moreover, given that orthographic representations, phonological encoding and memory for phonologic forms interact (Schiller, 1998; Bassetti, 2006; Escudero et al., 2008), a possible explanation would be that, throughout training, the ID_V trainees were relating the sounds they heard to previously stored representations of real words (e.g. the ones provided by the response choices). According to Schiller (1998), orthography primes phonological representations and words are acquired by associating the orthographic form to a phonological form (Pierce, 2014; Piske, 2008). On the other hand, the DIS training did not make use of either the phonetic symbol or any orthographic representation. Due to the absence of labels, the DIS_V learners had to rely solely on the acoustic signal whilst discriminating the target sounds. Consequently, previously stored representations of real words might not have been activated during training and no lexical contrast for the testing stimuli was created (Escudero et al., 2008). Overall, the results indicate that the ID method is more suitable than a categorical DIS task as a training method for vowels, in line with some previous findings (Nozawa, 2015). The difference between DIS and ID may be further explained by the fact that an ID task enhances between-category distinctions rather than within-category distinctions (Jamieson & Moroson, 1986) and taps into higher levels of phonological encoding than the category DIS task (Iverson et al., 2003; Iverson et al., 2012).

When looking at the performance of the consonant trained groups (ID_C and DIS_C) with real word stimuli, it was observed that neither group generalized the consonant learning to real words to a statistically significant degree, despite the notable amount of gain (13.5%) obtained by the ID_C group. It is hypothesized that the lack of significant effect reported here is partly connected to the high scores obtained in consonant identification embedded in real words at the outset of the study and partly connected to the short duration of the training regime. Even though this improvement did not reach significance, the numerical improvement indicates that the ID_C benefitted more than the DIS_C group from the training received, performing numerically better when identifying real word stimuli. Connecting these results to the argument developed previously regarding the presence/absence of labels, it can be said that the numerical advantage observed for the ID_C group might be connected to the presence of the response alternatives in the identification task and the focus on form that they provided during training (Saito, 2015).

Another noteworthy outcome is the fact that learners identified the vowel target sounds better when embedded in real words (generalization 1) than when embedded in non-words (main pre-test/post-test comparison). This may indicate that learners found it easier to recognize the vowels when they were found in words that they recognized. These results relate to previous findings from lexical decision task studies, which report that learners perform faster and more accurately when identifying real words than non-word stimuli (Solé, 2013). According to Solé (2013), L2 contrasts that are not easily distinguishable in non-words, but can be differentiated in real words, indicate that L2 phonological categories may be formed after lexical categories, which are learned as a whole. Interestingly, the same effect was not observed with stop consonants, since stops were identified to a similar extent in real words and non-words. It is possible that L2 learners may have formed separate phonological categories for consonants and therefore, the lexical status of the word (word or non-word) is not at play, whereas this may not be the case for vowels. Possible reasons may be connected to the fact that consonant sounds are more relevant for lexical processing than vowels (Nespor, Peña & Mehler, 2003) and to the fact that consonants are perceived and learned with more ease than vowels (Pereira, 2014).

6.6.2. Generalization 2 – Novel non-words

Generalization 2 assessed whether the improvement that resulted from the training regime received was stimuli-specific or effective enough to promote generalization to novel non-words produced by training talkers. Thus, novel non-words produced by familiar talkers were tested a week later. Recall that all four experimental groups were found to improve from pre-test to post-test in their identification of trained sounds, that is ID_V and DIS_V on vowels and ID_C and DIS_C on initial consonants. Moreover, both DIS training groups (DIS_V and DIS_C) significantly improved their perception of untrained segments from pre-test to post-test. No significant effect on final consonant stimuli was evidenced as a result of direct consonant training , most probably due to the greater perceptual difficulty that this type of stimuli entails (Jang, 2014) and the need for a longer training regime (Flege, 1989). Remarkably, the DIS_V trainees were the only group to show improvement on their perception of final consonant stimuli through a cross-training effect. That is, a vowel training regime improved learners' final consonant perception (see section 6.3.1 for more details).

In all cases, the gain on the trained segments was maintained or even increased a week later when tested on different words produced by familiar talkers (generalization 2b). The positive generalization results (strong generalization in all cases) evidenced by the experimental groups on trained segments confirm that both training methods here applied (ID and categorical DIS) promoted generalization effects that went beyond the training stimuli, providing evidence of robustness of learning (Logan & Pruitt, 1995). This result corroborates previous findings in the literature that attest that both training methods (ID and categorical DIS) are effective (Flege, 1995b). Moreover, this result goes in line with previously reported positive generalization effects for trained segments in HVPT regimes (Lacabex et al., 2009; Rato, 2014; Pereira, 2014; among several others).

Results revealed that in the case of vowels, the perception of the novel non-words seemed to be slightly easier to the learners, as even the CG performed numerically better at this stage than at post-test. This indicates that these tokens posed less of a challenge for the learners than the testing tokens. As the tokens were comparable to the testing tokens in nature (they were also CVC non-words containing a balanced exposure of the target sounds), one possible explanation might be connected with voice quality and intelligibility of each talker. As previously explained (Method section 5.3.2), the eight most intelligible talkers were chosen to be part of the study and the four most intelligible ones were assigned as training talkers. This distribution of talkers may have played a role in the vowel results of generalization 2 (novel stimuli). Recall that even the CG, which was not exposed to these talkers during training, performed better when identifying the stimuli produced by the training talkers. This may indicate that L2 perception may be somewhat talker-specific, as certain voices may be easier to understand than others, due to specific voice quality characteristics. This issue will be further discussed in the next section (6.6.3), which deals with the generalization effects to absolute novel talkers.

Regarding the perception of untrained sounds, only the DIS_C generalized (strong generalization) the cross-training learning outcomes to novel non-word stimuli, confirming the efficacy of this training method and the strength of the cross-training effect. On the other hand, the DIS_V training group was not found to generalize the improvement on untrained segments (final consonants) to novel non-word stimuli. It is possible that the improvement previously observed (from pre-test to post-test) was stimuli-specific, not being consolidated enough to promote generalization to novel tokens (Logan & Pruitt, 1995). Generalization test 3 and the retention effects will shed some light on this specific case, as they also reflect on the robustness of the cross-training effect.

6.6.3. Generalization 3 – Absolute novel talkers

Generalization 3 investigated whether the improvement observed as a result of the training regimes received generalized to novel voices. Recall that the main findings of the study already confirm generalization to novel talkers (generalization 3a), since the testing tokens were produced by different talkers from the ones used in training. Thus, generalization to novel talkers has taken place for both training methods, supporting the effectiveness of both ID and categorical DIS as L2 training tools for vowel and initial consonant training. In addition, a further generalization test (generalization 3b) assessing if learning generalized to *absolute novel talkers*, that is, talkers that were not heard neither at the testing nor at the training phase, was carried out. Thus, familiar words produced by two absolute new talkers (T7 and T8) were tested in week 9. Results suggest that such generalization was talker-dependent, since generalization of vowel segments only took place for T8, and generalization of consonant segments only took place for T7. The cases in which generalization took place will be discussed first, followed by the cases in which generalization did not occur.

In the case of the vowel sounds, generalization to T8 occurred for the three experimental groups which had shown a significant improvement from pre-test to post-test (ID_V, DIS_V and DIS_C). However, while the ID_V group showed a *strong* generalization effect, the DIS_V and the DIS_C showed a *weak* generalization effect[25].This result may provide some further evidence that an ID method is more effective than a DIS method when training vowels sounds (Logan & Pruitt, 1995). Interestingly, the cross-training effect observed for the DIS_C was also observed in the stimuli produced by absolute novel talkers, confirming the effectiveness of the DIS training regime on the untrained sounds (vowels). This result might indicate that making use of vowel duration differences as a cue to identify consonant sounds might have promoted L2 vowel category learning (Logan & Pruitt, 1995).

[25] Recall that *weak generalization* = (post-test = generalization, but generalization = pre-test), and *strong generalization* = [(post-test = or < generalization) + (generalization > pre-test)].

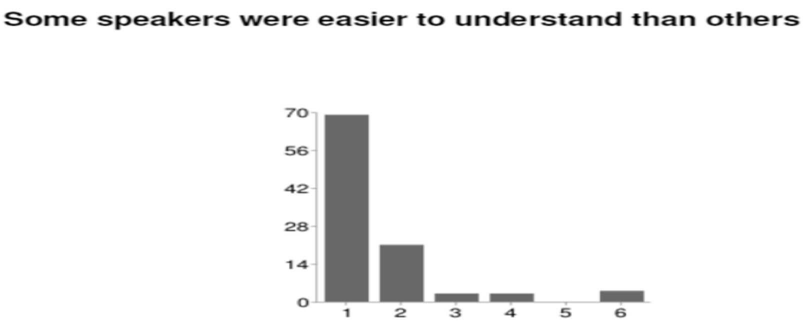
A possible explanation might lie in the fact that preceding vowel duration is a relevant and sufficient cue for final stop perception (Raphael, 1972) used by NES (Flege, 1989).

Similarly, when looking at the consonant stimuli, the three experimental groups that had shown improvement on consonants (ID_C, DIS_C and DIS_V) were able to generalize such improvement when identifying tokens produced by T7. For both consonant-trained groups there was strong generalization of learning, while weak generalization was found for the cross-training effect observed by the DIS_V. Thus, in line with the main results of this study, which already assess generalization to novel talkers (generalization 3a), both consonant trained groups were found to promote generalization to the same extent when training consonant sounds (Flege, 1995b). Moreover, the cross training effect on consonants generated by the vowel training regime (DIS_V) sheds some light on the strength of the cross-training effect here observed, since it generalized to an absolute novel talker.

Moving our attention to the cases in which generalization did not take place, it was observed that the three experimental groups that had improved on vowels from pre-test to post-test (ID_V, DIS_V, DIS_C) failed to generalize their vowel learning to tokens produced by T7. Similarly, when looking at the consonant results, the three experimental groups that had shown improvement on consonants (ID_C, DIS_C and DIS_V) were not successful with tokens produced by T8. As previously explained, specific properties of these talkers' speech impeded such transfer (T7 for vowel sounds and T8 for initial consonant sounds). According to Peterson and Barney (1952), "one of the major sources of variability in the speech signal comes from talker-specific attributes, where differences across talkers can result in vastly different acoustic signals". Moreover, Creel and Bregman (2011) add that these talker-specific acoustic features might affect the identification of speech sounds either positively or negatively.

In fact, evidence that some talkers may pose more of a difficulty to the L2 learners than others was found in the current study. Impressions about the training were collected from all trainees by means of a post-training survey (See methods section 5.4.3.3 for an explanation and Appendix H for the results of the survey). One of the questions included in the survey required students to express their level of agreement or disagreement regarding the following statement *"some speakers were easier to understand than others".* Figure 6.32 illustrates the students' answers. Recall that answers 1-3 represent different degrees of agreement with the statement and answers 4-6 represent different degrees of disagreement with the statement. Thus, it can be observed that the majority (93%) of the learners agreed that some talkers were easier to understand than others, as most answers corresponded to "1", that is, "I strongly agree".

Figure 6.32. Learners' opinions on the post-training survey regarding the talkers they were exposed to during the study



Having established that generalization to absolute novel talkers was talker-specific, it becomes relevant to explore if differences in talker intelligibility were related to specific properties of each talker's speech. When looking at the vowel results more closely, it is observed that two specific vowels drove the lack of success when identifying vowels produced by T7, namely /ɪ/ and /ɜː/. An acoustic analysis showed that the lax vowel sound /ɪ/ produced by T7 was produced with a relatively high F1 value, overlapping with the sound /e/ (Hillenbrand, Getty, Clark & Wheeler, 1995) and making this sound more confusable with /e/

to non-native ears. In fact, according to Feldman, Griffiths, Goldwater and Morgan (2013), phonetic categories usually overlap, and as a consequence, the phonetic boundaries are not necessarily well-defined and more challenging to L2 perception. As regards to the central vowel sound /ɜː/, the tokens produced by T7 were relatively shorter than tokens produced by other talkers of the study and it may have resembled the sound /ʌ/ to non-native ears. In fact, /ʌ/ was the most commonly selected answer by the participants (43% of the time).

Similarly, it was observed that results obtained for T8's initial voiced stops, particularly /d/ and /b/, obtained the lowest identification scores (28% and 25% respectively). An acoustic analysis of T8's production revealed an interesting pattern. Looking at the voiced alveolar stop first, the VOT values produced by T8 were notably higher than the VOT values produced by the other generalization talker (T7), testing talkers (T1 and T2) and training talkers (T3, T4, T5 and T6). By way of illustration, the short lag VOT values for each talker produced in the same non-word (deedge) are shown in Table 6.29. The short lag values for the tokens produced by T8 were notably high across /d/ tokens (26-32 ms), falling at the end point of the short lag range, since the short lag in English voiced stops usually range from zero to 30ms (Berry, 2004). Thus, it is possible that learners were more likely to misidentify T8's voiced stops due to the longer VOT.

Table 6.29. Short lag VOT values produced by all talkers for the non-word "deedge"

	Talker	VOT value - non-word "deedge"
Generalization talkers	**T8**	**32 ms**
	T7	18 ms
Testing talkers	T1	20 ms
	T2	12 ms
Training talkers	T3	13 ms
	T4	25 ms
	T5	18 ms
	T6	23 ms

Importantly, recall that in the learners' L1 inventory, voiceless stops are produced with a short lag of approximately 0-10 ms (Lisker & Abramson, 1964), and thus, the short lag voiced stops produced by the NES talkers were predicted to be perceived through the L1 frame (Flege, 1995; Best, 1995). While this prediction was more evident at the outset of the study, the fact that the L2 leaners improved their perception of initial voiced stops significantly from pre-test to post-test (produced by T1 and T2) indicates that they succeeded in learning this L1-L2 difference to some extent. However, the tokens produced by T8, which contained higher VOT values, might have distracted the non-native ears. In this case, the L2 learners perceived the voiced stops as the voiceless counterparts. Regarding the other two stops, /b/ was also produced with longer VOT by T8 than by T7, which may also account for the talker effect.

In addition, as previously reported, one of the testing talkers made use of a *voicing-lead* mode instead of a *short-lag* mode when producing the voiced stops, which made L2 perception easier to the Spanish/Catalan ears during testing (section 6.3.3). In English, both short lag and/or voicing lead might take place when producing voiced stops (Lisker & Abramson, 1964; Smith, 1975; Berry, 2004). According to Smith (1975), voicing lead takes place more commonly before high vowels and in certain places of articulation (namely, alveolar and labial). Thus, the fact that some talkers of the study used a voicing lead mode and some talkers used a short lag mode might have made voiced stop identification confusing to the L2 learners'. These findings might indicate that bimodal patterns of pronunciation that fall within the native speaker range might be more challenging to L2 learners during training, as the different pronunciation patterns might mislead learners' perceptual ability of the specific target-sound.

Moreover, it is relevant to recall that the stimuli of this study were validated by NES at the outset, and every single token was perceived as representative of the different target sounds of the target language variety. Thus, it can be concluded that the characteristics in T7 and T8's speech that posed problems to the Catalan learners did not pose problems to NES who validated the stimuli. This result provides evidence that generalization to absolute new talkers might be talker-specific. This is because certain characteristics present in some talkers' speech might diverge from the characteristics that the learners were exposed to during training, despite falling into the range of within-category variability tolerated by native listeners. One reason for the effect observed here could lie in the fact that this HVPT only exposed learners to four different talkers throughout the training regime, which might not have been a sufficient number of different talkers to provide learners with a wide enough range of natural variability of certain phonetic categories.

All in all, the results described in this section answer the last part of RQ.3, which deals with the generalization effects of the training methods under investigation. On the one hand, it was observed that generalization to novel talkers was promoted by both training methods (DIS and ID) for trained sounds and additionally for untrained sounds by the DIS method (as observed from pre-test to post-test). On the other hand, generalization to absolute novel talkers was found to be talker-specific. In line with the main results of the study, the ID trainees performed better with vowel sounds. The ID_V was the only method to promote a strong generalization effect for vowels, whereas the other two methods (DIS_C and DIS_V) promoted a weak generalization effect. Conversely, both ID_C and DIS_C promoted strong generalization to initial consonants, corroborating that both methods are equally effective for training initial consonants. Having discussed the results of generalization to novel talkers and absolute novel talkers, the results of the delayed post-test are reported next.

6.6.4. Delayed post-test

This phase of the study assessed whether the improvements observed as a result of training were maintained after the training regime was over. To that effect, two months after the post-test, participants performed the delayed post-test. Importantly, participant mortality (Mackey & Gass, 2005) took place in this latter phase of the study. That is, out of the total number of participants (89), only 63 completed this task.

As for the vowel training groups (ID_V and DIS_V), the direct effect of training (vowel learning) for both experimental groups and the indirect effect of training (final consonant learning) for the DIS_V were still observed two months after the completion of the training regime. Similarly, the consonant training groups also maintained the post-test scores at the delayed post-test phase, that is, the initial consonant scores for DIS_C and ID_C and the vowels scores for DIS_C. These results revealed that learning acquired through training was maintained over a period of two months. It confirms that training effects can be retained even after the training regime is over, in line with several previous studies (Bradlow et al., 1997,1999; Lively et al., 1994; Wang, 2002; Wang & Munro, 2004; Nishi & Kewley-Port, 2007). According to Flege (1995b), if knowledge acquired during training is retained over time, it might indicate that robust L2 categories have been established in the L2 learners' perceptual space. Moreover, it confirms the relevance of phonetic training as an L2 teaching tool.

All in all, the results of the delayed post-test answer RQ. 4, confirming that both training methods (ID and Categorical AX DIS) were able to promote long term effects and are effective when training vowel and/or initial consonant perception, in line with Flege's (1995b) findings on final consonant stimuli. Apart from promoting long term effects on the trained sounds, a categorial AX DIS method also promoted long term effects on the untrained sounds that had originally shown improvement with training (vowels for the DIS_C and final consonants for the DIS_V). This result confirms that the cross- training effect observed with the untrained but explicitly exposed sounds was consolidated. Moreover, since the effects were retained over time, it might be an indicator of L2 category formation (Flege, 1995b).

Part III: Beyond perception

6.7. Production results

Production was assessed at two times, at pre-test and at post-test. Production results were analysed by means of native speaker judgments. Munro (2008) points out the relevance of using listeners' ratings when analysing L2 speech: "From the standpoint of communication, there is no useful way to assess accentedness […] except through listener responses of some sort" (p. 200). In the present study, twelve native English speakers performed a series of rating tasks that included a subset of all the stimuli so that each stimulus was evaluated by four different native English raters (refer back to section 5.6.2.2 for further details). Recall that due to the large amount of stimuli involving the three types of segments (vowels, initial consonants and final consonants), only the initial voiceless stops and the final voiced stops were elicited (see section 5.6.2.2 for details). Seven identification tests with category goodness ratings were created for the vowel sounds and eight identification tests with category goodness ratings were created for the consonant sounds (four for the initial consonants and four for the final consonants) and the rating scale ranged from 1 (difficult to recognize as the selected sound) to 9 (easy to identify as the selected sound).

In order to assess if the raters agreed in their judgement for the vowels and consonants produced by the L2 learners at pre-test and post-test, a reliability analysis using an intra-class correlation coefficient (ICC) with a level of "absolute agreement" was conducted on the rating scores. According to Cronbach (1951) "a reliability coefficient demonstrates whether the test designer was correct in expecting a certain collection of items to yield interpretable statements about individual differences" (p. 297).

The results of the reliability analysis revealed a robust rater-agreement in all cases, as the Cronbach's alpha values were high. Cronbach's alpha reliability coefficient usually varies from 0 and 1, and, by convention, a Cronbach's Alpha from 0.40 to 0.59 indicates a *moderate* inter-rater reliability, from 0.60 to 0.79 indicates a *substantial* inter-rater reliability, and a value higher than 0.80 indicates an *outstanding* inter-rater reliability (Landis & Koch, 1977). The results, presented in Table 6.30, confirm that the four raters agreed on the analysis in all tests. In the case of *vowel test 5*, *vowel test 6* and *final consonant test 4*, the interrater reliability was *substantial* and in all other cases it was *outstanding.*

Table 6.30. Cronbach´s Alpha values for each vowel and consonant tests

Rating Test	Raters	Cronbach's Alpha (α)
Vowel test 1	4	α=.883
Vowel test 2	4	α=.831
Vowel test 3	4	α=.854
Vowel test 4	4	α=.818
Vowel test 5	4	α=.741
Vowel test 6	4	α=.788
Vowel test 7	4	α=.905
Initial consonant test 1	4	α=.888
Initial consonant test 2	4	α=.840
Initial consonant test 3	4	α=.859
Initial consonant test 4	4	α=.818
Final consonant test 1	4	α=.884
Final consonant test 2	4	α=.884
Final consonant test 3	4	α=.861
Final consonant test 4	4	α=.737

Having confirmed that the raters were in agreement, the median rating score for each token at pre-test and post-test was calculated (see method section 5.6.2.2 for more details on the calculations). As a result, the production gain scores were obtained by subtracting the pre-test scores from the post-test scores and were further submitted to statistical analysis. The production results (goodness ratings scores) on the trained and untrained sounds will be first shown for the vowel training groups followed by the production results on the trained and untrained sounds for the consonant training groups.

6.7.1. Vowel training groups

The median ratings that resulted from the 9-point likert scale on the trained and untrained sounds at pre-test and post-test for the vowel training groups are shown in Table 6.31, as well as the gain scores which were subsequently submitted to statistical analyses.

Table 6.31. Median rating (1-9) at pre-test, post-test and production gain scores for trained and untrained sounds in real words obtained by the vowel trained groups and the control group

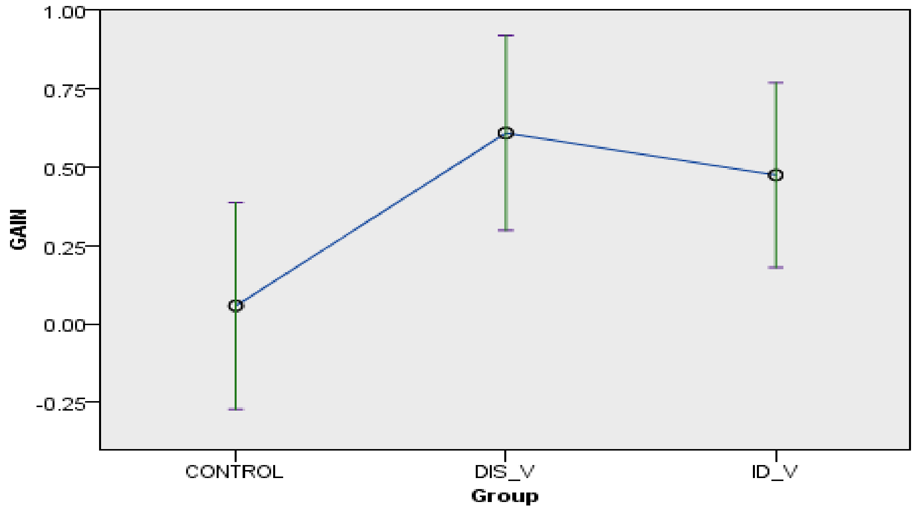
	CONTROL		DIS_V		ID_V	
Trained sounds Vowels	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	4.5	1.4	4.7	.78	4.6	1.5
POST	4.2	.89	5.1	1.1	5.2	1.1
GAIN	-0.3	.67	0.4	.89	0.6	.93
Untrained sounds Initial consonants	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	5.4	.97	5.9	1.1	5.8	.48
POST	5.6	.78	6.3	.44	6.4	.58
GAIN	0.2	3.8	0.4	1.2	0.6	.61
Untrained sounds Final consonants	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	3.0	.78	2.6	.56	3.2	.58
POST	3.4	.69	3.6	1.1	3.4	1.1
GAIN	0.4	1.3	1.0	1.2	0.2	1.2

As observed in Table 6.31, the scores attributed to the control group on their production of vowels resulted in a numerical decrease from pre-test to post-test. The vowel training groups, on the other hand, were given a higher rating score for vowel production after the training regime took place. More specifically, according to the NES raters, the DIS_V group's median scores improved by 0.4 and the ID_V improved by 0.6.

Turning our attention to the results for the untrained sounds, it is observed that in line with the perception results, the initial consonants were the segments in which the learners showed a better performance from the outset of the study, and the final consonants were the segments which learners produced less accurately in general at pre-test. Moreover, the three groups showed a numerical increase in the production of the untrained segments (initial and final consonants) from pre-test to post-test, ranging from (0.2 - 1.0). It is relevant to note that the highest improvement was obtained by DIS_V (1.0), which may be connected with DIS_V's lowest scores at pre-test, and that the three groups seem to be numerically comparable at post-test (3.6-3.4). The numerical improvement obtained by the controls (0.2 for initial consonants and 0.4 for final consonants) may reflect a task familiarity effect and an effect of the formal phonetic instruction received during the semester, which might have improved learners' awareness of the target sounds here investigated.

The gain scores for trained and untrained sounds were submitted to a GLMM, with group (ID_V, DIS_V, CG), segment (vowel, initial consonant, final consonant) and a group by segment interaction as fixed effects and participants as random effects. The model resulted in a significant effect of group ($F(2, 153) = 3.11$, $p < .05$), no effect of segment ($F(2, 153) = 1.41$, $p > .05$) and no group by segment interaction ($F(4, 153) = 1.66$, $p > .05$). The group effect is accounted for by the fact that the DIS_V marginally outperformed the control group in the overall segment perception, which was confirmed by a pairwise comparison with a Bonferroni correction (*p=.052*). However, the overall performance obtained by the two vowel experimental groups did not differ from one another and, the ID_V and the CG did not differ significantly. The overall group performance is illustrated in Figure 6.33.

Figure 6.33. Production gain scores obtained by vowel trained groups and control group



Even though segment did not reach significance, it is interesting to notice that the trained segments (vowels) were not the segment that obtained the highest numerical improvement from pre-test to post-test in production. This is partly due to the decrease obtained by the control group's performance on the vowel segments, and to the fact that the DIS_V showed more improvement, albeit non-significantly, on their production of consonant sounds (untrained segments) than on their production of vowel sounds (trained sounds). Figure 6.34 below shows the overall segment production gain scores from pre-test to post-test.

The performance obtained for each group on each individual segment can be seen in Figure 6.35. Despite the seemingly different results obtained for the ID_V group compared to the other groups, the lack of a significant interaction may be related to the relatively small differences in performance obtained on the different segments by the different groups.

Figure 6.34. Production gain scores for each segment type across vowel trained groups and control group

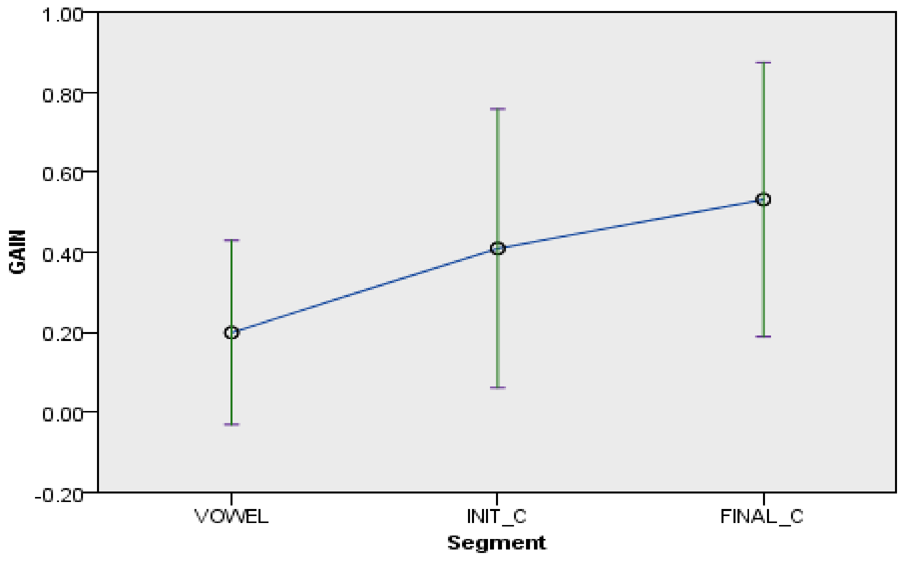
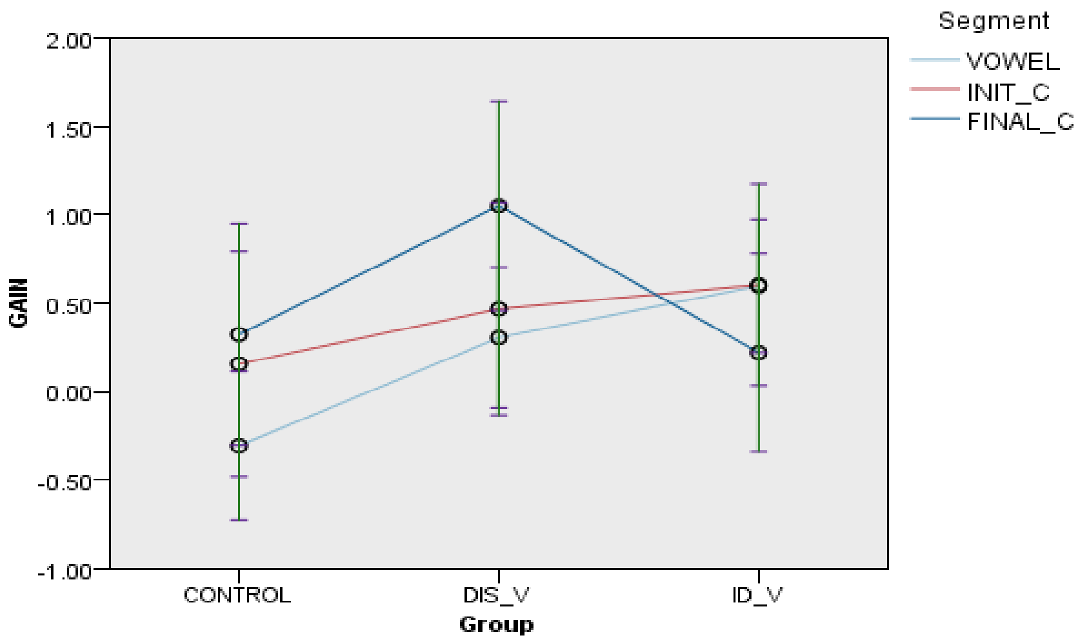


Figure 6.35. Group by segment production gain scores on trained and untrained sounds obtained by vowel trained groups and control group



6.7.1.1. Production gain scores on trained sounds (vowels)

When separating the effect and looking at the trained sounds only, the results yield a significant main effect of group ($F(2, 153) = 5.09$, $p < .01$), and pairwise comparisons with a sequential Bonferroni correction reveal that only the ID_V group significantly outperformed the controls ($p < .01$) in their improved production of vowel segments as perceived by native speakers of English. This result might indicate that the ID method was not only efficient in improving the learners perception of vowel sounds significantly, as it was also able to modify learners production as perceived by native English speakers. The individual vowel sound performance for trained segments (vowels) in real words obtained by the vowel trained groups and the controls can be seen in Table 6.32.

Table 6.32. Production gain scores for individual vowel sounds (trained segment) in real words obtained by the vowel trained groups and the control group

	CONTROL			DIS_V			ID_V		
	PRE	POST	GAIN	PRE	POST	GAIN	PRE	POST	GAIN
SOUND	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)
æ	4.7 (2.7)	4.1 (1.8)	-0.6 (2.5)	4.6 (1.8)	5.8 (2.8)	1.2 (2.7)	4.0 (2.5)	4.5 (2.1)	0.5 (2.6)
ɜː	6.1 (1.8)	5.9 (2.8)	-0.2 (2.7)	6.3 (2.5)	6.4 (2.1)	0.1 (2.6)	5.9 (2.8)	6.5 (1.6)	0.6 (2.7)
ɪ	4.8 (2.5)	4.4 (2.9)	-0.4 (2.4)	3.9 (2.5)	3.9 (2.9)	0.0 (2.4)	4.7 (2.7)	5.1 (2.2)	0.4 (2.8)
iː	5.1 (1.7)	4.9 (3.2)	-0.2 (2.4)	5.8 (2.1)	6.1 (2.6)	0.3 (2.2)	6.3 (1.7)	6.6 (3.2)	0.3 2.4
ʌ	2.0 (2.2)	1.8 (2.8)	-0.1 (1.6)	3.3 (2.6)	3.3 (2.2)	0 (.54)	2.3 (1.7)	3.4 (3.2)	1.1 (2.4)

Note. M= Median

Similarly to the results observed for perception, some vowel sounds seemed to obtain better results than others and improved to different degrees. The vowel that obtained the lowest production ratings from the outset was the vowel /ʌ/, followed by the vowels /ɪ/ and /æ/. The two highest rated vowel sounds were /iː/ and /ɜː/. These results differ from the previously reported perception results mostly regarding two sounds: /ɜː/, which was comparatively less successfully identified than other vowels, and /ɪ/, which was better perceived than other vowels. Both /ʌ/ and /æ/ seem to pose difficulties to learners both in perception and production and the tense sound /iː/ was the least challenging, particularly in production. While the control group showed no improvement in vowel sounds, the two experimental groups seemed to improve in at least some of the target sounds, in line with the better performance of the ID and DIS groups observed in the analyses of segment types. The DIS_V's improvement ranged from 0-1.2, and the ID_V's improvement ranged from 0.3-1.1.

6.7.1.2. Production gain scores on untrained sounds (initial and final consonants)

The simple effects tests conducted on the production gain scores obtained for the untrained segments yielded no group effect (initial consonants: $F(2, 153) = 0.55$, $p > .05$; final consonants: ($F(2, 153) = 2.33$, $p > .05$). Tables 6.33 and 6.34 present the individual consonant sound production performance at pre-test, post-test and the corresponding gain scores obtained by the vowel training groups (DIS_V, ID_V) and the control group.

Looking at the results for each stop (recall that only voiceless stops were elicited), the bilabial stop obtained somewhat lower scores than the other two stops, particularly for the control group. The controls in fact showed minimal improvement from pre-test to post-test, except for /p/. The two trained groups showed a similar amount of (numerical) improvement across sounds (0.6-0.8) with the exception of /k/ in the case of the DIS_V.

Table 6.33. Production gain scores for individual initial consonant sounds (untrained segment) in real words obtained by the vowel trained groups and the control group

	CONTROL			DIS_V			ID_V		
	PRE	**POST**	**GAIN**	**PRE**	**POST**	**GAIN**	**PRE**	**POST**	**GAIN**
SOUND	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M (SD)	M(SD)	M(SD)
p_	4.6 (2.7)	5.1 (2.5)	0.5 (1.6)	5.7 (2.9)	6.2 (1.4)	0.5 (2.6)	5.4 (2.2)	5.9 (2.1)	0.5 (1.4)
t_	5.9 (2.3)	6.0 (2.1)	0.1 (0.6)	5.7 (2.8)	6.5 (2.0)	0.8 (2.5)	6.0 (2.3)	6.6 (2.4)	0.6 (2.1)
k_	5.8 (2.8)	5.7 (2.3)	-0.1 (2.9)	6.1 (2.2)	6.2 (2.1)	0.1 (1.0)	5.9 (1.8)	6.5 (2.2)	0.6 (2.8)

Note. M= Median

The production scores for the individual voiced final stops are shown in Table 6.34. Learners' productions of final consonant sounds were rated considerably low at the outset of the study in comparison with vowels and initial consonants. The bilabial was the sound that obtained the poorest rating score across all groups (1.7-1.8), followed by the alveolar (2.7-3.3) and the voiced velar stop was the best rated of the three sounds (4.1-5.0).

Table 6.34. Gain scores for individual final consonant sounds (untrained segment) in non-words obtained by the vowel trained groups and control group

	CONTROL			DIS_V			ID_V		
	PRE	**POST**	**GAIN**	**PRE**	**POST**	**GAIN**	**PRE**	**POST**	**GAIN**
SOUND	M (SD)	M (SD)	M (SD)	M (SD)	M(SD)	M (SD)	M (SD)	M (SD)	M (SD)
_b	1.8 (2.7)	2.4 (2.2)	0.6 (2.8)	1.7 (1.7)	2.3 (2.9)	0.6 (2.8)	1.7 (2.9)	1.6 (2.1)	-0.1 (2.8)
_d	3.3 (2.9)	3.0 (2.7)	-0.3 (1.9)	3.3 (1.5)	3.0 (2.5)	-0.3 (2.9)	2.7 (2.9)	3.8 (2.2)	1.1 (2.8)
_g	4.1 (2.5)	5.0 (3.0)	0.8 (3.6)	4.1 (2.3)	4.9 (2.1)	0.8 (2.8)	5.0 (2.7)	4.6 (2.8)	-0.4 (2.4)

Note. M= Median

An explanation for why different segments patterned somewhat differently could be related to specific characteristics of the sounds or to the nature of the words in which they were elicited (word familiarity, presence of minimal pairs). In any case, analysing the effect of training on individual segments lies beyond the scope of this study. Still, the general pattern observed in the production of initial and final stops goes in line with the previously reported perceptual results. That is, that participants performed better with initial voiceless stops, but were less successful with final voiced stops.

6.7.2. Consonant training groups

Similarly to the data presented for the vowels, the median ratings obtained for the consonant training group and the corresponding gain scores on the trained and untrained sounds at pre-test and post-test for the consonant training groups are shown in Table 6.35.

Table 6.35. Median rating (1-9) at pre-test, post-test and production gain scores for trained and untrained sounds in real words obtained by the consonant trained groups and the control group

	CONTROL		DIS_C		ID_C	
Trained sounds Initial consonants	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	5.4	.69	6.1	.92	6.0	.86
POST	5.6	.81	7.0	.76	6.8	.67
GAIN	0.2	1.5	0.9	.75	0.8	1.3
Trained sounds Final consonants	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	3.0	.43	2.9	.42	2.9	.56
POST	3.4	.78	3.2	.69	3.6	.67
GAIN	0.4	1.4	0.3	1.4	0.7	1.1
Untrained sounds Vowels	**Median**	**SD**	**Median**	**SD**	**Median**	**SD**
PRE	4.5	.69	5.2	.72	4.3	1.1
POST	4.2	.72	5.4	.68	4.6	.68
GAIN	-0.3	.67	0.2	.61	0.3	1.0

As shown in Table 6.35, the largest gain among all groups was experienced by the DIS_C and ID_C in their production of initial consonants (0.9 and 0.8 gain, respectively). The amount of gain obtained by the control group in initial consonant production, which may reflect a task effect and the enhanced metalinguistic knowledge obtained from the phonetics course they were enrolled in, was 0.2. The production of final consonants obtained the lowest rating scores at pre-test and the improvement observed as a result of the consonant training received was 0.7 for the ID_C and 0.3 for the DIS_V. The control group also improved in final consonant production as perceived by native English speakers (0.4). The production of vowel sounds (untrained segments) by the consonant training groups showed little improvement, 0.2 for DIS_C and 0.3 for ID_C, and no improvement for the controls. The relevance of the difference between pre-test and post-test scores for the different groups is analysed next.

The gain scores obtained for the three segments were entered in a GLMM, in which the fixed effects were group (ID_C, DIS_C, CG) and segments (vowel, initial consonant and final consonant) and participants were treated as random effect. The model resulted in a marginally significant effect of group ($F(2, 144) = 2.99$, $p = .054$), a significant effect of segment ($F(2, 144) = 4.14$, $p < .05$) and no group by segment interaction ($F(4, 144) = 0.56$, $p > .05$). The marginally significant effect of group is explained by the fact that the ID_C group marginally outperformed the control group ($p = .057$, in the Bonferroni pairwise comparison) in the overall production (Figure 6.36). Regarding the significant effect of segment, pairwise comparisons with a Bonferroni correction indicate that the initial consonant improvement was significantly greater than the improvement obtained for the vowel segments ($p < .05$), which can be observed in Figure 6.37.

Figure 6.36. Production gain scores obtained by consonant trained groups and control group

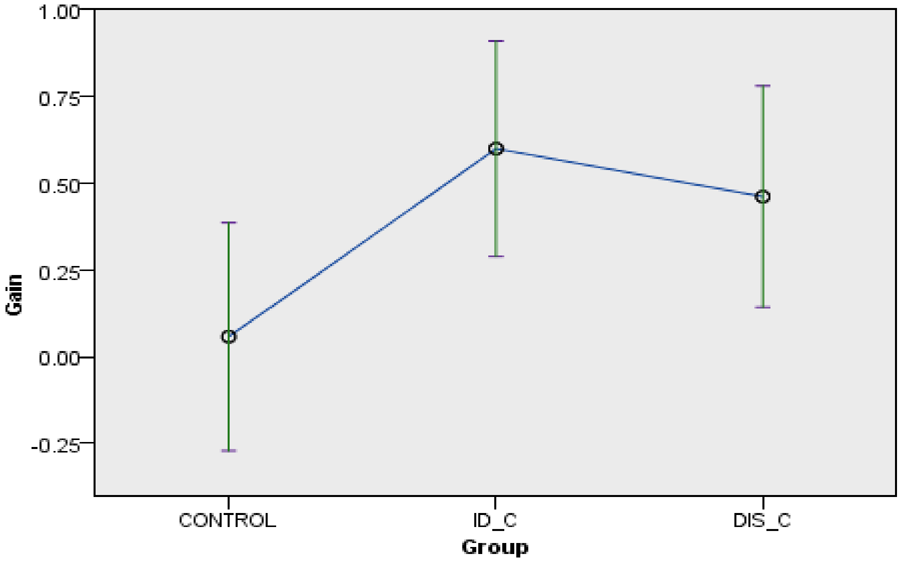


Figure 6.37. Production gain scores for each segment type across consonant trained groups and control group

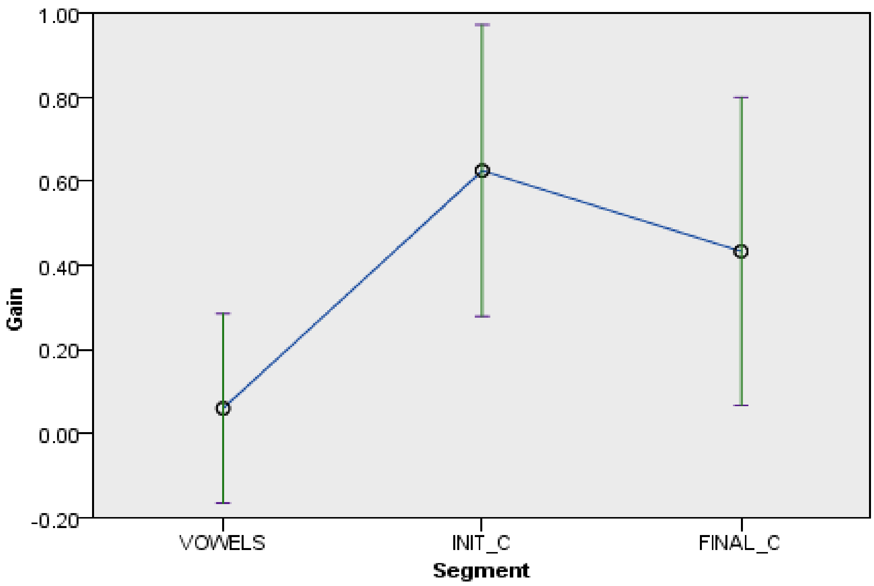
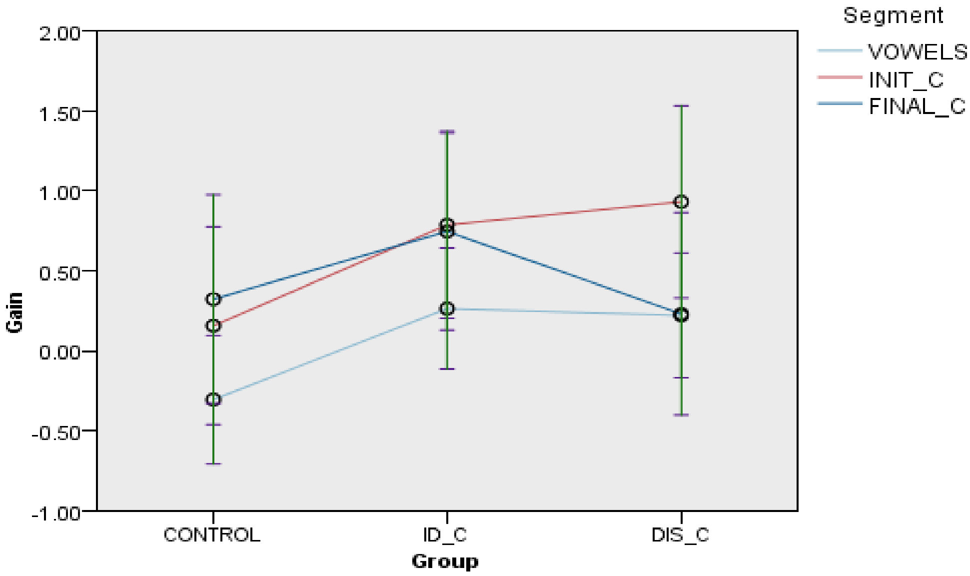


Figure 6.38 presents the production gain scores on trained and untrained sounds obtained by consonant trained groups (ID_C, DIS_C) and the control group (CG). The lack of

a segment by group interaction shows that the differences among segment types for all three groups were comparable.

Figure 6.38. Group by segment production gain scores on trained and untrained sounds obtained by consonant trained groups and control group



6.7.2.1. Production gain scores on trained sounds (initial and final consonants)

Separating the effect and focusing on the results for the trained sounds only, no effect of group was found for initial consonants ($F(2, 144) = 1.63$, $p > .05$), nor for final consonant segments ($F(2, 144) = 0.73$, $p > .05$), as expected from the lack of an interaction in the general analysis described above. The results for each individual sound for the two consonant trained groups (ID_C and DIS_C) will be presented next. Tables 6.36 and 6.37 show the rating scores obtained by consonant trainees and the controls for the three initial stops at pre-test, post-test and the corresponding gain scores.

Table 6.36. Production gain scores for individual initial consonant sounds (untrained segment) in real words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE M (SD)	POST M (SD)	GAIN M (SD)	PRE M (SD)	POST M (SD)	GAIN M (SD)	PRE M (SD)	POST M(SD)	GAIN M(SD)
p_	4.6 (2.7)	5.1 (2.5)	0.5 (1.6)	5.7 (2.2)	6.9 (2.0)	1.2 (1.9)	6.1 (1.8)	6.6 (2.2)	0.4 (1.9)
t_	5.9 (2.3)	6.0 (2.1)	0.1 (0.6)	6.5 (2.2)	7.0 (2.1)	0.5 (2.0)	6.0 (1.8)	6.8 (2.2)	0.8 (1.9)
k_	5.8 (2.8)	5.7 (2.3)	-0.1 (2.9)	6.3 (2.2)	7.3 (2.1)	1.0 (2.0)	5.9 (2.2)	7.0 (2.1)	1.1 (2.0)

Note. M= Median

Looking at the individual voiceless stops, it can be seen that overall both experimental groups obtained better ratings at post-test than an pre-test. The experimental groups' performance at pre-test ranged from 5.7 to 6.5, and it was enhanced to 6.6-7.3 after training, showing an improvement of 0.4 to 1.2, varying according to target sound and group. The greatest improvement was obtained by the DIS trainees on /p/ and /k/ and by ID_C for /k/, reaching relatively high rating scores at post-test (7 for DIS_C and 7.3 for ID_C) considering that the rating scale ranged from 1-9. Moreover, both experimental groups showed a positive numerical improvement on the alveolar stop /t/ from pre-test to post-test (0.5 for DIS_C and 0.8 for ID_C). Interestingly, the previously reported consonant results obtained by the vowel training groups patterned similarly to the results obtained by the consonant training groups. That is, /p/ also tended to obtain lower results than the other stops, which might probably be connected to the generally shorter VOT of bilabial stops (Yavaş, 2007), making aspiration a less salient cue than in the alveolar and velar stops. Moreover, despite the non-significant result of training on consonants, it is interesting to note that the overall improvement of the consonant trained groups with initial consonants is greater than the improvement observed by the vowel trained group, showing a tendency for perceptual training to affect learners' production of initial stops in the right direction.

Table 6.37. Gain scores for individual final consonant sounds (trained segment) in non-words obtained by the consonant trained groups and control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE M (SD)	POST M (SD)	GAIN M (SD)	PRE M (SD)	POST M(SD)	GAIN M (SD)	PRE M (SD)	POST M (SD)	GAIN M (SD)
_b	1.8 (2.7)	2.4 (2.2)	0.6 (2.8)	1.7 (1.7)	1.9 (2.6)	0.2 (2.3)	1.6 (2.3)	2.5 (2.4)	0.8 (2.2)
_d	3.3 (2.9)	3.0 (2.7)	-0.3 (1.9)	3.4 (2.3)	3.2 (2.6)	-0.1 (2.4)	2.6 (2.1)	3.6 (1.9)	1.0 (1.7)
_g	4.1 (2.5)	5.0 (3.0)	0.8 (3.6)	3.6 (2.3)	4.4 (2.6)	0.8 (2.3)	4.2 (2.2)	4.9 (2.0)	0.7 (1.9)

Note. M= Median

As seen in Table 6.37, the voiced bilabial was the sound which obtained the poorest rating score across all groups at both testing phases (1.6-1.8 at pre-test; 1.9-2.5 at post-test), followed by the voiced alveolar stop (2.6-3.4 at pre-test and 3-3.6 at post-test) and the final voiced velar was the best rated of the three sounds at both phases (3.6- 4.2 at pre-test and 4.4-5.0 at post-test). When looking at the gain scores, the largest improvement was obtained by the ID_C group for the alveolar sound (1.0), which is partially connected to the low score obtained at the outset. The sound that improved the most across the three groups was the final voiced velar stop (0.7-0.8), followed by the improvement obtained for the bilabial stop (0.2-0.8).

In line with the results found for perception, final voiced stops obtained lower scores than initial stops, showing that the former pose a greater difficulty to Catalan learners in both perception and production. Interestingly, the DIS_V group is the experimental group that obtained the greatest gain when producing final consonant stimuli, even greater than the consonant training groups. These results will be further developed in the discussion session.

6.7.2.2. Production gain scores on untrained sounds (vowels)

A simple effects test on the untrained segments revealed no effect of group ($F(2, 144) = 3.23$, $p > .05$) for vowels. The individual vowel sound performance for the vowel segments (untrained segments) in real words obtained by the consonant trained groups and the controls can be seen in Table 6.38.

Table 6.38. Production gain scores for individual vowel sounds (untrained segment) in real words obtained by the consonant trained groups and the control group

	CONTROL			DIS_C			ID_C		
SOUND	PRE M (SD)	POST M (SD)	GAIN M (SD)	PRE M (SD)	POST M (SD)	GAIN M (SD)	PRE M (SD)	POST M (SD)	GAIN M (SD)
æ	4.7 (2.7)	4.1 (1.8)	-0.6 (2.5)	5.0 (2.9)	4.6 (2.6)	-0.4 (2.1)	3.5 (2.8)	4.3 (2.6)	0.8 (2.0)
ɜ:	6.1 (1.8)	5.9 (2.8)	-0.2 (2.7)	5.8 (2.7)	6.6 (2.6)	0.9 (2.0)	5.9 (2.5)	6.1 (2.4)	0.2 (1.7)
ɪ	4.8 (2.5)	4.4 (2.9)	-0.4 (2.4)	6.4 (2.5)	6.1 (2.2)	-0.3 (1.8)	4.5 (2.5)	4.1 (2.4)	-0.5 (1.7)
i:	5.1 (1.7)	4.9 (3.2)	-0.2 (2.4)	5.7 (2.6)	6.2 (2.2)	0.5 (1.9)	6.0 (2.5)	6.7 (2.5)	0.7 (1.9)
ʌ	2.0 (2.2)	1.8 (2.8)	-0.1 (1.6)	3.4 (2.4)	3.8 (2.4)	0.4 (1.9)	1.8 (2.4)	1.9 (2.3)	0.1 (1.9)

Note. M= Median

As observed in Table 6.38, at pre-test, the vowel that obtained the lowest production ratings was the vowel /ʌ/, followed by the vowels /æ/ and /ɪ/ (with exception to the DIS_C group's score). The two highest rated vowel sounds were /iː/ and /ɜː/, in line with the results previously reported for the vowel trainees and SLM predictions[26]. Turning our attention to the gains scores, it can be seen that the largest improvement (0.9) was observed for the DIS_C group on the production of the central vowel sound /ɜː/, followed by the improvement obtained by the ID_C with /æ/ (0.8) and /iː/ (0.7). The controls showed no improvement. Thus, despite differences among groups and vowel sounds, the trained groups (unlike the controls) showed a general tendency to improve at least with some vowels, even if the differences between groups did not reach significance.

6.7.3. Relationship between perception and production

As this study assessed both perception and production of L2 consonants and vowels, it becomes relevant to investigate if there was a relation between the improvement that leaners obtained in perception and the learning that learners obtained in production. A positive correlation between the gains obtained in perception and production would imply that the same participants who improved the most in one modality also improved the most in the other modality, showing that both skills are correlated and progress hand in hand.

As the production task assessed the learners' ability to produce the target sounds in real words, the production results were contrasted with the results from perception of target sounds embedded in real words only. Moreover, as the production results only looked at initial voiceless consonants and final voiced consonants, only these specific sounds were considered in the correlation analyses. Thus, the perceptual gain scores obtained for each segment (vowel, initial voiceless stops and final voiced stops) were contrasted with the production gain rating scores for the exact same segments. The scores were submitted to Pearson (product-moment) correlation tests for each individual group, resulting in three

[26] According to the SLM predictions, the near identical vowel sound /iː/ and the new vowel sound /ɜː/ should be the best produced L2 vowel sounds by the Catalan learners.

Pearson correlation tests per group (5 groups * 3 segments = 15 tests). The results for the 15 different tests can be seen in Table 6.39.[27]

Table 6.39. Perception and production Pearson (product-moment) correlation scores per group

Group	segment	Pearson's r	Significance
ID_V	Vowels	.416	(p= .068)
	Initial C	.404	(p=.077)
	Final C	.346	(p=.135)
DIS_V	Vowels	.011	(p=.965)
	Initial C	.116	(p=.647)
	Final C	.265	(p=.288)
ID_C	Vowels	.204	(p=.416)
	Initial C	.012	(p=963)
	Final C	-.312	(p=207)
DIS_C	Vowels	.171	(p=.512)
	Initial C	-.028	(p=.916)
	Final C	.284	(p=.269)
CG	Vowels	.054	(p=.842)
	Initial C	.198	(p=.463)
	Final C	-0.30	(p=.912)

Pearson correlation test results yielded no significant results. This indicates that the learners' perception improvement was unrelated to their production scores. This result is in line with previous training studies that found no correlation between perception and production (Bradlow et al., 1997; Pereira, 2014; Iverson et al., 2012), and contradict other previous studies that did find a correlation (Rochet, 1995; Wong, 2012).

[27] Correlation analyses were conducted for all groups and segments, eventhough only the ID_V improved in both domains (perception and production) and thus, a significant result was only expected for this specific experimental group on the performance of trained segments. Further correlation analyses were conducted on percentage and median scores from pre-test data and post-test data for the ID_V. There was no correlation in both cases, having pre-test (r= 0.89, n= 20, p=.710) and post test (r=.184, n=20, p =.436).

6.8. Discussion III - Production

One of the issues that this study aimed to explore was the effect of perceptual training on the production of L2 sounds (RQ.5) and whether both training methods (ID and Categorical DIS) applied in this study promoted production improvement to the same extent (Sub-RQ.5.1 and Sub-RQ.5.2). The experimental groups and the controls were rated on their L2 English production before and after the training phase by 12 native speakers of Southern British English (NES), who identified and rated the production of target sounds (vowels and consonants) embedded in real word stimuli. Even though training strictly made use of non-word stimuli, production abilities were tested by means of real word stimuli due to practical and word familiarity reasons. The results found for the vowel training groups will be discussed first, followed by the discussion of the consonant training groups' results. Then, the relationship between perception and production will be discussed, and related to previous findings in the literature.

6.8.1. Vowel training groups

The vowel experimental groups and the controls were contrasted on their overall production abilities of target sounds embedded in real words. Results revealed that while the DIS_V group marginally outperformed the controls on the overall production (i.e. both trained and untrained segments combined), only the ID_V outperformed the CG on the production of the trained segments (vowels). The fact that the DIS_V group promoted an overall improvement in production goes in line with the perceptual results found in this study, which have shown that the DIS_V method promoted significant gains on trained/attended segments as well as on the untrained/unattended segments. This cross-training effect and overall improvement observed by the DIS_V group might be explained by the same possible reasons previously mentioned when discussing the perception results. These were: (i) the presentation of two physically present stimuli in each trial in the discrimination task, (ii) the fact that the discrimination task does not involve the use of labels to identify the stimuli and (iii) the fact that vowels are affected by neighbouring consonants and vice-versa. The DIS_V trainees were thus overall better understood and perceived by NES after a perceptual training regime. This finding goes in the same direction as some previous findings that a perceptual training regime can have a positive impact on learners' production abilities (Yamada et al. 1996; Flege, 1989; Bradlow et al., 1997; Hardison, 2003; Rochet et al., 2005; Hazan et al., 2005; Lambacher et al., 2005; Lengeris, 2008; Iverson et al., 2012; Thomson, 2011; Pereira, 2014; Rato, 2014).

However, when focusing on the production performance of the trained sounds only (vowels), the ID_V group was the only group that outperformed the CG. These results add to previous evidence that the ID method is a more efficient method than the DIS methods to train vowel segments (Nozawa, 2015) and corroborate previous findings in the L2 vowel training literature that found transfer from perceptual training methods to production of vowel sounds without any explicit production training (Lengeris, 2008; Iverson et al., 2012; Thomson, 2011; Pereira, 2014; Rato, 2014). Another important factor that may explain why only the ID_V promoted production gains for trained segments might be connected to the use of real word stimuli. Recall that the ID_V was the only vowel training group to promote perceptual gains of L2 sounds embedded in real words. Hence, the production results are consistent with the perceptual results for real word stimuli. In line with what was argued for real word perception, the presence of labels in the ID_V training might have been crucial when facilitating L2 vowel learning in real word stimuli. This is connected to the fact that focus on form enhances L2 perception (Saito, 2015), and perceptual gains lead to consequent production gains (Flege, 1995).

Moreover, previous studies have shown that previously stored lexical representations associated with orthography might influence perception either positively or negatively (Pierce, 2014). In training studies, the influence is said to be positive, since the trial by trial feedback combined with orthography provides learners with tools to adjust their previously stored L2 phonetic categories towards the native standard (Iverson et al., 2012). In the case of the training carried out in the present study, it might be the case that the orthography present in the labels might have promoted lexical associations during training. Thus, the ID trainees may have been able to contrast the information of the native input with the information of their previously stored lexical categories. Thus, the ID_V training regime might have provided learners with the opportunity to correct inaccurate previously stored associations connected to spelling.

In addition, recall that the production task (indirectly) included spelling of each word. As explained in the method section (5.4.2), a picture naming task was used. However, the orthographic representation of each word was briefly shown alongside the picture, and disappeared before the speaker was prompted to produce the word. This procedure was incorporated in order to guarantee that each word was elicited correctly. However, the inclusion of the orthography in the picture naming task might have conditioned the outcomes of the present study. It would make sense that only the group that was trained with the presence of orthography and the consequent focus on form that it entails improved in this task. Notice, however, that none of the words used in the labels of the ID tasks were the same as the words used in production (see Table 5.14 for the production list and Figure 5.5 for the vowel ID task response options). Nevertheless and importantly, the gains observed in production were considerably low compared to the perceptual gains. Thus, generalization to production was not found to be consistent with the perceptual gains of the present investigation. This might be due to two reasons: the nature of the training regime itself, since it involved only perceptual training, or the fact that learners' production skills might take longer to be developed than learners' perceptual skills.

6.8.2. Consonant training groups

Turning our attention to the consonant trained groups, their overall production revealed a marginally significant improvement by the ID_C group. While this marginally significant effect is connected to the improvement on the two consonant segment types (initial consonant and final consonant) combined, it is also connected to a slight numerical improvement on the untrained segments (vowels). Since no cross-training effect was observed for this group in perception, it is not likely that the slender numerical improvement on vowels is linked to a cross-training effect promoted by this training method. Instead, it might be connected to (i) the lower scores obtained at pre-test for vowel identification, and (ii) the fact that the better production of initial consonants at post-test obtained by the ID_C group might have affected the perception of this group's vowel sounds by the NES judges. It is known that the perception of vowels might be affected by the neighbouring consonants (Strange, Edman & Jenkins, 1979). In fact, most NES raters mentioned that it was a difficult task to judge sounds individually, as they felt like they were being influenced by the pronunciation of the word as whole. Thus, the more native-like production of the voiceless stops, possibly due to the longer VOT values, might have positively influenced the perception of L2 vowel sounds.

Regarding the performance on the trained segments only, the results revealed that neither of the two perceptual training methods (ID and DIS) were efficient in improving learners' production skills significantly. This lack of significance found with consonant segments is most probably connected to the short duration of the training regimes conducted as part of this study. Recall that the total duration of each session (30-40 minutes) was divided into two equal parts in order to train initial and final consonant stimuli and, thus, learners were exposed to each segment (initial vs. final consonant) only for 15-20 minutes for five sessions. Thus, in the case of initial consonants, the short duration of the training was extensive enough to promote perceptual improvements, but not to promote significant gains in production.

Although the consonant production results found in the present investigation did not reach significance, the numerical improvement indicate a positive tendency in the right direction, that is, learners appeared to be better perceived by NES after training. This tendency supports the efficacy of training methods on VOT perception (Carney et al., 1977; Pisoni et al., 1982; Aliaga-García & Mora, 2009; Collet et al., 2013; Everitt, 2015; Alves & Luchini, 2016) and suggests that a longer period of exposure to specialized input might have been able to promote larger and possibly significant production gains, in line with previous studies (Flege, 1989; 1995b).

6.8.3. Production of individual sounds

Recall that this thesis considers vowels and consonants as a global measure and thus a detailed analysis of each individual sound lies beyond the scope of the present investigation. Nevertheless, a brief discussion of the general patterns observed for individual sounds in production will be developed next for the sake of completion. Thus, the results for individual sounds in production will be compared to the patterns found in perception and will be discussed in light of the SLM's predictions. The performance of individual vowel sounds will be discussed first, followed by the performance of individual consonant sounds.

The overall gain for vowel production assigned by native English listeners for the experimental groups was 0.4 for the DIS_V, 0.6 for the ID_V, 0.2 for the DIS_C and 0.3 for the ID_C. Looking at the individual vowel performance, it is interesting to note that the most challenging sounds for the L2 learners were /ʌ/, /ɪ/ and /æ/. The difficulty in producing the vowel pair /ʌ/-/æ/ is connected to what has been previously explained regarding the perceptual level. That is, the fact that this vowel pair would be perceived in the frame of a "single-category assimilation" pattern and consequently its predicted discriminability would be poor (Best, 1995). It was hypothesized that this discriminability prediction would be transferred to the identification of these specific sounds, as learners would not be able to distinguish one from the other accurately (Rallo-Fabra & Romero, 2012). Moreover, as perception difficulties may hinder production (Flege, 1995), production of this vowel pair was expected to pose difficulties to Catalan learners of English (Aliaga-García & Mora, 2009; Fullana & MacKay, 2003). Apart from the high results obtained by the DIS_C, the results found in the production of the lax high front vowel /ɪ/ for the remaining experimental groups patterned differently from its perceptual performance. While it had not been a problematic sound for Spanish/Catalan learners at the perceptual level, it posed a challenge in production to most trainees. The reason for this discrepancy in the perception and production of this specific sound might be connected to the fact that perception may precede production (Flege, 1995) and learners may not yet have attained production accuracy for the sound /ɪ/. However, the positive perceptual results may indicate that learners' production of this sound will eventually be accurately developed as a consequence of the enhanced perception (Flege, 1995).

Conversely, the two best rated vowel sounds were the sounds /iː/ and /ɜː/. The tense high front vowel /iː/ was the least problematic sound both in perception and production due to the presence of a highly similar L1 Catalan sound /iː/ (Cebrian, 2006, 2015; Cebrian et al., 2011; Rallo-Fabra & Romero, 2012), and the fact that identical or "near identical" sounds are easily perceived and consequently produced in a L2 (Flege, 1995). On the other hand, the performance of the central sound /ɜː/ patterned differently in perception and production. While this sound was poorly perceived when embedded in non-words at the outset of the study, learners' production of it obtained comparatively high ratings by NES from the outset of the study. The relative greater success in the production of this sound can be accounted for by the fact that this sound has no counterpart in the L1 vowel inventory (Cebrian et al., 2011; Cebrian, 2015). This is in accordance with Flege's SLM, since the more dissimilar L1-L2 sounds are more likely to be acquired more accurately given enough L2 experience (Flege, 1995). It is possible that different sounds pattern differently regarding perception and production, as discussed in the next section.

Turning our attention to the consonant sounds, despite the comparatively higher rating scores obtained with initial voiceless stops at the outset of the study, numerical gains (albeit non-significant) were observed. Both consonant training groups were better rated after training than before training on all initial stops. The largest gain was observed for the velar stop (1.0 for the DIS_C and 1.1 for the ID_C), which is probably connected to the fact that the velar stops tend to have the longest lag VOT of all stops (Lisker & Abramson, 1967; Docherty, 1992; Cho & Ladefoged, 1999), and thus, the salience of the aspiration as a cue becomes more marked (Burnham, 1986). For the bilabial sound /p/, the DIS_C showed a large improvement (1.2), which is most probably connected to the lower scores at pre-test. As regards the production of the alveolar stop /t/, the ID_V rating scores improved by 0.8. Interestingly, in line with the perception results, the voiceless stops were comparatively highly rated (especially at post-test). This may indicate that aspirated voiceless stops can be accurately acquired by L2 learners whose languages do not present aspiration despite the original difficulty that Romance language learners often experience with acquiring this feature (Alves & Zimmer, 2015).

Regarding the final voiced stops, recall that the two consonant training regimes applied in this study may not have been sufficient to modify learners' perception. In fact, the only experimental group that was found to improve significantly on the perception of final consonant sounds was the DIS_V through a cross training effect. Thus, no transfer to production should be expected from the consonant trainees. In line with the results for perception, the rating scores obtained at both phases of the study were considerably low. As previously discussed, these sounds pose a lot of difficulty to Catalan learners of English in perception and production, due to the similar, but not identical realization in the L1 and the L2 (Flege, 1995). While in Catalan, the voicing contrast in final obstruents is neutralized and obstruents are always voiceless in prepausal position (Recasens, 1993; Cebrian, 2000), in English, final obstruents may be phonetically devoiced and the voicing contrast is mostly cued by the preceding vowel duration (Raphael, 1972; Roach, 2000). Thus, the low scores obtained at post-test by the consonant trainees indicate that the perceptual training received was not effective in modifying learners' production abilities, in line with the results for perception.

6.8.4. Relationship between perception and production

Perception and production improved in different ways in the present study, since perception and production gains were not found to correlate. This comes as no surprise considering that two out of five vowel sounds patterned differently in perception and production, namely the sounds /ɪ/ and /ɜː/[28]. This fact provides further evidence that these two

[28] The remaining three sounds patterned similarly in perception and production and in line with initial predictions of the study.

abilities may develop separately (Bradlow et al., 1997; Iverson et al., 2012). Moreover, apart from the learners' performance with the sound /ɜː/, which was better produced in real words than perceived when embedded in non-words, the performance with all other target sounds point to the fact that perceptual learning may take place prior to production gains (Flege, 1995). The case of /ɜː/ seems to provide evidence to the opposite fact, that is, that production might develop before perception for specific sounds (Sheldon & Strange, 1982; Llisterri, 1995; Kluge et al. 2007; Hattori & Iverson, 2010)[29]. It is possible that different sounds are acquired differently, some being perceived accurately before the learners are able to acquire its accurate articulatory production and vice-versa. In fact, as previously mentioned, the literature provides evidence for three different possible outcomes regarding the perception and production interface. That is: (i) that perception precedes production (Flege, 1984; Flege & Hillenbrand, 1984; Flege,1995), (ii) that both skills are developed simultaneously (Best, 1995; Fowler, 1996; Rochet, 1995; Hazan et al., 2005, Thomson, 2008) and (iii) that production precedes perception (Flege & Eefting, 1987; Sheldon & Strange, 1982; Llisterri, 1995; Kluge et al. 2007; Hattori & Iverson, 2010). Thus, it might be the case that different types of sounds are acquired differently by different L2 learners. This hypothesis would explain the different results found in different previous studies and the disparity with the sound /ɜː/ in the current thesis. The production advantage observed in the present study in the case of the most dissimilar sound, /ɜː/, may indicate that a production advantage over perception might be possible in the case of "dissimilar" sounds, since they are the sounds to be acquired with more ease (Flege, 1995). All in all, these results provide further evidence to the fact that the relationship between the perception and the production of L2 sounds is rather intricate (Sheldon & Strange, 1982; Llisterri, 1995; Kluge et al. 2007; Baker & Trofimovich, 2005).

[29] Poorer perception was evidenced mostly when stimuli involved non-words, since learners identified this sound embeded in real words more accurately (identification ranging from 67.1 % -71.9%).

Summary

The production findings answer RQ.5, showing that for vowel training (Sub-RQ. 5.1), only the ID_V task promoted significant gains on production of the trained segments without any specific production training technique, in line with some previous studies (Rochet, 1995; Yamada et al., 1996; Bradlow et al., 1997, 1999; Hardison, 2003, 2004; Hazan et al., 2005; Iverson & Ewans, 2009; Thomson, 2011; Pereira, 2014). Interestingly, the DIS_V task was efficient in promoting an overall production gain when combining trained and untrained segments. This finding provided further support to the cross-training effect endorsed by this training method. In the case of the consonant training groups (Sub-RQ.5.2), the results of the present thesis show that the two perceptual training methods (ID_C and DIS_C) were not able to promote production gains. As previously argued, this is most likely connected to the short duration of the consonant training regimes, particularly the final voicing contrast. Nonetheless, it is relevant to mention that other reasons might have driven the lack of greater improvement in production. Recall that production measurements were limited to a few words per sound and elicited in citation form. Perhaps other styles of tasks may have shown more improvement and possibly more data would be needed to show differences from pre-test to post-test.

Part IV – Students' opinions about phonetic training

6.9. Results of the post-training survey

The last research question of the present study was concerned with students' impressions of phonetic training as an EFL tool to improve learners' pronunciation. Moreover, the study also aimed at finding out what training method students preferred (ID or DIS). In order to answer these questions, the experimental trainees were asked to complete an online survey (refer back to section 5.4.3.3 for more details) after the completion of the training sessions (week 7). The survey evaluated different aspects of the training through a 6-point likert scale, where 1 = *strongly agree*, and 6 = *strongly disagree*. As the purpose of this survey was informative, an even-numbered scale was chosen, in order to avoid a neutral answer and to force participants to position themselves when answering each question. Thus, responses 1-3 represented different degrees of agreement and responses 4-6 represented different degrees of disagreement. The responses are reported next, combining the three levels of agreement (1-3) and/or disagreement (4-6) as a single measure. General results and comments that apply to all training groups will be presented first, followed by the results and comments that are group-specific (ID_V, ID_C, DIS_V, DIS_C). All individual responses can be seen in Appendix H.

6.9.1. General results

Generally speaking, the results of this survey were positive, as 94% of the students agreed that a subject like *English Phonetics and Phonology 1* should include ear training exercises and 84% agreed that the sessions performed at the lab were very useful to help them work on their pronunciation. Furthermore, 91% of the students agreed that their ability to perceive English sounds improved and 79% claimed that they felt more confident when producing English sounds after having attended the sessions. Additionally, 90% of the participants found the lab sessions interesting and 89% found them enjoyable. Regarding the adequacy of the duration of the training sessions, 80% of the learners stated that the duration was not too long and 87% reported having managed to maintain high concentration levels throughout the sessions. However, 56% of the students reported feeling tired after the sessions. Regarding the participants' motivation, 96% stated that they felt motivated when their answers were right and 79% claimed that they felt frustrated when they were wrong. Interestingly, 93% of the students claimed that "some speakers were easier to understand than others", a fact that, as discussed above, is related to the results obtained in the tests of generalization to new talkers (section 6.6.3).

Concerning the use of non-words, 74% of the participants answered that they would have preferred to have heard real words instead of non-words during training. With respect to the training setting, less than half of the students 45% would have preferred to do the exercises from home in their free time. At the end of the survey, students were asked to anonymously express their opinion regarding the training regime. A few representative comments have been selected and are listed below (1) in order to illustrate the participants' general opinion regarding the training received. All the remaining comments can be seen in Appendix H. These selected comments were selected in order to illustrate the different views on phonetic training as an L2 training tool. Interestingly, they reflect some of the results previously reported throughout this section, as they provide additional qualitative data supporting some of the issues previously discussed (i.e. usefulness of the training methods, perceptual improvement as an outcome of the sessions, tiredness and length of the training regime applied).

(1) Selected open question responses:

- “In my opinion I would like to do this exercise during all the semester”;
- “Sessions were very useful and interesting”;
- “I found this activity really useful; I notice that my ear is more aware of different sounds now”;
- “I would recommend it to any of the future students attending *phonetics and phonology I* subject”;
- “I liked them, but I found them a bit too long and a bit tiring, but they really helped me”;
- “I understand the length of the exercise since it follows a scientific pattern, but as a student I've found it way too long”
- “I did not like doing it at the beginning, but I think it ended up helping me a lot”

6.9.2. Group specific results

Some group specific questions were included with the intention of assessing if there were differences between ID and DIS groups regarding whether they enjoyed the activity they were performing, whether they found it easy and/or difficult and what segments they focused their attention on. Moreover, a few group specific comments will also be shown in order to complement and support the students’ opinions and shed some light on which training method was preferred by the trainees.

6.9.2.1. ID vs. DIS

The trainees’ opinions on different aspects of the training regime will be presented according to the training method they were exposed to (ID or DIS). The answers report on aspects such as the difficulty, usefulness and tiredness levels experienced during the training regime as well as whether participants paid attention to untrained sounds in addition to trained ones, which constitutes one of the key considerations of the present thesis. The opinions for the vowel trainees will be described first, followed by the opinions provided by the consonant trainees.

6.9.2.1.1. Vowel trained groups

The comparison between the two training methods will focus on the answers to three main questions, which are listed in Table 6.40. As shown in the table, a total of 85% of the ID_V trainees and 73% of the DIS_V trainees stated that they liked the type of activity they performed, showing a somewhat greater acceptance of the ID method**.** Moreover, while only 50% of the ID_V trainees agreed that "The type of exercise I was doing was difficult", a total of 78% of the DIS_V trainees positively responded to the same statement. Regarding the focus of their attention concerning the trained and untrained segments (consonants vs. vowels), 69.2% of the DIS_V answered that they only paid attention to vowels during the vowel training sessions and the remaining 30.8% of students claimed that they paid attention to both vowels and consonants. As for the ID_V group, an exact 50% of the participants mentioned that they paid attention to the trained segment only (vowels) and the other half claimed to pay attention to both vowels and consonants during the sessions.

Table 6.40. Vowel trainees' opinions toward the training method received (ID vs. DIS)

Statement	ID_V	DIS_V
I liked the activity I performed	Yes - 85%	Yes -73%
The type of exercise I was doing was difficult	Yes - 50%	Yes -78%
I only paid attention to vowels (trained segment)	Yes - 50%	Yes -69.2%

As for the specific comments that the vowel trainees provided, it is interesting to mention that some vowel trainees specified that they would have liked to have had some practice with consonants. Another interesting comment came from one of the DIS_V participants in relation to the absence of labels in the DIS test. The student mentioned that *"it would be useful to know which specific sound we are listening to"*. This specific comment indicates that DIS trainees may have missed having more focus on form during the training regime received.

6.9.2.1.2. Consonant trained group

Table 6.41 shows the responses of the consonant trained groups regarding the three questions that compared both methods. As can be observed, 85% of the ID_C trainees enjoyed the activity they were performing in contrast to only 38% of the DIS_C trainees. Regarding the difficulty of the task, 55% of the ID_C trainees agreed that this method was difficult and 57.2% of the DIS_C trainees agreed that this method posed some difficulty for them. From the DIS_C trainees, 52.4% mentioned that they paid attention to the trained segment only (consonants) and 47.6% of the students mentioned that they were paying attention to both vowels and consonants during the training regime. Concerning the ID_C, 30% claimed to have paid attention to the trained segments (consonants) only, whereas 70% of the participants said that they paid attention to both segments.

Table 6.41. Consonant trainees' opinions toward the training method received (ID vs. DIS)

Statement	ID_C	DIS_C
I liked the activity I performed	85%	38%
The type of exercise I was doing was difficult	55%	57.2%
I only paid attention to consonants (trained segment)	30%	52.4%

Furthermore, similarly to the comments left by the vowel trainees, the consonant trainees mentioned two main things. On the one hand, they would have liked to have practiced the untrained segment (vowels) as well as the trained segments, and, on the other hand, they found the training regime too long, in particular the ID_C group.

6.10. Discussion IV – Post-training survey

This last part of the study investigated students' impressions on phonetic training as an EFL training tool and it aimed at assessing which method (ID vs. DIS) was favoured by Catalan learners of English. Students' impressions were assessed in the present study with the intention to add some qualitative data to the present investigation, and further our understanding of the effects of each training method used in the present study (ID vs. DIS). Moreover, some group-specific questions shed some light on students' views regarding phonetic training and each specific training method.

The general results revealed that the trainees see phonetic training as a useful tool in order to improve L2 pronunciation, and their overall impression of the methodology was positive. Most students found the sessions interesting, enjoyable and a good addition to the formal instruction received during the semester. In general, students were happy with the training being carried out in the lab at university rather than at home and tended to be pleased with the length of each training session. However, the majority of students reported feeling tired at the end of the sessions. This generally positive feedback from students is very encouraging, since phonetic training is a potential source of target language exposure and immediate corrective feedback in the foreign language context (Iverson et al., 2012; Thomson & Derwin, 2014), where specialized input quantity and quality are restricted (Muñoz, 2008; Derwing & Munro, 2015).

Furthermore, students reported that they would have preferred to have listened to real words instead of non-word stimuli during the sessions, which might be connected to the extra difficulty that perceiving sounds in non-word stimuli entails. This difficulty became evident in the case of vowels in this study, since students performed better when identifying real word stimuli rather than non-word stimuli from the outset of the study. As previously discussed (section 6.6.1), the fact that vowels were better identified in real words than in non-words might provide evidence to the formation of separate phonological and lexical categories for these specific sounds (Solé, 2013).

Another result that became evident from the answers provided by the learners was the fact that some speakers were easier to understand than others. This result, which was previously discussed in section 6.6.3, confirms previous findings that talker specific attributes provide a source of speech variability (Peterson & Barney, 1952) that affects L2 sound perception (Creel & Bregman, 2011). In training studies, this intra-talker variability is crucial, since high variability phonetic training (HVPT) has been found to obtain better results than low variability training (LVPT) (Lively et al., 1993; Pierce, 2014; Wong, 2014).

6.10.1. Group specific results

Students' impressions on the two training methods (ID vs. ID) were gathered with the intention to shed some light on the likeability, difficulty and usefulness of each task. Overall, the ID methodology was preferred by the participants (vowel trainees and consonant trainees). However, this preference was more pronounced for the consonant trainees. Only 38% of the DIS_C trainees reported having enjoyed the activity they were performing during training, whereas 85% of the ID_C reported having enjoyed performing this specific task and 73% in the case of DIS_V. These results go in line with Flege's (1995b) finding about learners' preference for the ID training method, precisely involving consonants. Two potential reasons for this task preference are explained next. The first possible reason lies in the absence of labels in the DIS task and the lack of focus on form that it entails. Recall that one of the DIS trainees expressed the desire of more focus on form by stating that "*it would be useful to know which specific sound we are listening to".* The absence of labels might have made the task less interesting for these learners, since they were learning the phonetic symbols in the phonetics course and they probably would have liked to have practiced labelling them during the training regime. Furthermore, the absence of labels probably makes the DIS task more challenging than the ID. In fact, the difficulty of the task is the second probable reason for the preference of the ID over the DIS by the Catalan learners. In general, more DIS trainees reported that the task was difficult than did ID trainees. That was true particularly for the vowel trained participants, specifically 78% (DIS) vs. 50% (ID) of the vowel trainees and 57.2% (DIS) vs. 55% (ID) of the consonant trainees.

These results taken together provide further evidence that the ID training method is preferred by the participants (Flege, 1995b) as most learners reported to enjoy this specific activity. Even though motivation lies beyond the scope of the present investigation, it might be the case that the greater success experienced by the ID methods is partly connected to learner motivation, since this is one of the non-linguistic factors that affect L2 learning (Calvo Benzies, 2014; Darcy et al., 2012; Piske et al., 2001).

One last aspect addressed at the post-training survey that is relevant for the discussion of the present investigation is connected to the attention paid to each segment during training. As previously explained, attention was oriented through explicit instruction (Pederson & Guion-Anderson, 2010; Alves & Luchini, 2016) in the present study. In this sense, two groups were vowel-oriented groups and two groups were consonant-oriented groups. From the main findings of this study, it was hypothesized that the DIS trainees paid attention to the word as whole, rather than only paying attention to the trained segments, which in turn promoted the cross-training effect. However, the qualitative results collected through the post training survey revealed that the DIS trainees were not consciously paying attention to the untrained segments, as expected. In fact, most of the trainees claimed to have paid attention to the trained segments only, especially in the case of vowels. A total of 69.2% of the DIS_V trainees claimed to have paid attention to vowels only. This fact sheds some light on the fact that by paying attention to the vowel sounds and most probably to the natural within-category variability present in vowel sounds (Peterson & Barney, 1952; Ladefoged, 2006), the learners may have been able to enhance their perception on final stops. This is so, since the preceding vowel duration is a stable and reliable cue to final stop perception (Raphael, 1972; Flege, 1989; Roach, 2000).

In the case of the consonant trainees, 47.6% claimed to have paid attention to the untrained segment, vowels. Even if this is not the majority of respondents, it constitutes of a notable part of the participants in this group. This result provides further evidence to the fact that some DIS_C trainees felt the need to rely on some specific characteristic of the vowels in order to identify the consonant sounds, as discussed in section (6.3.2). Interestingly, more ID trainees claimed to have paid attention to the untrained segments than the DIS trainees (50% for the ID_V and 70% for the ID_C). While this fact is not reflected in the numerical results of the study, it provides evidence to the fact that explicit instruction alone is not sufficient to direct learners' full attention to specific target sounds (cf. Pederson & Guion-Anderson, 2010; Alves & Luchini, 2016). It could be that the learners' were unconsciously attracted to the more stable cue.

In summary, the post training survey results answer RQ.6, which concerns learners' opinions towards phonetic training. The qualitative findings confirmed that learners' overall reaction towards phonetic training was positive, in line with previous findings (Rato, 2014) and that ID tends to be preferred over DIS (Flege, 1995b), answering Sub-RQ.6.1. Collecting students' opinions after a training regime can be a good method to complement the quantitative results obtained through training. It can shed some light on non-linguistic factors that might influence the training outcomes, and it can enrich our understanding of the type of learning that different training methods promote.

7. General discussion and conclusions

The current chapter seeks to provide the reader with a general discussion and conclusions of the study by offering a general overview of the main outcomes of the study and the implications for L2 phonetic training research. A visual summary of the results of the present investigation is provided first in Table 7.1. Then, the main objectives and methods applied in the current investigation are reviewed, and the main findings will be discussed in accordance with the research questions and the extant literature. Limitations of the present investigation and suggestions for further research are subsequently offered, alongside the main conclusions.

Table 7.1. Summary of results

Test	Vowels	Initial consonants	Final consonants
Main Perception + Gen 3a	**VT_Gs :** ID_V > DIS_V > CG ****CT _ Gs:** DIS_C > CG, ID_C	**CT_Gs:** ID_C = DIS_C > CG **VT_Gs:** ns	**CT_Gs:** ns ****VT_Gs:** DIS_V > CG, ID_V
Gen 1- Real words	**VT_Gs:** ID_V > DIS_V, CG **CT_Gs:** ns	**CT_Gs:** ns **VT_Gs:** ns	**CT_Gs:** ns **VT_Gs:** ns
Gen 2 - Novel non-words*	**ID_V** = strong gen **DIS_V** = Strong gen ****DIS_C**= strong gen	**ID_C**= strong gen **DIS_C** = strong gen	**ns**
Gen 3b - Novel talker (T7)*	**ns**	**ID_C**= strong gen **DIS_C** = strong gen	****DIS_V** = weak gen
Gen 3b - Novel talker (T8)*	**ID_V** = strong gen, **DIS_V** = weak gen ****DIS_C**= weak gen	**ns**	**ns**
Retention*	**ID_V** = yes **DIS_V** = yes ****DIS_C** = yes	**ID_C** = yes **DIS_C**= yes	****DIS_V** = yes
Production* (*****DIS_V**=marginally significant global improvement)	**VT_Gs:** ID_V > DIS_V,CG **CT_Gs: ns**	**ns**	**ns**

Note. VT_Gs= Vowel training groups, CT_Gs= Consonant training groups, ns= no significant differences, gen= generalization. *Follow-up effects (generalization, retention and/or transfer to production) are reported for groups that showed improvement from pretest to posttest only.** Instances of cross-training effect. *** Global improvement (vowels + consonants).

7.1. General discussion

The aim of the present work was to investigate the impact of two different HVPT training methods (ID vs. Categorical DIS) on the perception and production of trained and untrained L2 sounds by Spanish/Catalan learners. More specifically, this investigation was intended to shed some light on the effect of each training method on the perception and production of five SBE English vowels (/iː ɪ æ ʌ ɜː/) and stop consonants in word initial and word final position. In order to do so, 100 Catalan learners of English were divided into 4 experimental groups and a control group. The trainees underwent a phonetic training regime that lasted 5 weeks. Training stimuli consisted of naturally produced CVC non-words designed to contain a balanced distribution of the target sounds under investigation. The experimental groups differed either in terms of training method (ID or DIS) or focus of training (consonants or vowels), resulting in four different groups. Thus, two groups were instructed to attend to vowels and two groups were instructed to attend to consonants, all embedded in the exact same set of stimuli. All groups were tested on both consonants and vowels, thus assessing the effects of the two perceptual training methods on attended and unattended segments.

Learners' perception and production skills were assessed before and after training by means of a forced-choice ID task and a picture naming task. In order to assess the training effectiveness and learning robustness, several different measures were considered: (i) the overall gain of perception and production of trained sounds, (ii) the overall gain of perception and production of untrained sounds, (iii) generalization to real words, novel non-words and novel talkers, and (iv) retention of learning over time. The main results of this study will be discussed next in relation to each specific research question and previous findings in the literature.

7.1.1. Vowel training effects on trained and untrained segments

The first research question addressed the effect of high variability phonetic training (HVPT) on the perception of trained and untrained sounds. More specifically, it aimed at finding out whether two vowel training regimes (ID_V and DIS_V) promoted perceptual gains on vowels (trained segments) and on stops word initially and word finally (untrained segments). Additionally, it aimed at discovering which method was more efficient in promoting gain. Results of the trained sounds revealed that HVPT on vowels positively affected the perception of trained segments, and this improvement was facilitated by both methods tested in the present study (ID and Categorical DIS). However, the ID task proved to be superior to a categorical DIS task when training vowels. It is possible that a task familiarity effect may have played a role in this result, since perception was tested by means of an identification task only, which may have given an advantage to the ID trainees. It should be noted however, that no advantage for the ID trainees was observed in the case of the consonant trained groups, as discussed below.

Nonetheless, the significant difference between ID_V and DIS_V may provide evidence as to the superiority of an ID training method over a categorical DIS training method as an L2 vowel training tool. This result is in line with the only previous study, conducted on a smaller scale, which compared similar methods when training L2 vowel perception (Nozawa, 2015). A possible explanation for this advantage may lie in the fact that DIS tasks have been said to promote within-category sensitivity and probably tap into low levels of phonological encoding that are not greatly affected by language experience. In contrast, it has also been stated that ID tasks might enhance between-category sensitivity and may tap into higher levels of phonological encoding that are crucial for L2 categorization (Jamieson & Moroson, 1986; Logan & Pruitt, 1995; Iverson et al., 2003; Iverson et al., 2008, Iverson et al., 2012).

Still, the findings of the present study suggest that a categorical DIS task is also effective when training L2 vowel perception, even if to a lesser extent than the ID task. This study is pioneering in assessing the effect of these two methods (ID vs. categorical AX DIS) on vowel perception and the findings challenge previous views on the lower efficacy of discrimination tasks that were solely based on auditory DIS tasks (Strange & Dittmann, 1984; Jamieson & Moroson, 1986; Logan & Pruitt, 1995). However, the results of the present investigation are more in agreement with a previous work that compared these same methods in a training study involving English final stops (Flege, 1995b). Flege found that both ID and categorical DIS were equally efficient methods, thus showing the efficacy of a categorical DIS task to train L2 stop perception. This may be because, contrary to the auditory DIS task, the categorical DIS task exposes learners to a greater range of acoustic variability (Polka, 1992), which in turn, may promote L2 categorization (Strange, 1992).

Another important and novel finding of the present study in relation to RQ.1 was the effect on untrained but implicitly exposed segments promoted by the DIS_V method. This result indicates that vowel training by means of a categorical DIS task was able to modify learners' perception of final consonant stimuli, showing a cross-training effect. Three possible explanations for this effect on the untrained/unattended sounds are offered. The first possible explanation lies in the difference in stimuli presentation offered by both tasks. That is, the ID task presents one stimulus at a time, which might force learners to directly compare the given stimulus with a previously stored mental representation of the corresponding sound category. In contrast, a categorical DIS task exposes learners to two physically present stimuli to be compared at every trial, which in turn might increase learners' awareness of the unattended segments. The second possible explanation is connected to the absence and/or presence of labels in each task. It is hypothesized that the absence of labels in the DIS task made leaners pay attention to the whole stimulus, focusing on trained and untrained segments. The third possible explanation is related to the fact that DIS tasks might enhance within-category variability awareness (Strange & Dittmann, 1984; Jamieson & Moroson, 1986; Logan & Pruitt, 1995). In addition, since the perception of final consonants is cued by the within-category properties present in the preceding vowel sounds, the Catalan learners might have been able to apply the enhanced awareness of the preceding vowel duration differences when identifying final consonants.

On the other hand, the forced-choice ID_V method strictly directed learners' attention to the target sounds. This is probably due to the presence of labels in this task, which forces learners to categorize each stimulus as one of the options provided (Jamieson & Morosan, 1986). The presence of specific labels in the ID task might limit L2 perception results (Polka, 1992; Logan & Pruitt, 1995; Beddor & Gottfried, 1995). In fact, the presence of labels in the vowel identification task and the absence of labels in the vowel discrimination task might make these two tasks different in the nature of levels of processing involved. This might be explained in terms of two different perceptual task types: overt and covert (Bohn, 2002). While discrimination tasks constitute an overt task type, in which both categories to be contrasted are physically presented during trials, identification tasks constitute a covert task type. In the latter, the only presented category is directly compared with a pre-existing memory representation, in the absence of another physically present category.

Furthermore, in the ID task, participants were provided with a more structured, or explicit way of learning the target sounds (sounds are presented to them alongside the information of specific symbols and common spellings for each sound). By contrast, in the DIS training task participants only heard the target sound in a stimulus paired with another sound either from the same category or a different category. The feedback they received strictly reflected their capacity to distinguish between the two sounds, not on their ability to relate the given sound to a symbol and /or spelling. Consequently, their learning of the sounds was not related to a specific form, but is more implicit. Thus, it could be argued that in the ID training, the participants were exposed more explicitly to the sounds of the target language, in contrast to the DIS training, since the latter did not provide them with any focus on form, making the exposure to the sounds more implicit.

Taken together, the findings from the first research question suggest that there is an advantage to discrimination training in the perception of untrained sounds. Thus, it shows that while ID tasks might overall be more effective for L2 vowel training, categorical DIS tasks seem to promote improvement with the final stop consonant present in the stimuli to a greater extent.

7.1.2. Consonant training effects on trained and untrained segments

As regards the second research question, which concerned the effects of the two consonant training methods (ID_C and DIS_C) on trained (initial and final consonants) and untrained sounds (vowels), it was observed that both training methods promoted significant gains in the perception of directly trained initial consonants to a similar extent. Recall that the present study is novel in assessing the effectiveness of these two methods (ID and categorical DIS) when training the perception of initial consonants, and the results are in agreement with Flege's (1995b) and Nozawa's (2015) findings for coda stops and coda nasals, respectively. Interestingly, despite a possible task familiarity effect advantage for the ID trainees, the DIS task promoted gains in the identification of initial stops that did not differ significantly from the gains obtained through the ID training task. The positive results found with the DIS_C group are particularly interesting when contrasted with the results found for the DIS_V group (RQ.1), since they show that a DIS task was able to promote greater results with directly trained initial consonants (DIS_C = 12.8% gain) than with directly trained vowels (DIS_V = 9.8% gain) in a shorter amount of time (i.e. half of the amount of training time dedicated to vowel training)[30]. These results provide further evidence that initial stop perception is easily modified through perceptual training (Carney et al., 1977; Pisoni et al., 1982; Collet et al., 2013), since the VOT difference between the voiced and voiceless stimuli is psychoacoustically robust (Burnham, 1986). Moreover, it indicates that the trainability of initial voicing contrasts might be relatively easy in comparison to other distinctions (Strange & Dittmann, 1984). Thus, training consonants and vowels might require different protocols (Nishi & Kewley-Port, 2007; Nozawa, 2015) and different amounts of training, since consonants and vowels do not entail the same degree of effort on the part of L2 learners (Aliaga-García & Mora, 2009; Pereira, 2014). Thus, the novelty of the balanced cross-design applied in the current thesis made the comparison of vowel and stop consonant training possible.

Furthermore, the similar performance of the two methods with initial consonants but not with vowels also points to an important outcome: the fact that ID is superior to DIS when training vowels, but not when training initial stops. The most probable explanation for this fact lies in the interplay of the skills enhanced by each method (low vs. high levels of phonological encoding) and the nature of each individual segment. That is, the phonetic and phonological differences between consonants and vowels (Pisoni, 1973). Recall that vowels and consonants have been found to be perceived differently (Fry et al., 1962; Pisoni, 1973; Strange, 2007), involve different levels of processing when learned (Pisoni, 1973) and activate different neuronal patterns when processed (Carreiras & Price, 2008; Engineer et al., 2015). According to Liberman, Cooper, Shankweiler and Studdert-Kennedy (1967), consonants are generally more acoustically stable than vowels, which facilitate perception. It might be the case that training through higher levels of phonological encoding might be more beneficial for vowel perception, since they are perceived more continuously rather than

[30] Recall that the training time was divided in half for consonants, training both initial and final consonants.

categorically. On the other hand, initial stops benefit equally from a task which trains higher level of phonological encoding and a task which trains lower level of phonetic processing. This might be connected to the acoustic stability found in consonants (Liberman et al., 1967), since lower levels of phonetic processing are connected to the physical properties of the sounds.

Conversely, learners' perception of final consonant stimuli was not significantly enhanced by any of the consonant training methods (cf. Flege, 1995b). Possibly, this outcome was due to the restricted training time (Flege, 1989; Flege, 1995b), indicating that training final consonant perception requires a longer amount of time and effort than training initial consonant perception. This is so especially when the final consonant voicing distinction is non-existent in the learners' L1 (Flege, 1989; Flege et al., 1992; Flege et al., 1995; Flege, 1998). It is possible that this difference lies in the fact that final stop voicing contrasts are cued by the preceding vowel duration in English (among other cues), and thus are usually perceived more continuously than categorically (Raphael, 1971; Pisoni, 1973). Interestingly, the only method that was able to enhance the identification of final consonant stimuli significantly was the DIS_V method (7.7% gain from pre to post-test). Through a cross-training effect, which most probably brought attention to the within-category variability properties of vowels that cue final consonant voicing, a vowel training regime promoted final consonant improvement. Thus, while these within-category differences found in vowels may not be relevant for training vowels (Jamieson & Moroson, 1986), they may be an essential cue for final stop training and identification. This is so since vowel duration is a reliable cue to the identification of final stop voicing (Raphael, 1972; Roach, 2000).

Moreover and remarkably, the DIS_C group also enhanced learning on the untrained segments (vowels). Thus, three experimental groups promoted vowel perception gains: ID_V (26.3%), DIS_V (9.8%) and DIS_C (9.7%). While the ID training was superior to the two

DIS methods under investigation, a noteworthy finding is that the DIS_C trainees improved on vowels despite not having received feedback on vowels. The controlled exposure to the target sounds might have been responsible for this outcome, since the group who was receiving feedback on consonant sounds (DIS_C) also improved on vowels to a similar extent. In fact, this enhanced ability with vowels as an effect of consonant training has previously been reported in a different area of study: L1 phonological therapy. In an unpublished study, Stemberger (J. Stemberger, personal communication, September 8^{th}, 2015) observed that, by training pre-schoolers with phonological impairment on consonant sounds, the infants improved their ability of producing consonant and vowel sounds.

On the whole, the results of both DIS training groups verify the cross-training effect with vowels and stops promoted by this method as a consequence of the balanced exposure to the target sounds, which is one of the main contributions of the present investigation and supports the tendency observed in a previous small scale study on vowels and nasal consonants (Nozawa, 2015). Furthermore, this novel finding provides further evidence of the fact that vowels affect consonants and vice-versa (Recasens & Mira, 2015; Steinlen, 2005). On the basis of these results, it could be concluded that the DIS method seems to be able to promote gains not only with the trained, but also with the untrained segments. A possible reason for this may be connected to the fact that final stops are cued by the preceding vowel, which may have allowed participants to focus their attention on this stable cue for final consonant perception (Raphael, 1972; Roach, 2000) rather than attempting to rely on the subtle final voicing distinction present in the stimuli.

In sum, the vowel and consonant results found in the present thesis suggest that different training methods (ID and categorical DIS) might play different roles when training vowels and consonants. An ID method was found to enhance L2 vowel perception to a greater extent than a categorical DIS method, whereas both methods were equally effective in modifying initial consonant perception. In addition, the DIS method was found to enhance the perception of the untrained but implicitly exposed sounds. Thus, in order to maximize the effect of the training regime and profit from the entire exposure of the CVC stimulus, a combination of ID and categorical AX DIS tasks is suggested. The effects of both training tasks seem to complement each other and suggest that combining more than one method of training in the same regime might promote a more ample L2 learning effect. This combination would not only allow learners to develop different perceptual abilities (within and between category sensitivity), as it would also promote learning on trained and untrained segments present in the stimuli, provided the training materials involve a balanced exposure to all target sounds.

7.1.3. Generalization of learning acquired through training

The results of the third research question, which investigated the generalizability of learning to novel dimensions (real word stimuli, novel non-words, novel talkers and absolute novel talkers) and the effect of each training method on the generalization of learning will be summarized and discussed next. In the case of generalization to real words, recall that learners performed better with real words than with non-words at the outset of the study, at least with vowels. This result is in agreement with previous findings from lexical decision task studies (Solé, 2013). According to Solé, better performance with real words rather than non-words might suggest that L2 phonological categories for these specific sounds (vowels in the case of the present study) might be formed after lexical categories.

As for the generalization results, only the ID_V group was able to transfer the knowledge obtained through training with non-words to the identification of real words. This result is in accordance with previous research showing that the ID method is efficient in promoting generalization from non-word to real words (e.g. Pierce, 2014; Nishi & Kewley-

Port, 2007). The remaining three groups were not found to promote this type of generalization, suggesting that the method (DIS) and the limited amount of training time applied in the case of the ID_C were not effective enough in promoting transfer from non-word stimuli to real words. The advantage found for ID_V in comparison to DIS_V might also be explained by the presence of phonetic symbols and/or orthography in the labels. It is hypothesized that both elements provided learners with extra tools and focus on form, which is beneficial for learning (Saito, 2015). Likewise, the orthographic representation might have facilitated lexical access during training and thus provided learners with an extra layer of information to rely on (Escudero et al., 2008). This may be because orthographic representations, phonological encoding and phonological memory may interact (Schiller, 1998; Bassetti, 2006, Escudero et al., 2008). Further research would ideally compare these two methods without the inclusion of orthography and/or phonetic symbols in the ID task (e.g. Pierce, 2014; Thomson, 2011), trying to make the two tasks more comparable. For example, abstract representations and/or pictures could be used as labels in the identification task instead of words or phonetic symbols, and the DIS task could include information about the specific sound categories presented in each trial.

In contrast to the results for generalization to real words, all four training groups (ID_V, ID_C, DIS_V, and DIS_C) generalized the learning acquired through training to novel non-words (i.e. novel words produced by familiar talkers) similarly (strong generalization in all cases)[31]. This result supports previous findings that both training methods (ID and DIS) are equally effective and promote robust learning to a similar extent (Flege, 1995b, cf. Strange & Dittmann, 1984; Jamieson & Moroson, 1986; Logan & Pruitt, 1995). Moreover, this result adds to the HVPT literature which reported positive generalization effects to novel stimuli (Lacabex et al., 2009; Rato, 2014, Pereira, 2014,

[31] Recall that *weak generalization* = post-test, but = pre-test, and *strong generalization* = (post-test = or < generalization) and (generalization > pre-test).

among several others).What is more, not only the trained sounds were found to generalize. The DIS_C trainees were able to generalize their vowel learning (untrained segments) to novel non-word stimuli, as a result of training. This suggests that a short training regime in consonants was effective in enhancing learners' perception of trained as well as untrained sounds, and provides further support for the cross-training effect as a result of a categorical DIS training method.

Another dimension that can be assessed in training studies is the generalization to novel talkers, that is, talkers who differ from the training talkers. As evidenced by the gain obtained from pre-test and post-test, all four methods were efficient in promoting this transfer. These results support the effectiveness of both ID and DIS as L2 training tools (Flege, 1995b) for vowel and initial consonant training. However, recall that the ID method was found to be superior to DIS for vowel training, and both were found to be equally effective when training initial stops.

Generalization to *absolute novel talkers*, that is, talkers heard neither at the training phase nor at the testing phase, was also observed. However, different training methods promoted different degrees of generalization, and only to some talkers. Interestingly, some specific characteristics present in the talkers' speech affected the perception by the Catalan learners (i.e. T8's slightly longer lag in voiced stops; T7's shorter duration of the sound /ɜː/, and high F1 value for the sound /ɪ/). This result suggests that generalization to new talkers might be talker-specific, since specific characteristics present in some talkers' speech might deviate from the previous exposure during training, in spite of falling into the range of within-category variability recognized by native listeners. This finding further supports the importance of using highly variable stimuli during training procedures.

Looking at the cases in which generalization to absolute novel talkers did take place, it was observed that the ID_V group promoted a *strong* generalization effect of the directly trained segments, vowels. However, the other two groups that had initially promoted perceptual gains on vowels (DIS_V and DIS_C), revealed a *weak* generalization effect. Once again, this result indicates that an ID method is more effective than a DIS method when training vowels sounds, as generalization indicates that robust learning took place (Logan & Pruitt, 1995). In line with the main results for initial consonant sounds, both consonant training groups were able to transfer their learning to *absolute new talkers* to a similar extent, since both training methods promoted *strong* generalization effects. This further piece of evidence points to the same fact previously argued: both training methods are equally effective in training initial stop consonants and promoting generalization of learning. This finding is also in accordance with the few previous studies that contrasted similar training methods when training consonants in coda position (Flege, 1995b, Nozawa, 2015).

Thus, the initial hypothesis for generalization, which predicted that both methods would promote generalization of learning to novel dimensions to a similar extent, was partially confirmed. The results of the current study showed that different methods promoted generalization of learning differently. While real word generalization was only promoted by the ID method specifically for vowels, both methods promoted generalization to novel non-word and novel talkers for both consonants and vowels. Interestingly, generalization to absolute novel talkers was found to be talker-specific and the results corroborate previous findings that an ID method promotes better results for vowels, while both ID and DIS stimulate consonant learning similarly. Furthermore, the cross-training effect was found to successfully generalize to new talkers and absolute novel talkers for both DIS training groups and additionally to novel non-words in the case of DIS_C.

7.1.4. Retention effects promoted by training

The fourth research question of the present study aimed at measuring the robustness of learning acquired during a five-week training period, by looking at the retention effects promoted by the two HVPT methods under investigation. It was hypothesized that this retention would take place, in line with previous findings (Bradlow et al., 1997, 1999; Lively et al., 1994; Wang, 2002; Wang & Munro, 2004; Nishi & Kewley-Port, 2007) and it would be promoted by both training methods to a similar extent. In fact, results confirmed that both training methods (ID and DIS) were equally effective in promoting retention effects that were still observed two months after the training regime was over. These findings support the initial prediction, confirming the robustness of the perceptual learning (Logan & Pruitt, 1995) and the pedagogical relevance of a HVPT method. Remarkably, the cross-training effect observed for both DIS groups (DIS_V on final consonants and DIS_C on vowels) was also retained for two months, showing that the learning of the untrained segments was robust.

7.1.5. Production gains as a transfer from perceptual training

The fifth research question (RQ. 5) sought to investigate whether the learning acquired through a five-week perceptual training regime would transfer to production. Production was evaluated by means of a series of rating tasks in which 12 native English speakers identified and rated the production data elicited at pre-test and post-test. Results revealed that the DIS_V group showed a marginal improvement on the overall production (vowels and consonant sounds combined) and the ID_V group was the only group that showed a significant effect on directly trained segments (vowels). This finding contributes to some previous studies that show that a perceptual training method might alter the production of L2 sounds, at least to some extent, without the need of explicit production training

(Yamada et al. 1996; Bradlow et al., 1997; Flege, 1989; Hardison, 2003; Rochet et al., 2005; Hazan et al. 2005; Lambacher et al, 2005; Lengeris, 2008; Iverson et al., 2012; Thomson, 2011; Pereira, 2014; Rato, 2014). It also sheds some light on the effect of each training method on the production of trained and untrained sounds. These findings suggest that the cross-training effects observed for the DIS_V trainees in perception seem to also have affected learners' overall production abilities, since NES perceived these learners to be marginally better after training than prior to it. Recall, however, that when analysing consonant perception only (cross-training), the results did not reach significance.

Focusing on the vowel segments only, the ID_V was the only experimental group which promoted significant gains on the production of vowel segments, which suggests the superiority of the ID method when compared to the DIS method when training vowel sounds. A possible reason could lie in the fact that since production assessment included real words, the orthographic representation present in the labels of the ID training might have played a role. Also, recall that the ID_V was the only group which promoted generalization to real word stimuli, so it makes sense that they were the only group to promote gains on the production of real words. Conversely, no improvement was observed for any of the consonant training groups, either on trained or untrained sounds. This effect might be connected to the limited time devoted to training each consonant segment (initial and final stops). In the case of initial stops, the short training regimes seem to have been sufficient in modifying learners' perceptual performance, but not in order to transfer learning to L2 production abilities. As for final consonant training, on the other hand, the short training regimes were not effective in promoting either perceptual or production gains.

Taken together, these findings fit the predictions of the SLM and the NLM, which postulate that perception gains occur prior to production gains and the former is a prerequisite for the latter. However, it seems that the learners are at the stage that perception is more developed than production, since the perceptual gains were overall greater than the production gains in the study and the improvement in both domains was not correlated. According to Iverson et al. (2012), perception and production may rely on different factors and possibly on different abstract representations. Thus, this result provides further evidence that the improvement in both domains does not seem to occur in parallel (Bradlow et al., 1997; Pereira, 2014; Iverson et al., 2012; cf. Rochet, 1995; Wong, 2012); that is, changes in perception and production seem to develop differently.

7.1.6. Students' opinions on phonetic training and different training methods

The sixth and last research question explored students' opinions about phonetic training as an L2 learning tool and about the specific training methods under investigation. Trainees filled out a training assessment survey (Appendix H) immediately after the completion of the training regime. Both hypotheses, which projected that learners would react positively towards phonetic training as a learning tool and that the ID method would be favoured over the DIS method, were confirmed. The qualitative results collected in this investigation established that the Catalan learners find phonetic training to be a positive complement to the phonetics and phonology classes, in line with previous research (Rato, 2014). However, the ID is preferred over a DIS method, especially in the case of stops (Flege, 1995b). Interestingly, the fact that learners preferred one method over the other might have generated different levels of motivation throughout the training regime. Knowing that motivation is one of the non-linguistic factors that affect L2 learning (Calvo Benzies, 2014; Darcy et al., 2012; Piske et al., 2001), it is possible that part of the superiority found for the ID trainees has been promoted by this non-linguistic factor.

7.2. Limitations of the present study and directions for further studies

The present work is not without limitations, which suggest directions for further exploration. First, a design limitation lies in the similarity of the task used for testing learners' perception and the task used for ID training. The probable task familiarity effect generated for the ID training groups might have influenced the outcomes of the present investigation. However, a significant advantage for the ID trainees was only observed for L2 vowel identification, as both ID_C and DIS_C patterned similarly when identifying initial stops. Thus, it remains unclear whether a task familiarity effect played a significant role in the outcomes of the current investigation, and it raises an intriguing question: why would a task familiarity effect have a greater impact on the perception of L2 vowels rather than L2 initial stops. A possible answer to this question lies in another methodological issue found in the present investigation, that is, the fact that the consonant ID task and the vowel ID tasks used different types of response options. Recall that the response options for the consonant ID task only included the phonemic symbol corresponding to each sound category, whereas the vowel testing task required learners to label the sounds by selecting the corresponding phonemic symbol accompanied by two words that contained that sound. In a way, the vowel ID trainees were not only trained on categorizing such sound, but they were also trained on associating the L2 category with the phonemic symbol and common spelling. By contrast, the consonant trainees did not go through this association training, as the grapheme-phoneme correspondence for stops is more straightforward than for vowels, and besides, no sample words were provided for each response label. Future studies could incorporate alternatives to the use of phonetic symbols and orthographic representation in ID tasks, such as labels made up by pictures (e.g. Pierce, 2014; Thomson, 2011).

Another way of controlling for a potential effect of task familiarity would be to test trainees using the two tasks used in training, that is, both ID and DIS[32]. In fact, a later training study using the same stimuli (Cebrian, Carlet & Gavaldà, 2016) tested vowel trainees in both abilities, and revealed that ID enhanced the identification of vowel sounds to a greater extent than the DIS method did, extending the findings of the current study. Interestingly, the ID method enhanced learners' vowel discrimination abilities to a similar extent as the DIS method did, in line with previous findings (Wayland & Li, 2008). Hence, the preliminary findings of Cebrian et al. (2016) confirm the superiority of the ID training method for L2 vowel identification, as this method was able to enhance both perceptual abilities (identification and discrimination) either to a similar or to a greater extent than the categorical DIS method did.

Another limitation of the present thesis is in relation to different training times for consonants and vowels. This is explained by the inaccurate assumption that initial and final consonants could be seen as a single global measure at the early stages of the study. Thus, the global amount of training time was divided into initial and final consonant training. The short length of training for consonants might have had a negative impact on the results, since no significant improvement was found for final consonant sounds and no production and/or generalization to real words was observed for initial stops. According to Flege (1989), learners should be exposed to a large number of trials in order for training to work, especially when this L2 feature is non-existent in the L1. Thus, initial and particularly final stops would ideally be trained for a longer period of time and with the exposure to more repetitions of fewer tokens.

Moreover, while the perceptual training involved non-words, production was assessed by means of eliciting real words only. It would be of interest to examine whether learners

[32] Both testing and training of ID and DIS were part of the original design of the current thesis. However, due to practical reasons, both tasks could not be implemented.

would have produced the trained non-word stimuli better than new real words. In addition, using non-words would avoid any possible influences of word familiarity.

Another factor that may have affected the results of the present work is the use of orthographic representation in the perceptual and in the production tasks. In the case of perception, pictures could be used instead of symbols and orthography, and in the case of production, a delayed repetition task could be used to avoid orthographic interference. Moreover, the evaluation of production involved only a limited set of words elicited through a controlled task. Ideally, production would be assessed by means of more tokens elicited in a more naturalistic task (e.g. picture description task, elicited narratives, conversational speech). This is because previous research points to the fact that "L2 production data elicited in experiments involving the use of written materials may not always reveal how accurately bilinguals can produce the sounds of an L2 in conversational speech" (Piske, Flege, Mackay & Meador, 2010, p. 195). Another issue that might have affected the outcomes of the present investigation is the fact that the participants were enrolled in a phonetics and phonology course and thus acquired meta-linguistic knowledge during the training regime. Ideally, it would be interesting to know the extent to which the knowledge of phonetics and phonology may have affected the results, and further test students without this knowledge. Furthermore, production data was analysed by means of NES ratings only. An acoustic analysis of the stimuli would be necessary to fully evaluate the effect of perceptual training on production. This would be informative regarding the specific kind of changes that each training method was able to promote after a period of five weeks[33]. In addition, contrasting the acoustic data with the NES ratings would allow an analysis of what aspects of production play a more relevant role in NES perception. Another limitation was that, due to the beginning of the new

[33] A preliminary acoustic analysis of the initial stop consonant data at pre-test and post-test revealed a similar finding to the one obtained through native speaker ratings. That is, a numerical improvement in the right direction (longer VOT for voiceless stops) for both consonant training groups (ID_C and DIS_C), albeit non-significant.

academic semester, the delayed post-test was carried out only two months after the post-test. It would be ideal to investigate the retention effects after a longer period of time. Finally, it could be interesting to explore possible factors affecting the different degrees of improvement, such as individual differences, inhibitory control, phonological memory, motivation and language aptitude. These factors might reveal possible reasons to account for the varying degrees of success in the present study, as previous research points to the fact that non-linguistic factors and individual differences in cognitive abilities affect the perception and production of L2 sounds (Mora, 2014; Aliaga-Garcia et al., 2011; Safronova & Mora, 2012; Darcy, Mora & Daidone, 2014).

7.3. Summary and conclusions

In summary, this study assessed the effects of two perceptual methods on trained (attended) and untrained (unattended) sounds, and it demonstrated positive changes in L2 learners' perceptual and production abilities as a result of high variability phonetic training (HVPT). More specifically, the present investigation provided evidence that identification training is more effective than discrimination training at improving perception and production (to a lesser extent) of L2 vowel sounds, in line with some previous studies. This was true both with the perception of nonsense words as well as when testing perception and production of real words. Nevertheless, categorical discrimination training was also effective in improving vowel perception, even if to a lesser extent than identification training. Besides this, both methods were similarly found to generalize and retain the learning acquired through training. As regards to training consonant sounds, performance with final stops was not successfully enhanced by any of the two short training regimes under investigation. On the other hand, both ID and categorical DIS methods were found to promote gain, generalization effects and retention effects to a similar extent when training initial stops differing in VOT. This suggests that modifying the perception of different types of segment might require different training procedures and amounts of training time.

Importantly, a novel finding of this study concerned the cross-training effect. That is, only categorical DIS discrimination training was found to positively affect the perception and marginally the production of untrained as well as trained sounds. Possibly, due to the absence of labels and the fact that learners were exposed to two physically present stimuli in each trial, the DIS method might have facilitated attention to segments beyond the target sounds. Conversely, by specifically directing trainees' attention to trained segments, ID training may have forced learners to focus only on the target sounds. Since the majority of training studies make use of CVC words, this may be an important result. This is because whilst training one type of segment, the implicit exposure to the untrained segments present in the stimuli might contribute to learners' overall perception. In line with these results, a combination of both tasks (ID and categorical DIS) is suggested in order to enhance different perceptual abilities and maximize the effects of training.

The post-training survey suggested that, despite the overall positive reaction towards phonetic training as an L2 learning tool, the trainees reacted more positively towards the ID method rather than the DIS method. This was especially true in the case of stops. Most notable in participants' comments was the lack of focus on form provided by the DIS method, since students would have liked to know what sounds they were asked about.

In conclusion, this study is innovative in that it compares the effects of two training methods (ID vs. DIS) on the perception and production of both vowels and stops word initially and finally through a cross-design. It succeeds in shedding some light on the interplay between the different training methods and the trainability of different segment types through a controlled and balanced design and thus it contributes to the literature on phonetic training by establishing further the potential benefits of perceptual training.

Moreover, it was novel in finding that a categorical DIS training method might enhance perception of untrained as well as trained segments, by showing that controlled exposure to stimuli might be able to enhance listeners' sensitivity even when their attention is not on the target sound. Finally, it evidenced the fact that HVPT can be an efficient tool to enhance learners' perception and production abilities, and it is welcomed by students. Unfortunately, despite the success and acceptance from part of the learners, phonetic training methods are rarely implemented in the classroom (Liberman, 2008; Pierce, 2014). There is a need to bridge HVPT research and teaching practices, by making sure that this powerful tool to train perception is pedagogically implemented.

References

Aliaga-García, C. (2007). The role of phonetic training in L2 speech learning. In *proceedings of phonetics teaching & learning conference, PTLC 2007* (pp.1-5). London, UK: PTLC.

Aliaga-García, C. (2009). Effects of audiovisual auditory and articulatory training on second-language (L2) vowel category learning. *The Journal of the Acoustical Society of America*, *125*(4), 2765-2765. doi:10.1121/1.4784692

Aliaga-García, C. (2010). Measuring perceptual cue weighting after training: A comparison of auditory vs. articulatory training methods. *In Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 12-18). Poznań, Poland: Adam Mickiewicz University press.

Aliaga-García, C. (2016). *Vowel learning: Perceptual versus articulatory training*. Paper presented at the 8th international conference of language acquisition. University of the Balearic Islands, Palma de Mallorca, Spain.

Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. *Recent research in second language phonetics/phonology: Perception and production*, 2-31.

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics, 47*, 1-14. doi:10.2478/psicl-2011-0002

Alves, U. K., & Luchini, P. L. (2016). Percepción de la distinción entre oclusivas sordas y sonoras iniciales del inglés (LE) por estudiantes argentinos: Datos de identificación y discriminación. *Lingüística, 32*(1), 25-39. doi:10.5935/2079-312x.20160002

Alves, U., & Zimmer, M. (2015). Perception and production of English VOT patterns by Brazilian learners - the role of multiple acoustic cues in a DST perspective. *Alfa*, *59*, 155-175.

Bach, S. (2012). *Coexistence of phonetic systems in Danish/English bilinguals: A study of the production of VOT categories.* (Unpublished MA thesis). University of Aarhus, Aarhus, DK.

Baker, A. (2013). Exploring teachers' knowledge of second language pronunciation techniques: teacher cognitions, observed classroom practices, and student perceptions. *TESOL Quarterly*, *48*(1), 136-163. doi:10.1002/tesq.99

Baker, W., & Trofimovich, P. (2005). Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech*, *48*(1), 1-27. doi:10.1177/00238309050480010101

Bassetti, B. (2006). Orthographic input and phonological representations in learners of Chinese as a foreign language. *Written Language & Literacy*, *9(1)*, 95-114. doi:10.1075/wll.9.1.07bas

Beddor, P., & Gottfried, T. (1995). Methodological issues in cross-language speech perception research with adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 207-232). Timonium, MD: York Press.

Benkí, J. (2005). Perception of VOT and first formant onset by Spanish and English speakers. In *Proceedings of the Fourth International Symposium on Bilingualism* (pp. 240-248). Arizona, USA: Arizona State University.

Berry, J. (2004). Control of short lag VOT (voice-onset time) for voiced English stops. *The Journal of the Acoustical Society of America*, *115*(5), 2465. doi:10.1121/1.4782396

Best, C. (1995). A direct realist view of cross-language speech perception. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 171-204). Timonium, MD: York Press.

Best, C. C., & Mcroberts, G. W. (2003). Infant perception of non-native consonant contrasts that adults assimilate in different ways. *Language and Speech, 46*(2-3), 183-216. doi:10.1177/00238309030460020701

Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 13-34). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Boersma, P. (2001). Praat, a system for doing phonetics by computer. *Glot International, 5(9/10)*, 341-345.

Bohn, O., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics, 11*(03), 303. doi:10.1017/s0142716400008912

Bohn, O., & Flege, J. E. (1992). The production of new and similar vowels by adult german learners of English. *Studies in Second Language Acquisition, 14*(02), 131. doi:10.1017/s0272263100010792

Bohn, O., & Flege, J. E. (1997). Perception and production of a new vowel category by adult second language learners. *Second Language Speech,* 53-74. doi:10.1515/9783110882933.53

Bohn, O., & Munro, M. J. (2007). The study of second language speech learning. *Language Experience in Second Language Speech Learning: Language Learning & Language Teaching,* 3-11. doi:10.1075/lllt.17.06mun

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 275-300). Timonium, MD: York Press.

Bohn, O-S. (2002). On phonetic similarity. In Burmeister, P. (Ed.), Piske, T. (Ed.), Rohde, A. (Ed.) An integrated view of language development: Papers in honor of Henning Wode (pp. 191-216). Trier: Wissenschaftlicher Verlag.

Bowen, C. (1998). Give me five. A broad-based approach to phonological therapy. In *Communications ahead. Speech-Language Therapists' Association Biennial Conference Proceedings* (pp.1-11). New Zealand; NZ: Dunedin University.

Bradlow, A. R. (2008). 10. Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ - /l/ contrast. *Studies in Bilingualism Phonology and Second Language Acquisition*, 287-308. doi:10.1075/sibil.36.14bra

Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics, 61*(5), 977-985. doi:10.3758/bf03206911

Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., & Tohkura, Y. (1997). Training Japanese listeners to identify English /r/ and /l/: IV. Some effects of perceptual learning on speech production. *The Journal of the Acoustical Society of America, 101*(4), 2299-2310. doi:10.1121/1.418276

Broadbent, D. E. (1958). Effects of noise on behaviour. In C.M. Harris (Ed.), *Handbook of Noise Control* (pp. 10-34). New York: McGraw-Hi.

Brown, C., & Matthews, J. (1997). The role of feature geometry in the development of phonemic contrasts. *Focus on Phonological Acquisition Language Acquisition and Language Disorders*, 67-78. doi:10.1075/lald.16.05bro

Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics, 7*(03), 207. doi:10.1017/s0142716400007542

Busà, M. G. (2008). New perspectives in teaching pronunciation. In *From Didactas to Ecolingua* (pp. 165-182). Trieste:Università di Trieste.

Calvo, Y. J. (2014). The teaching of pronunciation in Spain: students' and teachers' views. In T. Pattison (Ed.), *IATEFL 2013 Liverpool Conference Selections (*pp. 106-108). Faversham, UK: IATEFL.

Calvo, Y. J. (2016). *The teaching and learning of English pronunciation in Spain. An analysis and appraisal of students' and teachers' views and teaching materials.* (Unpublished doctoral dissertation). University of Santiago de Compostela, Santiago, Spain.

Carbonell, J. F. & Llisterri, J. (1999). 'Catalan.' In international phonetic association (Ed.), *Handbook of the international phonetic association* (pp. 61-65). Cambridge, UK: Cambridge University Press.

Carlet, A., & Rato, A. (2015). Non-native perception of English voiceless stops. In E. Babatsouli & D. Ingram (Eds.), *Proceedings of the International Symposium on Monolingual and Bilingual Speech 2015* (pp. 57-67). Chania, Greece: Institute of Monolingual and Bilingual Speech.

Carney, A. E., Widin, G. P., & Viemeister, N. F. (1977). Noncategorical perception of stop consonants differing in VOT. *The Journal of the Acoustical Society of America, 62*(4), 961-970. doi:10.1121/1.381590

Carreiras, M., & Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex, 18*(7), 1727-1735. doi:10.1093/cercor/bhm202

Casillas, J. V., Díaz, Y., & Simonet, M. (2015). Acoustics of spanish and English coronal stops. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: Glasgow Universiy.

Catford, J. C., & Pisoni, D. B. (1970). Auditory vs. articulatory training in exotic sounds. *The Modern Language Journal*, *54*(7), 477. doi:10.2307/321767

Cebrian, J. (2000). Transferability and productivity of L1 rules in Catalan-English interlanguage. *Studies in Second Language Acquisition, 22*(01), 1-26. doi:10.1017/s0272263100001017

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

Cebrian, J. (2007). Old sounds in new contrasts: L2 production of the English tense-tax vowel distinction. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the 16th International Congress of Phonetic Sciences* (pp.1637-1640). Saarbrucken: University of Saarland.

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

Cebrian, J., & Carlet, A. (2014). Second-language learners' identification of target-language phonemes: A short-term phonetic training study. *Canadian Modern Language Review*, *70*(4), 474-499. doi:10.3138/cmlr.2318

Cebrian, J., Carlet, A. & Gavaldà, N. (Nov, 2016). *Incorporating perceptual training in an English phonetics course: Implementation, outcome and students' impressions*. Paper presented at TELLC: Teaching English language, literature and culture, held at the Universitat Autònoma de Barcelona, Barcelona, Spain.

Cebrian, J., Mora, J.C. & Aliaga-García C. (2011). Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Wrembel, M., Kul, M., Dziubalska-Kołaczyk, K. (Eds.), *Achievements and perspectives in the acquisition of second language speech: New Sounds 2010* (pp. 41-52). Frankfurt am Main: Peter Lang.

Celce-Murcia, M., Brinton, D., & Goodwin, J. (1996). *Teaching pronunciation*. Cambridge: Cambridge University Press.

Cenoz, J., & García-Lecumberri, M. L. (1999). The acquisition of English pronunciation: learners' views. *International Journal of Applied Linguistics, 9*(1), 3-15. doi:10.1111/j.1473-4192.1999.tb00157.x

Cerviño, E., & Mora, J. C. (2009). Duration as a phonetic cue in the categorization of /iː/-/ɪ/ and /s/-/z/ by spanish/Catalan learners of English. *The Journal of the Acoustical Society of America*, *125(4),* 2764-2764.doi.org/10.1121/1.4784683

Cho, T., & Ladefoged, P. (1999). Variation and universals in VOT: evidence from 18 languages. *Journal of Phonetics, 27*(2), 207-229. doi:10.1006/jpho.1999.0094

Christensen, L. A., & Humes, L. E. (1997). Identification of multidimensional stimuli containing speech cues and the effects of training. *The Journal of the Acoustical Society of America*, *102*(4), 2297-2310. doi:10.1121/1.419639

Collentine, J. (2009). Study abroad research: Findings, implications, and future directions. *The Handbook of Language Teaching*, 218-233. doi:10.1002/9781444315783.ch13

Collet, G., Colin, C., Serniclaes, W., Hoonhorst, I., Markessis, E., Deltenre, P., & Leybaert, J. (2013). Changes in voicing perception by adult French speakers after identification training. *Applied Psycholinguistics, 36*(02), 463-483. doi:10.1017/s0142716413000313

Cooke, M., García-Lecumberri, M.L., Maidment, J., & Ericsson, A. (2005). The web transcription tool. Retrieved February 18, 2017, from http://www.wtt.org.uk/.

Cortés Pomacóndor, S. (2003). Transfer in L2 sound production. In D. Recansens, M. J. Solé, & J. Romero (Eds.), *Proceedings of the 15th International Congress of Phonetic Sciences* (pp. 1097-1100). Barcelona: Causal Productions.

Cortés Pomacóndor, S. (2007). *Representations and transfer processes in L2 speech production: Evidence from Catalan learners of English.* (Unpublished doctoral dissertation). University of Edinburgh, Edinburgh, UK.

Council of Europe (2001). *The common European framework of reference for languages: learning, teaching, assessment*. Cambridge, UK: Cambridge University Press.

Creel, S. C., & Bregman, M. R. (2011). How talker identity relates to language processing. *Language and Linguistics Compass, 5*(5), 190-204. doi:10.1111/j.1749-818x.2011.00276.x

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297-334. doi:10.1007/bf02310555

Cruttenden, A. (2014). *Gimson's pronunciation of English*. Abingdon: Routledge.

Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. Lavelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011* (pp. 93-108). Ames, IA: Iowa State University.

Darcy, I., Mora, J. C., & Daidone, D. (2014). Attention control and inhibition influence phonological development in a second language. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (Vol. 5, pp. 115-129). Montreal: Concordia Working Papers in Applied Linguistics (COPAL).

Derwing, T. M. (2008). Curriculum issues in teaching pronunciation to second language learners. *Studies in Bilingualism Phonology and Second Language Acquisition*, 347-369. doi:10.1075/sibil.36.17der

Derwing, T. M. (2010). Utopian goals for pronunciation teaching. In *Proceedings of the 1st pronunciation in second language learning and teaching conference* (pp. 24-37). Iowa, USA: Iowa State University.

Derwing, T. M., & Munro, M. J. (2015). *Pronunciation fundamentals: evidence-based perspectives for L2 teaching and research* (Vol. 42). Netherlands: John Benjamins Publishing Company.

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

Earle, F. S., & Myers, E. B. (2013). Building phonetic categories: an argument for the role of sleep. *Frontiers in psychology, 5*, 1192-1192. doi:10.3389/fpsyg.2014.01192

Ellis, R. (2005). Principles of instructed language learning. *System, 33*(2), 209-224. doi:10.1016/j.system.2004.12.006

Engineer, C. T., Rahebi, K. C., Buell, E. P., Fink, M. K., & Kilgard, M. P. (2015). Speech training alters consonant and vowel responses in multiple auditory cortex fields. *Behavioural Brain Research, 287*, 256-264. doi:10.1016/j.bbr.2015.03.044

Escudero, P. (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.A. Thesis). University of Edinburgh, Edinburgh, UK.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition, 26*(04), 551-585. doi:10.1017/s0272263104040021

Escudero, P., Hayes-Harb, R., & Mitterer, H. (2008). Novel second-language words and asymmetric lexical access. *Journal of Phonetics, 36*(2), 345-360. doi:10.1016/j.wocn.2007.11.002

Evans, B., & Martín-Alvarez, L. (2016). *Age-related differences in second-language learning? A comparison of high and low variability perceptual training for the acquisition of English /ɪ/-/ i/ by Spanish adults and children.* Paper presented at New Sounds. International symposium on the acquisition of Second Language Speech, Aarhus University: Aarhus.

Everitt, C. (2015). *Accent Imitation on the L1 as a task to improve L2 pronunciation.* (Unpublished MA thesis). University of Barcelona: Barcelona, Spain.

Fabra, L. R. (2005). Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test. *VIAL, Vigo International Journal of Applied Linguistics, (2),* 75-92.

Feldman, N. H., Griffiths, T. L., Goldwater, S., & Morgan, J. L. (2013). A role for the developing lexicon in phonetic category acquisition. *Psychological Review, 120*(4), 751-778. doi:10.1037/a0034245

Ferrand, C. (2007). *Speech science: An integrated approach to theory and clinical experience*. New York: Pearson.

Flege, J. (1995). Second language speech learning: Theory, findings and problems. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 233-277). Timonium, MD: York Press.

Flege, J. E. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics, 16,* 425-442. doi:10.1017/S0142716400066029

Flege, J. E. (1984). The detection of french accent by american listeners. *The Journal of the Acoustical Society of America, 76*(3), 692-707. doi:10.1121/1.391256

Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/–/d/ contrast: Performance before and after training. *The Journal of the Acoustical Society of America, 86*(*5),* 1684-1697.

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

Flege, J. E. (1998). Second-language learning: The role of subject and phonetic variables. *In Proceedings of SpeechTechnology in Language Learning Conference* (pp.1-8). Marholmen, Sweden: ESCA.

Flege, J. E. (2003). Assessing constraints on second-language segmental production and perception. *Phonetics and Phonology in Language Comprehension and Production,6,* 319-355. doi:10.1515/9783110895094.319

Flege, J. E., & Davidian, R. D. (1984). Transfer and developmental processes in adult foreign language speech production. *Applied Psycholinguistics, 5*(04), 323. doi:10.1017/s014271640000521x

Flege, J. E., & Eefting, W. (1987). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication, 6*(3), 185-202. doi:10.1016/0167-6393(87)90025-2

Flege, J. E., & Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The Journal of the Acoustical Society of America, 76*(3), 708-721. doi:10.1121/1.391257

Flege, J. E., & Mackay, I. R. (2004). Perceiving vowels in a second language. *Studies in Second Language Acquisition, 26*(01), 1-34. doi:10.1017/s0272263104026117

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

Flege, J. E., Frieda, E. M., Walley, A. C., & Randazza, L. A. (1998). Lexical factors and segmental accuracy in second language speech production. *Studies in Second Language Acquisition, 20*(02), 155-187. doi:10.1017/s0272263198002034

Flege, J. E., Munro, M. J., & Fox, R. A. (1994). Auditory and categorical effects on cross-language vowel perception. *The Journal of the Acoustical Society of America, 95*(6), 3623-3641. doi:10.1121/1.409931

Flege, J. E., Munro, M. J., & Mackay, I. R. (1995). Factors affecting strength of perceived foreign accent in a second language. *The Journal of the Acoustical Society of America, 97*(5), 3125-3134. doi:10.1121/1.413041

Flege, J. E., Munro, M. J., & Skelton, L. (1992). Production of the word-final English /t/-/d/contrast by native speakers of English, Mandarin, and Spanish. *The Journal of the Acoustical Society of America*, *92(1),* 128-143.

Flege, J.E. (2016). *The role of phonetic category formation in second language speech acquisition.* Paper presented at New Sounds-International symposium on the acquisition of Second Language Speech, Aarhus University: Aarhus.

Fowler, C. A. (1986). An event approach study of speech perception from a directrealist perspective. *Journal of Phonetics*, *14*, 3-28. doi:10.1177/002383098803100302

Fowler, C. A. (1996). Listeners do hear sounds, not tongues. *The Journal of the Acoustical Society of America, 99*(3), 1730-1741. doi:10.1121/1.415237

Fox, R. A., Flege, J. E., & Munro, M. J. (1995). The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis.

The Journal of the Acoustical Society of America, 97(4), 2540-2551. doi:10.1121/1.411974

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech, 5(4),* 171-189. doi:10.1177/002383096200500401

Fullana, N., & Mora, J. C. (2009) Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production.* (pp. 97-117). Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Fullana, N., & MacKay, I. R. (2003). Production of English sounds by EFL learners: The case of /i/ and /ɪ/. In *Proceedings of the 15th international congress of phonetic sciences*, ICPhS (pp. 1525-1528). Barcelona, Spain: ICPhS organizing comittee.

Fullana, N., & MacKay, I. R. (2008). 'I said made (mate?)': Catalan/Spanish bilinguals' production of English word-final obstruents. *Canadian Acoustics*, *36(3),* 118-119.

García-Lecumberri, M. L., & Gallardo, F. (2003). English FL sounds in school learners of different ages. In M. P. García Mayo & M. L. García-Lecumberri (Eds.), *Age and the Acquisition of English as a Foreign Language* (pp.115-135). Clevedon: Multilingual Matters.

Gibson, J. J., & Gibson, E. J. (1955). What is learned in perceptual learning? A reply to Professor Postman. *Psychological Review*, *62*(6), 447-450. doi:10.1037/h0048196

Gonet, W., & Święciński, R. (2012). More on the voicing of English obstruents: voicing retention vs. voicing loss. *Research in Language, 10*(2), 183-199. doi:10.2478/v10015-011-0035-8

Gordon, J., & Darcy, I. (2016). The development of comprehensible speech in L2 learners: A classroom study on the effects of short-term pronunciation instruction. *Journal of Second Language Pronunciation*, *2*(1), 56-92. doi:10.1075/jslp.2.1.03gor

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics*, *28*(4), 267-283. doi:10.3758/bf03204386

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. *Language Experience in Second Language Speech Learning Language Learning & Language Teaching*, 57-77. doi:10.1075/lllt.17.09gui

Hacquard, V., Walter, M. A., & Marantz, A. (2007). The effects of inventory on vowel perception in French and Spanish: An MEG study. *Brain and Language*, *100*(3), 295-300. doi:10.1016/j.bandl.2006.04.009

Hardison, D. M. (1999). Bimodal speech perception by native and nonnative speakers of English: Factors influencing the McGurk effect. *Language Learning*, *49*, 213-283. doi:10.1111/0023-8333.49.s1.7

Hardison, D. M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(04), 495-522. doi:10.1017/s0142716403000250

Hardison, D. M. (2004). Generalization of computer-assisted prosody training: Quantitative and qualitative findings. *Language Learning & Technology*, *8(1)*, 34-52.

Hardison, D. M. (2005). Second-language spoken word identification: Effects of perceptual training, visual cues, and phonetic environment. *Applied Psycholinguistics*, *26*(04), 579-602. doi:10.1017/s0142716405050319

Hardison, D. M. (2014). Changes in second-language learners' oral skills and socio-affective profiles following study abroad: A mixed-methods approach. *Canadian Modern Language Review*, *70*(4), 415-444. doi:10.3138/cmlr.2202

Harris, D. P. (1969). *Testing English as a second language*. New York: McGraw-Hill Book Co.

Hattori, K., & Iverson, P. (2008). English/r/-/l/ pronunciation training for Japanese speakers. *The Journal of the Acoustical Society of America*, *123(5),* 3327-3327. doi:10.1121/1.3021295

Hattori, K., & Iverson, P. (2009). English /r/-/l/ category assimilation by Japanese adults: Individual differences and the link to identification accuracy. *The Journal of the Acoustical Society of America*, *125*(1), 469-479. doi:10.1121/1.3021295

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

Hazan, V., Sennema, A., Iba, M., & Faulkner, A. (2005). Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication*, *47*(3), 360-378. doi:10.1016/j.specom.2005.04.007

Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel-related differences in laryngeal articulatory and phonatory function. *Journal of Speech Language and Hearing Research, 41*(4), 712-724. doi:10.1044/jslhr.4104.712

Hillenbrand, J. M., Clark, M. J., & Houde, R. A. (2000). Some effects of duration on vowel recognition. *The Journal of the Acoustical Society of America*, *108*(6), 3013-3022. doi:10.1121/1.1323463

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical society of America*, *97(5),* 3099-3111. doi:10.1121/1.409456

Hirata, Y., & Kelly, S. D. (2010). Effects of lips and hands on auditory learning of second-language speech sounds. *Journal of Speech, Language, and Hearing Research, 53*(2), 298-310. doi:10.1044/1092-4388(2009/08-0243)

Hodson, B. W., & Paden, E. P. (1983). *Targeting intelligible speech: A phonological approach to remediation*. San Diego, CA: College-Hill Press.

Højen, A., & Flege, J. E. (2006). Early learners' discrimination of second-language vowels. *The Journal of the Acoustical Society of America, 119*(5), 3072-3084. doi:10.1121/1.2184289

Hwang, H., & Lee, H. Y. (2015). The effect of high variability phonetic training on the production of English vowels and consonants. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences* (pp. 1041-1). Glasgow, UK: Glasgow University.

IBM Corp. Released 2013. *IBM SPSS Statistics for Windows*, Version 22.0. Armonk, NY: IBM Corp.

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

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America, 122*(5), 2842-2854. doi:10.1121/1.2783198

Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America, 126*(2), 866-877. doi:10.1121/1.3148196

Iverson, P., & Kuhl, P. K. (1995). Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. *The Journal of the Acoustical Society of America, 97*(1), 553-562. doi:10.1121/1.412280

Iverson, P., & Kuhl, P. K. (1996). Influences of phonetic identification and category goodness on American listeners' perception of /r/ and /l/. *The Journal of the Acoustical Society of America, 99*(2), 1130-1140. doi:10.1121/1.415234

Iverson, P., Ekanayake, D., Hamann, S., Sennema, A., & Evans, B. G. (2008). Category and perceptual interference in second-language phoneme learning: An examination of English /w/-/v/ learning by Sinhala, German, and Dutch speakers. *Journal of Experimental Psychology: Human Perception and Performance, 34*(5), 1305-1316. doi:10.1037/0096-1523.34.5.1305

Iverson, P., Hazan, V., & Bannister, K. (2005). Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English /r/-/l/ to Japanese adults. *The Journal of the Acoustical Society of America, 118*(5), 3267-3278. doi:10.1121/1.2062307

Iverson, P., Kuhl, P. K., Akahane-Yamada, R., Diesch, E., Tohkura, Y., Kettermann, A., & Siebert, C. (2003). A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition, 87*(1), B47-B57. doi:10.1016/s0010-0277(02)00198-1

Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics, 33*(01), 145-160. doi:10.1017/s0142716411000300

Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics, 40*(4), 205-215. doi:10.3758/bf03211500

Jang, M. (2014). Perception of English Consonants in Different Prosodic Positions by Korean Learners of English. *Phonetics and Speech Sciences, 6*(1), 11-19. doi:10.13064/ksss.2014.6.1.011

Kartushina, N., Hervais-Adelman, A., Frauenfelder, U. H., & Golestani, N. (2015). The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America, 138*(2), 817-832. doi:10.1121/1.4926561

Kent, R. D., Read, C., & Kent, R. D. (1992). *The acoustic analysis of speech (Vol. 58).* San Diego: Singular Publishing Group.

Kissling, E. M. (2013). Teaching pronunciation: Is explicit phonetics instruction beneficial for FL learners? *The Modern Language Journal, 97*(3), 720-744. doi:10.1111/j.1540-4781.2013.12029.x

Kivistö-de Souza, H. (2016). *Phonological Awareness and Pronunciation in a Second Language.* (Unpublished doctoral dissertation). Universitat de Barcelona, Barcelona, Barcelona, Spain.

Kivistö-de Souza, H., & Carlet, A. (2014). Vowel inventory size and the use of temporal cues in non-native vowel perception by Catalan and Danish EFL learners. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (Vol. 5, pp. 322-336). Montreal: Concordia Working Papers in Applied Linguistics (COPAL).

Kluge, D. C., Rauber, A. S., Reis, M. S., & Bion, R. A. H. (2007). The relationship between the perception and production of English nasal codas by Brazilian learners of English. In *Interspeech* (pp. 2297-2300). Antwerp, Belgium: ISCA.

Kuhl, P. K. (1993). Early linguistic experience and phonetic perception: Implications for theories of developmental speech perception. *Journal of Phonetics*, 21, 125-139.

Kuhl, P. K. (2000). A new view of language acquisition. *Proceedings of the National Academy of Sciences, 97(22),* 11850-11857.

Kuhl, P., & Iverson, P. (1995). Linguistic experience and the "Perceptual Magnet Effect". In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 121-154). Timonium, MD: York Press.

Kushch, O., Gluhareva, D., Llanes, J. & Prieto, P. (2016). *The role of beat gesture perception and production in L2 pronunciation training.* Paper presented at New Sounds. International symposium on the acquisition of Second Language Speech, Aarhus University, Aarhus.

Labov, W. (1972). "The social stratification of (r) in New York city department stores". In *Sociolinguistics Patterns* (pp. 43-54). Philadelphia: University of Pennsylvania Press.

Lacabex, E. G., & García-Lecumberri, M.L. (2010). Investigating training effects in the production of English weak forms by Spanish learners. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 137-43). Poznań, Poland: Adam Mickiewicz University press.

Lacabex, E. G., García-Lecumberri, M. L., & Cooke, M. (2008). Identification of the contrast full vowel-schwa: training effects and generalization to a new perceptual context. *Ilha do Desterro A Journal of English Language, Literatures in English and Cultural Studies*, (55). 173-196. doi:10.5007/2175-8026.2008n55p173

Ladefoged, P. (1962). Subglottal activity during speech. In *Proceedings of the Fourth International Congress of Phonetic Sciences* (pp. 73-91). The Hague: Mouton.

Ladefoged, P. (1971). *Preliminaries to linguistic phonetics*. Chicago, USA: University of Chicago Press.

Ladefoged, P. (2006). Representing linguistic phonetic structure [unfinished manuscript], Retrieved on January 31, 2012 from http://www. linguistics. ucla. edu/people/ladefoge/PhoneticStructure. pdf

Lambacher, S. G., Martens, W. L., Kakehi, K., Marasinghe, C. A., & Molholt, G. (2005). The effects of identification training on the identification and production of American English vowels by native speakers of Japanese. *Applied Psycholinguistics, 26*(02), 227-247. doi:10.1017/s0142716405050150

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi:10.2307/2529310

Larson-Hall, J. (2008). Weighing the benefits of studying a foreign language at a younger starting age in a minimal input situation. *Second Language Research, 24*(1), 35-63. doi:10.1177/0267658307082981

Lee, H., & Hwang, H. (2016). Gradient of learnability in teaching English pronunciation to Korean learners. *The Journal of the Acoustical Society of America, 139*(4), 1859-1872. doi:10.1121/1.4945716

Lengeris, A. (2008). The effectiveness of auditory phonetic training on Greek native speakers's perception and production of southern British English vowels. In *Proceedings of ISCA Tutorial and Research Workshop on Experimental Linguistics, ExLing 2008* (pp. 133-136). Athens: ISCA.

Leow, R. P. (2012). Attention, awareness, and noticing in second language acquisition. *The encyclopedia of applied linguistics* (pp. 1-7). Oxford, UK: Wiley-Blackwell.

Lerdpaisalwong, S. (2015). *Perception training of Thai learners: American English consonants and vowels*. (Unpublished doctoral dissertation). The University Of Wisconsin-Milwaukee, Wisconsin, USA.

Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly, 39*(3), 369-377. doi:10.2307/3588485

Levis, J. M., & Grant, L. (2003). Integrating pronunciation into ESL/EFL classrooms. *TESOL Journal, 12*(2), 13-19. doi:10.1002/j.1949-3533.2003.tb00125.x

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review, 74(6),* 431-461.

Liberman, M. (2008). HVPT /blog: filed under Language teaching and learning/. Retrieved on January 5, 2017 from http://languagelog.ldc.upenn.edu/nll/?p=328

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: acoustical measurements. *WORD, 20*(3), 384-422. doi:10.1080/00437956.1964.11659830

Lively, S. E., Logan, J. S., & Pisoni, D. B. (1993). Training Japanese listeners to identify English /r/ and /l/. II: The role of phonetic environment and talker variability in learning new perceptual categories. *The Journal of the Acoustical Society of America, 94*(3), 1242-1255. doi:10.1121/1.408177

Lively, S. E., Pisoni, D. B., Yamada, R. A., Tohkura, Y., & Yamada, T. (1994). Training Japanese listeners to identify English /r/ and /l/. III. Long-term retention of new phonetic categories. *The Journal of the Acoustical Society of America, 96*(4), 2076-2087. doi:10.1121/1.410149

Llisterri, J. (1995). Relationships between speech production and speech perception in a second language. In *Proceedings of the XIIIth International Congress of Phonetic Sciences* , ICPhS (Vol. 4, pp. 92-99). Stockholm, Sweden: Stockholm University.

Logan, J. S., Lively, S. E., & Pisoni, D. B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America, 89*(2), 874-886. doi:10.1121/1.1894649

Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 351-378). Timonium, MD:York Press.

Long, M. (1991). Focus on Form. *Studies in Bilingualism Foreign Language Research in Cross-Cultural Perspective,* 39-52. doi:10.1075/sibil.2.07lon

Long, M. H., & Larsen-Freeman, D. (1991). *An introduction to second language acquisition research.* Harlow: Longman.

MacKay, D. G. (1987). Perceptual sequencing and higher level activation. In *The Organization of Perception and Action* (pp. 62-89). New York: Springer.

Mackey, A., & Gass, S. (2005). *Second Language Research: Methodology and Design*. New York and London: Routledge.

Martínez-Flor, A., & Usó-Juan, E. (2006). A comprehensive pedagogical framework to develop pragmatics in the foreign language classroom: The 6Rs approach. *Applied Language Learning*, *16(2),* 39-63.

Mascaró, J. (1991). Iberian spirantization and continuant spreading. *Catalan working papers in linguistics, 1*, 167-179.

Mccandliss, B. D., Fiez, J. A., Protopapas, A., Conway, M., & Mcclelland, J. L. (2002). Success and failure in teaching the [r]-[l] contrast to Japanese adults: Tests of a Hebbian model of plasticity and stabilization in spoken language perception. *Cognitive, Affective, & Behavioral Neuroscience, 2*(2), 89-108. doi:10.3758/cabn.2.2.89

Mcclaskey, C. L., Pisoni, D. B., & Carrell, T. D. (1983). Transfer of training of a new linguistic contrast in voicing. *Perception & Psychophysics, 34*(4), 323-330. doi:10.3758/bf03203044

Mclaughlin, M. W. (1987). Learning from experience: Lessons from policy implementation. *Educational Evaluation and Policy Analysis, 9*(2), 171-178. doi:10.2307/1163728

Monje-Sangüesa (2016). *The second time around: The effect of FI upon return from SA.* (Unpublished MA Thesis). University Pompeu Fabra: Barcelona, Spain.

Mora, J. C. (2008). Learning context effects on the acquisition of a second language phonology. *A portrait of the young in the new multilingual Spain*, 241-263.

Mora, J. C. (2014). Inter-subject variation in L2 speech perception and cognitive abilities. In Casesnoves, R., Forcadell, M. & Gavaldà, N. (Eds.), Ensqueda la paraula. Estudis de lingüística aplicada en honor de M. Teresa Turell (pp.83-101). Barcelona, Institut Universitari de Lingüística Aplicada.

Mora, J. C., & Fullana, N. (2007). Production and perception of English /i/-/ɪ/ and /æ/-/ʌ/ in a Formal setting: Investigating the effects of experience and starting age. In *Proceedings of the 16th international congress of phonetic sciences* (Vol. 3, pp. 1613-1616). Saarbrücken: Universitat de Saarlandes.

Morley, J. (1991). The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly, 25*(3), 481-520. doi:10.2307/3586981

Morrison, G. S. (2009). L1-Spanish speakers' acquisition of the English /ɪ/-/i/ contrast II: Perception of vowel inherent spectral change. *Language and Speech*, *52(4)*, 437-462. doi:10.1177/0023830909336583

Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, *29*(4), 578-596. doi:10.1093/applin/amm056

Muñoz, C. (2011). Input and long-term effects of starting age in foreign language learning. *IRAL - International Review of Applied Linguistics in Language Teaching, 49*(2), 113-133. doi:10.1515/iral.2011.006

Munro, M. J. (2008). Foreign accent and speech intelligibility. *Studies in Bilingualism Phonology and Second Language Acquisition*, 193-218. doi:10.1075/sibil.36.10mun

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning, 58*(3), 479-502. doi:10.1111/j.1467-9922.2008.00448.x

Murphy, J. (2014). Teacher training programs provide adequate preparation in how to teach pronunciation. *Pronunciation myths: Applying second language research to classroom teaching*, 188-224.

Nespor, M., Peña, M., & Mehler, J. (2003). On the different roles of vowels and consonants in speech processing and language acquisition. *Lingue e linguaggio*, *2(2)*, 203-230.

Neumann, O. (1996). *Handbook of perception and action: attention*. London: Academic Press.

Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech Language and Hearing Research, 50*(6), 1496-1509. doi:10.1044/1092-4388(2007/103)

Nishi, K., & Kewley-Port, D. (2008). Nonnative speech perception training using vowel subsets: Effects of vowels in sets and order of training. *Journal of Speech Language and Hearing Research, 51*(6), 1480-1493. doi:10.1044/1092-4388(2008/07-0109)

Nobre-Oliveira, D. (2007). Effects of perceptual training on the learning of English vowels in non-native settings. In *Proceedings of the 5th International symposium on the acquisition of second language speech, New Sounds* (Vol. 5, pp. 382-389). Florianopolis: Federal University of Santa Catarina.

Nozawa, T. (2015). Effects of training methods and attention on the identification and discrimination of American English coda nasals by native Japanese listeners. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: the University of Glasgow. ISBN 978-0-85261-941-4

Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America, 127*(2), EL54-EL59. doi:10.1121/1.3292286

Pereira, Y. I. (2014). *Perception and production of English vowels by Chilean learners of English: Effect of auditory and visual modalities on phonetic training* (Unpublished doctoral dissertation). University College London, London, UK.

Pérez-Vidal, C., Juan-Garau, M., & Mora, J. C. (2011). The effects of formal instruction and study abroad contexts on foreign language development: The SALA project. In C. Sanz, & R. P. Leow (Eds.), *Implicit and explicit conditions, processes and knowledge in SLA and bilingualism* (pp. 115-138). Washington D. C.: Georgetown University Press.

Perrachione, T. K., Lee, J., Ha, L. Y., & Wong, P. C. (2011). Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America, 130*(1), 461-472. doi:10.1121/1.3593366

Petersen, S. E., & Posner, M. I. (2012). The attention system of the human brain: 20 years after. *Annual Review of Neuroscience, 35*(1), 73-89. doi:10.1146/annurev-neuro-062111-150525

Peterson, G. E., & Barney, H. L. (1951). Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America, 24(2),* 175-184. doi:10.1121/1.1917300

Pierce, L. (2014). *Learning novel vowel contrasts: experimental methods in classroom applications* (Unpublished doctoral dissertation). University of Illinois at Urbana-Champaign: Urbana-Champaign, USA.

Piske, T. (2008). Phonetic awareness, phonetic sensitivity and the second language learner. In *Encyclopedia of language and education* (pp. 1912-1923). New York, US: Springer Publishing.

Piske, T., Flege, J. E., Mackay, I. R., & Meador, D. (2010). Investigating native and non-native vowels produced in conversational speech. Achievements and perspectives in the acquisition of second language speech. In *Proceedings of the 6th International Symposium on the Acquisition of Second Language Speech, New Sounds 2010* (pp. 195-205). Switzerland: Peter Lang.

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

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics, 13*(2), 253-260. doi:10.3758/bf03214136

Pisoni, D. B., Aslin, R. N., Perey, A. J., & Hennessy, B. L. (1982). Some effects of laboratory training on identification and discrimination of voicing contrasts in stop consonants. *Journal of Experimental Psychology: Human Perception and Performance, 8*(2), 297-314. doi:10.1037//0096-1523.8.2.297

Pisoni, D. B., Lively, S. E., & Logan, J. S. (1994). Perceptual learning of non-native speech contrasts: Implications for theories of speech perception. In: Nusbaum H, Goodman J (Eds.), *Development of Speech Perception: The Transition from Recognizing Speech Sounds to Spoken Words* (pp. 121-166). Cambridge: MIT.

Pisoni, D., & Lively, S. (1995). Variability and invariance in speech perception: A new look at some old problems in perceptual learning. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp.433-459). Timonium, MD: York Press.

Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics, 52*(1), 37-52. doi:10.3758/bf03206758

Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, *32*(1), 3-25. doi:10.1093/acprof:oso/9780199791217.003.0003

Proceedings of the annual conference of the Canadian Acoustics Association, Canadian Acoustics, 37(3), 138-139.

Protopapas, A., & Calhoun, B. (2000). Adaptive phonetic training for second language learners. In *Proceedings of the 2nd International Workshop on Integrating Speech Technology in Language Learning* (pp. 31-38). Dundee: Dundee Univesity.

Rallo-Fabra, L., & Romero, J. (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics, 40(3),* 491-508. doi:10.1016/j.wocn.2012.01.001

Raphael, L. J. (1971). An electromyographic investigation of the tense-lax feature in some English vowels. *Haskins Laboratories Status Report on Speech Research SR, 25(26),* 131-140.

Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America, 51*(4B), 1296-1303. doi:10.1121/1.1912974

Rato, A. A. S. (2014). *Cross-language perception and production of English vowels by Portuguese learners: the effects of perceptual training.* (Unpublished doctoral dissertation). Universidade do Minho, Braga, Portugal.

Rato, A., & Rauber, A. (2015). The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In the Scottish Consortium for ICPhS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. Glasgow, UK: Glasgow University.

Rauber, A., Rato, A., Kluge, D., & Santos, G. (2012). TP (Version 3.1).[Software]. *Brazil: Worken. [http://www.worken.com.br/tp_regfree.php?l=i]*.

Recasens, D. (1991). *Fonètica descriptiva del Català: Assaig de caracterització de la pronúncia del vocalisme i consonantisme del català al segle XX.* Barcelona: Institut d'estudis Catalans.

Recasens, D. (1993). *Fonètica i fonologia*. Barcelona:Enciclopèdia Catalana.

Recasens, D., & Mira, M. (2015). Place and manner assimilation in Catalan consonant clusters. *Journal of the International Phonetic Association, 45*(02), 115-147. doi:10.1017/s0025100315000080

Redford, M. A., & Diehl, R. L. (1999). The relative perceptual distinctiveness of initial and final consonants in CVC syllables. *The Journal of the Acoustical Society of America, 106*(3), 1555-1565. doi:10.1121/1.427152

Reis, M., & Nobre-Oliveira, D. (2007). Effects of perceptual training on the identification and production of English voiceless plosives aspiration by Brazilian EFL learners. In *Proceedings of the 5th International symposium on the acquisition of second language speech, New Sounds* (Vol. 5, pp. 398-407). Florianopolis: UFPR.

Roach, P. (2000). *Phonetics and phonology: A Practical course*. Cambridge, UK: Cambridge Press.

Roach, P. (2004). British English: Received pronunciation. *Journal of the International Phonetic Association*, *34*(2), 239-245. doi:10.1017/s0025100304001768

Rochet, B. (1995). Perception and production of second-language speech sounds by adults. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 379-410). Timonium, MD: York Press.

Safronova, E., & Mora, J. C. (2012). Acoustic and phonological memory in L2 vowel perception. In *Proceedings of 22nd EUROSLA* (pp.384-390). Poznań, Poland: Adam Mickiewicz University press.

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

Saito, K. (2015). Variables affecting the effects of recasts on L2 pronunciation development. *Language Teaching Research*, *19*(3), 276-300. doi:10.1177/1362168814541753

Saito, K., & Hanzawa, K. (2015). Developing second language oral ability in foreign language classrooms: The role of the length and focus of instruction and individual differences. *Applied Psycholinguistics, 37*(04), 813-840. doi:10.1017/s0142716415000259

Saito, K., & Lyster, R. (2011). Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning, 62*(2), 595-633. doi:10.1111/j.1467-9922.2011.00639.x

Schiller, N. O. (1998). The effect of visually masked syllable primes on the naming latencies of words and pictures. *Journal of Memory and Language*, *39*(3), 484-507. doi:10.1006/jmla.1998.2577

Schmidt, R. (*2001*). “Attention”. In *P. Robinson* (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge: Cambridge University Press.

Schmidt, R. W. (1990). The role of consciousness in second language learning. *Applied Linguistics, 11*(2), 129-158. doi:10.1093/applin/11.2.129

Schwab, S., & Llisterri, J. (2011). Are French speakers able to learn to perceive lexical stress contrasts. In the Consortium for ICPhS 2011 (Ed.). *Proceedings of the 17th International Congress of Phonetic Sciences* (pp. 1774-7). Hong Kong, China: City University of Hong Kong.

Schwab, S., & Llisterri, J. (2014). Does training make French speakers more able to identify lexical stress? In *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (Vol. 5, pp. 624–636). Montreal: Concordia Working Papers in Applied Linguistics (COPAL).

Schwartzhaupt, B., Alves, U. K., & Ana Beatriz Arêas Da Luz Fontes. (2015). The role of L1 knowledge on L2 speech perception: investigating how native speakers and Brazilian learners categorize different VOT patterns in English. *Revista de Estudos da Linguagem, 23*(2), 311-334. doi:10.17851/2237-2083.23.2.311-334

Sheldon, A., & Strange, W. (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics, 3*(03), 243-261. doi:10.1017/s0142716400001417

Smith, B. L. (1975). Effects of vocalic context, place of articulation, and speaker's sex on “voiced” stop consonant production. *The Journal of the Acoustical Society of America, 58*(S1), S61-S61. doi:10.1121/1.2002223

Smith, M. S. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition, 15*(02), 165-179. doi:10.1017/s0272263100011943

Solé, M. J. (1991). Stress and rhythm in English. In *Revista Alicantina de Estudios Ingleses*: Vol. 4 (pp. 145-162). Alicante: Gráficas Díaz.

Solé, M. J., & Estebas, E. (2000). Phonetic and phonological phenomena: VOT: A cross-language comparison. In *Proceedings of the 18th AEDEAN Conference* (pp. 437-44). Vigo: University of Vigo.

Solé, M.J. (2013). Phonological vs lexical categories in an L2. In *Proceedings of the 6th Phonetics and Phonology in Iberia Conference* (pp.58-59). Lisbon, Portugal: Lisbon University.

Steinlen, A. K. (2005). *The influence of consonants on native and non-native vowel production: a cross-linguistic study* (Vol. 30). Tübingen: Gunter Narr.

Strange, W. (1992). Learning non-native phoneme contrasts: Interactions among subject, stimulus, and task variables. In Tohkura, E., Vatikiotis-Bateson, E., & Sagisaka, Y (Eds.) *Speech perception, production and linguistic structure* (pp.197-219). Tokyo, Japan: Ohmsa.

Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 35-55). Amsterdam/Philadelphia: John Benjamins Publishing Company.

Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-l/ by Japanese adults learning English. *Perception & Psychophysics, 36*(2), 131-145. doi:10.3758/bf03202673

Swingley, D. (2009). Onsets and codas in 1.5-year-olds' word recognition. *Journal of Memory and Language, 60*(2), 252-269. doi:10.1016/j.jml.2008.11.003

Thomson, R. I. (2008). Language experience in second language speech learning: In honor of James Emil Flege (review). *The Canadian Modern Language Review/La revue canadienne des languesvivantes*, *65(2)*, 344-346.

Thomson, R. I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal, 28*(3), 744-765. doi:10.11139/cj.28.3.744-765

Thomson, R. I. (2012). Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning, 62*(4), 1231-1258. doi:10.1111/j.1467-9922.2012.00724.x

Thomson, R. I., & Campagna, P. (2010). The effect of task type on native speaker judgments of L2 accented speech. *Proceedings of the annual conference of the Canadian Acoustics Association, 38(3)*, 122-123.

Thomson, R. I., & Derwing, T. M. (2014). The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, *36*(3), 326-344. doi:10.1093/applin/amu076

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-97). Ames, IA: Iowa State University.

Thomson, R. I., & Isaacs, T. (2009). Within-category variation in L2 English vowel learning.

Trapp, N. L., & Bohn, O. S. (2000). Training Danish listeners to identify English word-final /s/ and /z/. In *New Sounds 2000: Proceedings of the Fourth International Symposium on the Acquisition of Second-language Speech*. Klagenfurt am Wörthersee, Austria: University of Klagenfurt.

VanPatten, B. (1990). Attending to form and content in the input. *Studies in Second Language Acquisition, 12*(03), 287-301. doi:10.1017/s0272263100009177

VanPatten, B. (1996). *Input processing and grammar instruction in second language acquisition*. New York: Ablex Publishing.

VanPatten, B. (2002). Processing the content of input–processing and processing instruction research: a response to Dekeyser, Salaberry, Robinson, and Harrington. *Language Learning*, *52(4)*, 825-831. doi:10.1111/1467-9922.00205

Wager, T. D., Jonides, J., & Smith, E. E. (2006). Individual differences in multiple types of shifting attention. *Memory & Cognition, 34*(8), 1730-1743. doi:10.3758/bf03195934

Wang, H., Ho, T., Yang, R., Shen, J., Bai, B., Hong, J., Lee, L. (1997). Complete recognition of continuous Mandarin speech for Chinese language with very large vocabulary using limited training data. *IEEE Transactions on Speech and Audio Processing, 5*(2), 195-200. doi:10.1109/89.554782

Wang, X. (2002). *Training Mandarin and Cantonese speakers to identify English vowel contrasts: long-term retention and effects on production.* (Unpublished doctoral dissertation). Simon Fraser University, Burnaby, Canada.

Wang, X., & Munro, M. J. (1999). The perception of English tense-lax vowel pairs by native Mandarin speakers: The effect of training on attention to temporal and spectral cues. In *Proceedings of the 14th international congress of phonetic sciences* (Vol. 3, pp. 125-128). Berkeley, CA: University of California.

Wang, X., & Munro, M. J. (2004). Computer-based training for learning English vowel contrasts. *System*, *32*(4), 539-552. doi:10.1016/j.system.2004.09.011

Wang, Y., Behne, D. M., & Jiang, H. (2008). Linguistic experience and audio-visual perception of non-native fricatives. *The Journal of the Acoustical Society of America, 124*(3), 1716-1726. doi:10.1121/1.2956483

Wang, Y., Jongman, A., & Sereno, J. A. (1998). Training American listeners to perceive and produce Mandarin tones. *The Journal of the Acoustical Society of America, 103*(5), 3090-3090. doi:10.1121/1.422940

Wang, Y., Jongman, A., & Sereno, J. A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America, 113*(2), 1033-1043. doi:10.1121/1.1531176

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics, 36*(2), 250-267. doi:10.1016/j.wocn.2007.06.004

Werker, J. F., & Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *The Journal of the Acoustical Society of America, 75*(6), 1866-1878. doi:10.1121/1.390988

Wong, J. W. S. (2014). The effects of high and low variability phonetic training on the perception and production of English vowels/e/-/æ/by Cantonese ESL learners with high and low L2 proficiency levels. In *Interspeech* (pp. 524-528). Singapore: ISCA.

Yamada, R. A., & Tohkura, Y. (1992). The effects of experimental variables on the perception of American English /r/ and /l/ by Japanese listeners. *Perception & Psychophysics, 52*(4), 376-392. doi:10.3758/bf03206698

Yamada, R.A, Tohkura, Y. I., Bradlow, A. R., & Pisoni, D. B. (1996). Does training in speech perception modify speech production? In *Proceedings of the Fourth International Conference on Spoken Language, ICSLP 96* (Vol. 2, pp. 606-609). Philadelphia, PA: IEEE.

Yavaş, M. (2002). Voice onset time patterns in bilingual phonological development. *Investigations in clinical phonetics and linguistics*, 36, 341-349.

Yavaş, M. (2007). Factors influencing the VOT of English long lag stops and interlanguage phonology. In *Proceedings of the 5th International symposium on the acquisition of second language speech, New Sounds* (Vol. 5, pp. 492-498). Florianopolis: Federal University of Santa Catarina.

Yavaş, M., & Wildermuth, R. (2006). The effects of place of articulation and vowel height in the acquisition of English aspirated stops by Spanish speakers. *IRAL – International*

Review of Applied Linguistics in Language Teaching, 44(3), 251-263. doi:10.1515/iral.2006.011

Ziegler, J. C., Besson, M., Jacobs, A. M., Nazir, T. A., & Carr, T. H. (1997). Word, pseudoword, and nonword processing: A multitask comparison using event-related brain potentials. *Journal of Cognitive Neuroscience, 9*(6), 758-775. doi:10.1162/jocn.1997.9.6.758

Appendix A - Consent form and privacy statement

UAB
Universitat Autònoma de Barcelona

Consent form and privacy statement

Name of the participant: ___________________________

Participant's code _________________________________

The results of this experiment, as well as your name and other personal information in the questionnaire will be treated anonymously and will not be distributed. Personal data will be used only for statistical purposes. Any subsequent publication will not make reference to individual personal information.

Thank you for your participation. In addition to getting credit for assignment 1, you are helping researchers understand the process of second language acquisition better.

Participant's statement

I have been informed of the procedure of the experiment; I understand that it will be performed over 10 sessions of approximately 30-35 minutes. I understand and agree that the completion of this experiment corresponds to "assignment 1's grade".

I have read the privacy statement and agree to participate in this study. I understand that the data collected, both personal data and the results of the experiment will be treated anonymously.

Signature and date: _____________________________________

Appendix B - Language background questionnaire

Linguistic background questionnaire

Participant's code………………………………………………………………………………………

Last name: .... Name:............................ NIA.....................

Email: ....Group/teacher...

A. General Information

1. Sex: F M
2. Age: ____
3. Do you have vision or hearing problems? ______________________________
4. Are you repeating this subject? ☐Yes ☐No
5. Major: ☐Grau ☐Grau combinat, specify language:_____________

B. Language Background

1. Your native language_______________________ Place of birth_____________________
2. Mother's native language:____________________ Place of birth ____________________
3. Father's native language:_____________________Place of birth _____________________
4. Are you Spanish/Catalan bilingual (learnt both languages before the age of 6)? ☐Yes ☐No

1.Do you regard yourself as:

A) Dominant in Spanish, because__________________
B) Dominant in Catalan, because__________________
C) I'm not Sp/Cat bilingual

2. What language do you use at home?

A) Catalan & Spanish
B) Catalan
C) Spanish
D) Other (specify which :) ___________________________

3. What is your second language (=learnt after the age of 5)?

A) English
B) Spanish
C) Catalan
D) Other (specify which :) ________________________

4. What is your third language (=learnt after you second language)?

A) English
B) Spanish
C) Catalan
D) Other (specify which :) ________________________
E) I don't have a third language

5. I consider myself fluent in the following languages:

A) Catalan
B) Spanish
C) English
D) Other (specify which :) ______________________________

6. Estimate the % of daily use of the following languages (leave blank if it does not apply, e.g., leave "At work" blank if you do not have a job).

a) Catalan:

At work (colleagues):	Never	5-25%	26-50%	51-75%	76-95%	Always
At university (classmates/teachers)	Never	5-25%	26-50%	51-75%	76-95%	Always
At home (family):	Never	5-25%	26-50%	51-75%	76-95%	Always
Socially (friends):	Never	5-25%	26-50%	51-75%	76-95%	Always

b) Spanish:

At work (colleagues):	Never	5-25%	26-50%	51-75%	76-95%	Always
At university (classmates/teachers)	Never	5-25%	26-50%	51-75%	76-95%	Always
At home (family):	Never	5-25%	26-50%	51-75%	76-95%	Always
Socially (friends):	Never	5-25%	26-50%	51-75%	76-95%	Always

English:

At work (colleagues):	Never	5-25%	26-50%	51-75%	76-95%	Always
At university (classmates/teachers)	Never	5-25%	26-50%	51-75%	76-95%	Always
At home (family):	Never	5-25%	26-50%	51-75%	76-95%	Always
Socially (friends):	Never	5-25%	26-50%	51-75%	76-95%	Always

Learning of English

5. Age of first exposure to English: _____Context: ☐At school ☐Outside school ☐Both
6. English instruction:
   a. Number of years of English instruction that you have received: ____
   b. English dialect spoken by your teachers

      American British

      Other: ________________ I don't know

   c. Were most of your English teachers native speakers of English? ☐Yes ☐ No
7. a. Have you ever taken an English pronunciation class? ☐Yes ☐ No

   b. If your answer is "yes", what course did you take, where and when?__________
8. Do you use English outside the university setting? (choose as many as you want)
   ☐ No, I don't use English outside the classroom
   ☐Yes, with my family ☐Yes, with native English speaking friends ☐Yes, reading/TV/internet ☐
   Other situation (specify) ________________________
9. I'm most familiar with this variety of English
   ☐British ☐American ☐ other (specify which):__________ ☐ None in particular
10. Have you ever spent time in an English-speaking country? If you have, fill out the following information about each place you have visited.

    i. Age: ________Place: ____________________Context: ____

    ii. Duration:_________________year(s)_____________________month(s) _________________________week(s)

    i. Age: _________Place: _____________________Context: ____

    ii. Duration:_________________year(s)_____________________month(s) _________________________week(s)

Appendix C - NES Questionnaire

Questionnaire filled out by the native English speakers who rated the production data

Questionnaire

Universitat Autònoma de Barcelona

Experiment date and time: ____________________________

Personal Information

- Name: ____
- Age: ____
- Occupation: ____
- Place of birth: ____
- Parents' place of birth: ____
- Place(s) of residence of the past 10 years:

Language information

- Native language: ____
- Parents' native language: ____
- Do you speak any other language fluently? _________________________________

  ____

Comments

Appendix D - Testing stimuli[34]

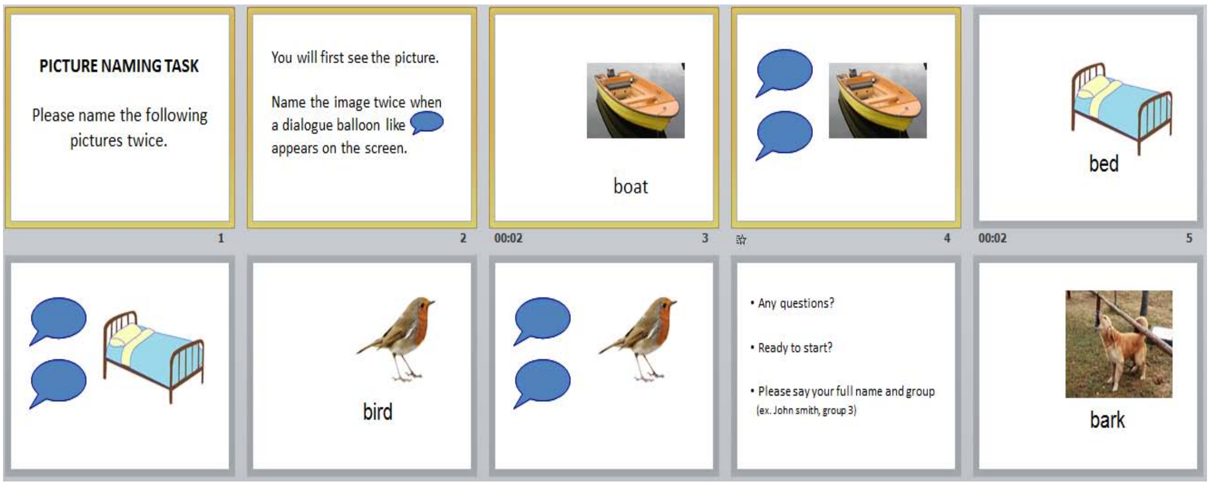
Practice stimuli					
Consonant (initial)		**Vowels**		**Consonant (final)**	
b_	bav	/ɑː/	barsh	_t	chet
d_	darge	/ɜː/	burf	_b	jeb
g_	gack	/i/	deedge	_d	Jeed
k_	kerch	/ʌ/	dudge	_k	vack
p_	pidge	/æ/	gack	_p	veep
t_	tudge	/ɪ/	kidge	_g	verg
		/e/	veck		
Testing stimuli					
Consonant (initial)		**Vowels**		**Consonant(final)**	
High	**Low**	**Voiced**	**Voiceless**	**Tense**	**Lax**
bidge	barsh	vab	vap	jurb	vib
deedge	dudge	zad	zat	jeed	zud
geedge	gerch	vag	vack	verg	vag
keedge	kerch	vub	vup	verk	vack
pidge	parsh	zud	zut	jurp	vip
teedge	tudge	vugg	vuck	Jeet	zut
		veeb	veep		
		jeed	jeet		
		veeg	veek		
Fillers		vib	vip		
veck		jid	jit		
vegg		vig	vick		
parsh		jurb	jurp		
barsh		jerd	jurt		
		verg	verk		

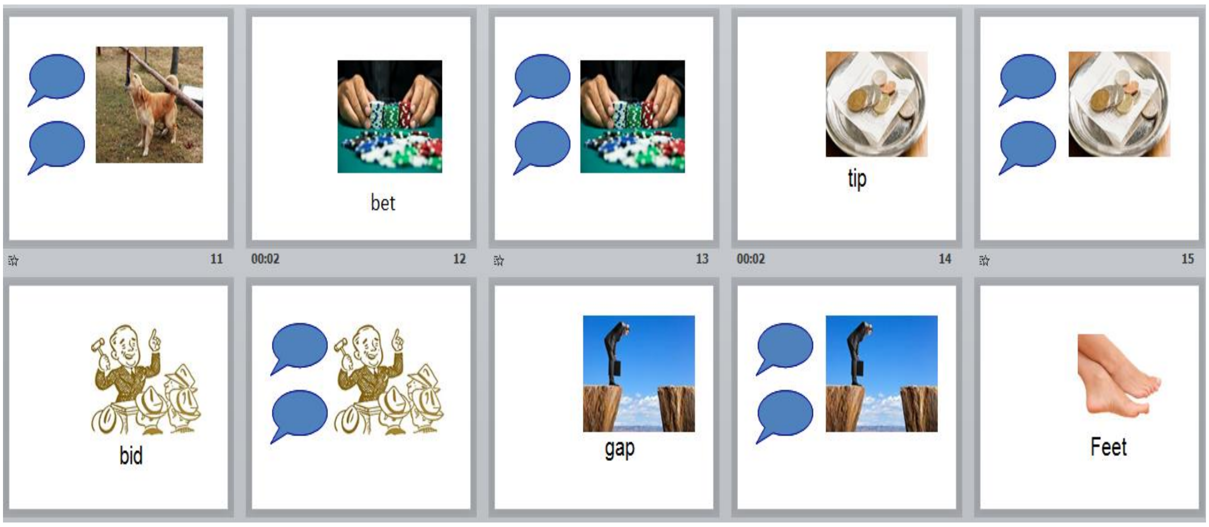
[34] Some nonsense words created for the experiment were found to have a meaning (i.e. chert, dirge) or a colloquial meaning in English (i.e. cack, burf, gertch) as listed in the urbandictionary.com . However, it was acknowledged that the learners participating in this study would not be familiar with such words and therefore no word knowledge bias would interfere with the experiment.

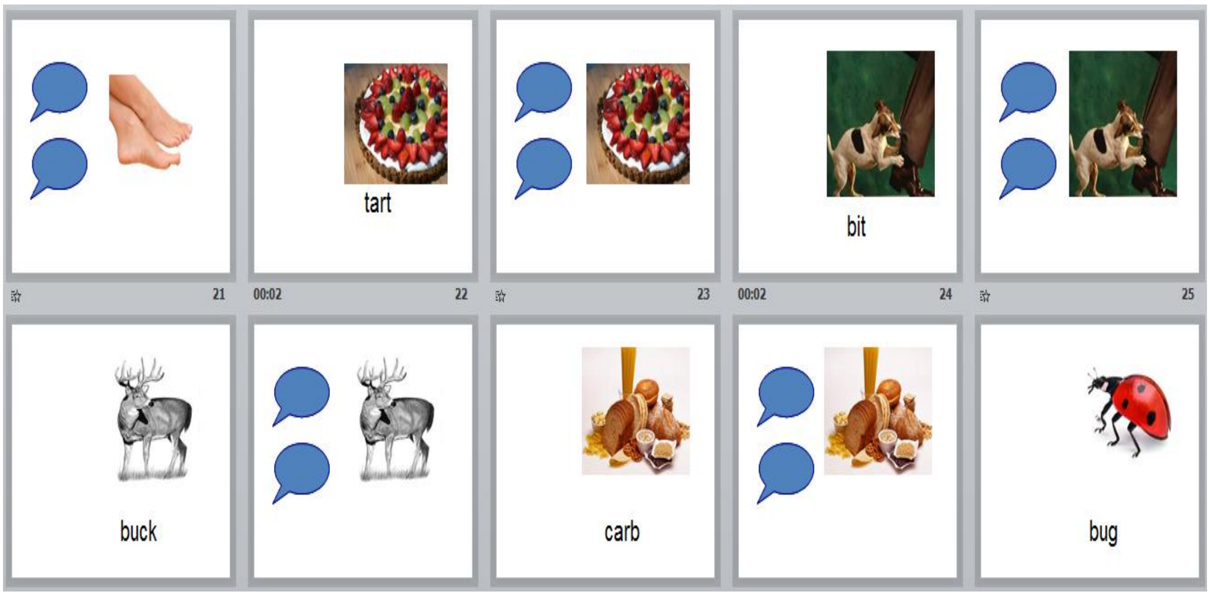
Appendix E - Generalization stimuli

Generalization to real words (Generalization 1)							
Consonant (initial)		**Consonant (final)**			**Vowels**		**Fillers**
High	**Low**	**Lax**	**Tense**		**_Vless**	**_Vced**	**Real words**
bet	bark	cab	carb		cap	cab	bark
dip	dart	bed	heard		pup	pub	tart
guilt	gap	bug	league		feet	feed	pet
kilt	cap	buck	leak		bit	bid	bed
pet	park	cap	carp		hurt	heard	
tip	tart	bet	hurt				
Generalization to new non-words (Generalization 2)							
Consonant (initial)			**Consonant (final)**			**Vowels**	
High	**Low**		**Lax**	**Tense**		**_Vless**	**_Vced**
pid	pag		dack	jeek		dack	pag
teeve	tacht		dut	vert		fip	pid
kiv	kav		Fip	kerp		geep	keeb
beesh	berve		bab	keeb		vert	derg
deeth	dup		jud	serd		dut	Jud
geep	gadge		sug	derg			
Generalization to novel talkers (Generalization 3)							
Consonant (initial)			**Consonant (final)**			**Vowels**	
High	**Low**		**Lax**	**Tense**		**_Vless**	**_Vcd**
teedge	tudge		zut	jeet		zat	zad
teedge	dudge		zud	jeed		vup	vub
pidge	parsh		vip	jerp		veek	veeg
bidge	barsh		vib	jerb		jit	jid
keedge	kerch		vack	verk		jerp	jerb
geedge	gerch		vagg	verg			

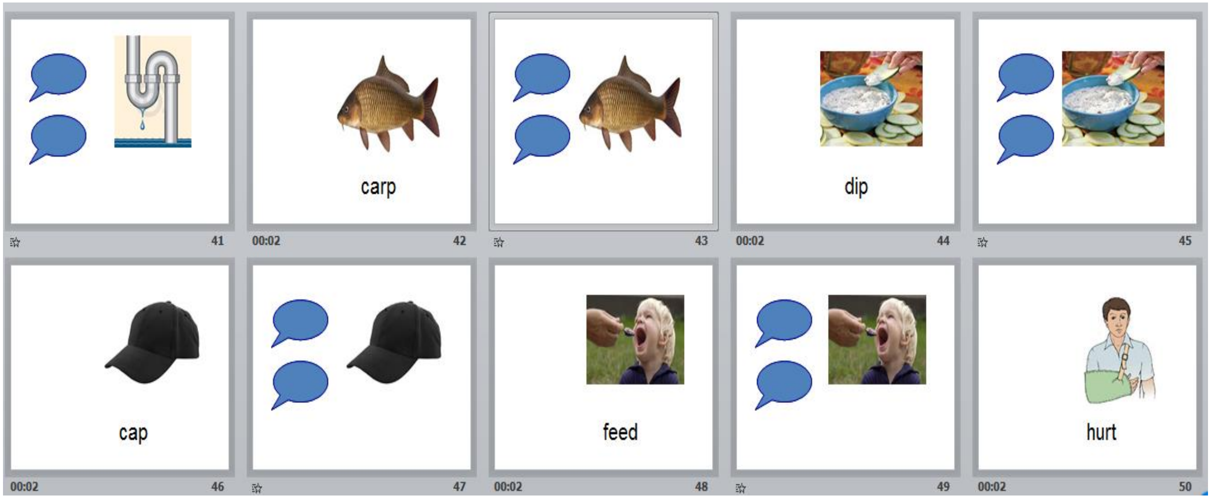
Appendix F - Production elicitation PowerPoint

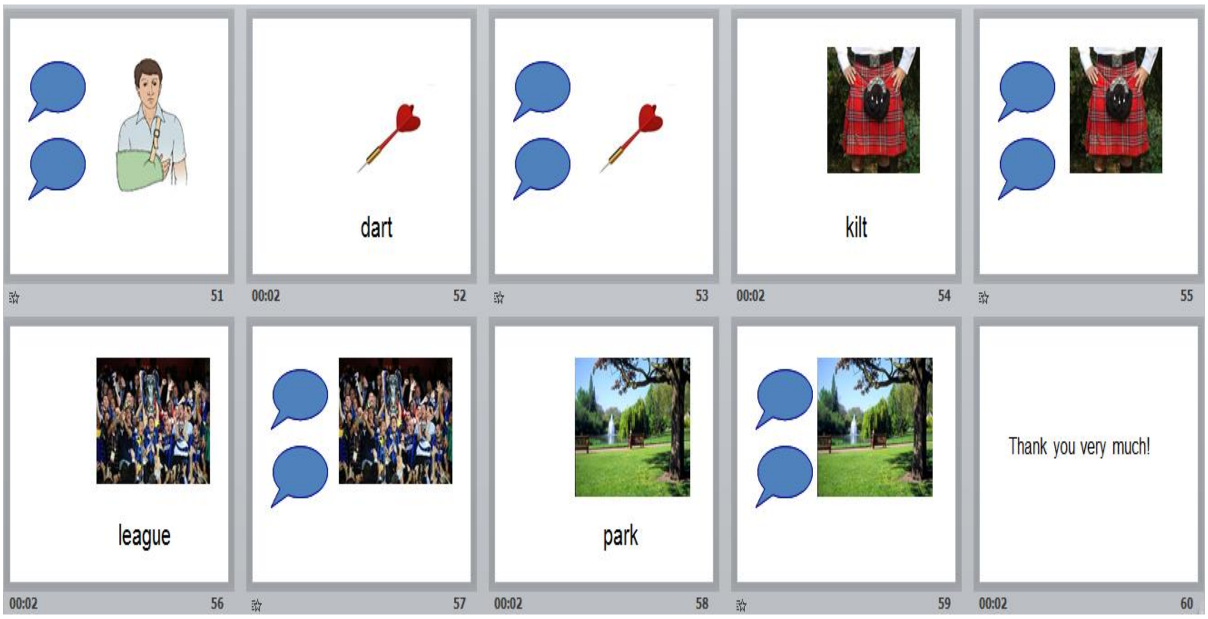







guilt
cab
31
00:02
32
33
00:02
34
35
heard
pet
leak


carp
dip
41
00:02
42
43
00:02
44
45
cap
feed
hurt
00:02
46
47
00:02
48
49
00:02
50


dart
kilt
51
00:02
52
53
00:02
54
55
league
park
Thank you very much!
00:02
56
57
00:02
58
59
00:02
60

Appendix G - Testing instructions

Instructions for the Identification experiments

You will be hearing a series of non-words and some real words. Your task is to follow the instruction that appears on the top green bar and click on the sound you think that matches the word/non-word spoken. The instructions will be as follows:

a) Identify the **INITIAL CONSONANT** you hear **(p_ , b_ , t_ , d_ , k_ , g_)**

b) Identify the **VOWEL** you hear **(/æ/, / ʌ/, / ɪ/, / iː/,/ ɜː/,/ e /,/ ɑː/)**

c) Identify the **FINAL CONSONANT** you hear **(_p, _b, _t, _d, _k, _g)**

Pay attention to the instructions before clicking on a response option. If you wish to hear the word again, click on the replay button. Be alert as the sound will be played as soon as you click to start the experiment.

When you complete the experiment, your results will be displayed on the screen. Copy the number of ***hits*** to your lab sign-up sheet.

When you've finished reading these instructions, please click on "*Application*", write your first and last names, and then press "*Start Identification Test*" to begin the experiment.

Appendix H - Post training survey results

angelicacarlet@gmail.com

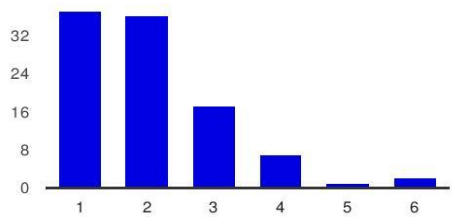
Edit this form

100 responses

View all responses

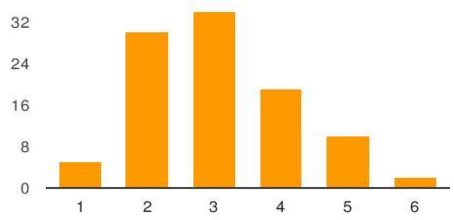
Summary

I found the lab sessions interesting



Strongly agree : 1	**37**	37%
2	**36**	36%
3	**17**	17%
4	**7**	7%
5	**1**	1%
Strongly disagree: 6	**2**	2%

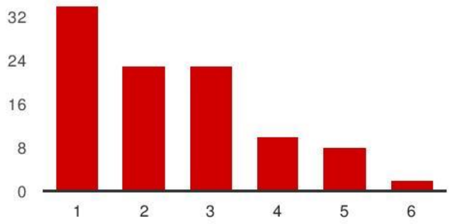
I found the lab sessions enjoyable



Strongly agree : 1	**5**	5%
2	**30**	30%
3	**34**	34%
4	**19**	19%

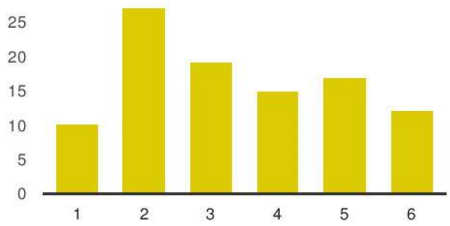
5	**10**	10%
Strongly disagree: 6	**2**	2%

The duration of the lab sessions was adequate (not too long)



Strongly agree : 1	**34**	34%
2	**23**	23%
3	**23**	23%
4	**10**	10%
5	**8**	8%
Strongly disagree: 6	**2**	2%

I felt tired after each session

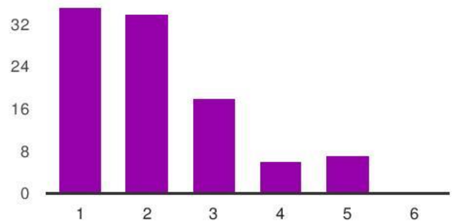


Strongly agree : 1	**10**	10%
2	**27**	27%
3	**19**	19%
4	**15**	15%
5	**17**	17%
Strongly disagree: 6	**12**	12%

I think a subject like "Phonetics and Phonology 1" should include ear training exercises like the lab sessions.

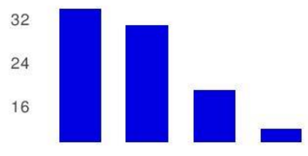
Strongly agree : 1	**84**	84%
2	**9**	9%
3	**1**	1%
4	**2**	2%
5	**4**	4%
Strongly disagree: 6	**0**	0%

I managed to stay focused and to concentrate on the tasks during the sessions



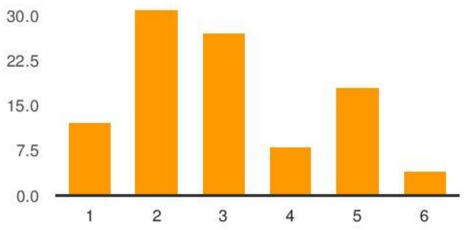
Strongly agree : 1	**35**	35%
2	**34**	34%
3	**18**	18%
4	**6**	6%
5	**7**	7%
Strongly disagree: 6	**0**	0%

I found the lab sessions very useful to help me work on my pronunciation



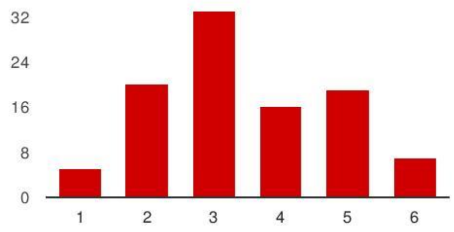
Strongly agree : 1	**34**	34%
2	**31**	31%
3	**19**	19%
4	**12**	12%
5	**1**	1%
Strongly disagree: 6	**3**	3%

I liked the type of exercise I was doing



Strongly agree : 1	**12**	12%
2	**31**	31%
3	**27**	27%
4	**8**	8%
5	**18**	18%
Strongly disagree: 6	**4**	4%

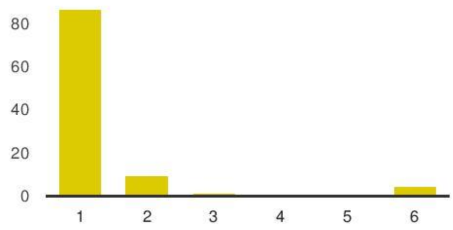
The type of exercise I was doing was difficult



Strongly agree : 1	**5**	5%

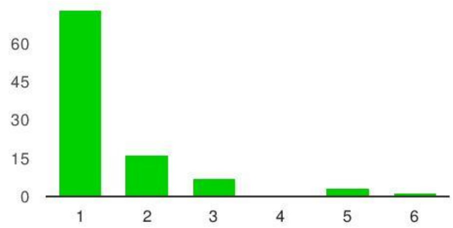
2	**20**	20%
3	**33**	33%
4	**16**	16%
5	**19**	19%
Strongly disagree: 6	**7**	7%

Instructions to the exercises were clear



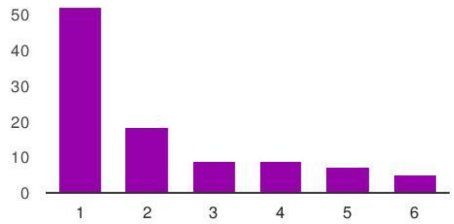
Strongly agree : 1	**86**	86%
2	**9**	9%
3	**1**	1%
4	**0**	0%
5	**0**	0%
Strongly disagree: 6	**4**	4%

I felt motivated when I got answers right



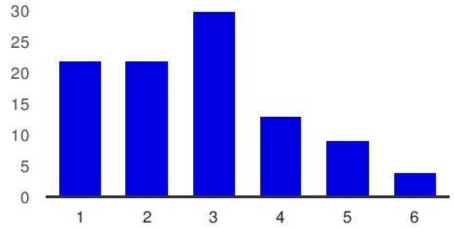
Strongly agree : 1	**73**	73%
2	**16**	16%
3	**7**	7%
4	**0**	0%
5	**3**	3%
Strongly disagree: 6	**1**	1%

I felt frustrated when I got answers wrong



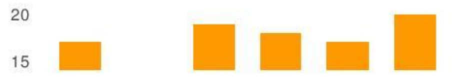
Strongly agree : 1	**52**	52%
2	**18**	18%
3	**9**	9%
4	**9**	9%
5	**7**	7%
Strongly disagree: 6	**5**	5%

I would have preferred to have listened to real words instead of non-words during the sessions



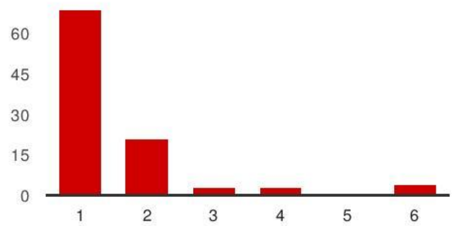
Strongly agree : 1	**22**	22%
2	**22**	22%
3	**30**	30%
4	**13**	13%
5	**9**	9%
Strongly disagree: 6	**4**	4%

I would have preferred to have been able to do the lab sessions from home at my own time



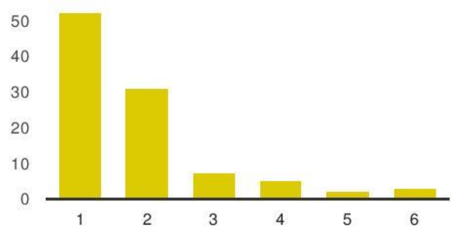
Strongly agree : 1	**17**	17%
2	**9**	9%
3	**19**	19%
4	**18**	18%
5	**17**	17%
Strongly disagree: 6	**20**	20%

Some speakers were easier to understand than others



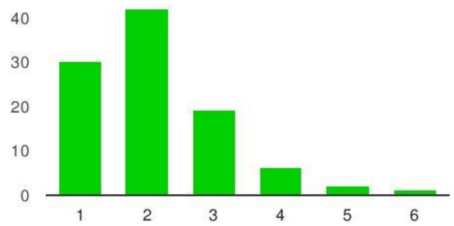
Strongly agree : 1	**69**	69%
2	**21**	21%
3	**3**	3%
4	**3**	3%
5	**0**	0%
Strongly disagree: 6	**4**	4%

After a few sessions, it became easier to identify/discriminate the sounds



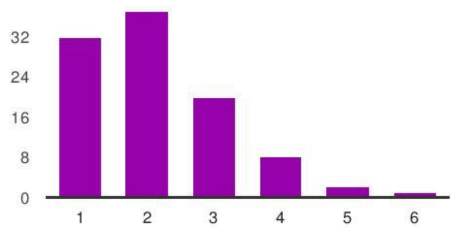
Strongly agree : 1	**52**	52%
2	**31**	31%
3	**7**	7%
4	**5**	5%
5	**2**	2%
Strongly disagree: 6	**3**	3%

My ability to perceive English sounds has improved thanks to the lab sessions.



Strongly agree : 1	**30**	30%
2	**42**	42%
3	**19**	19%
4	**6**	6%
5	**2**	2%
Strongly disagree: 6	**1**	1%

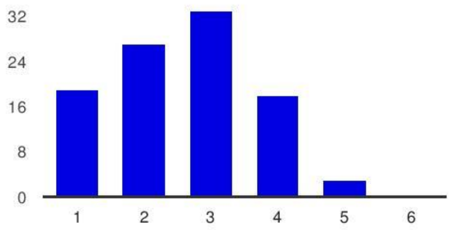
I feel more confident in identifying/discriminating English sounds after the lab sessions



Strongly agree : 1	**32**	32%
2	**37**	37%
3	**20**	20%
4	**8**	8%

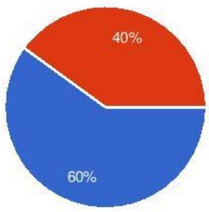
5	**2**	2%
Strongly disagree: 6	**1**	1%

I feel more confident in producing English sounds after the lab sessions



Strongly agree : 1	**19**	19%
2	**27**	27%
3	**33**	33%
4	**18**	18%
5	**3**	3%
Strongly disagree: 6	**0**	0%

I have come up with some "strategies" to identify/discriminate the sounds



Yes	**60**	60%
No	**40**	40%

If you answered yes to the question above, please state what strategy you used.

In the final consonant I tried to identifiy the final sound, and if the sound was the same in the other word I putted same or different.

Long/Short vowels

the enotonation

listen it carefully or twice

paying attention at the duration of the vowel as well as at the backness

I used some words as reference/example of some sounds

Repeat the words in my head to identify the sound

the difference beetween long vowels and short ones as well as words like dutch or deutch and so on.

Trying to listen if the sound was tense or lax, and/or front or back

In order to differenciate the /i/ sound, I gave more importance if it was long or short.

listen to the lenght of some vowels,

I found some connections among the sounds I heard and I tried to put them together in a 'Mind Chart'

For instance after vociless sounds we usually find short vowel

some "a" vowels sound more lika an "o" and their sound is shorter

I compared ^ to an o

Repeat the "word", or the sound in this case, and see if the lenght was the same, the backnes of the vowel was the same and if the mouth was too open or not. Just comparing one sound repeating it in my head, to the other one that I listened.

IDENTIFY THE VOWEL PROPERLY

memorising some word recordings and therefore identifying them as real words

repeat every sound by myself

Observe things like if the consonant is aspirated

to keep in mind the first sound as if I was producing it so as to see if the second one differed

When trying to distinguish between voiced and voiceless sounds I try to focus and hear if there is a plosion or not after the air is released.

listen to the recording twice with my eyes closed and repeating the same sound myself.

long and short vowels to identify voiced or voiceless ending consonants.

The vowels were shorter before a voiceless consonant.

I've developed a intuitive sistem of discrimination

I identify if it was any kind of aspiration to choose between p/b t/d and k/g

The length of the vowels

Finding out if the vowels are shorter or longer

Difference in the preceding sound

long and short vowels before a consonant, the aspiration at the beginning of a word

I tried to pay attention to the aspiration of inicial stops

I just focuse on the sound required and I compare it with the second sound. Sometimes the lenght of the vowels is useful.

To listen to the sound patiently

a lot of concentration

repeat the sounds for myself

To close the eyes and to think with hard attention the words given.

identify the final consonant by hearing if the vocal was long or not

I tried to imagine both speakers producing the sound and then notice if it was the same or different.

listen to vowels

depending on the speaker the I was different

To identify the longer and shorter vowels I tried to count seconds, and to the different /ae/

and /^/ sounds i tried to imagine if they were smiling or not.

I remember the sound of the first person and when the second person makes the other sound I make the first speaker's sound at the same time.

I try to remember those sounds which I always fail

When I had doubts about a vowel I tried to think of the phonemes and how each was pronounced to see if the sounds I heard corresponded to a same phoneme or not.

reproduce the first sound in my mouth and if I had to change it with the second one it meant it was different

Try to erase the words that were not asked to identify and focus just on the asked vowel or consonant.

I tried to hear if they do some aspiration after the voiceless sounds

Remembering some pairs of sounds

I started discriminating between b and p due to the aspiration

i repeat all the word in my own, to know the difference

Closing my eyes to focus on listening.

There were some pairs of non-words, whose vowel quality would always be the same or different every time they appeared.

Some sounds were repeated troughout the session, so I realized that, when I was wrong, I could check the correct answer and I was not wrong again

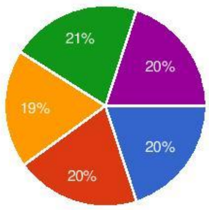
I have try to identify the long vowels

I tried switching the vowels and see if they would still remain the same or they sounded different

I more or less tried to discriminate between long and short vowels, and I also mentally and silently repeated the segments I heard

Consider the context surrounding the vowels and pay attention to the seech patterns.

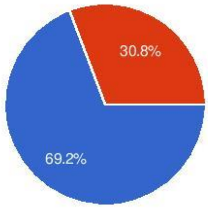
My training group was



B _ Identification (vowels)	**20**	20%
C_ Identification (consonants)	**20**	20%
D_Discrimination (vowels)	**19**	19%
E_Discrimination (consonants)	**21**	21%
F_ Cross_Discrimination (vowels)	**20**	20%

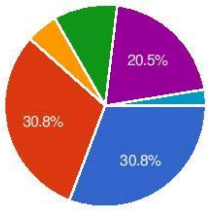
GROUPS D and F_ DISCRIMINATION vowels

I was answering about vowels and..



I only paid attention to vowel sounds	**27**	69.2%
I paid attention to vowel and consonant sounds	**12**	30.8%

The most difficult vowel pair to differentiate was



1 -/æ - ʌ/	**12**	30.8%
2 -/ɪ - i:/	**12**	30.8%
3- /e - ɜ:/	**2**	5.1%
4 - /ɜ:- ɑ:/	**4**	10.3%
Two (or more) pairs were equally difficult	**8**	20.5%
All were equally difficult	**1**	2.6%

If you answered "two (or more) pairs were equally difficult", please specify which ones

I could not say which ones specifically. However if I have to choose, i would rather say the e - 3:, and 3: - a:

1 and 2

1, 3 and 4

1-4

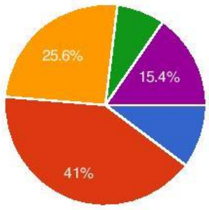
2, 3

1, 2

2, 4

1, 2 and 4, but especially 2

The easiest vowel pair to differentiate was



1- /æ - ʌ/	**4**	10.3%
2 - /ɪ - iː/	**16**	41%
3 -/e - ɜː/	**10**	25.6%
4 - /ɜː- ɑː/	**3**	7.7%
Two (or more) pairs were equally easy	**6**	15.4%
All were equally easy.	**0**	0%

If you answered "two (or more) pairs were equally easy", please specify which ones.

3 and 4

2-3

1, 3

2,4

1, 3 and 4

Please write any comments you may have regarding the activity or the lab sessions in general

I enjoyed the lab sessions; and I think that this activity would help other students who have problems with the differentiation of the sounds.

It is been useful!

There were some /ɪ/ that I thought that were long and then were short. I would have put my hand into the fire that they were both short, in some cases.

It was a bit long and when you were in the 200th exercise, you were a bit tired.

I think that the lab pratices sessions have been interesting and it could also be interesting to recording and listening ourselves and also, but quite impossible, to look at VOT or our vocal folds

It is very useful to learn every sound, but maybemaybe it would be useful to know which especific sound we are listening

Definately I consider this has been very useful to improve our pronounciation

I'd have liked to practise consonants too.

I found very useful this activity because I had problems to identify the different vowels (I used to see some of them as they were the same). However, I think that there were lots of exercises and when I was more or less at the middle I was feeling tired and had trouble to understand the sounds. In my opinion it would be better to do more sessions with less exercises, so that we get used to the sounds.

They were too long

Found it very useful to improve our vowel identification skills.

at the end of the 7 lab sessions , you can see the difference of the first day beetwen the las day

I think there were too many stimuli because at the end I began to make more mistakes.

It would be great if they did this all the years for Phonetics and Phonology I, it helps a lot, rather than just reading examples from a book. I would also recommend they put more "different" words, I know it may be difficult to do it, but a lot of them seemed the same.

These lab sessions have been quite helpful to my identifying vowels and, too, at comprehending vowel quality.

I think there are too many words, I have headache after the lab sessions.

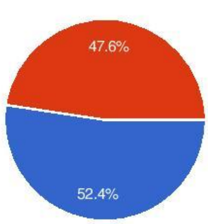
Although it is a bit repetitive and exhausting, it is a really useful activity that helped to improve my pronunciation. Thank you!

It was different, quite refreshing

I think they are ideal to help students learn identifying both consonants and vowels. As far as I am concerned it should be compulsory for everybody taking English Phonetics 1 because the exercises help us train our ear on how identificate sounds and that is really useful in order to pass the subject.

GROUP E _ DISCRIMINATION Consonants

I was answering about consonants and..



I only paid attention to consonant sounds	**11**	52.4%
I paid attention to vowel and consonant sounds	**10**	47.6%

The most dificult INITIAL consonant pair to differentiate was

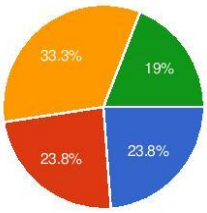
1- /p-b/ Initial	8	38.1%
2- /k-g/ initial	8	38.1%
3- /t-d/ initial	2	9.5%
Two (or more) pairs were equally difficult	2	9.5%
All were equally difficult	1	4.8%

If you answered "two (or more) pairs were equally difficult", please specify which ones.

2-3

/t-d/ inital and /k-g/ initial

The easiest INITIAL consonant pair to differentiate was



1- /p-b/ Initial	5	23.8%
2 -/k-g/ initial	5	23.8%
3-/t-d/ initial	7	33.3%
Two (or more) pairs were equally easy	4	19%
All were equally easy	0	0%

If you answered "two (or more) pairs were equally easy", please specify which ones.

2 and 3

1 AND 3

/P-B/ AND /K-G/

/P-B/ AND /T-D/

The most dificult FINAL consonant pair to differentiate was

1- /p-b/ final	**4**	19%
2 - /k-g/ final	**6**	28.6%
3- /t-d/ final	**3**	14.3%
Two (or more) pairs were equally difficult	**4**	19%
All were equally difficult	**4**	19%

If you answered "two (or more) pairs were equally difficult", please specify which ones.

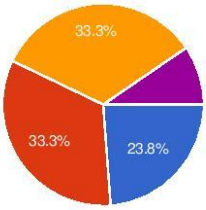
/p-b/ and /t-d/

p-b and t-d

1 AND 3

/P-B/ AND /K-G/

The easiest FINAL consonant pair to differentiate was



1- /p-b/ final	**5**	23.8%
2- /k-g/ final	**7**	33.3%
3- /t-d/ final	**7**	33.3%
Two (or more) pairs were equally easy	**0**	0%
All were equally easy	**2**	9.5%

If you answered "two (or more) pairs were equally easy", please specify which ones.

they weren't easy but I had to choose one

Please write any comments you may have regarding the activity or the lab sessions in general

Sessions very useful and interesting

I would like to have practiced vowels and consonats

I found this activity really useful, I do noticed my ear is more aware of different sounds now. This should continue to be done as a complement to the main lectures.

There are some sounds that doesn't work, they aren't really words or non-words, they are just a 'piiip' sound.

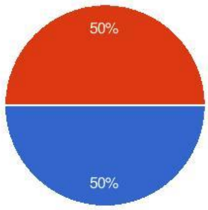
In my opinion I would like to do this exercise during all the semester.

I found it very useful

I think it should be better to provide real practices, with the teacher and a group of students, because with the computer you get tired so easily and it is always the same

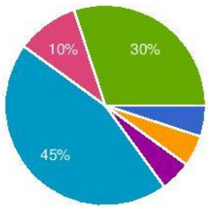
Group B_ IDENTIFICATION VOWELS

I was answering about vowels and..



I only paid attention to vowel sounds	**10**	50%
I paid attention to vowel and consonant sounds	**10**	50%

The easiest vowel to Identify was



1- /æ/	**1**	5%
2- /ʌ/	**0**	0%
3- /ɑ:/	**1**	5%
4- /ɪ/	**0**	0%
5- /i:/	**1**	5%
6- /e/	**9**	45%
7- /ɜ:/	**2**	10%

Two (or more) vowels were equally easy	**6**	30%
All were equally easy	**0**	0%

If you answered "two (or more) vowels were equally easy", please specify which ones.

ae and e

5, 6.

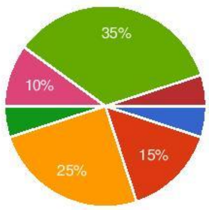
1,4,5,6,7,

2-6

short i, long i

5 /i:/ 4 /ɪ/ 6/e/

The most difficult vowel to Identify was



1- /æ/	**1**	5%
2- /ʌ/	**3**	15%
3- /ɑ:/	**5**	25%
4- /ɪ/	**1**	5%
5- /i:/	**0**	0%
6- /e/	**0**	0%
7- /ɜ:/	**2**	10%
Two (or more) vowels were equally difficult	**7**	35%
All were equally difficult	**1**	5%

If you answered "two (or more) vowels were equally difficult", please specify which ones.

1 and 3

1, 2

1-3-4-5

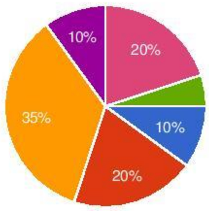
i confuse 2 and 3 sometimes.

1 and 2

4, 5

1, 2, 3, 7

The easiest vowel pair to differentiate was



1 -/æ - ʌ/	**2**	10%
2 -/ɪ - iː/	**4**	20%
3- /e - ɜː/	**7**	35%
4 - /ɜː- ɑː/	**0**	0%
5 - /æ -ɑː/	**2**	10%
6 - /ʌ -ɑː/	**0**	0%
Two (or more) pairs were equally easy	**4**	20%
All were equally easy	**1**	5%

If you answered "two pairs were equally easy", please specify which ones.

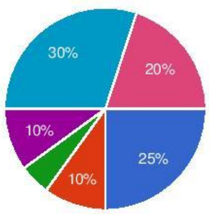
2, 3.

2,3,5

3-5-6

2 and 3

The most difficult vowel pair to differentiate was



1 -/æ - ʌ/	**5**	25%
2 -/ɪ - iː/	**2**	10%
3- /e - ɜː/	**0**	0%
4 - /ɜː- ɑː/	**1**	5%
5 - /æ -ɑː/	**2**	10%
6 - /ʌ -ɑː/	**6**	30%
Two (or more) pairs were equally difficult	**4**	20%
All were equally difficult	**0**	0%

If you answered "two (or more) pairs were equally difficult", please specify which ones.

2, 4

2-5

1, 5 and 6

1, 5, 6

Please write any comments you may have regarding the activity or the lab sessions in general

I found it really useful to have some example of different vowels

This kind of activity would be useful for each year of the degree or even the first one. Phonetics is an essential key which has to be taught from the very beginning of the learning process of no matter what language.

it was very useful

i found them very usefull and i think my skills at listening are better now

I didn't like doing it at the beginning, but I think it ended up helping me a lot.

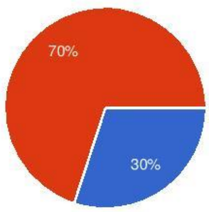
I found the activity very useful. I would recommend it to any of the future students attending "phonetics and phonology I" subject. I strongly think that it is very important to work on practice exercices concerning phonetic practice in order to have a larger perspective on how words are pronounced by native RP speakers.

I like them but I found them a bit too long and a bit tiring, but they really helped me.

I am satified because of having done these sessions ^^

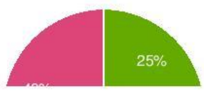
GROUP C_ IDENTIFICATION CONSONANTS

I was answering about consonants and..



I only paid attention to consonant sounds	**6**	30%
I paid attention to vowel and consonant sounds	**14**	70%

The easiest INITIAL consonant to identify was



1 - p	**2**	10%
2 - b	**0**	0%
3 - t	**1**	5%
4 - d	**1**	5%
5 - k	**1**	5%
6 - g	**2**	10%
Two (or more) consonants were equally easy	**8**	40%
All were equally easy	**5**	25%

If you answered "two (or more) consonants were equally easy", please specify which ones.

3 and 5

4-3-1

3, 4, 5, 6

1,3,5

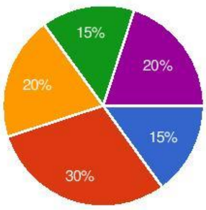
1, 3, 5

k and g

d and t

p, t, k

The easiest INITIAL consonant pair to differentiate was



1- /p-b/ Initial	**3**	15%
2 -/k-g/ initial	**6**	30%
3-/t-d/ initial	**4**	20%
Two (or more) pairs were equally easy	**3**	15%
All were equally easy	**4**	20%

If you answered "two (or more) pairs were equally easy", please specify

which ones.

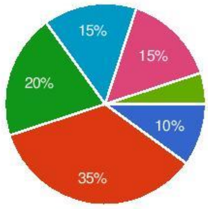
2, 3

g and k

/p-b/ and /t-d/

p, t, k

The most difficult INITIAL consonant to identify was



1 - p	**2**	10%
2 - b	**7**	35%
3 - t	**0**	0%
4 - d	**4**	20%
5 - k	**0**	0%
6 - g	**3**	15%
Two (or more) consonants were equally difficult	**3**	15%
All were equally difficult	**1**	5%

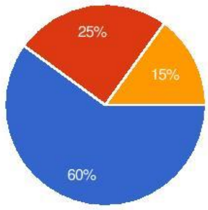
If you answered "two (or more) consonants were equally difficult", please specify which ones.

2 and 6

p-b

p and b

The most dificult INITIAL consonant pair to differentiate was

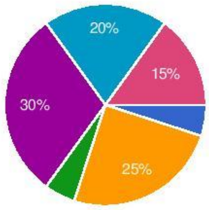


1- /p-b/ Initial	**12**	60%
2- /k-g/ initial	**5**	25%

3- /t-d/ initial	**3**	15%
Two (or more) pairs were equally difficult	**0**	0%
All were equally difficult	**0**	0%

If you answered "two (or more) pairs were equally difficult", please specify which ones.

The easiest FINAL consonant to identify was



1 - p	**1**	5%
2 - b	**0**	0%
3 - t	**5**	25%
4 - d	**1**	5%
5 - k	**6**	30%
6 - g	**4**	20%
Two (or more) consonants were equally easy	**3**	15%
All were equally easy	**0**	0%

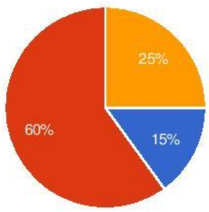
If you answered "two (or more) consonants were equally easy", please specify which ones.

g and k

p, b

k, g

The easiest FINAL consonant pair to differentiate was



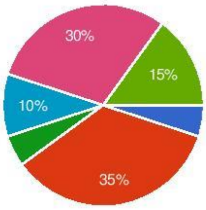
1- /p-b/ final	**3**	15%

2- /k-g/ final	**12**	60%
3- /t-d/ final	**5**	25%
Two (or more) pairs were equally easy	**0**	0%
All were equally easy	**0**	0%

If you answered "two (or more) pairs were equally easy", please specify which ones.

g and k

The most difficult FINAL consonant to identify was



1 - p	**1**	5%
2 - b	**7**	35%
3 - t	**0**	0%
4 - d	**1**	5%
5 - k	**0**	0%
6 - g	**2**	10%
Two (or more) consonants were equally difficult	**6**	30%
All were equally difficult	**3**	15%

If you answered "two (or more) consonants were equally difficult", please specify which ones.

1, 2, 4, 6

2 and 4

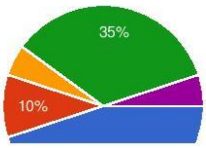
6-5-1-2

1,2,3,4

2,6

t and d

The most dificult FINAL consonant pair to differentiate was



1- /p-b/ final	**9**	45%
2 - /k-g/ final	**2**	10%
3- /t-d/ final	**1**	5%
Two (or more) pairs were equally difficult	**7**	35%
All were equally difficult	**1**	5%

If you answered "two (or more) pairs were equally difficult", please specify which ones.

1, 3

1, 2

1 and 3

2-1

1,3

1,2

p-b/t-d

Please write any comments you may have regarding the activity or the lab sessions in general

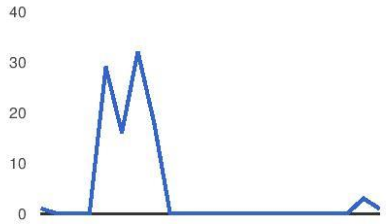
I think this is a good and useful activity. Sometimes I find it quite long but I can really notice I am improving. I am happy.

I understand the lenght of the excercice since it follows a scientific pattern, but as a student I've found it way too long; 576 times making the same routine.

Some of the sounds were confusing and sometimes they seem the same

As there are too much words to listen I feel sometimes freaked out

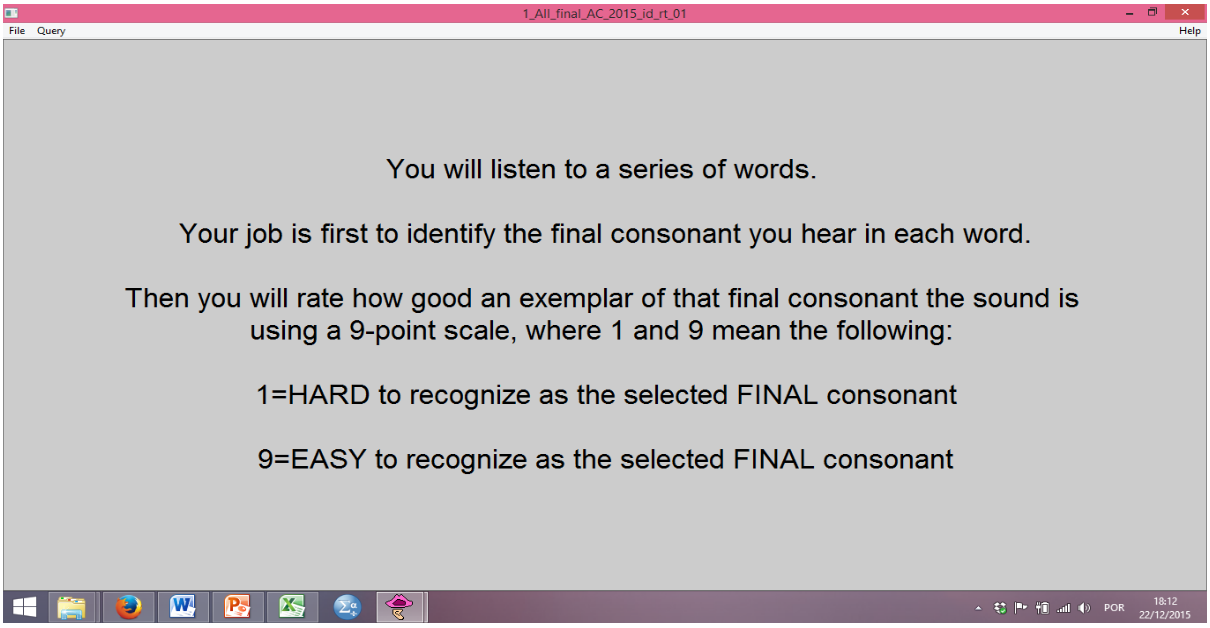
Number of daily responses



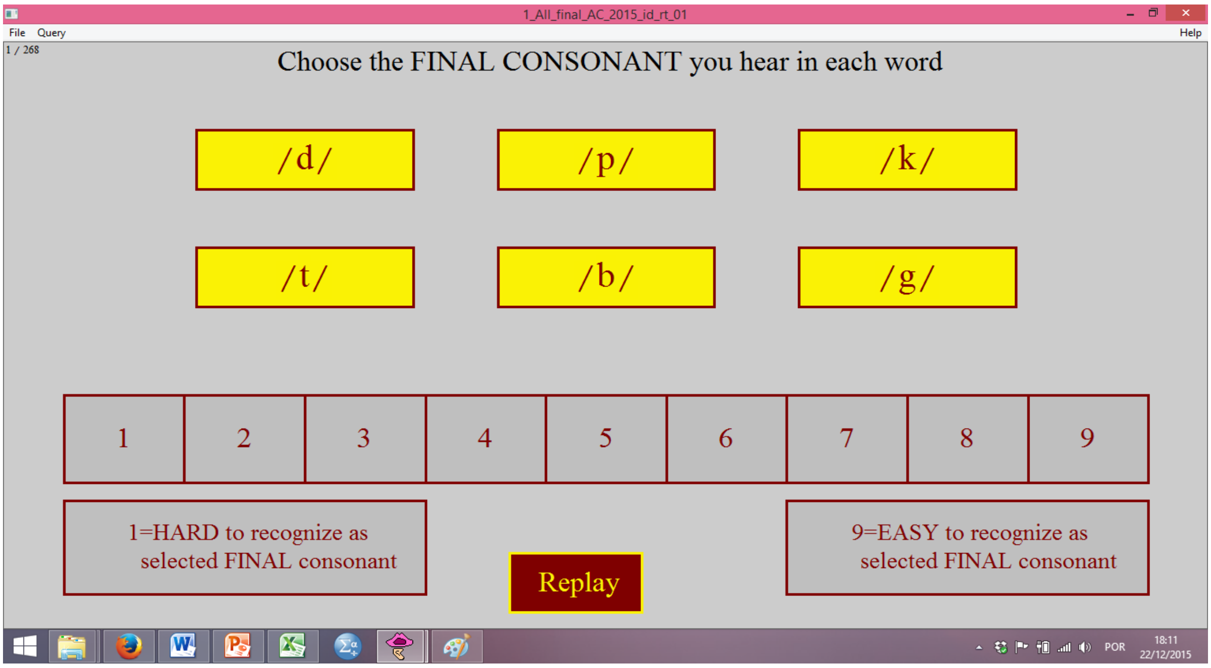
Appendix I – Production data rating tests

1) Final consonant rating test

a) Instructions for raters

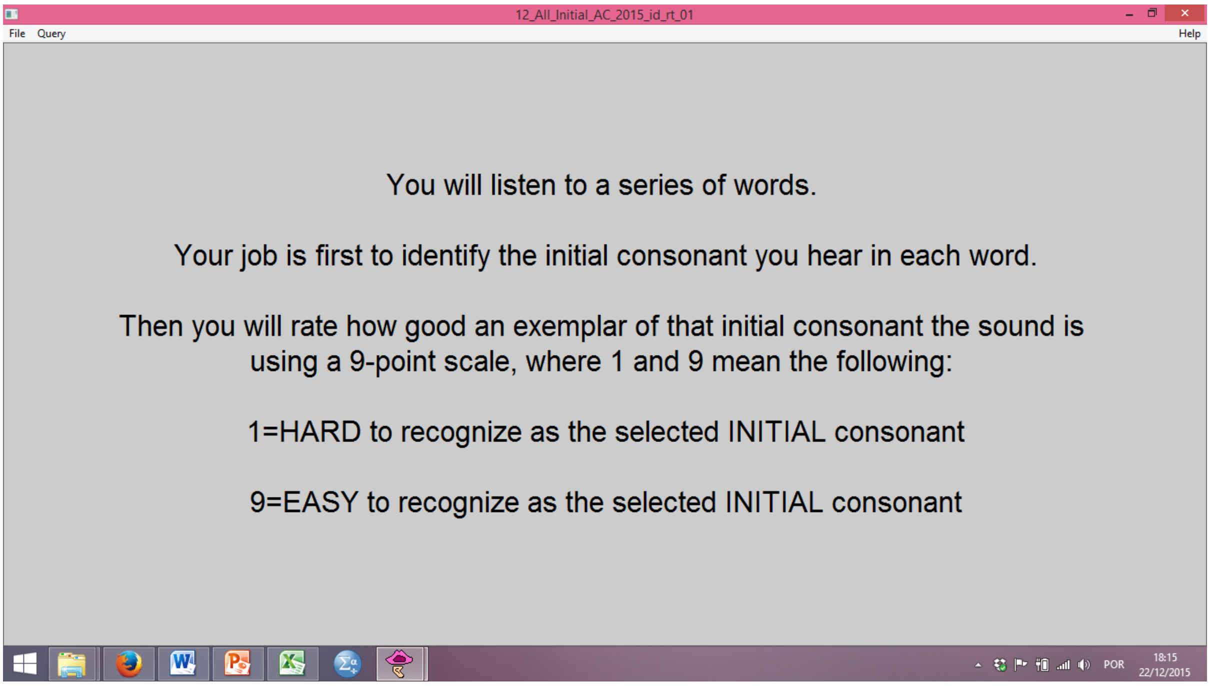


b) Response options and rating scale

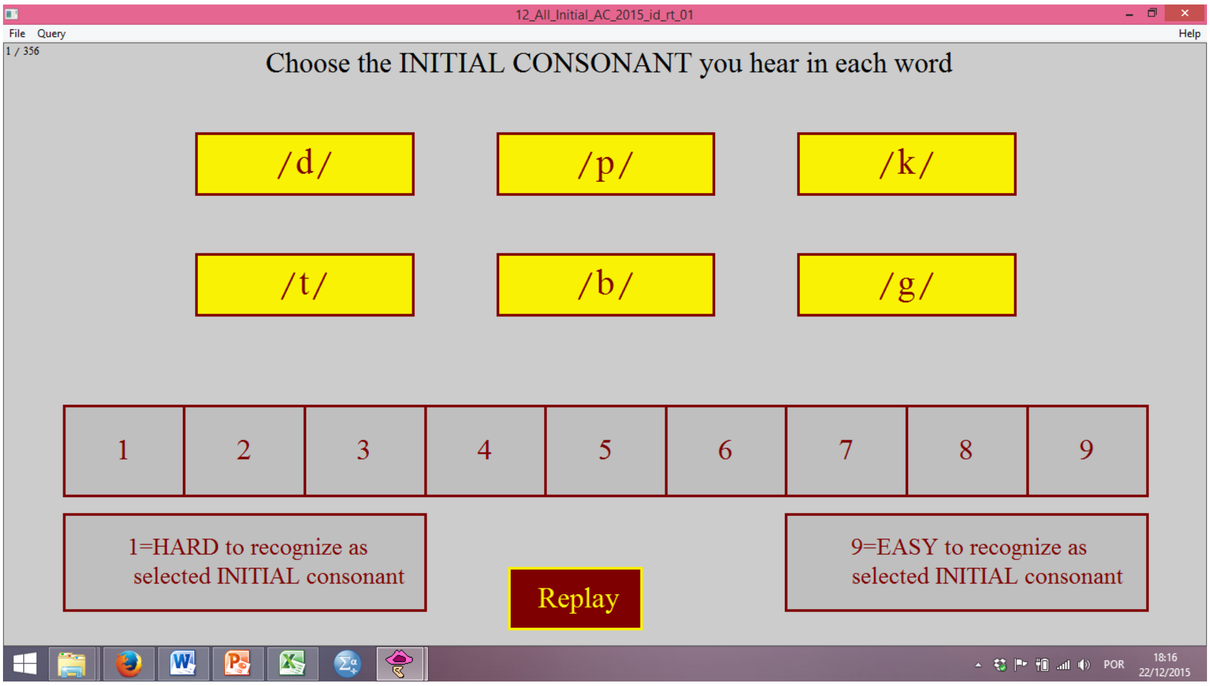


2) Initial consonant rating test

a) Instructions for raters

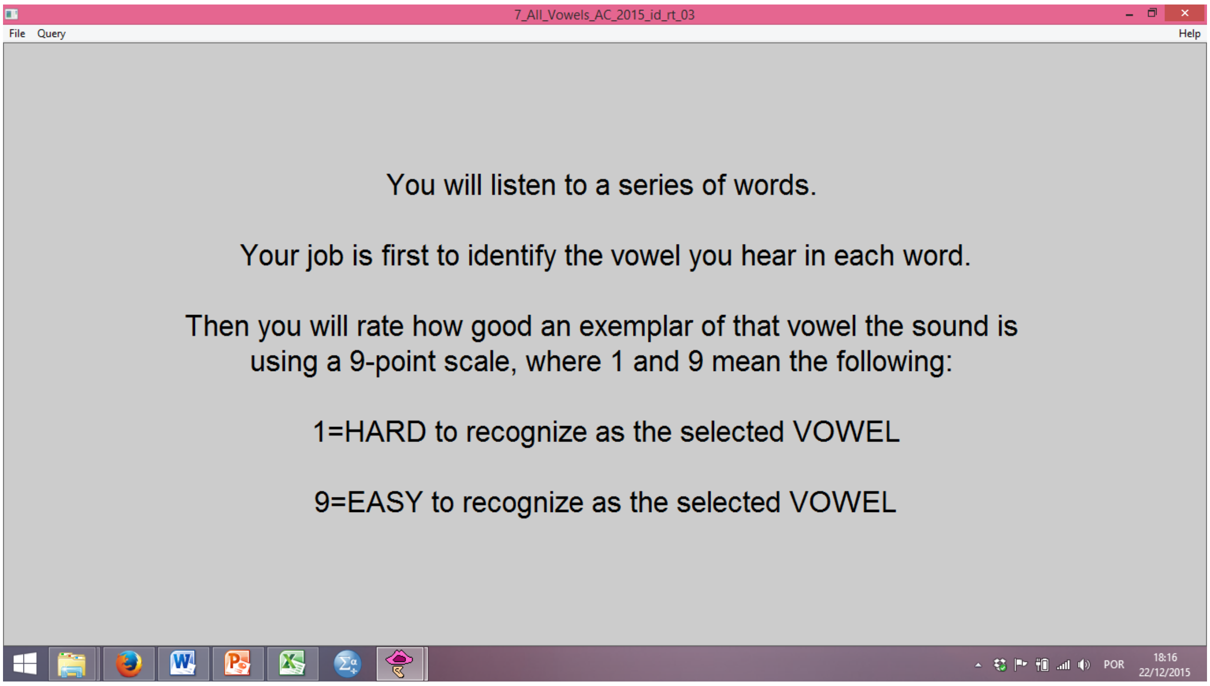


b) Response options and rating scale



3) Vowel rating test

a) Instructions for raters



b) Response options and rating scale